Supplementary Materials for
Self-Supervised Cryo-Electron Tomography Volumetric Image Restoration from Single Noisy Volume with Sparsity Constraint

Zhidong Yang[1,2,3], Fa Zhang[1,*], Renmin Han[2,*]
[1]High Performance Computer Research Center, ICT, CAS
[2]Research Center for Mathematics and Interdisciplinary Sciences, Shandong University
[3]University of Chinese Academy of Sciences

Content

This supplementary material provides the following contents:

- Proof of preliminaries on noise model from 2D projection to 3D volumetric image (Section 1).
- The sources of datasets (Section 2).
- Measures used in simulated experiment (Section 3).
- Visual demonstration rule of the 3D volumes (Section 4).
- 3D visualization of the three simulated datasets in article (Section 5.1).
- 3D visualization of the four real-world datasets in article (Section 5.2).
- Additional ablation study on L_{smooth} (Section 5.3).

1. Proof of Preliminaries

1.1. Fourier-Slice Theorem for 3D Reconstruction

If $\varphi(\boldsymbol{x})$, where $\boldsymbol{x} = (x, y, z)^T$, represents a spatial potential function defined in $\mathbb{R}^3_n$ and $\varphi_{\boldsymbol{R}_n}(\boldsymbol{x}) = \varphi(\boldsymbol{R}_n^{-1}\boldsymbol{x})$ represents the spatial state of object $\varphi(\boldsymbol{x})$ rotated from the coordinate system $x - y - z$ by a rotation matrix $\boldsymbol{R}_n$, then the projection of the object along z-axis is given by

$$\boldsymbol{P}_{\boldsymbol{R}_n\varphi}(x, y) = \int_{-\infty}^{\infty} \varphi_{\boldsymbol{R}_n}(x, y, z)\mathrm{d}z. \tag{1}$$

The 2D Fourier transform of $\boldsymbol{P}_{\boldsymbol{R}_n\varphi}$ can be calculated as

$$\mathcal{P}_{\boldsymbol{R}_n\varphi}(\mu, v) = \iint_{-\infty}^{\infty} \boldsymbol{P}_{\boldsymbol{R}_n\varphi}(x, y)e^{-j2\pi(\mu x+vy)}\mathrm{d}x\mathrm{d}y. \tag{2}$$

If $\varPhi_{\boldsymbol{R}_n}(\mu, v, w)$ denotes the 3D Fourier transform of $\varphi_{\boldsymbol{R}_n}$, the Fourier-slice theorem [1] is given as

$$\mathcal{P}_{\boldsymbol{R}_n\varphi}(\mu, v) = \varPhi_{\boldsymbol{R}_n}(\mu, v, 0). \tag{3}$$

Thus, given $\hat{\varPhi}_{\boldsymbol{R}_n}(\boldsymbol{w}) = \varPhi_{\boldsymbol{R}_n}(\boldsymbol{R}_n^{-1}\boldsymbol{w})$ with $\boldsymbol{w} = (\mu, v, 0)^T$ measured with different $\boldsymbol{R}_n, n = 0, \cdots, N-1$, we can reconstruct $\varphi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ by inverse Fourier transform

$$\iiint_{-\infty}^{\infty} \sum_{n=0}^{N-1} \hat{\varPhi}_{\boldsymbol{R}_n}(\mu, v, w)e^{j2\pi(\mu x+vy+wz)}\mathrm{d}\mu\mathrm{d}v\mathrm{d}w. \tag{4}$$

*All correspondence should be addressed to Fa Zhang (zhangfa@ict.ac.cn) and Renmin Han (hanrenmin@sdu.edu.cn).

1.2. Noise model from 2D projection to 3D volume

Obviously, the Fourier-slice theorem also works in discrete condition. To simply the discussion, we assume that a projection image $\boldsymbol{I}_n(x, y)$ in cryo-ET is a discrete observation of the projection $\boldsymbol{P}_{\boldsymbol{R}_n\varphi}(x, y)$ (assume it has been discretized) with additive Gaussian noise $\boldsymbol{N}(x, j)$, i.e. $\boldsymbol{I}_n = \boldsymbol{P}_{\boldsymbol{R}_n\varphi} + \boldsymbol{N}_n$. Consequently, the Fourier transform ($\mathcal{F}$) of projection image $\boldsymbol{I}_n$ can be defined as $\mathcal{F}(\boldsymbol{I}_n) = \mathcal{P}_{\boldsymbol{R}_n\varphi} + \mathcal{F}(\boldsymbol{N}_n)$. We can draw a conclusion that **The additive Gaussian noise in 2D projection remains as Gaussian noise in 3D reconstruction**.

The proof of this Lemma is divided into two stages. We need to firstly prove that **The discrete Fourier transform of Gaussian noise is still a Gaussian process**.

Lemma 1. The discrete Fourier transform of Gaussian noise is still a Gaussian process.

Proof. Given random signals $X = \{x_n \sim N(0, \sigma^2)\}$ with probability density $P_{x_n}(x) = G_\sigma(x)$, the discrete Fourier transform of $\{x_n\}$ are defined as

$$X_k = \sum_{n=0}^{N-1} [x_n \cos(2\pi nk/N) - jx_n \sin(2\pi nk/N)]. \tag{5}$$

Considering the symmetry, in the following, we only discuss the real part of X_k. Letting $y_n = x_n \cos(2\pi nk/N)$, we have $\mathcal{R}X_k = \sum_{n=0}^{N-1} y_n$. Obviously, $Y = c_{n,k} \cdot X$ with $c_{n,k} = \cos(2\pi nk/N)$. The probability density of Y can be written as $P_{y_n}(y) = \frac{1}{c_{n,k}} G_{\sigma c_{n,k}}(\frac{y}{c_{n,k}})$, and the probability density of $\mathcal{R}X_k$ becomes

$$P_{\mathcal{R}X_k}(x) = (G_{\sigma c_{0,k}} \otimes G_{\sigma c_{1,k}} \cdots \otimes G_{\sigma c_{N-1,k}})(x). \tag{6}$$

Applying Fourier transform to Equation 6, we have

$$\begin{aligned}
\mathcal{F}(P_{\mathcal{R}X_k})(v) &= \prod_{n=0}^{N-1} \mathcal{F}(G_{\sigma c_{n,k}})(v) = \prod_{n=0}^{N-1} G_{1/(\sigma c_{n,k})} \\
&= \prod_{n=0}^{N-1} \left[\left(\frac{\sigma^2 c_{n,k}^2}{2\pi} \right)^{\frac{1}{2}} \exp\left(\frac{-v^2}{2/\sigma^2 c_{n,k}^2} \right) \right] \\
&= C \cdot \exp\left[\frac{-v^2\sigma^2}{2} \sum_{n=o}^{N-1} \cos(\frac{2\pi nk}{N})^2 \right],
\end{aligned} \tag{7}$$

where $C = \left(\frac{\sigma^2}{2\pi}\right)^{\frac{N}{2}} \left(\prod_{n=0}^{N-1} c_{n,k}\right)$ is a constant, and $\sum_{n=o}^{N-1} \cos(\frac{2\pi nk}{N})^2 = \frac{N}{2}$. Consequently,

$$\mathcal{F}(P_{\mathcal{R}X_k}) \propto \exp\left(\frac{-v^2\sigma^2}{2} \cdot \frac{N}{2} \right) = G_{(2/\sigma^2 N)^{1/2}}, \tag{8}$$

and we get $P_{\mathcal{R}X_k} = \mathcal{F}^{-1}(G_{(2/\sigma^2 N)^{1/2}}) = G_{\sigma\sqrt{N/2}}$.

Here, we could get the similar conclusion for $P_{\mathcal{I}X_k}$ and so for P_{X_k}. Therefore, X_k is still a Gaussian process. □

Lemma 2. The additive Gaussian noise in 2D projection remains as Gaussian noise in 3D reconstruction.

Proof. Given the observed projection $\{\boldsymbol{I}_i\}$ and its corresponding rotation matrix $\{\boldsymbol{R}_i\}$, we have $\mathcal{F}(\boldsymbol{I}_n)(\mu, v) = \mathcal{P}_{\boldsymbol{R}_n\varphi}(\mu, v) + \mathcal{F}(\boldsymbol{N}_n)(\mu, v)$ for all $n = 0, \cdots, N-1$, where $\mathcal{P}_{\boldsymbol{R}_n\varphi}(\mu, v) = \Phi_{\boldsymbol{R}_n}(\mu, v, 0)$.

Thus, given $\hat{\mathcal{I}}_n(\boldsymbol{w}) = \mathcal{F}(\boldsymbol{I}_n)(\boldsymbol{R}_n^{-1}\boldsymbol{w})$ with $\boldsymbol{w} = (\mu, v, 0)^T$, we will have $\hat{\mathcal{I}}_n(\boldsymbol{w}) = \hat{\Phi}_{\boldsymbol{R}_n}(\boldsymbol{w}) + \hat{\mathcal{N}}_n(\boldsymbol{w})$, where $\hat{\mathcal{N}}_n(\boldsymbol{w})$ is a 3D Fourier transform defined on a spatial plane with values same as $\mathcal{F}(\boldsymbol{N}_n)(\mu, v)$ but rotated from the $x - y$ plane by $\boldsymbol{R}_n$. Consequently, a volumetric image can be reconstructed from $\{\hat{\mathcal{I}}_n\}$ as[1]

$$\begin{aligned}
\boldsymbol{V}(\boldsymbol{x}) = \mathcal{F}^{-1}\left(\sum_{n=0}^{N-1} \hat{\mathcal{I}}_n \right) &= \mathcal{F}^{-1}\left[\sum_{n=0}^{N-1} \left(\hat{\Phi}_{\boldsymbol{R}_n} + \hat{\mathcal{N}}_n \right) \right] \\
&= \boldsymbol{\Phi}(\boldsymbol{x}) + \mathcal{F}^{-1}\left(\sum_{n=0}^{N-1} \hat{\mathcal{N}}_n \right)
\end{aligned} \tag{9}$$

[1]Here, the interpolation process has been ignored for the convenience of discussion.

where $\mathbf{\Phi}(\boldsymbol{x})$ is the discrete version of $\varphi(\boldsymbol{x})$. Here, it should be found that $\sum_{n=0}^{N-1} \hat{\mathcal{N}}_n$ is composed of Gaussian random variables and so thus for its $\mathcal{F}^{-1}$. Therefore, we have

$$\boldsymbol{V}(\boldsymbol{x}) = \mathbf{\Phi}(\boldsymbol{x}) + \boldsymbol{N}(\boldsymbol{x}), \tag{10}$$

where $\boldsymbol{V}(\boldsymbol{x})$, $\mathbf{\Phi}(\boldsymbol{x})$ and $\boldsymbol{N}(\boldsymbol{x})$ are all defined in $\mathbb{R}^3$. □

In conclusion, for additive noise, the noise modeling in 3D volume remains stable as in 2D projections. So that our SC-Net can directly accept a 3D volumetric image as input for training under such assumption.

2. Sources of Datasets

Real-world datasets. The first two datasets and the fourth dataset Vesicle, Mitochondria and VEEV are provided by the Institute of Biophysics, Chinese Academy of Sciences. The third dataset Centriole is downloaded from IMOD tutorial (http://bio3d.colorado.edu/imod/files/tutorialData-1K.tar.gz).

3. Metrics in Simulated Experiment

PSNR and SSIM are selected as measures for our evaluations in simulated experiment. Eq.11 and Eq.12 give the mathematical formulations of PSNR and SSIM, respectively.

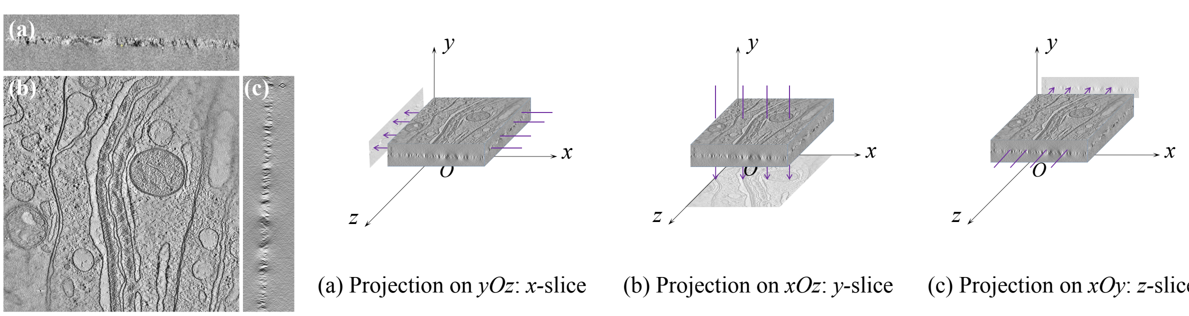
$$PSNR\left(\tilde{\boldsymbol{V}}, \boldsymbol{V}^g\right) = 20 \log_{10}\left(\frac{MAX_I}{MSE}\right) \tag{11}$$

$$SSIM\left(\tilde{\boldsymbol{V}}, \boldsymbol{V}^g\right) = \frac{\left(2\mu_{\tilde{\boldsymbol{V}}}\mu_{\boldsymbol{V}^g} + C_1\right)\left(2\sigma_{\tilde{\boldsymbol{V}}\boldsymbol{V}^g} + C_2\right)}{\left(\mu_{\tilde{\boldsymbol{V}}}^2 + \mu_{\boldsymbol{V}^g}^2 + C_1\right)\left(\sigma_{\tilde{\boldsymbol{V}}}^2 + \sigma_{\boldsymbol{V}^g}^2 + C_2\right)} \tag{12}$$

In the formulas shown above, $\tilde{\boldsymbol{V}}$ represents the denoised output image and $\boldsymbol{V}^g$ represents the ground truth image. In our experiments, pixel values of each image are normalized to $[0, 1]$. Hence the parameter MAX_I in Eq.11 is set to 1. For Eq.12, according to the study in [2], $C1 = (K_1 * L)^2$ and $C2 = (K_2 * L)^2$ with $K_1 = 0.01$, $K_2 = 0.03$ and $L = 255$ for 8-bits images. However, because the maximum of pixel values is 1 according to our normalization, L is set to 1 in our experiments.

4. Visual Demonstration Rule of 3D Volumes

To comprehensively present the results of different methods, we follow the visual demonstration rule shown in Figure 1. We project the reconstructed volumetric image on xOy, xOz and yOz plane separately and name them by their projection axis. That is, x-slice, y-slice and z-slice. As for y-slice, we select the slice from the center of volume, which has more abundant structural information than other positions.



(a) Projection on yOz: x-slice (b) Projection on xOz: y-slice (c) Projection on xOy: z-slice

Figure 1: Visual demonstration rule of the 3D volumes. We project the volumetric image into yOz, xOz and xOy plane separately to show the visual results in three different views, which can get a complete insight of image in 3D space.

5. Additional Information for Experimental Results

5.1. Additional Information for Simulated Data

To comprehensively analyze the visual results of simulated data, we select additional areas on z-slice. Table 1 shows the selected coordinates for each data. Figure 2-4 show the 3D visual results for simulated data.

Table 1: Selected coordinates for visual results of simulated data.

Dataset	Belt	Synapse	SARS
x-coord.	1	1	1
y-coord.	150	65	60
z-coord.	513	513	513

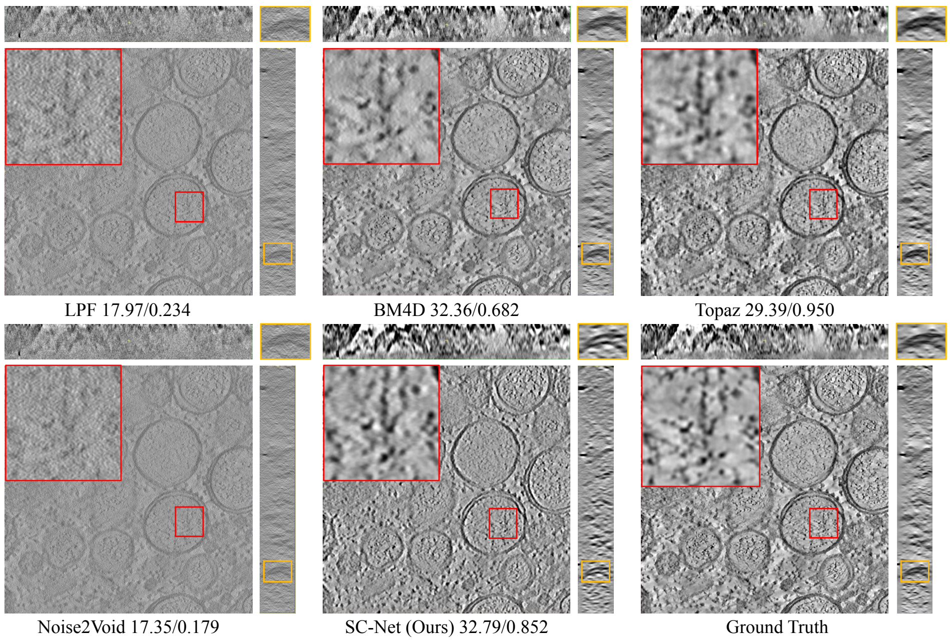


Figure 2: 3D Visual Results of SARS. Selected coordinates are x=1, y=150, z=513. Shown metrics: PSNR(dB)/SSIM.

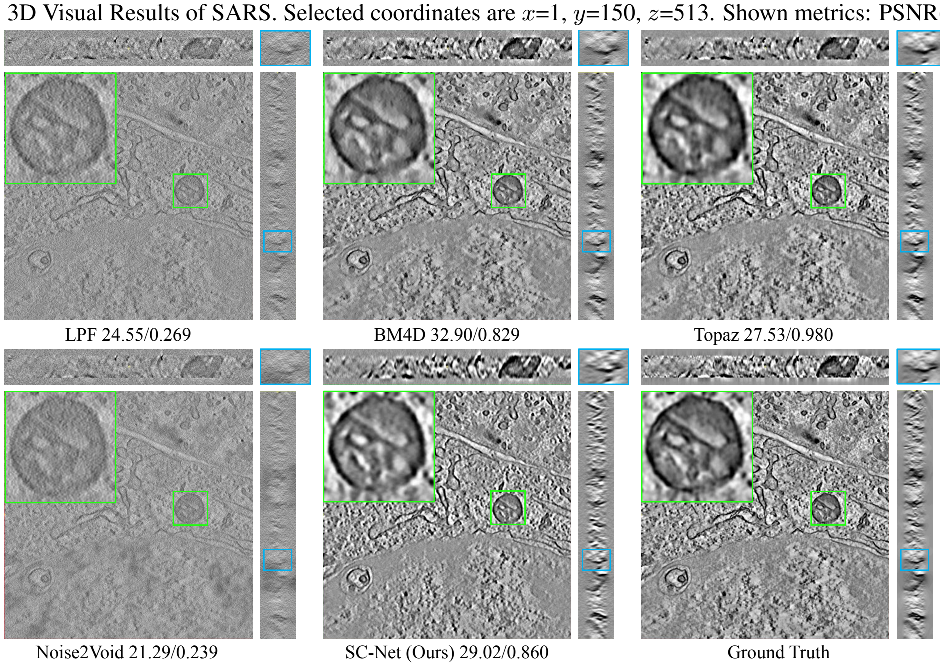


Figure 3: 3D Visual Results of Synapse. Selected coordinates are x=1, y=65, z=513. Shown metrics: PSNR(dB)/SSIM.

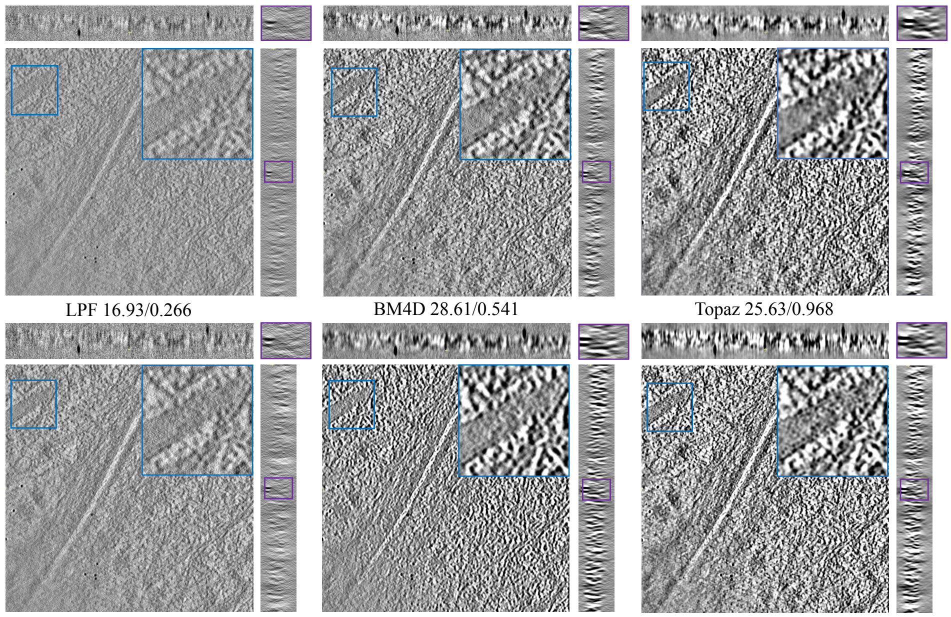


Figure 4: 3D Visual Results of Belt. Selected coordinates are x=1, y=60, z=513. Shown metrics: PSNR(dB)/SSIM.

5.2. Additional Information for Real-world Data

To comprehensively show the results of real-world data, we select different x, y and z coordinates for each data. Table 2 shows the selected coordinates for each data. Figure 5-8 show the 3D results on selected coordinates. Comparing with Topaz, our method can produce an enhanced result without introducing grid artifacts.

Table 2: Selected coordinates for visual results of real-world data.

Dataset	Vesicle	Mitochondria	Centriole	VEEV
x-coord.	360	905	95	650
y-coord.	150	150	120	120
z-coord.	513	740	670	1120

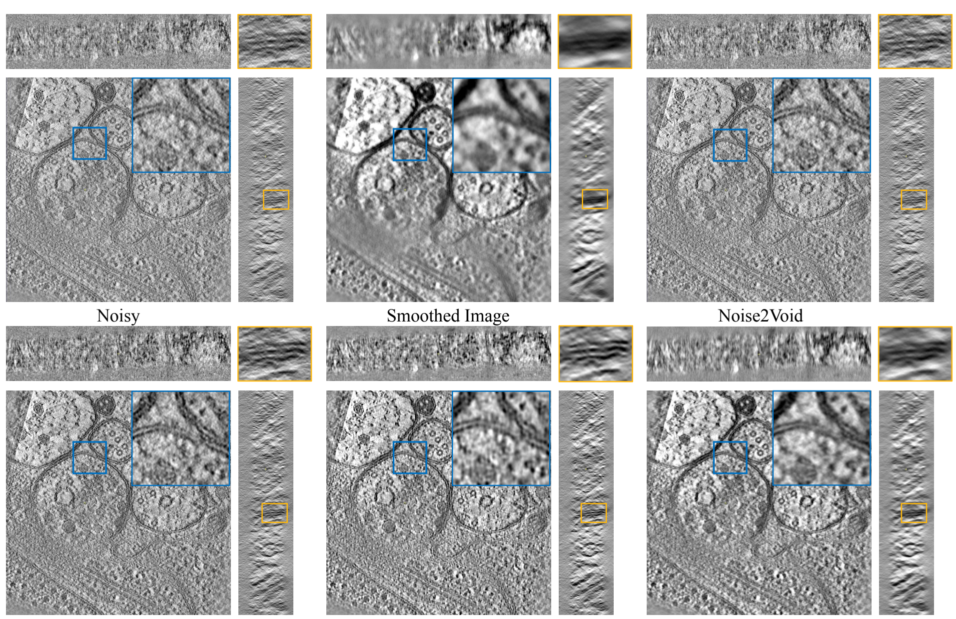


Figure 5: 3D Visual Results of Vesicle. Selected coordinates are x=360, y=150, z=513.

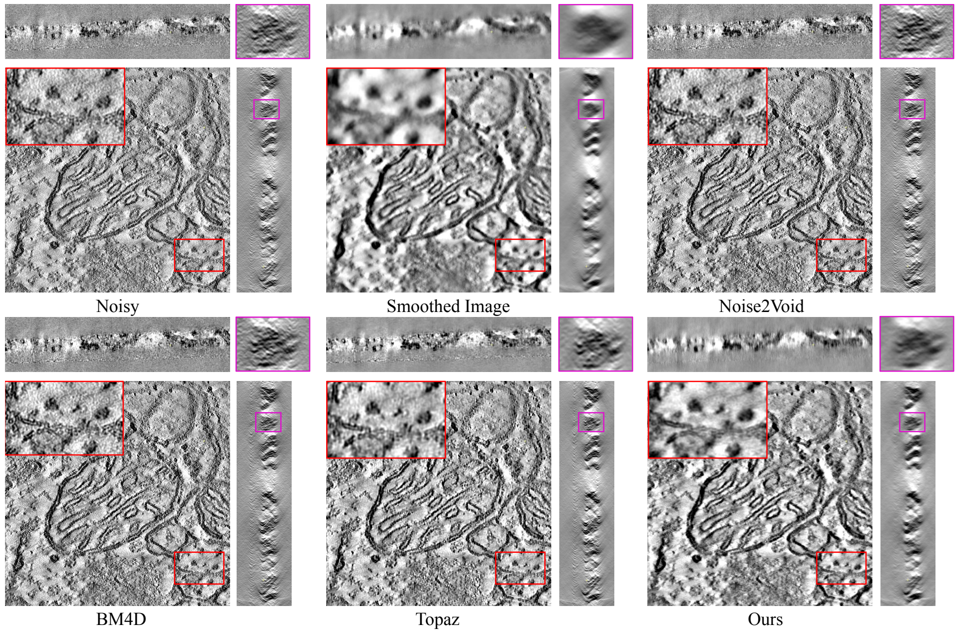


Figure 6: 3D Visual Results of Mitochondria. Selected coordinates are x=905, y=150, z=740.

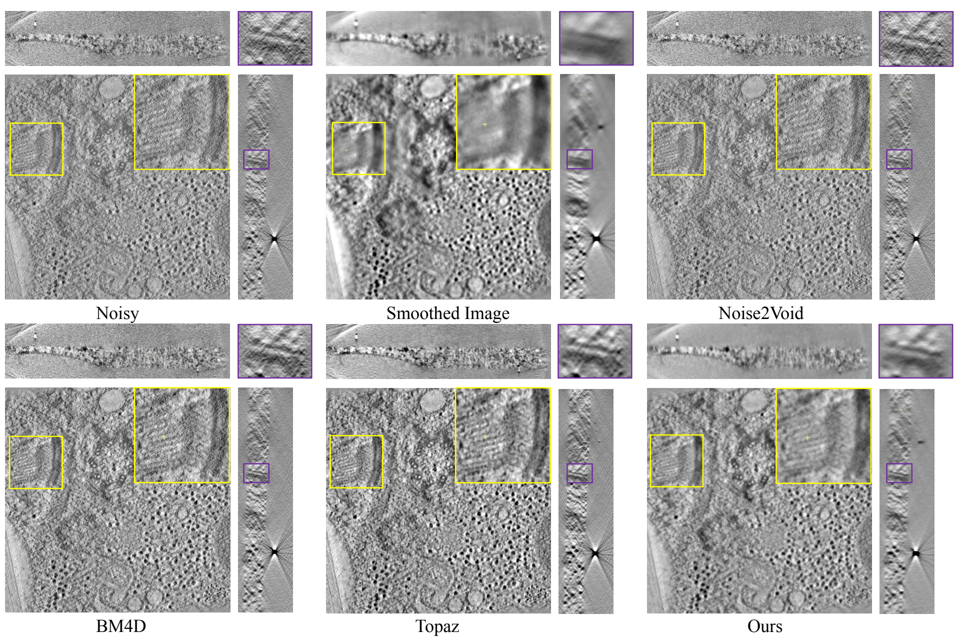


Figure 7: 3D Visual Results of Centriole. Selected coordinates are x=95, y=120, z=670.

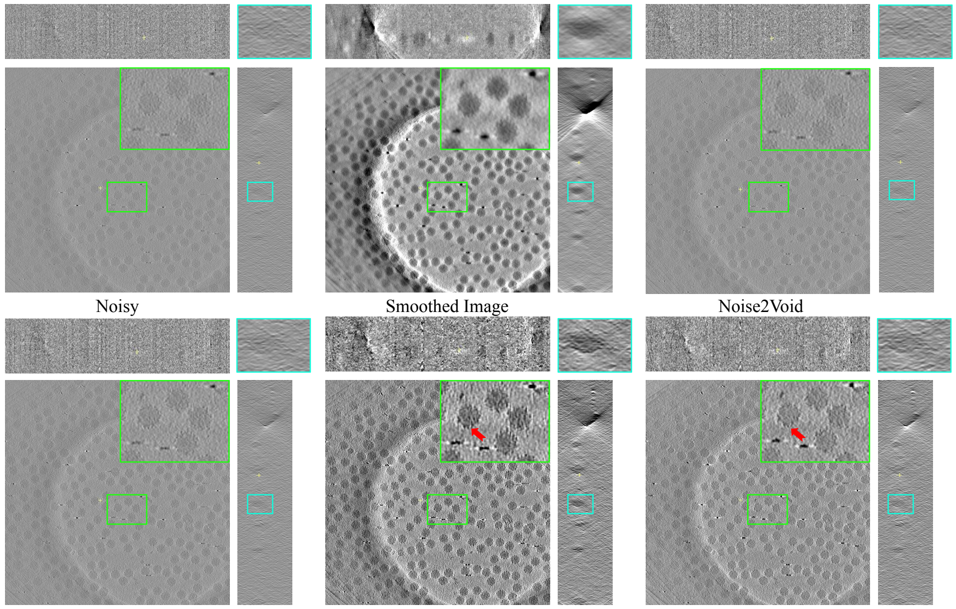


Figure 8: 3D Visual Results of VEEV. Selected coordinates are x=650, y=120, z=1120. The area pointed out by the red arrow has proved that, when reconstructing the membrane structure of particles, our method can produce a more reliable result without introducing artifacts comparing with Topaz.

5.3. Additional Ablation Study on L_{smooth}

In this study, we test our SC-Net in two forms: complete loss function with L_{smooth} (i.e., with L_{smooth}) and the loss function without L_{smooth} (i.e., Non-L_{smooth}). Table 3 shows the quantitative analysis of output from these two models and Fig 9 shows the visual results. Results prove that L_{smooth} can provide noticeable improvement for the performance of noise smoothing and structure preservation when training data is not sufficient.

Table 3: PSNR(dB)/SSIM Results for Ablation Study on Sparsity Constraint under the noise with σ=10.

Dataset	Noisy	Non-L_{smooth}	With L_{smooth}
Synapse	27.79/0.503	24.56/0.680	**30.03/0.932**
Belt	22.76/0.423	26.28/0.629	**28.61/0.914**

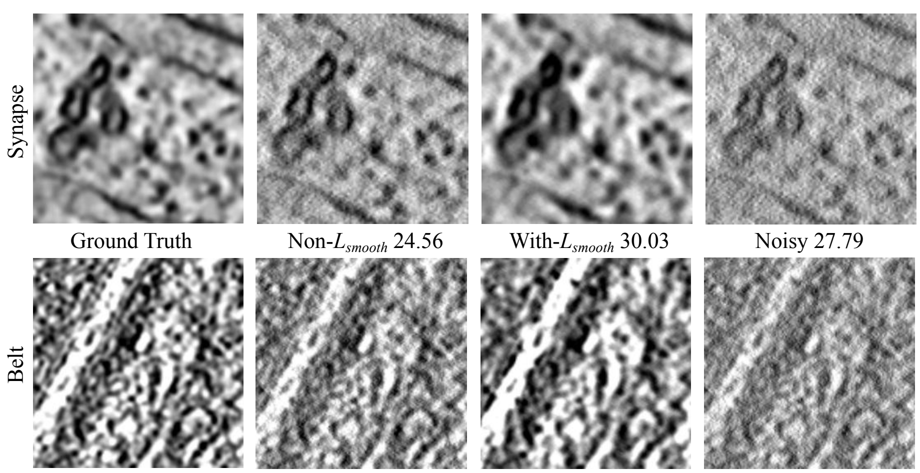


Figure 9: Visual results of comparisons between volumes filtered by SC-Net with sparsity constraint loss and the one without sparsity constraint loss. (noise intensity: $\sigma = 10$, metric:PSNR(dB)).

References

[1] Ren Ng. Fourier slice photography. In *ACM SIGGRAPH 2005 Papers*, pages 735–744. 2005. 1

[2] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 3