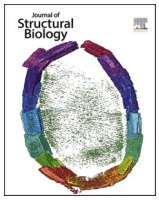
Journal of Structural Biology 216 (2024) 108056

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

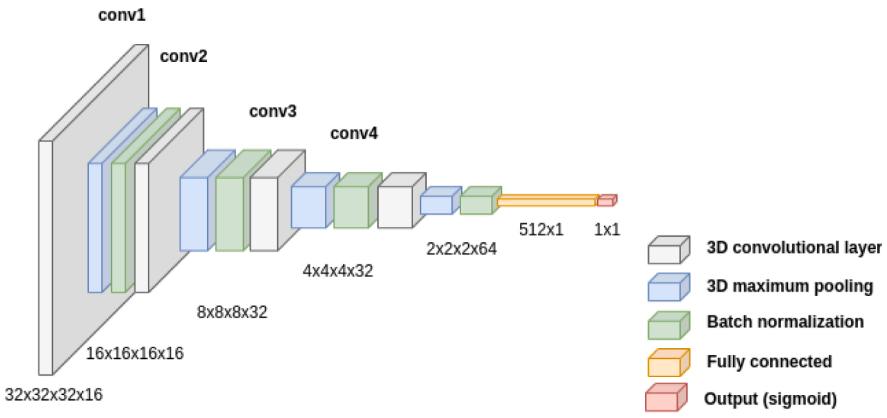


Fig. 1. Structure of the neural network for misalignment detection from fiducial markers.

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:

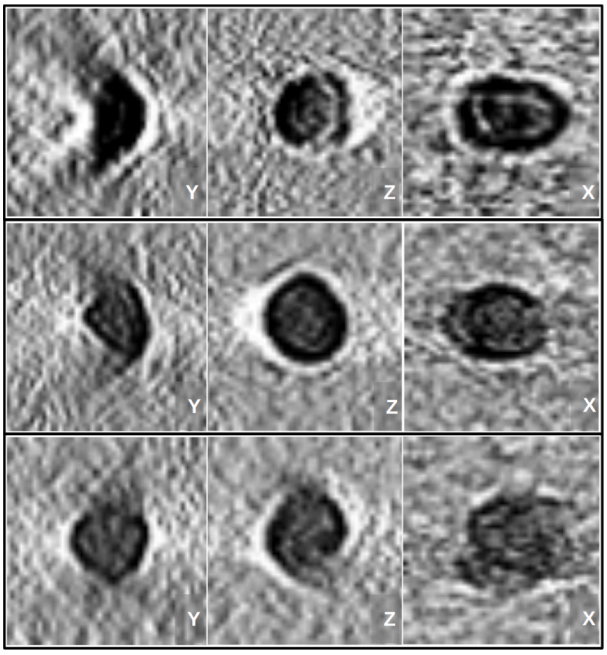


Fig. 2. Slices of the same fiducial marker reconstruction under different misalignments. The left column shows the *XZ*-plane, the center column the *XY*-plane, and the right column the *YZ*-plane. The first row presents bananas (incremental shift in *X*), the second line presents twisters (rotation 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.

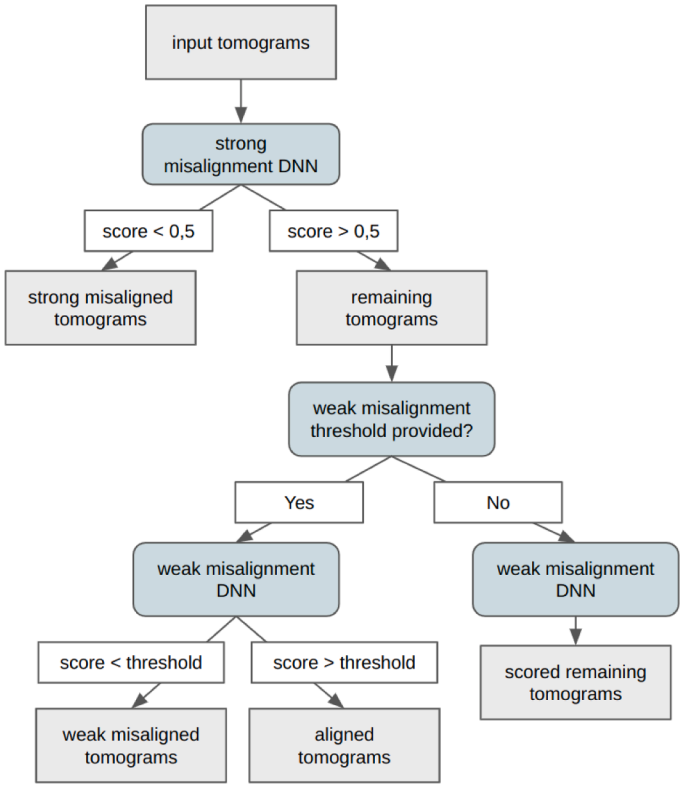


Fig. 3. Workflow for tomogram alignment classification.

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).

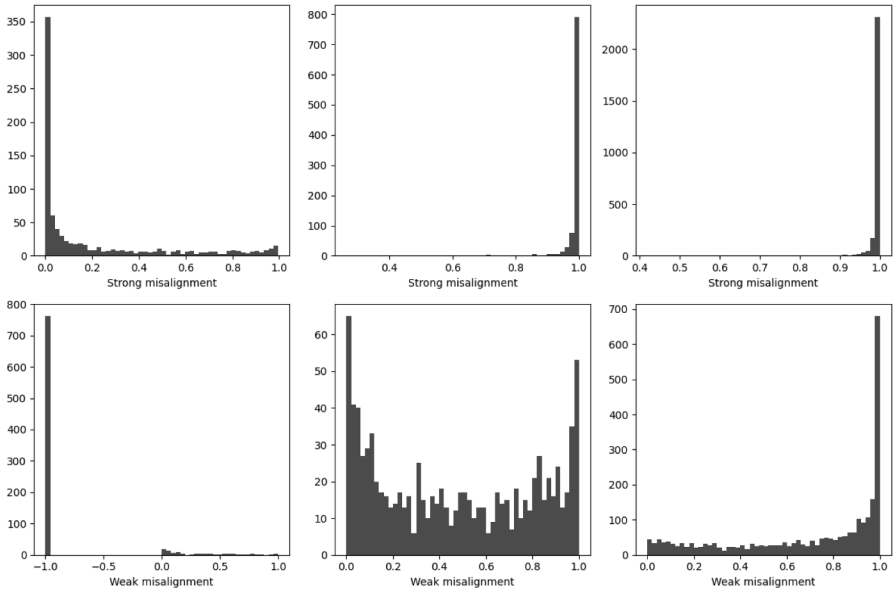


Fig. 4. Histogram distributions of the misalignment scores retrieved by both networks. The alignment algorithm provides the coordinates of the studied markers. The first row corresponds to the scores obtained by the first network (strong misalignment) and the second row to the second network (weak misalignment). Left, central, and right columns correspond to strongly misaligned, weakly misaligned, and aligned datasets respectively.

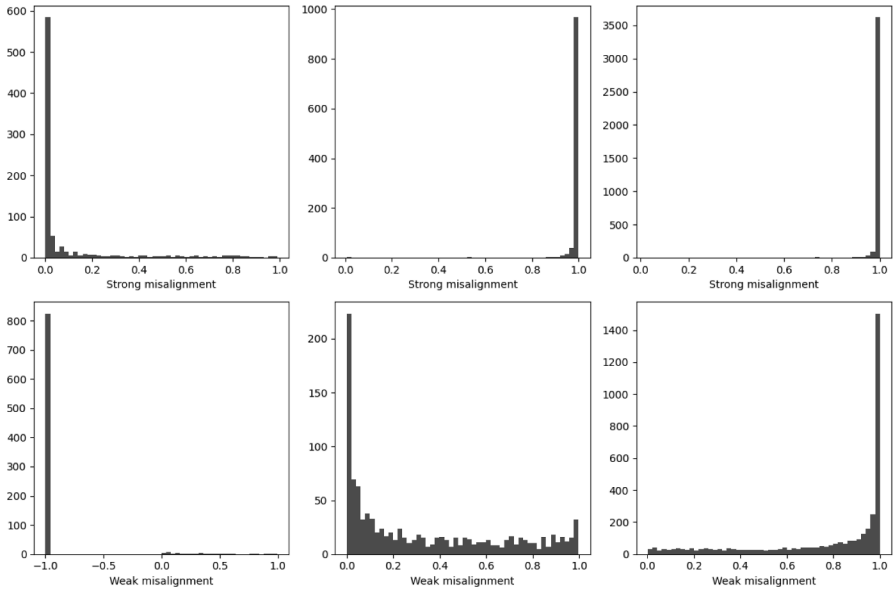


Fig. 5. Histogram distributions of the misalignment scores retrieved by both networks. The fiducial marker location algorithm provides the coordinates of the studied markers. The first row corresponds to the scores obtained by the first network (strong misalignment) and the last one to the second network (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

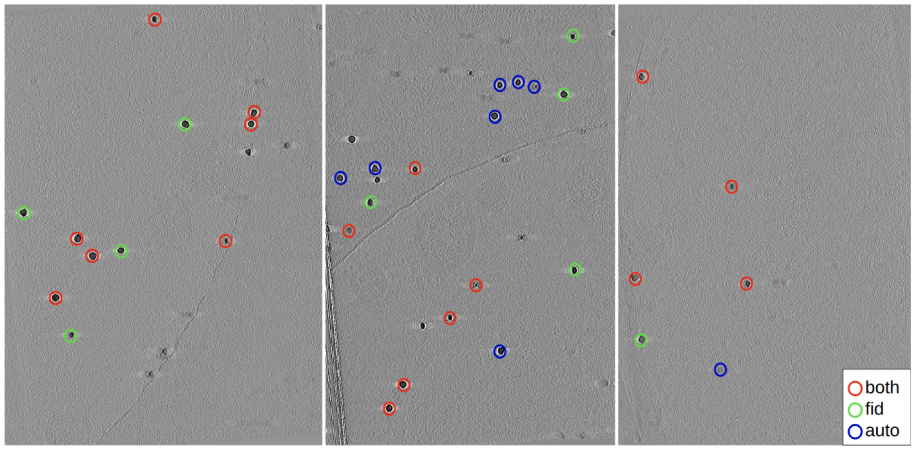


Fig. 6. Comparison of the high-contrast objects automatically found in three different tomograms: aligned (left), weakly misaligned (center), and strongly misaligned (right). The color code is: green (found by tilt series alignment but not automatically), blue (found as a high-contrast feature but not 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

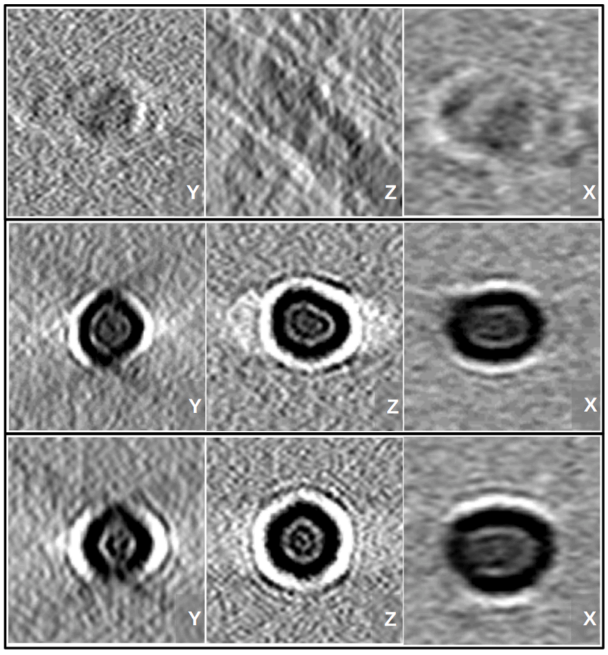


Fig. 7. Slices of fiducial markers that exemplify the three different tomogram sets. The left column shows the XZ-plane, the center column the XY-plane, and the right column the YZ-plane. The first row presents a marker from the strongly misaligned set; the second row presents one from the weakly misaligned 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

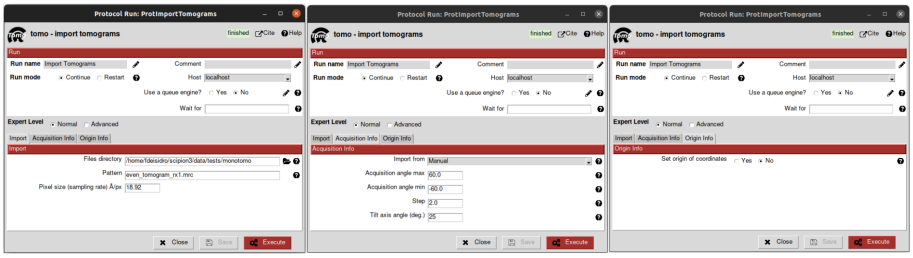


Figure 1: Import tomograms protocol form, including the views of its three slices.

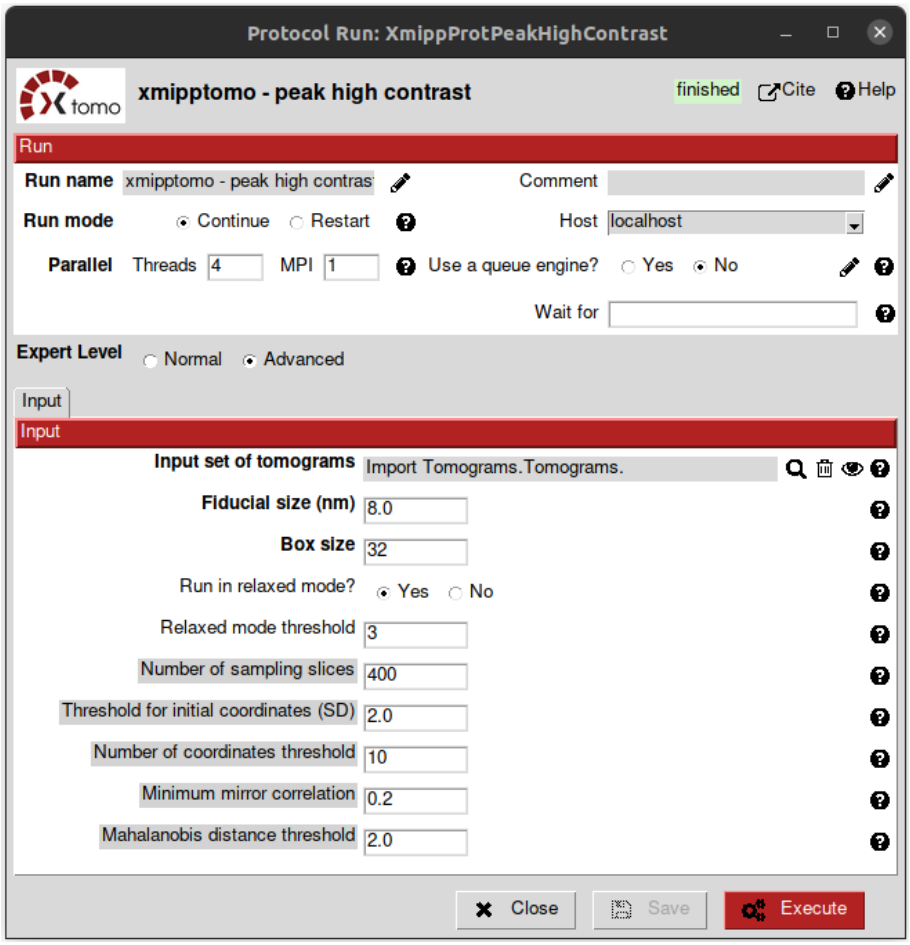


Figure 2: Peak high contrast protocol form.

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.

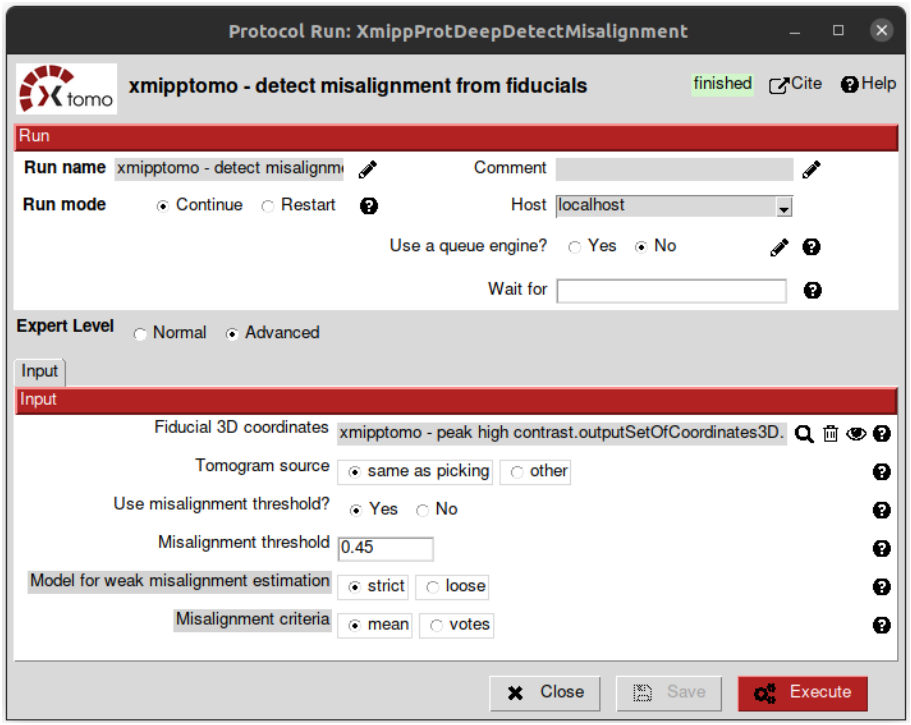


Figure 3: Detect misalignment protocol form.

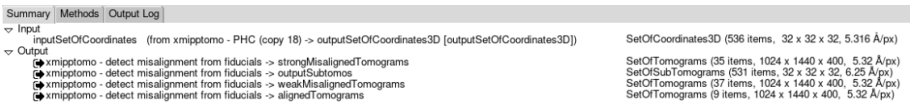


Figure 4: Detect misalignment output summary.