1. Brief plan

The basic plan of the method development of the approach was to segment the particle from the micrograph's images provided in the cryoppp_lite dataset utilizing the particle coordinate from the star file and extract the particle image. Then use these images to train a model that outputs the predicted electron density map.

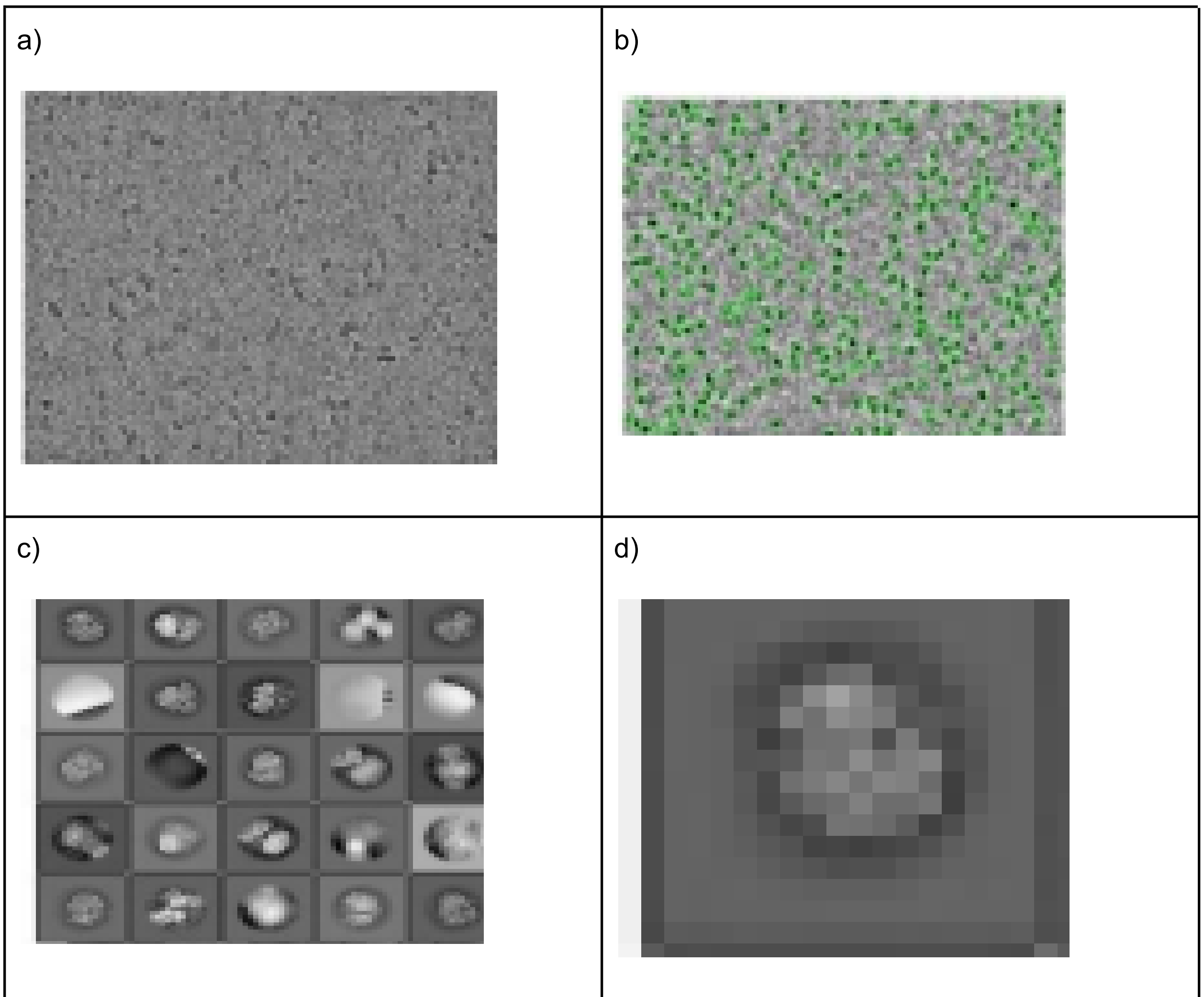


FIgure 1: a) A micrograph b) Particles are highlighted in green c) Many of the picked particles are displayed d) Showing one of the many picked particles. All these images are used to

illustrate the idea.

2. **Input**

The segmented particle images were preprocessed using various image processing techniques. Firstly, standardization was applied where the values of the images were subtracted by the mean and divided by the standard deviation. Secondly, contrast-limited adaptive histogram equalization (CLAHE) was used to improve the contrast and enhance the details in images. Lastly, the guided filter was employed to further enhance details and decrease noise. The final output of various particle images are displayed in Figure 2. In addition, clustering was also applied using resnet-152 during the early phases of experimentation to select diverse input data.

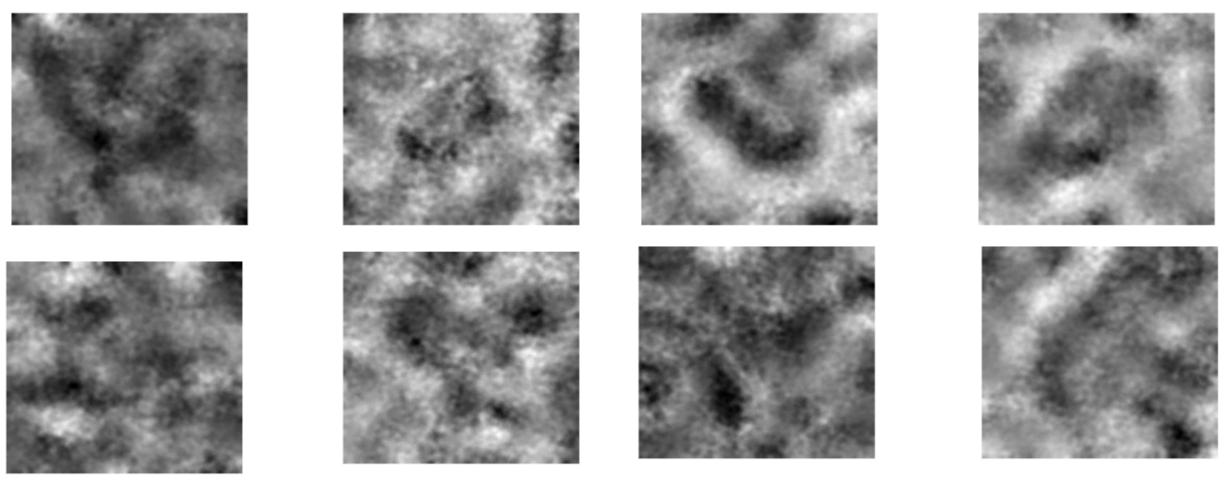
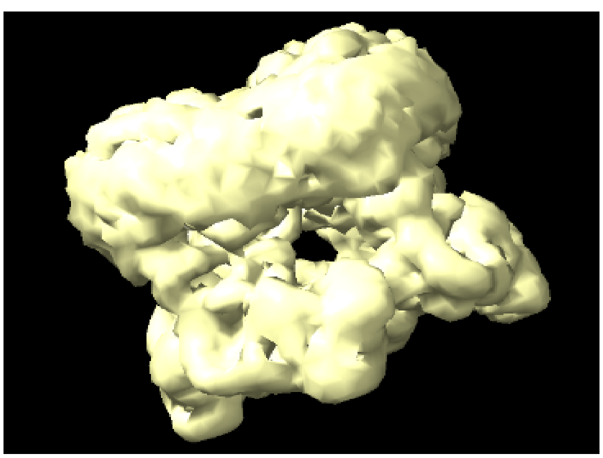


Figure 2: Displaying the final output of the image preprocessing techniques on multiple particles.

3. Labels

The voxel representation of the electron density maps were used as labels. At first, the map's dimension was reduced to 256x256x256 and each grid had a voxel size of 1. Later on, it was decreased to a dimension of 128x128x128. Then the map is normalized, so that all the voxel values are between 0 and 1. To summarize the normalization process, the whole 3d-array was divided by 95 percentile value, and then values greater than 1 were converted to 1 and negative values were all converted to 0, this is the same approach as cryo2struct. Also, I conduct an experiment by thresholding the label by converting all values below 0.9 to 0 and greater than or equal to 0.9 to 1. Thresholding does not affect the structure rather filters out noises from the map and also helps to get lower losses during training.

a)



b)

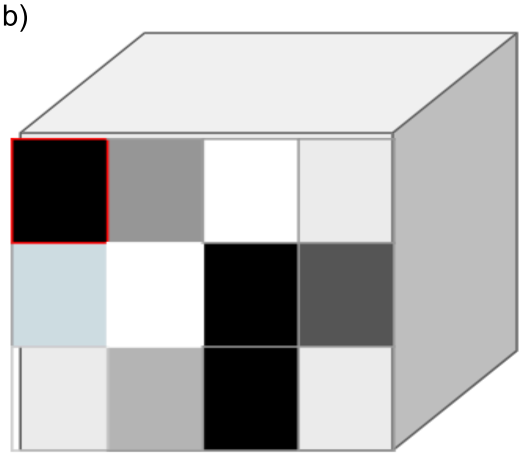


Figure 3: a) Displaying the target 10005 on the left b) 3d array that is representing a map in a 3d-array

4. Approaches

4.1. Only image and U-net approach

The first approach used an U-net architecture as displayed in Figure-4 where the label's dimension was 256x256x256. The input for this approach was a stack of 256 randomly selected images for each target.

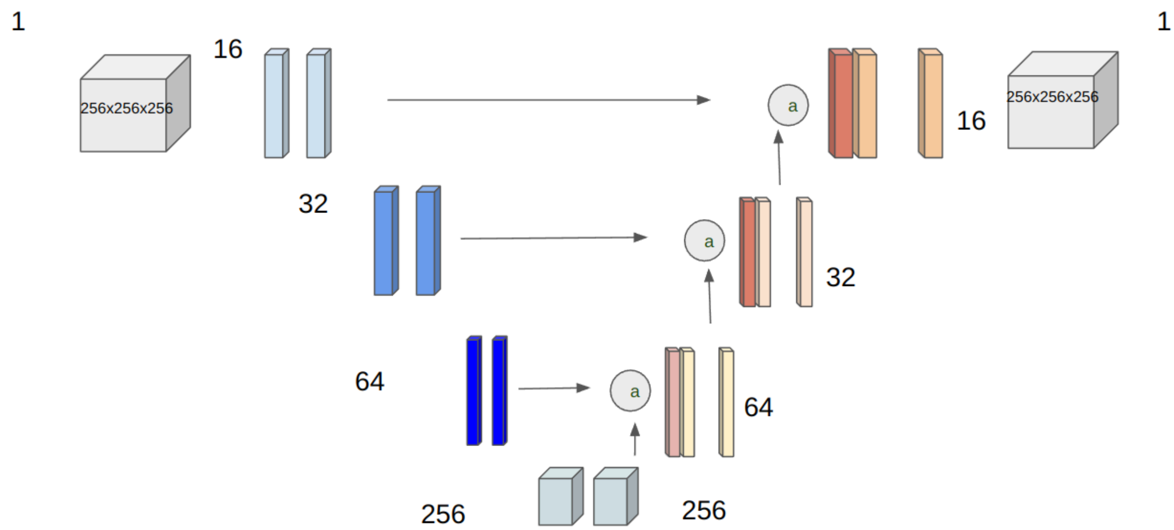


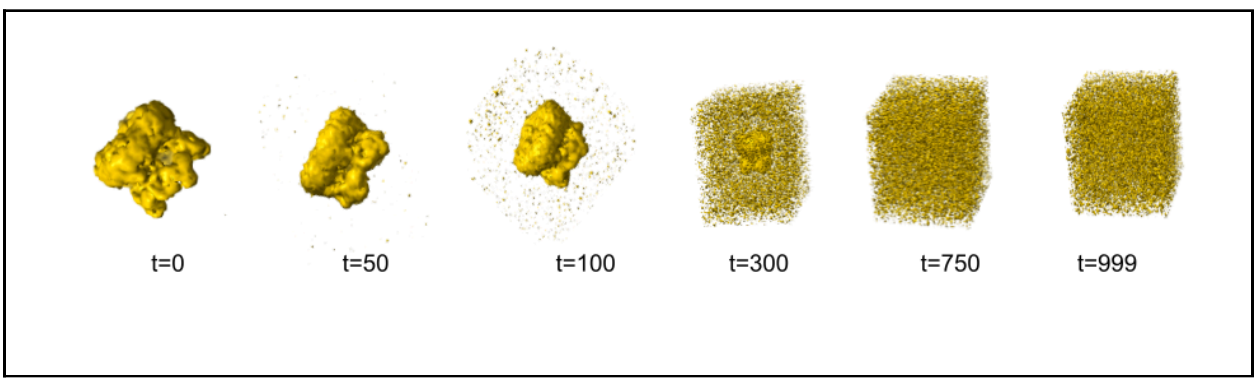
Figure 4: Displaying the architecture of an U-net where a denotes attention mechanism. The first layer has 16 convolutional layers and then 32 layers and followed by 64 layers and finally a bottleneck with 256 dense layers. The decoder network has the number of convolutional layers but in reverse orders and finally the output would be a tensor of 256x256x256. Relu was used as an activation function in all the layers except for the last layer where sigmoid was used.

4.2. Diffusion and Stack of Images

The second approach was using a diffusion model. In this approach noisy images are fed to the network and the model is trained to learn the noises in the input. So there are 2 steps, the forward diffusion process, and the reverse diffusion process. In the forward process, noises are gradually added to the image and in the reverse step starts with a randomly initialized noise which is then fed n times to the network to predict noises and denoise the structure as shown in Figure 5.

To train the model at first the input images are processed using the preprocessing technique used above and then a stack of randomly selected 256 images is selected. Then the stack of the images was concatenated with the noisy structure, the noise level of the noisy structure would depend on the timestep. It is noteworthy that for all the DISPR network mse loss and variational lower bound loss is used.

Later on, the code-base was replaced with the code-based used in the DISPR paper to increase the reliability of the code base as previous approaches were not yielding any results. The DISPR network also used a similar architecture as Figure-4, the only difference is they had more residual blocks on the downsampling and upsampling blocks.



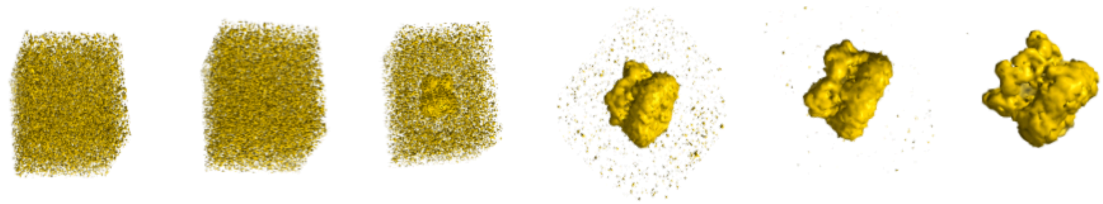


Figure 5. Displaying the forward and backward process of the diffusion model.

4.3. Using Cryosparc 2-D classes as input.

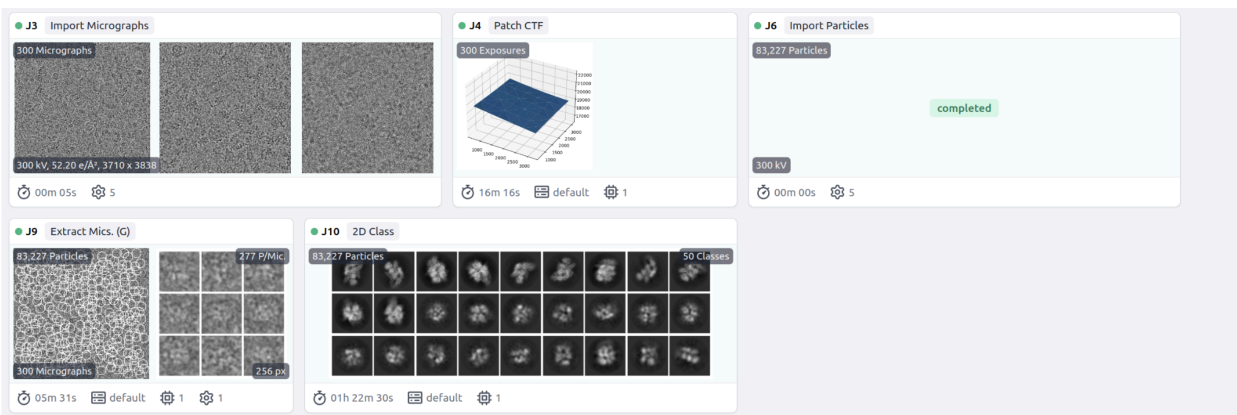


Figure 6. To generate the 2D classes the micrographs are imported, then patch CTF is applied followed by importing of the particles then the extract particle steps, and finally, the 2d-class is generated.

To simplify the problem with noisy images, the 2D classes were used that were generated using the cryosparc. Furthermore, the cryosparc-cli was interfaced with Python so that it could be automated. The general process to generate 2d-class is shown in Figure 6. Then based on the resolution top-20 class images were selected. Then using the best class images inputs were prepared in 3 different ways as shown in Figure 7. Then the prepared input is fed into the DISPR network to generate the final output which is the electron density map.

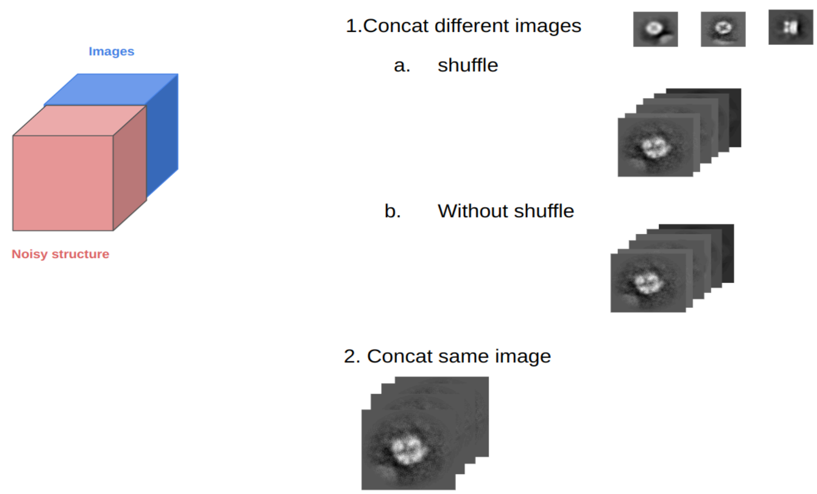


Figure 7: Displaying the input preparation with the 2d classes. 1. a. Uses the stack of the best class images with shuffling and 1.b. Uses the best images without shuffling and 2. Uses one of the top-20 images which is stacked with the same image as used in DISPR. In all the approaches the image stack is shaped to have a dimension of 128x128x128 and then is concatenated with the noisy structure that had a dimension of 128x128x128 as a result the total size of the input is 2x128x128x128. The network is illustrated in Figure 7.

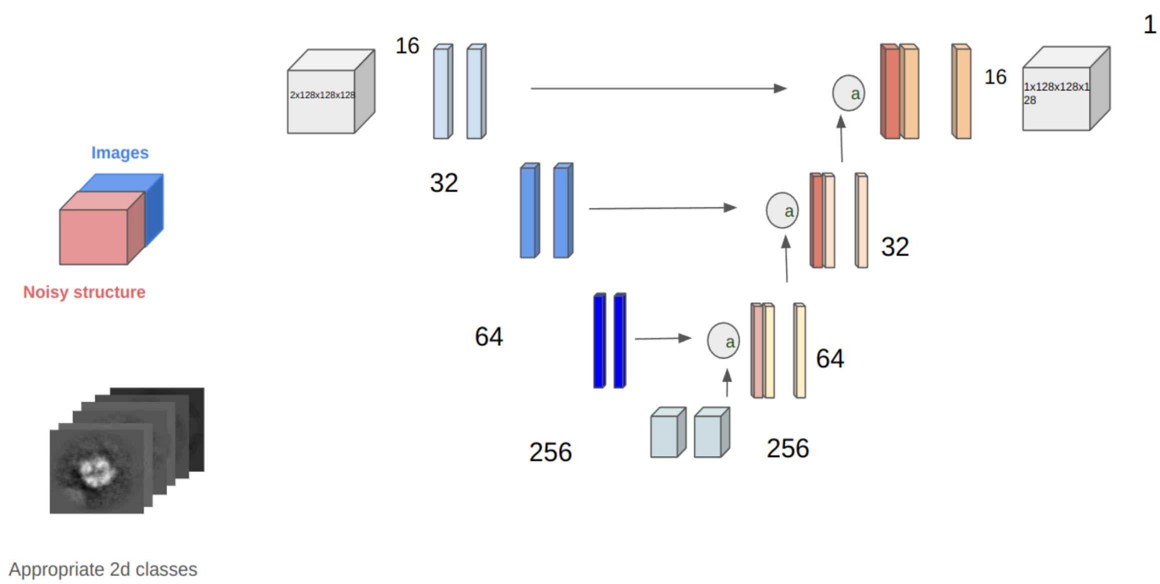


Figure 8. Displaying the cryosparc 2d-class input and the dispr network.

4.4. Using particle stack

With the particle stack, the inputs were prepared using the help of an autoencoder. At first autoencoder models were trained separately for each target and after the training the encoder was used to generate the 8-dim latent encodings as shown in Figure 9. Then all the encodings are stacked and reshaped to a tensor of 128x128x128 which is then fed into the DISPR network like the previous method.

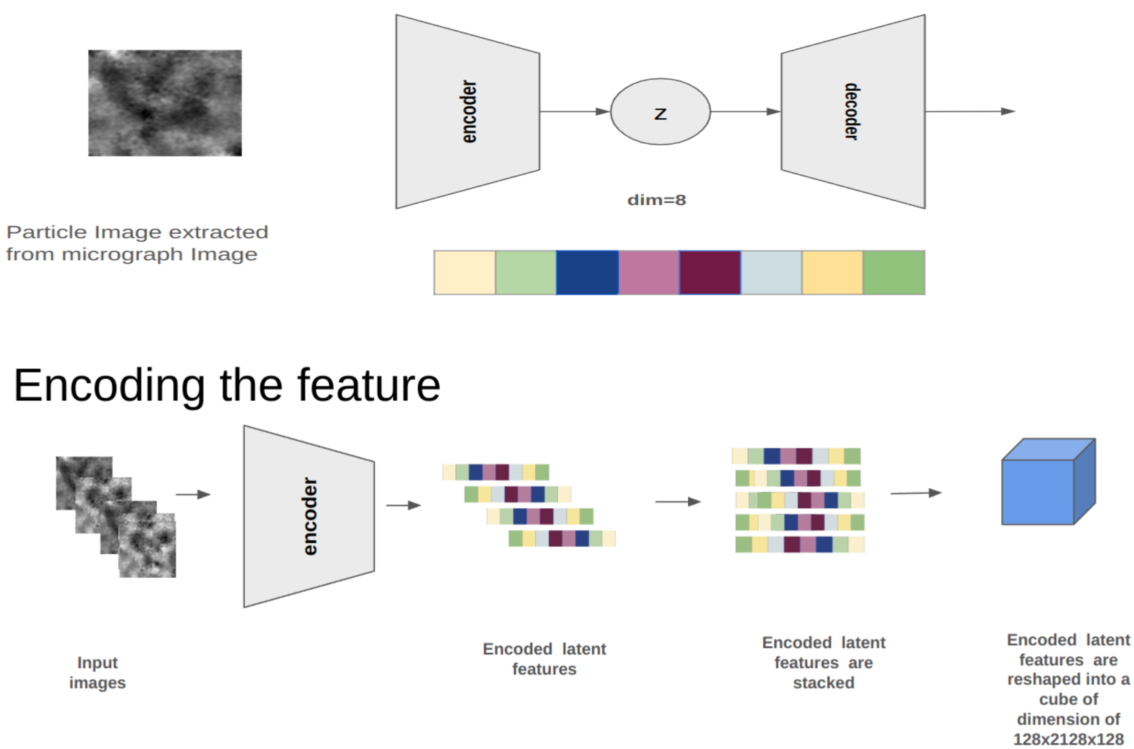
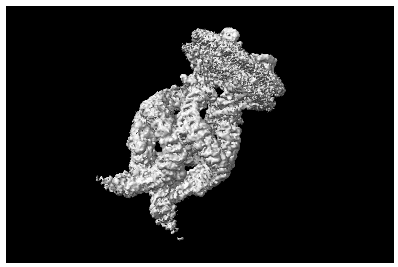
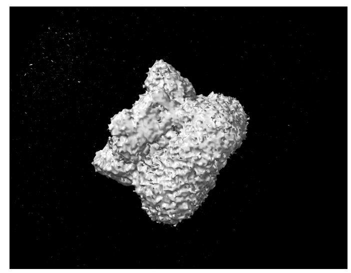
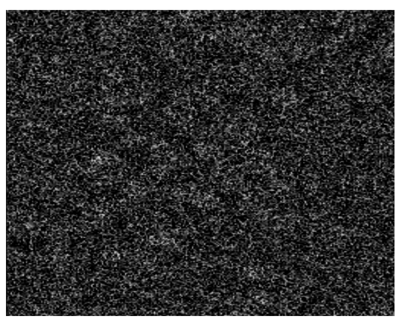
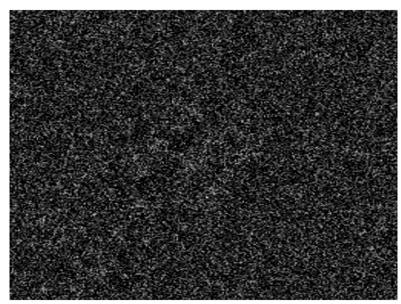
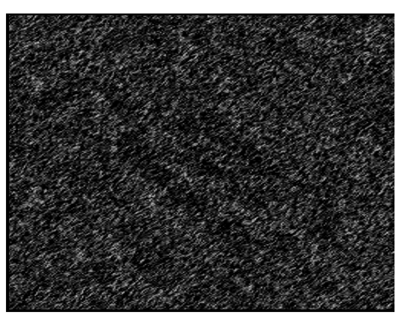
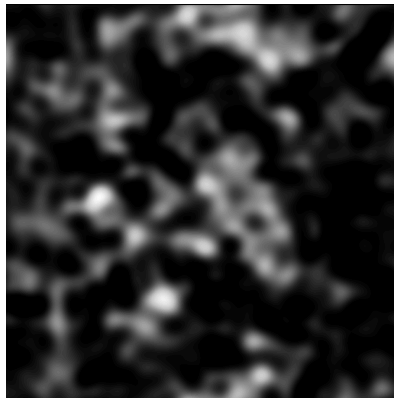
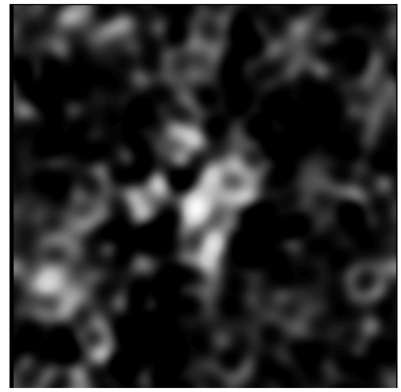
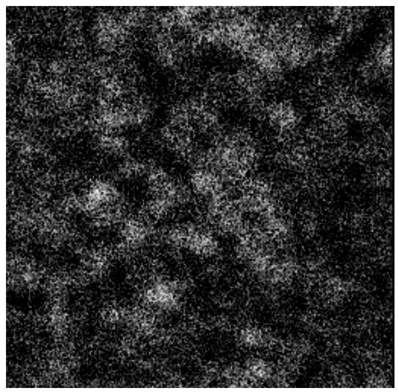
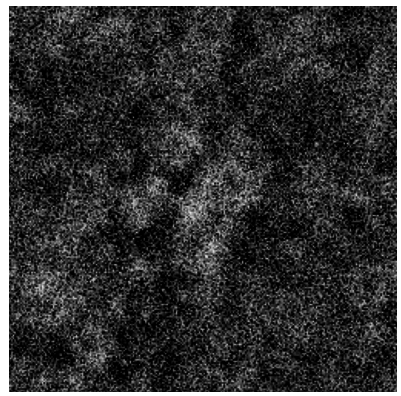
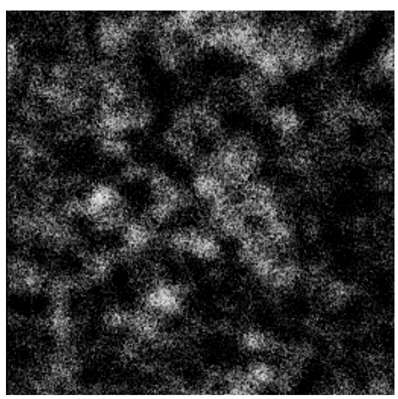
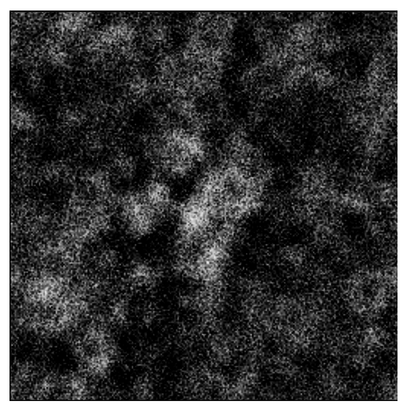


Figure 9: Displaying the input preparation using particle stack.

5.**Further Input Preprocessing**

Commonly the particle stacks were processed using a contrast transfer function to remove high frequency noise which is displayed in Figure 10 c. Later on the focus tool was used to do the denoising and it shows better clarity of the image, but I would suggest to use the weighted average of the original image and denoised to preserve the features of the particle more, the effect of various weighted averages are shown on Figure10.

a) 3D Structure		
b) Original Image of a particle from particle stack		
c) Contrast transfer function		

d) Denoised using Focustools		
e) (Denoised + Original)/2		
f) (2* Denoised + Original)/3		

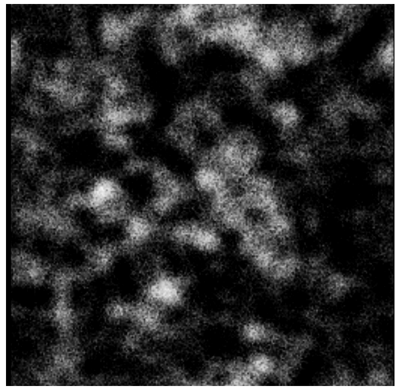
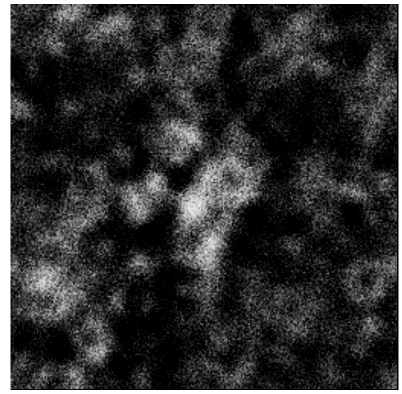
g) (4* Denoise + Original)/5		

Figure10: Illustration the 3d density map,original image of particle, denoised image using focus tools and result of different weighted average of them.

Important paper:

- Donnat C, Levy A, Poitevin F, Zhong ED, Miolane N. Deep generative modeling for volume reconstruction in cryo-electron microscopy. J Struct Biol. 2022 Dec;214(4):107920. doi: 10.1016/j.jsb.2022.107920. Epub 2022 Nov 8. PMID: 36356882; PMCID: PMC10437207. (Review paper)
- Zhong, E.D., Bepler, T., Berger, B. *et al.* CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat Methods* 18, 176–185 (2021). https://doi.org/10.1038/s41592-020-01049-4
- Punjani, A., Rubinstein, J., Fleet, D. *et al.* cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* 14, 290–296 (2017). https://doi.org/10.1038/nmeth.4169

- Levy A, Poitevin F, Martel J, Nashed Y, Peck A, Miolane N, Ratner D, Dunne M, Wetzstein G. CryoAI: Amortized Inference of Poses for Ab Initio Reconstruction of 3D Molecular Volumes from Real Cryo-EM Images. Comput Vis ECCV. 2022 Oct;13681:540-557. doi: 10.1007/978-3-031-19803-8_32. Epub 2022 Oct 23. PMID: 36745134; PMCID: PMC9897229.