Validation algorithms in cryo-electron tomography

Federico P. de Isidro Gómez

Programa de Doctorado en Ingeniería Informática y de Telecomunicación

Centro Nacional de Biotecnología (CNB) - CSIC

Madrid, 2025

Agradecimientos

Percibo que existe un debate manido respecto a qué es lo importante del viaje, si el camino o el destino, cuando para mí se hace evidente que la respuesta es la compañía. Una vez alcanzado el destino con el camino ya recorrido me gustaría agradecer...

A Jose María por darme esta oportunidad y hacer posible este trabajo, y a Carlos por empezar a educar a un adolescente hace más de 10 años y no haber dejado de hacerlo. Ahora sé más cosas, aunque lo de adolescente no se me haya pasado del todo.

A mi recién adquirida familia en Cambridge y mi nuevo laboratorio, por hacer este solapamiento vital algo más tolerable y hacerme sentir en casa haya donde no es.

Al B13, camara de tortura y refugio ante la tempestad. Docenas de racimos de veces han sido las que me he sentido agradecido de teneros de compañía en este viaje. Sois la verdadera ganancia de esta etapa de mi vida y no lo que se recoge en estas páginas.

A Vilas por ser amigo y mentor, la injusticia de la vida ha hecho que tu nombre no aparezca en la portada, no puedo más que darte un reconocimiento de consolación en este lugar.

A Sara, por obligarme a hacer un ingente número de estupideces, todas en contra de mi voluntad, y de las que no me arrepiento.

A Patri, por ayudarme a entenderme y a mantener el récord de más vueltas paseadas alrededor del CNB.

A Joako, por sentarse conmigo en el borde del precipicio, sin decir nada, porque nada había que decir.

A Javi y Olayo, por el regalo de vivir con la certeza de la compañía independiente del lugar, el apoyo independiente de la dificultad y la entrega independiente del precio a pagar.

A mi padre por enseñarme que lo bien hecho bien parece y que quien algo quiere algo le cuesta.

A mi madre por poner ciencia en mi, enseñarme el poder de la tozudez y lo cobarde que hay en crear decisiones excluyentes allá donde nunca hubo por qué elegir.

A mi hermana, verdadero adalid científico de la familia, recordatorio diario de que todo se puede conseguir con la cabeza alta y la moral intacta.

Y en general, a todos los que me habéis acompañado, porque al final lo importante era todo lo que estaba sucediendo fuera del laboratorio.

"Como no estás experimentado en las cosas del mundo, todas las cosas que tienen algo de dificultad te parecen imposibles. Confía en el tiempo que suele dar dulces salidas a muchas amargas dificultades."

- Miguel de Cervantes

Muchas gracias.

Abstract

Cryo-electron tomography (cryo-ET) is a powerful technique that allows scientists to study the structure of biological samples at the nanometric scale. This technique enables the examination of samples in their native environment and broadens the range of specimens that can be studied, informing about the underlying processes within their biological context. To fully leverage the potential of this technique, it is necessary to use data processing methods capable of extracting relevant biological information from the raw microscope images. However, these methods can introduce errors that degrade the data and hinder the possibility of obtaining accurate biological conclusions.

This dissertation targets the validation of cryo-ET data processing. To achieve this, we propose a series of automatic methods designed to ensure the quality and robustness of the results obtained by each processing method.

The conducted research begins with a general overview of the data processing pipeline, proposing validation solutions for each step. This general approach ensures the overall quality of the processing and facilitates the detection of critical steps. Thus, the subsequent study focuses on the alignment of tilt series, due to its instability and its core role regardless of the processing strategy being employed.

For this purpose, two standalone methods are proposed, capable of assessing the quality of the alignment. The first method, based on deep neural networks, allows for the detection of alignment errors in tomographic reconstructions. To overcome a series of limitations detected during the development of this work, a second method has been developed. This method performs the misalignment detection directly on the tilt series.

The set of methods presented in this work provides scientists with the necessary tools for better data processing, ensuring the biological conclusions drawn from the sample under study and maximizing the utilization of the data obtained.

Resumen

La criotomografía electrónica (cryo-ET, por sus siglas en inglés) es una técnica poderosa que permite a los científicos estudiar la estructura de muestras biológicas a escala nanométrica. Esta técnica permite el estudio de muestras en su entorno nativo y amplía el rango de especímenes de estudio, informando sobre los procesos subyacentes dentro de su contexto biológico. Para aprovechar al máximo el potencial de esta técnica, es necesario utilizar métodos de procesamiento de datos capaces de extraer información biológica relevante de las imágenes crudas del microscopio. Sin embargo, estos métodos pueden introducir errores que degradan los datos y dificultan la posibilidad de obtener conclusiones biológicas precisas.

Esta tesis se enfoca en la validación del procesamiento de datos de cryo-ET. Para lograr esto, proponemos una serie de métodos automáticos diseñados para asegurar la calidad y robustez de los resultados obtenidos por cada método de procesamiento.

La investigación realizada comienza con una visión general del flujo de procesamiento de datos, proponiendo soluciones de validación para cada paso. Este enfoque general asegura la calidad global del procesamiento y facilita la detección de pasos críticos. Así, el estudio a continuación se enfoca en el alineamiento de series de tilt, debido a su inestabilidad y su papel fundamental independientemente de la estrategia de procesamiento empleada.

Para este propósito, se proponen dos métodos independientes, capaces de evaluar la calidad del alineamiento. El primer método, basado en redes neuronales profundas, permite la detección de errores de alineación en reconstrucciones tomográficas. Para superar una serie de limitaciones detectadas durante el desarrollo de este trabajo, se ha desarrollado un segundo método. Este método realiza la detección de desalineación directamente en las series de tilt.

El conjunto de métodos presentados en este trabajo proporciona a los científicos las herramientas necesarias para un mejor procesamiento de datos, asegurando las conclusiones biológicas extraídas de la muestra de estudio y maximizando la utilización de los datos obtenidos.

Contents

Chapter 1

Introduction

The field of structural biology seeks to elucidate the three-dimensional structure of macromolecules. Understanding how their spatial arrangement influences biological processes, functional mechanisms, and molecular interactions is essential for revealing their roles in cellular function. This has a great impact beyond molecular biology. The knowledge obtained in this field has led to significant improvements in several areas, from deciphering the pathogenesis of diseases to developing targeted therapeutics and drug discovery Robertson et al. (2020); Van Drie and Tong (2022).

This discipline has experienced massive development in recent decades due to different imaging techniques and software developments. From advances in X-ray crystallography in the decade of 1990s to the most recent developments in prediction models for protein structures, including developments in nuclear magnetic resonance, these new tool sets have allowed scientists to elucidate molecular complexes to near-atomic resolution. Among all these developments, cryo-electron microscopy (cryo-EM) plays a particular role, addressing limitations imposed by previous techniques and becoming the most widely used technique in structural biology today.

Cryo-EM is a multidisciplinary technique that requires collaboration in various scientific fields for practical use. The process begins with preparing the sample and purifying it through several biochemical processes. Then, the sample is loaded onto a grid and rapidly frozen to preserve its natural state in a volume of vitreous ice in a process called vitrification. A schematic of this process is shown in Figure 1.1.

The sample is radiated posteriorly using an electron microscope. A beam of electrons illuminates the sample, capturing structural information and creating thousands of images. The acquired images are produced in transmission and are focused on the sample using magnetic lenses. Later, another set of magnetic lenses projects the image onto a detector. A schematic of the microscope's structure is provided in Figure 1.2.

The current generation of detectors can acquire images at a high-speed rate, as each

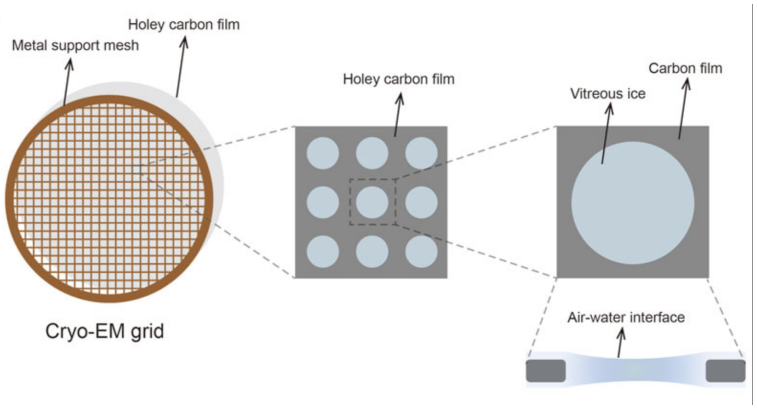


Figure 1.1: Simplified schematic of the structure of a cryo-EM grid with a loaded sample. Figure modified from Xu and Dang (2022).

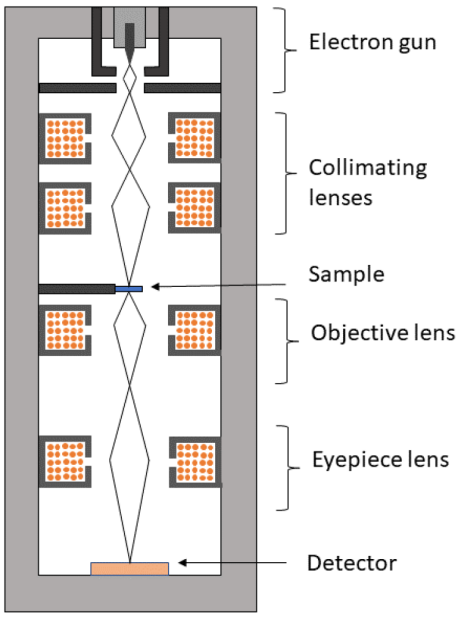


Figure 1.2: Simplified schematic of the structure of an electron microscope detailing each component.

acquired image contains the isolated impacts of electronsMcMullan et al. (2016). These images are known as frames. During the exposure time, multiple frames are captured, forming what is referred to as movies, which contain the raw data of the sample images and serve as the starting point for image processing. The image acquisition process is exemplified in Figure 1.3.

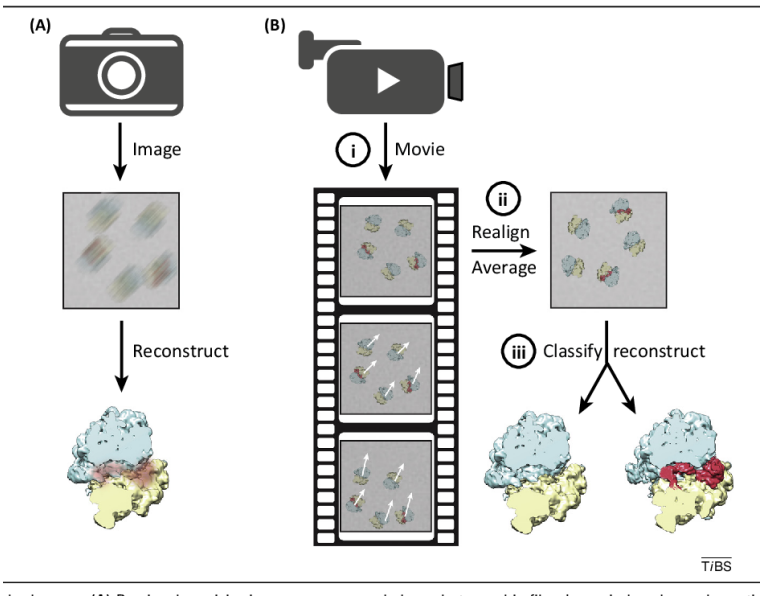


Figure 1.3: Comparison between different acquisition modes. (A) Single image (micrograph) recording mode. (B) Movie recording mode, resulting in a set of frames. Figure referenced from Bai et al. (2013).

Radiation damage restricts the number of images that can be acquired and reduces their contrast. Consequently, the images exhibit an extremely low signal-to-noise ratio (SNR) and contrast, significantly impacting the high-frequency details. To address these issues, dose-dependent low-pass filters are commonly applied to eliminate unreliable high frequencies Grant and Grigorieff (2015).

The goal of the image processing workflow is to use these images to generate a 3D representation of the specimen. To achieve this, advanced computational methods are applied to process the raw data obtained from the microscope and determine the three-dimensional structure of the molecule under study. The increase in microscopes' throughput and the stability of image processing methods in recent years has led to rapid, high-quality elucidation of structures, with particular interest in automating these methods to reduce human intervention.

Cryo-EM of purified samples has significantly advanced our understanding of molecular structures, but it imposes a set of limitations, from the requirement of a previous purification of the sample to the assumption that all the elements from the sample belong to the same specimen. Building on these advancements, cryo-electron tomography (cryo-

ET) offers a complementary approach by providing three-dimensional reconstructions of cellular structures in their native context.

Cryo-ET enables studying a broader range of samples, allowing for examining complexes in their native environments. In addition to purified samples, this technique can study entire cells, tissues, and large cellular assemblies. Although the broader field of view required for these samples compromises resolution, it provides valuable information about their native biological context, such as the organization of macromolecules within cells, which is crucial for answering biological questions.

Due to this new paradigm, some differences arise compared to classical cryo-EM. Although sample preparation remains essentially unchanged, with plunge freezing and high-pressure freezing being the most popular vitrification approaches, a new limiting factor to consider is ice thickness Navarro (2022). Due to the increased range of diversity of the sample, some specimens might grow in thickness up to several hundred of nanometers, and in those cases, the thickness must be reduced Lučić et al. (2013). The most popular approach is applying FIB-milling to the sample, a sputtering technique that removes material from the sample to produce a thin lamella.

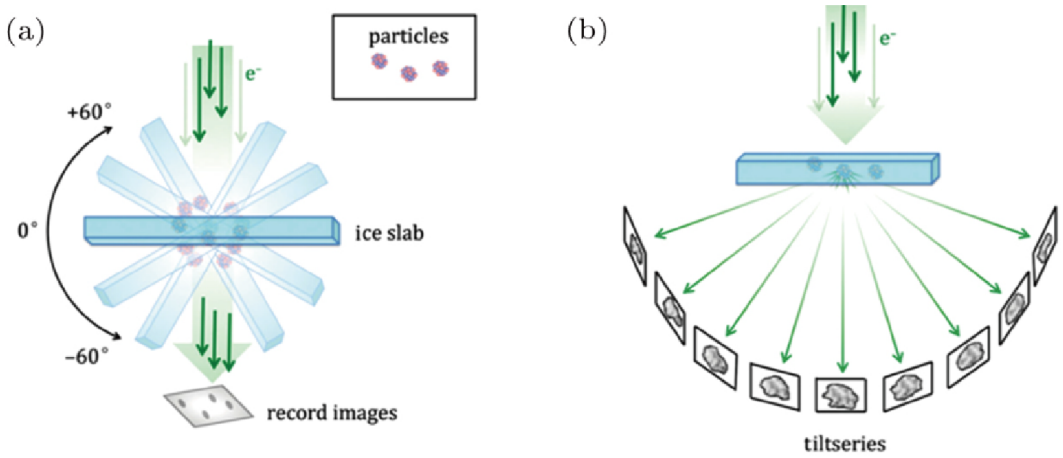


Figure 1.4: Schematic of the tomographic acquisition. (a) Tilting of the sample stage for image recording. (b) The collection of the projective images composes the tilt series. Figure modified from Galaz-Montoya and Ludtke (2017).

The core difference in cryo-ET consists of the tilting of the sample under the microscope, which allows for different views of the same specimen, exemplified in Figure 1.4. This breaks one of the assumptions made in cryo-EM, where all the particles are presumed to be equivalent copies of the same macromolecule. Thus, in cryo-ET, since projective information of the same sample is obtained in different directions, its structure can be directly reconstructed. However, two assumptions still apply to the cryo-ET image formation process Vulović et al. (2014).

1. The projection assumption (PA): where all the images obtained from the sample are assumed to be obtained along the beam direction. Thus, the tilting process produces a series of projections of the sample from various directions, determined by the tilt angles.

2. Weak Phase Object approximation (WPOA): where the electrons are so energetic that only one single scattering event is assumed to happen, being its interaction with the sample mostly elastic without any loss of energy. This approximation has profound implications in the imaging process, where the obtained images are produced due to a phase affection of the radiated electrons Unwin et al. (1971), leading to extremely low contrast. To mitigate these effects, the sample is purposely placed out of focus, but the tradeoff is that the microscope's Contrast Transfer Function (CTF) must be explicitly considered in the image processing. Another option to alleviate this effect is using phase plates Danev and Baumeister (2017).

A notable challenge in this field is that the sample is not radiated in all directions due to mechanical and dose limitations, and the tilt step is limited. Thus, the maximum tilt of the sample is limited to around 60°, with a typical angle step of typically 2-3°. These constraints cause artifacts reconstructions due to the missing wedge in the Fourier space and the limited tilting step. The effects of these two phenomena are exemplified in Figure 1.5.

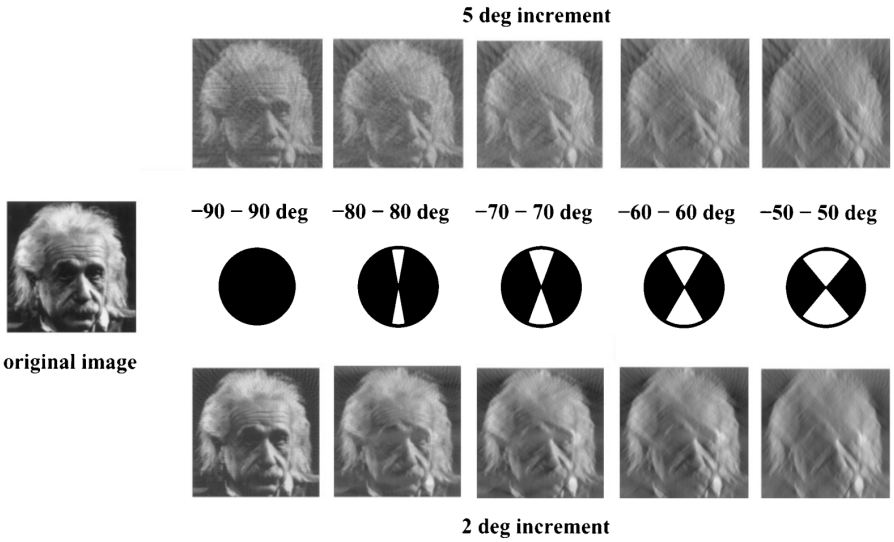


Figure 1.5: Illustration of the missing wedge and tilt step effects. The deteriorating influence of the missing gap depends on the tilt range and tilt increment. Figure referenced from Koster et al. (1997).

Another limitation is the dose accumulation during acquisition, as the interaction of electrons with the sample damages it. Radiation damage increases with each projective

image taken, making the order of image acquisition an important aspect of the workflow. Various acquisition schemes have been analyzed to improve the quality of the reconstruction step, combining different doses and acquisition geometries Turonova et al. (2020). Three of these schemes are shown in Figure 1.6.

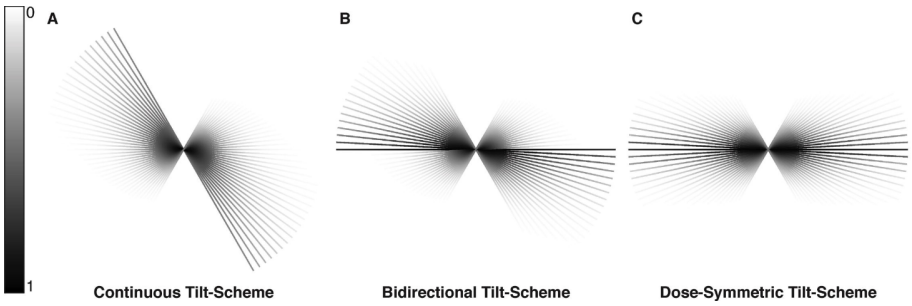


Figure 1.6: Examples of different acquisition schemes. The intensity is proportional to the accumulated dose (and, therefore, to the acquisition order). Figure modified from Hagen et al. (2017).

The new image acquisition schemes that tilt the sample, along with the broader range of biological specimens that can be studied, lead to the development of a new set of image processing tools tailored to the unique nature of these images and their specific constraints. Unlike classical cryo-EM, the image processing pipeline is not linear. Instead, different tools will be employed depending on the sample's characteristics and the specific biological questions the scientist poses. However, the first image processing steps are commonly independent of the final application or the specimen under study that yields from the raw images of the microscope to the final tomographic reconstruction. Subsequently, the initial steps of the workflow are outlined below.

1. Tilt series movie alignment: As in classical cryo-EM, a raw set of images is obtained from the microscope at each tilt angle, called movies. These sets of stacks of images are known as tilt series movies. The sets of movies are aligned and averaged in a single image to increase the SNR and correct for motion and deformations of the sample due to the electron beam Brilot et al. (2012), obtaining the final tilt series composed of one image per tilt angle. Since this is an equivalent problem to the one in classical cryo-EM, the tools employed to solve it are inherited from this field Li et al. (2013); Scheres (2014); Grant and Grigorieff (2015); Abrishami et al. (2015); Zheng et al. (2017); Strelák et al. (2020).

2. Tilt series preprocessing: Although the flexibility inherent in cryo-ET allows for some steps to be replaced at different stages of the workflow, there are a set of operations that, while not mandatory, are typically applied to the tilt series before calculating the three-dimensional reconstruction.

(a) X-ray filter: Due to the interaction of electrons with the sample, particularly significant for heavy atoms, X-ray photons may be generated. When these photons interact with the detector, they can produce an artifact in the value of some pixels in the image. These X-ray photons can be generated from two different interactions. First, an electron may interact with an inner shell electron of a sample atom, promoting an outer electron to this new energy level, releasing an X-ray photon. Secondly, if the electron is deflected by an interaction with a nucleus of the sample, the energy loss in this process is released as an X-ray photon. A schematic of the two sources of this phenomena is provided in Figure 1.7.

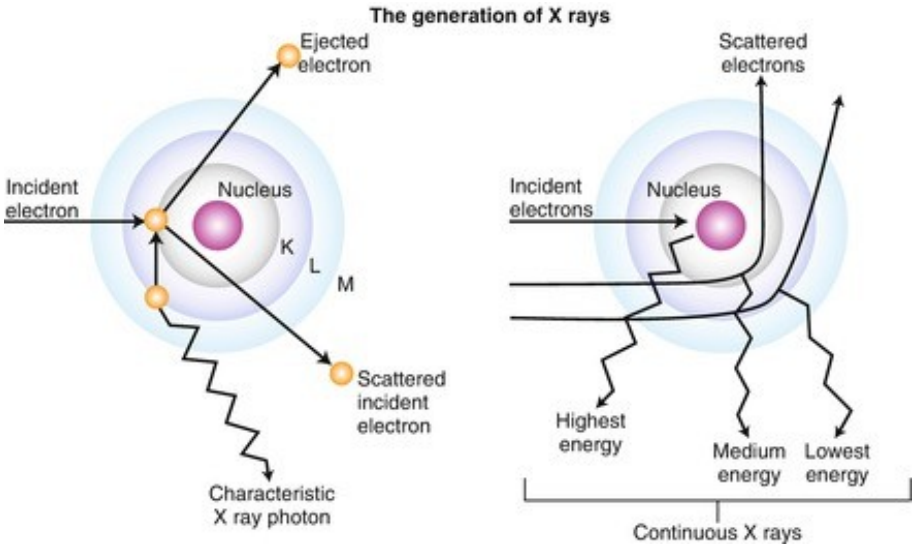


Figure 1.7: Graphical explanation of X-ray photon sources in cryo-ET acquisitions. The left diagram illustrates the formation of X-ray photons due to the promotion of an outer electron. The right diagram depicts the formation of X-ray photons resulting from the interaction of an electron with the nuclei. Figure referenced from Pelberg and Pelberg (2015)

(b) Dose filter: As in classical cryo-EM, images are corrected based on the accumulated dose to avoid artifacts from radiation damage to the sample. In the case of cryo-ET, each tilt image is corrected individually, considering the acquisition order, since the sample accumulates radiation with each angular sampling.

(c) Resampling: It is common practice in cryo-ET to rescale images before further processing, analogous to the resizing of particles typically done in cryo-EM. This is because the subsequent steps can be computationally intensive, and aligning resampled data can save resources without significantly affecting the results. Moreover, this step is especially beneficial if the scientist is not interested in a particularly high resolution. This is usually performed by the binning

operation, averaging a squared number of pixels, which is computationally less expensive and avoids interpolations. A schematic of the effect of the binning operation is shown in Figure 1.8.

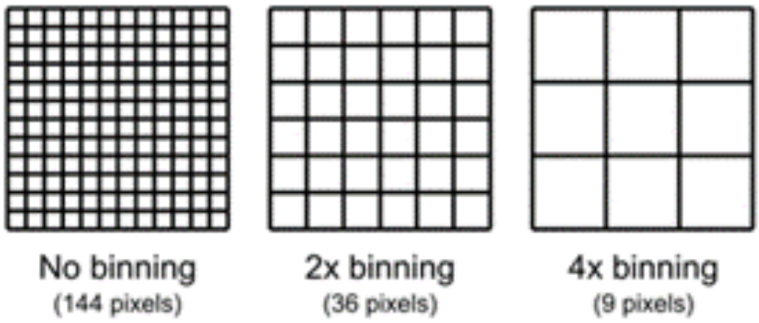


Figure 1.8: Schematic of the binning operation. The same image is rescaled to 3 different sizes.

3. CTF estimation: The raw images obtained from the microscope are affected by defocus and other microscope aberrations, modeled by the CTF in the frequency domain Kirkland (2010). An accurate determination of the CTF allows the proper characterization of the aberrations of the microscope, enabling its posterior correction Fernández et al. (2006); Voortman et al. (2011); Su (2019). The geometry of the acquisition imposes a set of contrains that might be exploited and considered by the algorithms Fernández et al. (2006); Xiong et al. (2009). However, some of them are directly inherited from classical cryo-EM.

   The sample tilting imposes a defocus gradient perpendicular to the tilt axis, with the gradient becoming more pronounced at higher tilt angles. Accounting for this gradient allows for a more accurate estimation of the CTF, rather than simplifying the defocus to a single averaged two-dimensional value Vilas et al. (2022); Rohou and Grigorieff (2015); Zhang (2016). This gradient in the defocus is considered in the model by factoring in that parallel stripes to the tilt axis in the image have similar defocus values and can be estimated independently.

4. Tilt series alignment: Before calculating the tridimensional reconstruction of the sample, it is necessary to reference all the images from the series to a common geometry. Ideally, the images stay consistent with the acquisition geometry, but in practice, it is necessary to solve the movements (shifts and rotations) the sample experiences in the acquisition process. These movements lead to unreliable artifact reconstructions unsuitable for further sample analysis. In the field, there are different mathematical methods to solve acquisition geometry, although the goal of them is to calculate the tilt axis and the sample displacements Messaoudi et al. (2007);

Heymann et al. (2008); Castano-Diez et al. (2010); Mastronarde and Held (2017); Fernandez et al. (2018); Sorzano et al. (2020); Fernandez and Li (2021); Seifer and Elbaum (2022); Zheng et al. (2022).

5. Tomographic reconstruction: Once the acquisition geometry is solved, it is possible to calculate the tridimensional reconstruction. The literature includes various reconstruction methods, as it is a classic problem in the image processing field and not restricted to cryo-ET. Despite the large number of algorithms, they can be classified into two families. First, Fourier Space and series expansion, where Weighted Back Projection (WBP) Radermacher (1992) is the standard in the field; and second, iterative algorithms Sorzano et al. (2017b), with notable algorithms such as ART Gordon et al. (1970), SIRT Gilbert (1972) or SART Andersen and Kak (1984). Generally, reconstruction programs offer the flexibility to select the reconstruction algorithm Kremer et al. (1996); Agulleiro and Fernandez (2010, 2015); Zheng et al. (2022).

   An interesting application exclusive to cryo-ET is the inclusion of CTF correction in the reconstruction process Turoňová et al. (2017), in contrast to the rest of the algorithms where it is assumed that CTF is corrected over the tilt images. Thus, this approach combines geometry and the CTF information to obtain a more accurate correction.

6. Noise reduction: Tomographic reconstructions are affected by a very low signal-to-noise ratio (SNR). To address this, algorithms can be applied to modify the reconstruction, preserving structural information while reducing noise. Incorporating this step into the workflow enhances the detection of structures and particles, thereby improving the understanding of the sample and enhancing the ability to draw biological conclusions. Denoising algorithms attempt to solve this problem. There are three prominent families of denoising algorithms Frangakis (2021): real-space algorithms, including classical image processing techniques such as Gaussian blur and mean or median filters, as well as more sophisticated ones like non-anisotropic diffusion Frangakis and Hegerl (2001); Moreno et al. (2018); transform-based filtering methods employing thresholding in the wavelet domain as a denoiser Huang et al. (2018); and neural networks-based Buchholz et al. (2019a,b); Bepler et al. (2020).

7. Reconstruction quality assessment: Another optional step in the workflow is the reliability quantification of the tomographic reconstruction. In classical cryo-EM SPA, Fourier Shell Correlation (FSC) is the most widely used metric for this assessment Harauz and van Heel (1986); Sorzano et al. (2017a); Cardone et al. (2013). Although its application to tomography is not straightforward, an equivalent metric

can be obtained by calculating two independent reconstructions from split frames of the tilt series movie Cardone et al. (2005). The resolution for which the FSC=0.143 is called the resolution of the map Rosenthal and Henderson (2003). It was shown that globalTan et al. (2017); Vilas and Tagare (2023) and localVilas et al. (2020c) resolutions are a directional magnitude. Additionally, anisotropy analysis of the reconstruction can be performed using conical FSC Diebolder et al. (2015). Other quality assessments, such as Noise-compensated Leave One Out (NLOO), compare several reconstructions of the same sample by removing one tilt image at a time, though this method is computationally expensive. Also, other metrics inform of the quality of the map as a measure of the local SNR Kucukelbir et al. (2014); Vilas et al. (2018), or in the map texture Ramirez-Aportela et al. (2019). Specific tools for the tomography paradigm have also been developed. For instance, MonoTomo Vilas et al. (2020b) estimates local resolution by performing hypothesis tests at multiple frequencies to determine if local amplitudes are greater or smaller than local noise amplitudes.

A graphical representation of these first steps of the workflow is provided in Figure 1.9.

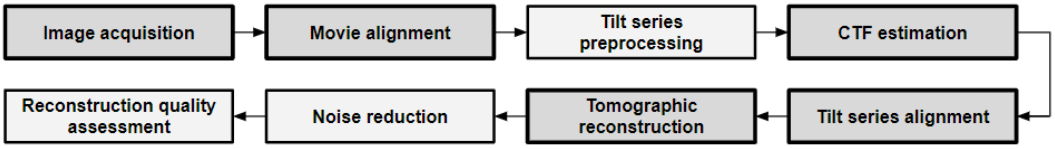


Figure 1.9: Basic tomogram reconstruction workflow. Core steps are highlighted in dark.

New developments in the field aim to automate these initial steps, eliminating the need for human intervention. This requires these steps to be robust, stable, and less prone to errors, especially considering they are common to any subsequent application. Consequently, developing new methods and algorithms that ensure the quality of these results is a relevant area to focus future work on, with a significant impact in the field. In this work, several developments aimed in this direction are presented.

Once the tomographic reconstruction is obtained, the nature of the sample will determine the subsequent steps. The scope of this technique encompasses a wide range of applications, from cell biologists interested in the distribution of structures within the cell and its landscape to structural biologists focused on high-resolution molecular structures. Among these applications, two main directions of image processing emerge in the field: cellular biology and structural biology. This second group is further divided into subtomogram averaging (STA) and per-particle-per-tilt (pPpT) analysis. A graphical representation of the alternative workflows is provided in Figure 1.11.

1. Cellular biology: This approach aims to better understand the cellular landscape and the biological environment. The resolution at which these samples are acquired is not typically that high, reducing the magnification searching for a wider field of view. The tools will aim to detect biological structures in the sample, exemplified in Figure 1.10.

   (a) Segmentators: Although previous quality analysis and denoising steps aimed to enhance understanding, the tomograms' low signal-to-noise ratio (SNR) complicates result interpretation. Segmentation algorithms aim to detect biological structures in tomographic reconstructions to address this. Historically, segmentation algorithms were largely manual. However, the growing volume of data has driven substantial progress in developing automated methods, with several algorithmic foundations from tensor voting Martinez-Sanchez et al. (2014) to deep neural networks de Teresa et al. (2022); Chen et al. (2017). The subsequent step involves annotators, whose purpose is to label the segmented structures, providing a biological identification for them as membranes, ribosomes, membrane proteins, filaments, or microtubules Xu et al. (2017); Lamm et al. (2022a); Dimchev et al. (2021).

   (b) Spatial distribution: Algorithms aimed to detect elements of interest and analyze their arrangement within the sample. This may include assessing the average concentration of a specimen across different regions of the sample or measuring the average distance of particles from a specific structure Zeng et al. (2023); Martinez-Sanchez et al. (2022), or analyzing the organization of biological structures Sazzed et al. (2023). Instead of studying the distribution of a specimen in the sample, other algorithms focus on the shape and size of structures in the samples, such as measuring the size and aspect ratio of vesicles or the thickness of cell envelopes Salfer et al. (2020); Frangakis (2022).

2. Structural biology: The aim of this approach is the elucidation of detailed biological structures. Here, the biological environment is receded into the background, and the inclusion of high-frequency information in the final reconstruction becomes the priority.

   The first step is the location of the structure of interest in the tomographic reconstruction, a process called picking. This is a sensitive process, as misidentifying particles can introduce incorrect information into the final reconstruction (false positives) or omit crucial details needed for high-resolution analysis (false negatives). These challenges are further exacerbated by a low SNR. Two categories of picking algorithms coexist in the field: nondirectional pickers that solve the three-dimensional

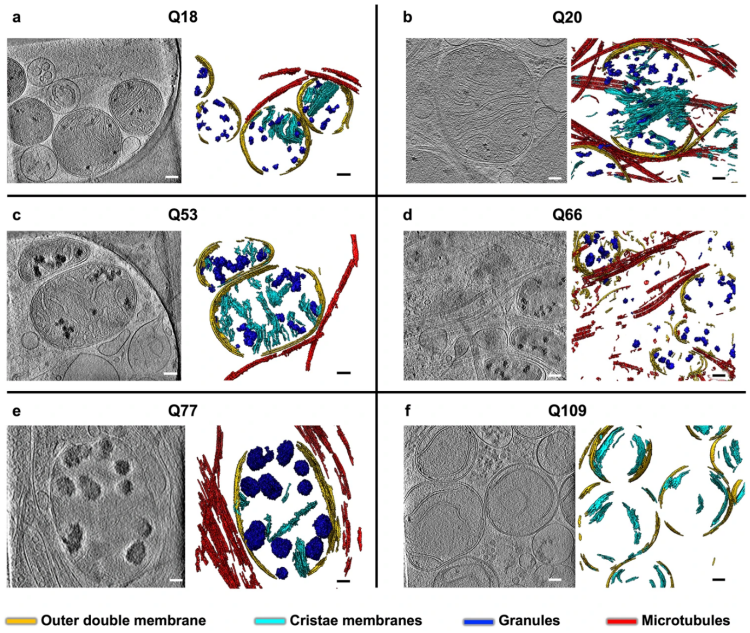


Figure 1.10: Segmentation and spatial contextualization of several biological structures from different samples. Image referenced from Wu et al. (2023).

location of the particle in the sample Wagner et al. (2019); de Teresa et al. (2022); and directional pickers, which also inform about the direction of the particle with respect to a structure (typically a mesh or surface representing some biological structure, such as a cellular membrane) Castano-Diez et al. (2017); Martinez-Sanchez et al. (2020); Lamm et al. (2022b).

Once the particles are located in the tomographic reconstruction, their information is combined in a single reconstruction to elucidate high-resolution details. Historically, only one approach, known as subtomogram averaging, existed in the field. However, in recent years, the per-particle-per-tilt approach has been introduced.

(a) Subtomogram averaging: The information from the subtomograms (three-dimensional subvolumes extracted directly from the tomographic reconstruction) is combined into a single reconstruction through alignment, classification, and averaging. The subtomograms are oriented to a common reference and subsequently classified into homogeneous groups representing the same protein in a specific conformation. Once these steps are completed, it is possible to average the information of all the subtomograms belonging to the same group after interpolating the calculated alignment. The averaging preserves the information while reducing the noise in the final volume. This step is very sensitive to the missing wedge, which affects both the tomographic reconstruction and the extracted subtomograms. Wan and Briggs (2016); Scaramuzza and Castaño-Díez

(2021); Bharat and Scheres (2016); Himes and Zhang (2018).

(b) Per-particle-per-tilt: This approach lies in performing the alignment, classification, and averaging steps on the tilt series particles (or sub-tilt series). These are two-dimensional images extracted from the tilt series at the locations where the particle of interest is projected at each tilt image. Equivalently to the STA approach, the tilt series particles are aligned to a common reference, classified into homogeneous groups, and posterior averaged, ideally taking advantage of the constraint imposed between sub-particles belonging to the same sub-tilt series. This approach has the advantage of eliminating the missing wedge bias present in STA, although at the cost of increasing the degrees of freedom. Galaz-Montoya et al. (2015, 2016); Zivanov et al. (2022); Tegunov et al. (2021).

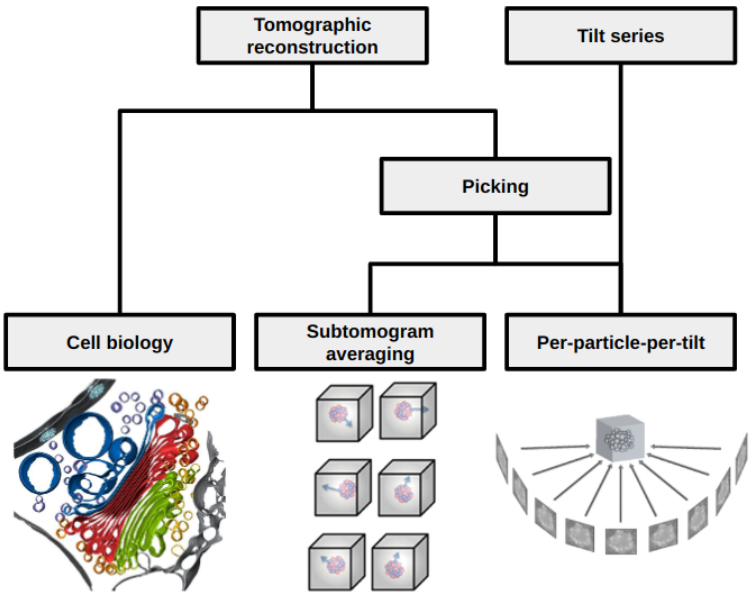


Figure 1.11: Schematic of the alternative applications once the tomogram reconstruction workflow is computed.

This classification is not a strict dichotomy in the field, and a combined approach aims to obtain both structural information about the specimen and its context within the sample. However, there are evident limitations in resolution and magnification. Obtaining high-resolution details of the structure requires a higher magnification, reducing the field of view, resulting in the loss of cellular context information. However, reducing the magnification increases the contextual information of the sample but makes the details in the final reconstruction inaccessible.

1.1 State of the art

The work presented in this dissertation introduces the development of algorithms for validating the processing of electron tomography data, with particular attention to the alignment of tilt series. The alignment of projective images for subsequent reconstruction is a classic problem in image processing. This was introduced in the previous section, being necessary to refer all the images that belong to the same reconstruction to a common reference before calculating the tomographic reconstruction. The alignment algorithm calculates the geometric transformations required for each image. However, the limited dose and consequent low SNR, combined with a specific acquisition scheme, introduce unique challenges in electron tomography.

From these limitations, the low SNR of the images results in the geometry being solved as an unstable problem. Due to this, it is a common practice in the field to add fiducial markers to the sample. These markers are small gold bead particles that, due to the high atomic number of this element compared to the elements of the sample, appear in the images as consistent, highly contrasted elements that are easily tracked through the tilt series. Thus, these methods can be classified based on the nature of the sample for which they were developed, being very popular to rely on the presence of fiducials. There are several developments in the field that approach this problem by considering either one or both approaches, with a more extensive set of algorithms focused on samples with fiducials.

1. Fiducial-based approaches: The first step for algorithms that rely on fiducials in the sample is their detection in each tilt image. Different preprocessing strategies might be applied to each tilt image depending on the approach followed by each algorithm. Some developments in the field target this problem by searching for the most robust detection possible Hou et al. (2024); Buckley et al. (2024). This is a sensitive step in the alignment procedure, since any false positive (confusing some structure in the sample with a fiducial) or negative (missing the location of some fiducial in the sample) spoils the subsequent calculation of the geometrical transformations.

   Once the fiducials are detected, it is possible to relate the correspondence of each fiducial in each image, defining the so-called fiducial or landmark chains. These constructions relate the position of the same fiducial in every tilt image, composing a set of common points among all the images. It is standard in the field that algorithms do not require complete chains to provide the location for the same fiducial in every image to calculate the alignment transformation. On the contrary, a set of incomplete fiducial chains is suitable for a correct alignment calculation if the coverage of the tilt series is sufficient. At this stage, it is common practice

to perform postprocessing strategies to ensure the quality of the chains. Examples include guaranteeing a minimum correlation for the fiducials belonging to the same chain, merging partial chains, or pruning some of them to avoid overrepresenting some sample areas.

With the relative positions for every fiducial determined, it is possible to calculate the relative geometrical transformations for every image before obtaining the tomographic reconstruction. The set of common points for each tilt image produces a set of relationships that enables the definition of an equation system that relates the three-dimensional coordinates in the sample with their associated coordinates in each tilt image through the projective transformation. This system is solved by minimizing the reprojection error, which is the difference between the detected fiducial coordinate in a tilt image and the projection of the calculated three-dimensional location of the fiducial, thereby solving the alignment for every tilt image. A schematic of the alignment process is shown in Figure 1.12.

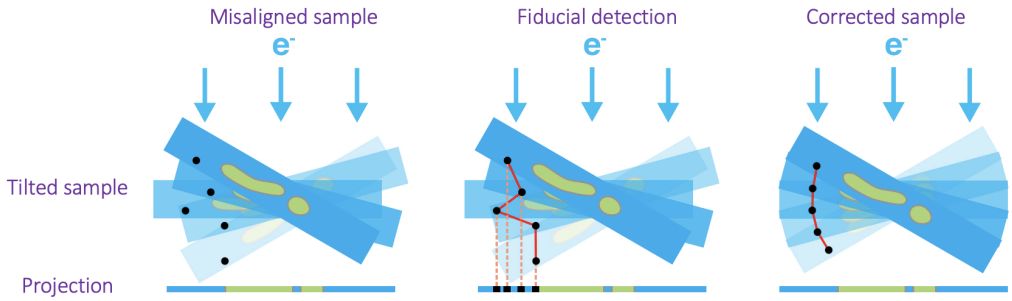


Figure 1.12: Schematic depicting the fiducial alignment process. Image referenced from Castaño-Díez et al. (2012).

2. Fiducialless-based approaches: This second group of algorithms addresses the alignment of those samples to which fiducials have not been included in the preparation process. Two main approaches to address this problem coexist in the field:

   (a) Patch-tracking alignment: These algorithms compare regions (patches) between the tilt images, usually by correlation. The number and size of patches and an accurate prior estimation of the tilt axis angle are sensitive parameters affecting the quality of results. Additionally, the preprocessing of the images, including filtering and accounting for the rotation of the sample, is crucial. Knowing the relative orientation of the different patches throughout the tilt series, it is possible to calculate the geometric transformations that relate all the series to a common reference.

(b) Reprojection Alignment: These algorithms calculate an initial reconstruction to compare its projection with the tilt series subsequently. Thus, the difference between the reprojection and the tilt image is used to refine the alignment in an iterative process until this difference is negligible. As in the previous case, an accurate prior estimation of the tilt axis angle is a critical step for correct alignment. Due to the design of these algorithms, after the subsequent estimation of the alignment with respect to a common geometry, the three-dimensional reconstruction is also calculated.

Both approaches might account for different effects on the images during acquisition, ranging from a simple linear transformation (in-plane rotation and shifting of the images) to refinement of the tilt angle, beam-induced motion, imposing a deformation and distortion fields, and refining a local alignment. Obviously, the greater the number of effects considered, the more accurate the solution to the problem. However, as a drawback, the system becomes more complex and increases the risk of converging to local minima, resulting in incorrect solutions to the alignment problem.

Although it is falling into disuse in the field, there is a particular type of acquisition known as dual-axis acquisition. Figure 1.13 provides a schematic of these two acquisition modes. In this case, the sample is not rotated with respect to a single tilt axis but rather to two perpendicular axes, obtaining two 'subtilt series' of the same sample. This has the advantage that, during reconstruction, there is a greater number of views of the sample, resulting in a more complete reconstruction space and reducing the number of artifacts.

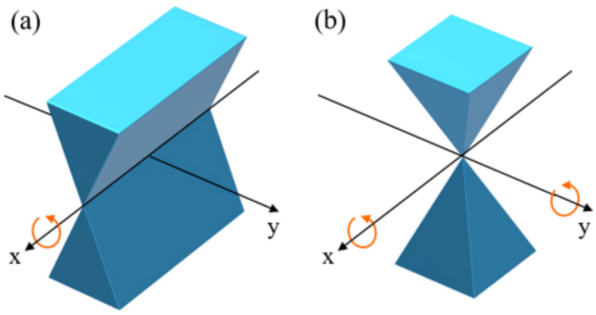


Figure 1.13: Comparison between different acquisition geometries. (a) Shapes of Missing Wedge for a uniaxial-tilt specimen holder. (b) Shapes of Missing Wedge for a double-tilt specimen holder.

Obviously, it presents limitations concerning the dose that the sample can withstand. Both approaches to the alignment problem (fiducial and fiducialless) are valid for this type of acquisition, although the acquisition geometry of each 'subtilt series' must be considered individually.

1.2 Objectives

In this dissertation, presented as a compendium of articles, a study of the cryo-electron tomography pipeline is addressed, seeking validation methods that enhance the stability and performance of current methods in the field. This corresponds to the first of the three publications in the compendium. Subsequently, the development of specific algorithms able to validate the quality of the alignment of a tilt series is studied in depth, leading to two new algorithms for this purpose, corresponding to the other two publications in the compendium. Thus, we define three main objectives for the achievement of the project:

1. To analyze each step of the cryo-electron tomography processing pipeline in detail, detecting instabilities in the workflow requiring to develop algorithms as well as validation strategies to ensure the robustness of the obtained results.

2. To define the mathematical basis of the tilt series alignment problem, paying particular attention to fiducial-based approaches.

3. To develop a set of new algorithms to validate tilt series alignment at different stages of the pipeline. These algorithms will provide an unsupervised automatic detection of alignment errors in cryo-electron tomography acquisitions.

All the new algorithms developed have been created as part of the Xmipp de la Rosa-Trevín et al. (2013); Strelak et al. (2021) software package and integrated into the Scipion framework de la Rosa-Trevin et al. (2016) within its tomography environment ScipionTomo Jimenez de la Morena et al. (2022), as well as other validation tools included in this dissertation, ensuring they are available and open to the cryo-EM community.

Chapter 2

Materials and methods

This section focuses on the works that constitute the research line encompassed by this dissertation, which includes three publications: two scientific articles and one book chapter. Additional complementary co-authored publications will be referenced to provide completeness and perspective on the published work, although they are not part of the main compendium. The theoretical and mathematical foundations are introduced, followed by a detailed exposition of each presented publication and the obtained results.

2.1 Theoretical foundations

2.1.1 Parameter estimation

Cryo-electron tomography is characterized by an extremely low signal-to-noise ratio (SNR), with typical values of 0.001 for the raw experimental microscope images. From a mathematical perspective, the aim of the image processing pipeline presented in the introduction section is to estimate a set of parameters that enables an optimal combination of the information contained in the raw images obtained by the microscope. The low SNR leads to the misestimation of these parameters, which can be decomposed into a combination of variance and bias in these estimations Sorzano et al. (2022), as formulated in Equation 2.1:

$$P_e = P_i + \Delta P, \tag{2.1}$$

where P_e is the estimated parameter and P_i the ideal one. ΔP accounts for the estimation errors that deviate from the ideal solution, defining variance as $\mathrm{Cov}(\Delta P)$, accounting for random errors with zero mean, and bias as $\mathrm{E}(\Delta P)$, accounting for systematic errors. Obviously, the value of ΔP is unkwown, as well as the affection of the measure by any bios or variance. Therefore, it is necessary to develop strategies for the robust detection

of imprecise estimations of the estimated parameters. The approaches to the validation of the estimated parameters can be classified into two different strategies:

- Consensus Approaches: These are based on the comparison of the results obtained from different estimation algorithms. This method is particularly effective if the source of errors in the estimation presents a significant covariance. However, if the estimations exhibit some bias, the different methods employed for the estimation should not fall into the same error. Thus, consistent estimations of the problem are assessed as correct, neglecting the possible case that two different algorithms fall into the same incorrect solution.

  The parallel execution of several algorithms (or multiple executions of the same algorithm in case it is non-deterministic) is not always feasible due to design constraints, computational load, or execution time. Therefore, validation strategies for the estimations obtained by single algorithms should be also considered.

- Standalone Approaches: Instead of comparing methods, a single estimation of the parameters is analyzed on its own. Thus, with a single estimation, it is possible to validate the quality of the results. This approach has the advantage of a lighter computational load since the parameters are estimated only once. Nonetheless, there are a series of trade-offs depending on the approach taken to this problem:

  1. Filtering Approaches: One approach consists of ensuring that the estimated parameters fall within a range that is thresholded by the scientist. The restrictiveness of the threshold is determined by the scientist, but this decision introduces a bias that affects subsequent processing and results
  2. Validation Algorithms: Another approach is based on the development of new algorithms that validate the calculated parameters. The underlying mathematical foundation of these algorithms must differ from that of the estimation algorithm to ensure that it does not fall into the same local minima, requiring a creative approach to the development of these new methods. Thus, although this is a very robust approach to the validation problem, it presents the drawback of requiring an investment from scientists for its development.

In the second section of this chapter, we introduce several validation algorithms following all the aforementioned approaches, both consensus and standalone, for different steps of the tomography pipeline. In the third section, we study in detail the tilt series alignment problem, providing two new validation algorithms that, in a standalone manner, automatically detect alignment errors both in the tomographic reconstruction and directly in the aligned tilt series.

2.1.2 Tilt series alignment

The tomographic acquisition introduces some undesired movements of the sample and imposes some rotation of the holder. Thus, it is necessary to correct for these movements to reference all the images from the tilt series to a common geometry. This involves correcting for relative movements between images and determining the rotation tilt axis, as it is depicted in Figure 2.1. These corrections are calculated in the tilt series alignment step.

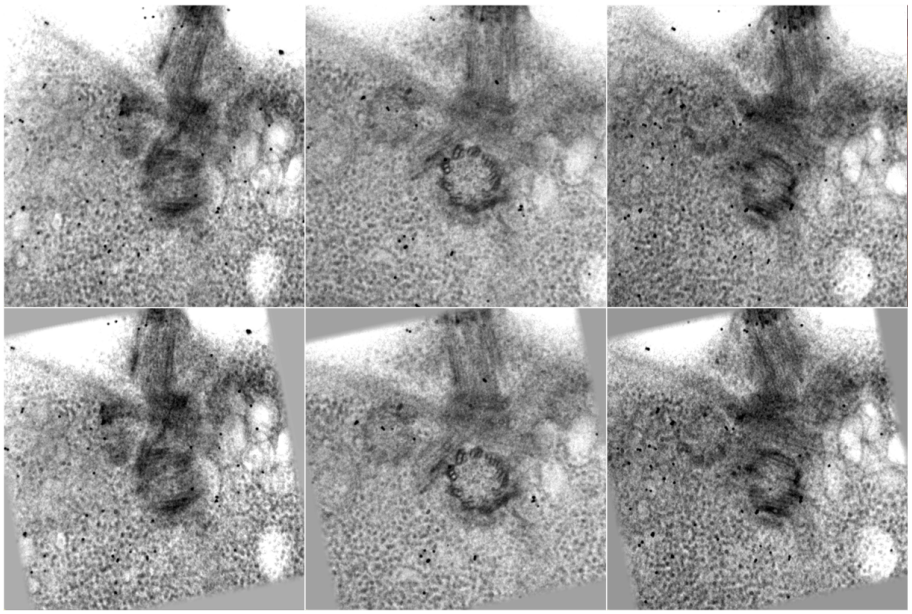


Figure 2.1: Illustration of the tilt series alignment process. The top row presents a misaligned tilt series, while the bottom row shows the results after alignment. Each column corresponds to different tilt angles: extreme negative, center, and extreme positive. Both shifts and rotations were applied to ensure smooth transitions between tilt images and vertical alignment of the tilt axis.

The first step consists of what is known as prealignment. Here, the aim is not to estimate the geometry of the acquisition but to ensure the smooth transition along the tilt series. Usually, this operation is performed by searching for the shifting transformation that maximizes the cross-correlation between adjacent images.

$$(I \star I')[x, y] \triangleq \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \overline{I[m, n]} I'[m + x, n + y]. \tag{2.2}$$

This calculation is sensitive to the tilt axis angle prior provided. The more accurate the estimation of the shifts and the better the correction of the transition between the different tilt images, the more robust the estimation of the acquisition geometry will be.

In the introduction section, three different approaches to the alignment of tilt series

have been introduced: fiducial-based and non-fiducial-based, with the latter having two variants: patch tracking and reprojection alignment. The validation methods introduced in the second section of this chapter focus on the fiducial-based samples.

Fiducials provide reliable points along the tilt series that serve as landmarks for the subsequent alignment calculation. For this same reason, fiducials are also useful when the aim is to detect alignment errors. The first step in calculating the alignment is the detection of the fiducial in every tilt image. The robustness of this detection is very important, since missing out on some fiducials or misdetecting artifacts and other elements as fiducials may lead to the spoilage of the alignment calculation (or equivalently, the detection of alignment errors).

Different approaches for this task are introduced in the state-of-the-art section and, indeed, it remains an open problem in the field, as evidenced by this dissertation. Nonetheless, there are common tools and algorithms that are typically applied to this problem. Applying filters that remove information at frequencies that do not contribute to the signal characteristic of the fiducials, noise reduction, or background subtraction operations are examples of some basic operations usually performed by fiducial detection algorithms.

Moreover, fiducials are high-contrast features in the image that are highly sensitive to the differentiation operation, as they appear as strong gradient elements in the tilt image, producing sharp edges. Consequently, a common operation for their enhancement is the Sobel operator

$$M(x,y) = \sqrt{(G_x(x,y) \star I(x,y))^2 + (G_y(x,y) \star I(x,y))^2}, \tag{2.3}$$

being $G_x(x,y)$ and $G_y(x,y)$, the gradient kernels in the vertical and horizontal directions, and $\star$ the convolution operator. Thus, this operation isolates regions of rapid intensity change, which correspond to the visual boundaries of fiducials

In order to detect those regions that represent a fiducial in the image, another operation is the convolution with a kernel that represents the typical shape of a gold bead. The common approach in the field is the assumption that fiducials present a bidimensional Gaussian shape. Thus, the convolution with a Gaussian kernel leads to the enhancement of features that present this distribution in the image.

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right), \tag{2.4}$$

where typically it is assumed that $\sigma_x^2 = \sigma_y^2 = \sigma^2$, imposing no skew in the definition of the kernel. Thus, this approach selectively amplifies image features that closely match the expected size and shape of gold beads.

Although there are other approaches in this field that target this problem by proposing

more sophisticated methods, the development of the work presented in this dissertation has employed the following approach, which has yielded robust results, as will be demonstrated in this section.

Once fiducials are detected in every tilt image, a correspondence is established between the different fiducials along the tilt series, composing the known landmark chains. These abstractions represent the position of the same fiducial along every tilt image from the series. As in the case of the tilt series prealignment, this step is very sensitive to an accurate tilt axis angle prior and an accurate performance of the prealignment itself. Thus, it is possible to relate every gold bead position in one image to the adjacent one by accounting for the expected movement due to the rotation of the sample and selecting the most suitable point in the following images.

Finally, once the landmark chains are composed, it is possible to solve for the geometry of the acquisition. In general terms, this process is carried out by calculating the three-dimensional coordinates of each fiducial used in the alignment. The projections of each three-dimensional coordinate in each of the tilt images must be compatible with the estimated acquisition geometry. In this way, each detected fiducial contributes to the calculation of the position of its corresponding three-dimensional coordinate. A schematic illustrating the contribution of the fiducials to the three-dimensional reconstruction is shown in Figure 2.2.

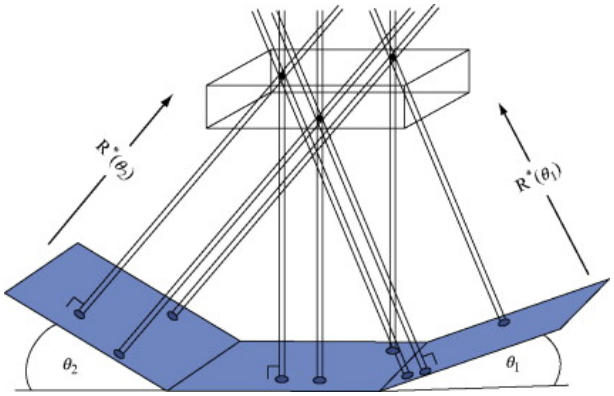


Figure 2.2: Schematic showing how gold beads are projected in 2D images. Tracking gold beads across images enables the recovery of the tridimensional location of the gold beads and the motion of the microscope. Image referenced from Amat et al. (2010).

Alignment algorithms are typically resilient to incomplete chains, where the position of the same gold bead is known for a subset of the tilt images belonging to the series. Moreover, being an overdetermined system (there are more equations than unknowns), the algorithms are also resilient to inconsistent (erroneous) constraints, and a least-squares solution that minimizes the sum of the squares of the residuals is calculated.

The mathematical definition, which is consistent with our earlier work Sorzano et al.

(2020), is as follows. Let $\mathbf{r}j$ be the coordinate of a tridimensional landmark and $\mathbf{p}ij$ the coordinate of its projection onto image i. The relationship between $\mathbf{r}j$ and $\mathbf{p}ij$ is

$$\mathbf{p}_{ij} = A_i \mathbf{r}_j + \mathbf{d}_i, \tag{2.5}$$

where A_i is the projection matric accountign for the tilting around the tilt axis and a subsequent in-plane rotation, and $\mathbf{d}_i$ is a bidimensional vector accounting for an in-plane shift of the image i. The projection matrix is defined as

$$A_i = H R_{\psi_i} R_{\theta_i, \mathbf{u}_{\text{axis}}} D_i, \tag{2.6}$$

where H is the matrix that projects 3D coordinates into its X, Y components, R_{ψ_i}, is a rotation matrix of ψ_i degrees around the Z axis (that is, the beam axis), D_i is any 3D deformation matrix, and $R_{\theta_i, \mathbf{u}_{\text{axis}}}$ is a rotation matrix of θ_i degrees around the tilt axis $\mathbf{u}_{\text{axis}}$. The tilt axis is described by the two Euler angles.

The goal of residual-based algorithms is to minimize the reprojection error

$$E = \sum_i \sum_{j \in V_i} \|\mathbf{p}_{ij} - (A_i \mathbf{r}_j + \mathbf{d}_i)\|^2, \tag{2.7}$$

reducing the magnitude of the calculated residuals.

This mathematical formulation links each 3D landmark to its observed 2D projection in every image, determining the alignment that minimizes the difference between predicted and actual fiducial locations in each tilt image.

A residual vector is a two-dimensional vector representing the detected marker's relative position in the tilt image compared to its calculated position after solving the series alignment. A residual graphical definition is shown in Figure 2.4. Thus, the set of residual vectors measures the quality of the computed alignment in a tilt series

$$\mathbf{e}_{ij} = \mathbf{s}_{ik} - \mathbf{p}_{ij}, \tag{2.8}$$

where $\mathbf{e}_{ij}$ is the residual vector defined by the detected landmark $\mathbf{s}_{ik}$ and its associated projection of the 3D coordinate $\mathbf{p}_{ij}$. These residuals will become particularly important as a metric to assess the alignment quality, especially in the second algorithm introduced in the third section of this chapter. A more detailed schematic of this mathematical definition is provided in Figure 2.4.

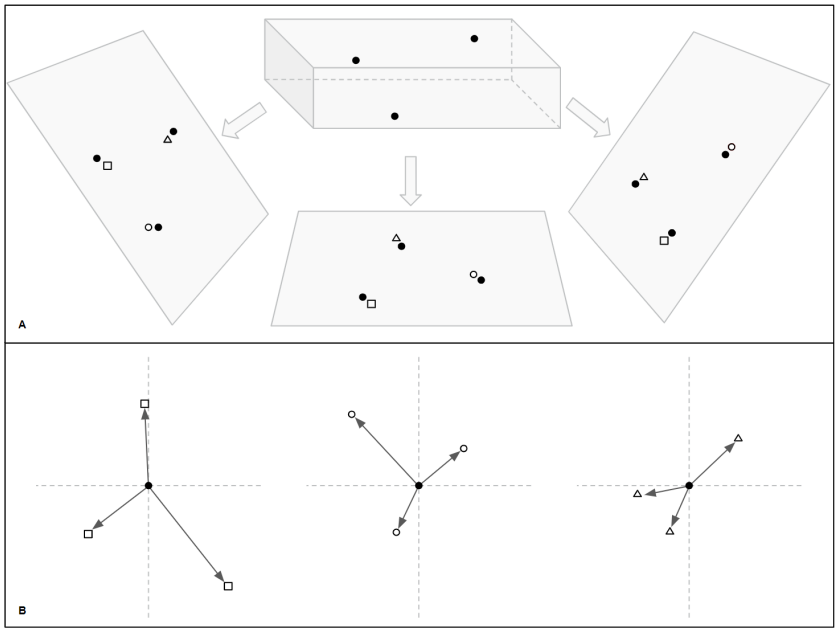


Figure 2.3: Schematic definition of residual vectors. (A) represents the projection of the three-dimensional coordinate onto every tilt image (black dots) and the detected fiducials in each tilt image (circles, squares, and triangles). Thus, the residual is defined as the vector whose origin is the projection of the fiducial coordinate and the endpoint is the detected landmark. (B) represents the calculated set of residuals for each fiducial in the sample. A hypothetical case with only three fiducials and three tilt images is presented for simplicity. Figure referenced from de Isidro-Gómez et al. (2025a).

2.2 Validation methods in cryo-ET

To ensure a comprehensive validation of the tomographic data processing pipeline, the first objective of this dissertation was to conduct a detailed analysis and proposal of validation strategies for each step. The scope of the development of validation methods initially focused on the reconstruction workflow, subsequently expanded to the subtomogram averaging pipeline. However, useful recommendations for cellular biology applications are also provided. Per-particle-per-tilt methods were not considered due to the advanced maturity of the software available at the time this work was conducted. In this work, we propose both established tools and the foundations for potential validation strategies that have yet to be developed.

The following section presents a summary of the different processing steps addressed in this study, along with a concise overview of the possible validation methods for each. The tools and methods presented in this chapter have been developed by the co-authors as a whole, although it has been the first author's task to coordinate all the work and propose validation solutions for each step of the workflow. Figure 2.5 provides a schematic representation of the validation workflow proposed in the first publication of the compendium

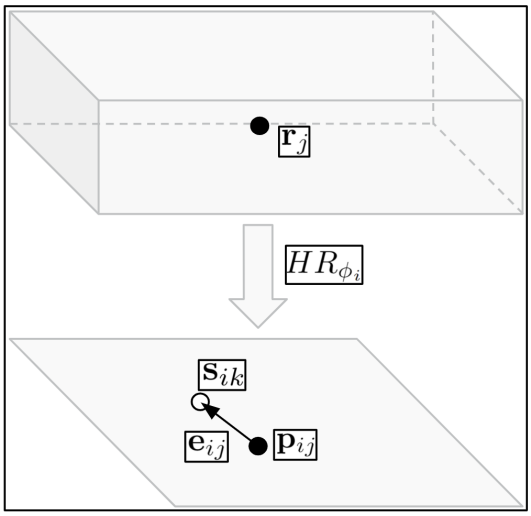


Figure 2.4: Schematic of the residual calculation algorithm, showing the projection of the fiducial coordinate onto the tilt image (black dots) and defining the residual as the vector from the projection to the nearest detected landmark (empty dot). Figure referenced from de Isidro-Gómez et al. (2025a).

de Isidro-Gómez et al. (2025b).

- Movie alignment: at each tilt angle, a stack of frames is acquired that subsequently must be aligned and averaged in a single image. Thus, we define Δx_i and Δy_i as the global relative shifts between the i-th and the $(i+1)$-th frame, representing the overall alignment for a movie

$$S = \begin{pmatrix} \Delta x_1 & \Delta x_2 & ... \\ \Delta y_1 & \Delta y_2 & ... \\ 1 & 1 & ... \end{pmatrix} \tag{2.9}$$

  where the 1 is added to express the shifts in homogeneous coordinates. The following validation method is proposed:

  - Consensus alignment: given two movie alignments for the same set of frames S_1 and S_2, a comparison can be made to determine the extent to which they are coincident. However, given that different alignment algorithms may differ in the origin of coordinates and axis conventions, first the affine transformation that best transforms one set of shifts into the other must be calculated

$$A = \arg\min \|S_1 - AS_2\|^2 = (S_1 S_2^T)(S_2 S_2^T)^{-1} \tag{2.10}$$

    establishing a direct transformation between both sets defined as $S_1' = AS_2$. Now that both transformations are aligned to a common reference, it is possible

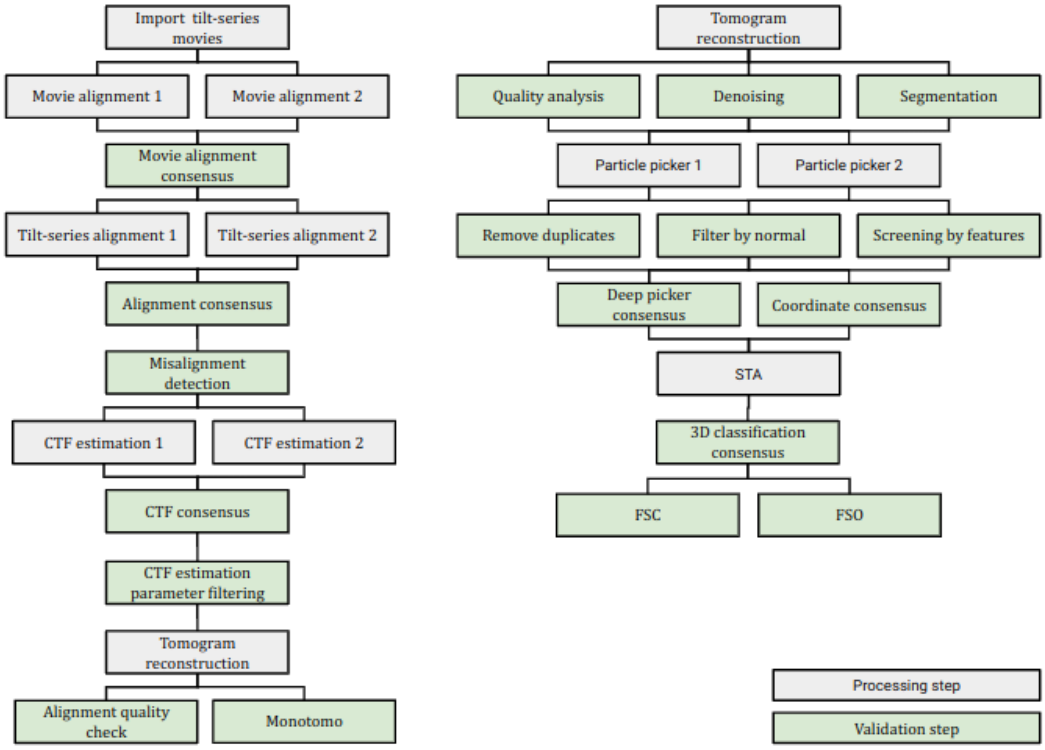


Figure 2.5: Schematic of the cryo-ET analysis workflow. The workflow on the left goes from the input tilt series movies to the reconstructed tomogram. The workflow on the right performs a subtomogram analysis. Image referenced from de Isidro-Gómez et al. (2025b).

to calculate the Pearson correlation coefficient of the global shifts in both X and Y directions, defined as $\rho_{\Delta x}$ and $\rho_{\Delta y}$ respectively. The final correlation between both trajectories is defined as

$$\rho = \min(\rho_{\Delta x}, \rho_{\Delta y}), \tag{2.11}$$

being possible to threshold the correlation value ρ to establish a minimum similarity between alignments S_1 and S_2.

This procedure quantifies how similarly are the relative displacements for the same set of frames calculated by two different algorithms. The author contributed to the development of this method.

- Tilt series alignment: tilting the sample during image acquisition introduces unintended movements that must be corrected by referencing all tilt images to a common geometry. As introduced in the previous chapter, we distinguish between two alignment strategies based on whether fiducial markers have been used in sample preparation. Regardless of the chosen strategy, the alignment process abstracts into a set of transformation matrices that map each tilt image onto the reference geometry. We denote this set of matrices as A_i, where there are T tilt images, indexed as $i = 1, 2, \ldots, T$, and defined in Equation 2.6.

  Thus, it is possible to calculate a new set of aligned tilt series $\tilde{I}_i$ interpolating each tilt image by applying each transformation

  $$\tilde{I}_i(M_i\mathbf{s}) = I_i(\mathbf{s}) \tag{2.12}$$

  The following validation methods are proposed:

  - Consensus alignemnt: as in the case of movie alignment, given two tilt series alignments $\{M_i^1\}$ and $\{M_i^2\}$, a comparison can be made to determine to what extent they are coincident. Again, to enable the comparison between both alignment the set of the affine transformation between both sets must be calculated. Given two aligned tilt series:

    $$\begin{aligned} \tilde{I}_i^1(M_i^1\mathbf{s}) &= I_i(\mathbf{s}) \\ \tilde{I}_i^2(M_i^2\mathbf{s}) &= I_i(\mathbf{s}) \end{aligned} \tag{2.13}$$

    it is possible to calculate a set matrices P_i that transform the coordinates of one alignment onto the other for each tilt image

$$P_i = M_i^1 (M_i^2)^{-1} \tag{2.14}$$

Nonetheless, this transformation matrix should be the same for all tilts, so we average all of them

$$P = \frac{1}{T} \sum_{i=1}^{T} M_{i,1} M_{i,2}^{-1} \tag{2.15}$$

at this stage it is possible to calculate the disagreement matrix ΔM_i between both sets

$$\Delta M_i = M_i^1 - P M_i^2 \tag{2.16}$$

From this matrix it is possible to obtain the associated rotation, ΔR_i, and translation, $\Delta \mathbf{t}_i$, and use them as thresholding variables to assess the coincidence of the alignments. The average alignment for those tilt series that fulfills the requirements can be calculated as

$$M_i^{12} = \frac{1}{2}(M_i^1 + P M_i^2) \tag{2.17}$$

Besides the average disagreement matrix, ΔM_i, the P matrix also provides valuable information for studying the concordance between the two alignments. Functionally, this matrix summarizes the cumulative disagreement between alignments, capturing both translational and rotational differences across the tilt series.

In cryo-ET, any offset displacement along the tilt axis (which, by convention, is usually oriented along the Y-axis) corresponds to two equivalent alignments. This occurs because a uniform translation of all the tilt images along the tilt axis results in equivalent reconstructions that differ only by a displacement. However, if the P matrix contains any other transformation, such as a rotation or a displacement along the X-axis, then the two alignments are no longer equivalent. Method developed by the author.

– Misalignment detection from residual vectors: following the definition of the residual in Equation 2.8, this set of vectors provides information about the mismatch between the projections of each calculated three-dimensional coordinate for each fiducial and the corresponding detected coordinates in each tilt image. Thus, it is possible to identify misalignments in the tilt series by analyzing the distribution of these residuals across the series. Further details of

this algorithm are not provided in this section, as it will be thoroughly defined in Section 2.3.2. Method developed by the author.

- Contrast transfer function (CTF): the set of acquired images is affected by the microscope's aberrations, making them non-ideal projections of the sample. Consequently, estimating and correcting these deviations from the ideal projection images is essential, which are modeled by the CTF. The main aberrations are defocus, spherical aberration, and astigmatism.

  The CTF is estimated from the Power Spectral Density (PSD) of every tit image, which presents a set of concentric fringes caused by the CTF sinusoidal component, called Thon rings Thon (1966). Consider a one-dimensional description of the Contrast Transfer Function (CTF) without high-frequency attenuation (a more detailed definition is provided in Sorzano et al. (2007)):

$$CTF(f;\Delta f) = -\sin(\chi(f;\Delta f)) + Q(f)\cos(\chi(f;\Delta f)) \tag{2.18}$$

  Here, f denotes the spatial frequency measured in Å^{-1}, Δf represents the defocus (the sample-to-detector distance), $Q(f)$ is the fraction of electrons undergoing inelastic scattering, and $\chi(f;\Delta f)$ is defined as

$$\chi(f,\Delta f) = \pi\lambda\left((\Delta f)f^2 + \frac{1}{2}C_s\lambda^2 f^4\right) \tag{2.19}$$

  where λ is the electron wavelength, and C_s is the spherical aberration of the microscope.

  Extending this one-dimensional CTF model to a two-dimensional definition involves two defoci, with their difference being referred to as astigmatism. The following validation methods are proposed:

  - CTF consensus: as in other consensus approach, two CTF estimations obtained from different algorithms are obtained from the same tilt image, presenting two different defoci values Δf_1 and Δf_2. Thus, a new function is defined as the difference between the two estimations

$$\epsilon(f;\Delta f_1,\Delta f_2) = |CTF(f;\Delta f_1) - CTF(f;\Delta f_2)| \tag{2.20}$$

    Due to the behavior of the CTF function, these differences are exacerbated as the frequency increases. This behavior can be observed in Figure 2.6. Thus, it is

possible to establish a frequency threshold $f_{\max}$ up to which the phase difference between both CTF estimations does not exceed a certain value $\Delta\chi_{\max}$.

$$|\chi(f;\Delta f_1) - \chi(f;\Delta f_2)| < \Delta\chi_{max} \quad \forall |f| < f_{max} \tag{2.21}$$

where typically, a threshold of $\Delta\chi_{\max} = \pi/2$ is chosen. The higher the value of $f_{\max}$, the more confident the CTF estimation becomes. For those estimations that meet these requirements, an appropriate strategy would be to average the defocus estimations Δf_1 and Δf_2

$$\bar{\Delta} f = \frac{1}{2}(\Delta f_1 + \Delta f_2) \tag{2.22}$$

In practical terms, this method informs the researcher up to which spatial frequency the CTF estimation remains reliable, helping assess whether the resolution of interest may be affected. The author contributed to the development of this method.

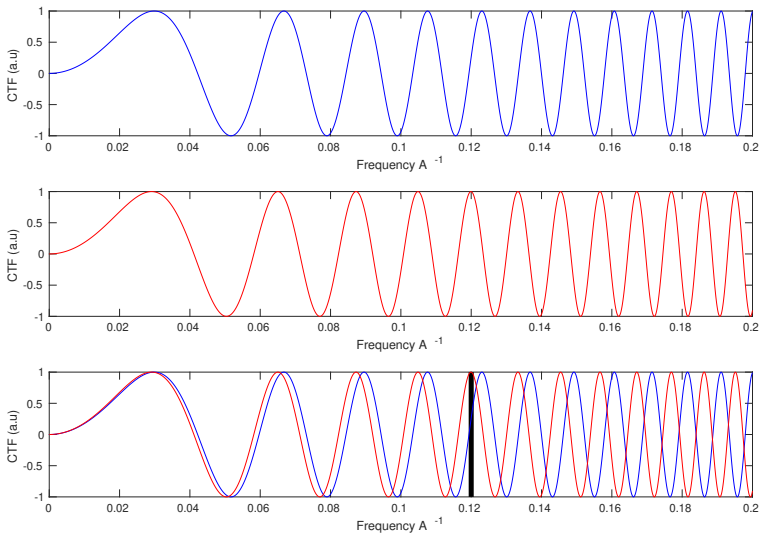


Figure 2.6: Two different CTF estimations: The first estimation (blue) considers a defocus of 27000 Å, and the second prediction considers a defocus of 30000 Å(red). At low frequencies (resolutions), both curves are similar, and both estimations can be accepted and provide a good CTF correction. However, When the frequency increases, the differences between them are significant. The vertical black line shows the resolution limit when the differences become significant. Image referenced from de Isidro-Gómez et al. (2025b).

– CTF parameter filtering: from all the parameters estimated by the CTF, it is possible to validate its estimation by establishing a range of acceptable val-

ues that jabthese parameters must satisfy. This is particularly useful since some parameters are targeted during acquisition (such as defocus), while others represent sources of error in the data that must be minimized (such as astigmatism). This approach has the advantage of applying quality criteria to the CTF estimation without requiring multiple CTF estimations for the same tilt series. The author contributed to the development of this method.

- Tomogram reconstruction: after all the previous processing the following step consists of solving the inverse problem of determining the tomographic three-dimensional structure compatible with the set of projections.

  Independently of the algorithm applied to solve the inverse problem introduced in the first chapter of this dissertation, it is possible to implement strategies to assess the quality of the tomographic reconstruction.

  - Detection of artifacts: reconstruction algorithms allocate the signal from the projection data into the tomographic reconstruction. However, incorrect alignment of the tilt series can lead to misallocation of this signal, producing artifacts in the reconstruction. These artifacts are particularly visible in regions with high contrast, with fiducial markers within the sample being of special interest.

    Thus, it is possible to develop strategies to analyze the fiducials in the reconstruction to detect these artifacts. The proposed methodology involves locating the fiducial markers in the sample (if they are not already provided) and training a deep neural network (DNN) to assess whether each fiducial reconstruction contains artifacts. This approach provides an overall evaluation of the tomogram's reconstruction quality. Further details of this algorithm are not included in this section, as it will be thoroughly defined in Section 2.3.1. Method developed by the author.

- Tomogram postprocessing: different global and local transformations can be applied to topographic reconstructions to improve, analyze, or enhance their features, thereby facilitating subsequent analysis. However, validating these transformations after reconstruction is not straightforward, as obtaining quantitative assessments is challenging and highly subjective, depending on the goals and conclusions the researcher seeks to derive from the reconstructed sample. Different outcomes of the validation methods presented below can be observed in Figure 2.7. There are several approaches for analyzing possible validation strategies:

  - Structural validation: this approach applies to both denoising and segmentation tools. Although the underlying algorithms and results may differ, both tools aim to enhance the signal relative to the noise and identify structural features in the sample. This common goal enables validation through the comparison of regions detected by different algorithms, making those results that are consistent across different methods more reliable. To the best of our knowledge, no validation strategy has yet been developed in this area.
  - Quality assessment validation: this approach applies to both global and local quality metrics. Global quality metrics, typically expressed in terms of overall resolution, should yield similar results for the same reconstruction. For local quality metrics, however, there is greater variability in how quality is measured (e.g., SNR, resolution, etc.), often reflecting the spatial reliability of information within specific regions of the reconstruction. In these cases, it is possible to assess whether different algorithms show consistency in these regions, allowing for the generation of a consensus-based local quality assessment map. As with the previous case, to the best of our knowledge, no implemented strategies are currently available for scientists to use.
  - Holistic approaches: A final line of development involves integrating the information from both of these approaches, where quality assessment provides a confidence measure for the data produced by denoising and segmentation algorithms

- Particle picking: following the structural analysis approach, the next task is to locate the particle of interest in the sample, which is affected by the low SNR of the tomographic reconstruction. The proposed validation methods also leverage the fact that the selected algorithm provides directional information:

  - Picking consensus: ideally, all particles in the tomogram should be identified in identical or nearly identical coordinates by the N different picking algorithms applied to the sample. However, this expected behavior may be disrupted due to the presence of false positives, false negatives, and mislocated particles. Therefore, following a consensus approach, a particle is considered a true positive if it is identified by a minimum subset of N pickers out of the total M algorithms, within a specific radius that defines the tolerance. This is exemplified in Figure 2.8.
  - Directional consensus: if the various particle-picking algorithms not only provide the location of particles of interest but also include directional information,

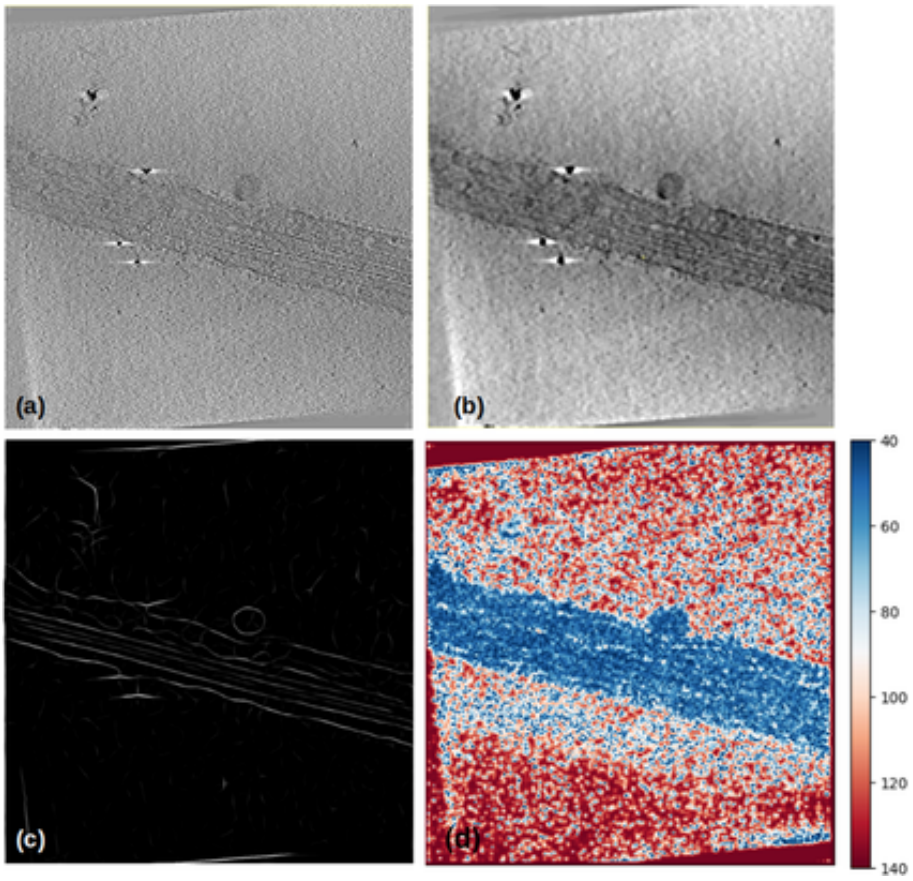


Figure 2.7: Different postprocessing algorithms applied to the same tomogram: (a) Original tomogram, (b) The tomogram is denoised with tomoEED Moreno et al. (2018), (c) A segmentation of the tomogram with tomosegmemTv Martinez-Sanchez et al. (2014), and (d) Local resolution tomogram obtained with MonoTomo Vilas et al. (2020b).

it becomes possible to reach a consensus on this directional estimation for the particles identified in the previous picking consensus. To enable direct comparison and averaging of the different solutions, the direction vectors are converted into quaternions. This approach simplifies the problem of averaging rotations into a straightforward and efficient eigendecomposition process Sorzano et al. (2014) Hu et al. (2020). A quaternion, defined as $q = a + bi + cj + dk$, can represent a rotation by an angle α around a specified 3D axis $\mathbf{u}$.

$$q_{\mathbf{u},\alpha} = \cos\left(\frac{\alpha}{2}\right) + \sin\left(\frac{\alpha}{2}\right)\frac{\mathbf{u}}{||\mathbf{u}||} \tag{2.23}$$

Considering a set of quaternions Q of the form

$$Q = (w_1\boldsymbol{q}_1, w_2\boldsymbol{q}_2, ..., w_n\boldsymbol{q}_n)$$

where $\boldsymbol{w}$ represents a vector of weights defining the weighted quaternion average, and $\boldsymbol{q}$ represents a quaternion vector whose rows stored the quaternion representing a given rotation.

From the previous representation, we can compute the self-adjoint positive semi-definite matrix QQ^T. The normalized eigenvector corresponding to the largest eigenvalue of the previous matrix corresponds to the weighted average of the set of quaternions previously defined. In other words, this formulation provides a principled way to identify and reinforce the preferred orientation consistently detected across different algorithms for each particle.

- Deep picking consensus: a more sophisticated approach to the consensus problem involves training a deep neural network capable of assessing the quality of the picked particles. To achieve this, the network can be fed with the extracted subtomograms from matching coordinates between the pickers as positive elements, while non-matching subtomograms and background noise are used as negatives. Once the network is trained, it operates as a standalone system capable of evaluating the quality of the detected elements in the sample.
- Filtering by map features: this standalone validation algorithm allows for the filtering of elements based on the local properties of the sample from which they are extracted. There is a wide variety of local properties that can be used for filtering the coordinates, such as local resolution, local defocus, distance to the carbon edge, or particle mass. Method developed by the author.
- Filtering duplicates: another standalone method enables the removal of coordinates that refer to the same element in the sample. A threshold radius is

defined to determine whether these coordinates correspond to the same element, which is particularly useful when the picking algorithm does not impose any geometrical constraints.

– Filtering by normal: when the picking algorithm provides directional information, it becomes possible to impose directional constraints in a standalone manner. These assessments allow the removal of coordinates based on their angle relative to a surface or mesh, ensuring normality with respect to a specific structure in the sample. Alternatively, coordinates can be filtered based on their tilt angle, ensuring they do not align with the direction of the missing wedge.

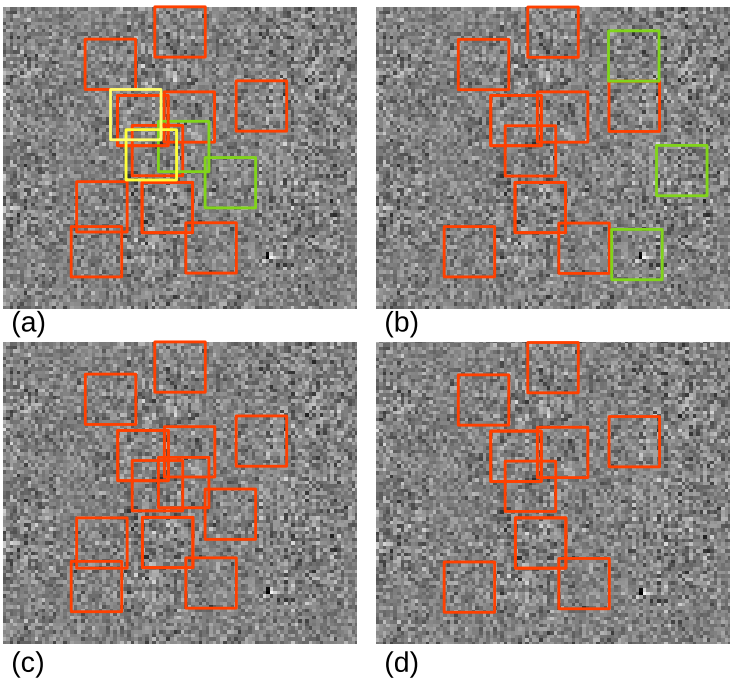


Figure 2.8: Illustrative examples of the picking consensus approaches: (a) Picked tomogram with an algorithm X, (b) the same picked tomogram with a different picking algorithm Y. (c) Validated particles in the picking (a) after applying the remove duplicates protocol (duplicated particles in yellow). (d) The consensus picked particles with the pickers A and B (non common particles in green). Image referenced from de Isidro-Gómez et al. (2025b).

- Subtomogram averaging: once the particles of interest are located in the sample, their information is combined into a single reconstruction. This is done through an iterative process in which all the subtomograms are aligned to a reference volume and classified into different conformational states. Both of these steps may undergo the following validation strategies:

– Alignment consensus: each alignment algorithm provides a geometrical transformation for every subtomogram V_i to its class representative. Let there be two different alignments provided by different algorithms:

$$\begin{aligned} \tilde{V}_i(M_i^1\mathbf{r}) &\approx V^1(\mathbf{r}) \\ \tilde{V}_i(M_i^2\mathbf{r}) &\approx V^2(\mathbf{r}) \end{aligned} \tag{2.24}$$

where V^1 and V^2 are the class representatives. However, these may not be aligned with each other. Then, we may find the transformation that aligns them as

$$V^1(M^{21}\mathbf{r}) \approx V^2(\mathbf{r}) \tag{2.25}$$

From all these relations, we know how to align the subtomogram to the first class representative using the second alignment matrix

$$\tilde{V}_i\left(M_i^2(M^{21})^{-1}\mathbf{r}\right) \approx V^1(\mathbf{r}) \tag{2.26}$$

Consequently, it must be

$$\Delta M_i = M_i^2(M^{21})^{-1} - M_i^1 \approx 0 \tag{2.27}$$

from which it is possible to calculate rotational ΔR and translational $\Delta\mathbf{t}$ transformations implied by ΔM_i, estimating the consistency of the alignment for all subtomograms.

The intuition mirrors other consensus strategies: it evaluates the agreement between alignment algorithms, with the key distinction that all comparisons are performed in 3D space and referenced to a common coordinate system.

– 3D classification consensus: in a similar trend to the previous approach, it is possible to achieve consensus not only on the alignment but also on the classification of each subtomogram to its associated class, although this is an unstable process Sorzano et al. (2022). Thus, given n classifications with $M_1, M_2, \ldots, M_n$ classes, the subgroups of particles that have been classified together are clustered. The maximum number of subgroups is:

$$m = \prod_{i=1}^{n} M_i \tag{2.28}$$

Stable classes tend to be preserved across different executions, while unstable classes tend to be distributed into many small subgroups. The fewer the number of subgroups, the more stable the classification is. Thus, consistently

classified particles are preferred when selecting the representative sets.

- Quality assessment of the average map: finally, once the average map is calculated, its quality is determined to estimate its reliability. There are several metrics of interest for this purpose:

  - Fourier Shell Correlation (FSC): being the standard in the field, this metric indicates the extent to which the frequency information is consistent between two half maps. Nonetheless, it rewards systematic errors committed during the previous image processing van Heel (2013); Henderson (2013); Sorzano et al. (2022).
  - Overfitting detection: inherited from Single Particle Analysis (SPA), it is possible to validate the reconstructed map by detecting possible overfitting in the reconstruction Chen et al. (2013); Heynmann (2015); Vargas et al. (2016); Heymann (2018), making its application straightforward to the tomography problem.
  - Local resolution: in a different trend from the previous two approaches, this method provides local information about the quality of the map and the expected detail of each region of the protein, rather than a global metric. It has been shown that local resolution is a necessary condition for visualizing structure but not sufficient Vilas et al. (2020a). Additionally, it is possible to assess the local-directional resolution, which informs not only about the quality of the region under study but also the directional loss of information Vilas et al. (2020c).
  - Fourier Shell Occupancy (FSO): in addition to the previous FSC, this metric not only informs about the information present at each frequency but also the frequency range in which the map presents anisotropy Vilas and Tagare (2023).

The summarized methods above propose validation tools that target each step of the cryo-electron tomography image processing pipeline. However, there are validation approaches that target the pipeline as a whole. These are some of these validation strategies:

- Pipeline parallelization: subsetting the frames of each tilt image allows for the generation of two equivalent tilt series movies, preserving the angular sampling, instead of averaging all these frames into a single one. It is not necessary to split the frames in halves (which would reduce the SNR by a factor of 2) but instead, generating two heterogeneous subsets is sufficient. From these two tilt series movies, it is possible to run two complete workflows in parallel, reaching consensus at each

step of the pipeline even with a single algorithm involved at each step. The main drawback of this approach is the increase in computational burden.

- Features validation: this approach relies on the use of a priori information about the sample instead of seeking algorithmic robustness. It is very specific to the sample. Examples of this kind of validation include comparing a previous in vitro reconstruction of the protein with an in vivo one, reconstructing different specimens in the sample in parallel, or comparing known structures from the literature with those present in the sample (e.g., vesicles, fiducials, etc.).

- Map back: this tool offers the possibility to reallocate the reconstructed volume into the original tomograms at the positions identified during the picking step. Thus, a new set of tomograms is generated with the specimen of interest highlighted in the original biological context. This informs the relative orientation of the reconstructed map to the biological structures present in the sample or detects structural variability along the tomogram if heterogeneity is present in the specimen.

To conclude the proposed validation methods, there are a set of more general topics that have been identified as future strategies to increase the robustness of cryo-ET image processing and their biological interpretation:

- Workflow traceability and availability: although it is common practice in the field to provide the elucidated three-dimensional structures of interest and the raw data, there is no tradition of uploading the followed image processing workflow. This is of special interest considering the variability in the underlying biological questions and the variability in the processing strategies, especially if the considerations introduced in this section are applied. Thus, developing a common framework for reporting the processing pipeline and making these available to the community are two important topics to foresee in the field. Scipion helps to keep track of the executed steps de la Rosa-Trevin et al. (2016); Jimenez de la Morena et al. (2022).

- Data management and standardization: the cryo-ET field is populated by a vast and increasing variety of different software packages, proposing various solutions to the steps that make up the image processing pipeline. This makes data standardization an increasing need in the field, with the FAIR approach (an acronym for Findable, Accessible, Interoperable, and Reusable) being of special interest Wilkinson et al. (2016).

- Combination of different techniques: due to the wide range of biological phenomena that cryo-ET encompasses, it is a technique that can particularly benefit from enriching its results with other methods to validate findings. Correlative microscopy,

mass spectrometry, nuclear magnetic resonance, protein structure prediction Jumper et al. (2021), and protein-protein interaction prediction Bryant et al. (2022) are some of the techniques that could yield interesting results when combined with cryo-ET.

2.3 Alignment errors detection in cryo-ET

Among all the steps in the cryo-ET pipeline, tilt series alignment has emerged as a key focus of this work. First, it is a core step independent of subsequent application (in cellular or structural biology), and it has also been identified as one of the most unstable steps.

Several algorithms with different approaches target this problem, with consensus approaches being particularly suitable for validating the results, as stated in the previous section. Nonetheless, this comparison requires that the results are provided in compatible formats, which is not the case for all tilt series alignment algorithms. Additionally, no standalone validation tool has been developed for this purpose.

Due to these issues, two new standalone algorithms for the assessment of the quality of tilt series alignment are presented in this section.

2.3.1 Alignment errors detection in tomograms

The first standalone algorithm targets the detection of alignment errors in tomographic reconstructions. The following developments belong to the second scientific publication of the compendium of this dissertation de Isidro-Gómez et al. (2024).

Alignment errors lead to the misallocation of signals in the final tomogram during the reconstruction process. This misallocation results in artifacted reconstructions, which are more noticeable in regions of the tomogram with high contrast features. Therefore, the proposed method focuses on fiducial markers to assess the alignment quality.

This algorithm is designed in two steps. First, the tomograms are scored, and those presenting strong misalignment are classified as misaligned and removed from the original dataset. Second, the remaining tomograms are then fed into a second, more sensitive classifier, which re-scores them with two possible outcomes: if a threshold is provided, the tomograms are reclassified into two groups; if not, only one scored set is calculated. This workflow is summarized in Figure 2.9.

Each of the scoring algorithms proposed is developed as a deep neural network (DNN) fed with a set of subtomograms centered around fiducial markers. Both networks present the same structure, differing only in their training. The network design consists of four convolutional layers with a rectified linear unit activation (ReLu), maximum pooling, and batch normalization applied at each step. This is followed by a fully connected layer of

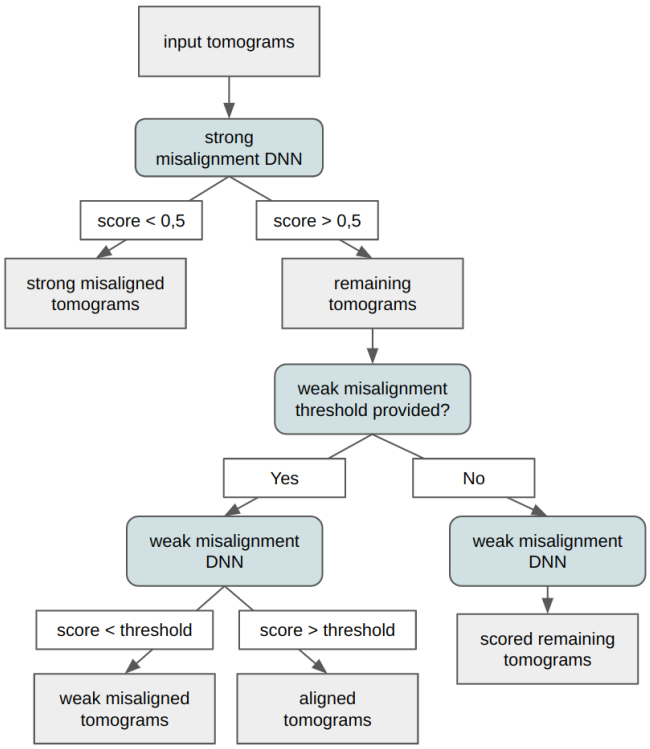


Figure 2.9: Workflow for tomogram alignment classification. Figure referenced from de Isidro-Gómez et al. (2024).

512 neurons with a dropout of 20%. Finally, a sigmoid activation function follows the last layer. This design involves 110.897 parameters (110.641 of them are trainable). An schematic of the network structure is provided in Figure 2.10.

Then, the algorithm is fed with a collection of small subtomograms centered around fiducial markers in the tomograms. The input data is standardized to fit each fiducial in a box size of $32 \times 32 \times 32$ voxels. Depending on its size, the extraction sampling rate is adjusted to provide an appropriate scope of the marker, including the potential misalignment artifact generated around it. Every subtomogram is then normalized to target a mean of zero and a standard deviation of one.

Finally, each fiducial is individually scored, ranging from 0 to 1, depending on the network's belief that the input subtomogram is correctly aligned (1) or presents some degree of misalignment (0). Once each subtomogram is scored, a global score for the tomogram as a whole is computed in two different ways, which the scientist can choose:

1. Averaging: The average of all the subtomogram scores.

2. Voting: The subtomograms are split into two groups based on whether their scores are closer to correct alignment or misalignment. The larger group determines the final classification of the tomogram.

Logically, the higher the number of fiducials located in the tomogram, the more robust and reliable the scoring and classification will be.

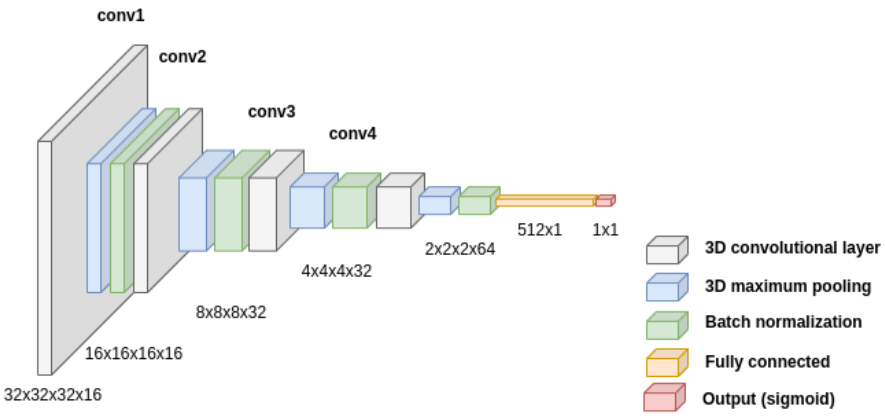


Figure 2.10: Structure of the neural network for misalignment detection from fiducial markers. Figure referenced from de Isidro-Gómez et al. (2024).

Some fiducial alignment algorithms provide, as a side result, the three-dimensional location of the fiducials included in the alignment calculation. This behavior is preferred because it avoids the need for separate location calculations and ensures more reliable positions (since errors in the alignment calculation will also lead to mislocation of the fiducials' coordinates). However, not all alignment algorithms provide this information.

To overcome this limitation, we developed a fiducial marker picker algorithm (Pick High Contrast algorithm, PHC) capable of detecting fiducials in the tomographic reconstruction. The robustness of this algorithm is crucial, as mislocation of the fiducial will lead to an inaccurate assessment of alignment quality. This algorithm executes a sequence of operations that, to reduce computational load, are performed on the two-dimensional slices of the tomogram whenever possible.

1. Tomogram preprocessing: A moving average of five slices across the tomogram is applied to increase the SNR and consistently reinforce the markers along several slices. Then, a bandpass filter centered at the frequency corresponding to the diameter of the fiducial and a 2D Laplacian filter are applied to every slice average.

2. 2D fiducial markers location: Each slice is z-score thresholded, keeping only outlier pixels. A connected components filter is then applied to remove small and non-circular regions.

3. 3D fiducial markers location: The 2D objects detected in every slice are combined into single 3D objects, clustering consistent objects in adjacent slices. The center of mass of the candidates is then refined by maximum correlation with their point symmetric object.

4. Quality filters: Finally, any of these filters may be applied:

    - Removal of duplicates: Coordinates inside a defined radius are assumed to refer to the same fiducial and are substituted by their average.
    - Removal of non-symmetric objects: Coordinates whose correlation between $V_i(\mathbf{r})$ and $V_i(-\mathbf{r})$ is smaller than a threshold are removed.
    - Removal of outlier objects: Those coordinates whose Mahalanobis distance between their radial profile and the average radial profile is above a threshold are removed.

A relaxed mode has been implemented, ensuring that if no coordinate satisfies the quality filters, none are removed, preventing an empty set. This approach is particularly beneficial for continuously feeding the quality assessment algorithm, even in cases where the tomograms are significantly misaligned and no detected features exhibit sufficient quality to pass the filters.

As a general remark, there is a trade-off between the quantity and quality of detected objects, making it necessary to balance having a large population of fiducials while keeping the false positive rate at a reduced level. This ensures that the study of the quality of

the alignment of each of these objects can be reliably made and is not performed over spurious objects in the tomogram.

To ensure the robustness and adaptability of the algorithm to different conditions and sources of fiducial coordinates, we propose iterative training of the network with different datasets. At each iteration, a new dataset is presented to the network while employing transfer learning from the previous iteration. This approach fine-tunes the training process and avoids local minima by presenting more challenging datasets to the network at each iteration, thereby making the weight refinement smoother.

The following public datasets have been processed as sources of data for training both networks: EMPIAR-10064 Khoshouei et al. (2017), EMPIAR-10164 Schur et al. (2016), EMPIAR-10364 Burt et al. (2020), and EMPIAR-10453 Turoňová et al. (2020). All datasets follow a similar acquisition pattern; each of them acquired with a tilting range from -60° to 60°, but composed of 41 images for datasets EMPIAR-10453 and EMPIAR-10164, and composed of 61 images for datasets EMPIAR-10064 and EMPIAR-10364.

These datasets comprise 302 tilt series in total that have been automatically aligned using IMOD Mastronarde and Held (2017); Kremer et al. (1996) in the Scipion framework de la Rosa-Trevin et al. (2016); Jimenez de la Morena et al. (2022). Each tilt series, and not the tomographic reconstruction, is then individually inspected to label the strength of the misalignment into three different groups: strong misalignment (61), weak misalignment (59), and properly aligned (182) tilt series. To assess which group a tilt series belongs to, we search for relative movements between the tilt images or a miscalculation of the tilt axis. To determine whether the misalignment is strong or weak, we look at whether the miscalculation in the alignment is minor (a few pixels) or more pronounced.

The networks are trained with subtomograms extracted from different calculated tomograms, accounting for possible effects of the CTF, the dose correction, and the combination of both. Although the authors consider that these effects are negligible compared to misalignment artifacts, this variability of the training data ensures that the network is sensitive to different preprocessing of the tilt series and has the side perk of increasing the size of the training dataset. After reconstruction, fiducials in each tomogram are located and extracted by employing the coordinates provided by the alignment algorithms and by the fiducial marker location algorithm aforementioned.

The distribution of extracted subtomograms obtained from each dataset and group can be observed in Table 2.1. The imbalance in the number of elements in each training population is addressed by a dynamic training system that ensures a constant proportionality of the training classes that feed the network in each training round.

We have extended the training dataset by incorporating a synthetic misaligned dataset generated by manipulating the alignment of 84 tilt series that had previously been cor-

Dataset	Alignment	TS	From alignment				From PHC			
			Raw	Dose	CTF	Dose+CTF	Raw	Dose	CTF	Dose+CTF
10064	Aligned	2	7	0	7	0	8	0	8	0
	Weak	2	17	0	17	0	12	0	11	0
10164	Aligned	32	351	351	351	351	1352	1250	1287	1251
	Weak	11	114	114	114	114	472	485	467	487
10364	Aligned	18	261	261	261	261	837	855	735	755
10453	Aligned	130	2015	2015	1995	1995	1626	2126	2034	2024
	Weak	46	817	817	817	817	567	892	839	839
	Strong	61	869	869	869	869	874	769	676	703

Table 2.1: Distribution of tilt series and subtomograms obtained from each training dataset, including its distribution in different groups depending on the alignment group, the source of the subtomogram coordinates, and, inside this, the preprocessing performed in each of them.

rectly classified as aligned by the network. A new tool has been developed to allow modifications to the alignment by introducing controlled misalignment, enabling independent adjustments to shifts and angles in different trends:

- Offset: A fixed transformation is applied to every tilt image in the series (p_0).
- Incremental: The transformation applied increments for every tilt image in the series at a fixed rate (p_1).
- Sine lobe: The transformation introduced in the series follows a half-sinusoidal period behavior, with both amplitude (p_2) and phase (p_3) configurable.
- Sinusoidal: The transformation introduced in the series follows a sinusoidal period behavior, with both amplitude (p_4) and phase (p_5) configurable.
- Gaussian: The transformation follows a Gaussian distribution for each tilt image, with a defined standard deviation (p_6).

Thus, the final displacement introduced for each shift and angle of every tilt image follows this expression

$$\Delta d_i = p_0 + p_1 \cdot i + p_2 \cdot \sin\left(\frac{i + p_3}{I}\pi\right) + p_4 \cdot \sin\left(\frac{i + p_5}{I}2\pi\right) + \mathcal{N}(0, p_6) \tag{2.29}$$

where i is the index position of the tilt image in the series and I is the total number of tilt images that compose the series.

Using these tools, we have expanded the initial training dataset with common misalignment patterns documented in the field. Assuming that the tilt axis is Y, the misalignment patterns that we have simulated are:

- Bananas: a constant offset in X added to all images in the series.
- Twisters: a constant offset in the in-plane rotation added to all images in the series.
- Birds: an incremental displacement in the Y direction.

Figure 2.11 shows an example of each one of these artifacts. For network training, a total of 4921 subtomograms presenting bananas, 3462 presenting birds, and 4105 presenting twisters have been added to the training dataset.

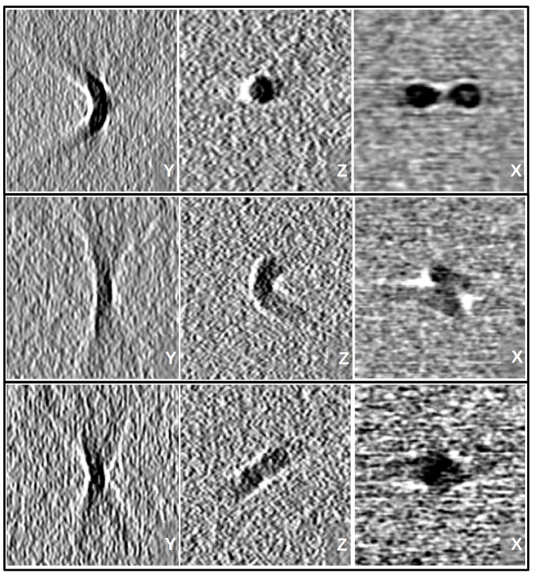


Figure 2.11: Slices of the same fiducial marker reconstruction under different misalignment patterns. The left column shows the XZ-plane, the center column the XY-plane, and the right column the YZ-plane. The first row presents bananas (incremental shift in X), the second line presents twisters (rotation of the tilt axis), and the last row birds (incremental shift in Y direction). Figure referenced from de Isidro-Gómez et al. (2024).

In the first training stage, the network is iteratively fed with the generated artifacts, the subtograms extracted from both aligned and strongly misaligned tomograms based on the coordinates obtained by the alignment algorithm, and finally, the fiducial markers' location algorithm. The weights of the network are updated at each iteration. The outcome model from this first training stage aims to detect strong misalignments in the tomograms, serving as the first of the two models in the presented classification algorithm.

Inheriting the model from the first training stage, the network is now fed with the subtograms extracted from both aligned and weakly misaligned tomograms, based on the coordinates obtained by the alignment algorithm and finally, the fiducial markers' location algorithm. Weights are inherited between iterations. The outcome model from this second training targets weak misalignment detection in the tomograms, serving as the second of the two models in the presented classification algorithm.

Before analyzing the performance of the classification algorithm, we study the performance of the fiducial marker location algorithm. To avoid dataset contamination, the relaxed mode is disabled so that only confidently detected fiducials are considered.

Comparing the results with the fiducial coordinates provided by the alignment algorithm, the number of fiducial markers located by the algorithm compared to those used to solve the alignment is 41.25% for the strongly misaligned subset, 81.07% for the weakly misaligned subset, and 68.18% for the aligned subset. It is expected that the performance of the algorithm is affected by the accuracy of the alignment calculation. The number of missed fiducials should increase in reconstructions that are poorly aligned, due to a lower SNR and their removal resulting from a low correlation with their mirror.

For a more detailed study, 10 tomograms from each group have been manually inspected. Figure 2.12 shows a comparison of the fiducials located by the proposed algorithms and the alignment algorithm in one example from each population of tomograms. The location presents a low false positive detection rate of 6.06% for the strongly misaligned subset, 0% for the weakly misaligned subset, and 0.83% for the aligned subset. In the case of misaligned tomograms, automatically detected fiducials provide more valuable information for classification, as the estimated coordinates provided by the alignment algorithm tend to be incorrectly placed in the tomogram.

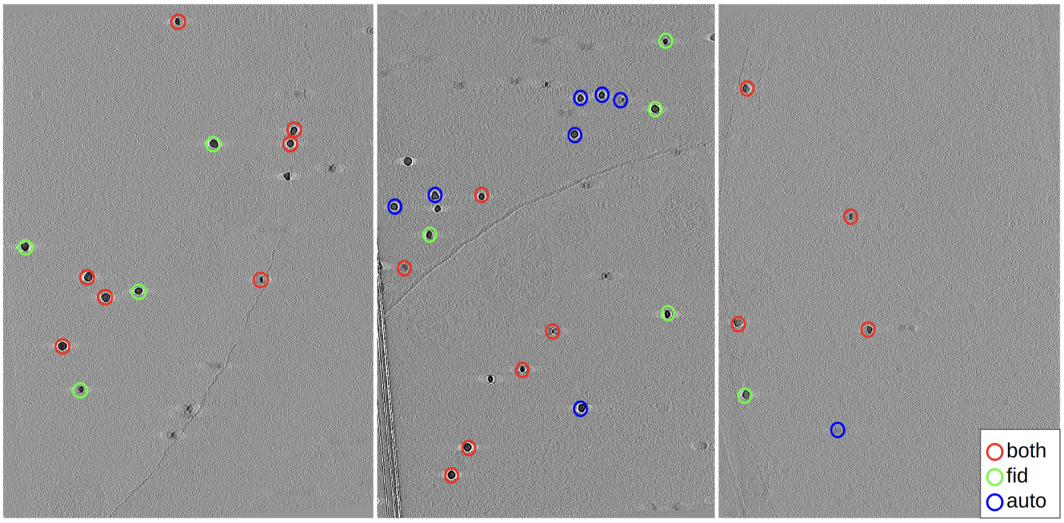


Figure 2.12: Comparison of the high-contrast objects automatically found in three different tomograms: aligned (left), weakly misaligned (center), and strongly misaligned (right). The color code is: green (found by tilt series alignment but not automatically), blue (found as a high-contrast feature but not by the tilt series alignment), and red (found by both). Figure referenced from de Isidro-Gómez et al. (2024).

Finally, we evaluate the performance of the alignment quality assessment algorithm, performing an algorithm calibration before obtaining results on real data. Several groups of bananas, twisters, and birds are generated with decreasing levels of misalignment in-

tensity, to determine the smallest change that the classification algorithm can correctly detect in at least 95% of the tomograms. These results are summarized in Table 2.2. Figure 2.13 shows an example of gold beads under these minimal deformations.

Artifact	Confidence	Misalignment
Bananas	98.81%	79 Å
Twisters	97.6%	2 degrees
Birds	98.77%	[-27, 27] Å

Table 2.2: Smallest misalignment detected in at least 95% of the subtomograms.

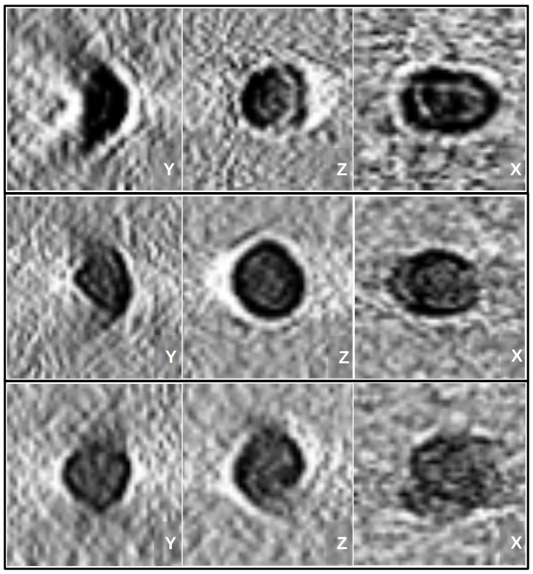


Figure 2.13: Slices of the same fiducial marker reconstruction under the minimum detectable conditions with 95% confidence. The left column shows the XZ-plane, the center column the XY-plane, and the right column the YZ-plane. The first row presents bananas (incremental shift in X), the second line presents twisters (rotation of the tilt axis), and the last row birds (incremental shift in Y direction). Figure referenced from de Isidro-Gómez et al. (2024).

Subsequently, the performance of the algorithm on real data is tested. For a more detailed study of the algorithm's performance, a preliminary analysis is conducted, focusing on the individual scored subtomograms. This study is performed separately for subtomograms whose coordinates are provided by the alignment algorithms or the fiducial marker location algorithms, in order to assess any possible differences in the performance of the algorithms based on the data source.

Regardless of the source of the coordinates, the first strong misalignment detection network exhibits binary behavior. This behavior indicates that each subtomogram clearly belongs to one of two categories: misaligned if there is strong misalignment present in

the tomogram, or aligned if either the misalignment is weak or the tomogram is properly aligned. As expected, almost all the strongly misaligned subtomograms are detected by the first network and are not input to the second.

However, the output of the second network presents a more continuous behavior, revealing that this second classification task is more complicated. Observing the distribution of the score histogram, there is a clear bias in the aligned set of fiducials, indicating that the network is able to detect features that are only present in this group. This bias is still evident in the weakly misaligned group, being more prominent when the source of the coordinates is the marker location algorithm.

The score distribution histograms are provided in Figure 2.14 and Figure 2.15, from which the described behaviors can be observed. The score distributions are qualitatively different so that the network is able to differentiate between both groups, with an F1 score of 0.77 when subtomogram coordinates are provided by the alignment and 0.85 if provided by the marker location algorithm, both cases using a threshold of 0.5. The F1 score is a predictive performance metric that, with a single value, indicates how well a test or model balances correctly identifying positive cases and avoiding false positives.

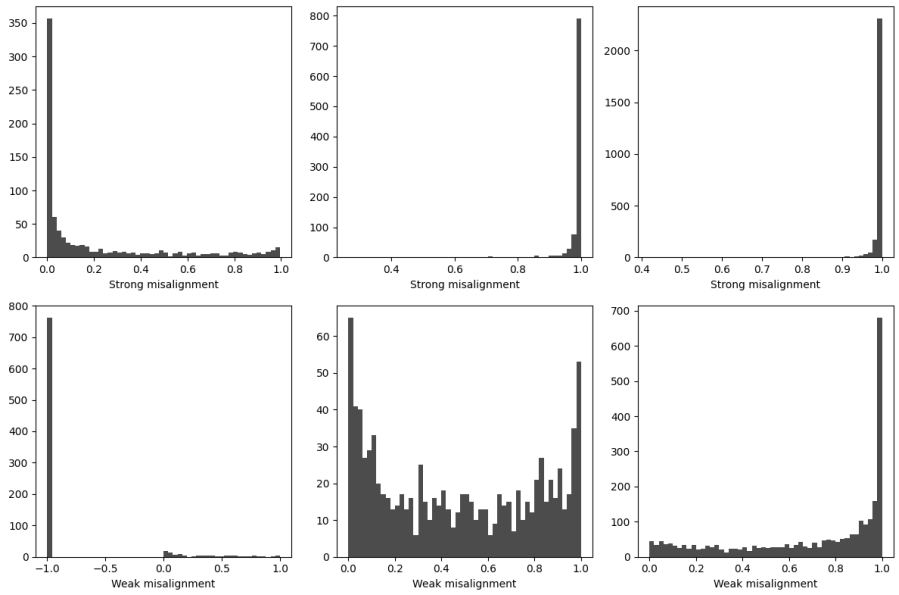


Figure 2.14: Histogram distributions of the misalignment scores retrieved by both networks. The alignment algorithm provides the coordinates of the studied markers. The first row corresponds to the scores obtained by the first network (strong misalignment) and the second row to the second network (weak misalignment). Left, central, and right columns correspond to strongly misaligned, weakly misaligned, and aligned datasets respectively. Figure referenced from de Isidro-Gómez et al. (2024).

Finally, the performance of the classification algorithm over tomograms is evaluated. From the two proposed classification results, the results are based on the average scores

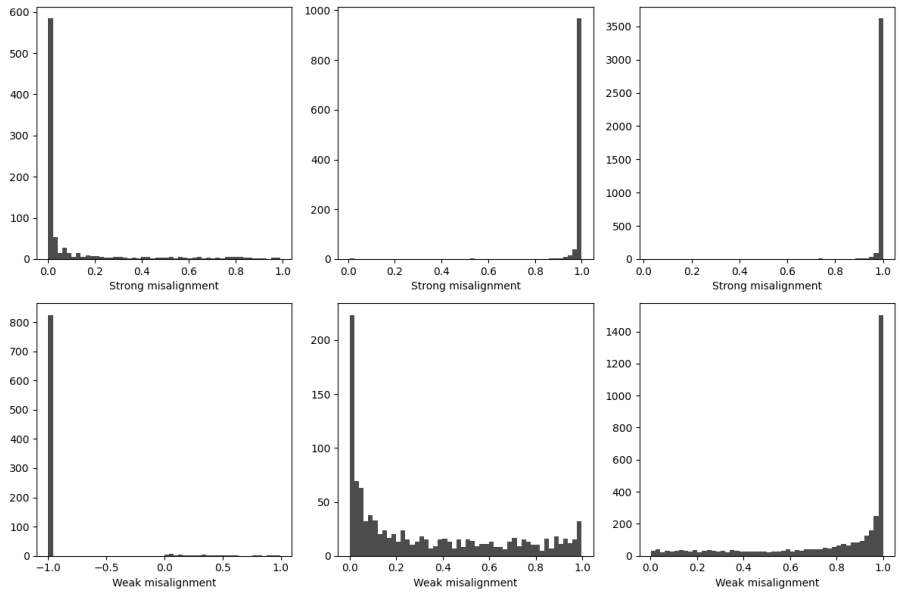


Figure 2.15: Histogram distributions of the misalignment scores retrieved by both networks. The fiducial marker location algorithm provides the coordinates of the studied markers. The first row corresponds to the scores obtained by the first network (strong misalignment) and the last one to the second network (weak misalignment). Left, central, and right columns correspond to strongly misaligned, weakly misaligned, and aligned datasets respectively. Figure referenced from de Isidro-Gómez et al. (2024).

instead of the voting system (although this is available to the scientist). We have imposed no bias in the misalignment classification, establishing a score threshold of 0.5 for both networks.

The confusion matrices from the classification of the experimental dataset are provided in Table 2.3 and Table 2.4, depending on whether the source of the fiducial coordinates is the alignment algorithms or the fiducial marker location algorithm.

From the results, it is observable that the algorithm is very robust in detecting strongly misaligned tomograms. The percentage of misclassified tomograms presenting strong misalignment is 1.63% if the coordinates come from the tilt series alignment and 1.72% if they come from the automatically detected markers. Moreover, the success rates for correctly aligned tomograms are 88.46% if coordinates come from the tilt series alignment and 91.21% if they come from the automatically detected markers, with none of them misclassified as strongly misaligned. Additionally, it is advantageous to run the algorithm in relaxed mode for alignment quality assessment so that quality filters do not remove fiducials from misaligned tomograms

The presented results validate the robustness of the classification algorithm for misalignment detection, although it is less reliable in the presence of weak misalignment. In light of these results, a deeper analysis was conducted to examine the detectability of

From alignment		Predicted		
		Aligned	Weak misalignment	Strong misalignment
Real	Aligned	161	21	0
	Weak misalignment	26	33	0
	Strong misalignment	1	4	56

Table 2.3: Confusion matrix comparing the ground truth from the tilt series and the classification obtained by the deep neural classification from the fiducial marker coordinates obtained from the tilt series alignment.

From PHC		Predicted		
		Aligned	Weak misalignment	Strong misalignment
Real	Aligned	166	16	0
	Weak misalignment	20	39	0
	Strong misalignment	1	5	52

Table 2.4: Confusion matrix comparing the ground truth from the tilt series and the classification obtained by the deep neural classification from the automatically detected fiducial markers.

subtle misalignment.

The main bias in this detection arises from the fact that the classification by individual inspection is performed on the aligned tilt series rather than on the tomograms. Consequently, subtle misalignments that are more apparent in the tilt series may be obscured in the tomographic reconstructions. This phenomenon can be observed in Figure 2.16, which presents a comparison of artifacts from these different populations.

Thus, the network is unable to discern differences in the subtracted tomograms that are observable in the aligned tilt series. To address this, we conducted a final experiment to determine whether this limitation stems from the network's learning capability or whether these differences are, in fact, imperceptible even to experienced scientists.

For this purpose, we performed a second round of classification conducted by two different scientists, this time exclusively inspecting the tomographic reconstructions. The scientists classified the reconstructions into two groups: those with misalignment and those without. Subsequently, we compared these classifications with the results obtained by the network. The outcomes of this comparison are summarized in Table 2.5.

From these results, the confusion matrix shows that both scientists exhibit a consistency of 95.36% and, when comparing their classifications with the network's estimations, the accuracy is 95.78% and 96.20%, respectively. These results were obtained by considering the union of the aligned and weakly misaligned sets from the deep neural network as aligned. This experiment highlights that, even for trained scientists, subtle misalignments

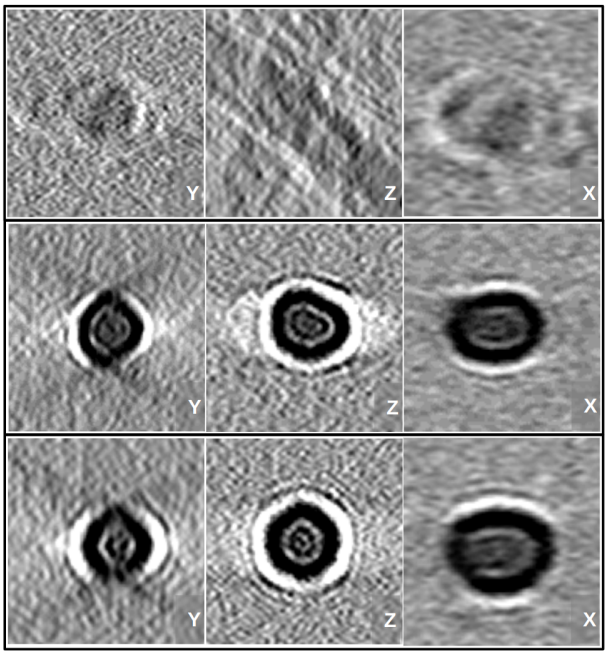


Figure 2.16: Slices of fiducial markers that exemplify the three different tomogram sets. The left column shows the XZ-plane, the center column the XY-plane, and the right column the YZ-plane. The first row presents a marker from the strongly misaligned set; the second row presents one from the weakly misaligned set, and the last row from the aligned set. Figure referenced from de Isidro-Gómez et al. (2024).

		Scientist 2	
		Aligned	Misaligned
Scientist 1	Aligned	167	5
	Misaligned	6	59
		DNN	
		Aligned + Weakly Misaligned	Strong Misaligned
Scientist 1	Aligned	169	3
	Misaligned	7	58
		DNN	
		Aligned + Weakly Misaligned	Strong Misaligned
Scientist 2	Aligned	170	3
	Misaligned	6	58

Table 2.5: Confusion matrices comparing the tomogram classification of two scientists and the deep-neural network algorithm.

in the tilt series may remain undetected in the tomographic reconstructions.

2.3.2 Alignment errors detection in tilt series

Considering the results of the previous publication and the conclusions drawn from it, our next work aims to detect misalignment in tilt series. The following developments belong to the second scientific publication of the compendium of this dissertation de Isidro-Gómez et al. (2025a).

Specifically, we focus on identifying single tilt image misalignment. Any method developed for this purpose must be capable of assessing the quality of the alignment both globally and locally for each tilt image. To achieve this, we will rely on the residual vectors introduced in Section 2.1.2. Residual-based alignment algorithms aim to minimize the reprojection error, as defined in Equation 2.7, by reducing the magnitude of the fiducials to achieve optimal alignment.

Knowing that each set of residuals provides information about the relative movements of each individual tilt series with respect to a common reference for the entire series, we will analyze the residuals to find a metric that assesses the quality of the alignment.

Similarly to the three-dimensional coordinates in the previous work, the residual vectors can be provided by the alignment algorithm or, if this information is not available, calculated subsequently. In this work, we develop a set of algorithms that allow the evaluation of alignment quality regardless of the available information. An overview of the workflow presented in this section is summarized in Figure 2.17.

If the residual vectors are provided by the alignment algorithm, the alignment quality assessment is triggered directly. If not, similar to the previous work, the three-dimensional coordinates might be provided. With this prior information, we developed algorithms to calculate residual vectors and feed quality assessment algorithms. In the unlikely event that even this information is not provided, the fiducials can be located in the tomographic reconstruction of the tilt series using the algorithms presented in the previous section. The only drawback of this case is that, unlike the previous methods, it would be necessary to calculate the tomographic reconstruction.

To assess the quality of alignment of the set of residual vectors, we propose a set of metrics designed to classify the tilt series into an alignment quality category. Each of these metrics aims to detect behaviors that, in the authors' opinion, should be characteristic enough to differentiate between aligned and misaligned tilt series. All metrics are calculated over a set of residuals to subsequently determine if they are good representatives of the alignment quality. These metrics are defined in detail in de Isidro-Gómez et al. (2025a).

1. Convex hull: to detect differences in the spatial distribution of residuals both in magnitude and directional bias by calculating the area and perimeter of the convex

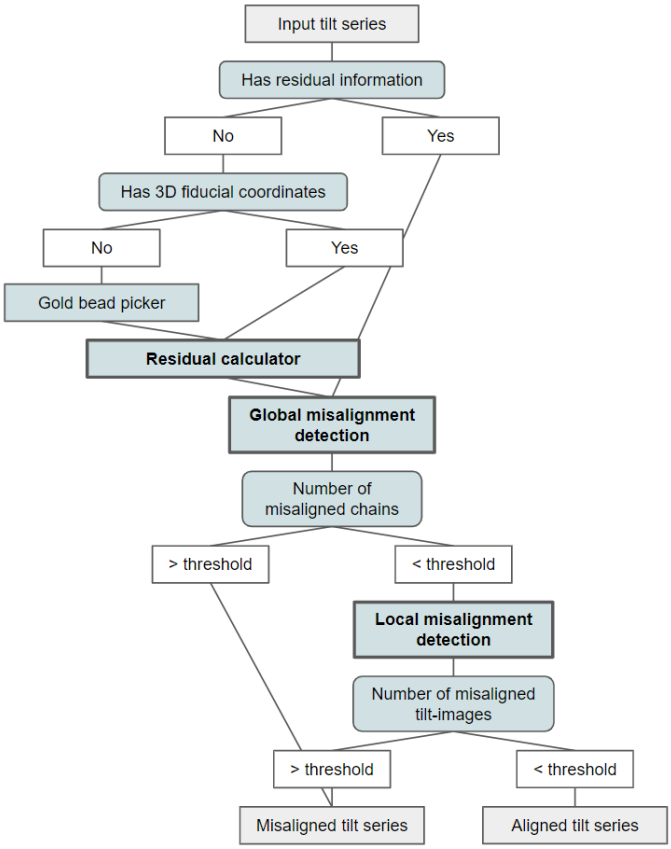


Figure 2.17: Scheme of the workflow for tilt series alignment classification. Boxes in bold highlight the steps solved by the algorithms presented in this work. Figure modified from de Isidro-Gómez et al. (2025a).

hull defined by the residual vectors Gonzalez (2009).

2. Binomial test: to detect the existence of a bias in the distribution sign of a set of residuals in any direction Zar (2010).

3. F-test: to detect differences in the directional dispersion of a population of residuals, searching for an skew in their spatial distribution Lowry (2014).

4. Augmented Dickey-Fuller test: to measure whether the residuals have a random walk distribution or exhibit some directional bias, implying a potential drift in the alignment Dickey and Fuller (1979).

5. Mahalanobis distance: Measure the conformity of each residual with the expected distribution when they originate from an aligned series Mahalanobis (2018).

From this set of proposed metrics, we aim to select one or a subset that robustly allows the assessment of alignment quality both global and per-tilt image. Nonetheless, as mentioned before, the alignment algorithm might not provide the residuals, and these must be determined in some other manner Zheng et al. (2022). Independently of the source of fiducials, the proposed metrics must properly characterize the alignment.

In case the scientist does not have the residuals, we propose an algorithm in this work to calculate them using the tilt series and the three-dimensional coordinates of the fiducials as input. This algorithm is composed of a two-step process: first, detecting the fiducials in the tilt series, a problem still open in the field Hou et al. (2024), and secondly, calculating the residuals associated with each provided three-dimensional fiducial coordinate.

In the first step of the algorithm, the set of operations described below are computed for every tilt series. Intermediate results at each step are exemplified in Figure 2.18 for a tilt image at 0° tilt (and at 60° in Figure 8 of the Supplementary Material in de Isidro-Gómez et al. (2025a)). These steps are more detailed in de Isidro-Gómez et al. (2025a).

1. Interpolation edges detection: Detect the interpolation borders after alignment using a Sobel filter. This enables the detection of the background of the image, allowing its removal in subsequent analyses, and identifies artifacts that might be present at the borders of the image (see Figure 4 from the Supplementary Material in de Isidro-Gómez et al. (2025a)).

2. Downsampling: Images are downsampled, targeting a specific size of the fiducials, to increase the signal-to-noise ratio and computational efficiency.

3. Landmark enhancement: Convolve with a fiducial kernel for enhancement. Fiducials are modeled as a two-dimensional Gaussian distribution. Finally, a rolling-ball background subtraction is applied to the image Sternberg (1983). Intermediate results of this step are shown in Figure 2.18 B.

4. Detect outlier elements: Mask out those pixels presenting z-scores lower than a given threshold from further analysis, keeping only the outlier values. Intermediate results of this step are shown in Figure 2.18 C.

   Then, the images are morphologically dilated, and a bandpass directional filter centered on the landmark target size is applied in several directions to preserve round objects (landmarks) while removing those with a high signal-to-noise ratio but presenting different shapes (such as carbon edges or membranes). Intermediate results of this step are shown in Figure 2.18 D.

5. Filter regions of interest: Mask out again the pixels presenting z-scores lower than a given threshold and subsequently label the image. Each labeled region is then morphologically analyzed:

   - Relative area: Regions significantly bigger or smaller than the expected area of the target landmark are removed.
   - Circularity: Regions whose shape differs significantly from a circle, measured as the ratio between the area of the region and the area of its circumscribed circle, are removed.

   Thus, after applying these filters, only those regions from the previous image that fulfill the morphological constraints are included as potential fiducials. Intermediate results of this step are shown in Figure 2.18 F.

6. Coordinates centering: Optionally, coordinates can be centered by the maximum shift obtained from the Fourier correlation of each landmark with its mirror.

Then, the second algorithm is fed with the locations of the fiducials detected in the tilt series and the three-dimensional coordinates of a set of fiducials provided by the alignment algorithms (or located in the tomogram). Thus, a set of residual vectors is calculated for each tilt image by applying the following series of operations for each of them.

1. Residual vector calculation: Each three-dimensional coordinate is projected onto the tilt image as defined in Equation 2.5. As illustrated in Figure 2.4, this step involves calculating the projection of the three-dimensional coordinates onto each tilt image (black dots).

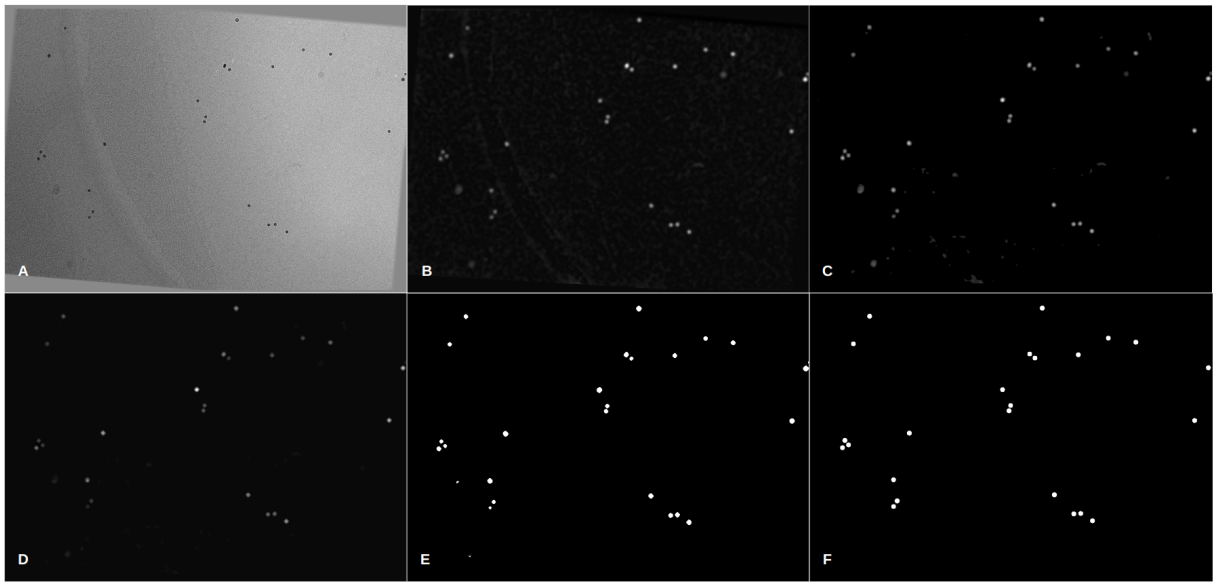


Figure 2.18: Intermediate results at different stages of the fiducial detection algorithm: (A) original tilt image, (B) result after landmark enhancement and background subtraction, (C) result after Z-score thresholding, (D) result after maximum pooling and directional filter, (E) labeled regions of interest after preprocessing, and (F) inpainting of the resulting detected landmarks after filtering the regions of interest. This tilt image belongs to tilt series E48g4_30 from EMPIAR 11457 dataset at 0°. Image referenced from de Isidro-Gómez et al. (2025a).

Then, with $\mathbf{s}_{ik}$ being the closest detected landmark in tilt image i to the projection $\mathbf{p}_{ij}$, the residual vector is defined in Equation 2.8. This is illustrated in Figure 2.4 as the vector originating from the detected projection (black dot) and ending at the closest detected landmark (empty dot).

2. Residual vector pruning: To ensure the calculation of residual vectors enables a robust assessment of the alignment quality, unreliable residual vectors are pruned. If the closest residual is missed in the detection, the end of the vector will point to the next closest one, mischaracterizing the alignment quality. This is especially problematic for aligned tilt series, as the vector modulus transitions from quasi-null to one of significant magnitude. Thus, those vectors whose origin and end regions have high correlations (both are gold beads) are removed from further analysis.

3. Metadata generation: The resulting vectors are collected, relating the three-dimensional coordinate, the landmark in each tilt image closest to each projection, and the residual vectors associating their relative positions.

Thus, if residuals are not provided, the presented algorithm will generate equivalent information to feed the alignment quality assessment algorithm. Nonetheless, it must be

ensured that the behavior of the selected metrics is equivalent regardless of the source of the fiducial. The presented metrics can be classified into three groups: statistical test-based metrics (binomial test, F test, and augmented Dickey-Fuller test), geometric metrics (perimeter and area of the convex hull), and distance metrics (Mahalanobis distance).

Whichever metric is chosen, it must be sensitive to the individual analysis of each tilt image. This implies that the results obtained must be robust even when the number of residuals available for characterization is reduced. This makes statistical measures less suitable. Additionally, the chosen metric must provide an easy threshold for automatic classification, which complicates the use of geometric measures as they are absolute. Furthermore, both the convex hull area and perimeter are sensitive to direction bias in the residuals, significantly decreasing their values if all residuals point in the same direction. An example of these limitations is shown in Figure 2.19.

In contrast, the Mahalanobis distance overcomes these limitations, as it measures the distance relative to the expected distribution of residuals,

$$D = \sqrt{(\mathbf{e} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{e} - \boldsymbol{\mu})}, \tag{2.30}$$

where $\mathbf{e}$ represents the residual vector, $\boldsymbol{\mu}$ represents the mean of the residual population, and $\boldsymbol{\Sigma}$ represents the covariance matrix. A perfect alignment implies a zero modulus for all residuals. When deviating from the ideal solution, the expected distribution is a bidimensional isotropic Gaussian distribution. This distribution has a covariance matrix,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}, \tag{2.31}$$

where $\sigma_x^2 = \sigma_y^2 = \sigma^2$, since it is isotropic. By default, a σ^2 value equal to one-third of the fiducial size is proposed, ensuring that 99.7% of the residuals fall within the fiducial radius in a Gaussian distribution. This provides a robust thresholding metric suitable for a low population of residuals.

The rationale behind why Mahalanobis distance stands out among all metrics is that it allows measurement of deviations from the expected residual distribution, both in magnitude and direction, while simplifying thresholding. This is achieved by adjusting the expected dispersion of the residuals and imposing isotropy in their distribution.

For the aforementioned reasons and based on the analysis conducted, the Mahalanobis distance yielded the most robust result for classification, enabling the classification of the tilt series. This is also shown in Figure 2.19. From these results, it can also be concluded that the residuals calculated by the proposed algorithm accurately characterize the alignment. The results obtained are equivalent to those sourced from the alignment

algorithm, demonstrating the robustness and reliability of the proposed method.

Once the Mahalanobis distance has been established as the metric of preference, it is calculated for each residual provided. Then, the quality of the alignment is globally assessed by analyzing the chains of residuals (vectors referring to the same fiducials along the series). If the percentage of chains (80% by default) whose average Mahalanobis distance exceeds 1, the series is flagged as misaligned.

If no global misalignment is detected, the tilt series is locally analyzed based on the set of residuals belonging to each tilt image. Two different criteria are proposed to assess quality: first, checking if the average Mahalanobis distance of the set of residuals exceeds 1, and secondly, a voting system that checks the percentage of fiducials (80% by default) that exceed a distance of 1. There is an extra parameter that allows for a certain number of misaligned tilt images that, if not exceeded, flags the misaligned images but keeps the series for further analysis.

Before this analysis and in case the provided residuals are significantly noisy, we allow the pruning of the residual vectors that exceed a Z-score threshold (by default 3 standard deviations).

Finally, the proposed algorithms are validated over three different public datasets: EMPIAR-10453 Turoňová et al. (2020), EMPIAR-11457 Ni et al. (2023), and EMPIAR-10364 Burt et al. (2020). This collection of datasets presents a significant variability in sampling rate and gold bead size, providing a comprehensive overview of the characteristics of available public cryo-electron tomography datasets. Results for all datasets are presented together and classifications using both sources or residuals are compared in Table 2.6 for comparison. An individual analysis for each dataset is presented in de Isidro-Gómez et al. (2025a).

Source of residuals		Predicted			
		Alignment		Automatic	
		Aligned	Misaligned	Aligned	Misaligned
Real	Aligned	292	0	285	7
	Misaligned	5	82	5	82

Table 2.6: Confusion matrix for alignment quality assessment.

The F1 score is a widely used metric for evaluating the performance of classification algorithms. For the entire dataset analyzed, the F1 score is 0.992 when residuals are provided by the alignment algorithm and 0.979 when the proposed algorithms automatically source them. For completeness, we also report the Jaccard index, which exhibits a more rapid decay, making it more sensitive to classification errors compared to the F1 score,

Figure 2.19: Metrics calculated for each tilt image from two series in the EMPIAR dataset 10453. The left column displays metrics from an aligned tilt series (TS_005), while the right column shows metrics from a misaligned tilt series (TS_233). For clarity and scale convenience, metrics are divided into three groups: the top row presents statistical tests, the middle row depicts geometrical measurements, and the bottom row illustrates Mahalanobis distances. Note that the geometrical metrics (convex hull area and perimeter) have different scales for clarity. Image referenced from de Isidro-Gómez et al. (2025a).

which demonstrates a flatter behavior. The Jaccard index is 0.983 when the alignment algorithm provides residuals and 0.959 when the proposed algorithms automatically source them.

As explained in the previous section, the F1 score quantifies the balance between correctly identifying positive cases and minimizing false positives. The Jaccard index, on the other hand, measures the similarity between two sets, specifically the actual and predicted outcomes. Both metrics are used to evaluate predictive performance.

In general, both the table and scores show that the performance of the classification algorithm is robust independently of the source of the residual vectors. False positives and false negatives are analyzed in detail in de Isidro-Gómez et al. (2025a); however, some typical patterns can be inferred.

First, we observe that the number of false positives is independent of the source of the residual vectors. In most cases, this type of misclassification arises due to limitations in characterization when the residual distribution is clustered within the sample or when the misalignment presents subtle movements that remain below the distance threshold. This is illustrated in Figures 5 and 6 of the supplementary material in de Isidro-Gómez et al. (2025a).

A noticeable difference in the classification comparison is observed in the number of false negatives, which, although minimal, is slightly higher when residual vectors are sourced from the proposed algorithms. In these cases, the algorithm suffers from a loss of contrast, particularly at high tilt angles. This results in residuals not being properly calculated for one or two tilt images, leading to the misclassification of the series. This behavior is exemplified in Figure 7 of the supplementary material in de Isidro-Gómez et al. (2025a).

Results indicate a subtle enhancement in performance when the source of the residual vectors is the alignment algorithm, especially when fiducial detection is compromised. When residual vectors are sourced from the alignment algorithm, only those involved in the alignment are used for quality assessment. This presents the advantage that if a landmark is undetected in a tilted image and excluded from the alignment, its residual vector is not reported. This is possible because some alignment algorithms use multiple partial fiducial chains without requiring a single chain covering the entire tilt series.

These algorithms have been developed aiming for a minimal configuration, mostly requiring only information known by the scientist. In summary, only a few options are of interest for modification: first, applying the robust choice for pruning the residual vectors, and second, choosing between voting or the mean as criteria to determine if the tilt series presents misalignment. Pruning residuals and the voting criterion are recommended if, when residuals are not provided by the alignment, the distribution is very noisy.

Compared to previous work, this method overcomes the limitation of subtle misalignment shadowing in tomographic reconstruction. Assessing the alignment quality directly on the tilt series allows for the detection of a reduced number of misaligned tilt images that might not produce strong enough artifacts in the reconstruction.

2.4 Other contributions

In addition to the research previously presented in this dissertation, there is a set of developments that are not focused on validation algorithms in tomography, but undoubtedly count as solid contributions to the field of cryo-electron microscopy.

In line with the last two publications of the compendium, there is a previous co-authored contribution that lays the general foundations and proposes a detailed model to account for the different effects to be considered in the tilt series alignment problem Sorzano et al. (2020).

As more general contributions, part of the work carried out during the doctorate involved software development in the Xmipp de la Rosa-Trevín et al. (2013); Strelak et al. (2021) and Scipion de la Rosa-Trevin et al. (2016) packages, specifically targeting the tomography environment ScipionTomo Jimenez de la Morena et al. (2022). These tasks included the maintenance and testing of the mentioned packages, both in tools oriented towards electron tomography and Single Particle Analysis (SPA).

In detail, regarding the tomography section, integration work has been carried out with various popular image processing packages in the field of electron tomography. In addition, contributions have been made to establish the ontology and abstractions of different data models within the ScipionTomo environment. Furthermore, other works have been conducted on the stability of CTF estimation and the development of DNN-based algorithms for stable defocus estimation in micrographs, which are currently in progress. None of these developments has led to first-authorship publications.

Chapter 3

Conclusions

The main objective of this dissertation, structured as a compendium of publications, is to address the validation of cryo-electron tomography data processing.

The research for this dissertation was approached with a general overview of the processing pipeline. During this exploratory phase, each step of the pipeline was studied to provide various validation strategies. These validation strategies can be classified into two groups: consensus approaches, where different results for the same step are compared to identify coinciding outcomes; and standalone approaches, where a single algorithm assesses the quality of the parameters estimated in that step. Thus, a comprehensive set of validation strategies was provided to scientists, allowing for more robust and stable data processing.

In addition to the algorithms and strategies developed in the initial exploratory phase, the importance of stability in the tilt series alignment step emerged as a clear target in the validation of the processing pipeline. Subsequently, during a specification process, we developed two different standalone validation algorithms, both aimed at detecting tilt series alignment errors.

The first method detects alignment errors in tomographic reconstructions using a DNN-based algorithm. The quality of the alignment is assessed by feeding the algorithm with fiducial markers that populate the sample. These markers are individually scored, and this information is then used to assess the quality of the alignment in the tomogram. For those cases where the location of these markers is not known, an algorithm for locating gold beads has also been developed, ensuring that scientists have all the necessary tools for alignment quality assessment. Finally, a detailed study of the possible limitations of this algorithm is provided, especially considering the shadowing of alignment errors due to the reconstruction process.

To overcome these limitations, a second algorithm has been designed to directly detect alignment errors in the tilt series. This algorithm follows a classical image processing

design, aiming to detect misalignment based on residual vectors, a measure of alignment quality. To ensure that scientists have all the necessary methods for alignment quality assessment, an algorithm capable of calculating these residual vectors is provided. This algorithm has demonstrated robust detection of misalignment in tilt series, even overcoming some cases where the previous algorithm failed. Similar to the previous work, this study provides an algorithm capable of calculating residual vectors to ensure scientists have the necessary methods for alignment quality assessment.

As a final remark, it is worth considering the potential applicability of the developments presented in this dissertation to other biomedical imaging domains. Although some applications are specific to cryo-electron tomography (such as validation of CTF estimation or tools based on gold bead fiducials) many of the methods address more general image processing challenges. These include alignment validation under low signal-to-noise conditions, detection of high-contrast features in noisy environments, and quality assessment of alignment based on residual vectors. Researchers from other fields are encouraged to explore and discuss the relevance and utility of these methods within their own areas of study.

3.1 Future work

The field of cryo-electron tomography is one of the main areas of evolution within electron microscopy. Therefore, the development of validation methods must be updated to the new methods and techniques that emerge in the field. In the case of tilt series alignment, the proposed tools provide good coverage of the problem, although we consider that there are possible future developments to overcome some limitations:

- We identify a significant potential for improvement in the development of new validation methods. This applies not only to the presented processing pipeline but also to new approaches in the field, such as the per particle per tilt strategy. This need is especially relevant when considering standalone methods, which are clearly lacking in the field.

- For the detection of alignment errors in tomographic reconstructions, a clear improvement would be to extend the assessment of alignment quality to fiducialless samples. The current method is trained on the gold beads present in the sample, making it unsuitable for this kind of samples. This development should be accompanied by a corresponding method capable of detecting features in the tomogram that allow for evaluation.

- Following a similar logic as in the previous step, the detection of alignment errors in tilt series could also be expanded. Although using residual vectors as an alignment quality report is not exclusive to fiducial samples, the method presented in this work for residual vector calculation is. New approaches could be explored to develop methods capable of providing a set of residual vectors for fiducialless samples.

Chapter 4

Conclusiones

El objetivo principal de esta tesis, estructurada como un compendio de publicaciones, es abordar la validación del procesamiento de datos de criotomografía electrónica.

El desarrollo de esta tesis comenzó con una aproximación general al pipeline de procesamiento. Durante esta fase exploratoria, se estudió cada paso del pipeline para proporcionar diversas estrategias de validación. Estas estrategias de validación se pueden clasificar en dos grupos: enfoques de consenso, donde se comparan diferentes resultados para el mismo paso a fin de identificar resultados coincidentes; y enfoques independientes, donde un solo algoritmo evalúa la calidad de los parámetros estimados en ese paso. De esta manera, se proporciona a los científicos un conjunto de estrategias de validación, permitiendo un procesamiento de datos más robusto y estable.

Además de los algoritmos y estrategias desarrollados en esta fase exploratoria inicial, la importancia de la estabilidad en el alineameinto de series de tilt surgió como un objetivo claro en la validación del procesamiento. Posteriormente, durante un proceso de especificación, desarrollamos dos algoritmos de validación independientes diferentes, ambos destinados a detectar errores de alineamiento de la serie de tilt.

El primer método detecta errores de alineamiento en reconstrucciones tomográficas utilizando un algoritmo basado en redes neuronales profundas (DNN). La calidad del alineamiento se evalúa alimentando el algoritmo con los marcadores fiduciarios presentes en la muestra. Estos marcadores se puntúan individualmente, y esta información se utiliza posteriormente para evaluar la calidad de la alineamiento en el tomograma. Para aquellos casos en los que no se conoce la ubicación de estos marcadores, también se ha desarrollado un algoritmo para localizar las partículas de oro, asegurando que los científicos tengan todas las herramientas necesarias para evaluar la calidad del alineamiento. Finalmente, se proporciona un estudio detallado de las posibles limitaciones de este algoritmo, especialmente considerando el encubrimiento de los errores de alineamiento debido al proceso de reconstrucción.

Para superar estas limitaciones, se ha diseñado un segundo algoritmo para detectar errores de alineamiento directamente en la serie de tilt. Este algoritmo sigue un diseño clásico de procesamiento de imagen, con el objetivo de detectar errores de alineamiento empleando vectores residuales, una medida de la calidad del alineamiento. Para asegurar que los científicos tengan todos los métodos necesarios para la evaluación de la calidad de la alineamiento, se proporciona un algoritmo capaz de calcular estos vectores residuales. Este algoritmo ha demostrado una detección robusta del desalineamiento en la serie de tilt, incluso superando algunos casos en los que el algoritmo anterior no produjo una clasificacion correcta. Similar al trabajo previo, este estudio presenta un algoritmo capaz de calcular vectores residuales, proporcionando a los científicos las herramientas necesarias para evaluar la calidad del alineamiento.

Como observación final, resulta pertinente plantear la posibilidad de aplicar los desarrollos presentados en esta tesis a otras áreas de la imagen biomédica. Aunque algunas aplicaciones son propias de la criotomografía electrónica, como la validación de la estimación del CTF o las herramientas basadas en la presencia de partículas de oro, muchos de los métodos abordan problemas generales de procesamiento de imágenes. Entre ellos se incluyen la validación de alineamiento en condiciones de baja relación señal-ruido, la detección de estructuras de alto contraste en entornos ruidosos y la evaluación de la calidad del alineamiento basada en vectores residuales. Se invita a investigadores de otras disciplinas a abrir el debate sobre la utilidad de estos métodos en sus respectivos campos de estudio.

4.1 Trabajo futuro

El campo de la tomografía electrónica es una de las principales lineas de desarrollo dentro de la microscopía electrónica. Por lo tanto, el desarrollo de métodos de validación debe actualizarse a los nuevos métodos y técnicas que surjan dentro del campo. Para el caso del alineamiento de series de tilt, las herramientas propuestas proporcionan una buena cobertura del problema, aunque consideramos que hay posibles desarrollos futuros para superar algunas limitaciones:

- Identificamos un potencial de mejora significativo en el desarrollo de nuevos métodos de validación. Esto se aplica no solo al pipeline de procesamiento presentado, sino también a nuevos enfoques en el campo, como la estrategia de *per particle per tilt*. Esta necesidad es especialmente relevante cuando se consideran métodos independientes, que claramente faltan en el campo.

- Para la detección de errores de alineamiento en reconstrucciones tomográficas, una

mejora clara sería extender la evaluación de la calidad del alineamiento a muestras sin marcadores fiduciarios. El método actual está entrenado focalizandose en las partículas de oro presentes en la muestra, lo que lo hace inadecuado para este tipo de muestras. Este desarrollo debe ir acompañado del correspondiente método capaz de detectar características en el tomograma que permitan dicha evaluación.

- Siguiendo una lógica similar a la del paso anterior, la detección de errores de alineamiento en la serie de tilt también podría expandirse. Aunque el uso de vectores residuales como informe de calidad de alineamiento no es exclusivo de las muestras con marcadores fiduciarios, el método presentado en este trabajo para el cálculo de dichos vectores sí lo es. Existe la posibilidad de explorar nuevos enfoques para desarrollar métodos capaces de proporcionar un conjunto de vectores residuales para muestras sin marcadores fiduciarios.

Chapter 5

List of publications

5.1 Compendium publications

1. **F.P. de Isidro-Gómez**, J. Vilas, E. Fernández-Giménez, O. Zarrabeitia, D. Marchán, M. Iceta, D. Herreros, P. Conesa, Y. Fonseca, J. Jiménez de la Morena, A. Cuervo, P. Losana, C. Sorzano, and J. Carazo, " Validation methods in cryo-electron tomography," in D. Hanein and N. Volkmann, editors, Cryo-electron Tomography, pp. 217–256, Elsevier, 2025.

2. **F. P. de Isidro-Gómez**, J. Vilas, P. Losana, J. Carazo, and C. O. S. Sorzano, "A deep learning approach to the automatic detection of alignment errors in cryo-electron tomographic reconstructions", Journal of Structural Biology, vol. 216, no. 1, p. 108056, 2024.

3. **F. P. de Isidro-Gómez**, J. Vilas, J. Carazo, and C. Sorzano, "Automatic detection of alignment errors in cryo-electron tomography", Journal of Structural Biology, vol. 217, no. 1, p. 108153, 2025.

5.2 Other publications

1. P. Conesa, Y. C. Fonseca, J. J. de la Morena, G. Sharov, J. M. de la Rosa-Trev´ın, A. Cuervo, A. G. Mena, B. R. de Francisco, D. Del Hoyo, D. Herreros, D. Marchan, D. Strelak, E. Fernández-Giménez, E. Ramírez-Aportela, **F. P. de Isidro-Gómez**, G. Sharov, I. Sánchez, J. Krieger, J. Jiménez de la Morena, J.L. Vilas, J.M. de la Rosa-Trevín, L. del Caño, M. Gragera, M. Iceta, M. Martínez, P. Losana, R. Melero, Y.C. Fonseca, R. Marabini, J.M. Carazo, C.O.S. Sorzano, "Scipion3: A workflow engine for cryo-electron microscopy image processing and structural biology," Biological Imaging, vol. 3, p. e13, 2023.

2. J. Jimenez de la Morena, P. Conesa, Y. Fonseca, **F. de Isidro-Gómez**, D. Herreros, E. Fernandez-Gimenez, D. Strelak, E. Moebel, T. Buchholz, F. Jug, A. Martinez-Sanchez, M. Harastani, S. Jonic, J. Conesa, A. Cuervo, P. Losana, I. Sanchez, M. Iceta, L. del Cano, M. Gragera, R. Melero, G. Sharov, D. Casta˜no-Diez, A. Koster, J. Piccirillo, J. Vilas, J. Oton, R. Marabini, C. Sorzano, and J. Carazo, "Scipiontomo: Towards cryo-electron tomography software integration, reproducibility, and validation," Journal of Structural Biology, vol. 214, no. 3, p. 107872, 2022.

3. C.O.S. Sorzano, J.L. Vilas, E. Ramírez-Aportela, D. del Hoyo, D. Herreros, E. Fernandez-Giménez, D. Marchán, **F. de Isidro-Gómez**, J.R. Macías, I. Sánchez, L. del Caño, Y. Fonseca-Reyna, P. Conesa, A. García-Mena, J. Burguet, J. García Condado, J. Méndez García, M. Martínez, A. Muñoz Barrutia, R. Marabini, J. Vargas, J.M. Carazo, "Image processing tools for the validation of cryoem maps," Faraday Discussions, vol. 240, pp. 210–227, 2022.

4. C.O.S. Sorzano, A. Jimenez-Moreno, D. Maluenda, M. Martinez, E. Ramirez-Aportela, R. Melero, A. Cuervo, J. Conesa, J. Filipovic, P.Conesa, L. del Caño, Y.C. Fonseca, J. Jiménez-de la Morena, P. Losana, R. Sáchez-García, D. Strelak, E. Fernandez-Gimenez, **F. de Isidro**, D.Herreros, J.L. Vilas, R. Marabini, J.M. Carazo, "On bias, variance, overfitting, gold standard and consensus in single-particle analysis by cryo-electron microscopy," Acta Crystallographica Section D, vol. 78, pp. 410–423, Apr 2022.

5. D. Strelak, A. Jiménez-Moreno, J.L. Vilas, E. Ramírez-Aportela, R. Sánchez-García, D. Maluenda, J. Vargas, D. Herreros, E. Fernández-Giménez, **F. de Isidro-Gómez**, J. Horacek, D. Myska, M. Horacek, P. Conesa, Y.C. Fonseca-Reyna, J. Jiménez, M. Martínez, M. Harastani, S. Jonic, J. Filipovic, R. Marabini, J.M. Carazo, C.O.S. Sorzano, "Advances in Xmipp for cryo–electron microscopy: From Xmipp to Scipion," Molecules, vol. 26, no. 20, 2021.

6. S. Sorrentino, J. Conesa, A. Cuervo, R. Melero, B. Martins, M.E. Fernández-Giménez, **F. de Isidro**, J. Jiménez, J.D. Studt, C.O.S. Sorzano, M. Eibauer, J.M. Carazo, O. Medalia. "Structural analysis of receptors and actin polarity in platelet protrusions,"Proceedings of the National Academy of Sciences, vol. 118, no. 37, p. e2105004118, 2021.

7. A. Jiménez-Moreno, L. del Caño, M. Martínez, E. Ramírez-Aportela, A. Cuervo, R. Melero, R. Sánchez-García, D. Strelak, E. Fernández-Giménez, **F. de Isidro-Gómez**, D. Herreros, P. Conesa, Y. Fonseca, D. Maluenda, J. Jiménez de la Morena, J.R.

Macías, P. Losana, R. Marabini, J.M. Carazo, C.O.S. Sorzano, "Cryo-em and single-particle analysis with scipion," JoVE (Journal of Visualized Experiments), no. 171, p. e62261, 2021.

8. C.O.S. Sorzano, D. Semchonok, S.C. Lin, J.L. Vilas, A. Jimenez-Moreno, M. Gragera, D. Maluenda, M. Martinez, E. Ramirez-Aportela, R. Melero, A. Cuervo, P. Conesa, L. del Caño, Y.C. Fonseca, R. Sánchez-García, D. Strelak, E. Fernández-Giménez, **F. de Isidro**, P. Kastritis, R. Marabini, B. Bruce, J.M. Carazo, "Algorithmic robustness to preferred orientations in single particle analysis by cryoem," Journal of Structural Biology, vol. 213, no. 1, p. 107695, 2021.

9. A. I. Gómez de Castro, M. Rheinstädter, P. Clancy, M. Castilla, F. de Isidro, J. I. Larruquert, T. de Lis-Sánchez, J. Britten, M. Cabero Piris, **F. P. de Isidro-Gómez**. "Graphite to diamond transition induced by photoelectric absorption of ultraviolet photons," Scientific Reports, vol. 11, no. 1, p. 2492, 2021.

10. C.O.S. Sorzano, A. Jiménez-Moreno, D. Maluenda, E. Ramírez-Aportela, M. Martínez, A. Cuervo, R. Melero, P. Conesa, R. Sánchez-García, D. Strelak, J. Filipovic, E. Fernández-Giménez, **F. P. de Isidro-Gómez**, D. Herreros, P. Conesa, L. del Caño, Y. Fonseca, J. Jiménez de la Morena, J.R. Macías, P. Losada, R. Marabini, J.M. Carazo, "Image processing in cryo-electron microscopy of single particles: the power of combining methods," in Structural Proteomics: High-Throughput Methods, pp. 257–289, Springer, 2021.

11. M. Martínez, A. Jiménez-Moreno, D. Maluenda, E. Ramírez-Aportela, R. Melero, A. Cuervo, P. Conesa, L. Del Caño, Y.C. Fonseca, R. Sánchez-García, D. Strelak, J.J. Conesa, E. Fernández-Giménez, **F. de Isidro**, C.O.S. Sorzano, J.M. Carazo, R. Marabini, "Integration of cryo-em model building software in scipion," Journal of Chemical Information and Modeling, vol. 60, no. 5, pp. 2533–2540, 2020.

12. C.O.S. Sorzano, **F. de Isidro-Gómez**, E. Fernández-Giménez, D. Herreros, S. Marco, J.M. Carazo, C. Messaoudi, "Improvements on marker-free images alignment for electron tomography," Journal of Structural Biology: X, vol. 4, p. 100037, 2020.

Bibliography

Abrishami, V., Vargas, J., Li, X., Cheng, Y., Marabini, R., Sorzano, C. O. S., and Carazo, J. M. (2015). Alignment of direct detection device micrographs using a robust optical flow approach. *Journal of Structural Biology*, 189:163–176.

Agulleiro, J. I. and Fernandez, J. J. (2010). Fast tomographic reconstruction on multicore computers. *Bioinformatics*, 27(4):582–583.

Agulleiro, J.-I. and Fernandez, J.-J. (2015). Tomo3d 2.0 – exploitation of advanced vector extensions (avx) for 3d reconstruction. *Journal of Structural Biology*, 189(2):147–152.

Amat, F., Castano-Diez, D., Lawrence, A., Moussavi, F., Winkler, H., and Horowitz, M. (2010). Alignment of cryo-electron tomography datasets. *Methods in enzymology*, 482:343–367.

Andersen, A. and Kak, A. (1984). Simultaneous algebraic reconstruction technique (sart): A superior implementation of the art algorithm. *Ultrasonic Imaging*, 6(1):81–94.

Bai, X., Fernandez, I., McMullan, G., and S.HW., S. (2013). Robust image alignment for cryogenic transmission electron microscopy. *eLife*, 2:e00461.

Bepler, T., Kelley, K., Noble, A. J., and Berger, B. (2020). Topaz-denoise: general deep denoising models for cryoem and cryoet. *Nature communications*, 11(1):5208.

Bharat, T. A. and Scheres, S. H. (2016). Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in relion. *Nature protocols*, 11(11):2054–2065.

Brilot, A. F., Chen, J. Z., Cheng, A., Pan, J., Harrison, S. C., Potter, C. S., Carragher, B., Henderson, R., and Grigorieff, N. (2012). Beam-induced motion of vitrified specimen on holey carbon film. *Journal of Structural Biology*, 177(3):630–637.

Bryant, P., Pozzati, G., and Elofsson, A. (2022). Improved prediction of protein-protein interactions using alphafold2. *Nature communications*, 13(1):1265.

Buchholz, T.-O., Jordan, M., Pigino, G., and Jug, F. (2019a). Cryo-care: content-aware image restoration for cryo-transmission electron microscopy data. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 502–506. IEEE.

Buchholz, T.-O., Krull, A., Shahidi, R., Pigino, G., Jékely, G., and Jug, F. (2019b). Content-aware image restoration for electron microscopy. *Methods in cell biology*, 152:277–289.

Buckley, G., Ramm, G., and Trépout, S. (2024). Golddigger and checkers, computational developments in cryo-scanning transmission electron tomography to improve the quality of reconstructed volumes. *Biological Imaging*, 4:e6.

Burt, A., Cassidy, C. K., Ames, P., Bacia-Verloop, M., Baulard, M., Huard, K., Luthey-Schulten, Z., Desfosses, A., Stansfeld, P. J., Margolin, W., et al. (2020). Complete structure of the chemosensory array core signalling unit in an e. coli minicell strain. *Nature communications*, 11(1):743.

Cardone, G., Grünewald, K., and Steven, A. C. (2005). A resolution criterion for electron tomography based on cross-validation. *Journal of Structural Biology*, 151(2):117–129.

Cardone, G., Heymann, J., and Steven, A. (2013). One number does not fit all: Mapping local variations in resolution in cryo-EM reconstructions. *Journal of Structural Biology*, 184:226–236.

Castano-Diez, D., Kudryashev, M., and Stahlberg, H. (2017). Dynamo catalogue: Geometrical tools and data management for particle picking in subtomogram averaging of cryo-electron tomograms. *Journal of Structural Biology*, 197(2):135–144. Electron Tomography.

Castano-Diez, D., Scheffer, M., Al-Amoudi, A., and Frangakis, A. S. (2010). Alignator: A gpu powered software package for robust fiducial-less alignment of cryo tilt-series. *Journal of Structural Biology*, 170(1):117–126.

Castaño-Díez, D., Kudryashev, M., Arheit, M., and Stahlberg, H. (2012). Dynamo: A flexible, user-friendly development tool for subtomogram averaging of cryo-em data in high-performance computing environments. *Journal of Structural Biology*, 178(2):139–151. Special Issue: Electron Tomography.

Chen, M., Dai, W., Sun, S. Y., Jonasch, D., He, C. Y., Schmid, M. F., Chiu, W., and Ludtke, S. J. (2017). Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. *Nature Methods*, 14(10):983–985.

Chen, S., McMullan, G., Faruqi, A., Murshudov, G., Short, J., Scheres, S., and Henderson, R. (2013). High-resolution noise substitution to measure overfitting and validate resolution in 3d structure determination by single particle electron cryomicroscopy. *Ultramicroscopy*, 135:24–35.

Danev, R. and Baumeister, W. (2017). Expanding the boundaries of cryo-em with phase plates. *Current Opinion in Structural Biology*, 46:87–94.

de Isidro-Gómez, F. P., Vilas, J., Losana, P., Carazo, J., and Sorzano, C. O. S. (2024). A deep learning approach to the automatic detection of alignment errors in cryo-electron tomographic reconstructions. *Journal of Structural Biology*, 216(1):108056.

de Isidro-Gómez, F., Vilas, J., Carazo, J., and Sorzano, C. (2025a). Automatic detection of alignment errors in cryo-electron tomography. *Journal of Structural Biology*, 217(1):108153.

de Isidro-Gómez, F., Vilas, J., Fernández-Giménez, E., Zarrabeitia, O., Marchán, D., Iceta, M., Herreros, D., Conesa, P., Fonseca, Y., Jiménez de la Morena, J., Cuervo, A., Losana, P., Sorzano, C., and Carazo, J. (2025b). Validation methods in cryo-electron tomography. In Hanein, D. and Volkmann, N., editors, *Cryo-electron Tomography*, pages 217–256. Elsevier.

de la Rosa-Trevin, J., Quintana, A., del Cano, L., Zaldivar-Peraza, A., Foche, I., Gutierrez, J., Gomez-Blanco, J., Burguet-Castells, J., Cuenca, J., Abrishami, V., Vargas, J., Oton, J., Sharov, G., Navas, J., Conesa, P., Vilas, J., Marabini, R., Sorzano, C., and Carazo, J. (2016). Scipion: a software framework toward integration, reproducibility, and validation in 3d electron microscopy. *Journal of Structural Biology*, 195:93–99.

de la Rosa-Trevín, J. M., Otón, J., Marabini, R., Zaldívar, A., Vargas, J., Carazo, J. M., and Sorzano, C. O. S. (2013). Xmipp 3.0: an improved software suite for image processing in electron microscopy. *Journal of Structural Biology*, 184(2):321–328.

de Teresa, I., Goetz, S., Mattausch, A., Stojanovska, F., Zimmerli, C., Toro-Nahuelpan, M., Cheng, D. W., Tollervey, F., Pape, C., Beck, M., Kreshuk, A., Mahamid, J., and Zaugg, J. (2022). Convolutional networks for supervised mining of molecular patterns within cellular context.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431.

Diebolder, C., Faas, F., Koster, A., and Koning, R. (2015). Conical fourier shell correlation applied to electron tomograms. *Journal of Structural Biology*, 190(2):215–223.

Dimchev, G., Amiri, B., Fäßler, F., Falcke, M., and Schur, F. K. (2021). Computational toolbox for ultrastructural quantitative analysis of filament networks in cryo-et data. *Journal of Structural Biology*, 213(4):107808.

Fernandez, J.-J. and Li, S. (2021). Tomoalign: A novel approach to correcting sample motion and 3d ctf in cryoet. *Journal of Structural Biology*, 213(4):107778.

Fernandez, J.-J., Li, S., Bharat, T. A., and Agard, D. A. (2018). Cryo-tomography tilt-series alignment with consideration of the beam-induced sample motion. *Journal of Structural Biology*, 202(3):200–209.

Fernández, J., Li, S., and Crowther, R. (2006). Ctf determination and correction in electron cryotomography. *Ultramicroscopy*, 106(7):587–596.

Frangakis, A. S. (2021). It's noisy out there! a review of denoising techniques in cryo-electron tomography. *Journal of Structural Biology*, 213(4):107804.

Frangakis, A. S. (2022). Mean curvature motion facilitates the segmentation and surface visualization of electron tomograms. *Journal of Structural Biology*, 214(1):107833.

Frangakis, A. S. and Hegerl, R. (2001). Noise reduction in electron tomographic reconstructions using nonlinear anisotropic diffusion. *Journal of Structural Biology*, 135(3):239–250.

Galaz-Montoya, J. G., Flanagan, J., Schmid, M. F., and Ludtke, S. J. (2015). Single particle tomography in eman2. *Journal of Structural Biology*, 190(3):279–290.

Galaz-Montoya, J. G., Hecksel, C. W., Baldwin, P. R., Wang, E., Weaver, S. C., Schmid, M. F., Ludtke, S. J., and Chiu, W. (2016). Alignment algorithms and per-particle ctf correction for single particle cryo-electron tomography. *Journal of structural biology*, 194(3):383–394.

Galaz-Montoya, J. G. and Ludtke, S. J. (2017). The advent of structural biology in situ by single particle cryo-electron tomography. *Biophysics reports*, 3:17–35.

Gilbert, P. (1972). Iterative methods for the three-dimensional reconstruction of an object from projections. *Journal of Theoretical Biology*, 36(1):105–117.

Gonzalez, R. C. (2009). *Digital image processing*. Pearson education india.

Gordon, R., Bender, R., and Herman, G. (1970). Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography. *Journal of Theoretical Biology*, 29(3):471–481.

Grant, T. and Grigorieff, N. (2015). Measuring the optimal exposure for single particle cryo-EM using a 2.6å reconstruction of rotavirus VP6. *elife*, 4.

Hagen, W. J., Wan, W., and Briggs, J. A. (2017). Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *Journal of structural biology*, 197(2):191–198.

Harauz, G. and van Heel, M. (1986). Exact filters for general geometry three dimensional reconstruction. *Optik*, 73:146–156.

Henderson, R. (2013). Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences*, 110(45):18037–18041.

Heymann, B. (2018). Single particle reconstruction and validation using bsoft for the map challenge. *Journal of Structural Biology*, 204(1):90–95.

Heymann, J. B., Cardone, G., Winkler, D. C., and Steven, A. C. (2008). Computational resources for cryo-electron tomography in bsoft. *Journal of structural biology*, 161(3):232–242.

Heynmann, B. (2015). Validation of 3dem reconstructions: The phantom in the noise. *AIMS Biophysics*, 2:21.

Himes, B. and Zhang, P. (2018). emclarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nature Methods*, 15(11):955–961.

Hou, G., Yang, Z., Zang, D., Fernández, J.-J., Zhang, F., and Han, R. (2024). Markerdetector: A method for robust fiducial marker detection in electron micrographs using wavelet-based template. *Journal of Structural Biology*, 216(1):108044.

Hu, M., Zhang, Q., Yang, J., and Li, X. (2020). Unit quaternion description of spatial rotations in 3d electron cryo-microscopy. *Journal of Structural Biology*, 212(3):107601.

Huang, X., Li, S., and Gao, S. (2018). Applying a modified wavelet shrinkage filter to improve cryo-electron microscopy imaging. *Journal of Computational Biology*, 25(9):1050–1058.

Jimenez de la Morena, J., Conesa, P., Fonseca, Y., de Isidro-Gomez, F., Herreros, D., Fernandez-Gimenez, E., Strelak, D., Moebel, E., Buchholz, T., Jug, F., Martinez-Sanchez, A., Harastani, M., Jonic, S., Conesa, J., Cuervo, A., Losana, P., Sanchez, I., Iceta, M., del Cano, L., Gragera, M., Melero, R., Sharov, G., Castaño-Diez, D., Koster, A., Piccirillo, J., Vilas, J., Oton, J., Marabini, R., Sorzano, C., and Carazo, J. (2022). Scipiontomo: Towards cryo-electron tomography software integration, reproducibility, and validation. *Journal of Structural Biology*, 214(3):107872.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021). Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589.

Khoshouei, M., Pfeffer, S., Baumeister, W., Förster, F., and Danev, R. (2017). Subtomogram analysis using the volta phase plate. *Journal of structural biology*, 197(2):94–101.

Kirkland, E. (2010). *Advanced Computing in Electron Microscopy*. Springer.

Koster, A. J., Grimm, R., Typke, D., Hegerl, R., Stoschek, A., Walz, J., and Baumeister, W. (1997). Perspectives of molecular and cellular electron tomography. *Journal of structural biology*, 120(3):276–308.

Kremer, J. R., Mastronarde, D. N., and McIntosh, J. (1996). Computer visualization of three-dimensional image data using imod. *Journal of Structural Biology*, 116(1):71–76.

Kucukelbir, A., Sigworth, F. J., and Tagare, H. D. (2014). Quantifying the local resolution of cryo-em density mapss. *Nature Methods*, 11:63–65.

Lamm, L., Righetto, R. D., Wietrzynski, W., Pöge, M., Martinez-Sanchez, A., Peng, T., and Engel, B. D. (2022a). Membrain: A deep learning-aided pipeline for detection of membrane proteins in cryo-electron tomograms. *Computer methods and programs in biomedicine*, 224:106990.

Lamm, L., Righetto, R. D., Wietrzynski, W., Pöge, M., Martinez-Sanchez, A., Peng, T., and Engel, B. D. (2022b). Membrain: A deep learning-aided pipeline for detection of membrane proteins in cryo-electron tomograms. *Computer Methods and Programs in Biomedicine*, 224:106990.

Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., Agard, D. A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods*, 10(6):584–590.

Lowry, R. (2014). Concepts and applications of inferential statistics. *Materials Sciences and Applications*, 6(6).

Lučić, V., Rigort, A., and Baumeister, W. (2013). Cryo-electron tomography: The challenge of doing structural biology in situ. *Journal of Cell Biology*, 202(3):407–419.

Mahalanobis, P. C. (2018). On the generalized distance in statistics. *Sankhyā: The Indian Journal of Statistics, Series A (2008-)*, 80:S1–S7.

Martinez-Sanchez, A., Baumeister, W., and Lučić, V. (2022). Statistical spatial analysis for cryo-electron tomography. *Computer Methods and Programs in Biomedicine*, 218.

Martinez-Sanchez, A., Garcia, I., Asano, S., Lucic, V., and Fernandez, J.-J. (2014). Robust membrane detection based on tensor voting for electron tomography. *Journal of Structural Biology*, 186(1):49–61.

Martinez-Sanchez, A., Kochovski, Z., Laugks, U., Meyer zum Alten Borgloh, J., Chakraborty, S., Pfeffer, S., W., B., and Lucic, V. (2020). Template-free detection and classification of membrane-bound complexes in cryo-electron tomograms. *Nature Methods*, 17:209–2016.

Mastronarde, D. N. and Held, S. R. (2017). Automated tilt series alignment and tomographic reconstruction in imod. *Journal of Structural Biology*, 197(2):102–113. Electron Tomography.

McMullan, G., Faruqi, A., and Henderson, R. (2016). Chapter one - direct electron detectors. In Crowther, R., editor, *The Resolution Revolution: Recent Advances In cryoEM*, volume 579 of *Methods in Enzymology*, pages 1–17. Academic Press.

Messaoudi, C., Boudier, T., Sorzano, C., and Marco, S. (2007). Tomoj: tomography software for three-dimensional reconstruction in transmission electron microscopy. *BMC Bioinformatics*, 8(288).

Moreno, J. J., Martinez-Sanchez, A., Martinez, J. A., Garzon, E. M., and Fernandez, J. J. (2018). TomoEED: fast edge-enhancing denoising of tomographic volumes. *Bioinformatics*, 34(21):3776–3778.

Navarro, P. P. (2022). Quantitative cryo-electron tomography. *Frontiers in Molecular Biosciences*, 9.

Ni, T., Mendonça, L., Zhu, Y., Howe, A., Radecke, J., Shah, P. M., Sheng, Y., Krebs, A.-S., Duyvesteyn, H. M., Allen, E., et al. (2023). Chadox1 covid vaccines express rbd open prefusion sars-cov-2 spikes on the cell surface. *Iscience*, 26(10).

Pelberg, R. and Pelberg, R. (2015). Concepts in radiation and radiation safety. *Cardiac CT Angiography Manual*, pages 1–17.

Radermacher, M. (1992). Weighted back-projection methods. In Frank, J., editor, *Electron Tomography: Three-Dimensional Imaging with the Transmission Electron Microscope*, pages 91–115. Springer US, Boston, MA.

Ramirez-Aportela, E., Mota, J., Conesa, P., Carazo, J., and Sorzano, C. (2019). *DeepRes*: a new deep-learning- and aspect-based local resolution method for electron-microscopy maps. *IUCrJ*, 6(6).

Robertson, M. J., Meyerowitz, J. G., and Skiniotis, G. (2020). Cryo-em as a powerful tool for drug discovery. *Bioorganic and Medicinal Chemistry Letters*, 30(22):127524.

Rohou, A. and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, 192(2):216–221.

Rosenthal, P. and Henderson, H. (2003). Determination of particle orientation,absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology*, 333:721–745.

Salfer, M., Collado, J. F., Baumeister, W., Fernández-Busnadiego, R., and Martínez-Sánchez, A. (2020). Reliable estimation of membrane curvature for cryo-electron tomography. *PLOS Computational Biology*, 16(8):e1007962.

Sazzed, S., Scheible, P., He, J., and Wriggers, W. (2023). Untangling irregular actin cytoskeleton architectures in tomograms of the cell with struwwel tracer. *International Journal of Molecular Sciences*, 24(24):17183.

Scaramuzza, S. and Castaño-Díez, D. (2021). Step-by-step guide to efficient subtomogram averaging of virus-like particles with dynamo. *PLoS Biology*, 19(8):e3001318.

Scheres, S. (2014). Beam-induced motion correction for sub-megadalton cryo-em particles. *eLife*, 3:e03665.

Schur, F. K., Obr, M., Hagen, W. J., Wan, W., Jakobi, A. J., Kirkpatrick, J. M., Sachse, C., Kräusslich, H.-G., and Briggs, J. A. (2016). An atomic model of hiv-1 capsid-sp1 reveals structures regulating assembly and maturation. *Science*, 353(6298):506–508.

Seifer, S. and Elbaum, M. (2022). Clusteralign: A fiducial tracking and tilt series alignment tool for thick sample tomography. *Biological Imaging*, 2:e7.

Sorzano, C., de Isidro-Gómez, F., Fernández-Giménez, E., Herreros, D., Marco, S., Carazo, J. M., and Messaoudi, C. (2020). Improvements on marker-free images alignment for electron tomography. *Journal of Structural Biology: X*, 4:100037.

Sorzano, C., Jonic, S., Núñez-Ramírez, R., Boisset, N., and Carazo, J. (2007). Fast, robust, and accurate determination of transmission electron microscopy contrast transfer function. *Journal of Structural Biology*, 160(2):249–262.

Sorzano, C., Vargas, J., Oton, J., Abrishami, V., de la Rosa-Trevin, J., Gomez-Blanco, J., Vilas, J., Marabini, R., and Carazo, J. (2017a). A review of resolution measures and related aspects in 3d electron microscopy. *Progress in biophysics and molecular biology*, 124:1–30.

Sorzano, C., Vargas, J., Oton, J., de la Rosa-Trevin, J., Vilas, J., Kazemi, M., Melero, R., del cano, L., Cuenca, J., Gomez-Blanco, J., and Carazo, J. (2017b). A survey of the use of iterative reconstruction algorithms in electron microscopy. *BioMed Research International*, 2017.

Sorzano, C. O. S., Jiménez-Moreno, A., Maluenda, D., Martínez, M., Ramírez-Aportela, E., Krieger, J., Melero, R., Cuervo, A., Conesa, J., Filipovic, J., Conesa, P., del Caño, L., Fonseca, Y. C., Jiménez-de la Morena, J., Losana, P., Sánchez-García, R., Strelak, D., Fernández-Giménez, E., de Isidro-Gómez, F. P., Herreros, D., Vilas, J. L., Marabini, R., and Carazo, J. M. (2022). On bias, variance, overfitting, gold standard and consensus in single-particle analysis by cryo-electron microscopy. *Acta Crystallographica Section D*, 78(4):410–423.

Sorzano, C. O. S., Vargas, J., de la Rosa-Trevín, J. M., Zaldívar-Peraza, A., Otón, J., Abrishami, V., Foche, I., Marabini, R., Caffarena, G., and Carazo, J. M. (2014). Outlier detection for single particle analysis in electron microscopy. In *Proc. Intl. Work-Conference on Bioinformatics and Biomedical Engineering, IWBBIO*, page 950.

Sternberg, S. R. (1983). Biomedical image processing. *Computer*, 16(01):22–34.

Strelak, D., Jiménez-Moreno, A., Vilas, J. L., Ramírez-Aportela, E., Sánchez-García, R., Maluenda, D., Vargas, J., Herreros, D., Fernández-Giménez, E., de Isidro-Gómez, F. P., Horacek, J., Myska, D., Horacek, M., Conesa, P., Fonseca-Reyna, Y. C., Jiménez, J., Martínez, M., Harastani, M., Jonić, S., Filipovic, J., Marabini, R., Carazo, J. M., and Sorzano, C. O. S. (2021). Advances in xmipp for cryo–electron microscopy: From xmipp to scipion. *Molecules*, 26(20).

Strelák, D., Filipovic, J., Jimenez-Moreno, A., Carazo, J. M., and Sanchez Sorzano, C. O. (2020). Flexalign: An accurate and fast algorithm for movie alignment in cryo-electron microscopy. *Electronics*, 9(6).

Su, M. (2019). goctf: Geometrically optimized ctf determination for single-particle cryo-em. *Journal of Structural Biology*, 205(1):22–29.

Tan, Y. Z., Baldwin, P. R., Davis, J. H., Williamson, J. R., Potter, C. S., Carragher, B., and Lyumkis, D. (2017). Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nature Methods*, 14(8):793–796.

Tegunov, D., Xue, L., Dienemann, C., Cramer, P., and Mahamid, J. (2021). Multi-particle cryo-em refinement with m visualizes ribosome-antibiotic complex at 3.5a in cells. *Nature Methods*, 18(2):186–193.

Thon, F. (1966). Zur defokussierunsabhängigkeit des phasenkontrastes bei der elektronenmikroskopischen abbildung. *Z. Naturforsch*, 21a:476–478.

Turonova, B., Hagen, W., Obr, M., Mosalaganti, S., Beugelink, J., Zimmerli, C., H., K., and Beck, M. (2020). Benchmarking tomographic acquisition schemes for high-resolution structural biology. *Nature Communications*, 11(876).

Turoňová, B., Sikora, M., Schürmann, C., Hagen, W. J., Welsch, S., Blanc, F. E., von Bülow, S., Gecht, M., Bagola, K., Hörner, C., et al. (2020). In situ structural analysis of sars-cov-2 spike reveals flexibility mediated by three hinges. *Science*, 370(6513):203–208.

Turoňová, B., Schur, F. K., Wan, W., and Briggs, J. A. (2017). Efficient 3d-ctf correction for cryo-electron tomography using novactf improves subtomogram averaging resolution to 3.4Å. *Journal of Structural Biology*, 199(3):187–195.

Unwin, P. N. T., Huxley, H. E., and Klug, A. (1971). Phase contrast and interference microscopy with the electron microscope. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 261(837):95–104.

Van Drie, J. H. and Tong, L. (2022). Drug discovery in the era of cryo-electron microscopy. *Trends in Biochemical Sciences*, 47(2):124–135.

van Heel, M. (2013). Finding trimeric hiv-1 envelope glycoproteins in random noise. *Proceedings of the National Academy of Sciences*, 110(45):E4175–E4177.

Vargas, J., Oton, J., Marabini, R., Carazo, J., and Sorzano, C. (2016). Particle alignment reliability in single particle electron cryomicroscopy: a general approach. *Scientific Reports*, 6:21626.

Vilas, J., Heymann, J., Tagare, H., Ramirez-Aportela, E., Carazo, J., and Sorzano, C. (2020a). Local resolution estimates of cryoem reconstructions. *Current Opinion in Structural Biology*, 64:74–78.

Vilas, J. and Tagare (2023). New measures of anisotropy of cryo-em maps. *Nature Methods*, in press.

Vilas, J. L., Carazo, J. M., and Sorzano, C. O. S. (2022). Emerging themes in cryoem-single particle analysis image processing. *Chemical Reviews*, 122(17):13915–13951.

Vilas, J. L., Gomez-Blanco, J., Conesa, P., Melero, R., de la Rosa Trevın, J. M., Oton, J., Cuenca, J., Marabini, R., Carazo, J. M., Vargas, J., and Sorzano, C. O. S. (2018). Monores: automatic and accurate estimation of local resolution for electron microscopy maps. *Structure*, 26:337–344.

Vilas, J. L., Oton, J., Messaoudi, C., Melero, R., Conesa, P., Ramirez-Aportela, E., Mota, J., Martinez, M., Jimenez, A., Marabini, R., Carazo, J. M., Vargas, J., and Sorzano, C. O. S. (2020b). Measurement of local resolution in electron tomography. *Journal of Structural Biology X*, 4:100016.

Vilas, J. L., Tagare, H. D., Vargas, J., Carazo, J. M., and Sorzano, C. O. S. (2020c). Measuring local-directional resolution and local anisotropy in cryo-EM maps. *Nature communications*, 11:55.

Voortman, L. M., Stallinga, S., Schoenmakers, R. H., van Vliet, L. J., and Rieger, B. (2011). A fast algorithm for computing and correcting the ctf for tilted, thick specimens in tem. *Ultramicroscopy*, 111(8):1029–1036.

Vulović, M., Voortman, L. M., van Vliet, L. J., and Rieger, B. (2014). When to use the projection assumption and the weak-phase object approximation in phase contrast cryo-em. *Ultramicroscopy*, 136:61–66.

Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., Raisch, T., Prumbaum, D., Quentin, D., Roderer, D., Tacke, S., Siebolds, B., Schubert, E., Shaikh, T. R., Lill, P., Gatsogiannis, C., and Raunser, S. (2019). Sphire-cryolo is a fast and accurate fully automated particle picker for cryo-em. *Communications Biology*, 2(1):218.

Wan, W. and Briggs, J. (2016). Chapter thirteen - cryo-electron tomography and subtomogram averaging. In Crowther, R., editor, *The Resolution Revolution: Recent Advances In cryoEM*, volume 579 of *Methods in Enzymology*, pages 329–367. Academic Press.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3(1):1–9.

Wu, G.-H., Smith-Geater, C., Galaz-Montoya, J. G., Gu, Y., Gupte, S. R., Aviner, R., Mitchell, P. G., Hsu, J., Miramontes, R., Wang, K. Q., et al. (2023). Cryoet reveals organelle phenotypes in huntington disease patient ipsc-derived and mouse primary neurons. *Nature communications*, 14(1):692.

Xiong, Q., Morphew, M. K., Schwartz, C. L., Hoenger, A. H., and Mastronarde, D. N. (2009). Ctf determination and correction for low dose tomographic tilt series. *Journal of Structural Biology*, 168(3):378–387.

Xu, M., Chai, X., Muthakana, H., Liang, X., Yang, G., Zeev-Ben-Mordehai, T., and Xing, E. P. (2017). Deep learning-based subdivision approach for large scale macromolecules structure recovery from electron cryo tomograms. *Bioinformatics*, 33(14):i13–i22.

Xu, Y. and Dang, S. (2022). Recent technical advances in sample preparation for single-particle cryo-em. *Frontiers in molecular biosciences*, 9:892459.

Zar, J. (2010). *Biostatistical Analysis*. Prentice Hall.

Zeng, X., Kahng, A., Xue, L., Mahamid, J., Chang, Y.-W., and Xu, M. (2023). High-throughput cryo-et structural pattern mining by unsupervised deep iterative subtomogram clustering. *Proceedings of the National Academy of Sciences*, 120(15):e2213149120.

Zhang, K. (2016). Gctf: Real-time ctf determination and correction. *Journal of Structural Biology*, 193(1):1–12.

Zheng, S., Wolff, G., Greenan, G., Chen, Z., Faas, F. G., Bárcena, M., Koster, A. J., Cheng, Y., and Agard, D. A. (2022). Aretomo: An integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction. *Journal of Structural Biology: X*, 6:100068.

Zheng, S. Q., Palovcak, E., Armache, J.-P., Verba, K. A., Cheng, Y., and Agard, D. A. (2017). Motioncor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*, 14:331–332.

Zivanov, J., Otón, J., Ke, Z., von Kügelgen, A., Pyle, E., Qu, K., Morado, D., Castano-Diez, D., Zanetti, G., Bharat, T. A., Briggs, J. A., and Scheres, S. H. (2022). A bayesian approach to single-particle electron cryo-tomography in relion-4.0. *eLife*, 11:e83724.

Appendices

Appendix A: Validation methods in cryo-electron tomography

CHAPTER 9

Validation methods in cryo-electron tomography

F.P. de Isidro-Gómez[1,2], J.L. Vilas[1], E. Fernández-Giménez[1,2], O.L. Zarrabeitia[1], D. Marchán[1,2], M. Iceta[1], D. Herreros[1,2], P. Conesa[1], Y.C. Fonseca[1], J. Jiménez de la Morena[1], A. Cuervo[1], P. Losana[1], C.O.S. Sorzano[1] and J.M. Carazo[1]

[1]Biocomputing Unit, Centro Nacional de Biotecnologia (CNB-CSIC), Campus Universidad Autonoma, Madrid, Spain; [2]University Autonoma de Madrid, Madrid, Spain

Introduction

Structural biology aims to elucidate the three-dimensional structure of cellular and subcellular components. The knowledge of the structure allows the prediction of the working mechanisms of the proteins in the cell environment, to infer their behavior, and therefore, to understand the biology of the cell. This has a great impact beyond molecular biology. For instance, it allows drug development (Robertson et al., 2020; Van Drie et al., 2022) based on structure, attempting to maximize the interaction with the target protein and minimize the interaction with others. Various structural techniques exist, such as X-ray diffraction, nuclear magnetic resonance, or cryo-electron microscopy (cryoEM). Each technique provides information of a very different nature, and not all of them can achieve the same reconstruction quality. The selection of the imaging technique depends on the information to be extracted from the sample. Electron tomography is a branch of cryoEM to study the cell and its components in their native state. A cryo-electron tomography (cryoET) project comprises different stages. Coarsely, they are sample preparation, data acquisition, and image processing.

The sample preparation stage comprehends the steps to achieve a sample ready to be introduced and studied with a transmission electron microscope. It is composed of the biochemical process to purify the sample if it is the case, the grid preparation, vitrification, and an FIB-milling process to

Cryo-electron Tomography
ISBN 978-0-443-18829-9
https://doi.org/10.1016/B978-0-443-18829-9.00009-8

reduce the sample thickness if necessary (Thompson et al., 2016). The vitrification is critical, the objective is to freeze the sample embedding it into a volume of amorphous ice. Plunge freezing and high-pressure freezing are the most popular vitrification approaches. However, thickness is a limiting factor in cryoET (Navarro, 2022); the ice layer must be as thin as possible. While in other cryo-electron microscopy (cryoEM) techniques, such as in Single-Particle Analysis, the specimen's size is as much as a few nanometers, in cryoET, the complexes such as cells are far from that range (Lučić et al., 2013). In those cases, the thickness has to be reduced. The most used approach is the FIB-milling process. This method is a sputtering technique that removes material from the sample to produce a thin lamella. The prepared sample is then introduced into the electron microscope to acquire the images. We will now focus on image processing, assuming that the sample preparation stage and image acquisition have been successfully performed. This third step allows for obtaining the three-dimensional representation of the sample.

A detailed explanation of the whole image processing pipeline is given further. In brief, the images are extremely noisy, making image processing a real challenge. Hence, validating the results is a pressing need. In this work, instead of reviewing the image processing algorithms for cryoET, we analyze the currently existing validation tools for this imaging technology. We will also propose other tools that could be implemented in the future. In our presentation of tools, we state the problem encountered at each step of the image processing workflow. We then explain the state of the art and possible future validation tools for each problem. Finally, we discuss the emerging validation topics that can appear in the near future.

The tools listed in this chapter complement the current heuristic approximations for cryoET validation used by different researchers. However, these latter depend on the particular skills of the scientist. Some of the most popular ones are visual inspection of the CTF profiles, the relative movement of the aligned tilt images, summing all the tilt images and checking that gold beads describe straight trajectories perpendicular to the tilt axis, manually locating the gold beads in the reconstructed tomogram and checking that they do not present artifacts, visual inspection of the location of the picked coordinates in the tomogram (or even resorting to manual-picking approaches), visual inspection of the extracted subtomograms (or their 2D projection, or their central slice) and the classes obtained from the subtomogram averaging (STA), and removal of the noisy classes or subtomograms whose orientation to the membrane in which they are

inserted is incorrect, and iterating the process until a significant amount of the miss-picked particles is removed.

The validation methods presented in this chapter focus on the most classical image processing pipeline in cryoET. Most of these validation tools are implemented in Scipion (de la Rosa-Trevin et al., 2016), and specifically in ScipionTomo, which is the section inside the Scipion framework dedicated to cryoET (Jimenez de la et al., 2022), using Xmipp (de la Rosa-Treví et al., 2013; Strelak et al., 2021). For each of them, we give the name of the Scipion protocol implementing it. We also propose possible validation methods to be implemented in the future. Many of these have been helpful in Single-Particle Analysis, and they can be easily extended to the handling of subtomograms.

The cryoET image processing workflow

The experiment begins by introducing the sample in the microscope and acquiring a set of images called tilt series. Each image is obtained at a specific tilt angle of the sample. The specimen under study is illuminated with electrons emitted by an electron gun. They are focused on the sample using magnetic lenses. Later, another set of magnetic lenses projects the image onto a detector. The acquired images are produced in transmission. The radiation damage limits the number of acquired images and their contrast. For this reason, the images have an extremely low signal-to-noise ratio (SNR) and contrast. This fundamentally affects the high frequencies of the image. Low-pass filters based on dose are common to remove unreliable high frequencies (Grant and Grigorieff, 2019). Some works have analyzed different acquisition schemas for the reconstruction step's quality. They combine different doses and acquisition geometries (Turonova et al., 2020).

The image formation is based on two hypotheses (Vulović et al., 2014).

1. The projection assumption (PA): All images are projections of the sample along the beam direction. Due to the tilting process, the result is a set of projections of the sample along different directions given by the tilt angles.
2. Weak Phase Object approximation (WPOA): This approximation considers that the energy of the electrons is so high that there is a single scattering event, and this event results in an elastic interaction (the electron's energy does not change after interacting).

The second hypothesis has profound consequences in imaging. Loosely speaking, the sample is transparent to the electron due to the high energy of

the electrons in combination with the nature of the light atoms that compose the sample. Note that the macromolecule and its surrounding ice atoms are light with relatively close atomic numbers. The image is produced by phase contrast (Unwin et al., 1971) resulting in extremely low contrast, complicating the identification of structural details. To alleviate this situation, the sample is defocused on purpose (leaving out the use of phase plates (Danev and Baumeister, 2017)). Defocusing the sample has a price to pay in terms of structural information. The images are blurred, and high resolution is lost. This must be considered during image processing. The Contrast Transfer function, or CTF, describes this and other microscope effects on the images.

Finally, the images are recorded employing sensors called direct electron detectors (McMullan et al., 2016). The current generation of detectors can acquire images extremely fast, as the acquired image contains the isolated impacts of electrons. These images are called frames. During the exposure time, several frames are acquired, making the so-called movie (one movie per tilt angle). In summary, the result of a tomography image acquisition with the electron microscope is a set of movies at different tilt angles. These movies contain the raw information of the sample and define the starting point of the image processing. The objective of the image processing workflow is to use these images to obtain a 3D specimen representation.

The first step of the workflow consists of correcting for the possible movement of the sample induced by the radiation of the electron beam (Brilot et al., 2012). Once motion and sample deformations are corrected, the frames are averaged, obtaining a single tilt image per movie. The tilt series is the set of all tilt images obtained at every angle. This step can be automated; therefore, it is common to consider the tilt series as the starting point of image processing.

Defocus and other microscope aberrations blur the acquired images. These aberrations are modeled in the Fourier space by the CTF (Kirkland, 2010). Accurate determination of the CTF is crucial for its posterior correction. Besides classical implementations of CTF estimators inherited from SPA, other approaches explicitly consider the cryoET case (Fernández et al., 2006; Xiong et al., 2009). However, the most widespread estimation uses a classical 2D correction as implemented in SPA (Vilas et al., 2022), particularly CTFFIND4 and gCTF (Rohou and Grigorieff, 2015; Zhang, 2016). These algorithms estimate the mean defocus, and the tilt angle with the distance to the tilt axis is used to perform a local correction. A more accurate estimation of the CTF may be achieved considering the acquisition

geometry. Once the CTF is correctly estimated, the aberrations introduced by the microscope are corrected (Fernández et al., 2006; Voortman et al., 2011; Su, 2019).

Ideally, the sample rotates around an axis common to all tilt images. However, the reality is slightly different, and the sample's undesired movements (small shifts and rotations) affect the acquired images. Thus, the set of images in the tilt series is misaligned with respect to each other. The consequences of the misalignment are unreliable or artifactual reconstructions or even the absence of any information in the reconstruction. Estimating and correcting the sample movements in the tilt series images are necessary to avoid this problem. This problem is solved in the tilt-series alignment step. Several proposed approaches differ in the mathematical methods used to solve this geometrical problem (Messaoudi et al., 2007; Castano-Diez et al., 2010; Mastronarde and Held, 2017; Fernandez et al., 2018; Fernandez and Li, 2021; Seifer and Elbaum, 2022; Zheng et al., 2022). These methods attempt to find the tilt axis in each image and correct the image displacements. The aligned images should present smooth transitions between consecutive images. The extremely low SNR complicates the alignment. To simplify this scenario, it is common to add fiducial markers to the sample, particularly small gold bead particles. The gold has a high atomic number in comparison with the elements of the sample, and they appear in the tilt images as consistent highly contrasted elements. They are easily tracked through the tilt series. Many tilt series algorithms rely on their presence to solve the misalignment.

Once the sample movement is corrected, it is possible to use several reconstruction methods to elucidate a three-dimensional map from the tilt series. There are different methods with variations in their underlying algorithms that will lead to reconstructions presenting different features. These reconstruction methods can be classified into two families: Fourier space and series expansion, iterative algorithms (Sorzano et al., 2017a). Weighted Back Projection (WBP) (Radermacher, 1992) dominates the first group. In the second group, ART (Gordon et al., 1970), SIRT (Gilbert, 1972), and SART (Andersen and Kak, 1984) are the most used. All reconstruction algorithms assume that the CTF has been corrected in the tilt series. A case of particular interest is novaCTF, a WBP algorithm that combines the reconstruction algorithm with a local CTF correction (Turoňová et al., 2017). Even if the CTF is corrected before or during the reconstruction, the reconstructed tomogram is affected because the

maximum tilt angle is limited to 60−70 degrees, resulting in the so-called missing wedge in Fourier space.

Up to this point, all the steps described are common to all cryoET workflows and result in a set of tomograms. Beyond this point, the tomography workflow takes multiple directions depending on the nature of the biological problem and the researcher's aim. The typical final step consists of tomogram segmentation to label its structures or regions of interest for applications focused on the cellular environment and the biological context of macromolecules. In contrast, further processing steps are required if the study aims to elucidate the structure of a protein of interest. This is the subtomogram averaging workflow.

In the subtomogram averaging workflow, the next step is particle picking, which attempts to identify the particles of interest in the tomogram. The proper identification of the proteins is critical for later steps. If the particles are wrongly selected, the protein reconstruction is inaccurate. Tomograms are affected by a low SNR, complicating the identification of particles, especially if they are small or embedded in cellular membranes. Generally, there are two kinds of picking algorithms: nondirectional (Wagner et al., 2019; de Teresa et al., 2023) and directional (Castano-Diez et al., 2017; Martinez-Sanchez et al., 2020; Lamm et al., 2022). The first kind only gives information on the three-dimensional localization of the particle in the tomogram, while directional pickers also give some extra information on the particle's orientation. This second kind of picker usually needs some surface or mesh of points from which the relative orientation of the particle is calculated. These pickers are particularly suitable for identifying particles located over a surface, such as in membrane proteins. This orientation information can be used later during their 3D alignment.

The next step in the workflow extracts the information for each particle identified in the tomogram. Two kinds of extraction can be carried out: (1) subtomograms, that is, a set of small subvolumes with the dimensions of the protein of interest; (2) tilt-series particles (also called subtilt series (Galaz et al., 2015) or pseudo-subtomograms (Zivanov et al., 2022)), that is, 2D images extracted from the tilt series where the protein of interest is located. Note that subtomograms or tilt-series particles are two different representations of the same protein in the tomogram. The first assumes that the tilt series images have been perfectly aligned, while the second allows further refinement of the tilt series alignment.

Then, subtomograms are aligned and classified into homogeneous groups in which all subtomograms represent the same protein in the same

Bibliography

conformation. These are the 3D alignment, classification, and averaging steps. Within a homogeneous class, subtomograms are combined by averaging all subtomograms after determining their proper relative orientation (3D alignment step) (Wan et al., 2016). The noise variance is reduced by averaging, and the underlying signal is maintained. The result is a map per class with an improved SNR. Note that all experimental subtomograms are affected by the missing wedge. This fact complicates the alignment. If the algorithm does not handle the missing wedge correctly, the subtomograms tend to be aligned and classified according to the direction of the wedge.

In the per-particle-per-tilt strategy, the goal is also to classify the input proteins into homogeneous classes and refine their relative orientation with respect to the class representative. However, the algorithm now has access to the individual projection, and consequently, it has more degrees of freedom to try to match the class representative to the acquired projections.

The workflow steps described are the basic steps for a cryoET analysis. These steps can be combined in multiple ways with various quality checks, consensus algorithms, etc., to create the specific workflow to solve a specific problem. Fig. 9.1 shows a proposed workflow for cryoET. Consensus

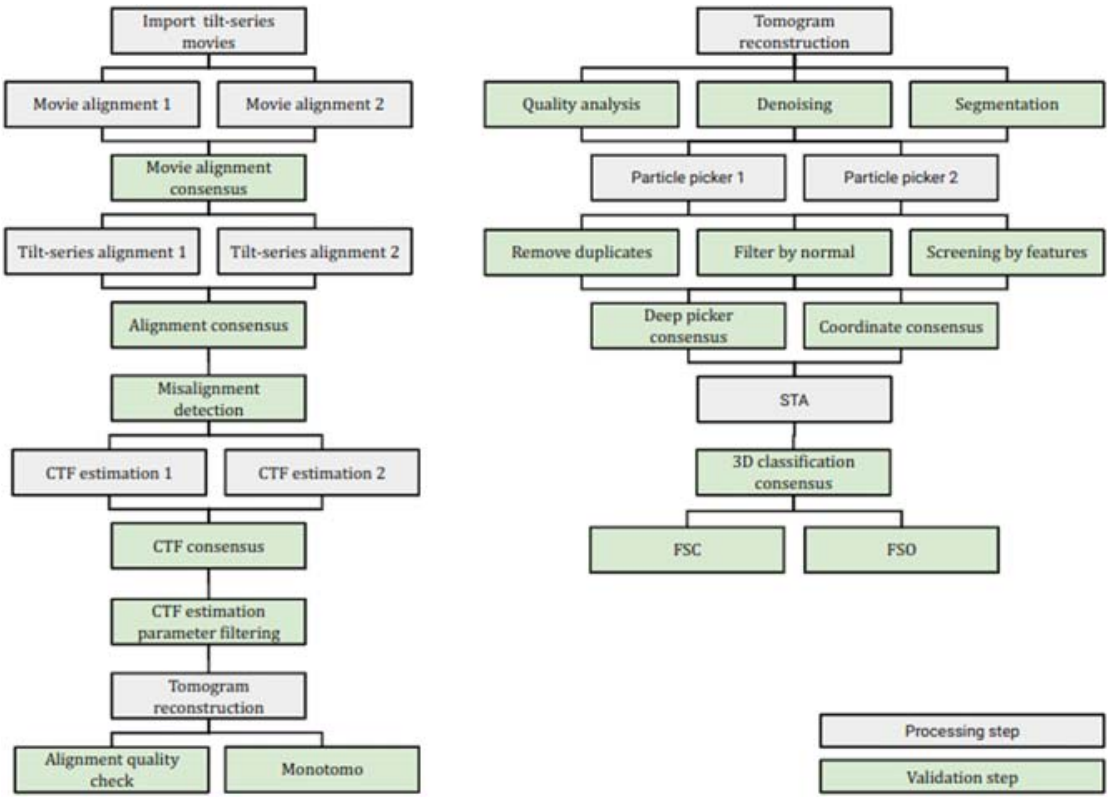


Figure 9.1 Example of a cryoET analysis workflow. The workflow on the left goes from the input tilt-series movies to the reconstructed tomogram. The workflow on the right performs a subtomogram analysis.

algorithms consist of different techniques to compare the results of different programs and algorithms that solve the same problem in order to obtain more robust results by the combination of them.

Movie alignment

Problem: At each of the tilt angles, a stack of frames is acquired. These frames must be combined into a single image per tilt. To do so, the relative movement (global and local) between images is calculated.

In this step, the relative movement of the frames from each movie acquired at each tilt angles is estimated and corrected. Many algorithms have been developed for this task (Li et al., 2013; Scheres, 2014; Grant and Grigorieff, 2019; Abrishami et al., 2015; Zheng et al., 2017; Strelák et al., 2020). After the alignment, the corrected frames are averaged to obtain a single image per tilt angle. The resulting tilt image presents a much higher SNR than the isolated frames. This is illustrated in Fig. 9.2.

Validation methods

- *Consensus alignment*: In estimating the relative movement of the sample through the movie frames, any possible miscalculation leads to a loss of signal in the reconstructed tomogram, decreasing its SNR, especially at high frequencies. This validation tool aims to detect any possible error in calculating the relative shifts between frames. A possible strategy is to compare the global frame shifts estimated by different algorithms, being more confident in those shifts consistent across different algorithms. Let us denote as $(\Delta x_i, \Delta y_i)$ the global shift between the i-th and the $(i+1)$-th frame. For the first movie alignment algorithm, we may arrange all these shifts into a single matrix with all the global shifts

$$S_1 = \begin{pmatrix} \Delta x_1 & \Delta x_2 & \dots \\ \Delta y_1 & \Delta y_2 & \dots \\ 1 & 1 & \dots \end{pmatrix} \tag{9.1}$$

where 1 is added to express the shifts in homogeneous coordinates. We construct similarly S_2 for the second movie alignment algorithm. Depending on the frame and spatial location chosen as the origin by the different algorithms and their axis conventions, the two global shift matrices may

Bibliography

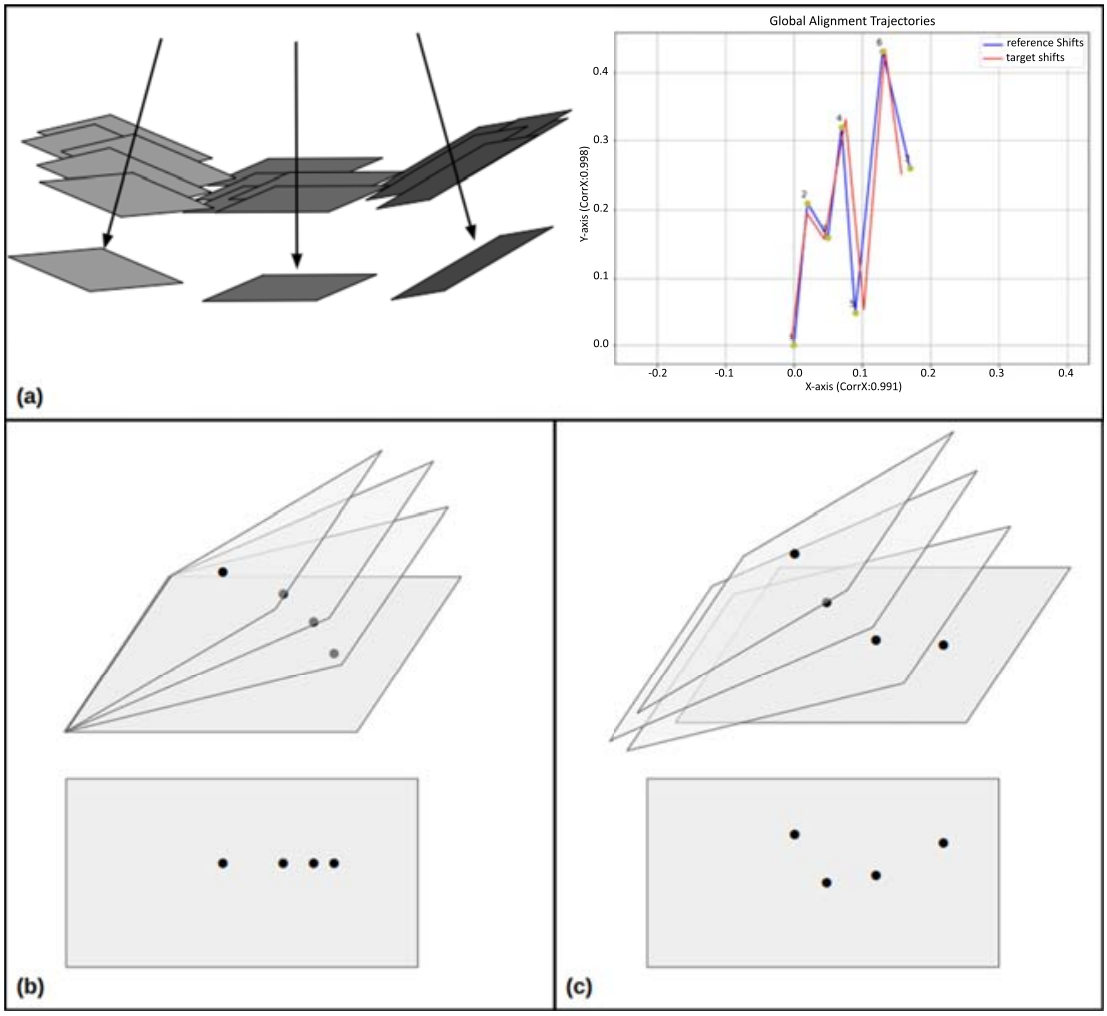


Figure 9.2 Tilt-series movies and tilt-series alignment schematic. (a) Tilt-series movies alignment schematic from three stacks of movies to three single tilt images (left) and the correlation chart of two different alignments for the same tilt-series movie stack. (b) Schematic of an aligned tilt series in which it can be observed a single marker along the series (top) and its projection in the central plane (bottom). (c) Same as (b), but in this case, the tilt series presents some misalignment. It can be observed that the disposition of the protected markers does not present a straight line (as in the case of the aligned tilt series).

have quite different values. However, we may look for the affine transformation that best transforms one set of shifts into the other

$$A = \arg\min \| S_1 - AS_2 \|^2 = \left(S_1 S_2^T\right)\left(S_2 S_2^T\right)^{-1} \tag{9.2}$$

Once we have calculated this affine transformation, we may transform S_2 to approximate S_1

$$S'_1 = AS_2 \tag{9.3}$$

We now calculate the Pearson correlation coefficient between the X global shifts in S_1 and S'_1. We will refer to it as $\rho_{\Delta x}$. We compute this correlation for the Y global shifts similarly. Finally, the correlation between the two global trajectories is defined as

$$\rho = \min\left(\rho_{\Delta x}, \rho_{\Delta y}\right) \tag{9.4}$$

The movie alignment performed by an algorithm is said to be validated if its correlation with a second algorithm, ρ, is above a given threshold (our default value is 0.5, but this value can be changed by the user). This algorithm is implemented in Scipion under the protocol `xmipp – movie alignment consensus`.

Tilt-series alignment

Problem: Tilting the sample during the image acquisition produces unintended movements in the sample that must be corrected with a tilt-series alignment algorithm before the tomographic reconstruction.

The methodology to solve this problem depends on the sample preparation process and the mathematical methods used to find the geometrical transformations that align the whole tilt series. Typically, small gold beads are added to the sample as markers called fiducials or landmarks. They appear in the images as high-contrast points, making the alignment process easier for posterior reconstruction. Even distinctive points of the sample, such as high-contrast regions, can be used as landmarks as long as they can be easily identified throughout the tilt series (Galaz-Montoya et al., 2015). Despite the differences among the different alignment algorithms (Mastronarde and Held, 2017; Messaoudi et al., 2007; Fernandez and Li, 2021), their strategy is similar: (1) detecting the landmarks in as many images of the tilt series as possible, (2) finding their correspondence (this landmark in the i-th image is this landmark in the $(i+1)$-th) generating the so-called landmark chains (made of the coordinates of the same landmark across multiple images), and (3) inferring the geometry of the movement from these landmark chains. This is ilustrated in Fig. 9.2.

In the case of fiducial-less data, a different family of algorithms is used whose strategy uses the whole image content and compares each tilt image to a reference tomogram, as in Single-Particle Analysis (Zheng et al., 2022).

Bibliography

However, in the current cryoET, the gold beads approach is much more frequent than the one exploiting the image content.

Whatever the strategy followed, the result of a tilt-series alignment algorithm is a set of 2D transformation matrices (typically encoding just rotation and translation, although they may also include other affine transformations) that bring each tilt image into a common coordinate system. Also, these matrices should align the sample tilt axis with the Y axis. We will refer to this set of matrices as $\{M_i\}$ and assume that there are T tilt images, that is, $i = 1, 2, \ldots, T$. If these matrices only describe rotations and translations (they are rigid transformation matrices), then their structure in homogeneous coordinates is of the form

$$M_i = \begin{pmatrix} R_i & \mathbf{t}_i \\ 0^T & 1 \end{pmatrix} \tag{9.5}$$

where R_i is the rotation needed for the i-th tilt image and $\mathbf{t}_i$ the necessary shift.

Once the transformation images are determined, they are used to produce an aligned tilt series whose images $\widetilde{I}_i$ are calculated from the original, unaligned images I_i

$$\widetilde{I}_i(M_i\mathbf{s}) = I_i(\mathbf{s}) \tag{9.6}$$

where $\mathbf{s}$ is the (x, y) coordinate in the coordinate system of the original image.

Validation methods

- *Misalignment detection based on fiducial residual vectors*: The question now is how to determine if the transformation matrices have been correctly identified. Most practitioners rely on visual inspection and their experience to detect this situation. However, we may take a more systematic approach.

Let us consider the 2D coordinate of a given landmark j on the aligned i-th tilt image, $\widetilde{\mathbf{s}}_{ij}$. This location must be compatible with the existence of a 3D coordinate in the tomogram, $\mathbf{r}_j$, whose projection on the aligned tilt image is $\widetilde{\mathbf{s}}_{ij}$. Because the tilt axis should be aligned with the Y axis, the relationship between the 3D coordinate and its 2D projection should be (Sorzano et al., 2014a)

$$\widetilde{\mathbf{s}}_{ij} = HR_i^Y\mathbf{r}_j \tag{9.7}$$

where R_i^Y the rotation matrix of the tilt angle around the Y axis, and H the 3D-to-2D projection matrix

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

However, in practice, the theoretical projection of the 3D landmark and its 2D location in the aligned image seldom coincide. We define the landmark residual as

$$\varepsilon_{ij} = \|\widetilde{\mathbf{s}}_{ij} - HR_i^Y \mathbf{r}_j\| \tag{9.8}$$

Studying these residuals makes it possible to detect misalignment in the tilt series. A residual can be large because the 2D location of the landmark has been incorrectly determined or because the image is misaligned. In the first case, only one landmark is affected within a particular image i. In the second, all landmarks are affected. Also, if the tilt series is aligned and we focus on a single 3D landmark, all its residuals should follow a random walk across images in the tilt series. If the residuals follow some special trajectory (for instance, a constant shift perpendicular to the tilt axis), this indicates tilt-series misalignment. Still, having identified tilt images with inconsistent alignments does not immediately invalidate the tomographic reconstruction. The quality of the reconstructed tomogram depends on the fraction of incorrectly aligned images (a single image will likely not spoil the tomogram, and this problem often goes unnoticed by the bare eye). Also, the per-particle-per-tilt approach during the subtomogram analysis may identify and correct these misalignment errors.

The protocol `xmipptomo - detect misalignment tilt-series` in Scipion allows the user to make decisions based on the evaluation of the residuals per landmark and tilt image. It also detects particular trajectory patterns whose specific description will be the subject of a subsequent manuscript.

- *Consensus alignment*: If we estimate the tilt-series alignment matrices twice using different algorithms, we may compare the consistency of their results and identify those tilt images for which the two alignments disagree. We cannot know which of the two is right if any, but we can guarantee that there has been a problem with that particular image in at least one of the algorithms.

Let us refer to the two estimations of the transformation matrices as $\{M_i^1\}$ and $\{M_i^2\}$. Comparing these two sets is far from trivial because the common coordinate system chosen by the first algorithm is not generally

Bibliography

the one chosen by the second algorithm. We need to find a matrix that transforms one into the other. To find this transformation, let us analyze deeper the alignment problem. Then, the aligned images, according to the two algorithms, are

$$\begin{aligned} \widetilde{I}_i^1\left(M_i^1\mathbf{s}\right) &= I_i(\mathbf{s}) \\ \widetilde{I}_i^2\left(M_i^2\mathbf{s}\right) &= I_i(\mathbf{s}) \end{aligned} \tag{9.9}$$

However, there must be a transformation P that transforms the coordinates of the second alignment into the coordinates of the first:

$$PM_i^2\mathbf{s} = M_i^1\mathbf{s} \tag{9.10}$$

and this transformation P must be the same for any coordinate $\mathbf{s}$. Therefore, it must be

$$P = M_i^1\left(M_i^2\right)^{-1} \tag{9.11}$$

That is, each tilt image gives an estimate of the necessary transformation matrix. However, this transformation matrix should be the same for all tilts, so we average all of them to get a more accurate estimate

$$P = \frac{1}{T}\sum_{i=1}^{T} M_{i,1}M_{i,2}^{-1} \tag{9.12}$$

We may now compare the disagreement between any two estimates:

$$\Delta M_i = M_i^1 - PM_i^2 \tag{9.13}$$

This matrix is also a rigid transformation with an associated rotation, ΔR_i, and translation, $\Delta\mathbf{t}_i$. We may identify those images within the tilt series whose rotational or translational disagreements are larger than a user-specified threshold.

For those tilt images where the two alignment estimates agree sufficiently (the rotational and translational error is below the user-specified threshold), we may further refine the alignment by combining both estimates into a single one

$$M_i^{12} = \frac{1}{2}\left(M_i^1 + PM_i^2\right) \tag{9.14}$$

This algorithm is implemented in Scipion under the protocol `tomo - tilt-series consensus alignment`.

Contrast transfer function (CTF)

Problem: The set of acquired images are affected by the defocus and aberrations of the microscope; therefore, they are not ideal projections of the sample. Estimating and correcting these deviations from the ideal projection images are necessary.

The Contrast Transfer Function (*CTF*) models the electron microscope's aberrations and, thus, the image formation. The aberrations are imperfections of the optical system that result in blurred images (Mahajan, 1991). The main aberrations are defocus, spherical aberration, and astigmatism. Other aberrations, such as trefoil and comma, may exist, but they are supposed to be small and not commonly estimated. Defocus, Δf, is the distance between the detector's position and the optical system's focal plane. In cryoET, defocus is introduced to produce contrast, but the price to pay is an image blur. The defocus is the most important parameter in the CTF estimation; hence, its estimation is critical.

For the sake of simplicity, let us concentrate here on a one-dimensional description of the *CTF* without attenuation at high frequency. The reader interested in a more complete model is referred to (Sorzano et al., 2007). Let f denote the spatial frequency measured in $\mathring{A}^{-1}$. Then, the *CTF* is a function whose oscillatory behavior can be written as

$$CTF(f;\Delta f) = -\sin\left(\chi(f;\Delta f)\right) + Q(f)\cos\left(\chi(f;\Delta f)\right) \tag{9.15}$$

where we have emphasized the *CTF*'s dependence on the frequency f and the defocus Δf, which changes from micrograph to micrograph and in cryoET even within the micrograph. $Q(f)$ is the fraction of electrons being inelastically scattered, and it is normally assumed to be a constant between 0.05 and 0.2, depending on the sample thickness. $\chi(f;\Delta f)$ is the argument of the *CTF* sinusoids and can be calculated as

$$\chi(f,\Delta f) = \pi\lambda\left((\Delta f)f^2 + \frac{1}{2}C_s\lambda^2 f^4\right) \tag{9.16}$$

with λ being the electron wavelength and C_s the microscope spherical aberrations, both parameters depend exclusively on the microscope and do not change from acquisition to acquisition. Extending this one-

Bibliography

dimensional *CTF* model to a two-dimensional *CTF* model involves two defoci, instead of one, and their difference is called astigmatism.

The *CTF* is measured from the acquired images. The Power Spectral Density (*PSD*) describes the energy distribution of the image in the frequency domain. It can be estimated through the modulus of the image's Fourier transform. It is related to the multiplication of the *PSD* of the ideal projection multiplied by the modulus squared of the *CTF*. Usually, the *PSD* looks like a set of concentric fringes caused by the *CTF* sinusoidal component, called Thon rings (Thon, 1966).

The aberrations and defocus are calculated from the Thon rings' minima frequency location. Unfortunately, the limitations in image acquisition complicate the *CTF* estimation. First, to limit the radiation damage, the electron dose must be reduced; typically, each image has been radiated with 2 $e/\text{Å}^2$, and the tilt series with around 90–240 $e/\text{Å}^2$ (Navarro, 2022). Second, cryoET algorithms for *CTF* rely on the Weak Phase Object Approximation (WPOA) (Vulović et al., 2014), which assumes that the sample is thin and the interaction of electron–matter is weak. When the sample is tilted, the effective thickness of the sample is increased, reducing the applicability of the WPOA. In addition, a tilted sample presents a defocus gradient. Each point of the image is affected by a different defocus depending on its distance to the tilt axis (generally at the center of the image). This gradient complicates the estimation of the CTF because there is not a single defocus value for the whole micrograph. Most CTF estimation algorithms estimate a single defocus per tilt image (Rohou and Grigorieff, 2015; Zhang, 2016) corresponding to its mean value. However, the CTF correction must take into account the defocus gradient. This is carried out considering geometrical constraints (Xiong et al., 2009).

Validation methods

- *Filtering by CTF parameters*: The estimation of the *CTF* consists of a set of parameters that properly characterize this function. Different pieces of software implement different mathematical models of the *CTF*. They all share some common parameters, most importantly, the defocus (two values) and some measure of the maximum resolution present in the micrograph. Thus, it is possible to validate the *CTF* estimation according to these parameters. We may set a range within which the defoci, astigmatism, and maximum resolution must be. Out-of-range parameters reveal either a problem with the micrograph or the *CTF* estimation

(algorithms are normally more unstable for high-tilt image estimations). In either case, the *CTF* estimation must be flagged as problematic. The researcher must decide what to do with a tilt series in which some of its tilt images have been identified with problematic *CTF*s (disregard the warnings, eliminate the problematic images, or discard the whole tilt series).

- *CTF consensus*: Another possible strategy to ensure a proper estimation of the *CTF* parameters is the direct comparison between two or more different estimations through a consensus algorithm. Let us assume that two different algorithms estimate the defoci Δf_1 and Δf_2 for the same micrograph. Then, we may evaluate how these different values affect the *CTF* function

$$\varepsilon(f;\Delta f_1,\Delta f_2)=\left|CTF(f;\Delta f_1)-CTF(f;\Delta f_2)\right| \tag{9.17}$$

At low frequencies, this difference will be small. However, they may significantly differ at high frequency, making *CTF* correction completely unreliable (see Fig. 9.3).

It is custom to put a threshold on the phase difference of the *CTF* instead of the *CTF* value. This way, two *CTF*s are said to agree up to a frequency f_{max} if

$$\left|\chi(f;\Delta f_1)-\chi(f;\Delta f_2)\right|<\Delta\chi_{max}\quad\forall\,|f|<f_{max} \tag{9.18}$$

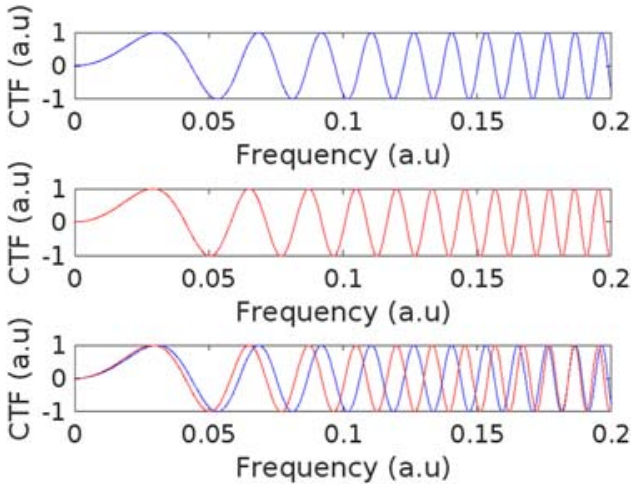


Figure 9.3 Two different CTF estimations: The first estimation (blue) considers a defocus of 27,000 Å, and the second prediction considers a defocus of 30,000 Å (red). At low frequencies (resolutions), both curves are similar, and both estimations can be accepted and provide a good CTF correction. When the frequency has increased, the differences between them are significant. The vertical black line shows the resolution limit when the differences become significant.

Typically, $\Delta\chi_{max} = \pi/2$ is chosen as threshold. Again, the researcher must evaluate the maximum resolution of agreement and decide how to deal with the problematic tilt images.

If we think of the defoci as noisy estimates of an underlying true defocus, then a sensible strategy is to average all estimates

$$\overline{\Delta}f = \frac{1}{2}(\Delta f_1 + \Delta f_2) \tag{9.19}$$

If Δf_1 and Δf_2 agree to high resolution, i.e., f_{max} is high, then we can be more confident that these two defocus estimates are affected by a small amount of random noise rather than a large error.

Scipion's protocol `tomo - CTF validate` implements this defocus consensus and allows the computation of the parameter averages.

Tomogram reconstruction

Problem: At this step, the tilt series have been aligned, their orientations are known, and the defocus and aberrations have been corrected. The problem now is the inverse problem of determining the three-dimensional structure, called a tomogram, compatible with the set of projections (tilt series).

Once the geometry of the movement of the sample is corrected, different reconstruction methods can be used to elucidate a three-dimensional volume from the projected tilt images based on different underlying algorithms.

The first approach to elucidate the three-dimensional structure of the sample employs the WBP (Radermacher, 1992), an algorithm based on the central-slice theorem (projection-slice theorem or Fourier slice theorem), which states that the 2D Fourier transform of each projection is a slice of the 3D Fourier transform of the volume being sought. This set of methods has the advantage of being fast and not very intensive in using computational resources.

The other classical approach is methods based on a series expansion (Sorzano et al., 2017a): the map in real space is divided into a sum of small basis functions whose intensity must be determined through the solution of a linear equation system. The equation system is solved iteratively, with the speed at which the system converges being adjustable. Although it is a more computationally demanding method, it can be easily parallelized in several processors.

During the reconstruction process, the microscope aberrations can be corrected explicitly considering the effect of the tilt on the defocus at a particular spatial location, as is done in (Turoňová et al., 2017).

Validation methods

- *Detection of artifacts in fiducials*: An incorrect tilt-series alignment results in an incorrect 3D reconstruction. Tilt-series alignment errors produce characteristic features in reconstructing high-contrast objects such as the fiducials (see Fig. 9.4). However, they affect the overall reconstruction, although they are more evident in those regions or features presenting a high contrast. It is not uncommon that, even for the trained eye, these artifacts are only visible in these regions remaining unnoticed for the rest of the reconstruction. These artifacts are independent of the reconstruction method since they come from errors in determining the sample movement geometry.

Thus, studying those high-contrast regions and using an image classification algorithm to decide whether these regions are artifactual are possible. There are plenty of options for image classification algorithms being machine learning the most popular one nowadays, especially deep neural networks (DNNs). The landmarks in Fig. 9.4 have been identified using a neural network available in Scipion through the protocol `xmipptomo - tomogram misalignment detection`. The details of this network will be published elsewhere. The same neural network strategy can be used to study CTF miscorrection artifacts, although this latter is not yet implemented.

Tomogram postprocessing

Problem: The reconstructed tomogram presents a poor SNR compromising the interpretation of the results. We must analyze the tomogram quality at its different locations. Then, we would like to increase the SNR by a denoising or sharpening step to simplify later steps of picking, segmenting, or simplifying the interpretation of the tomogram. Segmentation will attempt to group high-SNR voxels and classify them according to their structure or biological function.

Bibliography

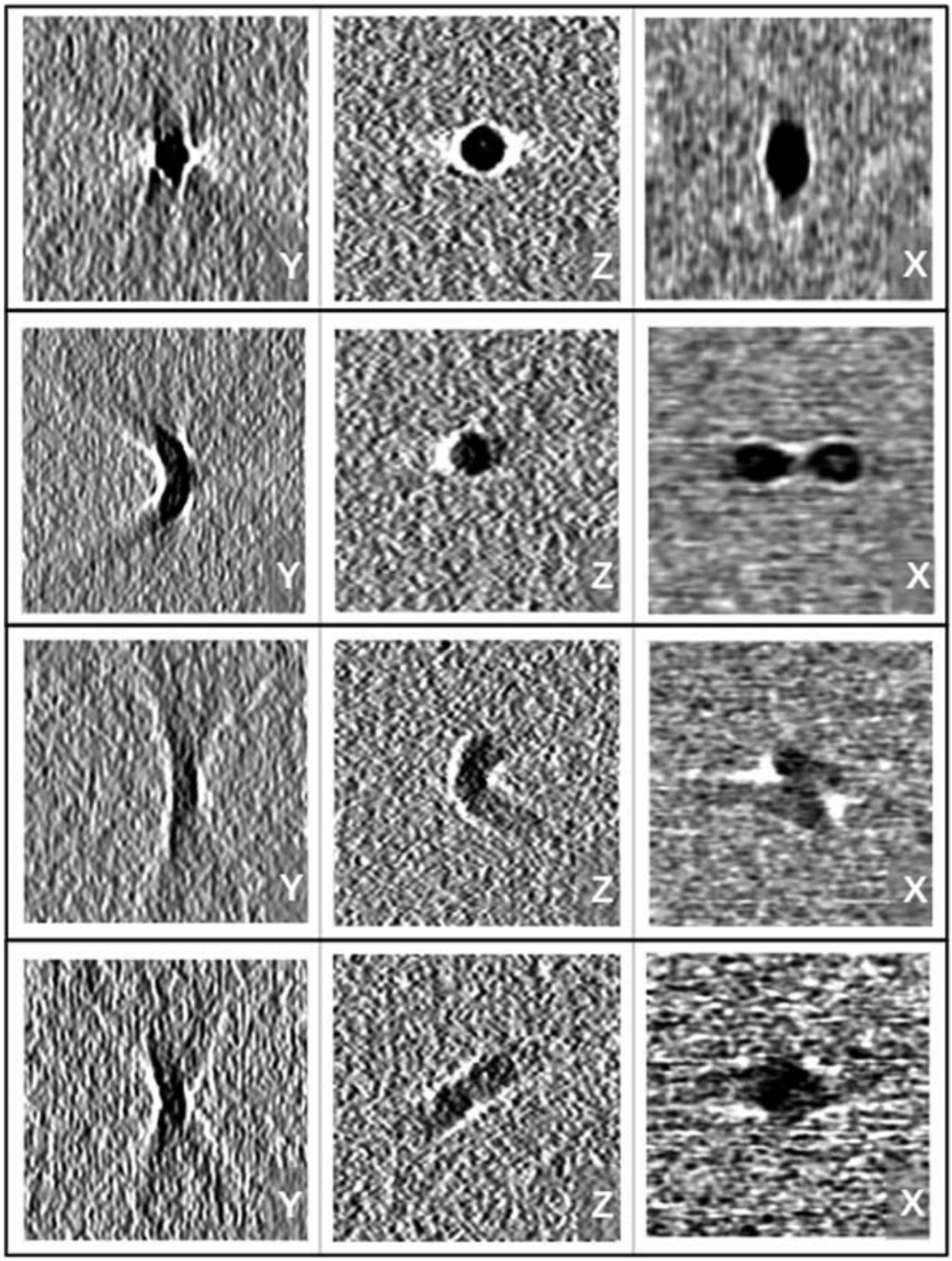


Figure 9.4 Slices along X, Y, and Z of the same gold bead reconstructed with different misalignment. The first row presents no misalignment. The second row presents bananas (incremental, or offset, shift in X direction). The third line presents twisters (rotation of the tilt axis). The last row birds (incremental shift in Y direction).

If image processing aims to study a cellular environment, then the postprocessing step will end the image-processing workflow. In contrast, if we are interested in knowing the structure of a macromolecular complex, then postprocessing will be an optional step that can help in the sub-tomogram averaging pipeline.

The postprocessing tools are diverse, but all are addressed to estimate the tomogram quality or highlight its signal for better visualization. Different postprocessing tools can be applied:

1. Quality Analysis: The objective is to quantify the reconstructed tomogram's reliability. Currently, there is no standard for estimating the quality of tomograms, and this problem has been an issue for many years. The Fourier Shell Correlation (FSC) is the most widely used metric in single particle analysis (Harauz and van Heel, 1986; Rosenthal and Henderson, 2003; Sorzano et al., 2017b). But its extension to electron tomography is not straightforward. The reason is the need to split the dataset into two independent subsets of equal size to obtain two independent reconstructions. In cryoET, this partition can be obtained by splitting the set of tilt images into two subsets, the odd and the even images (split according to the acquisition order). The resulting odd and even tilt series would present an angular sampling equal to twice the original. This allows obtaining two independent tomograms; therefore, an FSC can be calculated (Cardone et al., 2005). Another approach is the method of Noise-compensated Leave One Out (NLOO) (Cardone et al., 2005), which rules out a single tilt image, reconstructs the tomogram without it, and projects the tomogram along the direction of the discarded image. Thus, an FSC between the experimental and projected images can be estimated. Unfortunately, the algorithm is computationally expensive. Other FSC metrics, such as a conical FSC under the same approach odd−even, have been proposed to analyze the resolution anisotropy (Diebolder et al., 2015). However, all these global methods aim to fit the spectral information into a single number. Tomograms present diverse information, and a better approach can be the measurement of local resolution. Unfortunately, the dimensions of the tomograms imply a high computational burden and extremely long computational times to estimate a local FSC. Instead, the only local resolution algorithm is MonoTomo (Vilas et al., 2020a). This method estimates the local resolution by performing hypothesis tests at multiple frequencies to determine if the local amplitudes

Bibliography

are greater or smaller than the local noise amplitudes. Moreover, it proposed splitting the movie frame set into odd and even instead of the tilt series.

2. Denoising: Recently, technological and computational advances have provided a considerable improvement in the quality of the acquired images and, consequently, the reconstructed tomograms. This directly impacts the interpretation of the tomograms and, therefore, the obtained biological conclusions. However, the extremely low SNR of the tomogram is still a limiting factor. Denoising algorithms attempt to solve this problem (Frangakis, 2021). These algorithms aim to modify the tomograms, keeping the structural information but removing noise. A cleaner tomogram has many advantages: (1) a better understanding of cellular or macromolecular information; (2) simplifying the segmentation step afterward; (3) or if subtomogram averaging is carried out, providing a cleaner tomogram to identify the protein of interest. There are three prominent families of denoising algorithms (Frangakis, 2021): real-space algorithms, transform-based filtering methods, and neural networks. The first group contains classical image processing techniques such as Gaussian blur and mean or median filters. A particularly sophisticated algorithm in this family is the nonanisotropic diffusion algorithm (Frangakis and Hegerl, 2001; Moreno et al., 2018). In the family of transform-based methods, thresholding in the wavelet domain is used as a denoiser (Huang et al., 2018). Recently, neural networks have gained tremendous popularity in denoising tasks thanks to CRYO-CARE (Buchholz et al., 2019a,b).
3. Segmentation: As was mentioned, the low SNR of the tomograms complicates the interpretation of the results. Previous steps of quality analysis and denoising looked to enhance such understanding. However, we would like a clean representation of the biological elements of interest in tomogram elements without any noise. Segmentation tools undertake this task. Occasionally, they can label the biological elements in the cell, such as membranes, ribosomes, membrane proteins, or microtubules. One of the main drawbacks of this family of algorithms is automation. The reason is the combination of low SNR and the variety of components in the tomograms. Many tools have been developed to solve this problem. Currently, deep neural networks have shown an excellent performance in undertaking segmentation tasks (de Teresa et al., 2023; Chen et al., 2017). However, many other tools are still

in use, such as the classic tomoSegMemTv based on tensor voting (Martinez-Sanchez et al., 2014) or recent solutions based on membrane curvature (Frangakis, 2022).

Proposed validation methods

Postprocessing methods are especially complex to validate. Validation algorithms generally use a set of measures of the same magnitude. Postprocessing tools' very diverse nature (quality analysis, denoising, or segmentation) implies the need for validation methods adapted to each task. Despite their differences, these three families attempt to look for the signal in the tomogram, but with different aims. Quality methods try to estimate the SNR, denoising search to highlight the signal over noise, and segmentation algorithms to label it. This common point suggests a possible validation tool based on a shared underlying signal. Thus, the local resolution shows the degree of spatial reliability in each position of the tomogram. The local resolution map should match the regions with structural information (resolution greater than background resolution), the denoised regions, and the segmented ones. Similarly, it is possible to use the original and denoised tomograms to carry out two independent segmentations to compare them and carry out a consensus. Unfortunately, this kind of validation is weak and, to our knowledge, is not used. An example of this kind of validation is shown in Fig. 9.5, where different postprocessing algorithms were applied to the same tomogram. All these methods should be in agreement showing signals in the same locations. In particular, we observe that the segmentation is not good enough because there is no agreement between the denoised tomogram and the local resolution tomogram.

In the case of quality analysis, the set of algorithms measures different magnitudes. In the case of global resolution, all methods should provide similar results. To the best of our knowledge, there are no methods to validate the performance of segmentation algorithms. However, the validation can be carried out very simply by means of a consensus. Ideally, all segmented tomograms should be identical. The most straightforward validation can be the tomogram obtained due to the intersection of all segmentation obtained by different segmentation algorithms. More complex approaches can consider a voting system or a comparison with a quality metric such as the local resolution. To our knowledge, none of these strategies have been implemented in any tool practitioners may use.

Bibliography

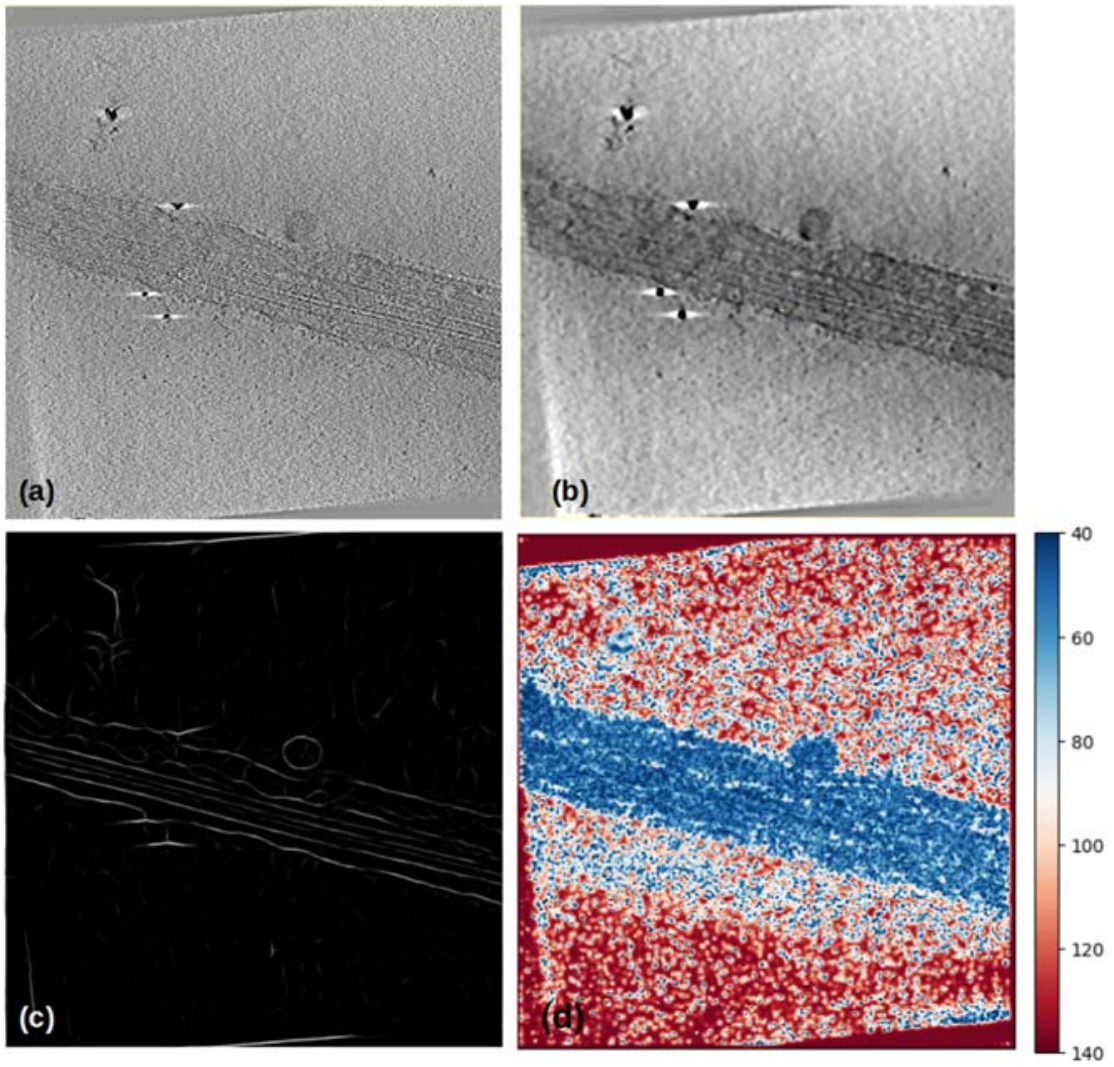


Figure 9.5 Different postprocessing algorithms applied to the same tomogram: (a) Original tomogram, (b) the tomogram is denoised with tomoEED (Moreno et al., 2018), (c) a segmentation of the tomogram with tomosegmemTv (Martinez-Sanchez et al., 2014), and (d) local resolution tomogram obtained with MonoTomo (Vilas et al., 2020a).

Particle picking

Problem: This step tries to find the proteins of interest in the tomogram. The identification of the proteins is not an easy task because of the tomogram's low SNR.

The different particles identified in the tomogram are subjected to subtomogram averaging to reveal a detailed structure of the protein or to study the spatial distribution of the particles in the tomogram. Traditionally, particle picking has been a challenge in cryoEM. False positives and negatives during the particle identification compromise the results of the subsequent analysis. Consequently, there is a broad variety of picking

algorithms, but we can classify them according to their degree of automation and directionality. Automation refers to the amount of user intervention needed to select the particles. Directionality refers to identifying the particles' orientation, for instance, with respect to a membrane.

Validation methods

- *Multiple picking*: The idea is to pick the tomograms with N different pickers. Ideally, all pickers should identify each particle with the same or very similar coordinates. Reality is far from this ideal situation, and the particles identified by each algorithm can wildly differ. To validate and obtain a reliable set of coordinates, a consensus approach can be carried out. We can consider different consensus criteria. For instance, a candidate coordinate is considered a true particle if identified by at least M of the N used pickers. Note that the maximum reliability occurs when $M = N$. To ensure that two pickers identify the same particle, the coordinates must be inside a sphere of a given radius (tolerance). In Fig. 9.6, we illustrate this criterion. This validation method is available in Scipion through the protocol `tomo - picking consensus`.
- *Deep picker consensus*: Following the previous line of reasoning, it is possible to design a neural network that identifies incorrectly identified particles (Sanchez-Garcia et al., 2018). The aim is to use the results of several pickers as the training set for the network, using the matching coordinates as positive elements and background noise as negatives. The nonmatching coordinates are used on the trained network to test if they are properly classified. Then, the network can be used to score the obtained coordinates. The protocol `xmipptomo - deep consensus picking D` in Scipion implements this idea.
- *Duplicate removal*: This kind of validation addresses the removal of repeated coordinates from single or multiple pickers. A repeated coordinate is defined as two or more coordinates close enough to each other to be considered to point to the same particle. The goal of this validation is to introduce a physical constraint that avoids the overlapping between adjacent particles. Some pickers, such as PySeg (Martinez-Sanchez et al., 2020), have the minimum distance between particles as one of their parameters. This validation method is available in Scipion through the protocol `tomo - remove duplicates`. In Fig. 9.6, we illustrate this criterion.

Bibliography

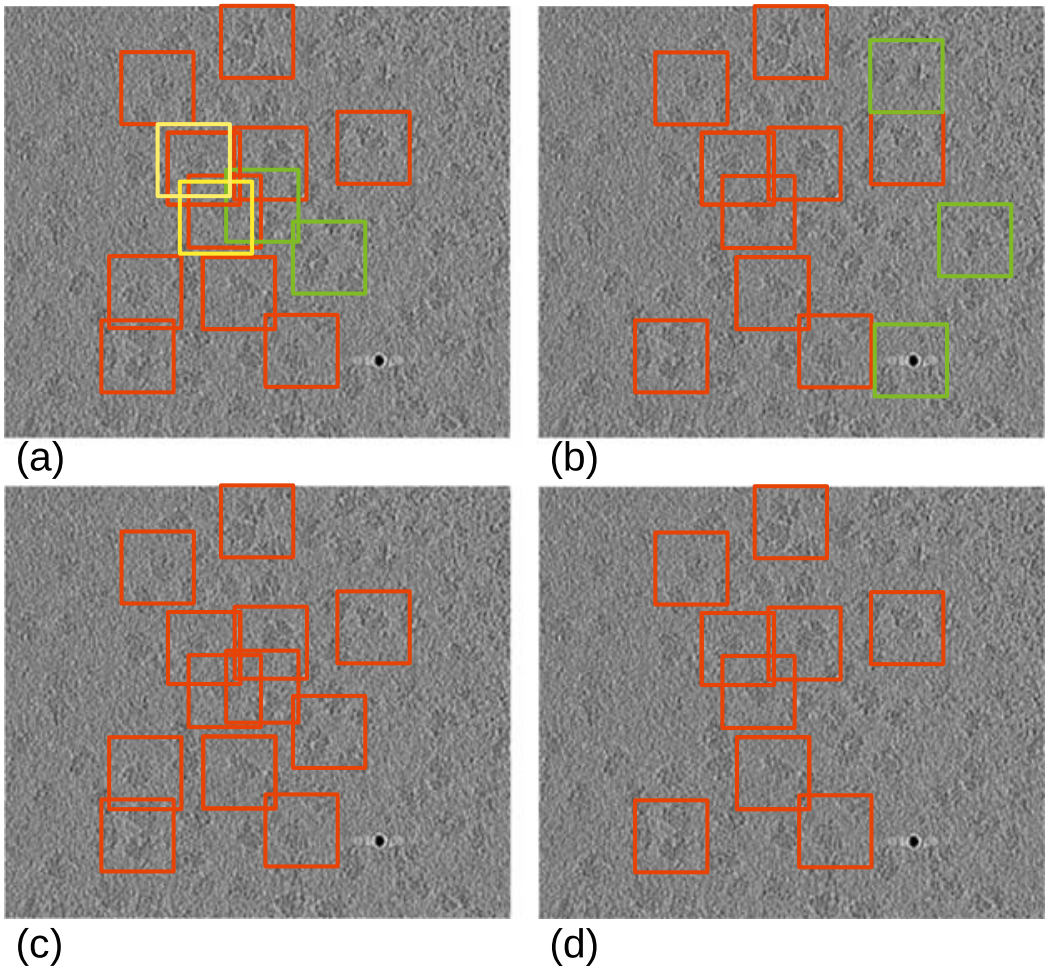


Figure 9.6 Illustrative examples of the picking consensus approaches: (a) Picked tomogram with an algorithm X, (b) the same picked tomogram with a different picking algorithm Y. (c) Validated particles in the picking (a) after applying the remove duplicates protocol (duplicated particles in yellow). (d) The consensus picked particles with the pickers A and B (noncommon particles in green).

- *Filtering by normal*: This filter takes advantage of the directional picking by introducing some priors or limitations about the orientation of the particles in the tomogram. For instance, the direction of the picked particles with respect to a surface or mesh, usually representing a membrane or vesicle. Thus, it is possible to filter a set of coordinates according to their direction with respect to this surface and remove those that do not verify a given normality criterion. As it happens with the removal of duplicates, some pickers such as PySeg (Martinez-Sanchez et al., 2020) contain this functionality for membrane picking. This validation method is available in Scipion through the protocol `tomo - filter by normal`. This protocol also offers the possibility of filtering particles by their tilt angle, removing those ones that are above a specified angular

threshold, and avoiding introducing particles that are in the direction of the missing wedge in further processing.

- *Directional consensus*: In cases where particles have been assigned directional information according to the local geometrical properties of their location in the tomogram, it is also possible to perform a directional consensus filtering based on a multipicking approach. To that end, the coordinate alignment information is converted into quaternions, allowing one to compare and average the angular alignments of several consensus coordinates before assigning them to their consensus representative. In particular, the conversion of a rotation into its quaternion representation simplifies the average rotation problem into a simple and fast eigen-decomposition. Quaternions were defined as an extension to imaginary numbers, instead of having an imaginary part, they have a 3D vector as an imaginary part. The one most similar to complex numbers is to write a quaternion as $q = a + bi + cj + dk$ where a is the equivalent of the real part, and b, c, d are the equivalent of the imaginary part (Sorzano et al., 2014b; Hu et al., 2020). Thus, it is possible to define a rotation of an angle α around a given 3D axis u using quaternions as

$$q_{\mathbf{u},\alpha} = \cos\left(\frac{\alpha}{2}\right) + \sin\left(\frac{\alpha}{2}\right)\frac{\mathbf{u}}{||\mathbf{u}||}$$

Considering a set of quaternions Q of the form

$$Q = (w_1\boldsymbol{q}_1, w_2\boldsymbol{q}_2, \ldots, w_n\boldsymbol{q}_n)$$

where $\boldsymbol{w}$ represents a vector of weights defining the weighted quaternion average, and $\boldsymbol{q}$ represents a quaternion vector whose rows stored the quaternion representing a given rotation. From the previous representation, we can compute the self-adjoint positive semidefinite matrix QQ^T. The normalized eigenvector corresponding to the largest eigenvalue of the previous matrix corresponds to the weighted average of the set of quaternions previously defined.

The protocol `tomo - consensus 3D coordinates` in Scipion implements this validation tool.

Proposed validation methods

- *Screening by features*: Particles are characterized by geometric features defining their shape. The goal is to screen the picked particles according to their geometrical or morphological features. This kind of validation is used in Single-Particle Analysis and can be extrapolated to cryoET

Bibliography

(Vargas et al., 2013). The idea is to determine statistical features from the picked particles and establish a threshold or Z-score to eliminate outliers. To our knowledge, this possibility is not implemented in any software package. Another possibility is to filter the coordinates according to a local property of the tomogram. For instance, its local resolution as estimated by MonoTomo (Vilas et al., 2020a), the local defocus of the particles, their distance to the carbon edge, or simply the particle's mass. This second approach aims to create a subset of particles with specific properties (for instance, good quality) instead of discarding wrong-picked particles. The protocol `xmipptomo - filter coordinates by map` in Scipion allows filtering the particles according to their local resolution.

Subtomogram averaging: Alignment and classification

Problem: Subtomogram averaging assumes that all subtomograms are identical copies with different orientations of the same reference protein and in the same conformational state. Thus, aligning and classifying the subtomograms are necessary to obtain an average map. The alignment seeks to place all particles in the same orientation, and the classification attempts to understand the protein's heterogeneity and discard wrong-picked particles. Once all subtomograms have been aligned and classified, their (weighted) average must be calculated to obtain a detailed protein structure.

Once the particles of interest are located in the tomogram, the aim is to obtain a high-resolution structure combining the information provided by each peaked subtomogram. To achieve this, there are two main tools to achieve the final average map: alignment and classification.

These two processes are combined to refine the data and produce one or several final maps. The aim is to sort all the particles into different homogeneous groups or classes according to their structural similarity. This step can also detect wrongly picked particles, such as background or other high-contrast features not of interest. Each particle is aligned to each class and assigned to the group it is more similar. This process can be repeated as many times as desired to achieve very homogeneous subgroups.

3D alignment and classification are a very computationally expensive process. Classification can be highly accelerated by classifying 2D slices or

projections of the subtomograms. Although this approach significantly reduces the complexity of the problem, the scientist must be aware that the solved problem does not represent the original dilemma's overall complexity.

Validation methods

- *Consensus alignment*: The subtomogram alignment to their class representative can be calculated multiple times using different algorithms or multiple executions of the same algorithm. Let us assume that the same subtomogram, V_i, has been aligned to its class representative in two different algorithms or runs. This alignment can be expressed using the same algebraic machinery we used in Eq. (9.5)

$$\begin{aligned} \widetilde{V}_i\left(M_i^1\mathbf{r}\right) &\approx V^1(\mathbf{r}) \\ \widetilde{V}_i\left(M_i^2\mathbf{r}\right) &\approx V^2(\mathbf{r}) \end{aligned} \tag{9.20}$$

where V^1 and V^2 are the class representatives. However, these may not be aligned with each other. Then, we may find the transformation that aligns them as

$$V^1(M^{21}\mathbf{r}) \approx V^2(\mathbf{r}) \tag{9.21}$$

From all these relations, we know how to align the subtomogram to the first class representative using the second alignment matrix

$$\widetilde{V}_i\left(M_i^2(M^{21})^{-1}\mathbf{r}\right) \approx V^1(\mathbf{r}) \tag{9.22}$$

Consequently, it must be

$$\Delta M_i = M_i^2(M^{21})^{-1} - M_i^1 \approx 0 \tag{9.23}$$

As in Sec. 4, we may calculate the ΔR and $\Delta\mathbf{t}$ implied by ΔM_i. In this way, we estimate the consistency of the alignment for all subtomograms. Scipion's protocol `xmippt - align volume and particles` estimates M^{21} and calculates the set of matrices $M_i^2(M^{21})^{-1}$, while the protocol `xmipp - compare angles` estimates ΔR and $\Delta\mathbf{t}$ for each subtomogram.

- *3D classification consensus*: Subtomogram classification is an unstable process (both between different algorithms and even executions of the same algorithm (Sorzano et al., 2022)). Although we cannot influence the stability of the classification, we may measure it. The aim is to compare

Bibliography

different classification results analyzing the distribution of particles among the classes and rewarding those particles that are classified together among the different results.

Given n classification results with M_1, M_2, …, M_n classes, we first calculate the subgroups of particles that have been classified together. The maximum number of subgroups is

$$m = \prod_{i=1}^{n} M_i \tag{9.24}$$

However, we may not necessarily reach this number of subgroups. The most stable the classification is, the smallest the number of subgroups. Stable classes tend to be preserved among the different executions, while unstable classes tend to be spread into many small subgroups. This algorithm is available in Scipion under the protocol `xmipp - consensus classes`. We are working on automatic ways to determine the number of classes that can be reliably identified. However, for the moment, these are exploratory lines of research, and the user must, at present, choose by him/herself the stable classes.

Quality of the average map

Problem: Once the average map is obtained, it is necessary to estimate its quality to determine how much reliable the average is.

Resolution by itself is an indirect validation of the information in the reconstruction. There is no universal definition of resolution, but it is usually defined as the size of the smallest reliable structural detail present in the reconstruction. The most widespread is the FSC (Harauz and van Heel, 1986; Sorzano et al., 2017b), which is based on two Gold Standard reconstructions, each of them with half of the set of picked particles (Scheres, 2012). The FSC calculates the normalized cross-correlation at different frequencies (resolutions) between two normalized Fourier shells of two independent reconstructions. In this sense, the FSC is a self-consistency measurement. The resolution for which the FSC = 0.143 is called the resolution of the map (Rosenthal and Henderson, 2003). As it happens with any magnitude, the estimation of resolution should be given with an error, this allows determining how reliable our measurement is (Beckers and Sachse, 2020; Penczek, 2020). When resolution analysis is restrained to a bounded region of the reconstruction, it is called local resolution

estimation. There are several local resolution algorithms: some based on a local FSC (Cardone et al., 2013), in a measure of the local SNR (Kucukelbir et al., 2014; Vilas et al., 2018), or in the map texture (Ramirez-Aportela et al., 2019). A review of local resolution methods and good practices for their use can be seen in (Vilas et al., 2020b). Local resolution informs about the difference in quality between different parts of the map. Generally, they are caused by flexibility, radiation damage, or angular assignment errors. It was shown that global (Tan et al., 2017; Vilas and Tagare, 2023) and local (Vilas et al., 2020c) resolutions are a directional magnitude. The resolution anisotropy implies that quality is a directional magnitude, usually related to preferred orientations in the sample.

Validation methods

- *Fourier Shell Correlation*. The FSC is the current standard resolution metric but rewards systematic errors during image processing. Overfitting is one of these errors. The gold standard attempts to prevent it. Despite that, we are not fully protected against it (van Heel, 2013; Henderson, 2013; Sorzano et al., 2022).
- *Detection of overfitting*. In SPA, some algorithms were developed to detect overfitting and validate the reconstructed map (Chen et al., 2013; Heynmann, 2015; Vargas et al., 2016; Heymann, 2018). Although not designed for subtomogram averaging, their generalization should be straightforward.
- *Local resolution*. Local resolution measurements provide a different kind of validation. It is expected that the range of local resolution should not be far from the global FSC resolution. Moreover, the blurred regions in the average map should match low-resolution values. It was shown that local resolution is a necessary condition for visualizing structure but not sufficient (Vilas et al., 2020b). For instance, it is necessary to have 3 Å resolution to observe an alpha-helix, but having a 3 Å resolution does not warranty its visualization. In contrast, the opposite statement is enough. Visualizing a given structural detail (i.e., alpha-helix) ensures a given resolution (in the alpha-helices case, below 5 Å). The second validation opportunity is the angular assignment of the particles or subtomograms. Commonly, there is a radial loose of resolution, it is the center of the map that presents a higher resolution than the periphery. This phenomenon can be measured by means of the local-directional resolution (Vilas et al., 2020c).

- *Fourier Shell Occupancy (FSO)*. It informs about the range of frequencies in which the map presents anisotropy (Vilas and Tagare, 2023) and how much information is present at each frequency.

Workflow validation strategies

Up to this point, all the proposed validation methods for the cryoET processing pipeline are focused on specific tools or steps in the workflow. Nevertheless, other validation strategies do not apply to a specific workflow step but to the whole image processing pipeline. The proposed strategies in this section are focused on this second approach. They can be applied up to the stage of the pipeline that the scientists consider reasonable, being advisable to apply them for every performed step.

Validation methods

- *Parallel workflow with frame subsets*: Typically, all the frames from each movie are combined to maximize the information in the final tilt image and increase the SNR. However, an alternative instead of combining all the available frames is to split them into two groups and combine each after the alignment calculation. This will produce two slightly different tilt series from the same data that can be processed in parallel. This split at the level of frames has the advantage that it does not increase the tilting step by a factor of 2, as caused by the split that separates tilt images in odd and even tilts. However, this gain in the tilting step is made at the expense of a decrease by a factor of 2 in the SNR of each tilt image. To avoid this, there are two possible strategies:
  1. Allowing the two tilt series to share some frames so that the SNR loss is not 2.
  2. Calculating some of the processing parameters (alignments, CTF information, picking coordinates, etc.) from the tilt series calculated using all frames instead of the two subtilt series.

  Another advantage of this approach is that there is no need for different algorithms. Both tilt series generated from the subset of movie frames can be processed with the same software, and the validation process is still valid since the difference in the processing is not coming from the underlying algorithm of the software but from minor discrepancies in the data with the particularity that these discrepancies should not affect the final obtained results.

Another obvious drawback common to some of the tools in this chapter is that this strategy implies a consistent disk space and computational power duplication at every performed step. Nevertheless, ensuring the quality of the final results might justify this duplication of space and time.

- *Validation via features*. As shown in the previous sections, the cryoET images and reconstructions present a low SNR. It complicates image processing and is the main reason for the need for validation. Despite the broad variety of algorithms discussed in this chapter, validation methods are uncommon in cryoET. An alternative is to use the own features of the data set. For instance, in one of the previous sections, we used the fiducial shape as a feature to validate the tilt series alignment. This approach can be generalized to other features and other steps of the workflow. In general, our tomograms contain different proteins. In some cases, we know the structure of some of them; in others, we know a characteristic of the protein (for instance, it is a membrane protein). These properties can be used as a validation tool. In a subtomogram averaging workflow, it is possible to reconstruct the structure of a well-known protein to ensure that the image processing is correct. Then, in the second step, the protein of interest is processed.
- *Map back*: It is a postprocessing tool for validation and visualization. This method consists of placing the final average subtomogram in the positions (coordinates) of the tomogram where subtomograms were picked, with their original orientation (alignment). The result of performing a map back will be the original tomograms modified to highlight the different kinds of molecules and their conformations. The Scipion protocol `xmipptomo - map back subtomos` has several "painting" options to highlight, binarize, or erase where the subtomograms are, combined with the possibility to remove the background. In this way, the result of the map back can be used as a validation tool to check the alignment of the particles when they have a reference, such as a membrane, as it allows to check whether the subtomograms placed back in the tomograms are oriented in the right direction or not. For example, imagine a membrane protein that appears in the outer part of a membrane. The subtomogram averages in the tomogram resulting from the map should also be pointing out of the membrane. However, if the particles have been poorly aligned (due to low signal), some may be "flipped" and point to the inner part of the membrane. As we know that the protein should point outward, we can conclude that those particles have been aligned incorrectly, and we can remove them or try to improve their alignment.

Bibliography

Emerging topics in cryoET validation

Finally, we review some emerging topics in cryoET validation that, in our opinion, should be considered as possible future strategies of validation that will increase the robustness of the results obtained in the future processing and their interpretation and the biological meaning extracted from them.

- *Workflow traceability and availability*. This topic is not exclusive to cryoET. However, considering it is a novel technique gaining popularity among scientists and the number of software tools and algorithms available is rapidly increasing, it especially benefits from these guidelines. Also, the authors know that following the recommendations in this chapter increases the workflow complexity, so the scientist is encouraged to keep a record of the following steps to provide the information needed for a more conscious posterior judgment of the results.

  First is workflow traceability. For instance, Scipion helps to keep track of the executed steps (de la Rosa-Trevin et al., 2016; Jimenez de la et al., 2022). This software proposes a framework in which several tools from different sources can be combined, not only facilitating the labor of the scientist but also avoiding translation errors between packages and ensuring interoperability between them. We highly recommend the usage of this tool, being an excellent solution to keep track of the executed steps, the parameters used in each execution, the inputs and output for each step, and, in general, the workflow arrangement.

  The second is workflow availability. Keeping track of all the image processing steps and their parameters undoubtedly helps researchers in their analysis task. Nonetheless, these benefits accrue not only to the scientist but not to the whole scientific community. Although the community has been increasing efforts to maintain public databases (Iudin et al., 2022), image processing workflows are seldom shared. We encourage researchers to report raw data, results, and information describing the workflows followed. This kind of software, such as Scipion, also facilitates this task, making sharing the workflow structure, parameters, and intermediate results easy.
- *Data management and standardization*. Nowadays, cryoET is suffering the revolution that SPA suffered some years ago. The number of algorithms is growing, as well as the number of elucidated structures. Much new software engineering has appeared for solving specific problems pushing the cryoET resolution limits. Thus, there are many packages such as IMOD (Kremer et al., 1996; Mastronarde and Held, 2017), Dynamo

(Castaño-Díez et al., 2012), EMAN (Galaz-Montoya et al., 2015), RELION (Zivanov et al., 2022), emClarity (Himes and Zhang, 2018), M (Tegunov et al., 2021), Xmipp (Strelák et al., 2020), among many others that cover the complete image processing workflow. The variety of software implies communication drawbacks and data conversion between different formats. To validate the result by reproducing the same result with two software packages, these problems have to be solved. This is the case with the consensus approaches. Some frameworks have been proposed to address these problems, such as Scipion-Tomo (Jimenez de la et al., 2022). However, data standardization is still a need. Proof of that is the previous paragraph, where the data reproducibility and availability were considered. We propose standardizing the cryoET data as FAIR (an acronym for findable, accessible, interoperable, and reusable). It simplifies and ensures the reproducibility of results and pushes forward new developments in cryoET.

- *Combination of experimental techniques*. Generally, cryoET reconstructions are only carried out with electron microscopy images. Other information sources are not used during image processing. However, richer information can be used to validate image processing. In this regard, correlative microscopy plays a prominent role. It can provide complementary information to cryoET and guide image processing. The recent advances in deep learning have also provided excellent tools such as AlphaFold (Jumper et al., 2021). Subtomogram averaging can take advantage of steps such as picking or initial model. Other experimental techniques such as protein−protein interaction or mass spectrometry can also be applied to understand the macromolecular behavior in the tomogram or help during image processing.

Acknowledgment

The authors acknowledge the economical support from the Ministry of Science, Innovation and Universities (BDNS n. 716450) to the Instruct Image Processing Center (I2PC) as part of the Spanish participation in Instruct-ERIC, the European Strategic Infrastructure Project (ESFRI) in the area of Structural Biology, Grant PID2022-136594NB-I00 funded by MCIN/AEI/10.13039/501100011033/and "ERDF A way of making Europe," by the "European Union" and Grant PRE2020−093527 funded by MCIN/AEI/10.13039/501100011033 and by "ESF Investing in your future," "Comunidad Autónoma de Madrid" through Grant: S2022/BMD-7232, European Union (EU) and Horizon 2020 through grant: HighResCells (ERC - 2018 - SyG, Proposal: 810057). This research work was also funded by the European Commision − NextGenerationEU(Regulation 2020/2094), through CSIC's Global Health Platform (PTI Salud Global).

Bibliography

References

Abrishami, V., Vargas, J., Li, X., Cheng, Y., Marabini, R., Sorzano, C.O.S., Carazo, J.M., 2015. Alignment of direct detection device micrographs using a robust optical flow approach. J. Struct. Biol. 189, 163–176.

Andersen, A., Kak, A., 1984. Simultaneous algebraic reconstruction technique (sart): a superior implementation of the art algorithm. Ultrason. Imag. 6 (1), 81–94.

Beckers, M., Sachse, C., 2020. Permutation testing of Fourier shell correlation for resolution estimation of cryo-em maps. J. Struct. Biol. 212 (1), 107579.

Brilot, A.F., Chen, J.Z., Cheng, A., Pan, J., Harrison, S.C., Potter, C.S., Carragher, B., Henderson, R., Grigorieff, N., 2012. Beam-induced motion of vitrified specimen on holey carbon film. J. Struct. Biol. 177 (3), 630–637.

Buchholz, T.-O., Jordan, M., Pigino, G., Jug, F., 2019a. Cryo-care: content-aware image restoration for cryo-transmission electron microscopy data. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, pp. 502–506.

Buchholz, T.-O., Krull, A., Shahidi, R., Pigino, G., Jékely, G., Jug, F., 2019b. Content-aware image restoration for electron microscopy. Meth. Cell Biol. 152, 277–289.

Cardone, G., Grünewald, K., Steven, A.C., 2005. A resolution criterion for electron tomography based on cross-validation. J. Struct. Biol. 151 (2), 117–129. https://www.sciencedirect.com/science/article/pii/S1047847705001085.

Cardone, G., Heymann, J., Steven, A., 2013. One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. J. Struct. Biol. 184, 226–236.

Castaño-Díez, D., Kudryashev, M., Arheit, M., Stahlberg, H., 2012. Dynamo: a flexible, user-friendly development tool for subtomogram averaging of cryo-em data in high-performance computing environments. J. Struct. Biol. 178 (2), 139–151 special Issue: Electron Tomography. https://www.sciencedirect.com/science/article/pii/S1047847711003650.

Castano-Diez, D., Scheffer, M., Al-Amoudi, A., Frangakis, A.S., 2010. Alignator: a gpu powered software package for robust fiducial-less alignment of cryo tilt-series. J. Struct. Biol. 170 (1), 117–126. https://www.sciencedirect.com/science/article/pii/S1047847710000262.

Castano-Diez, D., Kudryashev, M., Stahlberg, H., 2017. Dynamo catalogue: geometrical tools and data management for particle picking in subtomogram averaging of cryo-electron tomograms. J. Struct. Biol. 197 (2), 135–144 electron Tomography. https://www.sciencedirect.com/science/article/pii/S1047847716301113.

Chen, S., McMullan, G., Faruqi, A., Murshudov, G., Short, J., Scheres, S., Henderson, R., 2013. High-resolution noise substitution to measure overfitting and validate resolution in 3d structure determination by single particle electron cryomicroscopy. Ultramicroscopy 135, 24–35.

Chen, M., Dai, W., Sun, S.Y., Jonasch, D., He, C.Y., Schmid, M.F., Chiu, W., Ludtke, S.J., 2017. Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. Nat. Methods 14 (10), 983–985.

Danev, R., Baumeister, W., 2017. Expanding the boundaries of cryo-em with phase plates. Curr. Opin. Struct. Biol. 46, 87–94.

de la Rosa-Trevín, J.M., Otón, J., Marabini, R., Zaldívar, A., Vargas, J., Carazo, J.M., Sorzano, C.O.S., 2013. Xmipp 3.0: an improved software suite for image processing in electron microscopy. J. Struct. Biol. 184 (2), 321–328.

de la Rosa-Trevin, J., Quintana, A., del Cano, L., Zaldivar-Peraza, A., Foche, I., Gutierrez, J., Gomez-Blanco, J., Burguet-Castells, J., Cuenca, J., Abrishami, V., Vargas, J., Oton, J., Sharov, G., Navas, J., Conesa, P., Vilas, J., Marabini, R., Sorzano, C., Carazo, J., 2016. Scipion: a software framework toward integration, reproducibility, and validation in 3D electron microscopy. J. Struct. Biol. 195, 93–99.

Bibliography

de Teresa, I., Goetz, S.K., Mattausch, A., Stojanovska, F., Zimmerli, C.E., Toro-Nahuelpan, M., Cheng, D.W., Tollervey, F., Pape, C., Beck, M., Kreshuk, A., Mahamid, J., Zaugg, J., 2023. Convolutional networks for supervised mining of molecular patterns within cellular context. Nat. Methods 20. https://doi.org/10.1101/2022.04.12.488077.

Diebolder, C., Faas, F., Koster, A., Koning, R., 2015. Conical Fourier shell correlation applied to electron tomograms. J. Struct. Biol. 190 (2), 215–223. https://www.sciencedirect.com/science/article/pii/S1047847715000726.

Fernández, J., Li, S., Crowther, R., 2006. Ctf determination and correction in electron cryotomography. Ultramicroscopy 106 (7), 587–596. https://www.sciencedirect.com/science/article/pii/S0304399106000222.

Fernandez, J.-J., Li, S., 2021. Tomoalign: a novel approach to correcting sample motion and 3d ctf in cryoet. J. Struct. Biol. 213 (4), 107778. https://www.sciencedirect.com/science/article/pii/S1047847721000836.

Fernandez, J.-J., Li, S., Bharat, T.A., Agard, D.A., 2018. Cryo-tomography tilt-series alignment with consideration of the beam-induced sample motion. J. Struct. Biol. 202 (3), 200–209. https://www.sciencedirect.com/science/article/pii/S1047847718300285.

Frangakis, A.S., Hegerl, R., 2001. Noise reduction in electron tomographic reconstructions using nonlinear anisotropic diffusion. J. Struct. Biol. 135 (3), 239–250. https://www.sciencedirect.com/science/article/pii/S1047847701944065.

Frangakis, A.S., 2021. It's noisy out there! a review of denoising techniques in cryo-electron tomography. J. Struct. Biol. 213 (4), 107804. https://www.sciencedirect.com/science/article/pii/S104784772100109X.

Frangakis, A.S., 2022. Mean curvature motion facilitates the segmentation and surface visualization of electron tomograms. J. Struct. Biol. 214 (1), 107833. https://www.sciencedirect.com/science/article/pii/S104784772200003X.

Galaz-Montoya, J.G., Flanagan, J., Schmid, M.F., Ludtke, S.J., 2015. Single particle tomography in eman2. J. Struct. Biol. 190 (3), 279–290. https://www.sciencedirect.com/science/article/pii/S104784771500101X.

Gilbert, P., 1972. Iterative methods for the three-dimensional reconstruction of an object from projections. J. Theor. Biol. 36 (1), 105–117.

Gordon, R., Bender, R., Herman, G., 1970. Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography. J. Theor. Biol. 29 (3), 471–481.

Grant, T., Grigorieff, N., 2019. Measuring the optimal exposure for single particle cryo-EM using a 2.6Í reconstruction of rotavirus VP6. Elife 4.

Harauz, G., van Heel, M., 1986. Exact filters for general geometry three dimensional reconstruction. Optik 73, 146–156.

Henderson, R., 2013. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. Proc. Natl. Acad. Sci. U. S. A. 110 (45), 18037–18041.

Heymann, B., 2018. Single particle reconstruction and validation using bsoft for the map challenge. J. Struct. Biol. 204 (1), 90–95.

Heynmann, B., 2015. Validation of 3dem reconstructions: the phantom in the noise. AIMS Biophy. 2, 21.

Himes, B., Zhang, P., 2018. emclarity: software for high-resolution cryo-electron tomography and subtomogram averaging. Nat. Methods 15, 955–961.

Hu, M., Zhang, Q., Yang, J., Li, X., 2020. Unit quaternion description of spatial rotations in 3d electron cryo-microscopy. J. Struct. Biol. 212 (3), 107601. https://www.sciencedirect.com/science/article/pii/S104784772030174X.

Huang, X., Li, S., Gao, S., 2018. Applying a modified wavelet shrinkage filter to improve cryo-electron microscopy imaging. J. Comput. Biol. 25 (9), 1050–1058.

Bibliography

Iudin, A., Korir, P.K., Somasundharam, S., Weyand, S., Cattavitello, C., Fonseca, N., Salih, O., Kleywegt, G., Patwardhan, A., 2022. EMPIAR: the electron microscopy public image archive. Nucl. Acid. Res. 51 (D1), D1503–D1511.

Jimenez de la Morena, J., Conesa, P., Fonseca, Y., de Isidro-Gomez, F., Herreros, D., Fernandez-Gimenez, E., Strelak, D., Moebel, E., Buchholz, T., Jug, F., Martinez-Sanchez, A., Harastani, M., Jonic, S., Conesa, J., Cuervo, A., Losana, P., Sanchez, I., Iceta, M., del Cano, L., Gragera, M., Melero, R., Sharov, G., Castaño-Diez, D., Koster, A., Piccirillo, J., Vilas, J., Oton, J., Marabini, R., Sorzano, C., Carazo, J., 2022. ScipionTomo: towards cryo-electron tomography software integration, reproducibility, and validation. J. Struct. Biol. 214 (3), 107872.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with alphafold. Nature 583–589.

Kirkland, E., 2010. Advanced Computing in Electron Microscopy.

Kremer, J.R., Mastronarde, D.N., McIntosh, J., 1996. Computer visualization of three-dimensional image data using imod. J. Struct. Biol. 116 (1), 71–76. https://doi.org/10.1006/jsbi.1996.0013. https://www.sciencedirect.com/science/article/pii/S1047847796900131.

Kucukelbir, A., Sigworth, F.J., Tagare, H.D., 2014. Quantifying the local resolution of cryo-em density mapss. Nat. Methods 11, 63–65.

Lamm, L., Righetto, R.D., Wietrzynski, W., Pöge, M., Martinez-Sanchez, A., Peng, T., Engel, B.D., 2022. Membrain: a deep learning-aided pipeline for detection of membrane proteins in cryo-electron tomograms. Comput. Meth. Progr. Biomed. 224, 106990. https://www.sciencedirect.com/science/article/pii/S0169260722003728.

Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., Cheng, Y., 2013. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat. Methods 10 (6), 584–590.

Lučić, V., Rigort, A., Baumeister, W., 2013. Cryo-electron tomography: the challenge of doing structural biology in situ. J. Cell Biol. 202 (3), 407–419. https://doi.org/10.1083/jcb.201304193.

Mahajan, V.N., 1991. Aberration theory made simple. SPIEL 6.

Martinez-Sanchez, A., Garcia, I., Asano, S., Lucic, V., Fernandez, J.-J., 2014. Robust membrane detection based on tensor voting for electron tomography. J. Struct. Biol. 186 (1), 49–61. https://www.sciencedirect.com/science/article/pii/S1047847714000495.

Martinez-Sanchez, A., Kochovski, Z., Laugks, U., Meyer zum Alten Borgloh, J., Chakraborty, S., Pfeffer, S., V. Lucic, B.W., 2020. Template-free detection and classification of membrane-bound complexes in cryo-electron tomograms. Nat. Methods 17, 209–2016.

Mastronarde, D.N., Held, S.R., 2017. Automated tilt series alignment and tomographic reconstruction in imod. J. Struct. Biol. 197 (2), 102–113. https://www.sciencedirect.com/science/article/pii/S1047847716301526.

McMullan, G., Faruqi, A., Henderson, R., 2016. Chapter one - direct electron detectors. In: Crowther, R. (Ed.), The Resolution Revolution: Recent Advances in cryoEM, Vol. 579 of Methods in Enzymology. Academic Press, pp. 1–17.

Messaoudi, C., Boudier, T., Sorzano, C., Marco, S., 2007. Tomoj: tomography software for three-dimensional reconstruction in transmission electron microscopy. BMC Bioinf. 8 (288).

Bibliography

Moreno, J.J., Martinez-Sanchez, A., Martinez, J.A., Garzon, E.M., Fernandez, J.J., 2018. TomoEED: fast edge-enhancing denoising of tomographic volumes. Bioinformatics 34 (21), 3776–3778. https://doi.org/10.1093/bioinformatics/bty435.

Navarro, P.P., 2022. Quantitative cryo-electron tomography. Front. Mol. Biosci. 9. https://www.frontiersin.org/articles/10.3389/fmolb.2022.934465.

Penczek, P.A., 2020. Reliable cryo-EM resolution estimation with modified Fourier shell correlation. IUCrJ 7 (6), 995–1008.

Radermacher, M., 1992. Weighted Back-Projection Methods. Springer US, Boston, MA, pp. 91–115. https://doi.org/10.1007/978-1-4757-2163-8_5.

Ramirez-Aportela, E., Mota, J., Conesa, P., Carazo, J., Sorzano, C., 2019. *DeepRes*: a new deep-learning- and aspect-based local resolution method for electron-microscopy maps. IUCrJ 6 (6).

Robertson, M.J., Meyerowitz, J.G., Skiniotis, G., 2020. Cryo-em as a powerful tool for drug discovery. Bioorg. Med. Chem. Lett. 30 (22), 127524.

Rohou, A., Grigorieff, N., 2015. CTFFIND4: fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. 192 (2), 216–221.

Rosenthal, P., Henderson, H., 2003. Determination of particle orientation,absolute hand, and contrast loss in single-particle electron cryomicroscopy. J. Mol. Biol. 333, 721–745.

Sanchez-Garcia, R., Segura, J., Maluenda, D., Carazo, J., Sorzano, C., 2018. *Deep Consensus*, a deep learning-based approach for particle pruning in cryo-electron microscopy. IUCrJ 5 (6), 854–865.

Scheres, S.H.W., 2012. A Bayesian view on cryo-EM structure determination. J. Mol. Biol. 415 (2), 406–418.

Scheres, S., 2014. Beam-induced motion correction for sub-megadalton cryo-em particles. Elife 3, e03665.

Seifer, S., Elbaum, M., 2022. Clusteralign: a fiducial tracking and tilt series alignment tool for thick sample tomography. Biolog. Imag. 2, e7. https://doi.org/10.1017/S2633903X22000071.

Sorzano, C., Jonic, S., Núñez-Ramírez, R., Boisset, N., Carazo, J., 2007. Fast, robust, and accurate determination of transmission electron microscopy contrast transfer function. J. Struct. Biol. 160 (2), 249–262. https://www.sciencedirect.com/science/article/pii/S104784770700202X.

Sorzano, C.O.S., Marabini, R., Vargas, J., Otón, J., Cuenca-Alba, J., Quintana, A., de la Rosa-Trevín, J.M., Carazo, J.M., 2014a. Interchanging Geometry Conventions in 3DEM: Mathematical Context for the Development of Standards. Springer New York, New York, NY, pp. 7–42. https://doi.org/10.1007/978-1-4614-9521-5_2.

Sorzano, C.O.S., Vargas, J., de la Rosa-Trevín, J.M., Zaldívar-Peraza, A., Otón, J., Abrishami, V., Foche, I., Marabini, R., Caffarena, G., Carazo, J.M., 2014b. Outlier detection for single particle analysis in electron microscopy. In: Proceedings of the International Work-Conference on Bioinformatics and Biomedical Engineering. IWBBIO, p. 950.

Sorzano, C., Vargas, J., Oton, J., de la Rosa-Trevin, J., Vilas, J., Kazemi, M., Melero, R., del cano, L., Cuenca, J., Gomez-Blanco, J., Carazo, J., 2017a. A survey of the use of iterative reconstruction algorithms in electron microscopy. BioMed Res. Int., 6482567

Sorzano, C., Vargas, J., Oton, J., Abrishami, V., de la Rosa-Trevin, J., Gomez-Blanco, J., Vilas, J., Marabini, R., Carazo, J., 2017b. A review of resolution measures and related aspects in 3d electron microscopy. Prog. Biophys. Mol. Biol. 124, 1–30.

Sorzano, C.O.S., Jiménez-Moreno, A., Maluenda, D., Martínez, M., Ramírez-Aportela, E., Krieger, J., Melero, R., Cuervo, A., Conesa, J., Filipovic, J., Conesa, P., del Caño, L., Fonseca, Y.C., Jiménez-de la Morena, J., Losana, P., Sánchez-García, R., Strelak, D., Fernández-Giménez, E., de Isidro-Gómez, F.P., Herreros, D., Vilas, J.L., Marabini, R., Carazo, J.M., 2022. On bias, variance, overfitting, gold standard and consensus in single-particle analysis by cryo-electron microscopy. Acta Crystallogr. D 78 (4), 410–423.

Bibliography

Strelák, D., Filipovic, J., Jimenez-Moreno, A., Carazo, J.M., Sanchez Sorzano, C.O., 2020. Flexalign: an accurate and fast algorithm for movie alignment in cryo-electron microscopy. Electronics 9 (6).

Strelak, D., Jiménez-Moreno, A., Vilas, J.L., Ramírez-Aportela, E., Sánchez-García, R., Maluenda, D., Vargas, J., Herreros, D., Fernández-Giménez, E., de Isidro-Gómez, F.P., Horacek, J., Myska, D., Horacek, M., Conesa, P., Fonseca-Reyna, Y.C., Jiménez, J., Martínez, M., Harastani, M., Jonić, S., Filipovic, J., Marabini, R., Carazo, J.M., Sorzano, C.O.S., 2021. Advances in xmipp for cryo−electron microscopy: from xmipp to scipion. Molecules 26 (20).

Su, M., 2019. goctf: geometrically optimized ctf determination for single-particle cryo-em. J. Struct. Biol. 205 (1), 22−29.

Tan, Y.Z., Baldwin, P.R., Davis, J.H., Williamson, J.R., Potter, C.S., Carragher, B., Lyumkis, D., 2017. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. Nat. Methods 14 (8), 793−796.

Tegunov, D., Xue, L., Dienemann, C., Cramer, P., Mahamid, J., 2021. Multi-particle cryo-em refinement with m visualizes ribosome-antibiotic complex at 3.5a in cells. Nat. Methods 18, 186−193.

Thompson, R.F., Walker, M., Siebert, C.A., Muench, S.P., Ranson, N.A., 2016. An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. Methods 100, 3−15.

Thon, F., 1966. Zur defokussierunsabhängigkeit des phasenkontrastes bei der elektronenmikroskopischen abbildung. Z. Naturforsch. 21a, 476−478.

Turoňová, B., Schur, F.K., Wan, W., Briggs, J.A., 2017. Efficient 3d-ctf correction for cryo-electron tomography using novactf improves subtomogram averaging resolution to 3.4 Å. J. Struct. Biol. 199 (3), 187−195. https://www.sciencedirect.com/science/article/pii/S1047847717301272.

Turonova, B., Hagen, W., Obr, M., Mosalaganti, S., Beugelink, J., Zimmerli, C., Beck, M., 2020. Benchmarking tomographic acquisition schemes for high-resolution structural biology. Nat. Commun. 11 (876).

Unwin, P.N.T., Huxley, H.E., Klug, A., 1971. Phase contrast and interference microscopy with the electron microscope. Philos. Trans. R. Soc. Lond. B Biol. Sci. 261 (837), 95−104.

Van Drie, J.H., Tong, L., 2022. Drug discovery in the era of cryo-electron microscopy. Trend. Biochem. Sci. 47 (2), 124−135.

van Heel, M., 2013. Finding trimeric hiv-1 envelope glycoproteins in random noise. Proc. Natl. Acad. Sci. U. S. A. 110 (45), E4175−E4177.

Vargas, J., Abrishami, V., Marabini, R., de la Rosa-Trevín, J.M., Zaldivar, A., Carazo, J.M., Sorzano, C.O.S., 2013. Particle quality assessment and sorting for automatic and semiautomatic particle-picking techniques. J. Struct. Biol. 183 (3), 342−353.

Vargas, J., Oton, J., Marabini, R., Carazo, J., Sorzano, C., 2016. Particle alignment reliability in single particle electron cryomicroscopy: a general approach. Sci. Rep. 6, 21626.

Vilas, J.L., Tagare, H.D., 2023. New measures of anisotropy of cryo-em maps. Nat. Methods 20, 1021−1024. https://doi.org/10.1038/s41592-023-01874-3.

Vilas, J.L., Gomez-Blanco, J., Conesa, P., Melero, R., de la Rosa Trevın, J.M., Oton, J., Cuenca, J., Marabini, R., Carazo, J.M., Vargas, J., Sorzano, C.O.S., 2018. Monores: automatic and accurate estimation of local resolution for electron microscopy maps. Structure 26, 337−344.

Vilas, J.L., Oton, J., Messaoudi, C., Melero, R., Conesa, P., Ramirez-Aportela, E., Mota, J., Martinez, M., Jimenez, A., Marabini, R., Carazo, J.M., Vargas, J., Sorzano, C.O.S., 2020a. Measurement of local resolution in electron tomography. J. Struct. Biol. X 4, 100016.

Vilas, J., Heymann, J., Tagare, H., Ramirez-Aportela, E., Carazo, J., Sorzano, C., 2020b. Local resolution estimates of cryoem reconstructions. Curr. Opin. Struct. Biol. 64, 74–78.

Vilas, J.L., Tagare, H.D., Vargas, J., Carazo, J.M., Sorzano, C.O.S., 2020c. Measuring local-directional resolution and local anisotropy in cryo-EM maps. Nat. Commun. 11, 55.

Vilas, J.L., Carazo, J.M., Sorzano, C.O.S., 2022. Emerging themes in cryoem-single particle analysis image processing. Chem. Rev. 122 (17), 13915–13951. https://doi.org/10.1021/acs.chemrev.1c00850.

Voortman, L.M., Stallinga, S., Schoenmakers, R.H., van Vliet, L.J., Rieger, B., 2011. A fast algorithm for computing and correcting the ctf for tilted, thick specimens in tem. Ultramicroscopy 111 (8), 1029–1036. https://www.sciencedirect.com/science/article/pii/S0304399111000878.

Vulović, M., Voortman, L.M., van Vliet, L.J., Rieger, B., 2014. When to use the projection assumption and the weak-phase object approximation in phase contrast cryo-em. Ultramicroscopy 136, 61–66. https://www.sciencedirect.com/science/article/pii/S0304399113002088.

Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., Raisch, T., Prumbaum, D., Quentin, D., Roderer, D., Tacke, S., Siebolds, B., Schubert, E., Shaikh, T.R., Lill, P., Gatsogiannis, C., Raunser, S., 2019. Sphire-cryolo is a fast and accurate fully automated particle picker for cryo-em. Commun. Biol. 2 (1), 218.

Wan, W., Briggs, J., 2016. Chapter thirteen - cryo-electron tomography and subtomogram averaging. In: Crowther, R. (Ed.), The Resolution Revolution: Recent Advances in cryoEM, Vol. 579 of Methods in Enzymology. Academic Press, pp. 329–367. https://www.sciencedirect.com/science/article/pii/S0076687916300325.

Xiong, Q., Morphew, M.K., Schwartz, C.L., Hoenger, A.H., Mastronarde, D.N., 2009. Ctf determination and correction for low dose tomographic tilt series. J. Struct. Biol. 168 (3), 378–387. https://www.sciencedirect.com/science/article/pii/S1047847709002433.

Zhang, K., 2016. Gctf: real-time ctf determination and correction. J. Struct. Biol. 193 (1), 1–12.

Zheng, S.Q., Palovcak, E., Armache, J.-P., Verba, K.A., Cheng, Y., Agard, D.A., 2017. Motioncor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods 14, 331–332.

Zheng, S., Wolff, G., Greenan, G., Chen, Z., Faas, F.G., Bárcena, M., Koster, A.J., Cheng, Y., Agard, D.A., 2022. Aretomo: an integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction. J. Struct. Biol. X 6, 100068. https://www.sciencedirect.com/science/article/pii/S2590152422000095.

Zivanov, J., Otón, J., Ke, Z., von Kügelgen, A., Pyle, E., Qu, K., Morado, D., Castano-Diez, D., Zanetti, G., Bharat, T.A., Briggs, J.A., Scheres, S.H., 2022. A bayesian approach to single-particle electron cryo-tomography in relion-4.0. Elife 11, e83724. https://doi.org/10.7554/eLife.83724.

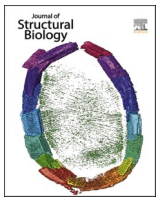
Appendix B: A deep learning approach to the automatic detection of alignment errors in cryo-electron tomographic reconstructions

Journal of Structural Biology 216 (2024) 108056

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi



Research Article

A deep learning approach to the automatic detection of alignment errors in cryo-electron tomographic reconstructions

F.P. de Isidro-Gómez [a,b], J.L. Vilas [a], P. Losana [a], J.M. Carazo [a,*], C.O.S. Sorzano [a,*]

[a] Biocomputing Unit, Centro Nacional de Biotecnologia (CNB-CSIC), Darwin, 3, Campus Universidad Autonoma, 28049 Cantoblanco, Madrid, Spain
[b] Univ. Autonoma de Madrid, 28049 Cantoblanco, Madrid, Spain

A B S T R A C T

Electron tomography is an imaging technique that allows for the elucidation of three-dimensional structural information of biological specimens in a very general context, including cellular *in situ* observations. The approach starts by collecting a set of images at different projection directions by tilting the specimen stage inside the microscope. Therefore, a crucial preliminary step is to precisely define the acquisition geometry by aligning all the tilt images to a common reference. Errors introduced in this step will lead to the appearance of artifacts in the tomographic reconstruction, rendering them unsuitable for the sample study. Focusing on fiducial-based acquisition strategies, this work proposes a deep-learning algorithm to detect misalignment artifacts in tomographic reconstructions by analyzing the characteristics of these fiducial markers in the tomogram. In addition, we propose an algorithm designed to detect fiducial markers in the tomogram with which to feed the classification algorithm in case the alignment algorithm does not provide the location of the markers. This open-source software is available as part of the Xmipp software package inside of the Scipion framework, and also through the command-line in the standalone version of Xmipp.

1. Introduction

Elucidation of three-dimensional (3D) structural information of biological specimens is one of the greatest assets of modern biology. Gathering structural and functional information helps scientists to understand the underlying organization of the different elements in the sample. Cryogenic electron tomography (cryo-ET) is an advanced imaging technique extensively employed in investigating biological complexes, enabling the study of the three-dimensional structure of various types of samples ranging from cellular environments to purified complexes. The technique involves freezing the specimen in its native hydrated state and imaging it using a Transmission Electron Microscope (TEM) under cryogenic conditions. Cryo-ET has become an essential tool for studying the structure and function of macromolecular complexes *in situ*, such as the machinery of cells and viruses, and has enabled researchers to gain unprecedented insights into biological processes at the molecular level. It has applications in various fields, including structural biology, virology, cell biology, and drug discovery (Robertson et al., 2020; Van Drie and Tong, 2022).

Electron tomography is based on the determination of three-dimensional structural information through the combination of projective images of the sample. A series of 2D projection images of the specimen are acquired at different viewing angles by tilting the specimen inside the TEM. The projective information in the series of images is computationally combined to reconstruct a three-dimensional volume, allowing for the visualization and analysis of the specimen's internal structures in full 3D. However, to combine the projective information accurately in three-dimensional space, it is imperative to correct for the sample's relative movements between images and estimate their acquisition orientation.

To make the alignment easier, high-contrast fiducial markers, such as colloidal gold particles, are frequently spread over the specimen to aid in correcting or compensating for any geometric distortions. By tracking the nanoparticles as the specimen is tilted, the precise orientation of the tilt axis can be established, and the 2D images can be aligned accordingly. Several approaches have been developed for the automatic detection and tracking of fiducial markers or other suitable landmarks along the tilt series to solve the geometry of the tilt series (Sorzano et al., 2009; Mastronarde and Held, 2017; Fernandez et al., 2018; Fernandez and Li, 2021; Seifer and Elbaum, 2022; Sorzano et al., 2020). Other approaches do not rely on the presence of trackable points to calculate the sample motion correction (Castano-Diez et al., 2010; Noble and Stagg, 2015). They are based on the successive matching of projections from partial reconstruction to align the tilt series (Zheng et al., 2022). The main focus of this work pertains to samples containing fiducial markers.

Once the geometry of the tilt series is solved, the tomographic reconstruction may be calculated. There are different reconstruction

* Corresponding author.
E-mail address: coss.eps@ceu.es (C.O.S. Sorzano).

https://doi.org/10.1016/j.jsb.2023.108056
Received 25 July 2023; Received in revised form 21 November 2023; Accepted 11 December 2023
Available online 14 December 2023

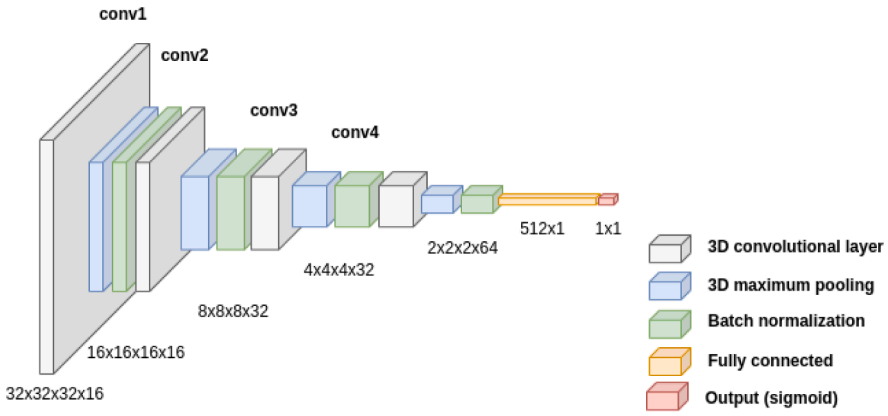


Fig. 1. Structure of the neural network for misalignment detection from fiducial markers.

methods with different underlying algorithms. Although the different approaches induce different characteristics in the final reconstruction, the eventual presence of misalignment errors always translates into reconstruction artifacts. Reconstruction methods can be classified into two families: direct methods and series expansion algorithms (Sorzano et al., 2017). Weighted Back Projection (WBP) (Radermacher et al., 1992) dominates the first group. In the second group, ART (algebraic reconstruction techniques) (Gordon et al., 1970), SIRT (simultaneous iterative reconstruction techniques) (Gilbert, 1972), or SART (simultaneous algebraic reconstruction techniques) (Andersen and Kak, 1984) are the most used. In this work, the method of preference for tomogram reconstruction is WBP because of its light computational requirements and popularity in the field.

Traditionally, scientists have relied on heuristic techniques to identify these artifacts. These methods typically involve visually inspecting the aligned tilt series, the tomogram, or performing various operations on them (such as filtering or projecting in different directions) to detect potential misalignment artifacts. Instead, this paper presents a deep-learning approach that automatically identifies misalignment with no needed intervention from the scientist. Our algorithm relies on studying the fiducials contained in the sample and the possible presence of artifacts due to misalignment. Therefore, to assess the quality of the alignment, we first need to locate these high-contrast regions within the tomogram. Fortunately, in many cases where a fiducial-based alignment is performed, the location of these fiducials is already defined during the alignment step. This process typically involves tracking individual gold beads throughout the tilt series. Consequently, the 3D location of the gold beads within the tomogram is directly known from the image alignment.

However, there are situations in which the fiducials 3D location is not available, either because the algorithm used for the alignment does not provide this information or because it is not available for the user. For those cases, we provide an additional algorithm to locate fiducials within the tomogram.

In summary, we address the need to automatically assess the performance of any tilt series alignment, whether it is fiducial-based or not. Thus, the task of manual inspection of the alignment results is avoided, which is both a tedious and error-prone process as it depends on the user's skill to identify artifacts in the tomogram. The algorithm presented in this work is an open-source tool that has been implemented in Xmipp (de la Rosa-Trevín et al., Nov 2013; Strelak et al., 2021) and it is accessible through its command-line interface. It is also accessible through the Scipion workflow engine (de la Rosa-Trevin et al., 2016), within its tomography environment ScipionTomo (Jimenez de la Morena et al., 2022).

2. Methods

This section presents our proposal of a neural network-based approach that will analyze and classify a set of fiducials present in the tomogram to measure the quality of the alignment performed over the tilt series. We introduce the neural network design, the training pipelines, and the tomogram classification strategies based on individual misalignment scoring of each fiducial. Additionally, we also provide an algorithmic approach to detect these fiducial markers in case this information is not available from the alignment step.

2.1. Classification of subtomograms

In the presented work, neural networks are used to classify each subtomogram based on the presence of misaligned artifacts. We have trained two neural networks with identical architecture. The first is trained to detect strong misalignment, while the second is trained to detect weak misalignment. It must be noted that the misalignment patterns that these two networks have to learn are completely different. In case of strong misalignment, the artifacts are easily discernible to the bare eye. However, in case of weak misalignment, the artifacts are very subtle, to the extent that even a trained researcher may find it challenging to determine whether a tomogram suffers from a slight misalignment. In this context, the classification error rate of the second network should be expected to be larger than that of the first one.

The input to our networks is a collection of small subtomograms centered around fiducial markers in the tomograms. The location of these markers may come from the 3D landmarks calculated by the tilt series alignment algorithm, the 3D coordinates extracted by the algorithm in the previous section, or both. We have standardized the input to our neural network to fit each fiducial in a box size of $32 \times 32 \times 32$ voxels. Depending on its size, the extraction sampling rate is adjusted so that it provides an appropriate scope of the marker, including the potential misalignment artifact generated around it. As is common in neural networks, we normalize each subtomogram to have a mean of zero and a standard deviation of one.

The network output ranges from 0 to 1 depending on the network's belief that the input subtomogram is correctly aligned (1) or presents some degree of misalignment (0).

The architecture of the two networks is shown in Fig. 1. The network design consists of four convolutional layers with a rectified linear unit activation (ReLu), maximum pooling, and batch normalization applied at each step. This is followed by a fully connected layer of 512 neurons with a dropout of 20%. Finally, a sigmoid activation function follows the last layer. This design involves 110.897 parameters (110.641 of them are trainable).

We have used the EMPIAR-10064 (Khoshouei et al., 2017), EMPIAR-10164 (Schur et al., 2016), EMPIAR-10364 (Burt et al., 2020), and,

Table 1
Distribution of tilt series and subtomograms obtained from each training dataset, including its distribution in different groups depending on the alignment group, the source of the subtomogram coordinates, and, inside this, the preprocessing performed in each of them.

			From alignment				From PHC			
Dataset	Alignment	TS	Raw	Dose	CTF	Dose + CTF	Raw	Dose	CTF	Dose + CTF
10064	Aligned	2	7	0	7	0	8	0	8	0
	Weak	2	17	0	17	0	12	0	11	0
10164	Aligned	32	351	351	351	351	1352	1250	1287	1251
	Weak	11	114	114	114	114	472	485	467	487
10364	Aligned	18	261	261	261	261	837	855	735	755
10453	Aligned	130	2015	2015	1995	1995	1626	2126	2034	2024
	Weak	46	817	817	817	817	567	892	839	839
	Strong	61	869	869	869	869	874	769	676	703

EMPIAR-10453 (Turoňová et al., 2020) collections of tomograms to train both networks. These datasets comprise 302 tilt series in total that have been automatically aligned using IMOD inside Scipion. We have manually labeled the degree of misalignment of each alignment with the following result among all datasets: strong misalignment (61), weak misalignment (59), and properly aligned (182). The distribution of the tilt series for each dataset between each group is summarized in Table 1. All the datasets follow a similar acquisition pattern; each of them acquired with a tilting from -60° to 60° but composed of 41 images for datasets EMPIAR-10453 and EMPIAR-10164, and composed of 61 images for datasets EMPIAR-10064 and EMPIAR-10364.

We classified the tilt series by visual inspection of the tilt image, not relying on the reconstructed tomograms for classification. In this inspection, we search for relative movements of the images along the series or a miscalculation of the tilt axis. To settle whether the misalignment is strong or weak, we look at whether the miscalculation in the alignment is minor (a few pixels) or more pronounced. In this regard, this classification is more accurate as small errors introduced in the alignment can be easily detected in the tilt series. Still, they are hardly noticeable in the tomogram.

The authors consider that the strength of misalignment artifacts is neglectable to the network compared to the possible effects of contrast transfer function (CTF) or dose correction to the data. Nonetheless, these calculations have been considered also in the training dataset with the purpose of making the network more robust to differences in the preprocessing of the input data. Thus, for each dataset, we have generated for each reconstructed tomogram its counterparts by applying CTF, dose correction, and a combination of both. This has not been possible for the EMPIAR-10064 dataset, since no dose information is provided (only CTF correction considered). This is also useful to increase the training dataset. In the case of dose information is available, it implies quadruplicating the available amount of data (only duplicating in case it is not).

Then, for each dataset with different preprocessing steps, we extracted the coordinates corresponding to the fiducial markers provided by the IMOD tilt series alignment algorithm (in total, across all datasets, there were 10482 from the strongly misaligned subset, 3758 from the weakly misaligned, and 3476 from the aligned one). Additionally, we extracted the coordinates provided by the fiducial marker detection algorithm (in total, across all datasets, there were 20773 from the strongly misaligned subset, 5071 from the weakly misaligned, and 3022 from the aligned one). The distribution of extracted subtomograms obtained from each dataset and group can be observed in Table 1.

As can be observed, there is an imbalance in the number of elements in each training set. To address this issue, a dynamic training system has been implemented, ensuring a constant proportionality among the different groups that input the network in each training round.

We have extended the training dataset incorporating synthetically generated misaligned tomograms to train the first network. These misaligned tomograms were generated by individually manipulating the alignment parameters of 84 tilt series that had previously been correctly classified as aligned by the network. For this alignment manipulation, the alignment parameters have been distorted with different kinds of alignment errors with varying strengths. Assuming that the tilt axis is *Y*, the misalignment patterns that we have simulated are:

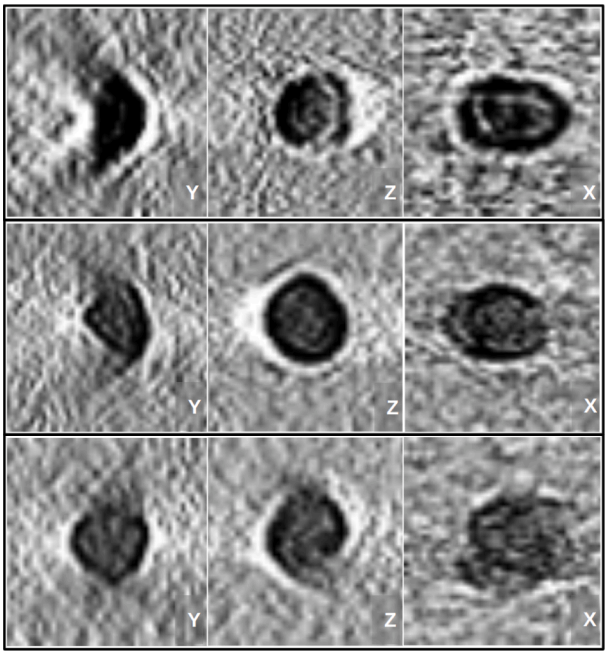


Fig. 2. Slices of the same fiducial marker reconstruction under different misalignments. The left column shows the *XZ*-plane, the center column the *XY*-plane, and the right column the *YZ*-plane. The first row presents bananas (incremental shift in *X*), the second line presents twisters (rotation of the tilt axis), and the last row birds (incremental shift in *Y* direction).

- Bananas: a constant offset in *X* added to all images in the series.
- Twisters: a constant offset in the in-plane rotation added to all images in the series.
- Birds: an incremental displacement in the *Y* direction.

Fig. 2 shows an example of each one of these artifacts. For the network training, a total of 4921 subtomograms presenting bananas, 3462 presenting birds, and 4105 presenting twisters, have been added to the training dataset.

The first network is trained to detect those subtomograms extracted from tomograms reconstructed from tilt series that exhibit strong misalignment or an artificially simulated one. The second network is trained to separate the weakly misaligned subtomograms from those correctly aligned. To take advantage of the knowledge acquired by the first network, we initialized the weights of the second network with the weights of the first one.

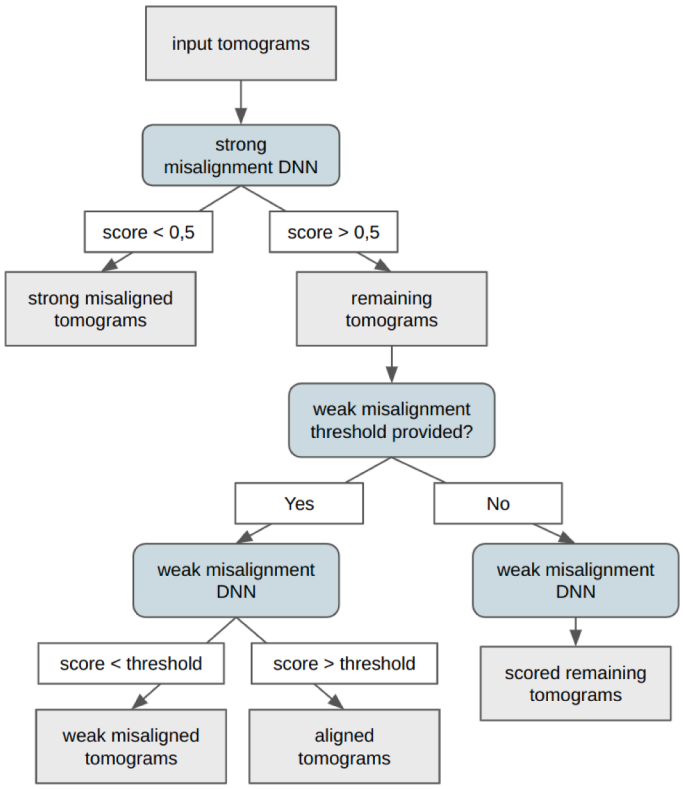


Fig. 3. Workflow for tomogram alignment classification.

2.2. Classification of tomograms

The neural networks of the previous section tell us whether a given subtomogram presents any degree of misalignment. However, we cannot judge a tomogram based on a single subtomogram. Instead, we should use all the fiducial markers identified in it. Two different alternative approaches are proposed. The first and recommended one is the calculation of the mean of the set of scores obtained from every subtomogram. Then, a tomogram is classified as misaligned or not based on whether its average score is below a given threshold. The second approach employs a voting system in which a tomogram is classified as misaligned if most subtomograms are classified as such, and vice versa.

The suggested threshold for the first network (strong misalignment) is 0.5. A suitable threshold for the second network is more difficult to find. 0.5 is possible but may be too strict or too loose depending on the dataset. For this reason, we let the user have the possibility of providing a different threshold. Also, a specific threshold may not be specified, leading to the evaluation of the tomogram based on the average score derived from the subtomograms it contains. Sorting by this average score, the user may decide later which tomograms to keep. This classification workflow is depicted in Fig. 3.

2.3. Location of fiducial markers

In case the location of the fiducial markers is not available, we propose a tool for detecting these markers in the tomographic reconstruction. Properly localizing fiducial markers in the tomogram is a key step in detecting alignment errors. A miscalculation of their position in the sample leads to an inaccurate assessment of the alignment quality of the extracted subtomograms, since the network relies on the presence or not of misalignment artifacts related to high Signal-to-Noise-Ratio (SNR) features (the fiducials).

This algorithm consists of several steps presented in the next paragraphs. One of our concerns with this algorithm was its execution time and memory consumption, as it has to deal with tomograms, which normally require lots of memory. For this reason, many steps are performed on 2D slices of the tomogram, which serve our detection purpose and, at the same time, largely increase the algorithm efficiency.

1. Tomogram preprocessing. We perform a moving average of 5 slices across the tomogram's Z direction to increase the SNR and reinforce the markers consistently along several slices, as fiducial markers are. Then, within each slice, a bandpass filter with raised cosine falloff centered at the frequency corresponding to the diameter of the gold beads (where the associated frequency is $1/D$, being D the diameter of the gold bead) is applied. Finally, we apply a 2D Laplacian filter.
2. 2D fiducial markers location. We start by identifying dark outliers in the preprocessed 2D slices. Outliers are those pixels whose intensity z-score is below a given value (in our experiments, 3). The intensity z-score is defined as the difference between the pixel intensity value and the slice mean, divided by the standard deviation of the slice. The outlier detection results in a binary slice with value 1 where the outlier dark pixels are located. We compute the connected components of this binary image and remove those objects whose size is smaller than a given threshold (normally computed as a fraction of the fiducial marker diameter) and whose circularity is below a given threshold (in our examples, 0.75). We calculate circularity as the ratio between the object area and the area of the minimum circle that contains the object.
3. 3D fiducial markers location. In this step, we combine the 2D objects detected in the previous step into separated 3D objects. To do so, each 2D object in each slice votes for the overlapping 2D objects in adjacent slices. Non-voted objects are removed. The voting process is iteratively repeated until no object is removed. The center of mass of the remaining objects are candidates to be the center of a fiducial marker.

   These centers are further refined by looking at the intensities in the tomogram. Let us refer as $V_i(\mathbf{r})$ to the subtomogram of size $2D \times 2D \times 2D$ extracted around the i-th center of mass. We look for the coordinate displacement $\Delta\mathbf{r}_i$ that maximizes the correlation between

Table 2
Smallest misalignment detected in at least 90% of the subtomograms.

Artifact	Confidence	Misalignment
Bananas	98.81%	79 Å
Twisters	97.6%	2 degrees
Birds	98.77%	[-27, 27] Å

$V_i(\mathbf{r})$ and $V_i(-\mathbf{r})$. This displaced coordinate is the center of the most point-symmetric object in that area. This coordinate should correspond to the center of the fiducial marker.

4. Quality filters. We may apply any of the quality filters described below:

- Removal of duplicates. Those coordinates whose distance is smaller than a given threshold (in our experiment, $1.5D$) are considered to belong to the same object and are substituted with their average.
- Removal of non-symmetric objects. We may remove the coordinates of objects whose correlation between $V_i(\mathbf{r})$ and $V_i(-\mathbf{r})$ is smaller than a given threshold (in our experiments, 0.1).
- Removal of outlier objects. For each object, we calculate the radial profile of $V_i(\mathbf{r})$, compute the average radial profile of all objects, and remove those coordinates whose radial profile has a Mahalanobis distance to the average profile larger than a given threshold (in our experiments, 2).

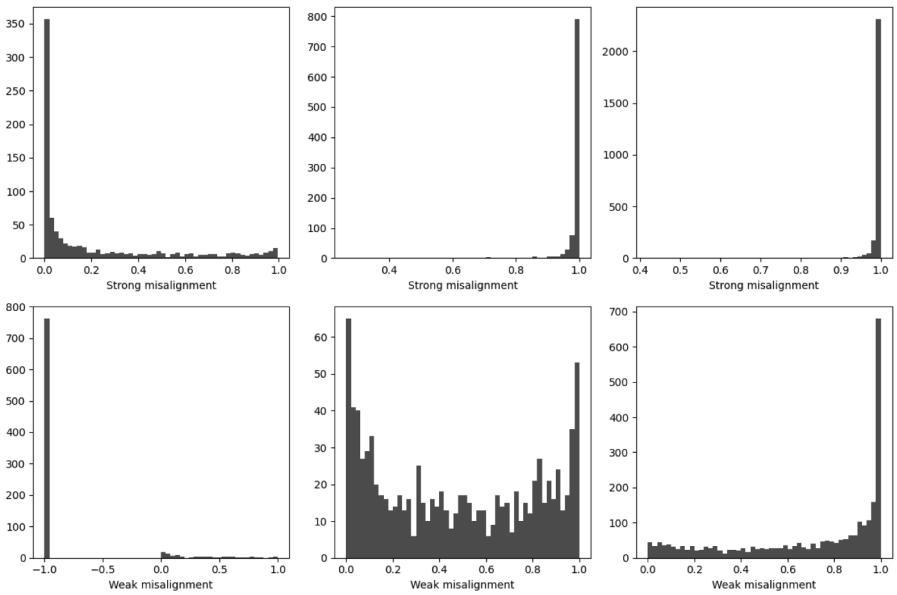


Fig. 4. Histogram distributions of the misalignment scores retrieved by both networks. The alignment algorithm provides the coordinates of the studied markers. The first row corresponds to the scores obtained by the first network (strong misalignment) and the second row to the second network (weak misalignment). Left, central, and right columns correspond to strongly misaligned, weakly misaligned, and aligned datasets respectively.

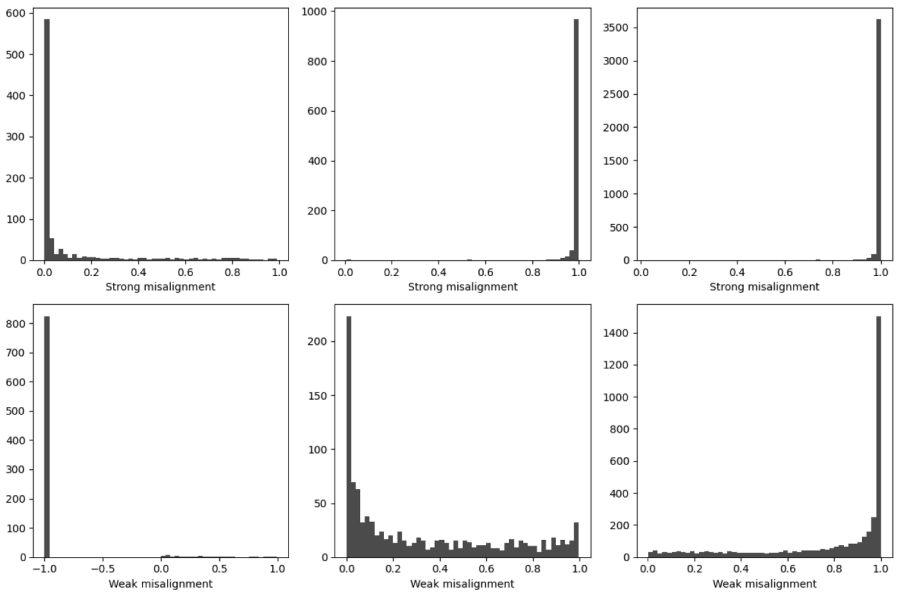


Fig. 5. Histogram distributions of the misalignment scores retrieved by both networks. The fiducial marker location algorithm provides the coordinates of the studied markers. The first row corresponds to the scores obtained by the first network (strong misalignment) and the last one to the second network (weak misalignment). Left, central, and right columns correspond to strongly misaligned, weakly misaligned, and aligned datasets respectively. The fiducial marker location algorithm provides the coordinates of the studied markers.

A relaxed mode has been implemented so that, in case none of the coordinates survive these quality filters, none of them is removed from the final output. This avoids an empty output when the quality filters are too strict or the data quality too poor.

Generally, there is a trade-off between the quantity and quality of detected objects. On the one hand, having a larger number of objects provides a stronger base for decision-making in the tomogram, as it increases the amount of evidence available. On the other hand, it is crucial to maintain a low false positive rate in object detection. This ensures that the study of the quality of the alignment of each of these objects can be reliably made and is not performed over spurious objects in the tomogram.

3. Results

This section presents the results of the different algorithms described in this work.

3.1. Algorithm calibration

In this section, we estimate the misalignment detection limits of our algorithm. To do so, we simulate bananas, twisters, and birds of different strengths. We start from the set of correctly aligned tilt series, see Section 2.1, and add small offsets to the different alignment parameters. We used the Scipion protocol `tomo - misalign tilt series` for this purpose. We keep reducing the offset until we find the minimum misalignment parameter for which the first neural network can still detect at least 90% of the tomograms as misaligned. The network is fed only with the fiducial detected by the marker location algorithm on the misaligned tomograms. Table 2 shows the results including confidence for the minimum offset, and Fig. 2 is an example of the gold beads under these minimal deformations. Note that smaller misalignments may still be detected, although not 90% of the time.

3.2. Experimental data: subtomogram scoring

We now evaluate the two networks on the subtomograms extracted from the three types of tomograms: strongly misaligned, weakly misaligned, and correctly aligned. The score histograms are presented in Fig. 4 and Fig. 5, in case the location of the subtomograms is provided by the alignment algorithm or automatically detected. The score ranges from 0 to 1, where low score indicates a subtomogram from a misaligned tomogram, while a high score indicates a subtomogram from a correctly aligned one.

The first network, the one detecting strong misalignment, is rather binary and its behavior is equivalent independently of the source of the data (fiducials from the alignment or the fiducial marker location algorithm). For this network, the subtomograms clearly belong to one of the two categories (strongly misaligned or correctly aligned, even if there is some weak misalignment). As expected, almost all the strongly misaligned subtomograms are detected by the first network and not input to the second. The output of the second network is much more continuous, with many scores between 0 and 1, revealing that this second classification task is much more difficult. Observing the score histograms, a clear bias is evident in the aligned set of subtomograms, indicating that the network is able to detect features that are only present in this group. This bias is still clear in the weakly misaligned group, although it is less prominent when the source of the coordinates is the alignment algorithm rather than the marker location algorithm. The score distributions are qualitatively different so that the network is able to differentiate between both groups, with an F1 score of 0.77 when subtomogram coordinates are provided by the alignment and 0.85 if provided by the marker location algorithm, both cases using a threshold of 0.5.

Table 3
Confusion matrix comparing the ground truth from the tilt series and the classification obtained by the deep neural classification from the fiducial marker coordinates obtained from the tilt series alignment.

From alignment		Predicted		
	Aligned	Weak misalignment	Strong misalignment	
Real	Aligned	161	21	0
	Weak misalignment	26	33	0
	Strong misalignment	1	4	56

Table 4
Confusion matrix comparing the ground truth from the tilt series and the classification obtained by the deep neural classification from the automatically detected fiducial markers.

From PHC		Predicted		
	Aligned	Weak misalignment	Strong misalignment	
Real	Aligned	166	16	0
	Weak misalignment	20	39	0
	Strong misalignment	1	5	52

3.3. Experimental data: tomogram classification

We finally evaluate the tomogram classification workflow in Fig. 3. In our experiment, the classification based on the average score is more robust than using the voting system (results not shown). Still, the voting system is available if the user is interested in possible comparisons. A threshold for both strong and weak misalignment networks of 0.5 has been used, meaning that we do not bias the request of evidence to resolve the alignment quality.

Tables 3 and 4 show the confusion matrices for the experimental dataset, depending on the source of coordinates (provided by the tilt series alignment or automatically detected by the fiducial markers location algorithm). The algorithm is very robust in detecting strongly misaligned tomograms in the dataset, classifying most of them as strongly or weakly misaligned. The percentage of misclassified tomograms presenting strong misalignment is 1.63% if coordinates come from the tilt series alignment and 1.72% if they come from the automatically detected markers. The percentage of tomograms properly classified as aligned is 88.46% if coordinates come from the tilt series alignment and 91.21% if they come from the automatically detected markers, and also, as a positive note, none of them have been classified as strongly misaligned.

These results also show that the performance of the weak misalignment is not as reliable as that of the strong misalignment classification. This result was expected because some weakly misaligned tilt series can be recognized at the tilt series level but not at the tomographic reconstruction.

3.4. Location of fiducial markers

The behavior of the fiducial maker location algorithm is presented in this section. Fig. 6 shows the algorithm's results in locating high-contrast objects in the tomogram. It shows the outcomes for three types of tomograms: aligned, weakly misaligned, and strongly misaligned tomograms.

To study the performance of this algorithm, 10 tomograms from each group have been manually inspected. In the case of the strongly misaligned tomograms, it has been ensured that the relaxed mode has not been executed to prevent any contamination of the dataset.

As expected, the performance of the location of fiducial markers depends on the quality of the tilt series alignment: the better the alignment, the better the location of these coordinates will be, especially considering the presence of strong misalignment. The number of fiducial markers located by the algorithm compared to the ones used to solve the

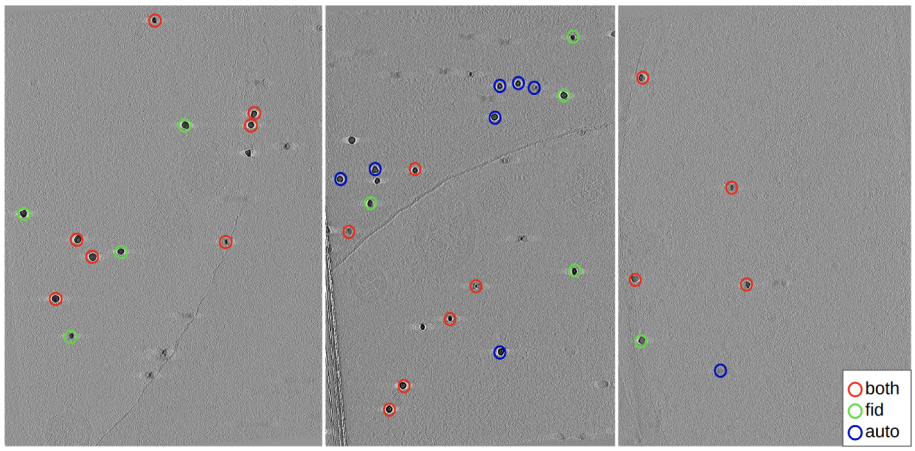


Fig. 6. Comparison of the high-contrast objects automatically found in three different tomograms: aligned (left), weakly misaligned (center), and strongly misaligned (right). The color code is: green (found by tilt series alignment but not automatically), blue (found as a high-contrast feature but not by the tilt series alignment), and red (found by both).

alignment is 41.25% for the strongly misaligned subset, 81.07% for the weakly misaligned, and 68.18% for the aligned one.

This behavior is not necessarily detrimental to misalignment detection. In fact, if the tilt series is misaligned, it will cause the detection of poorly resolved markers that can be easily detected by the neural network as indicators of misalignment. The number of markers for a misaligned tomogram is normally low. There are two main reasons for this: firstly, they go undetected due to a lower SNR; and secondly, even if some markers are detected, most of them are removed due to a lack of correlation with their mirror.

The performance of the location of fiducial markers is acceptable, especially in the case of not strongly misaligned tomograms. The number of detected gold beads is sufficiently high, with a low false positive detection rate of 6.06% for the strongly misaligned subset, 0% for the weakly misaligned, and 0.83% for the aligned one. We have also observed that in the case of misalignment, the automatically detected fiducial markers provide more valuable information to estimate the misalignment than the 3D coordinates obtained by the tilt series alignment algorithm. The reason is that, as a result of misalignment, the 3D coordinates from the tilt series alignment tend to be inaccurately placed in the tomogram.

In the case of strong misalignment, most objects found do not pass through the quality filters. In this case, it might be advantageous to run the algorithm in a relaxed mode in order not to remove them by any filter and use them as the input to the neural networks.

4. Discussion and conclusions

This paper introduces a new tool for automatically detecting misalignment errors in cryo-electron tomographic reconstructions. Just as it has been happening in single particle analysis in cryo-electron microscopy, the use of automatic tools in tomography is gaining popularity among users. This is particularly prevalent in the initial stages of the workflow, especially when considering the steps previous to the tomographic reconstruction. There are two main reasons for this trend: first, automatic tools for pre-processing, alignment, CTF estimation, and reconstruction have significantly improved in terms of their robustness, yielding higher-quality results with no need for scientist interaction; second, high-throughput acquisition systems and streaming processing are also becoming a reality in tomography. Thus, the availability of automatic tools for detecting misalignment errors is a key step in developing robust pipelines for data processing in tomography, especially considering that the reconstructed tomogram is the starting point for many subsequent analysis steps.

Analyzing the strong misalignment detection, it can be observed that

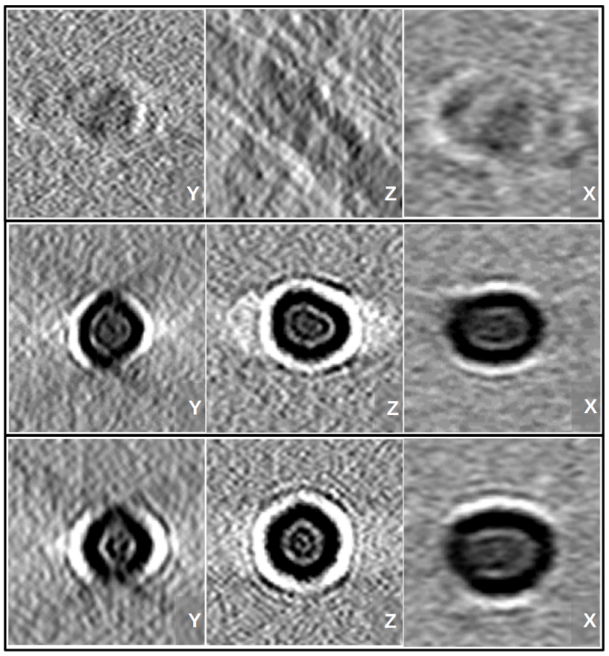


Fig. 7. Slices of fiducial markers that exemplify the three different tomogram sets. The left column shows the XZ-plane, the center column the XY-plane, and the right column the YZ-plane. The first row presents a marker from the strongly misaligned set; the second row presents one from the weakly misaligned set, and the last row from the aligned set.

the algorithm performs slightly better when the source of the coordinates is the alignment algorithm. This behavior is expected since these coordinates are the ones the tilt series alignment algorithm relies on when solving the alignment of the acquisition geometry. Additionally, in cases where the alignment errors are so severe that no structure is recognized in the reconstruction, the marker location algorithm cannot identify any structure with sufficient contrast. As a result, tomograms are not classified due to the lack of coordinates.

On the other hand, in weak misalignment detection, the opposite behavior is observed, with a slight overperformance of the algorithm when the source of the coordinates is the marker location algorithm itself. The authors believe that this behavior can be explained because the alignment algorithm imposes the limitation that markers used for alignment must be present in all tilt images (or visible at zero degrees of

Table 5
Confusion matrices comparing the tomogram classification of two scientists and the deep-neural network algorithm.

		Scientist 2	
		Aligned	Misaligned
Scientist 1	Aligned	167	5
	Misaligned	6	59
	DNN		
		Aligned + Weakly Mis.	Strong Mis.
Scientist 1	Aligned	170	2
	Misaligned	11	54
	DNN		
		Aligned + Weakly Mis.	Strong Mis.
Scientist 2	Aligned	172	1
	Misaligned	9	55

tilting). Consequently, fiducials located by the marker location algorithm might be reconstructed from regions of the tomogram where a lower number of tilt images contribute to the reconstruction, illuminating weak misalignment errors that appear only in a few images. Also, fiducials that are located in regions of the tomogram far from the tilt axis typically present artifacted reconstructions, which might be another reason for this behavior.

Nevertheless, as mentioned in the methods section, not all alignment algorithms use fiducial markers to calculate the alignment or report their three-dimensional coordinates. For this reason, it is necessary to provide automatic tools to work with this other kind of tomograms.

The most challenging problem encountered has been the detection of weak misalignment errors. During the reconstruction process, it is important to note that the combination of projective information onto the third-dimensional space does not ensure that small displacements of some tilt series lead to detectable artifacts in the final tomographic reconstruction. Thus, it is not expected that misalignments detected over the aligned tilt series can also be detected in the tomographic reconstruction, especially the more subtle ones. We illustrate this problem in Fig. 7, showing three gold beads from the three different sets, where it is possible to visualize the subtle difference between aligned and weakly misaligned fiducial markers. The difference between the weakly misaligned and correctly aligned fiducials is small, and the neural network finds it hard to distinguish them.

To check if two human experts can distinguish them, two different scientists have driven a second classification round, but only inspecting the tomogram in this case. Each scientist has generated two sets, splitting the original set of tomograms into aligned and misaligned. When comparing these classifications, it can be observed from the confusion matrix that they present a consistency of 95.36%. When comparing the classification of both scientists with the deep-neural network algorithm, it is observed that for both of them, the accuracy of their estimation is respectively 94.51% and 95.78%. These results are obtained considering as aligned the union of the aligned and weakly misaligned sets from the deep-neural network. With this, the authors exemplified how, even from trained scientists, detecting subtle movements in the tilt series might be unrecognized in the tomographic reconstruction. The results from this comparison are summarized in Table 5.

Another core feature of this work consists of its automatic behavior. With the growing popularity of automatic image processing methods and high-throughput acquisition systems in the field, the automatic detection of artifactual reconstructions is a critical step in the cryo-ET processing pipeline. The developed work ensures this detection, without the need for the manual intervention of the user.

It is interesting to comment, as a potential application, on the opportunity that this software brings to study local alignment errors. Since the alignment score is reported for each gold bead individually, it is possible to compare the alignment quality between different regions in the tomographic reconstruction. Although it requires a certain consistency in local alignment errors along the series in the same region of the sample and a homogeneous and numerous enough distribution of gold beads, the authors find it an interesting research path for future work.

In this work, we have presented one automatic algorithm for the localization of fiducial markers in a tomographic reconstruction and another one for the assessment of the misalignment of the reconstructed tomogram. These artifactual reconstructions should be identified and discarded to ensure reliable further processing. Alternatively, the alignment quality might be improved by realigning automatically setting different parameters, using a different algorithm, or even by manual adjustment.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

data and sofware are both public and avilable via GitHub and EMPIAR.

Acknowledgements

The authors acknowledge the economic support from MICIN to the Instruct Image Processing Center (I2PC) as part of the Spanish participation in Instruct-ERIC, the European Strategic Infrastructure Project (ESFRI) in the area of Structural Biology, Grant PID2019-104757RB-I00 funded by MCIN/AEI/10.13039/501100011033/ and “ERDF A way of making Europe” by the European Union, and to the European Union (EU) and Horizon 2020 through the grant: HighResCells (ERC-2018-SyG, Proposal: 810057).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jsb.2023.108056.

References

Andersen, A., Kak, A., 1984. Simultaneous algebraic reconstruction technique (sart): A superior implementation of the art algorithm. Ultrason. Imaging 6 (1), 81–94.
Burt, A., Cassidy, C.K., Ames, P., Bacia-Verloop, M., Baulard, M., Huard, K., Luthey-Schulten, Z., Desfosses, A., Stansfeld, P.J., Margolin, W., et al., 2020. Complete structure of the chemosensory array core signalling unit in an e. coli minicell strain. Nat. Commun. 11 (1), 743.
Castano-Diez, D., Scheffer, M., Al-Amoudi, A., Frangakis, A.S., 2010. Alignator: A gpu powered software package for robust fiducial-less alignment of cryo tilt-series. J. Struct. Biol. 170 (1), 117–126.
de la Rosa-Trevin, J., Quintana, A., del Cano, L., Zaldivar-Peraza, A., Foche, I., Gutierrez, J., Gomez-Blanco, J., Burguet-Castells, J., Cuenca, J., Abrishami, V., Vargas, J., Oton, J., Sharov, G., Navas, J., Conesa, P., Vilas, J., Marabini, R., Sorzano, C., Carazo, J., 2016. Scipion: a software framework toward integration, reproducibility, and validation in 3d electron microscopy. J. Struct. Biol. 195, 93–99.
de la Rosa-Trevín, J.M., Otón, J., Marabini, R., Zaldívar, A., Vargas, J., Carazo, J.M., Sorzano, C.O.S., 2013. Xmipp 3.0: an improved software suite for image processing in electron microscopy. J. Struct. Biol. 184, 321–328. Nov.
Fernandez, J.-J., Li, S., 2021. Tomoalign: A novel approach to correcting sample motion and 3d ctf in cryoet. J. Struct. Biol. 213 (4), 107778.
Fernandez, J.-J., Li, S., Bharat, T.A., Agard, D.A., 2018. Cryo-tomography tilt-series alignment with consideration of the beam-induced sample motion. J. Struct. Biol. 202 (3), 200–209.
Gilbert, P., 1972. Iterative methods for the three-dimensional reconstruction of an object from projections. J. Theor. Biol. 36 (1), 105–117.
Gordon, R., Bender, R., Herman, G., 1970. Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography. J. Theor. Biol. 29 (3), 471–481.
Jimenez de la Morena, J., Conesa, P., Fonseca, Y., de Isidro-Gomez, F., Herreros, D., Fernandez-Gimenez, E., Strelak, D., Moebel, E., Buchholz, T., Jug, F., Martinez-Sanchez, A., Harastani, M., Jonic, S., Conesa, J., Cuervo, A., Losana, P., Sanchez, I., Iceta, M., del Cano, L., Gragera, M., Melero, R., Sharov, G., Castaño-Diez, D., Koster, A., Piccirillo, J., Vilas, J., Oton, J., Marabini, R., Sorzano, C., Carazo, J., 2022. Scipiontomo: Towards cryo-electron tomography software integration, reproducibility, and validation. J. Struct. Biol. 214 (3), 107872.

Khoshouei, M., Pfeffer, S., Baumeister, W., Förster, F., Danev, R., 2017. Subtomogram analysis using the volta phase plate. J. Struct. Biol. 197 (2), 94–101.

D.N. Mastronarde and S.R. Held, "Automated tilt series alignment and tomographic reconstruction in imod," Journal of Structural Biology, vol. 197, no. 2, pp. 102–113, 2017. Electron Tomography.

Noble, A.J., Stagg, S.M., 2015. Automated batch fiducial-less tilt-series alignment in appion using protomo. J. Struct. Biol. 192 (2), 270–278.

M. Radermacher, Weighted Back-Projection Methods, pp. 91–115. Boston, MA: Springer, US, 1992.

Robertson, M.J., Meyerowitz, J.G., Skiniotis, G., 2020. Cryo-em as a powerful tool for drug discovery. Bioorg. Med. Chem. Lett. 30 (22), 127524.

Schur, F.K., Obr, M., Hagen, W.J., Wan, W., Jakobi, A.J., Kirkpatrick, J.M., Sachse, C., Kräusslich, H.-G., Briggs, J.A., 2016. An atomic model of hiv-1 capsid-sp1 reveals structures regulating assembly and maturation. Science 353 (6298), 506–508.

Seifer, S., Elbaum, M., 2022. Clusteralign: A fiducial tracking and tilt series alignment tool for thick sample tomography. Biological Imaging 2, e7.

Sorzano, C.O.S., Messaoudi, C., Eibauer, M., Bilbao-Castro, J.R., Hegerl, R., Nickell, S., Marco, S., Carazo, J.M., 2009. Marker-free image registration of electron tomography tilt-series. BMC Bioinformatics 10, 124.

C. Sorzano, J. Vargas, J. Oton, J. de la Rosa-Trevin, J. Vilas, M. Kazemi, R. Melero, L. del cano, J. Cuenca, J. Gomez-Blanco, and J. Carazo, "A survey of the use of iterative reconstruction algorithms in electron microscopy," BioMed Research International, vol. 2017, 2017.

Sorzano, C., de Isidro-Gómez, F., Fernández-Giménez, E., Herreros, D., Marco, S., Carazo, J.M., Messaoudi, C., 2020. Improvements on marker-free images alignment for electron tomography. J. Struct. Biol.: X 4, 100037.

D. Strelak, A. Jiménez-Moreno, J.L. Vilas, E. Ramírez-Aportela, R. Sánchez-García, D. Maluenda, J. Vargas, D. Herreros, E. Fernández-Giménez, F.P. de Isidro-Gómez, J. Horacek, D. Myska, M. Horacek, P. Conesa, Y.C. Fonseca-Reyna, J. Jiménez, M. Martínez, M. Harastani, S. Jonić, J. Filipovic, R. Marabini, J.M. Carazo, and C.O.S. Sorzano, "Advances in xmipp for cryo–electron microscopy: From xmipp to scipion," Molecules, vol. 26, no. 20, 2021.

Turoňová, B., Sikora, M., Schürmann, C., Hagen, W.J., Welsch, S., Blanc, F.E., von Bülow, S., Gecht, M., Bagola, K., Hörner, C., et al., 2020. In situ structural analysis of sars-cov-2 spike reveals flexibility mediated by three hinges. Science 370 (6513), 203–208.

Van Drie, J.H., Tong, L., 2022. Drug discovery in the era of cryo-electron microscopy. Trends Biochem. Sci. 47 (2), 124–135.

Zheng, S., Wolff, G., Greenan, G., Chen, Z., Faas, F.G., Bárcena, M., Koster, A.J., Cheng, Y., Agard, D.A., 2022. Aretomo: An integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction. J. Struct. Biol.: X 6, 100068.

A deep learning approach to the automatic detection of alignment errors in cryo-electron tomographic reconstructions

Supplementary material

F.P. de Isidro-Gómez[1,2], J. L. Vilas[1], P. Losana[1],J.M. Carazo[1*], C.O.S. Sorzano[1*]

December 22, 2023

[1] Biocomputing Unit, Centro Nacional de Biotecnologia (CNB-CSIC), Darwin, 3, Campus Universidad Autonoma, 28049 Cantoblanco, Madrid, Spain
[2] Univ. Autonoma de Madrid, 28049 Cantoblanco, Madrid, Spain
* Corresponding author

User guide

In this section, we present a simplified user guide to use the software presented in this work. This guide includes both available sources available to the user: Xmipp standalone command-line mode, and the implementation of this sofware in the Scipion framework.

0.1 Xmipp standalone

To work with the presented tools in the standalone version it is only necessary to have installed the Xmipp software package. Through the command line, the user has access to the full functionality of this software.

To execute the high contrast peaker through the command line it is enough to execute the `xmipp_image_peak_high_contrast_program` followed by the options summed up in 1. This is an example command for the execution of this program:

```
xmipp_image_peak_high_contrast_program --vol tomo1.mrc -o outputCoords.xmd
--boxSize 32 --fiducialSize 80 --sdThr 2 --mirrorCorrelationThr 0.2
--mahalanobisDistanceThr 2 --numberSampSlices 400 --numberOfCoordinatesThr
10 --samplingRate 18.92 --relaxedModeThr 3
```

To execute the misalignment detection program through the command line it is enough to execute the `xmipp_deep_misalignment_detection` followed by the options summed up in 2. This is an example command for the execution of this program:

```
xmipp_deep_misalignment_detection --modelPick 0 --subtomoFilePath
subtomoCoords.xmd --misaliThr 0.45
```

Parameter	Description	Default
vol	File path to input volume	
o	File path to output coordinates file	
samplingRate	Sampling rate of the input tomogram (A/px)	
fiducialSize	Size of the fiducial markers in Angstroms (A)	100
boxSize	Box size of the peaked fiducials	32
numberSampSlices	Number of sampling slices to calculate the threshold value	10
sdThr	Number of STD away the mean to consider that a pixel has an outlier value	5
numberOfCoordinatesThr	Minimum number of points attracted to a coordinate	10
mirrorCorrelationThr	Minimum correlation of a coordinate with its mirror	0.1
mahalanobisDistanceThr	Minimum Mahalanobis distance	2
relaxedModeThr	Minimum number of peaked coordinates to disable a filter	3

Table 1: Table of parameters from the peak high contrast algorithm

0.2 Scipion framework

To work with the presented tools inside the scipion framework it is necessary to have installed both the Xmipp and the Scipion software packages. This is procedure is simplified since the installation of Xmipp is triggered when installing Scipion. Since this is a simplified tutorial, the input information for the presented workflow is the already reconstructed tomograms. Nonetheless, more extended documentation and tutorials about tomography data processing can be found at the Scipion documentation landing page:

https://scipion-em.github.io/docs/release-3.0.0/docs/user/tomography-tutorials.html

In particular the "Tomogram Reconstruction" explains to the user how to obtain a set of reconstructed tomograms from a set of raw tilt-series. The first step is to import the tomograms into a Scipion project (using the `tomo - import tomograms`). In figure 3 it is shown an example of this protocol from for its execution. The output of this protocol consists of a set of tomograms with which it is possible to start the processing.

Once the tomograms are correctly imported into the Scipion project, it is possible to run the peak high contrast algorithm through the Scipion protocol `xmipptomo - peak high contrast`. In figure 3 it is shown an example of this protocol from for its execution. The output of this protocol is a set of three-dimensional coordinates assigned to each tomogram from the set. The number of peaked coordinates is reported in the summary section and it can be visualized with the available viewers offered by Scipion.

Parameter	Description
modelPick	choose model for weak misalignment estimation. Strict model (0) is picked in order to avoid false positives. In case loose (1) model is chosen, less good aligned tomograms are lost. As a tradeoff, the number of false positives will increase.
subtomoFilePath	file path of the xmd file containing the coordinates of the extracted subtomos. This is the output got when extracting with xmipp_tomo_extract_subtomograms.
misaliThr	Threshold to settle if a tomogram presents weak or strong misalignment. If this value is not provided two output set of tomograms are generated, those discarded which present strong misalignment and those which do not. If this value is provided the second group of tomograms is split into two, using this threshold to settle if the tomograms present or not a weak misalignment.
misalignmentCriteriaVotes	Define criteria used for making a decision on the presence of misalignment on the tomogram based on the individual scores of each subtomogram. If this option is not provided (default) the mean of this score is calculated. If provided a voting system based on if each subtomo score is closer to 0 o 1 is implented

Table 2: Table of parameters from the deep misalignment detection algorithm

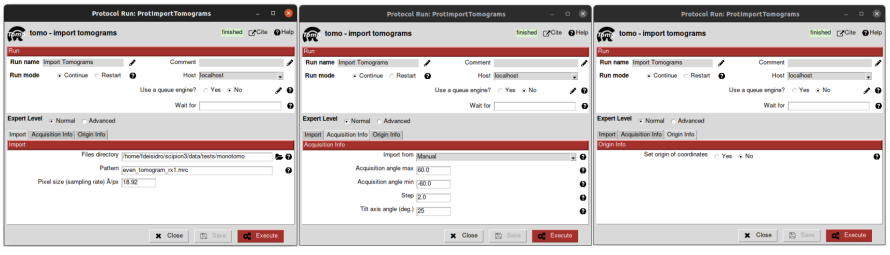


Figure 1: Import tomograms protocol form, including the views of its three slices.

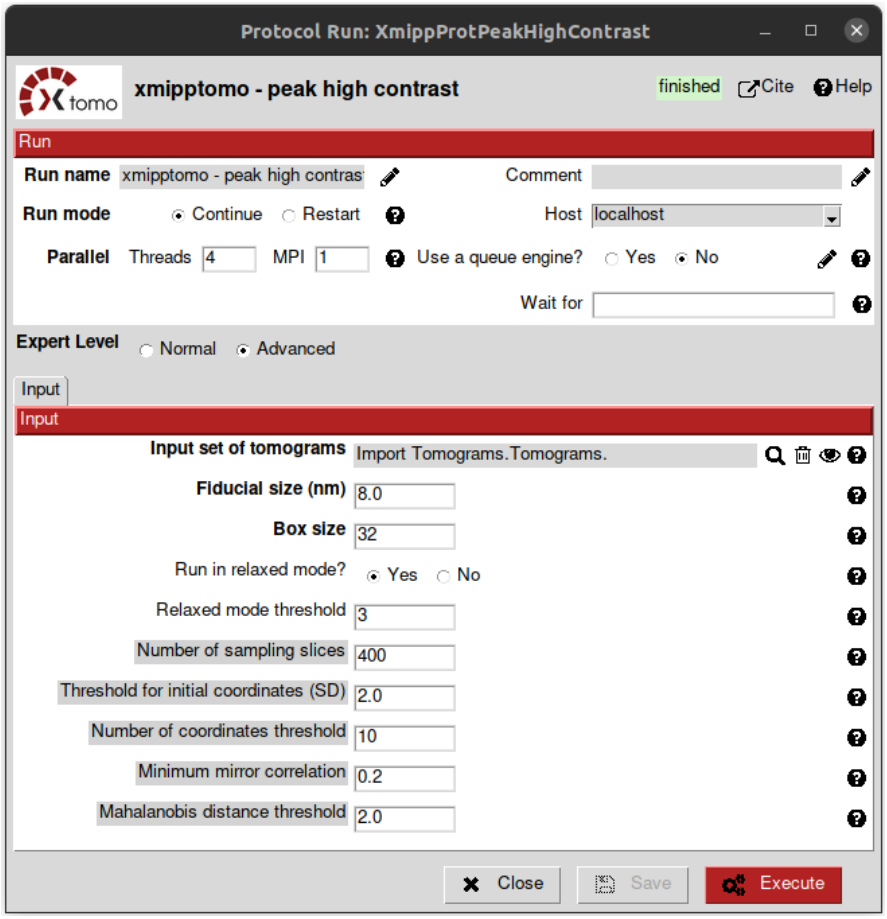


Figure 2: Peak high contrast protocol form.

Finally, once the fiducial markers are located in the tomogram, the degree of misalignment presented by the tomograms can be studied. The misalignment detection protocol is run through the Scipion protocol `xmipptomo - detect misalignment form fiducials`. In figure 3 it is shown an example of this protocol from for its execution. The distribution of strong and weak misaligned, and aligned tomograms is reported in the summary section along with the whole set of fiducial markers subtomograms used for the tomogram classification. This set of subtomograms includes the alignment scores provided by the deep neural networks from the algorithm. Figure 4 it is shown a capture of the summary reported by the detect misalignment protocol, including all kinds of possible outputs.

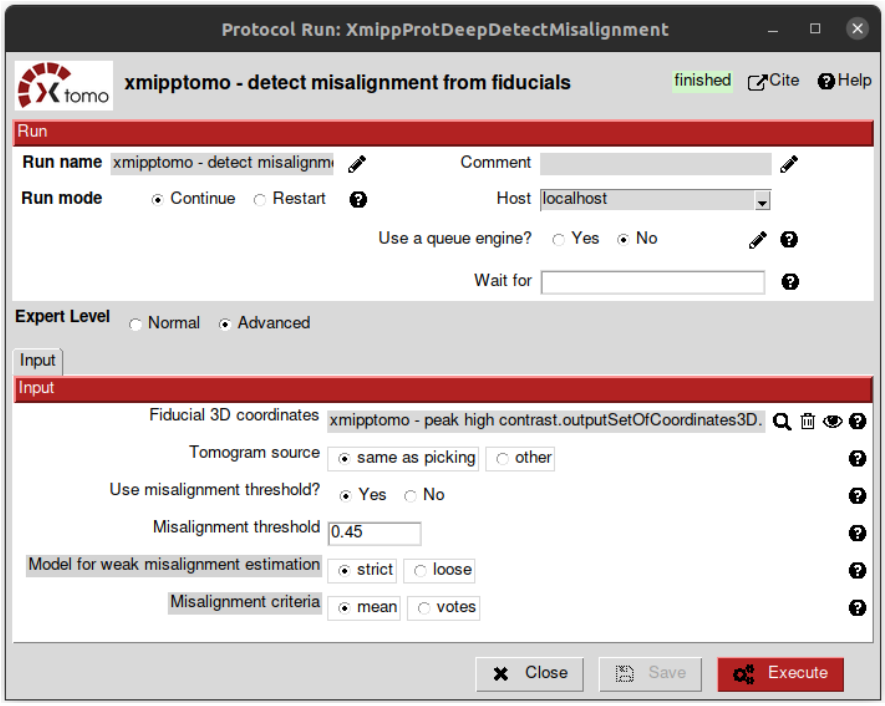


Figure 3: Detect misalignment protocol form.

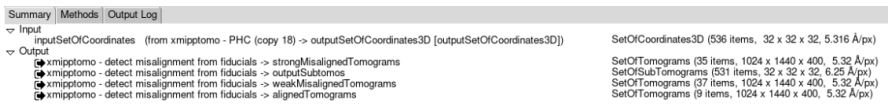


Figure 4: Detect misalignment output summary.

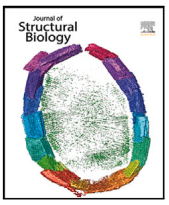
Appendix C: Automatic detection of alignment errors in cryo-electron tomography

Journal of Structural Biology 217 (2025) 108153

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi



Research article

Automatic detection of alignment errors in cryo-electron tomography

F.P. de Isidro-Gómez [a,b], J.L. Vilas [a], J.M. Carazo [a], C.O.S. Sorzano [a,*]

[a] Biocomputing Unit, Centro Nacional de Biotecnologia (CNB-CSIC), Darwin, 3, Campus Universidad Autonoma, 28049 Cantoblanco, Madrid, Spain
[b] University Autonoma de Madrid, 28049 Cantoblanco, Madrid, Spain

ARTICLE INFO

Edited by Steven J. Ludtke

Keywords:
Cryo-electron tomography
Tilt-series alignment
Electron tomography
Electron microscopy
Image processing
Structural biology

ABSTRACT

Cryo-electron tomography is an imaging technique that allows the study of the three-dimensional structure of a wide range of biological samples, from entire cellular environments to purified specimens. This technique collects a series of images from different views of the specimen by tilting the sample stage in the microscope. Subsequently, this information is combined into a three-dimensional reconstruction. To obtain reliable representations of the specimen of study, it is mandatory to define the acquisition geometry accurately. This is achieved by aligning all tilt images to a standard reference scheme. Errors in this step introduce artifacts into the final reconstructed tomograms, leading to loss of resolution and making them unsuitable for detailed sample analysis. This publication presents algorithms for automatically assessing the alignment quality of the tilt series and their classification based on the residual errors provided by the alignment algorithms. If no alignment information is available, a set of algorithms for calculating the residual vectors focused on fiducial markers is also presented. This software is accessible as part of the Xmipp software package and the Scipion framework.

1. Introduction

The three-dimensional (3D) structural analysis of biological specimens is a significant milestone in modern biology. The integration of structural and functional information provides scientists with the necessary tools to understand the underlying biological organization of the sample. Cryogenic electron tomography (cryo-ET) is a sophisticated imaging technique extensively utilized in exploring biological complexes. It allows the study of the structure of macromolecules without losing the information of their biological context. This technique enables the study of various biological samples, from cellular environments to purified complexes.

As in classical cryo-electron microscopy (cryoEM), in its variant of single-particle analysis (SPA), the biological specimen in its native hydrated state is rapidly frozen and then imaged via a Transmission Electron Microscope (TEM) under cryogenic conditions, with the main difference that the sample is tilted in the image acquisition process. Thus, cryo-ET has become an essential tool for studying the structure and function of macromolecular complexes *in situ*, such as cell machinery and viruses. Tomography has allowed a higher level of detail of biological processes at the molecular level, with applications in multiple disciplines, such as structural biology, virology, cell biology, and drug discovery (Robertson et al., 2020; Van Drie and Tong, 2022).

In detail, cryo-ET allows the determination of the three-dimensional structure of the sample by combining projective information from different views. Thus, by rotating the sample holder inside the TEM, it is possible to acquire a set of two-dimensional projective images. Then, the projective information in this set of images is subsequently combined into the three-dimensional space, elucidating the internal structure of the sample under study. However, correcting the sample's relative movements and estimating the acquisition's geometry are mandatory before this reconstruction. This is the aim of the tilt series alignment algorithms. Any error introduced in this step will compromise the final reconstruction, resulting in the loss of detailed structural information. This work presents a set of algorithms to evaluate the alignment quality.

In some sample preparation protocols, gold bead particles are included as high-contrast markers acting as reference points in the tilt series acquisition. Some alignment algorithms exploit these points to make alignment calculations easier and more robust by tracking their position along the series and solving the geometry of the acquisition. These are known as fiducial-based algorithms (Sorzano et al., 2009; Castano-Diez et al., 2010; Mastronarde and Held, 2017; Fernandez et al., 2018; Fernandez and Li, 2021; Seifer and Elbaum, 2022; Sorzano et al., 2020; Coray et al., 2024; Xu et al., 2024). A different set of methods is employed in samples where these particles are absent, known as fiducial-less algorithms (Zheng et al., 2022; Galaz-Montoya et al., 2015). This family of algorithms is based on an iterative process

* Corresponding author.
E-mail address: coss.eps@ceu.es (C.O.S. Sorzano).

https://doi.org/10.1016/j.jsb.2024.108153
Received 25 July 2024; Received in revised form 25 November 2024; Accepted 2 December 2024
Available online 16 December 2024

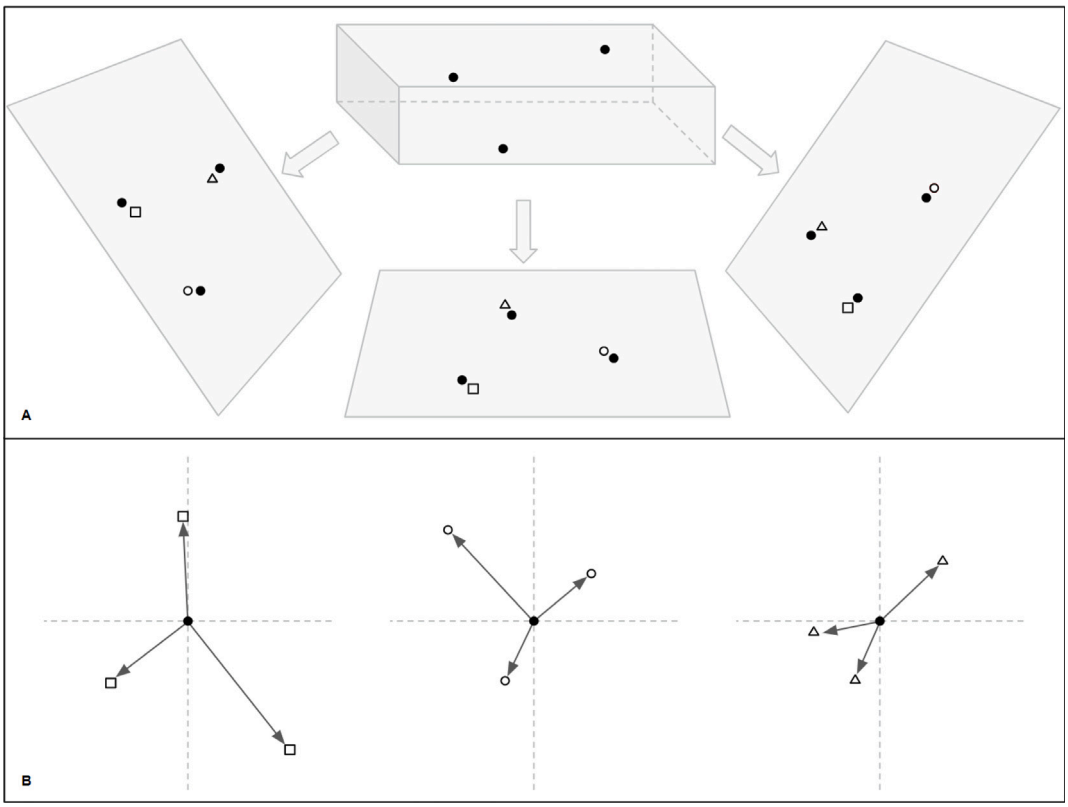


Fig. 1. Schematic definition of residual vectors. (A) represents the projection of the three-dimensional coordinate onto every tilt image (black dots) and the detected fiducials in each tilt image (circles, squares, and triangles). Thus, the residual is defined as the vector whose origin is the projection of the fiducial coordinate and the endpoint in the detected landmark. (B) represents the calculated set of residuals for each fiducial in the sample. A hypothetical case with only three fiducials and three tilt images is presented for simplicity.

of alignment and reconstruction in which the reconstructed volume is reprojected and matched to the acquired tilt series.

The study of the performance of the alignment algorithms followed in this work is based on the calculation of residual error vectors. A residual vector is a two-dimensional vector representing the detected marker's relative position in the tilt image compared to its calculated position after solving the series alignment. Thus, the set of residual vectors measures the quality of the computed alignment in a tilt series. A schematic of this definition is shown in Fig. 1.

The goal of residual-based algorithms is to minimize the magnitude of the residual vectors, thus minimizing the reprojection error. Using residual error vectors as a quality metric to report alignment errors is common, even for fiducial-less approaches. In cases where the alignment algorithm does not provide this measurement, this work introduces a set of algorithms for calculating these residual vectors. However, in that case, it will only be possible if the sample includes fiducial markers.

The alignment quality assessment algorithm is tested using the residual vectors from the alignment algorithm and those calculated by the new algorithm introduced in this work. This approach evaluates the algorithm's ability to classify tilt series accurately and the performance of the residual calculator to generate residuals that reflect alignment quality as effectively as the alignment algorithm.

A common practice in the field is to use heuristic techniques to detect the presence of misalignment. These techniques could include visual inspection of aligned tilt series, artifacts in the reconstructed tomogram, or visualization of the trajectories of high-contrast points in the protected tilt series. This work presents an automatic alternative to the misalignment detection problem that does not require user intervention and offers an autonomous quality alignment classification of input tilt series.

Previous work took advantage of the artifacts observed in the tomographic reconstruction of tilt series that present alignment errors (de Isidro-Gómez et al., 2024). This approach, although practical, presents a double pitfall: first, the calculation of the tomographic reconstruction is required, implying a computational payload; and second, the reconstruction process might shadow some of the most subtle alignment errors, becoming imperceptible in the tomogram. The algorithms presented in this work focus directly on the tilt series to assess the alignment quality, so these two pitfalls are avoided.

In summary, we address the need to automatically assess the performance of any tilt series alignment by working directly on the tilt series and avoiding tomographic reconstruction. Hence, the need for manual inspection of the alignment results is avoided, freeing the users from the manual, laborious, and error-prone task of analyzing the alignment quality of the series. The algorithm presented in this work has been implemented in Xmipp (de la Rosa-Trevín et al., 2013; Strelak et al., 2021). It is also accessible through the Scipion workflow engine (de la Rosa-Trevin et al., 2016), within its tomography environment ScipionTomo (Jimenez de la Morena et al., 2022) under the protocol name `xmipptomo - detect misaligned ts`.

2. Methods

The goal of residual-based algorithms is to minimize the reprojection error by reducing the magnitude of the calculated residuals (Sorzano et al., 2020)

$$E = \sum_i \sum_{j \in V_i} \|\mathbf{p}_{ij} - (A_i \mathbf{r}_j + \mathbf{d}_i)\|^2, \quad (1)$$

being $\mathbf{r}_j$ the jth three-dimensional coordinate and $\mathbf{p}_{ij}$ its projection onto tilt image i. The matrix A_i is the projection matrix accounting for the tilt around the tilt axis and a subsequent in-plane rotation,

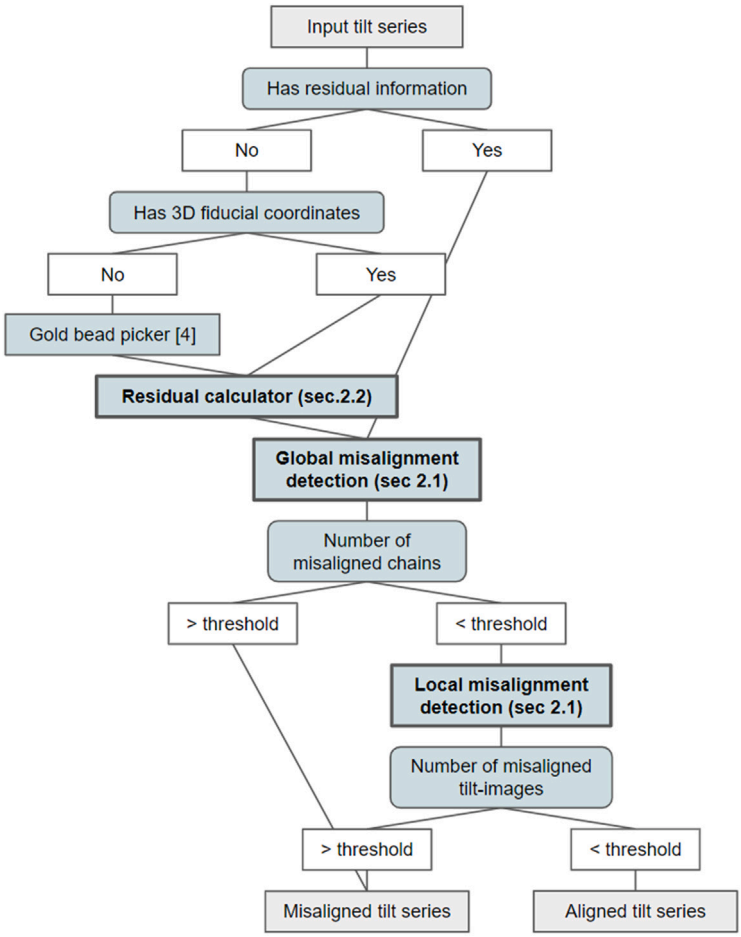


Fig. 2. Scheme of the workflow for tilt series alignment classification. Boxes in bold highlight the steps solved by the algorithms presented in this work.

while $\mathbf{d}_i$ is a 2D vector accounting for an in-plane shift of the ith image. Errors in minimizing this function with respect to A_i may lead to misaligned tilt series. Based on the authors' observations, errors in the pre-alignment of the tilt series (where only shifts are typically corrected via cross-correlation) and suboptimal detection of the landmarks in the tilt images are common sources of alignment errors, in addition to the presence of local minima typical of any optimization problem.

This section introduces the algorithms developed to detect alignment errors based on residual vectors and their characterization. First, we present an algorithm for analyzing these residuals in search of possible misalignments in the series. If the alignment algorithm does not provide the residual vectors, a set of algorithms for their calculation based on detecting fiducial markers in the sample is introduced. A schematic of the proposed workflow is introduced in Fig. 2.

2.1. Residual-based alignment errors detection

This algorithm aims to detect alignment errors through the study of residual vectors, as these provide a measure of the relative position of the detected fiducial in the tilt image with respect to its projected location after solving the alignment, thus providing a measure of the alignment quality. Software packages widely use residual errors in quality measure reports. Software packages widely use residual errors in quality measure reports. It presents the advantage that they are available using different alignment strategies, from fiducial-based algorithms (using gold beads as landmarks) to patch-tracking algorithms (used in fiducial-less samples).

In the development of this work, several measurements and statistics have been studied to characterize the alignment errors present in the tilt series from the reported residual vectors. All calculated statistics are saved and reported to make this information available to the user and any posterior data curing or statistic-based filtering of the analyzed tilt series. All these parameters might be calculated globally or per tilt. The statistics reported are as follows:

1. Convex hull: if the set of residuals represents a set of points C, the convex hull is defined as the unique and minimal convex subset containing C. The residual vectors are characterized by reporting the area and perimeter convex polygon defined by this subset of points (Gonzalez, 2009).
2. Binomial test: used to detect deviations in the sign of a set of residuals (Zar, 2010). The hypothesis is that residuals are randomly distributed, meaning there is no bias in the direction of the residuals, and their sign is equally likely ($p = 0.5$). This hypothesis is tested by defining as a statistic the probability of observing k positive-signed (or negative) residuals in a total set of n elements, as described in Eq. (2)

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}. \tag{2}$$

   Since residuals are two-dimensional vectors, two tests can be performed, one in X and another in Y direction. Thus, this hypothesis is tested positive if the reported p-value of the binomial test is lower than a significance level (typically 0.05).
3. F-test: used to detect differences in the directional variances of a population of residuals. First, the covariance matrix of the set of residuals is calculated, defined in Eq. (3)

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_x^2 & \sigma_{x,y} \\ \sigma_{y,x} & \sigma_y^2 \end{pmatrix}, \tag{3}$$

the eigenvectors of the covariance matrix determine the directions of maximum variance in the residual population, while the eigenvalues, λ_1 and λ_2, determine the dispersion of the residuals along these axes (Lowry, 2014). The hypothesis to be tested is that the difference in the dispersion over both directions is not significant. For this, the statistic of the F test is defined as the quotient of the two eigenvalues, defined in Eq. (4)

$$F = \frac{\lambda_1}{\lambda_2}, \tag{4}$$

this hypothesis is tested positive if the reported p-value of the F-test is less than a significance level (typically 0.05).

4. Augmented Dickey–Fuller test: to test if a population of residuals is stationary (it does not have a unit root). If the set of residuals is stationary, no trend or seasonal effect is present in the data. This means the residuals are randomly distributed, and the mean and variance should be constant along the tilt series. In other words, this test hypothesizes that the residuals present a random walk behavior (Dickey and Fuller, 1979). The statistic of the augmented Dickey–Fuller test is defined in Eq. (5)

$$DF_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}, \tag{5}$$

where $\hat{\gamma}$ is the autoregressive term, a quantification of the influence of the previous value of the series on the current one, and $SE(\hat{\gamma} = \frac{\sigma}{\sqrt{n}})$ is the standard error of $\hat{\gamma}$ where σ is the standard deviation and n the number of observations of the sample. Thus, if the statistic value is more negative than the critical value, the hypothesis is rejected, which means that the series is stationary. Intuitively, this test aims to measure whether the residuals have a random walk distribution or exhibit some directional bias, implying a potential drift in the alignment.

5. Mahalanobis distance: a measure of the distance of each residual to a specific distribution defined by

$$D = \sqrt{(\mathbf{e}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{e}-\boldsymbol{\mu})}, \tag{6}$$

where $\mathbf{e}$ represents the residual vector, $\boldsymbol{\mu}$ represents the mean of the residual population, and $\boldsymbol{\Sigma}$ represents the covariance matrix (Mahalanobis, 2018). We assume a set of residual vectors follows a two-dimensional Gaussian distribution if the tilt image does not present a large misalignment. Note that $\boldsymbol{\mu} = 0$ for our specific case, since the Gaussian residual distribution has zero mean.

Thus, each residual is weighted based on its Mahalanobis distance to a two-dimensional isotropic distribution with a covariance matrix $\boldsymbol{\Sigma}$, whose variance is characterized by the fiducial size. This is defined in Eq. (7)

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}, \tag{7}$$

where $\sigma_x^2 = \sigma_y^2 = \sigma^2$, assuming that the residuals do not have any skew in their distribution. By default, a σ^2 value equal to one-third of the fiducial size is proposed, ensuring that 99.7% of the residuals fall within the fiducial radius in a Gaussian distribution. This threshold is meant to be modified if a more relaxed or strict scenario is pursued. If the residual vectors come from fiducial-less samples, this parameter may be adjusted to allow for the smallest relative movements that ensure a sufficiently accurate reconstruction.

However, for the automatic detection of alignment errors, it is essential to select a statistic (or a set of them) that allows for a robust assessment of the tilt series alignment. From the experiments performed to detect the statistics that hold the most information to characterize the residuals, it has been concluded that the Mahalanobis distance is the most resilient quality metric. This is further explained in the Results section.

Both global and per-tilt misalignment detection are driven once every residual vector is weighted by its Mahalanobis distance. First, a global analysis is performed, and every chain of landmark residuals is analyzed. Each chain comprises a set of residuals referenced to the same landmark at every tilt image, with as many chains as fiducials used in the alignment. If the proportion of chains that exhibit an average Mahalanobis distance greater than 1 exceeds a specified threshold, the tilt series is flagged as misaligned, and the analysis is concluded. This threshold is set at 0.8, meaning that at least 80% of the chains must be classified as aligned.

If global misalignment is not detected, an equivalent analysis is performed for each tilt image with two possible criteria to report misalignment. First, misalignment is flagged if the average Mahalanobis distance of all residuals in the image exceeds one. Alternatively, a voting criterion is used to determine whether the percentage of residuals with a Mahalanobis distance greater than one exceeds a certain threshold (by default 80%). Additionally, a parameter allows the user to specify the maximum number of tilt images that can exhibit misalignment. The entire series is flagged for global misalignment if this threshold is exceeded. If not, only the misaligned tilt images are excluded from further processing.

Finally, we would like to suggest modifying the previous approach, which is more resilient to noisy residual vectors. For this, we use the Z-score, defined in Eq. (8)

$$Z = \frac{|e| - \mu}{\sigma}. \tag{8}$$

Before computing the global and per-tilt average residual distances, residuals whose Z-score exceeds a given threshold (the default parameter is three standard deviations) are removed. This behavior is particularly interesting if the alignment algorithm does not provide the residual vectors and another algorithm must calculate them, as introduced in the next section. This additional filter makes the assessment of alignment quality more robust if the calculation of the residuals is noisy.

Apart from those mentioned above, any tilt series for which at least two three-dimensional coordinates are not provided is excluded from this analysis and classified as misaligned. This is because it is impossible to solve the alignment of a tilt image with only a single projection point if the tilt axis orientation has to be calculated. Calculating the shifts and rotations for every image to reference them to a common geometry is necessary.

2.2. Residual vectors calculation

Scientists do not always have access to alignment quality metrics, particularly the alignment residual vectors. Sometimes, they rely solely on alignment transformations or aligned tilt series. However, to our knowledge, only a few alignment algorithms do not provide any of this information (Zheng et al., 2022). Thus, we introduce algorithms to calculate the residual vectors needed to feed the algorithm proposed in the previous section. To achieve this, the three-dimensional coordinates of a subset of fiducials in the sample must be provided. In an earlier work (de Isidro-Gómez et al., 2024), we introduced an algorithm for detecting these gold beads in case this information is unknown to the user. However, this implies the drawback of calculating the tomographic reconstruction of the tilt series.

The robustness of this step is critical to detecting alignment errors in tilt series. The number of landmark coordinates provided and the quality of the calculated residuals will drive the performance of the misalignment detection algorithm. Nonetheless, several quality control steps have been introduced to ensure the algorithm's robustness to outliers.

This section presents two algorithms. The first algorithm aims to detect fiducial markers, a problem that remains an open challenge in the field (Hou et al., 2024). To achieve this, the steps described below will be executed on each tilt image of the series. The results obtained at each step of the algorithm are exemplified in Fig. 3.

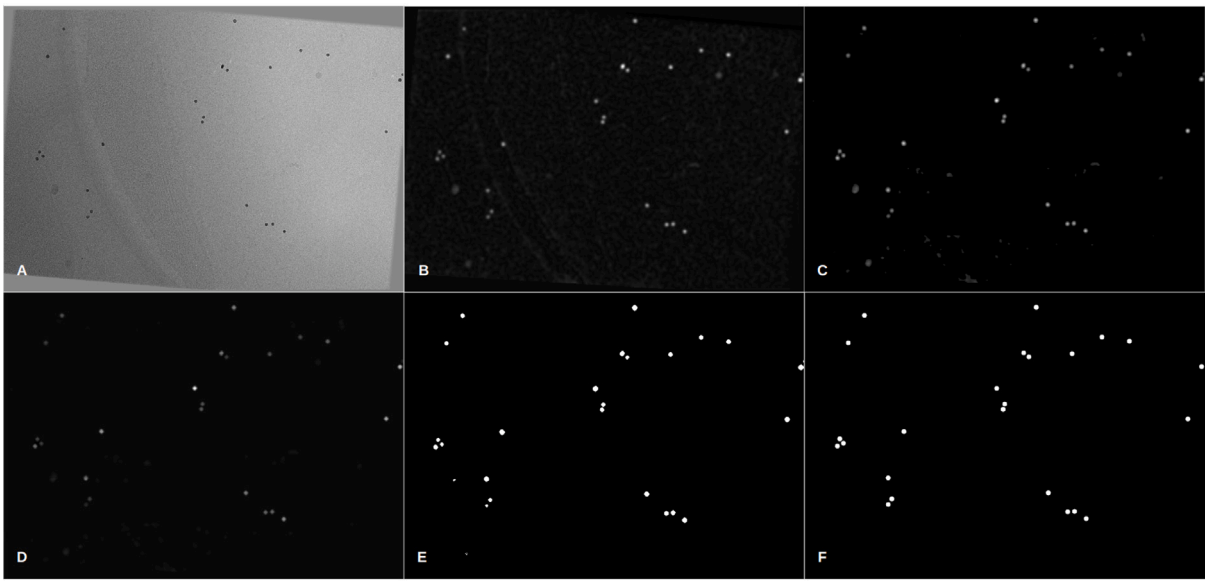


Fig. 3. Intermediate results at different stages of the fiducial detection algorithm: (A) original tilt image, (B) result after landmark enhancement and background subtraction, (C) result after Z-score thresholding, (D) result after maximum pooling and directional filter, (E) labeled regions of interest after preprocessing, and (F) inpainting of the resulting detected landmarks after filtering the regions of interest. This tilt image belongs to tilt series E48g4_30 from EMPIAR 11457 dataset at 0°. An equivalent figure at 60° its provided in the Supplementary Material.

1. Interpolation edges detection: Due to image interpolation during the alignment process, sharp edges are introduced in the images. Also, in some cases, artifacts are observed on the border of the tilt images (see Supplementary Material, Fig. 4 and Movie 1). Thus, when the images are aligned, these artifacts are reallocated in sensitive regions, even spoiling the postprocessing of the images. To solve this, we detect the background of the images, calculating the gradient magnitude image with the use of a Sobel filter defined in Eq. (9)

$$M(x,y) = \sqrt{(G_x \star I(x,y))^2 + (G_y \star I(x,y))^2}, \tag{9}$$

being G_x and G_y, the gradient kernels in the vertical and horizontal directions, and $\star$ the convolution operator. Then, the contour of the images is removed and set to the background value, removing any possible interpolation artifact.
2. Downsampling: The images are downsampled, targeting a specific size of the fiducials. This operation increases the signal-to-noise ratio and computational efficiency. Also, it allows the definition of convolutional kernels to be used in posterior steps.
3. Landmark enhancement: The downsampled images are convolved with a fiducial kernel for enhancement. Fiducials are modeled as a two-dimensional Gaussian distribution as defined in Eq. (10)

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right), \tag{10}$$

as in Eq. (7), $\sigma_x^2 = \sigma_y^2 = \sigma^2$, imposing no skew in the definition of the kernel. The sigma value is adjusted to one-third of the landmark target size from the previous downsampling step. Finally, a rolling-ball background subtraction is applied to the image, with the ball radius set by default to two times the landmark target size (Sternberg, 1983). Customization of this parameter is allowed, ensuring the preservation of fiducial candidates while removing background information. Intermediate results of this step are shown in Fig. 3 B.
4. Detect outlier elements. All the pixels inside the interpolation limits calculated in the first step are analyzed. Those presenting z-scores lower than a given threshold are masked from further analysis, keeping only the outlier values. Intermediate results of this step are shown in Fig. 3 C.

   Then, the images are morphologically dilated (comparable to a maximum pooling operation) to keep the landmark regions more homogeneous. After the thresholding, non-fiducial high-contrast elements might not have been masked out. A band-pass directional filter is applied to ensure the detection's robustness. The image is directionally filtered in as many directions as the user inputs (8 directions by default) with an angle amplitude of 10° for the complete cone. Each direction is also bandpass filtered, centered in the landmark target size. Finally, each direction is combined in a weighted mark applied to the image. Thus, round objects (landmarks) are preserved while removing those with a high signal-to-noise ratio but presenting different shapes (such as carbon edges or membranes). Intermediate results of this step are shown in Fig. 3 D.
5. Filter regions of interest. Finally, the pixels from the previous filtered and masked image with a z-score lower than the given threshold are removed from any further processing. Intermediate results of this step are shown in Fig. 3 D. The image is subsequently labeled, and two different criteria morphologically analyze each region:

   - Relative area: those regions significantly bigger or smaller than the expected area of the target landmark are removed.
   - Circularity: those regions whose shape differs significantly from a circle, measured as the ratio between the area of the regions and the area of its circumscribed circle are removed.

   Thus, after applying these filters, only those regions from the previous image that fulfill the morphological constraints are included as potential fiducials. Intermediate results of this step are shown in Fig. 3 F.
6. Coordinates centering. Optionally, coordinated can be centered by the maximum shift obtained from the Fourier correlation of each landmark with its mirror.

The second algorithm is fed with the landmarks' location at each tilt image. A set of residual vectors is calculated for each tilt image, computing the distance between the projection of each three-dimensional

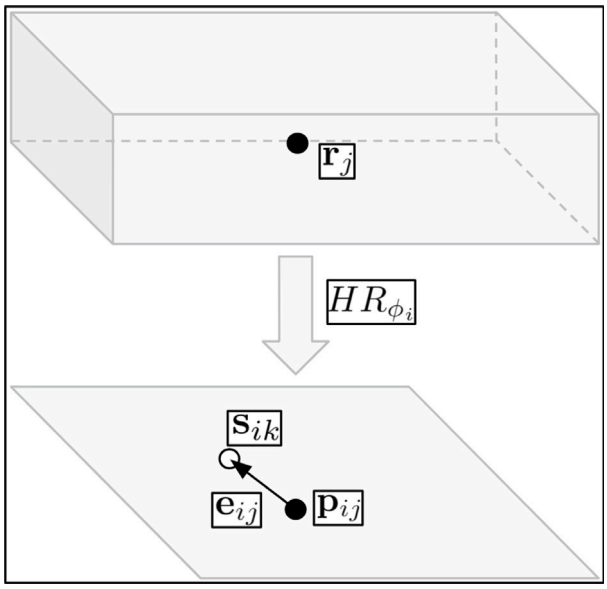


Fig. 4. Schematic of the residual calculation algorithm, showing the projection of the fiducial coordinate onto the tilt image (black dots) and defining the residual as the vector from the projection to the nearest detected landmark (empty dot).

coordinate to its closest detected fiducial. As with the previous algorithm, the set of operations applied to each tilt image is outlined in the following paragraphs.

1. Residual vector calculation. First, the set of three-dimensional coordinates is projected onto each tilt image. Let $\mathbf{r}_j$ be the jth three-dimensional coordinate and $\mathbf{p}_{ij}$ its projection onto tilt image i

$$\mathbf{p}_{ij} = H R_{\phi_i} \mathbf{r}_j, \tag{11}$$

where H is the matrix that projects the three-dimensional coordinate into its X and Y components, and R_{ϕ_i} is a rotation matrix of ϕ_i degrees corresponding to the tilt angle of each tilt image. Notice that no parameter for in-plane shift or rotation is introduced in the projection, unlike in Eq. (1). Although these parameters are typically included in alignment models, errors in estimating these parameters lead to the misalignment we aim to detect. As illustrated in Fig. 4, this step involves calculating the projection of the three-dimensional coordinates onto each tilt image (black dots).
Being $\mathbf{s}_{ik}$ the closest detected landmark in tilt image i to the projection $\mathbf{p}_{ij}$, the residual vector is defined in Eq. (12)

$$\mathbf{e}_{ij} = \mathbf{s}_{ik} - \mathbf{p}_{ij}, \tag{12}$$

which is illustrated in Fig. 4 as the vector whose origin is in the detected projection (black dot) and ends in the closest detected landmark (empty dot).
2. Residual vector pruning. A robust detection of misalignment requires a good estimation of the residual vectors. However, suppose the fiducial corresponding to the projection of one of the three-dimensional coordinates is not detected. In that case, the resulting vector will not adequately characterize the quality of the tilt series alignment. This is especially pernicious in an aligned tilt series, as the vector modulus transitions from quasi-null to one of significant magnitude.
However, if it is an aligned series, the resulting vector will originate from a landmark (the one that has not been correctly detected) and end at another (the next closest one that has been detected). Note that this does not occur if the tilted image is not correctly aligned. In this way, the correlation of the regions at the origin and the end of the vector can be calculated. If the correlation is significant, it validates that the residual originates and ends in a fiducial, removing this residual from future processing.
3. Metadata generation. The vectors that survived the previous pruning are collected into metadata relating the three-dimensional coordinate, the landmark in each tilt image closest to each projection, and the residual vectors associating their relative positions. With this information, it is possible to feed the misalignment detection algorithm introduced in the previous section.

3. Results

This section presents the results of the introduced methods. First, an analysis of the quality metrics introduced in this work is provided, followed by the performance of these methods on three different public datasets available in the EMPIAR public archive (Iudin et al., 2022).

All datasets are processed uniformly before analysis. The tilt series are aligned using the IMOD software package (Mastronarde and Held, 2017) (correcting for both shift and angle) and then visually inspected for classification, segregating the series with the correct alignment from those without.

To consider a tilt series aligned, the transition between the tilt images that compose it must be smooth, and the common landmarks between the different images must describe rectilinear trajectories perpendicular to the tilt axis (which by convention is positioned vertically and centered in each image). Focusing on high-contrast elements common to all images and observing the trajectories they describe facilitate this identification. (See also Supplementary Material, Movies 1 and 2 and Fig. 3).

The maximum number of allowed misaligned tilt images is also set to zero. Thus, if a single image is labeled misaligned, the entire series is labeled misaligned. This criterion maximizes the algorithm's level of stringency. The rationale behind this approach is to ensure that in our experiments, no error in the classification algorithm is shadowed for any tilt image. Thus, all errors are evident in the confusion matrices in the following sections.

For each dataset analyzed, a confusion matrix is presented, divided into two sections based on whether the residual vector source that feeds the classification algorithm is the alignment algorithm (IMOD) or the residual vector calculation algorithm introduced in this work (Automatic).

This work does not aim to analyze the performance of the IMOD alignment algorithms. A better parameter tuning and tilt series preprocessing, such as dose filtering or CTF correction (which localizes the signal in the real space (Glaeser et al., 2021)), could lead to improved performance. Also, some of these tuned choices are sample-dependent. The aim is to provide algorithms that can detect alignment errors in an automatic processing pipeline, where the same parameters are used for all acquired tilt series.

Authors are aware of alternative methods in tomography for tilt series alignment. However, IMOD is considered a standard in the field and provides a comprehensive report of the residual alignment vectors, allowing performance comparison with the methods presented in this work. Thus, we consider this work of special relevance for pipelines that include IMOD as their tilt series alignment algorithm. All of this processing pipeline has been executed in the ScipionTomo framework, ensuring its reproducibility and validation.

Regarding the computational load of the presented algorithms, both exhibit very short execution times. Misalignment detection algorithm execution is almost instantaneous, while the calculation of residuals, although slightly more costly, takes only a few seconds. Additionally, there are no significant differences in execution time relative to image size, as a downsampling process is applied to the images.

3.1. Quality metrics analysis

In the Methods section, a set of metrics is introduced to evaluate the alignment quality of a tilt series that can be grouped into three different conceptual groups: statistical test-based metrics (binomial test, F test, and augmented Dickey–Fuller test), geometric metrics (perimeter and area of the convex hull), and distance metrics (Mahalanobis distance). These metrics are calculated for both chains of residuals associated with the same fiducial along the tilt series, as well as for each set of residuals belonging to the same tilt image.

A typical behavior observed in alignment algorithms is that only a subset of tilt images, particularly those at high tilt, are not well-referenced to the common geometry. This results in the metrics computed over chains of residuals overshadowing the subset of misaligned images if their number is not significant enough. This imposes the limitation that any metric selected for automatic quality assessment must accurately characterize both global misalignment of the tilt series and local misalignment, computed individually for each tilt image. This limitation is more significant than it might initially appear, as some alignment algorithms require only a few fiducials (sometimes just 2 or 3) to resolve the alignment, leading to a reduced set of residuals on which to perform the analysis.

Given these limitations, it has been observed that the Mahalanobis distance provides results that allow for the assessment of alignment quality, as demonstrated in the following results sections. In the ideal case where the alignment is perfectly solved, all residuals should have zero modulus. However, when the perfect solution is abandoned, the residual vectors consistently have a non-zero modulus. If the calculated solution begins to deviate from the ideal, the residual vectors will start to form a two-dimensional pattern. If the calculated alignment is sufficiently close to the ideal solution, it is expected to resemble a two-dimensional Gaussian distribution with a reduced standard deviation. As introduced in the methods section, the standard deviation of the typical distribution of residual vectors for an aligned series is set to one-third of the fiducial radius. This defines a distribution where the majority (99.7%) of the residual vectors are expected to have a modulus smaller than the fiducial radius. If the researcher wishes to be more stringent in the assessment of the tilt series alignment, this value can be reduced. Thus, the Mahalanobis distance effectively detects outlier elements that unexpectedly deviate from this expected distribution.

The effectiveness of the Mahalanobis distance in characterizing the alignment quality is observed in Fig. 5 (left), where the distribution of values is significantly different when comparing an aligned tilt series with a misaligned one. Establishing a threshold for images with an average distance of their residuals greater than one is also straightforward.

Among the metrics introduced in the Methods section, those based on statistical tests do not adapt well to the local alignment quality assessment. As mentioned above, in those cases where the number of fiducials used to assess misalignment is limited, it is not possible to obtain significant results from a statistical test since the sample is not sufficiently large. Furthermore, these tests search for an uneven distribution of the signs of the residual or a bias in its direction. Although these metrics are informative, they are prone to false positives (a misaligned tilt series with high-modulus residuals may present an isotropic distribution) and false negatives (an aligned tilt series with low-modulus residuals may present bias in its direction). The limitation of these metrics to characterize the quality of alignment is exemplified in Fig. 5 (left), where no significant differences in their behavior are evident when comparing the results obtained from an aligned series with those of a misaligned one.

Finally, metrics based on the geometry of the fiducials are absolute metrics. This introduces the limitation of complex thresholding when using these metrics to determine if a set of residuals belongs to a misaligned tilt series. This behavior can be observed in Fig. 5 (right), where, although there is a significant difference in the results obtained for these metrics (especially for the area of the convex hull), the problem of selecting a threshold is also evident. In addition, these metrics are sensitive to possible directional bias in the residuals. If the residual vectors point in a similar direction, the area and perimeter of the convex hull significantly decrease when compared to an isotropic distribution of the fiducials.

Although the Mahalanobis distance is chosen as the metric for automatic alignment quality assessment, all these metrics are reported for every tilt image and residual chain. This approach provides the user with the ability to perform additional filtering of the classifications obtained according to these metrics.

Table 1
Confusion matrix for EMPIAR-10453 dataset.

EMPIAR 10453		Predicted			
Source of residuals		IMOD		Automatic	
		Aligned	Misaligned	Aligned	Misaligned
Real	Aligned	175	0	175	0
	Misaligned	3	59	2	60

3.2. EMPIAR-10453

The first dataset analyzed corresponds to entry 10453 of the EMPIAR database (Turoňová et al., 2020). The sample presents a SARS-Cov-2 spike at a pixel size of 1.33 Åand gold beads of approximately 10 nm. This dataset comprises 237 tilt series and, after a one-by-one visual inspection, 61 present misalignment. Misalignment analysis using the presented algorithms was carried out, and the confusion matrices of the results obtained are presented in Table 1. In this table, the predicted classification and ground truth are compared. The table is divided based on the source of the residual vectors: the alignment algorithm (left) or the algorithm introduced in this work (right).

The F1 score is used to measure the performance of the classification algorithm. This metric is defined as

$$F_1 = 2\frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}}, \tag{13}$$

where TP, FP, and FN are the number of true positives, false positives, and false negatives elements from the confusion matrices, respectively.

For completeness, the Jaccard index is included as an alternative performance measure. Unlike the F1 score, which tends to flatten, the Jaccard index exhibits rapid decay, making it more sensitive to classification errors. It is defined as:

$$J = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{14}$$

and, as in the previous case, TP, FP, and FN are the number of true positives, false positives, and false negatives elements from the confusion matrices, respectively.

For the IMOD residuals, three misaligned tilt series are misclassified, leading to an F1 score of 0.992 and a Jaccard index of 0.9831. In addition, when the residual vectors are provided by the algorithm presented in the work, two misaligned tilt series are misclassified; the F1 score is 0.995, and the Jaccard index is 0.988.

It should be noted that only one false positive is shared between both sources of residuals. This tilt series exhibits an unusual distribution of the fiducials, all clustered in one corner of the images (see Supplementary Material, Fig. 5). This distribution results in low-magnitude residuals, but some instability in the series alignment can be observed, becoming more pronounced in those regions away from the fiducial cluster. This is a clear example of the well-known importance of a homogeneous distribution of the fiducials.

Fig. 5. Metrics calculated for each tilt image from two series in the EMPIAR dataset 10453. The left column displays metrics from an aligned tilt series (TS_005), while the right column shows metrics from a misaligned tilt series (TS_233). For clarity and scale convenience, metrics are divided into three groups: the top row presents statistical tests, the middle row depicts geometrical measurements, and the bottom row illustrates Mahalanobis distances. Note that the geometrical metrics (convex hull area and perimeter) have different scales for clarity.

3.3. EMPIAR-11457

The second dataset analyzed corresponds to entry 11457 of the EMPIAR database (Ni et al., 2023). The sample presents ChAdOx spikes (AZD2816) with a pixel size of 2.18 Åand gold beads of approximately 12 nm. This dataset comprises two acquisitions at two different tilt axis angles (10° and 85° degrees) with 134 tilt series. For simplicity, the results are presented together. After a one-by-one visual inspection, 25 tilt series present misalignment. Both acquisitions are processed in parallel and follow the same steps. The confusion matrices for the classification based on the residual vectors provided by the alignment algorithm and those calculated by the algorithm introduced in this work are summarized in Table 2.

It can be observed that for the alignment algorithm residuals, two misaligned tilt series are misclassified, leading to an F1 score of 0.990 and a Jaccard index of 0.9802. Also, when the residual vectors are provided by the algorithm presented in the work, three misaligned tilt series are misclassified, and another four aligned ones are also misclassified, leading to an F1 score of 0.964 and a Jaccard index of 0.9314.

Table 2
Confusion matrix for EMPIAR-11457 dataset.

EMPIAR 11457		Predicted			
Source of residuals		IMOD		Automatic	
		Aligned	Misaligned	Aligned	Misaligned
Real	Aligned	99	0	95	4
	Misaligned	2	23	3	22

This dataset was particularly challenging, presenting a very low contrast in their high-tilt images. This significantly complicates fiducial detection, a key step in calculating residual vectors. Nonetheless, the false positive ratio obtained in both approaches does not differ significantly. Although the relative difference is high, the absolute rate is low for the false negative ratio. Thus, no tilt series presenting misalignment is included in any further processing.

It is interesting to note that two of the false negatives present in both classifications exhibit interesting behaviors. One presents a subtle movement along the entire tilt series, but the displacement never

Table 3
Confusion matrix for EMPIAR-10364 dataset.

EMPIAR 10364		Predicted			
Source of residuals		IMOD		Automatic	
		Aligned	Misaligned	Aligned	Misaligned
Real	Aligned	18	0	15	3
	Misaligned	0	0	0	0

exceeds the threshold. As for the other, the residuals also do not exceed the threshold since two very close three-dimensional coordinates were used to align the series (see Supplementary Material, Figure 6). This demonstrates that, at least for an extreme case like this, the residuals are not indicative of alignment quality and are compatible with those from an appropriately aligned series.

3.4. EMPIAR-10364

The third dataset analyzed corresponds to entry 10364 of the EMPIAR database (Burt et al., 2020). The sample presents a *Escherichia coli* minicells at a pixel size of 2.24 Å, presenting gold beads of approximately 9 nm. This dataset comprises 18 tilt series, and after their one-by-one visual inspection, there is no present misalignment. The confusion matrices for the classification based on the residual vectors provided by the alignment algorithm and those calculated by the algorithm introduced in this work are summarized in Table 3.

No tilt series is misclassified for the alignment algorithm residuals, leading to an F1 score and a Jaccard index of 1. When residual vectors are provided by the residual vector calculation algorithm presented in the work, three aligned tilt series are misclassified, leading to an F1 score of 0.909 and a Jaccard index of 0.833.

The alignment quality assessment shows better performance for this dataset when the algorithm is fed with residual vectors from the alignment algorithm. As observed in the previous dataset, there is a significant loss of contrast, particularly at high tilt angles. This loss complicates the detection of gold beads throughout the series, especially those classified as misaligned (see Supplementary Material, Figure 7). Evidence of this is that for the three misclassified tilt series, only the image at −60° is flagged as misaligned.

4. Discussion and conclusion

The increase in throughput in cryo-electron tomography acquisition has significantly increased the amount of data that scientists must analyze to elucidate the structures of interest in samples. The immediate consequence in the field is the automation of image processing methods to handle this enlarged data volume. This is especially critical in the initial steps of the pipeline, particularly up to the tomographic reconstruction step, which is essential for any tomography processing application (segmentation, subtomogram averaging, per particle per tilt, etc.).

Among all the steps performed up to this point, tilt series alignment is one of the most crucial and unstable. It is also the most time-consuming task due to data curation and manual quality checks. Therefore, in this work, we present an automatic tool for data curation at this step of the pipeline.

This new software directly analyzes the tilt series before tomographic reconstruction. This contrasts with our previous work (de Isidro-Gómez et al., 2024), which required tomograms, providing an advantage in computational load and data handling. Additionally, this approach allows for detecting alignment errors that might be obscured in the tomographic reconstruction.

To achieve this purpose, the software has been designed to require minimal configuration, which requires only data known to the user. Essentially, only a few options are of interest for modification: first, applying the robust choice for pruning the residual vectors, and second, choosing between voting or the mean as criteria to determine if the tilt series presents misalignment.

The datasets analyzed in this publication aim to present good variability in sampling rate and gold bead size, offering a comprehensive scope of the characteristics of available public cryo-electron tomography datasets. In our testing, we compared the performance of the classification algorithm when fed with either the residuals provided by the alignment algorithm or the residuals calculated by the algorithm introduced in this work. The results show that the performance of the classification algorithm is subtly enhanced when the source of the residual vectors is the alignment algorithm in cases where fiducial detection is compromised (for example, the case of dataset EMPIAR 10364), using the alignment algorithm as a source of residuals offers the advantage that the reported information only considers those residual vectors involved in the alignment. If a landmark is undetected in a tilted image and excluded from the alignment, its residual vector is not reported. This is possible because some alignment algorithms use multiple partial fiducial chains without requiring a single chain covering the entire tilt series.

Additionally, applying robust pruning of residuals or using the voting criteria is recommended if the distribution of vectors is noisy, as this complicates the alignment assessment. For scientists using this software in the ScipionTomo framework, this behavior is configured by default when the residual calculation algorithm sources vectors, although users are always free to change it. Both ways of accessing the software are explained in a guide included in the Supplementary Material.

CRediT authorship contribution statement

F.P. de Isidro-Gómez: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **J.L. Vilas:** Writing – review & editing, Supervision, Software, Formal analysis, Conceptualization. **J.M. Carazo:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **C.O.S. Sorzano:** Writing – review & editing, Validation, Supervision, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge the financial support from the Ministry of Science, Innovation, and Universities, Spain (BDNS n. 716450) to the Instruct Image Processing Center (I2PC) as part of the Spanish participation in Instruct-ERIC, the European Strategic Infrastructure Project (ESFRI) in the area of Structural Biology. This work was also supported by Grant PID2022-136594NB-I00 funded by MICIU/AEI/10.13039/501100011033 and "ERDF A way of making Europe", by the European Union and Comunidad Autónoma de Madrid, Spain through Grant S2022/BMD-7232. Additional support came from the European Union (EU) and Horizon 2020 through the HighResCells grant (ERC - 2018 - SyG, Proposal: 810057) and from the European Union (EU) and Horizon Europe through the Fragment Screen grant (Proposal: 101094131).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jsb.2024.108153.

Data availability

Data and code are open to the public.

References

Burt, A., Cassidy, C.K., Ames, P., Bacia-Verloop, M., Baulard, M., Huard, K., Luthey-Schulten, Z., Desfosses, A., Stansfeld, P.J., Margolin, W., et al., 2020. Complete structure of the chemosensory array core signalling unit in an E. coli minicell strain. Nature Commun. 11 (1), 743.

Castano-Diez, D., Scheffer, M., Al-Amoudi, A., Frangakis, A.S., 2010. Alignator: A GPU powered software package for robust fiducial-less alignment of cryo tilt-series. J. Struct. Biol. 170 (1), 117–126.

Coray, R., Navarro, P., Scaramuzza, S., Stahlberg, H., Castaño-Díez, D., 2024. Automated fiducial-based alignment of cryo-electron tomography tilt series in Dynamo. Structure.

de Isidro-Gómez, F.P., Vilas, J.L., Losana, P., Carazo, J.M., Sorzano, C.O.S., 2024. A deep learning approach to the automatic detection of alignment errors in cryo-electron tomographic reconstructions. J. Struct. Biol. 216 (1), 108056.

de la Rosa-Trevín, J.M., Otón, J., Marabini, R., Zaldívar, A., Vargas, J., Carazo, J.M., Sorzano, C.O.S., 2013. Xmipp 3.0: an improved software suite for image processing in electron microscopy. J. Struct. Biol. 184 (2), 321–328.

de la Rosa-Trevin, J.M., Quintana, A., del Cano, L., Zaldivar-Peraza, A., Foche, I., Gutierrez, J., Gomez-Blanco, J., Burguet-Castells, J., Cuenca, J., Abrishami, V., Vargas, J., Oton, J., Sharov, G., Navas, J., Conesa, P., Vilas, J.L., Marabini, R., Sorzano, C.O.S., Carazo, J.M., 2016. Scipion: a software framework toward integration, reproducibility, and validation in 3D electron microscopy. J. Struct. Biol. 195, 93–99.

Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. J. Amer. Statist. Assoc. 74 (366a), 427–431.

Fernandez, J.J., Li, S., 2021. TomoAlign: A novel approach to correcting sample motion and 3D CTF in CryoET. J. Struct. Biol. 213 (4), 107778.

Fernandez, J.J., Li, S., Bharat, T.A.M., Agard, D.A., 2018. Cryo-tomography tilt-series alignment with consideration of the beam-induced sample motion. J. Struct. Biol. 202 (3), 200–209.

Galaz-Montoya, J.G., Flanagan, J., Schmid, M.F., Ludtke, S.J., 2015. Single particle tomography in EMAN2. J. Struct. Biol. 190 (3), 279–290.

Glaeser, R.M., Nogales, E., Chiu, W., 2021. Single-Particle Cryo-EM of Biological Macromolecules. IOP Publishing.

Gonzalez, R.C., 2009. Digital Image Processing. Pearson Education India.

Hou, G., Yang, Z., Zang, D., Fernández, J.-J., Zhang, F., Han, R., 2024. MarkerDetector: A method for robust fiducial marker detection in electron micrographs using wavelet-based template. J. Struct. Biol. 216 (1), 108044.

Iudin, A., Korir, P.K., Somasundharam, S., Weyand, S., Cattavitello, C., Fonseca, N., Salih, O., Kleywegt, G.J., Patwardhan, A., 2022. EMPIAR: the electron microscopy public image archive. Nucleic Acids Res. 51 (D1), D1503–D1511.

Jimenez de la Morena, J., Conesa, P., Fonseca, Y.C., de Isidro-Gomez, F.P., Herreros, D., Fernandez-Gimenez, E., Strelak, D., Moebel, E., Buchholz, T.O., Jug, F., Martinez-Sanchez, A., Harastani, M., Jonic, S., Conesa, J.J., Cuervo, A., Losana, P., Sanchez, I., Iceta, M., del Cano, L., Gragera, M., Melero, R., Sharov, G., Castano-Diez, D., Koster, A., Piccirillo, J.G., Vilas, J.L., Oton, J., Marabini, R., Sorzano, C.O.S., Carazo, J.M., 2022. ScipionTomo: Towards cryo-electron tomography software integration, reproducibility, and validation. J. Struct. Biol. 214 (3), 107872.

Lowry, R., 2014. Concepts and applications of inferential statistics. Mater. Sci. Appl. 6 (6).

Mahalanobis, P.C., 2018. On the generalized distance in statistics. Sankhyā Ind. J. Stat. A (2008-) 80, S1–S7.

Mastronarde, D.N., Held, S.R., 2017. Automated tilt series alignment and tomographic reconstruction in IMOD. J. Struct. Biol. 197 (2), 102–113.

Ni, T., Mendonça, L., Zhu, Y., Howe, A., Radecke, J., Shah, P.M., Sheng, Y., Krebs, A.-S., Duyvesteyn, H.M.E., Allen, E., et al., 2023. ChAdOx1 COVID vaccines express RBD open prefusion SARS-CoV-2 spikes on the cell surface. Iscience 26 (10).

Robertson, M.J., Meyerowitz, J.G., Skiniotis, G., 2020. Cryo-EM as a powerful tool for drug discovery. Bioorgan. Med. Chem. Lett. 30 (22), 127524.

Seifer, S., Elbaum, M., 2022. ClusterAlign: A fiducial tracking and tilt series alignment tool for thick sample tomography. Biol. Imag. 2, e7.

Sorzano, C.O.S., de Isidro-Gómez, F., Fernández-Giménez, E., Herreros, D., Marco, S., Carazo, J.M., Messaoudi, C., 2020. Improvements on marker-free images alignment for electron tomography. J. Struct. Biol. X 4, 100037.

Sorzano, C.O.S., Messaoudi, C., Eibauer, M., Bilbao-Castro, J.R., Hegerl, R., Nickell, S., Marco, S., Carazo, J.M., 2009. Marker-free image registration of electron tomography tilt-series. BMC Bioinformatics 10, 124.

Sternberg, S.R., 1983. Biomedical image processing. Computer 16 (01), 22–34.

Strelak, D., Jiménez-Moreno, A., Vilas, J.L., Ramírez-Aportela, E., Sánchez-García, R., Maluenda, D., Vargas, J., Herreros, D., Fernández-Giménez, E., de Isidro-Gómez, F.P., Horacek, J., Myska, D., Horacek, M., Conesa, P., Fonseca-Reyna, Y.C., Jiménez, J., Martínez, M., Harastani, M., Jonić, S., Filipovic, J., Marabini, R., Carazo, J.M., 2021. Advances in xmipp for cryo–electron microscopy: From xmipp to scipion. Molecules 26 (20), 6224.

Turoňová, B., Sikora, M., Schürmann, C., Hagen, W.J.H., Welsch, S., Blanc, F.E.C., von Bülow, S., Gecht, M., Bagola, K., Hörner, C., et al., 2020. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. Science 370 (6513), 203–208.

Van Drie, J.H., Tong, L., 2022. Drug discovery in the era of cryo-electron microscopy. Trends Biochem. Sci. 47 (2), 124–135.

Xu, Z., Li, H., Wan, X., Fernández, J.-J., Sun, F., Zhang, F., Han, R., 2024. Markerauto2: A fast and robust fully automatic fiducial marker-based tilt series alignment software for electron tomography. Structure.

Zar, J.H., 2010. Biostatistical Analysis. Prentice Hall.

Zheng, S., Wolff, G., Greenan, G., Chen, Z., Faas, F.G.A., Bárcena, M., Koster, A.J., Cheng, Y., Agard, D.A., 2022. AreTomo: An integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction. J. Struct. Biol. X 6, 100068.

Automatic detection of alignment errors in cryo-electron tomography

Supplementary material

F.P. de Isidro-Gómez[1,2], J. L. Vilas[1], J.M. Carazo[1*], C.O.S. Sorzano[1*]

November 23, 2024

[1] Biocomputing Unit, Centro Nacional de Biotecnologia (CNB-CSIC), Darwin, 3, Campus Universidad Autonoma, 28049 Cantoblanco, Madrid, Spain
[2] Univ. Autonoma de Madrid, 28049 Cantoblanco, Madrid, Spain
[*] Corresponding author

1 User guide

A simplified user guide to use the software presented in this work is introduced in this section. This guide includes both available sources available to the user: Xmipp standalone command-line mode and its integration in the Scipion framework.

1.1 Xmipp standalone

To work with the tools presented in the standalone version, only the Xmipp software package must be installed. Through the command line, the user has access to the full functionality of this software. The instructions for installing Xmipp can be found in github repository of Xmipp.

The xmipp program `xmipp_tomo_calculate_landmark_residuals` is responsible of detecting misaligned tilt series based on its residual vectors. The parameters of this program are summarized in Table 1.1. This is an example command for the execution of this program:

```
xmipp_tomo_detect_misalignment_residuals -i inputTs.mrcs --inputResInfo vResMod.xmd -o
alignmentReport.xmd --samplingRate 3.00 --fiducialSize 100.00
```

Parameter	Description
–inputResInfo	Input file containing residual models
-o	Output location for alignment report
–samplingRate	Pixel size of the input tilt series
–fiducialSize	Gold bead size in Angstroms
–numberTiltImages	Number of tilt-images in the series
–removeOutliers	Remove outlier residuals (robust mode)
–voteCriteria	Use a votting criteria (instead of average)

Table 1: Xmipp detect misalignment from residuals algorithm parameters.

If residual vectors are not provided, they can be calculated using the xmipp program `xmipp_tomo_calculate_landmark_`. The parameters of this program are summarized in Table 1.1. This is an example command for the execution of this program:

```
xmipp_tomo_calculate_landmark_residuals -i inputTs.mrcs --tlt inputTlt.tlt --inputCoord
inputFiducialCoordinates.xmd -o vResMod.xmd --samplingRate 3 --fiducialSize 100.00 --thrSDHCC
3.00 --targetLMsize 8.00
```

1.2 Scipion framework

To work with the tools presented inside the scipion framework, it is necessary to have installed both the Xmipp and the Scipion software packages. This procedure is simplified since the installation of Xmipp

Parameter	Description
-i	Input tilt series
–tlt	Input tilt angle file
–inputCoord	Input 3D coordinates file
-o	Output location for residual models file
–samplingRate	Pixel size of the input tilt series
–fiducialSize	Gold bead size in Angstroms
–thrSDHCC	Threshold Z-score to consider a pixel value an outlier
–numberFTdirOfDirections	Number of directions to apply the bandpass filter
–targetLMsize	Target fiducial size for downsampling

Table 2: Xmipp calculate residuals from landmarks algorithm parameters.

is triggered when installing Scipion. The instructions for installing Scipion with Xmipp can be found in official webpage of Scipion. Since this is a simplified tutorial, the input information needed to feed the presented protocols is assumed to be available inside the Scipion project. However, more extensive documentation and tutorials on tomography data processing can be found in Scipion documentation landing page.

In particular, the "Tomogram Reconstruction" tutorial explains to the user how to obtain a set of aligned tilt series from a raw set of movies.

All algorithms presented in this work are maintained under the same Scipion protocol called `xmipptomo - detect misaligned TS`. Depending on the information that the user inputs to the protocol, it will trigger the different algorithms needed to assess the quality of the alignment.

Some software provide the residual information needed to directly study the alignment quality of the tilt series. This is, for example, the case of IMOD that is also the one used in the aforementioned tutorial. If the software provides this information, it is available inside the Scipion framework and thus can be input directly to the protocol. If this is the use case, an example of the configuration of the configuration of the protocol is available in Figure 1.

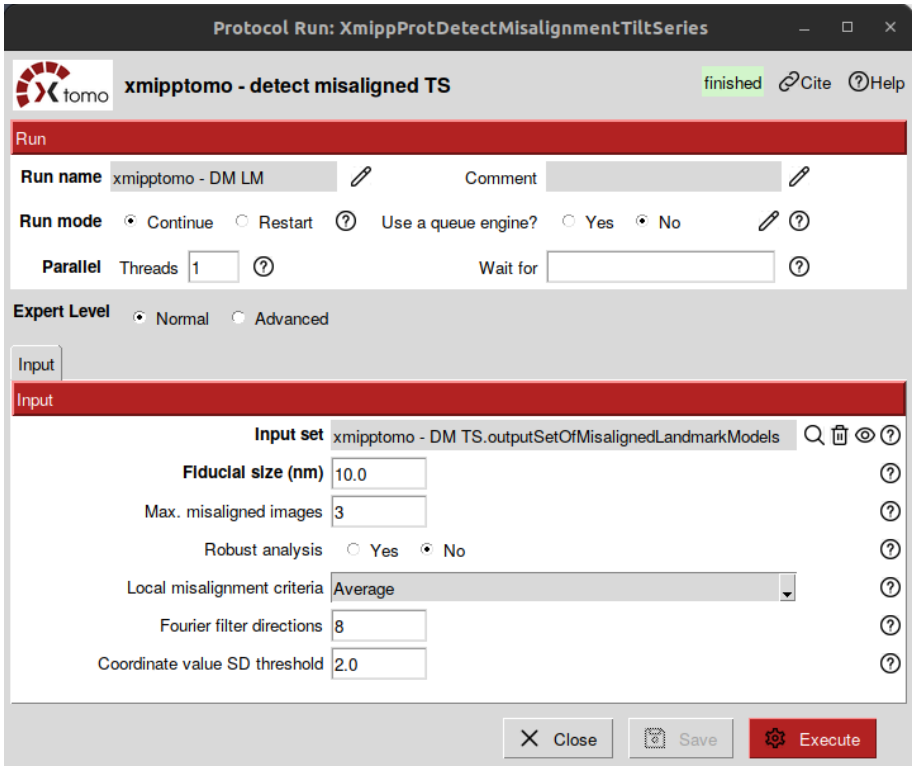


Figure 1: Detect misalignment protocol using residual vector models as input.

In case the residual information is not available is still possible to assess the quality of the alignment using the tilt series and the three dimensional coordinates of the fiducials. If this is the use case, an

example of the configuration of the configuration of the protocol is available in Figure 2.

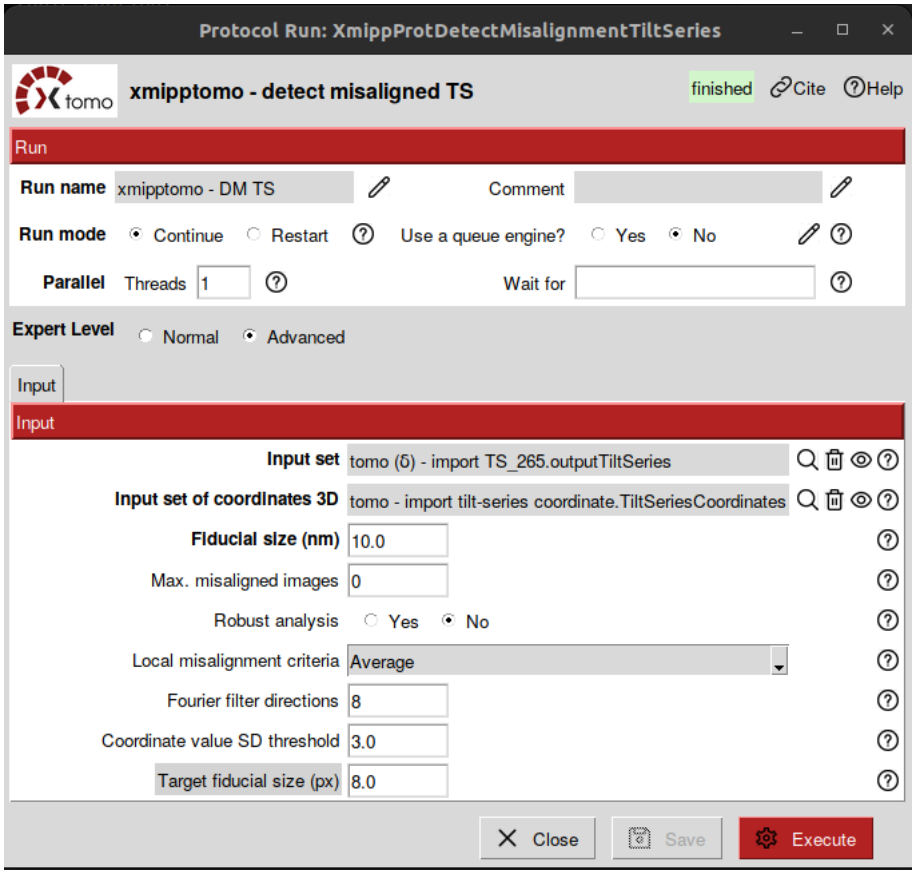


Figure 2: Detect misalignment protocol using the tilt series and the three dimensional coordinates of the fiducials as input.

The only option available in the protocol that is not in the Xmipp standalone version is the Maximum number of misaligned images. This option is inherent to the Scipion framework and sets the tilt series as misaligned if the number of detected misaligned tilt images exceeds this threshold.

2 Supplementary Movies

Two movies have been appended to the Supplementary Material to provide a clearer comparison between a tilt series with properly estimated geometry and one without:

1. Movie 1 presents and example of an aligned tilt series. This is tilt series TS_019 from EMPIAR 10453 dataset.

2. Movie 3 presents and example of a misaligned tilt series. This is tilt series TS_229 from EMPIAR 10453 dataset.

In addition, the calculated average Mahalanobis distance for each of these tilt series is shown in Figure 3.

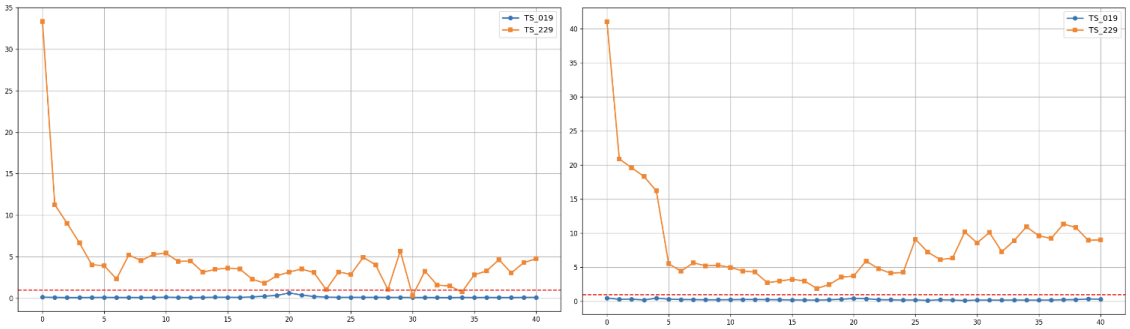


Figure 3: Associated average Mahalanobis distance for each tilt image belonging to Movies 1 and 2. The left plot shows the distribution of distances for each tilt image when the Mahalanobis distances are calculated over the residuals provided by IMOD. The right plot shows the same information when the residual calculation algorithm introduced in this work is the source of the residuals.

3 Supplementary Figures

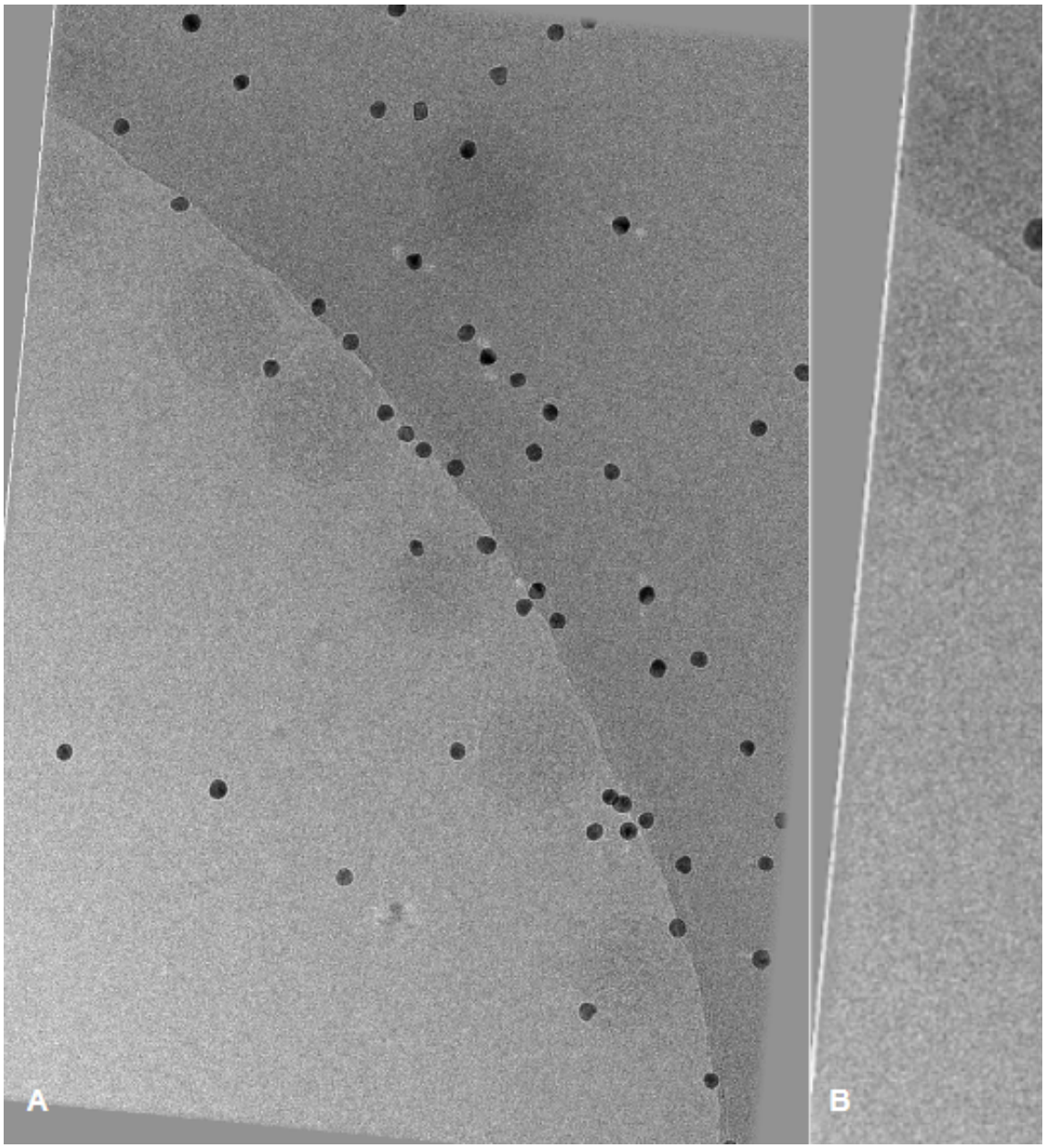


Figure 4: Example of border artifacts. (A) shows a tilt image presenting these artifacts in tilt series TS_005 from EMPIAR 10453, and (B) shows the detail.

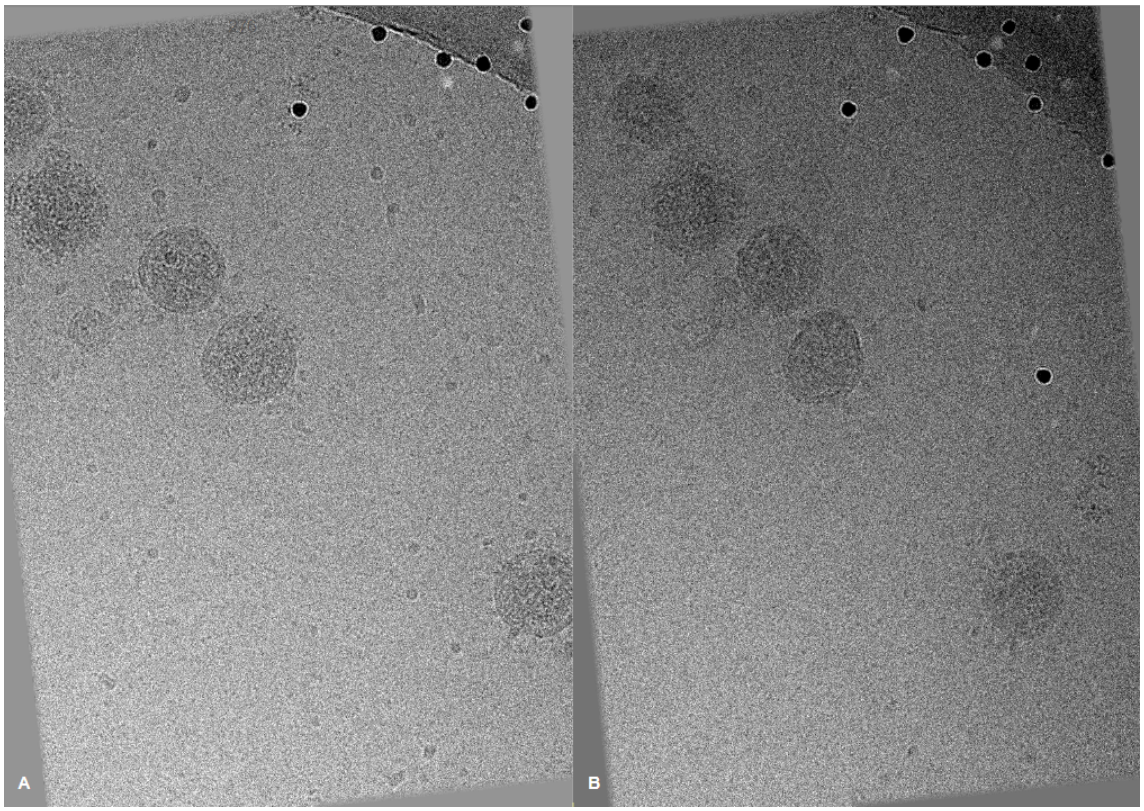


Figure 5: Tilt series TS_276 from EMPIAR 10453 dataset, false positive example from alignment quality assessment algorithm. This tilt series shows an unexpected distribution of residuals, being all of them clustered in one corner of the sample. (A) and (B) shows tilt images at 0° and 42° tilt angle respectively.

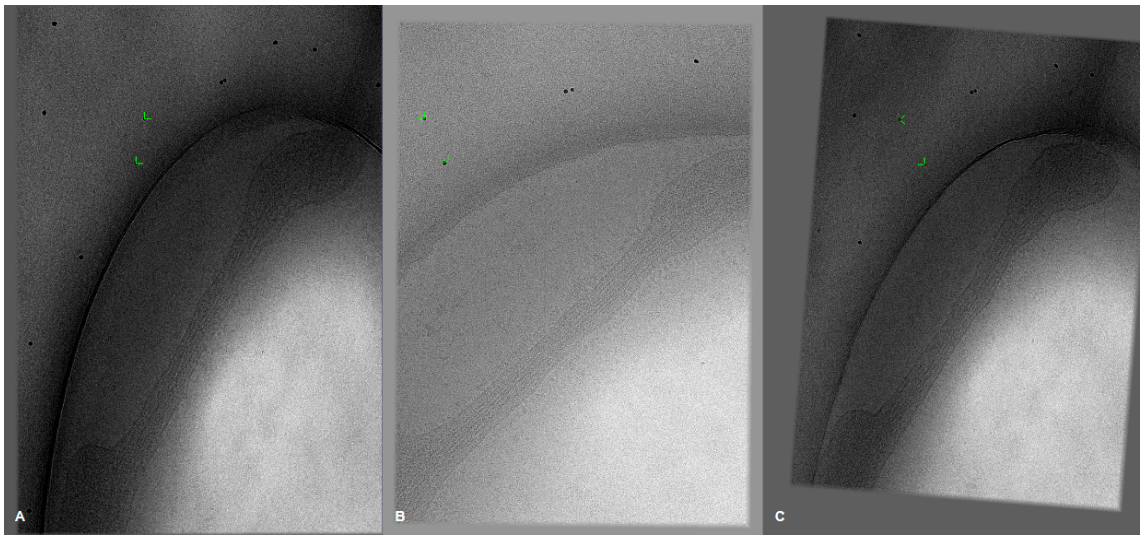


Figure 6: Tilt series E48g4_18 from EMPIAR 11457 dataset, false positive example from alignment quality assessment algorithm. This tilt series was aligned only using two very close fiducials, leading to low modulus residual vectors along the series. Residuals are plotted as yellow arrows but since their modulus is low, only the arrowhead is visible.

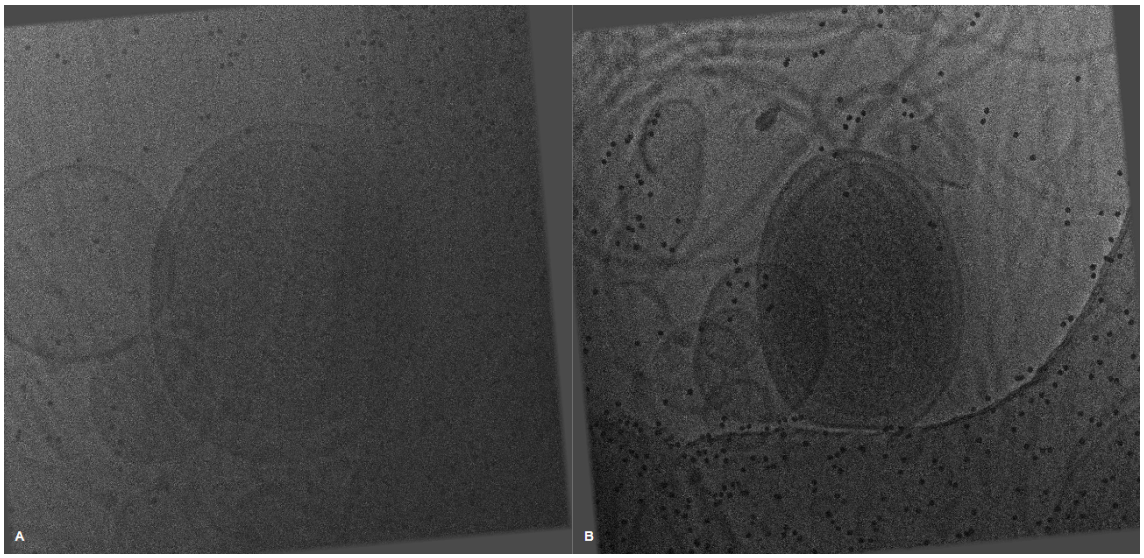


Figure 7: Tilt series TS002 (A) and TS017 (B) from EMPIAR 10364 dataset, both tilt images at -60°. Tilt series TS002 is a false negative example from the alignment quality assessment algorithm. The low contrast at high tilt for this tilt series difficult the fiducial detection and the posterior assessment of the alignment. The reduction in contrast is compared to TS017 tilt series, which has been correctly classified.

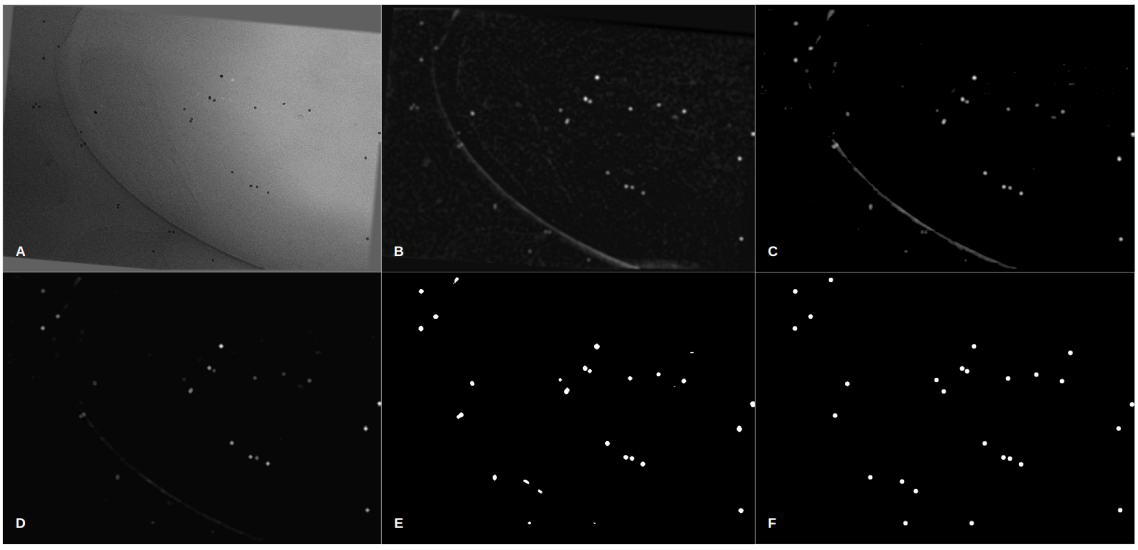


Figure 8: Intermediate results at different stages of the fiducial detection algorithm: (A) original tilt image, (B) result after landmark enhancement and background subtraction, (C) result after Z-score thresholding, (D) result after maximum pooling and directional filter, (E) labeled regions of interest after preprocessing, and (F) inpainting of the resulting detected landmarks after filtering the regions of interest. This tilt image belongs to tilt series E48g4_30 from EMPIAR 11457 dataset at 60°.