# SUPPLEMENTARY: GUM-NET: UNSUPERVISED GEOMETRIC MATCHING FOR FAST AND ACCURATE 3D SUBTOMOGRAM IMAGE ALIGNMENT AND AVERAGING

**Annonymous**

## Contents

## List of Figures



## List of Tables

# S1 Background

## S1.1 Cryo-electron tomography

Scientists have long been trying to understand biological processes through isolating and purifying macromolecules from a cell. Powerful techniques such as X-ray crystallography and single-particle cryo-EM have been providing near atomic resolution structures to infer macromolecules' function and mechanism *in vitro*. Despite the enormous success, it is coming to a realization that the complex networks among cellular components rather than individual structures define the ultimate function [1]. Fluorescence imaging, on the other hand, provides the localization and interaction information of labeled cellular molecules. However, the resolution is much lower, and the efficiency and protein structural perturbation of fluorophore-labeling post challenges. As our knowledge of the cell gets deeper, cryo-electron tomography (cryo-ET) emerges as a revolutionary technique that studies all macromolecular structures, their spatial organizations, and interactions with other subcellular components in single cells. Consequently, cryo-ET becomes the foundation of the emerging field of *in situ* structural biology. Through the process of sample vitrification, thinning, electron imaging, and data analysis, scientists can recover the subcelluar structural map inside a single cell, at the resolution and coverage not attainable by any other techniques [2]. Moreover, by resolving the molecular differences between healthy and disease states, cryo-ET is expected to assist not only structural biology discoveries but also medical diagnostics in the future [3]. For example, in 2014, cryo-ET was applied for the first time to human clinical samples to elucidate human ciliary structural defects in patients with primary ciliary dyskinesia, which the conventional diagnosing tool EM failed 30 % of the time [4]. Later, Wang et al. [5] demonstrated the effectiveness of using cryo-ET as a non-invasive tool to identify ovarian cancer patients by imaging their platelets. They build a simple model using the number of mitochondria and length of microtubules in cryo-ET images and correctly predicted 20 of 23 cases. Other studies have identified cellular structural changes in disease states such as Leigh syndrome [6], Huntington's disease [7], and virus infection [8].

Because the raw cryo-ET data is noisy and unlabeled, advanced data processing techniques specifically designed for cryo-ET are needed. In the past decade, due to the development of better instruments and data collection software, cryo-ET data is collected at an increasingly faster pace. Cryo-ET enters into the rim of high-throughput techniques [9]. Nevertheless, the traditional exhaustive methods have very high computational and manual quality control costs because macromolecules are of random orientation and displacement inside a tomogram. For example, without parallel computation, scanning one structural template over one tomogram will take about a month because the parametric space of 3D rotation and translation is huge (around hundreds of millions times correlation computation). Nowadays, many public EM repositories have been developed such as the Electron Microscopy Data Bank (EMDB) [10] and the Electron Microscopy Public Image Archive (EMPIAR) [11]. As more and more cryo-ET datasets are collected and open to the public, improving the efficiency and accuracy of cryo-ET data analysis methods becomes urgent and important.

Since 2017, deep learning-based cryo-ET data analysis methods have been proposed and attracted a lot of attention. However, the majority of them are simple supervised models for cryo-ET data classification [12] or semantic segmentation [13, 14]. Creating valid training data requires intensive computation and manual quality control, which still limits these methods at a proof-of-principle stage. On the other hand, one unsupervised model [15] has been proposed to coarsely filter subtomograms by dimensionality reduction using an autoencoder and clustering the latent representations using K-means. To facilitate grouping together the subtomograms containing the same structure but in different orientations, they proposed a pose normalization step to coarsely normalize the orientation and displacement of structures inside a subtomogram before feeding into the autoencoder. Because both the pose normalization and clustering steps are separated from the trainable autoencoder model, this model's accuracy is low. In practice, this model is usually applied as a pre-processing filtering step before doing subtomogram alignment and averaging, which is the main focus of our proposed Gum-Net model.

## S1.2 Missing wedge effect

The three-dimensional cryo-tomographic image is obtained by imaging the cell sample through a series of tilt projections. The tilt projections are subsequently feed into a reconstruction algorithm to produce 3D tomographic reconstruction. Because of the increasing effective sample thickness during tilting, to prevent excessive electron beam damage to the cell sample, the tilt angle range is limited typically to $\pm 60^\circ$ with $1^\circ$ step size [16]. This results in a double V-shaped missing value region (a.k.a. missing wedge) of Fourier coefficients of the reconstructed tomogram in Fourier space. The missing wedge effect also produces image distortion in the spatial domain such as the elongation of features along the direction of the missing wedge axis. Fortunately, in a tomogram, different copies of the same macromolecular structure in different orientations will have different missing wedge distortions. Therefore, by aligning and averaging multiple subtomograms of the same structure, the missing wedge effect can be eliminated to recover a higher resolution structure. However, as a form of image distortion, the missing wedge effect must be taken into account during subtomogram

alignment. Figure S1 shows a noise-free proteasome structure reconstructed from different tilt angle ranges. $\pm 90^\circ$ tilt angle range means there is no missing wedge angle. $\pm 60^\circ$ tilt angle range means $30^\circ$ missing wedge angles. 2D slices representation of a 3D image is the series of 2D images broken at the z-axis.

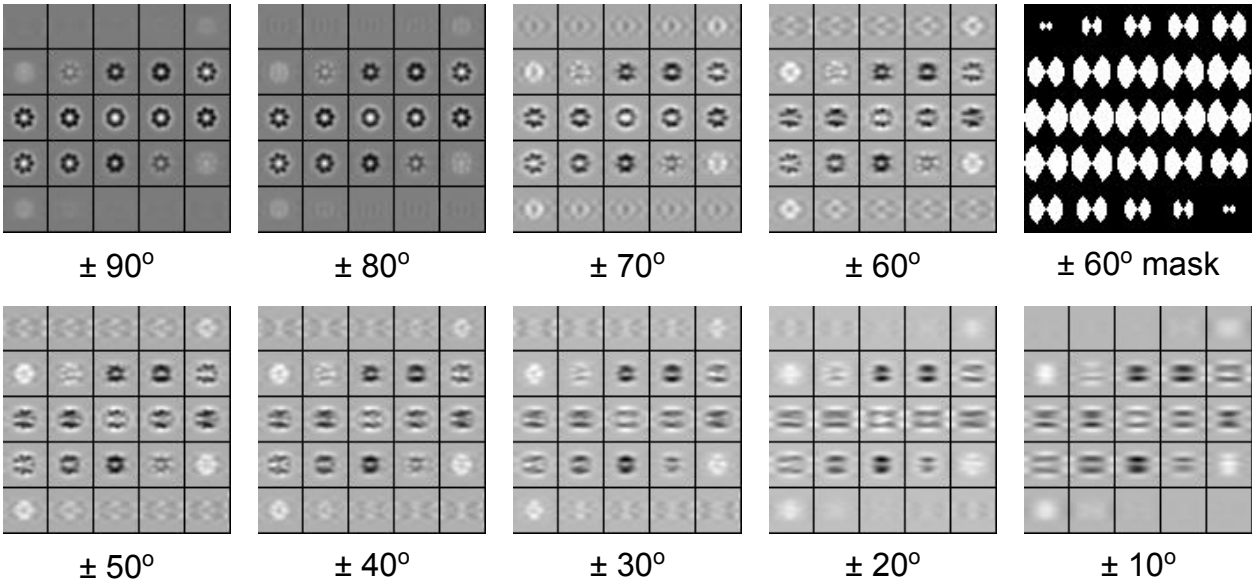


Figure S1: 2D slices representation of a noise-free proteasome structure (PDB ID: 1KP8) reconstructed from different tilt angle ranges. The structure becomes more distorted as the missing angle increases. $\pm 60^\circ$ mask is a spherical missing wedge mask for the structure with $\pm 60^\circ$ tilt angle range.

## S2 Implementation Details

The two most popular state-of-the-art subtomogram alignment methods are (1) high-throughput subtomogram alignment (H-T align) [17], and (2) fast and accurate subtomogram alignment (F&A align) [18]. These two subtomogram alignment methods are traditional geometry-based methods which do not involve a training process and thus do not need labeled training data either. We note that there are two major approaches to tackle the missing wedge effects. One is to estimate the missing values, such as in our spectral data imputation step, and the other is to restrict the correlation measure to observed regions, such as in the two baseline methods [17, 18]. The relationship between the two approaches has been discussed in our previous work [19], where we mathematically proved that some of the latter approach is equivalent to special cases of the former approach.

H-T align has one parameter to tune, the cardinality of the set of suboptimal rigid transformations computed under a translation-invariant upper-bound. We use default parameter 36.

F&A align has three parameters to tune: the bandwidth of the spherical function, the maximal distance allowed to shift, and the maximal frequency in the calculation. We use the default bandwidth parameter [4, 64] and 16 (half of the image size) as the maximal distance allowed to shift. We tuned the maximal frequency to 8 to achieve the highest accuracy for the testing dataset of SNR 100 containing 5000 pairs of simulated subtomograms.

To demonstrate the improvement using proposed modules, we performed three ablation studies with existing modules in [20, 21]. Namely, Gum-Net Max Pooling (Gum-Net MP) and Gum-Net Average Pooling (Gum-Net AP) are equipped with convolution and max pooling or average pooling operations for feature extraction without the proposed DCT spectral pooling & filtering layers. As max pooling achieves more local transformation invariance [22], we expect Gum-Net MP's performance to be worse than Gum-Net AP because local transformation invariance is not desirable for geometric matching. Gum-Net Single Correlation (Gum-Net SC) is equipped with only one correlation layer for computing the correlation map instead of using the proposed Siamese correlation layer. All the other components, including the number of convolution layers for feature extraction and the training processes, of the three Gum-Net baselines remain unchanged. In Gum-Net, four DCT spectral pooling & filtering layers were inserted between five convolution layers, each resizing the feature map to $26^3$, $18^3$, $12^3$, and $8^3$ sequentially. The cropped spectra were of size $22^3$, $15^3$, $10^3$, and $7^3$, respectively. After the fifth convolution layer, the two input feature maps to the Siamese matching module were of size $6^3$. The two correlation maps were each processed with two convolution layers, flattened, and concatenated. Then, after two fully connected layers, the output is the estimated 6 rigid transformation parameters. For the Gum-Net baseline MP and AP, the spectral pooling & filtering layers were replaced with two max pooling layers (or average pooling) to resize the feature map to the same size $8^3$ for the fifth convolution layer. For the Gum-Net baseline SC, only $c_{ab}$ was computed in the matching module. One model was trained for 500 epochs on the simulated alignment dataset of SNR 100 using the Adam optimizer with a learning rate of $1 \cdot 10^{-5}$ and decay $2 \cdot 10^{-8}$ [23]. To speed up convergence, this model was used as initialization and fine-tuned for real datasets and other simulated datasets for only 200 epochs a learning rate of $1 \cdot 10^{-6}$ and decay $5 \cdot 10^{-9}$. No external pre-trained models or additional supervision was used. All models was trained on a computer with four NVIDIA GeForce Titan X Pascal GPU cores. For transforming a subtomogram given the 3D rigid transformation parameters, Gum-Net and all the five baselines use the trilinear interpolation method.

Figure S2 shows the architecture of the proposed Gum-Net.

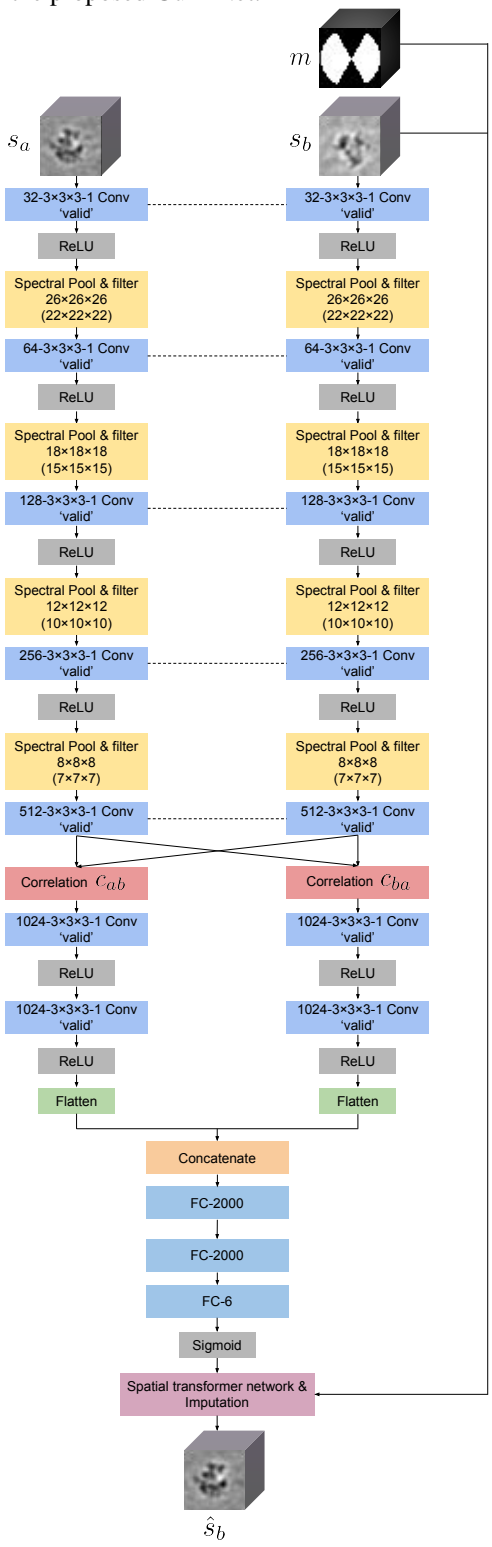


Figure S2: Gum-Net architecture. Each colored box represents one layer. '$32-3\times3\times3-1$ Conv 'valid' denotes a 3D convolutional layer with 32 filters, kernel size $3\times3\times3$, stride 1, and valid padding (no padding). 'Spectral Pool & filter $26\times26\times26$ $(22\times22\times22)$' denotes a spectral pooling & filtering layer with $26\times26\times26$ output size and $22\times22\times22$ cropping size. 'FC-2000' denotes a fully connected layer with 2000 neurons. The dash line denotes the connected layers share weights.

Figure S3 shows the architecture of the Gum-Net MP.

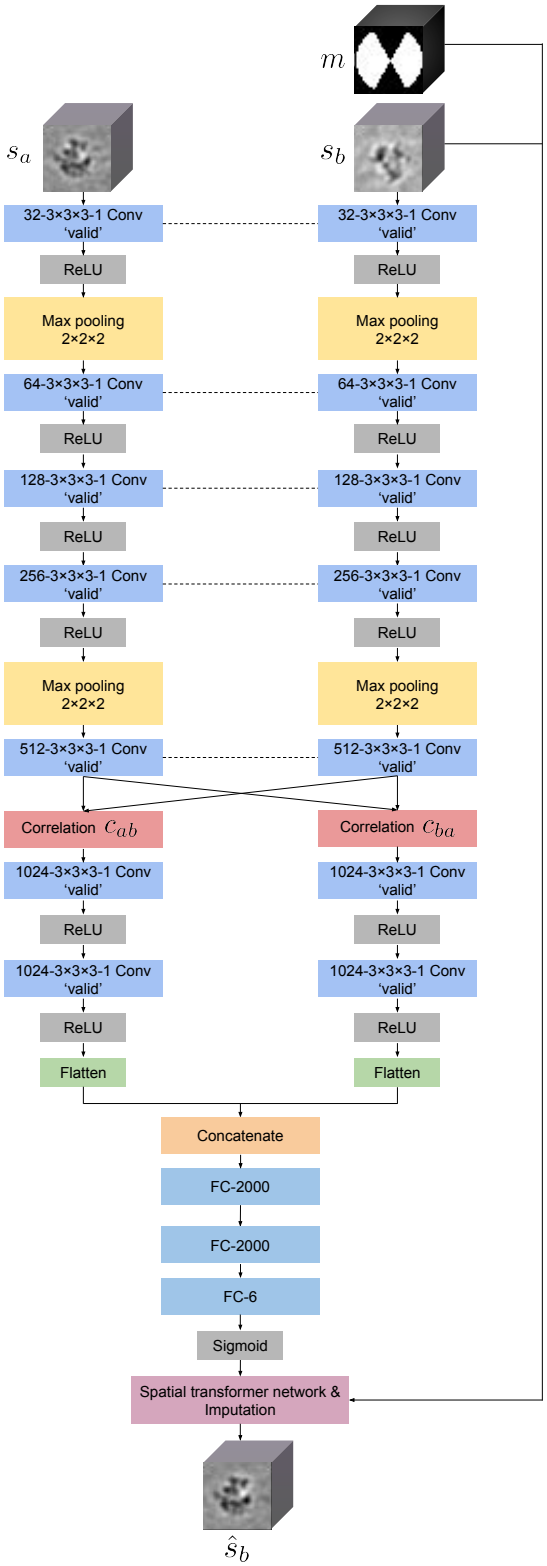


Figure S3: Gum-Net MP architecture. Each colored box represents one layer. '$32-3\times3\times3-1$ Conv 'same" denotes a 3D convolutional layer with same padding (padding to same size as input). 'Max pooling $2\times2\times2$' denotes a max pooling layer with pooling factor $2\times2\times2$.

Figure S4 shows the architecture of the Gum-Net AP.

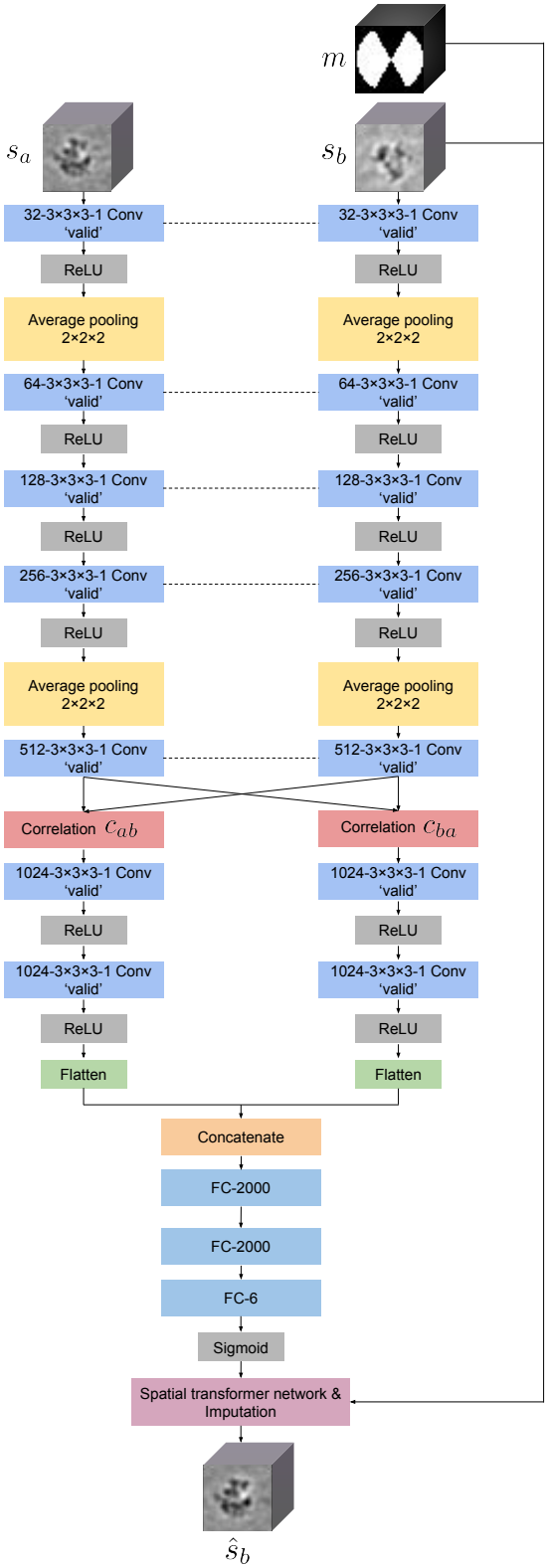


Figure S4: Gum-Net AP architecture. Each colored box represents one layer. 'Average pooling $2 \times 2 \times 2$' denotes an average pooling layer with pooling factor $2 \times 2 \times 2$.

Figure S5 shows the architecture of the Gum-Net SC.

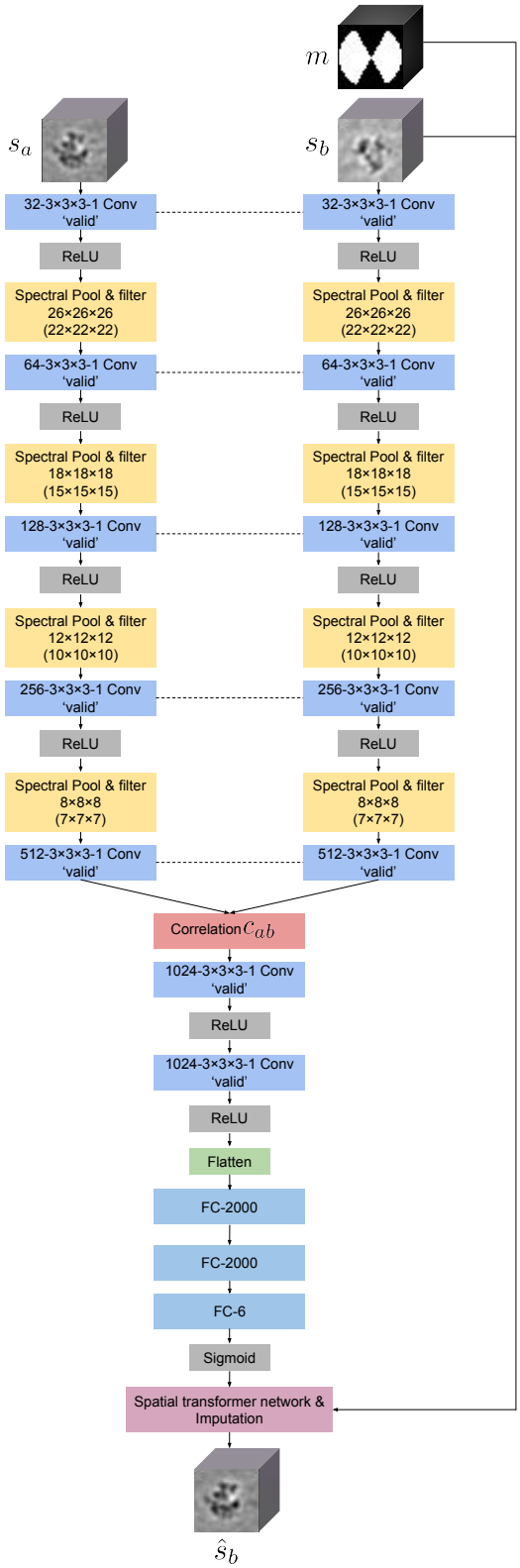


Figure S5: Gum-Net SC architecture. Each colored box represents one layer. Different from Gum-Net, only one correlation map $c_{ab}$ is computed in the matching module.

# S3 Experiments

## S3.1 Example input subtomograms from real datasets

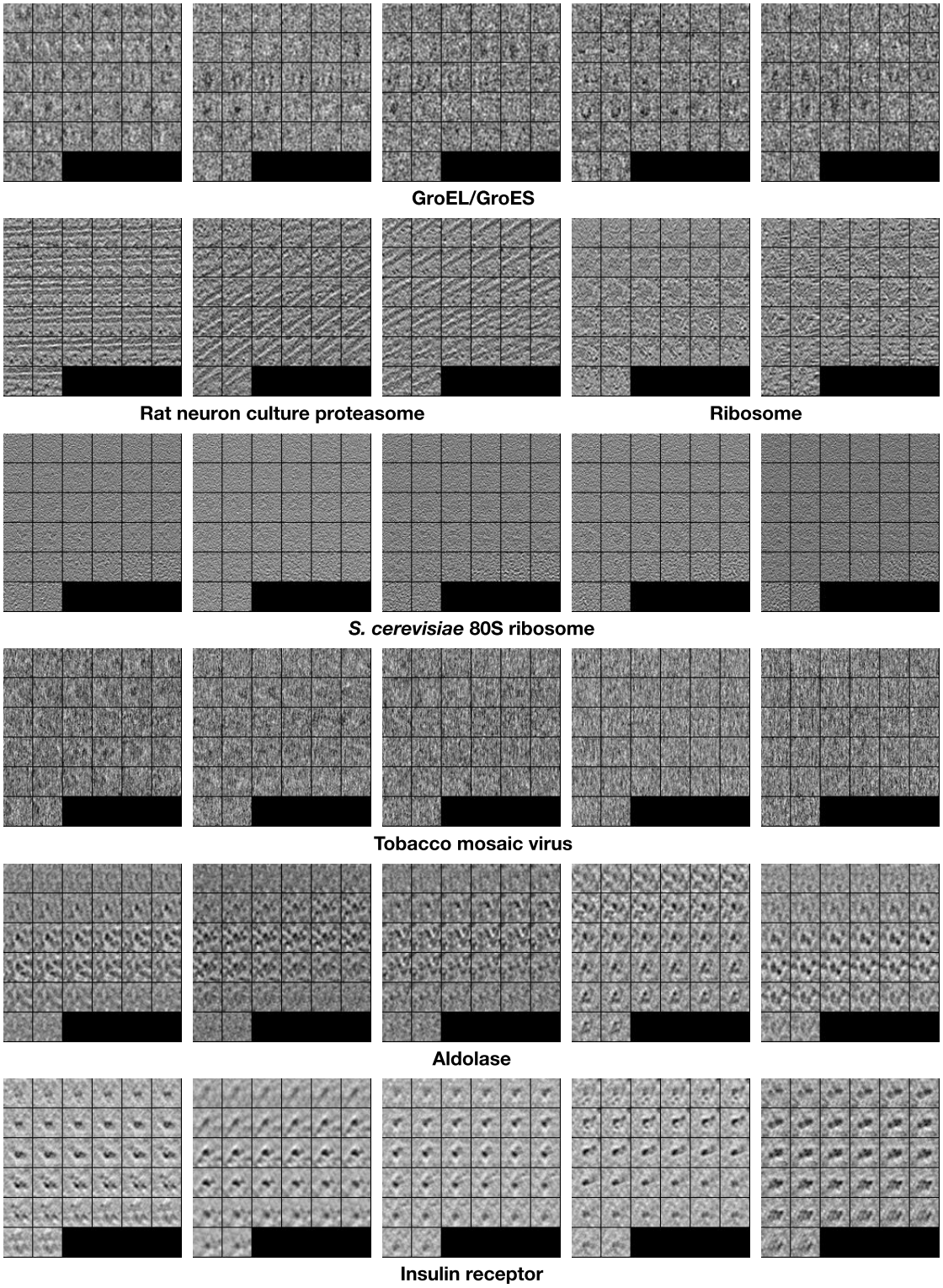


Figure S6: 2D slices representation of five example subtomograms in six real datasets.

### S3.2 Example input subtomograms from simulated datasets

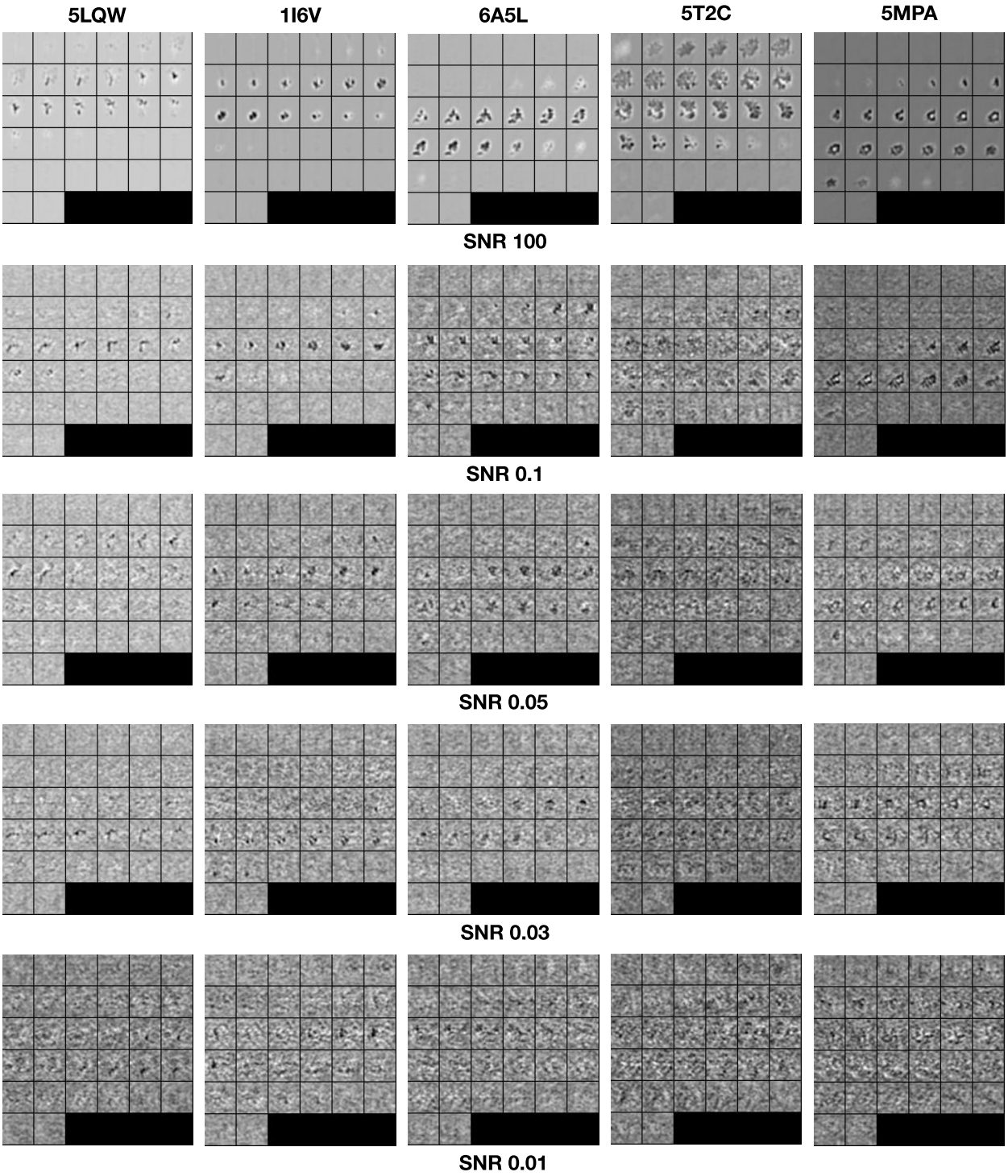


Figure S7: 2D slices representation of five example inputs $s_a$ (no missing wedge) from simulated datasets at different SNRs. There are five structures in each dataset: spliceosome (PDB ID: 5LQW), RNA polymerase-rifampicin complex (1I6V), RNA polymerase II elongation complex (6A5L), ribosome (5T2C), and capped proteasome (5MPA).

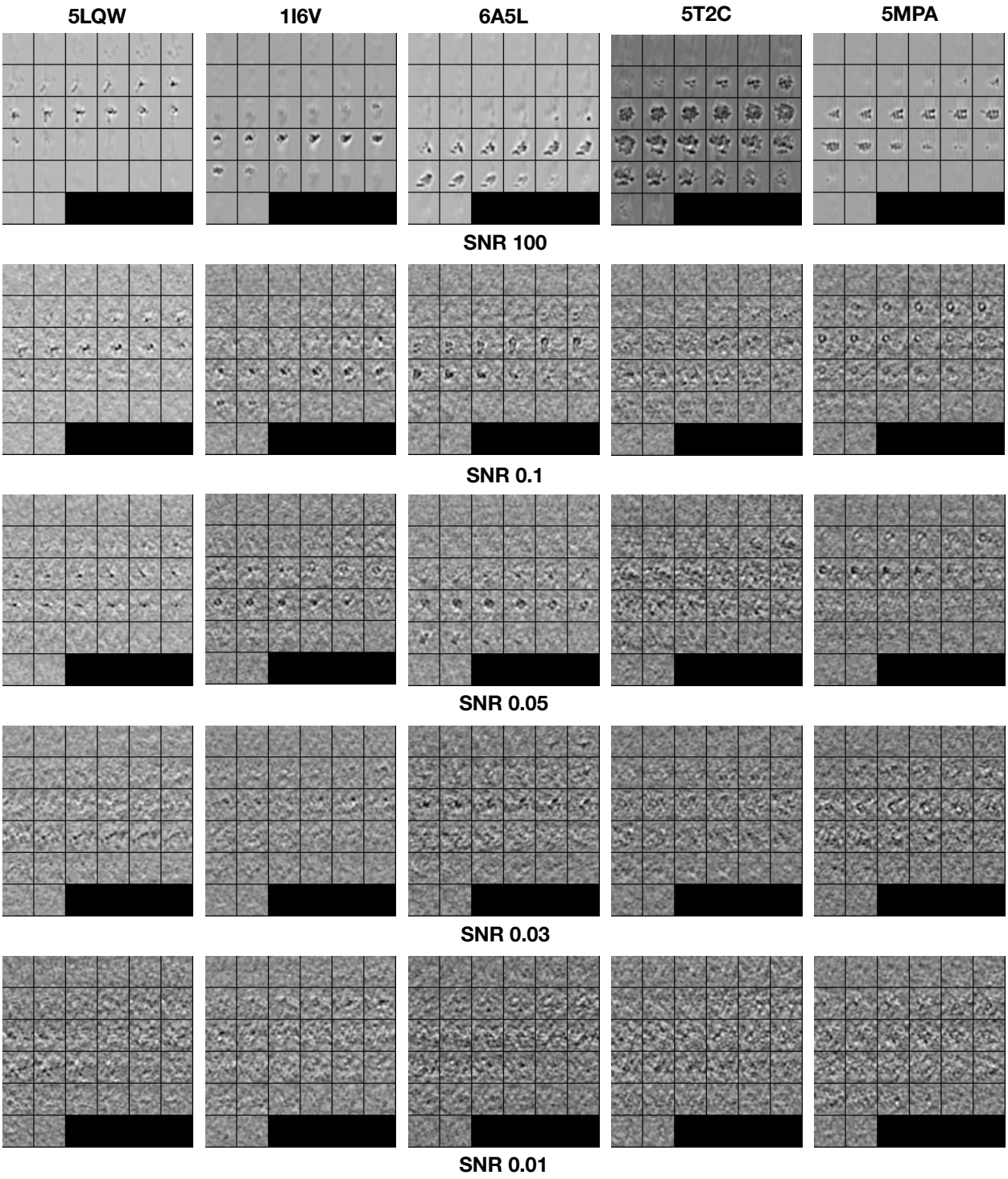


Figure S8: 2D slices representation of five example inputs $s_b$ (30° missing wedge) from simulated datasets at different SNRs. There are five structures in each dataset: spliceosome (PDB ID: 5LQW), RNA polymerase-rifampicin complex (1I6V), RNA polymerase II elongation complex (6A5L), ribosome (5T2C), and capped proteasome (5MPA).

### S3.2.1 Simulation process

To simulate realistic subtomograms, we first downloaded from Protein DataBank [24] the five representative structures: spliceosome (PDB ID: 5LQW), RNA polymerase-rifampicin complex (1I6V), RNA polymerase II elongation complex (6A5L), ribosome (5T2C), and capped proteasome (5MPA). The downloaded .pdb files describe the 3D structure of macromolecular complexes using atomic coordinates, secondary structure assignments, and atomic connectivity. To transform the 3D structure description to 3D volumes of the electron density map representing the structural templates, we applied the *Situs 2.0* PDB2VOL program [25]. The generated 3D volumes of the five structural templates are of size $32^3$ voxels with voxel size 1.2 nm and resolution 1.2 nm. To achieve the target resolution, the atomic structure is convolved by a Gaussian kernel with a standard deviation half of the target resolution.

Given a structural template, we simulate subtomograms by mimicking the experimental condition and process of imaging a cellular specimen and reconstructing a cryo-tomogram. Since the macromolecule complex inside a subtomogram is of random orientation and displacement, we first randomly rotate and translate (up to 7 voxels) the structural template along each of the three axes. Then we project the 3D rotated and translated volume to a series 2D projection images with the specified tilt angle ranges and angular increment of $1^\circ$. For $s_b$ ($30^\circ$ missing wedge), the tilt angle range is set as $\pm 60^\circ$. For $s_a$ (no missing wedge), the tilt angle range is set as $\pm 90^\circ$. Electron optical factors including the contrast transfer function (CTF) and the modulation transfer function (MTF) need to be simulated at this stage. Images in a typical transmission electron microscope are modulated in a spatial frequency-dependent manner described by the CTF [26]:

$$\text{CTF}(f) = A(\sin(\pi\lambda f^2(\Delta z - 0.5\lambda^2 f^2 c_s)) + B\cos(\pi\lambda f^2(\Delta z - 0.5\lambda^2 f^2 c_s))), \tag{1}$$

where $f$ denotes the spatial frequency, $\lambda$ denotes the electron wavelength, $\Delta z$ denotes the defocus, $c_s$ denotes the spherical aberration, $A$ denotes the defocus-dependent envelope function, and $B$ denotes the fraction of amplitude contrast. Simulating the imaging inaccuracies due to CTF is essential for producing realistic simulated cryo-ET data.

Modulation refers to the contrast between bright and dark regions of an image [27]. The MTF, equivalent to the optical transfer function without phase effects, describes how much contrast in the original cellular specimen is maintained by the transmission electron microscopy:

$$\text{MTF}(f) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} l(x) e^{i2\pi f x} \mathrm{d}x, \tag{2}$$

where $f$ denotes the spatial frequency, $l$ denotes the line spread function, and $x$ denotes the spatial distance. Since the spatial frequency content of the specimen is not strictly transferred to the image, taking into account the MTF is vital to cryo-ET data simulation.

We convolved the projection images with the CTF and MTF to obtain the simulated electron micrographic images. The CTF and MTF parameters are similar to the experimental settings with voltage as 300kV, defocus as $-2.0\,\mu\text{m}$, and spherical aberration 2.7mm. Gaussian noise is added to the electron micrographic images to achieve the target signal-to-noise ratio (SNR). Finally, the electron micrographic images are back-projected to obtain the reconstructed subtomogram [28]. All the five datasets, each containing 10500 subtomogram pairs, were simulated in this manner.

### S3.3 Subtomogram alignment results details

The following tables show the alignment accuracy of the five specific structures from manuscript Table 1. At SNR close to experimental condition (0.1 to 0.01), Gum-Net generally outperformed all baselines on aligning all the five structures. The comparison between Gum-Net and the three Gum-Net ablation baselines demonstrated that both the proposed feature extraction module and the Siamese matching module are effective for improving the subtomogram alignment accuracy. Gum-Net performs slightly worse than the three ablation baselines on aligning ribosome structure at SNR 0.01. The potential reason is the training instability at a low SNR. In the future, we will continue to test different training and optimization strategy to reduce the instability.

| Method | SNR 100 | SNR 0.1 | SNR 0.05 | SNR 0.03 | SNR 0.01 |
|---|---|---|---|---|---|
| H-T align | 0.06±0.02, 1.03±0.63 | 0.61±0.87, 2.64±3.55 | 1.62±1.14 6.08±4.92 | 2.15±0.88, 8.49±4.72 | 2.38±0.56, 11.36±5.13 |
| F&A align | 0.08±0.13, 1.09±1.14 | 0.64±0.97, 2.96±3.99 | 1.68±1.16, 6.32±4.91 | 2.12±0.89, 8.39±4.79 | 2.35±0.59, 11.2±5.00 |
| Gum-Net MP | 0.54±0.61, 2.18±2.53 | 1.02±0.70, 4.07±3.16 | 1.25±0.78, 4.89±3.30 | 1.38±0.75, 5.41±3.31 | 1.65±0.78, 6.79±3.08 |
| Gum-Net AP | 0.39±0.54, 1.67±2.22 | 0.87±0.65, 3.56±2.78 | 1.12±0.74, 4.45±3.00 | 1.29±0.74, 5.07±3.09 | 1.60±0.81, 6.69±3.11 |
| Gum-Net SC | 0.51±0.59, 2.02±2.43 | 0.96±0.71, 3.83±3.13 | 1.22±0.79, 4.76±3.28 | 1.38±0.76, 5.28±3.33 | 1.65±0.78, 6.82±3.20 |
| Gum-Net | 0.27±0.54, 1.13±2.03 | **0.47±0.57, 1.94±2.26** | **0.68±0.64, 2.61±2.25** | **0.93±0.68, 3.62±2.32** | **1.38±0.78, 5.65±3.31** |

Table S1: Spliceosome (5LQW) subtomogram alignment accuracy on five datasets with SNR specified. In each cell, the first term is the mean and standard deviation of the rotation error and the second term, the translation error. We highlighted Gum-Net results that are significantly better ($p < 0.001$) than all baselines by the paired sample t-test.

| Method | SNR 100 | SNR 0.1 | SNR 0.05 | SNR 0.03 | SNR 0.01 |
|---|---|---|---|---|---|
| H-T align | 0.63±0.99, 3.15±4.27 | 1.67±1.06, 6.31±5.01 | 2.09±0.87, 7.65±4.56 | 2.22±0.74, 8.10±4.43 | 2.40±0.57, 10.93±4.97 |
| F&A align | 0.67±1.00, 3.22±4.24 | 1.71±1.08, 6.63±4.96 | 2.06±0.90, 7.76±4.67 | 2.23±0.74, 8.48±4.62 | 2.37±0.56, 10.94±4.98 |
| Gum-Net MP | 0.92±0.78, 3.53±3.31 | 1.38±0.75, 5.25±3.53 | 1.50±0.76, 5.70±3.65 | 1.59±0.76, 6.08±3.54 | 1.66±0.77, 7.06±3.39 |
| Gum-Net AP | 0.83±0.79, 3.22±3.25 | 1.25±0.76, 4.75±3.37 | 1.39±0.76, 5.35±3.49 | 1.53±0.75, 5.81±3.46 | 1.65±0.77, 7.02±3.35 |
| Gum-Net SC | 0.90±0.80, 3.39±3.27 | 1.26±0.77, 4.83±3.58 | 1.42±0.77, 5.43±3.62 | 1.53±0.76, 5.73±3.47 | 1.68±0.76, 6.96±3.52 |
| Gum-Net | **0.56±0.78, 2.22±3.05** | **0.75±0.77, 2.99±3.17** | **0.87±0.76, 3.49±3.31** | **1.05±0.71, 3.96±2.77** | **1.42±0.78, 5.66±3.53** |

Table S2: RNA polymerase-rifampicin complex (1I6V) subtomogram alignment accuracy on five datasets with SNR specified.

| Method | SNR 100 | SNR 0.1 | SNR 0.05 | SNR 0.03 | SNR 0.01 |
|---|---|---|---|---|---|
| H-T align | 0.09±0.10, 1.11±0.82 | 0.94±0.95, 3.75±4.03 | 1.74±1.02, 6.31±4.60 | 2.21±0.75, 8.69±4.56 | 2.37±0.55, 11.58±5.02 |
| F&A align | 0.16±0.34, 1.31±1.62 | 1.06±1.06, 4.31±4.41 | 1.85±0.99, 6.99±4.85 | 2.18±0.79, 8.69±4.55 | 2.39±0.58, 11.31±4.83 |
| Gum-Net MP | 0.66±0.69, 2.52±2.73 | 1.13±0.74, 4.27±3.09 | 1.30±0.75, 4.80±3.11 | 1.45±0.76, 5.45±3.09 | 1.66±0.77, 6.99±3.28 |
| Gum-Net AP | 0.48±0.58, 1.83±2.00 | 0.98±0.67, 3.72±2.74 | 1.20±0.72, 4.45±2.85 | 1.40±0.74, 5.29±3.02 | 1.64±0.77, 6.97±3.33 |
| Gum-Net SC | 0.60±0.64, 2.24±2.33 | 1.07±0.73, 4.02±3.03 | 1.26±0.76, 4.56±3.07 | 1.47±0.77, 5.48±3.14 | 1.65±0.76, 6.89±3.33 |
| Gum-Net | 0.30±0.55, 1.08±1.71 | **0.46±0.54, 1.80±1.90** | **0.71±0.63, 2.55±2.12** | **1.12±0.73, 3.93±2.45** | **1.45±0.76, 5.94±3.32** |

Table S3: RNA polymerase II elongation complex (6A5L) subtomogram alignment accuracy on five datasets with SNR specified.

| Method | SNR 100 | SNR 0.1 | SNR 0.05 | SNR 0.03 | SNR 0.01 |
|---|---|---|---|---|---|
| H-T align | 0.06±0.02, 0.99±0.60 | 1.16±1.04, 4.43±4.21 | 2.13±0.84, 8.79±4.77 | 2.34±0.61, 10.59±4.98 | 2.36±0.59, 11.56±4.91 |
| F&A align | 0.05±0.03, 0.98±0.61 | 1.54±1.12, 6.39±5.19 | 2.17±0.80, 9.39±5.09 | 2.35±0.58, 10.81±4.93 | 2.40±0.55, 11.81±4.89 |
| Gum-Net MP | 1.31±1.10, 4.49±4.18 | 1.58±0.83, 5.51±3.07 | 1.71±0.80, 6.28±3.16 | 1.70±0.80, 6.72±3.13 | 1.70±0.78, 8.27±3.58 |
| Gum-Net AP | 0.64±0.9, 2.36±3.22 | 1.30±0.79, 4.71±2.76 | 1.58±0.80, 5.94±3.05 | 1.63±0.81, 6.70±3.20 | 1.68±0.78, 8.14±3.51 |
| Gum-Net SC | 0.77±0.93, 2.73±3.37 | 1.41±0.79, 4.90±2.94 | 1.63±0.79, 5.98±3.11 | 1.66±0.80, 6.54±3.15 | 1.71±0.77, 8.35±3.64 |
| Gum-Net | 0.43±0.87, 1.67±3.31 | **0.73±0.81, 2.70±2.87** | **1.19±0.84, 4.23±3.01** | **1.43±0.79, 5.67±2.96** | 1.76±0.75, 10.46±5.10 |

Table S4: Ribosome (5T2C) subtomogram alignment accuracy on five datasets with SNR specified.

| Method | SNR 100 | SNR 0.1 | SNR 0.05 | SNR 0.03 | SNR 0.01 |
|---|---|---|---|---|---|
| H-T align | 0.65±0.95, 2.81±3.44 | 1.72±0.99, 6.65±4.55 | 2.08±0.88, 7.47±4.46 | 2.16±0.81, 8.42±4.47 | 2.38±0.58, 11.22±5.03 |
| F&A align | 0.69±0.97, 3.02±3.72 | 1.73±1.01, 6.69±4.71 | 1.97±0.94, 7.26±4.67 | 2.24±0.79, 8.59±4.69 | 2.39±0.56, 11.33±4.88 |
| Gum-Net MP | 1.08±0.85, 3.98±3.48 | 1.40±0.80, 5.52±3.60 | 1.43±0.78, 5.63±3.44 | 1.53±0.76, 6.12±3.45 | 1.68±0.77, 7.30±3.33 |
| Gum-Net AP | 0.66±0.61, 2.51±2.31 | 1.05±0.69, 4.28±2.92 | 1.19±0.73, 4.78±3.04 | 1.37±0.73, 5.64±3.22 | 1.66±0.77, 7.10±3.27 |
| Gum-Net SC | 0.74±0.66, 2.77±2.53 | 1.12±0.76, 4.47±3.30 | 1.24±0.78, 4.92±3.40 | 1.38±0.77, 5.71±3.43 | 1.66±0.78, 7.16±3.35 |
| Gum-Net | **0.48±0.67, 1.86±2.53** | **0.68±0.64, 2.61±2.46** | **0.89±0.72, 3.13±2.68** | **1.12±0.72, 4.25±2.73** | **1.46±0.78, 6.22±3.38** |

Table S5: Capped proteasome (5MPA) subtomogram alignment accuracy on five datasets with SNR specified.

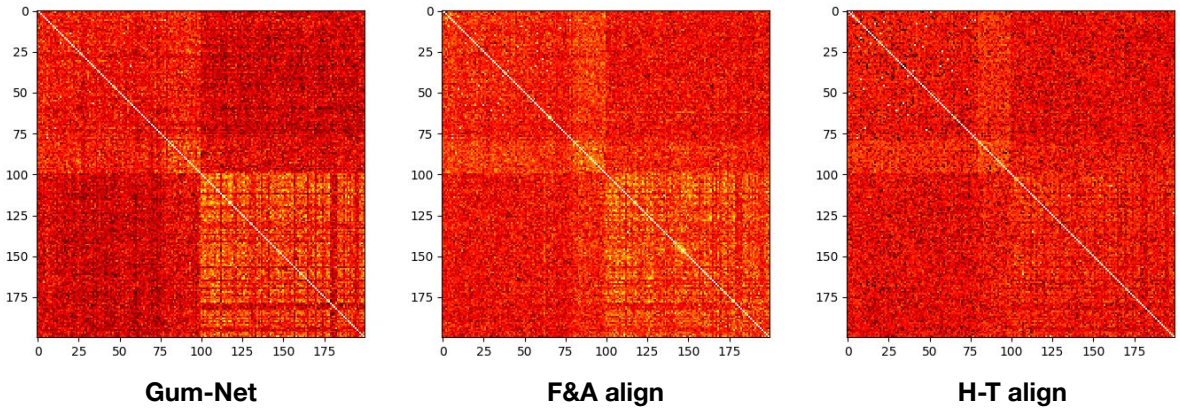


Figure S9: Heatmaps pf the pairwise subtomogram alignment correlation matrix for the rat neuron culture dataset by three methods.

We use heatmaps to visualize the pairwise subtomogram alignment experimental results (manuscript Section 4.3) from the rat neuron culture dataset. The first 100 subtomograms contain ribosomes and the rest 100 subtomograms contain capped proteasomes. Clearly, the correlation matrix computed by Gum-Net shows better clustering patterns, which results in a 92% accuracy applying the complete-linkage hierarchical clustering algorithm with $k = 2$. In comparison, F&A align has a much lower accuracy of 65% and H-T align, 53.5%.

### S3.4 2D slices representation

The 2D slices representation of the structures in manuscript Figure 3 and 4 are plotted here for a more detailed view.

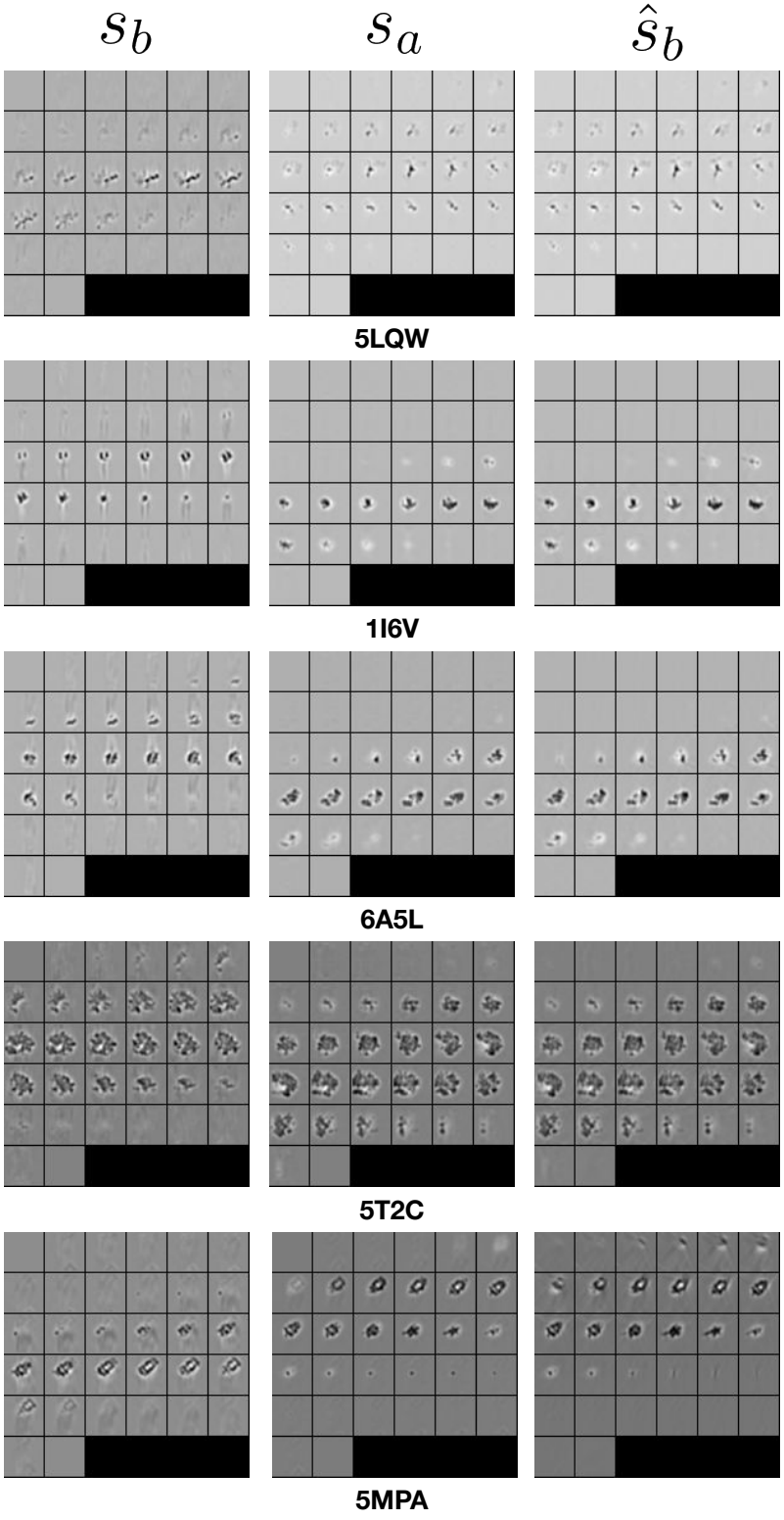


Figure S10: 2D slices representation of the structures in manuscript Figure 5.

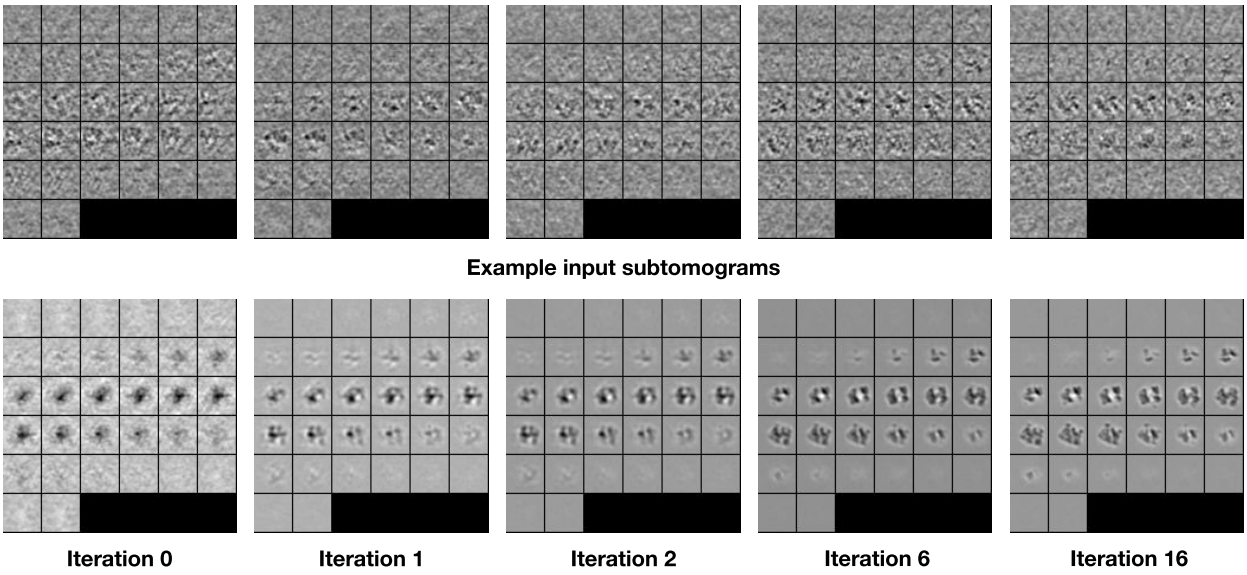


Figure S11: 2D slices representation of the structures in manuscript Figure 4.

### S3.5 Mathematical definition of metrics

#### S3.5.1 SNR

Subtomogram signal-to-noise ratio (SNR) is defined in Equation 3.61 in [29]. When two subtomogram containing the same structure at a certain SNR are optimally aligned, we can calculate the Pearson's correlation $c$ between them. The SNR is defined as:

$$\text{SNR} = \frac{c}{1-c} \tag{3}$$

The higher the SNR, the better the signal. When there is no noise, the expected Pearson's correlation $c$ will be equal to 1 and the SNR will reach infinity. Where the subtomograms are of pure noise, the expected Pearson's correlation $c$ will be equal to zero and the SNR will be zero.

#### S3.5.2 FSC

Fourier shell correlation (FSC) is defined as [30]:

$$\text{FSC}(r) = \frac{\sum_{r_i \in r} F_1(r_i) \cdot F_2(r_i)^*}{\sqrt{\sum_{r_i \in r} |F_1(r_i)|^2 \cdot \sum_{r_i \in r} |F_2(r_i)|^2}} \tag{4}$$

where $F_1$ is the complex structure factor for subtomogram average 1, $F_2^*$ is the complex conjugate of the structure factor for subtomogram average 2, and $r_i$ is the individual voxel element at radius $r$ of the corresponding shell in Fourier space. We chose the standard FSC 0.143 cutoff as the resolution value, which measures the $r$ corresponds to the correlation coefficient equal to 0.143 [31, 32]. For simulated data, the subtomogram average is compared with the noise-free structural template used to generate the simulated subtomograms. For real data, we do not have an accurate structural template because even if the structure is known, it may appear differently in the real data due to structural dynamics, different species, and different experimental conditions. Therefore, we use the gold-standard procedure [33] by dividing the real subtomograms into two independent sets, averaging them separately, and comparing subtomogram average 1 and subtomogram average 2 using the Fourier shell correlation.

## References


[1] Martin Beck and Wolfgang Baumeister. Cryo-electron tomography: can it reveal the molecular sociology of cells in atomic detail? *Trends in cell biology*, 26(11):825–837, 2016.

[2] Xiao-Chen Bai, Greg McMullan, and Sjors HW Scheres. How cryo-em is revolutionizing structural biology. *Trends in biochemical sciences*, 40(1):49–57, 2015.

[3] Roman I Koning, Abraham J Koster, and Thomas H Sharp. Advances in cryo-electron tomography for biology and medicine. *Annals of Anatomy-Anatomischer Anzeiger*, 217:82–96, 2018.

[4] Jianfeng Lin, Weining Yin, Maria C Smith, Kangkang Song, Margaret W Leigh, Maimoona A Zariwala, Michael R Knowles, Lawrence E Ostrowski, and Daniela Nicastro. Cryo-electron tomography reveals ciliary defects underlying human rsph1 primary ciliary dyskinesia. *Nature communications*, 5:5727, 2014.

[5] Rui Wang, Rebecca L Stone, Jason T Kaelber, Ryan H Rochat, Alpa M Nick, K Vinod Vijayan, Vahid Afshar-Kharghan, Michael F Schmid, Jing-Fei Dong, Anil K Sood, et al. Electron cryotomography reveals ultrastructure alterations in platelets from patients with ovarian cancer. *Proceedings of the National Academy of Sciences*, 112(46):14266–14271, 2015.

[6] Stephanie E Siegmund, Robert Grassucci, Stephen D Carter, Emanuele Barca, Zachary J Farino, Martí Juanola-Falgarona, Peijun Zhang, Kurenai Tanji, Michio Hirano, Eric A Schon, et al. Three-dimensional analysis of mitochondrial crista ultrastructure in a patient with leigh syndrome by in situ cryoelectron tomography. *iScience*, 6:83–91, 2018.

[7] FJB Bäuerlein, A Mishra, I Dudanova, M Hipp, R Klein, FU Hartl, W Baumeister, R Fernández-Busnadiego, et al. Structural characterization of mutant huntingtin inclusion bodies by cryo-electron tomography. *Microscopy and Microanalysis*, 22(S3):80–81, 2016.

[8] Sheng Cao, José O Maldonado, Iwen F Grigsby, Louis M Mansky, and Wei Zhang. Analysis of human t-cell leukemia virus type 1 particles by using cryo-electron tomography. *Journal of virology*, 89(4):2430–2435, 2015.

[9] Louie D Henderson and Morgan Beeby. High-throughput electron cryo-tomography of protein complexes and their assembly. In *Protein Complex Assembly*, pages 29–44. Springer, 2018.

[10] Ardan Patwardhan. Trends in the electron microscopy data bank (emdb). *Acta Crystallographica Section D: Structural Biology*, 73(6):503–508, 2017.

[11] Andrii Iudin, Paul K Korir, José Salavert-Torres, Gerard J Kleywegt, and Ardan Patwardhan. Empiar: a public archive for raw electron microscopy image data. *Nature methods*, 13(5):387, 2016.

[12] Min Xu, Xiaoqi Chai, Hariank Muthakana, Xiaodan Liang, Ge Yang, Tzviya Zeev-Ben-Mordehai, and Eric P Xing. Deep learning-based subdivision approach for large scale macromolecules structure recovery from electron cryo tomograms. *Bioinformatics*, 33(14):i13–i22, 2017.

[13] Muyuan Chen, Wei Dai, Stella Y Sun, Darius Jonasch, Cynthia Y He, Michael F Schmid, Wah Chiu, and Steven J Ludtke. Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. *nature methods*, 14(10):983, 2017.

[14] Emmanuel Moebel, Antonio Martinez, Damien Larivière, Julio Ortiz, Wolfgang Baumeister, and Charles Kervrann. 3d convnet improves macromolecule localization in 3d cellular cryo-electron tomograms. 2018.

[15] Xiangrui Zeng, Miguel Ricardo Leung, Tzviya Zeev-Ben-Mordehai, and Min Xu. A convolutional autoencoder approach for mining features in cellular electron cryo-tomograms and weakly supervised coarse segmentation. *Journal of structural biology*, 202(2):150–160, 2018.

[16] Lubomír Kováčik, Sami Kerïeche, Johanna L Höög, Pavel Jda, Pavel Matula, and Ivan Raška. A simple fourier filter for suppression of the missing wedge ray artefacts in single-axis electron tomographic reconstructions. *Journal of structural biology*, 186(1):141–152, 2014.

[17] Min Xu, Martin Beck, and Frank Alber. High-throughput subtomogram alignment and classification by fourier space constrained fast volumetric matching. *Journal of structural biology*, 178(2):152–164, 2012.

[18] Yuxiang Chen, Stefan Pfeffer, Thomas Hrabe, Jan Michael Schuller, and Friedrich Förster. Fast and accurate reference-free alignment of subtomograms. *Journal of structural biology*, 182(3):235–245, 2013.

[19] Min Xu, Jitin Singla, Elitza I Tocheva, Yi-Wei Chang, Raymond C Stevens, Grant J Jensen, and Frank Alber. De novo structural pattern mining in cellular electron cryotomograms. *Structure*, 2019.

[20] Ignacio Rocco, Relja Arandjelovic, and Josef Sivic. Convolutional neural network architecture for geometric matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6148–6157, 2017.

[21] Ignacio Rocco, Relja Arandjelović, and Josef Sivic. End-to-end weakly-supervised semantic alignment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6917–6925, 2018.

[22] Jiahuan Zhou, Weiqi Xu, and Ryad Chellali. Analysing the effects of pooling combinations on invariance to position and deformation in convolutional neural networks. In *2017 IEEE International Conference on Cyborg and Bionic Systems (CBS)*, pages 226–230. IEEE, 2017.

[23] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[24] Helen M Berman, Philip E Bourne, John Westbrook, and Christine Zardecki. The protein data bank. In *Protein Structure*, pages 394–410. CRC Press, 2003.

[25] Pablo Chacón and Willy Wriggers. Multi-resolution contour-based fitting of macromolecular structures. *Journal of molecular biology*, 317(3):375–384, 2002.

[26] Giulia Zanetti, James D Riches, Stephen D Fuller, and John AG Briggs. Contrast transfer function correction applied to cryo-electron tomography and sub-tomogram averaging. *Journal of structural biology*, 168(2):305–312, 2009.

[27] Syed Naeem Ahmed. 7 - position-sensitive detection and imaging. In Syed Naeem Ahmed, editor, *Physics and Engineering of Radiation Detection (Second Edition)*, pages 435 – 475. Elsevier, second edition edition, 2015.

[28] J Bernard Heymann, Giovanni Cardone, Dennis C Winkler, and Alasdair C Steven. Computational resources for cryo-electron tomography in bsoft. *Journal of structural biology*, 161(3):232–242, 2008.

[29] Joachim Frank. *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state*. Oxford University Press, 2006.

[30] Marin Van Heel and Michael Schatz. Fourier shell correlation threshold criteria. *Journal of structural biology*, 151(3):250–262, 2005.

[31] Florian KM Schur, Wim JH Hagen, Alex De Marco, and John AG Briggs. Determination of protein structure at 8.5 å resolution using cryo-electron tomography and sub-tomogram averaging. *Journal of structural biology*, 184(3):394–400, 2013.

[32] Thomas H Sharp, Abraham J Koster, and Piet Gros. Heterogeneous mac initiator and pore structures in a lipid bilayer by phase-plate cryo-electron tomography. *Cell reports*, 15(1):1–8, 2016.

[33] Sjors HW Scheres and Shaoxia Chen. Prevention of overfitting in cryo-em structure determination. *Nature methods*, 9(9):853, 2012.