Supporting Information

Single-Image Super-Resolution Improvement of X-ray Single-Particle Diffraction Images Using a Convolutional Neural Network

Atsushi Tokuhisa,*‡,† Yoshinobu Akinaga,‡,†,§ Kei Terayama,¶ ‡ Yuji Okamoto‡ and Yasushi Okuno,*‡,†,⊥

‡RIKEN Center for Computational Science, 7-1-26, Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan

†RIKEN Medical Sciences Innovation Hub Program, 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan

§VINAS Co., Ltd., Keihan Dojima Bldg., Dojima 2131, Kita-ku, Osaka 530-0003, Japan

¶Graduate School of Medical Life Science, Yokohama City University, 1-7-29, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

∥RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihombashi, Chuo-ku, Tokyo 103-0027, Japan
S1. Description of Figure 1 as an overview

Data were prepared with synthetically generated labeled data between grand truth high-resolution (GTHR) images and low-resolution (LR) images in silico space using molecular dynamics simulation and diffraction image simulator. The SRCNN model was trained using GTHR images and LR images prepared in silico space, considering shot noise, molecular orientations, multiconformation, and X-ray intensity fluctuations. An SRCNN image can be predicted using a trained SRCNN model from the single-particle experimental image. Various structural analyses could be performed using the SRCNN image, such as the template matching method and HIO method.

S2. Protocol for 70S ribosome MD simulation

The structure of the 70S ribosome was obtained from the PDB (PDB ID: 4v9o), which contains 50S ribosomal proteins (chain CC-C5), 30S ribosomal proteins (chain DB-DU), an elongation factor G (chain DV), 23S rRNA (chain CA), 5S rRNA (chain CB), and 16S rRNA (chain DA). We removed the viomycin molecule because its parameters were not covered by the employed force field. The residues missing in the DN (Ala36–Arg41) and DV (Asn37–Ala48) were complemented using the software provided by the molecular operating
environment (Chemical Computing Group, Montreal, Canada) structure preparation module, version 2016.08.

Energy minimization and MD simulations were conducted using modules of GROMACS v. 2016.357 by adopting the AMBER ff99SB force field to relax the built target molecules, according to the following protocol. A target molecule was placed in a rectangular space 20 Å larger than the target molecule in the X, Y, and Z directions. It was solvated with 150 mM NaCl solution using the genion module of GROMACS. The energy of the entire system was minimized using the steepest descent algorithm. Initial velocities were randomly generated according to the Maxwell–Boltzmann distribution at 298.15 K. An equilibration MD simulation of 50 ps was conducted. During equilibration, crystallized atoms, including water molecules, were restrained to the minimized structure with a harmonic potential of 1000 kJ mol\(^{-1}\) nm\(^{2}\). A production run of conformational sampling with a 200 ns MD simulation was conducted. The trajectory from 0.5 to 200 ns of the production run was used in the subsequent analysis of the 70S ribosome. Then, we selected three structures with RMSDs of 0, 5, and 10 Å with respect to the initial structure.

**S3. Procedure of simulating X-ray diffraction images**

The EMV of photons arriving at a detector pixel of \(s(k)\) were as follows:

\[
s(k) = s(Ah) = s(A; \xi, \varphi) = I_i r_{ce}^2 \omega(\xi) |F(k)|^2,
\]

where \(I_i\) was the incident X-ray intensity and was expressed in photons pulse\(^{-1}\) \(\mu\)m\(^{-2}\), \(r_{ce}\) was the classical electron radius, and \(F(k)\) was the structure factor. \(K\) and \(h\) were the momentum transfer vectors in the molecular fixed coordinate system (MFCS) and the detector fixed coordinate system (DFCS), respectively. \(A\) was the \(3 \times 3\) matrix describing the molecular orientation of the sample molecule to the incident X-ray beam, which was given in the Eulerian angle \((\alpha, \beta, \gamma)\) by the following expression:
\[
A = A_3(-\gamma)A_2(-\beta)A_3(-\alpha)
\]

\[
A_3(\alpha) = \begin{pmatrix}
\cos \alpha & -\sin \alpha & 0 \\
\sin \alpha & \cos \alpha & 0 \\
0 & 0 & 1
\end{pmatrix},
A_2(\beta) = \begin{pmatrix}
\cos \beta & 0 & \sin \beta \\
0 & 1 & 0 \\
-\sin \beta & 0 & \cos \beta
\end{pmatrix}
\]

The surface vector of the Ewald sphere \( \mathbf{h}(\xi, \varphi) \) in DFCS was as follows:

\[
\mathbf{h}(\xi, \varphi) = \lambda^{-1} \left[ \mathbf{E} - A_3(\varphi + \pi)A_2(\xi) \right] \mathbf{e}_3
\]

\[
= \lambda^{-1} (\sin \xi \cos \varphi, \sin \xi \sin \varphi, 1 - \cos \xi)^t, \quad \mathbf{e}_3 = (0,0,1)^t,
\]

where \((\xi, \varphi)\) was a polar coordinate. Here, \( \mathbf{h} \) became 0 at the origin \( \xi = 0 \) and \( \varphi = 0 \). The magnitude of momentum transfer of \( k \) was determined by the scattering angle of \( \xi \) and the incident X-ray wavelength of \( \lambda \), and the wave number \( k = |k| \text{Å}^{-1} \) was given by the following equation using \( \xi = 2\theta \), where \( \theta \) was the scattering angle:

\[
k = \frac{2}{\lambda} \sin \frac{\xi}{2}.
\]

\( \omega \) was the solid angle per pixel and is expressed by the following equation using the incident X-ray wavelength \( \lambda \), molecular length \( L \), scattering angle \( \xi \), and linear oversampling ratio \( \sigma \):

\[
\omega(\xi) = \left( \frac{\lambda}{\sigma L} \right)^2 \cos^2 \xi.
\]

The EMV diffraction images were simulated with the EMV of \( s(k) \) without any noise effect. For the experimental diffraction images, we simulated the probability of an integral number of photons given in the pixel by replacing the function \( s(k) \) with a stochastic function \( s_Q(k) \) that assumed only integral values according to the Poisson distribution given by the quantum mechanical probability because only integral numbers of photons are counted in real experiments.

In fact, here \( \lambda = 1 \text{Å} \) corresponds to an X-ray energy with 12.4 keV, \( \sigma = 1 \) or 2, and \( L = 270 \text{Å} \). Additionally, the maximum wavenumber of each simulated diffraction image was set to 0.2 \( \text{Å}^{-1} \), and the diffraction images corresponding to a 5 \( \text{Å} \) structural resolution were calculated.
EMV diffraction images from 5,120 molecular orientations, which filled the possible molecular orientations evenly, were prepared with an image size of 220 × 220 pixels for one structure and used as the EMV image dataset. The experimental diffraction images were prepared by adding Poisson noise corresponding to the scattered X-ray intensity to the EMV diffraction images. Noise sources presented in the experimental diffraction images, such as inelastic scattering, beam stopping, and detector reading noise, were not considered here.

**S4. Anscombe-transformation**

To confirm the effect of the Anscombe transformation in the SRCNN model, we examined the case of incorporating the Anscombe transformation in the pre-processing. The intensity of each pixel was converted using the following formula for the Anscombe transformation:

\[
s_{Q,AT}(k) = 2.0 \times \sqrt{s_Q(k) + \frac{3.0}{8.0}}
\]  

(S7)

The image quality improvement by SRCNN was performed using precisely the same procedure described in the methods section except with and without applying Anscombe transformation. Figure S1 presents the results of the similarity improvement by the SRCNN model with and without Anscombe-transformation. There is a slight difference in the \( R_c \) score between the two cases, confirming a remarkably high correlation. This result indicates that applying Anscombe transformation in the pre-processing of diffraction images for the SRCNN training is not essential.
Figure S1. Change in similarity improvement by SRCNN model with or without Anscombe-transformation. X: $R_c$ score of SRCNN with Anscombe transformation. Y: $R_c$ score of SRCNN without Anscombe transformation.

S5. Diffraction image data sets

Three data sets were prepared as follows for the training, validation, and test, respectively:

**Training data:** The training data is used in the training process to update the model parameters through backpropagation. EMV images were created as the GTHR images from a set of $N = 640$ molecular orientations for each molecular structure. Diffraction images of $N$ molecular orientations were prepared for each conformation with $S$ types considering structural fluctuations, i.e., we prepared $N \times S$ EMV images in total. Meanwhile, we prepared $N \times S \times I \times 5$ experimental images as LR images, where $I$ is the number of X-ray intensities and 5 is the number of randomly generated noise patterns.

**Validation data:** The validation data, which was used to monitor the training process of each epoch, was prepared in the same way as the training data, with another set of $N = 320$ molecular orientations.

**Test data:** The test data was used to evaluate the trained model. This test data was created
from another set of \( N = 320 \) orientations, which did not appear in the training process, with varying intensities and a single noise pattern.

For each of the data sets described above, the molecular orientations are selected to fill all directions in space evenly. There is no common orientation between different image sets. Table S1 lists the 18 data sets employed here.

**Table S1.** Diffraction image data sets for the training/validation/test data were prepared for the SRCNN model. We prepared a variety of annotated synthetic data sets for training/validation/testing. Three different conformations with root mean square deviation (RMSD) of 0, 5, and 10 Å were prepared using the MD simulation described in Section S1. Diffraction images in 640 and 320 different molecular orientations were prepared for each case. Seven data sets with different X-ray incident intensity with A: \( 5 \times 10^{12} \), B: \( 8 \times 10^{12} \), C: \( 1 \times 10^{13} \), D: \( 3 \times 10^{13} \), E: \( 5 \times 10^{13} \), F: \( 1 \times 10^{14} \), and G: \( 1 \times 10^{15} \) photons pulse\(^{-1}\) μm\(^{-2}\) were prepared. We also show the number of partial images for training/validation.

| Dataset | RMSD [Å]: Training/validation/testing | Number of orientations: Training/validation/testing | X-ray incident intensity \( I_i \) [photons pulse\(^{-1}\) μm\(^{-2}\)]: Training/validation/testing | Number of total partial images: Training/validation |
|---------|--------------------------------------|-----------------------------------------------|-------------------------------------------------|-----------------------------------------------|
| 1       | 0/0/0                                | 640/320/320                                   | A/A/A                                           | 204,800/48,000                                |
| 2       | 5, 10/5, 10/0                        | 640/320/320                                   | A/A/A                                           | 409,600/96,000                                |
| 3       | 0/0/0                                | 640/320/320                                   | C/C/C                                           | 204,800/48,000                                |
|   |   |   |   |
|---|---|---|---|
| 4 | 5, 10/5, 10/0 | 640/320/320 | C/C/C | 409,600/96,000 |
| 5 | 0/0/0 | 640/320/320 | E/E/E | 204,800/48,000 |
| 6 | 5, 10/5, 10/0 | 640/320/320 | E/E/E | 409,600/96,000 |
| 7 | 0/0/0 | 640/320/320 | F/F/F | 204,800/48,000 |
| 8 | 5, 10/5, 10/0 | 640/320/320 | F/F/F | 409,600/96,000 |
| 9 | 0/0/0 | 640/320/320 | A, B/A, B/A, B | 409,600/96,000 |
| 10 | 5, 10/5, 10/0 | 640/320/320 | A, B/A, B/A, B | 819,200/192,000 |
| 11 | 0/0/0 | 640/320/320 | C, D/C, D/C, D | 409,600/96,000 |
| 12 | 5, 10/5, 10/0 | 640/320/320 | C, D/C, D/C, D | 819,200/192,000 |
| 13 | 0/0/0 | 640/320/320 | A, C/A, C/A, C | 409,600/96,000 |
| 14 | 5, 10/5, 10/0 | 640/320/320 | A, C/A, C/A, C | 819,200/192,000 |
| 15 | 0/0/0 | 640/320/320 | E, F/E, F/E, F | 409,600/96,000 |
A: $5 \times 10^{12}$, B: $8 \times 10^{13}$, C: $1 \times 10^{13}$, D: $3 \times 10^{13}$, E: $5 \times 10^{13}$, F: $1 \times 10^{14}$, G: $1 \times 10^{15}$

S6. Two-channel SRCNN model

Figure S2 shows a diagram of the adopted network model of SRCNN. The first layer ($f_1 \times f_1$ convolution) of the SRCNN model corresponded to the operation of “patch extraction of LR image and representation,” the second layer ($f_2 \times f_2$ convolution) corresponded to the operation of “embedding a partial image of LR image into the feature vector, and the third layer ($f_3 \times f_3$ convolution) corresponded to the operation of “searching for the corresponding GSHR image from the feature vector of the patch image of LR image.”
**Figure S2.** Two-channel SRCNN model. In the training phase, the partial images extracted from a whole diffraction image are used for data augmentation. LR images with the incident X-ray intensity data and the EMV images as GTHR images are used for training using the mean square error of EMV images for the loss function. In the first layer, patch extraction (yellow square) and representation of the LR image are conducted. In the second layer, nonlinear mapping from the LR image to the HR image is conducted. In the third layer, reconstruction of the HR image is conducted. In the test phase, the whole experimental image is input into the trained SRCNN model with the intensity data, and the whole SRCNN-predicted HR image is obtained.

**S7. Hyperparameter optimization**

Hyperparameter optimization was performed using the experimental design. To improve the generalization performance using mixed intensity data of various incident X-ray densities (dataset 18 in Table S1), 108 combinations of 5 hyperparameters were examined (see Table S2)
for the 2ch model using the hold-out method (300 images for training, 60 images for validation, and 100 images for testing). The $R_c$ values of the experiment and SRCNN were calculated for each hyperparameter set, and the average $R_c$ score $\bar{R}_c$ was calculated for each incident X-ray intensity. To select optimal hyperparameter values $\Delta \bar{R}_c = (\bar{R}_c$ of SRCNN) − ($\bar{R}_c$ of Experiment) were calculated.

**Table S2.** Searched hyperparameter values based on the experimental design.

| Hyperparameter          | Searched value |
|------------------------|----------------|
| Filter size: $f_1$     | 5, 9, 17       |
| Filter size: $f_2$     | 1, 3           |
| Filter size: $f_3$     | 3, 5, 9        |
| Set of feature map: $n_1$ | 64, 128, 256 |
| Set of feature map: $n_2$ | 32, 64        |

Figure S3 presents the hyperparameter optimization results, where the points above and below the diagonal line indicate that the diffraction image was improved and degraded by the SR model, respectively. The maximum value of $\Delta \bar{R}_c$ was 0.1598 for $f_1 = 5, f_2 = 1, f_3 = 3, n_1 = 256,$ and $n_2 = 64.$ The results represented in the following sections were obtained using these parameter values.
Figure S3. Hyperparameter search results. X: Average $R_c$ scores of the experimental diffraction images. Y: Average $R_c$ scores of the SRCNN diffraction images. Each plotted point corresponds to a different hyperparameter set, whose color denotes a different X-ray incident intensity in photons pulse$^{-1}$ μm$^{-2}$. 
S8. Learning curve of single intensity training

This paper employs early stopping, which is commonly used in the machine learning approach and stops the training on that criterion based on the MSE value. As described in the S6 section, training of the SRCNN network is performed using partial images, i.e., patch extraction. Therefore, the MSE is obtained from the partial image. On the other hand, the $R_c$ value is obtained from the whole image. Consequently, it should be emphasized that the two values are not directly comparable. However, there is some correlation between MSE and $R_c$. For example, the $R_c$ values from the SRCNN images estimated by the initial network model are minimal (0.021 on average compared to 0.063 for experimental images), confirming that it is not functional. On the other hand, in some cases, including shown here, due to sufficient image data provided, both MSE and the $R_c$ converged to a certain degree in the initial epoch stage, and the performance was adequate.

**Figure S4.** The learning curve of the independent training for each different X-ray intensity, in the case of $1 \times 10^{13}$ photons pulse$^{-1}$ $\mu$m$^{-2}$. The solid red line corresponds to the loss of training data. The blue dotted line corresponds to the loss of validation data.
S9. Poisson-NLPCA images

Poisson-NLPCA images are produced using the MATLAB code which is available online\textsuperscript{51}. The input images are the test images described in Sec. S5. The following parameters are used: Patch width=20, nb_axis=4, nb_iterations=20, nb_clusters=14, eps_stop=1e-1, epsilon_cond=1e-3, double_iteration=0, anscombe=0, newton=0, parallel=1. The reader is referred to Ref. 52 for theoretical background of Poisson-NLPCA.

S10. Real-space analysis by the HIO method

Phase retrieval calculation using the HIO method was performed on a 2D diffraction image prepared using the following procedure, respectively. In the HIO method, the Fourier transform is repeatedly performed under real-space and wavenumber-space constraints. In a noise-free ideal diffraction image that satisfies the oversampling condition, the phase is completely obtained in wavenumber space. The real image is reconstructed by performing the above iterative calculations. However, when applied to an actual experimental diffraction image, the convergence of the solution is affected by the sphericity of the Ewald sphere and the influence of noise.

The SRCNN model for upscaling diffraction images was trained using the experimental diffraction images, which did not satisfy the phase retrieval condition. The EMV images created as GTHR images at $\sigma = 2$ satisfied the phase retrieval condition as supervisory data. The experimental diffraction images with dimensions of 55×55 pixels at $\sigma = 1$ were upscaled to 110×110 pixels at $\sigma = 2$ using nearest neighborhood interpolation while maintaining the patterns, and phase retrieval was performed using the upscaled images as a comparison. On the other hand, the SRCNN image was predicted by the trained SRCNN model using $\sigma = 2$ EMV images as a GTHR and upscaled experimental images with dimensions of 110×110 pixels. A
phase retrieval calculation was performed using an SRCNN image as the predicted HR image. Each phase retrieval calculation set the support size to $55 \times 55$ pixels in the image center, and 10,000 iterations were performed. The real image with the minimum wavenumber space error was adopted as the reconstructed real-space image. Considering the initial phase dependence of the convergent solution, 100 phase retrieval calculations were performed by changing the initial phase for investigation.
S11. Generalization achievable by the SRCNN model with respect to molecular species

Figure S5. Improvement of the diffraction images of different molecular species by the constructed SRCNN model. From left to right: 70S ribosome, 80S ribosome, and β-galactosidase. Top row: similarity improvement. X: $R_c$ score of the experiment; Y: $R_c$ score of SRCNN. The EMV, SRCNN, and experimental diffraction images are also shown from top to
We tested the constructed SRCNN model for different molecular species. The SRCNN model described in section S6 was trained using diffraction images of the 70S ribosome. After that, we applied the same model to the experimental diffraction images of other molecular species, namely β-galactosidase and 80S ribosomes. Figure S5 reveals that the diffraction images were improved for all molecular species at incident X-ray intensities with almost all molecular orientations. In particular, the application to the 80S ribosome yielded an improved diffraction image compared to that of the 70S ribosome. In contrast, molecular orientation dependency was observed in the diffraction image improvement of β-galactosidase, and $R_e$ was relatively broadly distributed at each intensity because β-galactosidase had a flat elliptical sphere. Further improvement of the generalization performance for molecular species is a subject of future studies. Because the speckle size of the diffraction image was inversely proportional to the molecular size, some ingenuity was required to perform training of mixed molecular species data of various sizes. For example, it was conceivable to study diffraction images using images standardized by the reciprocal of the molecular size; alternatively, it was possible to perform training in which the molecular size was given as an additional input parameter of the model.