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

Zhidong Yang¹²³, Fa Zhang¹,∗, Renmin Han²,∗
¹High Performance Computer Research Center, ICT, CAS
²Research Center for Mathematics and Interdisciplinary Sciences, Shandong University
³University 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 $\varphi(x)$, where $x = (x, y, z)^T$, represents a spatial potential function defined in $\mathbb{R}^3$ and $\varphi_{R_n}(x) = \varphi(R_n^{-1}x)$ represents the spatial state of object $\varphi(x)$ rotated from the coordinate system $x - y - z$ by a rotation matrix $R_n$, then the projection of the object along $z$-axis is given by

$$P_{R_n\varphi}(x, y) = \int_{-\infty}^{\infty} \varphi_{R_n}(x, y, z)dz. \quad (1)$$

The 2D Fourier transform of $P_{R_n\varphi}$ can be calculated as

$$\mathcal{F}_{R_n\varphi}(\mu, v) = \int_{-\infty}^{\infty} P_{R_n\varphi}(x, y)e^{-j2\pi(\mu x + vy)}dx dy. \quad (2)$$

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

$$\mathcal{F}_{R_n\varphi}(\mu, v) = \Phi_{R_n}(\mu, v, 0). \quad (3)$$

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

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{n=0}^{N-1} \hat{\Phi}_{R_n}(\mu, v, w)e^{j2\pi(\mu x + vy + wz)}d\mu dv dw. \quad (4)$$

*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 $I_n(x, y)$ in cryo-ET is a discrete observation of the projection $P_{R_n \varphi} (x, y)$ (assume it has been discretized) with additive Gaussian noise $N(x, y)$, i.e. $I_n = P_{R_n \varphi} + N_n$. Consequently, the Fourier transform ($\mathcal{F}$) of projection image $I_n$ can be defined as $\mathcal{F}(I_n) = P_{R_n \varphi} + \mathcal{F}(N_n)$. 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 $\{x_n\}$ is defined as

$$X_k = \sum_{n=0}^{N-1} x_n \cos(2\pi nk/N) - j x_n \sin(2\pi nk/N).$$

(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 $\Re X_k = \sum_{n=0}^{N-1} y_n$. 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 $P_y(y) = \frac{1}{c_{n,k}} G_{\sigma c_{n,k}}(c_{n,k})$, and the probability density of $\Re X_k$ becomes

$$P_{\Re X_k}(x) = (G_{\sigma c_{n,k}} \otimes G_{\sigma c_{1,k}} \cdots \otimes G_{\sigma c_{N-1,k}})(x).$$

(6)

Applying Fourier transform to Equation 6, we have

$$\mathcal{F}(P_{\Re X_k})(v) = \prod_{n=0}^{N-1} \mathcal{F}(G_{\sigma c_{n,k}})(v) = \prod_{n=0}^{N-1} G_{1/(\sigma c_{n,k})}$$

(7)

$$= \prod_{n=0}^{N-1} \left[ \left( \frac{\sigma^2 c_{n,k}}{2\pi} \right)^{\frac{1}{2}} \exp \left( -\frac{v^2}{2\sigma^2 c_{n,k}} \right) \right]$$

$$= C \cdot \exp \left( -\frac{v^2 \sigma^2}{2} \sum_{n=0}^{N-1} \cos\left( \frac{2\pi nk}{N} \right)^2 \right),$$

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 $\sum_{n=0}^{N-1} \cos\left( \frac{2\pi nk}{N} \right)^2 = \frac{N}{2}$. Consequently,

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

(8)

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

Here, we could get the similar conclusion for $P_{\Im X_k}$ 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 $\{I_n\}$ and its corresponding rotation matrix $\{R_n\}$, we have $\mathcal{F}(I_n)(\mu, v) = P_{R_n \varphi}(\mu, v) + \mathcal{F}(N_n)(\mu, v)$ for all $n = 0, \cdots, N - 1$, where $P_{R_n \varphi}(\mu, v) = \hat{\Phi}_{R_n}(\mu, v, 0)$.

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

$$V(x) = \mathcal{F}^{-1} \left( \sum_{n=0}^{N-1} \hat{I}_n \right) = \mathcal{F}^{-1} \left( \sum_{n=0}^{N-1} \left( \hat{\Phi}_{R_n} + \hat{N}_n \right) \right)$$

(9)

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

\[
V(x) = \Phi(x) + N(x),
\]

where \( V(x), \Phi(x) \) and \( N(x) \) 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.

\[
PSNR(V^g, \hat{V}) = 20 \log_{10} \left( \frac{MAX_I}{MSE} \right) \tag{11}
\]

\[
SSIM(V^g, \hat{V}) = \frac{(2\mu_{\hat{V}}\mu_{V^g} + C_1)(2\sigma_{\hat{V}V^g} + C_2)}{(\mu_{\hat{V}}^2 + \mu_{V^g}^2 + C_1)(\sigma_{\hat{V}}^2 + \sigma_{V^g}^2 + C_2)} \tag{12}
\]

In the formulas shown above, \( \hat{V} \) represents the denoised output image and \( V^g \) 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], \( C_1 = (K_1 \ast L)^2 \) and \( C_2 = (K_2 \ast L)^2 \) 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.

![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.](image-url)
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.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Belt</th>
<th>Synapse</th>
<th>SARS</th>
</tr>
</thead>
<tbody>
<tr>
<td>x-coord.</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>y-coord.</td>
<td>150</td>
<td>65</td>
<td>60</td>
</tr>
<tr>
<td>z-coord.</td>
<td>513</td>
<td>513</td>
<td>513</td>
</tr>
</tbody>
</table>

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.
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.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Vesicle</th>
<th>Mitochondria</th>
<th>Centriole</th>
<th>VEEV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$-coord.</td>
<td>360</td>
<td>905</td>
<td>95</td>
<td>650</td>
</tr>
<tr>
<td>$y$-coord.</td>
<td>150</td>
<td>150</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td>$z$-coord.</td>
<td>513</td>
<td>740</td>
<td>670</td>
<td>1120</td>
</tr>
</tbody>
</table>
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 $L_{\text{smooth}}$

In this study, we test our SC-Net in two forms: complete loss function with $L_{\text{smooth}}$ (i.e., with $L_{\text{smooth}}$) and the loss function without $L_{\text{smooth}}$ (i.e., Non-$L_{\text{smooth}}$). Table 3 shows the quantitative analysis of output from these two models and Fig 9 shows the visual results. Results prove that $L_{\text{smooth}}$ 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 $\sigma=10$.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Noisy</th>
<th>Non-$L_{\text{smooth}}$</th>
<th>With $L_{\text{smooth}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synapse</td>
<td>27.79/0.503</td>
<td>24.56/0.680</td>
<td>30.03/0.932</td>
</tr>
<tr>
<td>Belt</td>
<td>22.76/0.423</td>
<td>26.28/0.629</td>
<td>28.61/0.914</td>
</tr>
</tbody>
</table>

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

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
