Single-image super-resolution improvement of X-ray single-particle diffraction images using convolutional neural network

Atsushi A. Tokuhisa‡,~,, Yoshinobu Y. Akinaga‡,~,&, Kei K. Terayama+,∥, Yasushi Y. 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 2 1 31, Kita-ku, Osaka 530-0003, Japan

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

∥RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihombashi, Chuo-ku, Tokyo 103-0027, Japan

#Center for Cluster Development and Coordination, Foundation for Biomedical Research and Innovation at Kobe, 6-3-5, Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan

⊥Graduate School of Medicine, Kyoto University, Shogoin-Kawaharacho, Sakyo-ku, Kyoto 606-8507, Japan

KEYWORDS femtosecond X-ray free electron laser; flexible biomolecule; single-particle analysis; structural analysis; super-resolution convolutional neural network

ABSTRACT Femtosecond X-ray pulse lasers are promising probes for elucidating the multi-conformational states of biomolecules because they enable snapshots of single biomolecules to be observed as coherent diffraction images. Multi-image processing using an X-ray free electron laser has proven to be a successful structural analysis method for viruses. However, some difficulties remain in single-particle analysis (SPA) for flexible biomolecules with sizes of 100 nm or less. Owing to the multi-conformational states of biomolecules and the noisy character of diffraction images, diffraction image improvement by multi-image processing is not always effective for such molecules. Here,

a single-image super-resolution (SR) model was constructed using a SR convolutional neural network (SRCNN). Data preparation was performed in silico to consider the actual observation situation with unknown molecular orientations, and fluctuation of molecular structure and incident X-ray intensity. It was demonstrated that the trained SRCNN model improved the single-particle diffraction image quality, which corresponded to an observed image with an incident X-ray intensity; i.e., approximately three to seven times higher than the original X-ray intensity, while retaining the individuality of the diffraction images. The feasibility of SPA for flexible biomolecules with sizes of 100 nm or less was dramatically increased by introducing the SRCNN improvement at the beginning of the variety structural analysis schemes.

1. INTRODUCTION

Femtosecond X-ray free electron lasers (XFELs)[1,2] are attracting attention as new probes for elucidating the multi-conformational states of biomolecules at room temperature[3,4] because they enable observation with low radiation damage of samples and short pulses. Although serial femtosecond crystallography has proven to be an effective method with low radiation damage,[5–7] removing the limitation of sample crystallization in single-particle analysis (SPA) remains challenging. Owing to the low

scattering ability of biomolecules, it is necessary to analyze diffraction images with low signal-to-noise (S/N) ratios in SPA. To obtain better diffracted images for biomolecular imaging applications, various experimental and analytical methods have been proposed,[8–16] such as the use of 100-nm focusing mirrors.[17] Thus, three-dimensional (3D) assembled structures have been reconstructed by processing multiple images obtained from relatively large, hard, and highly symmetric viruses.[18–20] However, for nanoscale flexible biomolecules, which are characterized by multi-conformational states, it is difficult to analyze a dataset of diffraction images from multi-conformational states and to reconstruct a clear 3D assembled structure by multi-image processing techniques. One of the key challenges in this area is realizing the SPA of nanoscale flexible biomolecules with sizes of ≤100 nm, such as ribosomes, nucleosomes, and membrane proteins, which are important for drug discovery.

In SPA, a single biomolecule is injected into a vacuum with an unknown molecular orientation and possibly different conformations for nanoscale flexible biomolecules. When the X-ray hits the sample well within the appropriate beam position, a snapshot of the instantaneous structure of a single biomolecule is acquired as a diffraction image without phase information, which is observed by a two-dimensional (2D) charge-coupled device (CCD) detector as pixel-wise information of the photon counts of a solid angle.

By discretizing in units smaller than a solid angle of the reciprocal of the molecular size in wave number space, which is called oversampling ratio, the phase information is retrieved by an algorithm such as the hybrid-input–output (HIO) algorithm.[21] However, the curvature of the Ewald sphere and noise affect the convergence of the phase retrieval calculation. In practical experiments, photons with detector-beam-sample relative geometry corresponding to a small solid angle 1/2 or less of the reciprocal of the molecular size are commonly measured. The high oversampling ratios required for phase recovery have resulted in a significant reduction in counting photons per pixel. If the requirement of high oversampling ratios is relaxed and observation is allowed at a larger solid angle, the number of achievable photons per pixel will increase significantly. Moreover, a noisy 2D diffraction image obtained by a single measurement is not sufficient to fully construct a 3D assembled structure, and in multiple measurements in nearly the same conformation, all possible molecular orientations are required for 3D assembled structure reconstruction. In the conventional scheme,[22] multiple diffraction images are used for S/N ratio improvement, orientation recovery, and phase recovery to reconstruct the 3D assembled structure. Moreover, because the multi-image processing technique loses the individuality of diffraction images that reflect the various conformational states of flexible biomolecules, only the average structure of the major

conformational states will be obtained. It is essential to elucidate the multi-conformational states of nanoscale flexible biomolecules expressed in individual noisy 2D diffraction images.

In image processing, in addition to general image complementation methods, super-resolution (SR) imaging, which estimates high resolution (HR) images from low resolution (LR) images, has been proposed. Two typical SR imaging strategies have been established: SR image reconstruction and example-based SR image reconstruction.[23] The former is often used as a multi-image processing technique, in which a HR image is composed of a set of LR images. The latter is a method of estimating a HR image by searching the database prepared in advance and is often used as a single-image SR approach. In recent years, research on single-image SR using convolutional neural networks (CNNs) has remarkably advanced. Dong et al.[24] developed a method called Super-Resolution Convolutional Neural Network (SRCNN), which employed CNN layers, and succeeded in achieving SR with higher accuracy than previous methods. Since SRCNN, various methods have been proposed, such as deeper network architectures,[25,26] methods for speeding-up,[27] improvements on loss functions such as perceptual loss and training strategies,[28] and generative adversarial network (GAN)-based methods.[29,30] Most of these methods are based on the framework of supervised learning, where an artificially

reduced LR image is created from a given grand truth high resolution (GTHR) image, and the system is trained to recover the original GTHR image from the LR image. Although these proposed methods have improved the performance of SR, it has been pointed out that neural network-based methods, especially GANs, do not necessarily reproduce the original image and generate artifacts.[31,32] Recently, unsupervised or weakly unsupervised learning,[33,34] where SR models are trained without corresponding HR images, or SR on few images,[35] have also been proposed.

In the 3D assembled structure reconstruction of viruses,[18–20] improvement of the S/N ratio of the diffraction images and orientation recovery were realized simultaneously by using the expand-maximize-compression algorithm,[36–38] which is a statistical algorithm for maximizing posterior probability as a multi-image processing technique. Pioneering techniques for X-ray diffraction image denoising were devised by Jin et al.,[39] whose method based on Poisson image denoising was previously applied to single-particle X-ray diffraction imaging, and piecewise principal component analysis was applied to the simulated single-particle X-ray diffraction image with an elongation factor of 2; the effect of denoising could be recognized by filtering the images with a low calculation cost. As conventional 3D reconstruction methods use many images to estimate an assembled 3D structure with resolution higher than that provided by a single 2D diffraction image, and

SRCNN has shown promising results in computer vision for a variety of single images, single-image SRCNN may be effective in single-particle biomolecular analysis. It is expected to improve the quality of single 2D diffraction images from LR to HR, and simultaneously solve the problems of noisy observation data and multi-conformational states of nanoscale flexible biomolecules. However, a practical problem is incomplete and chaotic experimental data with unknown molecular orientations exist from different conformations for nanoscale flexible biomolecules, which makes CNN training difficult. Preparation of large-scale annotated data in real-world space is essential for supervised learning.

This study is the first investigation of the possibility of using single-image SR techniques for X-ray diffraction image improvement to overcome the difficulties associated with the SPA of flexible biomolecules with sizes of 100 nm or less. To compensate for the incomplete and chaotic experimental data describing nanoscale flexible biomolecules, we propose using synthetically-generated labeled data in silico by using molecular dynamics simulations and virtual X-ray diffraction image simulator. These virtual diffraction image data enable us to employ various SR methods based on the supervised learning framework. To avoid excessive artifacts here,[31,32] we use the SRCNN model, which is a relatively simple network. Training network models using labeled synthetic data has been adapted

to a variety of methods and has been reported to improve performance, increase generalization performance, and reduce the amount of real-world data used in training.[40–42] If an improved diffraction image is obtained from an individual noisy diffraction image by the trained SRCNN model in silico, the applicability of SPA to nanoscale flexible biomolecules is dramatically increased. For example, the number of necessary diffraction images for the 3D assembled structure reconstruction is reduced. Moreover, the multi-conformational states of nanoscale flexible biomolecules is evaluated by using improved individual 2D diffraction images, which reflect the instantaneous structures of biomolecules, as input data of a pipeline for flexible 100-nm biomolecule structural analysis schemes, such as the HIO method for 2D real-image reconstruction and the diffraction template matching method for estimating plausible 3D structural models (see Figure 1).[43,44]

In the diffraction template matching method, we retrieve a plausible 3D structural model based on a single noisy 2D diffraction pattern by data assimilation concept. The basic idea is to compensate for the incompleteness of a single diffraction image to reconstruct a 3D structure in virtual space. First, various candidate structural models will be prepared by "conformational sampling." Second, "image matching" will be conducted using candidate diffraction images simulated in possible molecular orientations from each

candidate structural model. Third, a “plausible” 3D structural model will be selected according to image similarity between the experimental diffraction image and syncretic candidate diffraction images using a similarity detection algorithm. In this method, the estimation accuracy depends on the improvement of the similarity detection of a pair of diffraction images. Improving similarity detection performance is one of the most important factors contributing to the attainable structural resolution. Then, we focus on the demonstration of two key technologies, the similarity detection, and the phase recovery to verify the effectiveness of SRCNN in contributing to multi conformational analysis here.

Section 2 describes methods for constructing SRCNN models with data preparation in silico and for quantitative evaluation of image improvement. Section 3 presents the main results of the diffraction image improvement achieved by the constructed SRCNN model, effects of the structural and incident X-ray intensity fluctuations on diffraction image improvement, real image analysis by a phase retrieval algorithm, and generalization achievable by the constructed SRCNN model. Section 4 discusses the main contributions and the limitations of this work and topics for future research.

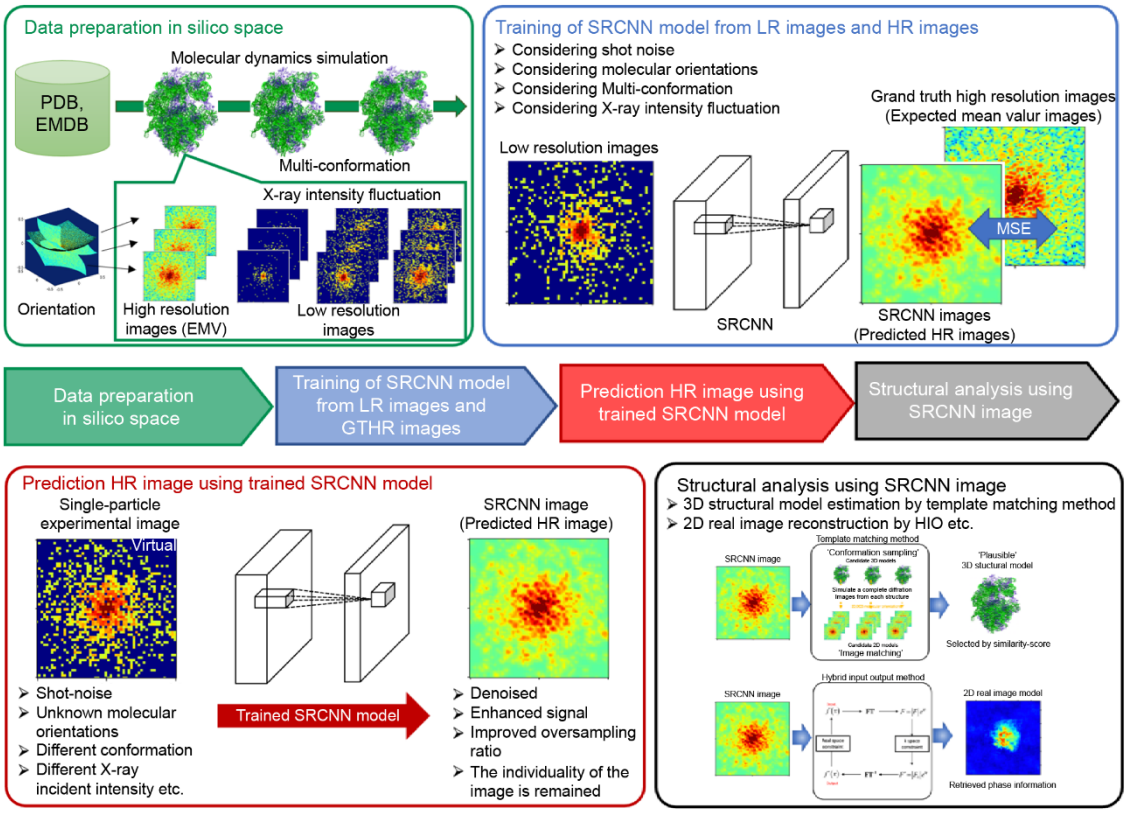

Figure 1. Schematic of a pipeline for 100-nm biomolecule structural analysis with single-image SRCNN. Data preparation with synthetically generated labeled data between grand truth high-resolution (GTHR) images and low-resolution (LR) images are performed in silico space using molecular dynamics simulation and diffraction image simulator. Training of the SRCNN model is performed using GTHR images and LR images prepared in silico space considering shot noise, molecular orientations, multi-conformation, and X-ray intensity fluctuations. A SRCNN image will be predicted using a trained SRCNN model from the single-particle experimental image. Many kinds of structural analysis could be performed using the SRCNN image, such as the template matching method and

the HIO method.

2. METHODS

To avoid a practical problem for supervised learning of data labeling between LR images and HR images, data preparation in silico was actively used here. Two types of synthetic diffraction images, the expected mean value (EMV) image as the GTHR images and experiment image as the LR images, are prepared using molecular dynamics simulation and a diffraction image simulator. Through this ingenuity, we constructed a trained SRCNN model that significantly improved the X-ray diffraction image quality for nanoscale flexible biomolecules. Then, we conducted studies on X-ray diffraction image improvement for ribosome molecules as a test case. A mixed dataset considering realistic experimental conditions,[45,46] incident X-ray intensity fluctuations, structural fluctuations, and various molecular orientations during observation was created in silico and data labeling and training was performed by linking with the EMV image and the experiment diffraction images (virtual). Supervised learning depended on the training data. For example, in a real experiment, the distance between the sample and detector might change. If one wanted to consider distance fluctuations, it was better to prepare images with different oversampling ratios in advance. Moreover, Gaussian noise from the detector might be added. Depending on the individual experimental situation, it was also possible

to incorporate factors other than those considered here when preparing data in silico. Additionally, the present training model was suitable as a pre-training network model for advance training more detailed individual experimental data and to train other molecule species. Biomolecules consisted of 20 types of amino acids and had several typical packing patterns. When we looked at a small patch of the diffraction image, we saw the similarity of the speckle pattern. When implementing patch learning as described later we expected some generalization performance, which would be specifically investigated in the section on generalization performance.

2.1. Data preparation and image dataset

Here, all diffraction image datasets were prepared by simulation as the following procedure. We simulated two types of X-ray diffraction images: EMV images as GTHR images and virtual experimental diffraction images as LR images, in which the quantum noise effect was considered. The latter were termed the experimental images. These diffraction images were prepared using the diffraction image simulator considering the curvature of the Ewald sphere described below using the structural model from the structural database of PDB. The EMV of photons arriving at a detector pixel of s(k), were as follows:

$$s(\mathbf{k}) = s(\mathbf{Ah}) = s(\mathbf{A}; \xi, \varphi) = I_i r_{ce}^2 \omega(\xi) |F(\mathbf{k})|^2. \quad (1)$$

Here, I_i was the incident X-ray intensity and was expressed in photons pulse^{-1} μm^{-2}, r_{ce} was the classical electron radius, and F (**k**) was the structure factor. The **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 × 3 matrix describing the molecular orientation of the sample molecule to the incident X-ray beam, which was given in the Eulerian angle (α, β, γ) by the following expression:

$$\mathbf{A} = \mathbf{A_3}(-\Upsilon)\mathbf{A_2}(-\beta)\mathbf{A_3}(-\alpha) \quad (2)$$

$$\mathbf{A_3}(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}, \mathbf{A_2}(\beta) = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \quad (3)$$

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

$$\mathbf{h}(\xi, \varphi) = \lambda^{-1}[\mathbf{E} - \mathbf{A}_3(\varphi + \pi)\mathbf{A}_2(\xi)]\mathbf{e}_3$$

$$= \lambda^{-1}(\sin\xi\cos\varphi, \sin\xi\sin\varphi, 1 - \cos\xi)^{\mathrm{t}}, \mathbf{e}_3 = (0,0,1)^{\mathrm{t}}, \quad (4)$$

where (ξ, φ) was a polar coordinate. Here, **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 ξ and the incident X-ray wavelength of λ, and the wave number k = |**k**| Å^{-1} was given by the following equation using $\xi = 2\theta$, where θ was the scattering angle:

$$k = \frac{2}{\lambda}\sin\frac{\xi}{2}. \quad (5)$$

ω was the solid angle per pixel and is expressed by the following equation using the incident X-ray wavelength λ, molecular length L, scattering angle ξ, and linear oversampling ratio σ:

$$\omega(\xi) = \left(\frac{\lambda}{\sigma L}\right)^2 \cos^3\xi. \quad (6)$$

The EMV diffraction images were simulated with the EMV of $s(\mathbf{k})$ itself without any noise effect. For the experiment diffraction images, we simulated the probability of an integral number of photons given in the pixel by replacing the function $s(\mathbf{k})$ with a stochastic function $s_Q(\mathbf{k})$ that assumed only integral value according to the Poisson distribution, because in real experiments, what is measured is an integral number of photons, given by the quantum mechanical probability.

In fact, here, λ = 1 Å corresponds to X-ray energy with 12.4 keV, σ = 1 or 2, and L = 270 Å. Additionally, the maximum wave number of each simulated diffraction image was set to 0.2 Å^{-1}, and the diffraction images corresponding to a 5 Å structural resolution were calculated. The EMV diffraction images from 5,120 different 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 experiment diffraction images were prepared by adding Poisson noise corresponding to the scattered

X-ray intensity to the EMV diffraction images. Noise sources that presented in the actual experiment diffraction images, such as inelastic scattering, beam stopping, and detector reading noise, were not considered here.

Considering the wide dynamic range of the diffraction images, as well as possible zero values at some pixels, we added one to each pixel value and took the logarithm as a preprocessing to obtain tractable numbers. The SRCNN output data were post-processed to obtain a normal diffraction image by inverse conversion.

For systematic model training and evaluation, we employed several subsets of the virtual images. The experiment images and EMV images were divided into three non-overlapping subsets: the training set was used to update network model weight factors through back-propagation during the iteration, the validation set was used to monitor the loss function in each epoch and stop the iteration, and the test set was used to evaluate the trained model based on the R_c score, which is described in detail in a later section. We adopted hold-out validation with a ratio of training data to validation data of 2:1. Details of these data sets are as follows:

Training data: EMV images as the GTHR images were created from a set of $N = 640$ molecular orientations for each molecular structure. Considering structural fluctuations, diffraction images of N molecular orientations were prepared for each conformation with

S types; i.e., we prepared $N \times S$ EMV images in total. Meanwhile, we prepared $N \times S \times I \times 5$ experiment images as LR images, where I was the number of X-ray intensities and 5 the number of randomly generated noise patterns.

Validation data: The validation data, which were used to monitor the training process of each epoch, were prepared similar to the training data, with another set of $N = 320$ molecular orientations.

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

In training the SRCNN model, partial images with 44 × 44 pixels were cropped from the training and validation images. From the training images, $K = 64$ patches were periodically taken from each image. For the validation images, $K = 30$ partial images were randomly selected from each image.

Here, we trained the SRCNN model using various combinations of training and validation sets. These combinations of training/validation sets were indexed by Dataset ID. Table 1 lists 18 datasets employed here. Dataset 18 was used for hyper-parameter optimization.

2.2. Adopted network model and SRCNN training

The SRCNN technique was a method that achieved high accuracy by replacing dictionary-based SR with a CNN. Figure 2 provides a diagram of an SRCNN. The network was composed of three convolution layers. We examined two different architectures: the single-channel (1ch) and the two-channel (2ch) models. The 1ch model took gray-scale diffraction image data as an input and output another gray-scale image. The 2ch model took a two-channel image as an input. In the first channel in the input, a gray-scale image data was stored. In the second channel, the incident X-ray intensity value was stored. While there were several ways to define the value of input channel 2 in the 2ch model, we used I_i in Eq. (1) here. Our preliminary calculations revealed that the 1ch model exhibited larger losses than the 2ch model and the loss function of the validation data tended to remain high after the loss function of the training data became small. Therefore, we employed the 2ch model here. 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." The network was implemented using Keras (ver2.2.4),[47] which was a framework for constructing deep

learning models.

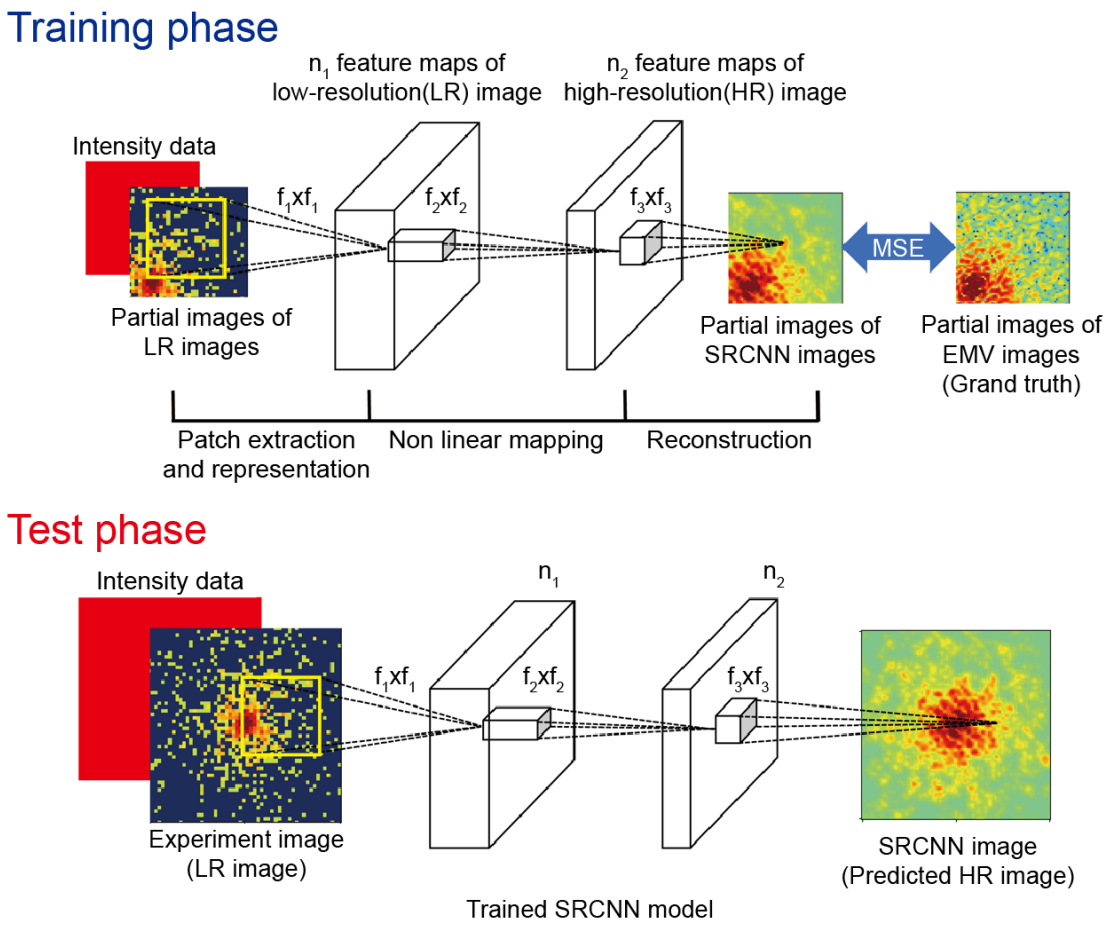

Figure 2. Two-channel SRCNN model. In the training phase, the partial images extracted from a whole diffraction image are used as data-augmentation. LR images with the incident X-ray intensity data and the EMV images as GTHR images are used for training using mean square error with respect to EMV images for the loss function. In the first layer, patch extraction (yellow square) and representation of the LR image is 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 experiment image is input to the trained SRCNN model together with the intensity data, and the whole SRCNN-predicted HR image will be obtained.

The training/validation/test data were prepared as described above and used to train and evaluate the SRCNN model using the hold-out method. The mean square error (MSE) with respect to the EMV images was adopted for the loss function. For the convolution of the intermediate layer, zero-padding was adopted to ensure that the input and output image sizes were the same, and a rectified linear unit was used for the activation function in the convolution layers. Mini-batch training with a batch size of 128 was performed for up to 100 epochs using the early stopping method. The Adam optimizer[48] with a learning rate of 0.0003 was used for the parameter update. The iteration was terminated when the loss function did not improve five times in a row. Once the iteration was terminated, the parameter set with the lowest loss value of the validation data was adopted as the trained model. In the test phase, an experimental image was given as the input to the trained network model, and the SRCNN image as the predicted HR image was obtained as the output. Then, the obtained SRCNN images were evaluated by the R_c with respect to the corresponding EMV images calculated using the similarity detection algorithm, which was described in detail in the next section.

2.4 Similarity detection algorithm

The similarity of a pair of diffraction images was evaluated using the following similarity determination algorithm[49,50] based on the correlation function.[22,51] We introduced a noise reduction mechanism to reduce the influence of quantum noise by the integral correlation patterns $I_{c,ij}(k,\alpha)$ in $k=\frac{2}{\lambda}\sin\frac{\xi}{2}$, and α. Here, s was the number of observed photons described in Eq. (1), $\bar{s}$ was the expected photon number, N_ξ was the discrete number of pixels on the concentric circles, $\xi = 2\theta$ represented the scattering angle, and λ was the incident X-ray wavelength. The rotation α of the molecule with respect to the incident X-ray axis appeared as a rotation around the center of the diffraction intensity image. By considering α, the correlation pattern was obtained immediately to acquire 360° rotating correlation coefficients to the 2D plane. Here, the correlation pattern c_{ij} (ξ, α) of a pair of diffraction images I and j was defined by the following equations:

$$c_{ij}(\xi,\alpha)=\frac{\Psi_{ij}(\xi,\alpha)}{\bar{s}_i(\xi)\bar{s}_j(\xi)}-1, \tag{4}$$

$$\Psi_{ij}(\xi,\alpha)=\frac{1}{N_\xi}\sum_{l=0}^{N_\xi-1}s_i\left(\xi,\frac{2\pi l}{N_\xi}\right)s_j\left(\xi,\frac{2\pi l}{N_\xi}+\alpha\right), \tag{5}$$

$$\bar{s}_i(\xi)=\frac{1}{N_\xi}\sum_{l=0}^{N_\xi-1}s_i\left(\xi,\frac{2\pi l}{N_\xi}\right). \tag{6}$$

Furthermore, we defined an integral correlation pattern using the following equations:

$$I_{c,ij}\left(k_{up},\alpha\right)=\int_0^{\xi_{up}}i_{c,ij}(\xi,\alpha)\,sin\xi d\xi=\lambda^2\int_0^{k_{up}}i_{c,ij}(k,\alpha)kdk, \tag{7}$$

$$i_{c,ij}(k,\alpha)=\frac{2\pi}{N_\xi}\sum_{\Delta\alpha=-2k_c/k}^{2k_c/k}c_{ij}(\xi,\alpha+\Delta\alpha). \tag{8}$$

By performing normalization with the autocorrelation term of each diffraction image, the similarity of the pair of diffraction images was quantified as the $R_{c,ij}(\hat{k}_{ij}, \hat{\alpha}_{ij})$ score:

$$R_{c,ij}(\hat{k}_{ij}, \hat{\alpha}_{ij}) = \frac{I_{c,ij}(\hat{k}_{ij}, \hat{\alpha}_{ij})}{\sqrt{I_{c,ii}(\hat{k}_{ii}, \hat{\alpha}_{ii})}\sqrt{I_{c,jj}(\hat{k}_{jj}, \hat{\alpha}_{jj})}}. \quad (9)$$

Here, $R_{c,ij}(\hat{k}_{ij}, \hat{\alpha}_{ij})$ was denoted as R_c for simplicity. In particular, when image *j* was the corresponding EMV image of image *i*, we called $R_{c,ij}(\hat{k}_{ij}, \hat{\alpha}_{ij})$ the R_c score of image *i*.

2.5. Gaussian Blur images

Gaussian Blur (GB) images were created by applying the GB filter to the experiment images using the GB function in the OpenCV library. The output image *B* obtained by applying a GB filter $G_s(k,l)$ with kernel $n \times n$ and standard deviation σ to an input image *A* was written as follows:

$$B(i,j) = \sum_{\substack{0 \le |k| \le \frac{n-1}{2}, \\ 0 \le |l| \le \frac{n-1}{2}}} G_\sigma(k,l) A(i+k, j+l) \quad (10)$$

where

$$G_\sigma(k,l) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{k^2+l^2}{2\sigma^2}\right). \quad (11)$$

The kernel size *n* and standard deviation *s* were set to 5 and 1.0, respectively.

2.6. 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. In this method, the Fourier transform was repeatedly performed under real-space and wavenumber space constraints. For an ideal diffraction image that satisfied the oversampling condition, the phase was retrieved in the wavenumber space, and the real image was reconstructed by performing the above iterative calculation. When applied to an actual experimental diffraction image, the convergence of the solution was affected by the sphericity of the Ewald sphere and the influence of noise. The experiment diffraction images with dimensions of 55×55 pixels at σ = 1 were upscaled to 110×110 pixels at σ = 2 using the nearest neighborhood interpolation while maintaining the patterns, and phase retrieval was performed using the upscaled images. The SRCNN image was predicted by the trained SRCNN model using σ = 2 EMV images and upscaled experiment images with dimensions of 110 × 110 pixels. A phase retrieval calculation was performed using an SRCNN image as the predicted HR image. In each phase retrieval calculation, the support size was set to 55 × 55 pixels in the image's center, 10,000 iterations were performed, and 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.

3. RESULTS AND DISCUSSION

3.1. 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 2), 108 combinations of five hyperparameters were examined (see Table 1), for the 2ch model by 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, $\Delta\overline{R_c} = (\overline{R_c}\ \text{of SRCNN}) - (\overline{R_c}\ \text{of Experiment})$ was calculated.

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

Table 1. Searched hyperparameter values based on the experimental design.

Figure S1 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 $\Delta\overline{R_c}$ was 0.1598 for f_1 = 5, f_2 = 1, f_3 = 3, n_1 = 256, and n_2 = 64. All the results represented in the following sections were obtained using these parameter values.

3.2. Diffraction image improvement by SRCNN

To evaluate the diffraction image improvement achievable by using an SRCNN, training was conducted using various training datasets with different molecular orientations and incident X-ray intensities (A: 5×10^{12}, C: 1×10^{13}, E: 5×10^{13}, F: 1×10^{14} photons pulse^{-1} μm^{-2}, corresponding to datasets 1, 3, 5, and 7, respectively, in Table 2). As an example, the learning curve for the intensity of 1×10^{13} photons pulse^{-1} is shown in Figure S2, which confirms that the loss function sufficiently converges during both validation and training. Figure 3 presents the results obtained in the test cases at four different incident X-ray intensities. The experiment, SRCNN, EMV, and GB images were shown at each incident X-ray intensity. Here, the images created in the simulation considering the quantum shot-noise effect were called the experiment images. Because the diffraction X-ray intensity was a function of the wavenumber with considerable attenuation (i.e., it

took a very wide range of values with a long tail), all diffraction images were displayed on a log scale. Focusing on the high wavenumber region, the experiment diffraction images had zero or single photon counts for most pixels because of insufficient diffraction intensity. Moreover, in the experiment images, the continuous speckle pattern could not be observed in most areas, except for the central regions of the images. However, in the diffraction images improved by the SRCNN, continuous speckle patterns could be seen from near the centers of the images, where the diffraction intensity was relatively high in the middle depending on the intensity. Moreover, in the GB images, continuous speckle patterns were observed near the images' centers. However, the speckle patterns in the GB images appeared to be different from those in the EMV images as GTHR images. This suggested that a simple image processing of Gaussian blur was not sufficient for structural analysis.

Data set	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
16	5, 10/5, 10/0	640/320/320	E, F/E, F/E, F	819,200/192,000
17	0/0/0	640/320/320	A, C, E/A, C, E/A, B, C, D, E, F	614,400/144,000
18	0/0/0	300/60/100	A, C, F, G/A, C, F, G/A, B, C, D, E, F,	384,000/36,000

			G	

A: 5×10^{12}, B: 8×10^{12}, C: 1×10^{13}, D: 3×10^{13}, E: 5×10^{13}, F: 1×10^{14}, G: 1×10^{15}

Table 2. Diffraction image datasets for the training/validation/test data were prepared for the SRCNN model. We prepared a variety of annotated synthetic datasets for training/validation/testing. Three different conformations with root mean square deviation (RMSD) of 0, 5, 10 Å were prepared. 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: 5×10^{12}, B: 8×10^{12}, C: 1×10^{13}, D: 3×10^{13}, E: 5×10^{13}, F: 1×10^{14}, and G: 1×10^{15} photons pulse^{-1} μm^{-2} were prepared. We also show the number of partial images for training/validation.

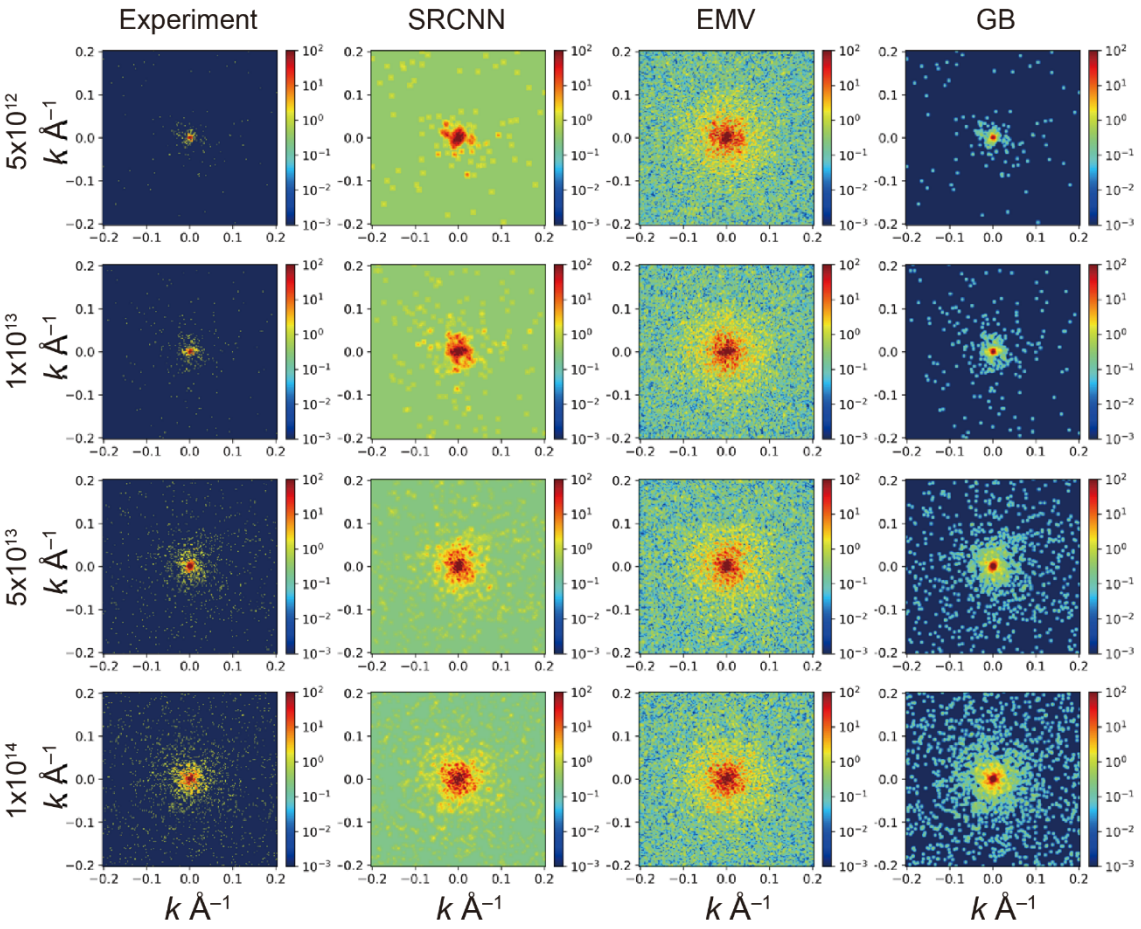

Figure 3. Diffraction images at incident X-ray intensities of 5×10^{12}, 1×10^{13}, 5×10^{13}, and 1×10^{14} photons pulse^{-1} μm^{-2} and in different molecular orientations. The experiment images, images improved by the SRCNN, EMV images, and GB images are shown, from left to right.

To evaluate the degree of diffraction image improvement quantitatively, the integral correlation pattern based on an EMV image was calculated using the similarity detection algorithm[49,50] and Figure 4 displays the results. Comparing the integral correlation patterns of the experiment (i.e., between the experiment and EMV images) and SRCNN

images (i.e., between the SRCNN and EMV images), a high correlation line clearly appeared in the SRCNN results for all incident intensities. The R_c score, which corresponded to the similarity of a pair of diffraction images described in Section 2.4.1, was evaluated. The results indicated that the R_c score of SRCNN was higher than that of the experiment images at all intensities: 5×10^{12} (R_c of experiment: 0.0387, R_c of SRCNN: 0.265, R_c of GB: 0.0898), 1×10^{13} (R_c of experiment: 0.0830, R_c of SRCNN: 0.327, R_c of GB: 0.159), 5×10^{13} (R_c of experiment: 0.242, R_c of SRCNN: 0.491, R_c of GB: 0.356), and 1×10^{14} (R_c of experiment: 0.348, R_c of SRCNN: 0.588, R_c of GB: 0.479). This finding indicated that the improved diffraction images obtained by using the SRCNN had patterns more similar to the corresponding EMV images than the experiment images to the EMV images. Focusing on the GB improvement, the R_c scores of the GB images were higher than those of the experiment images. However, they were lower than those of the SRCNN images, and the degree of improvement was not high compared to that achieved using the SRCNN.

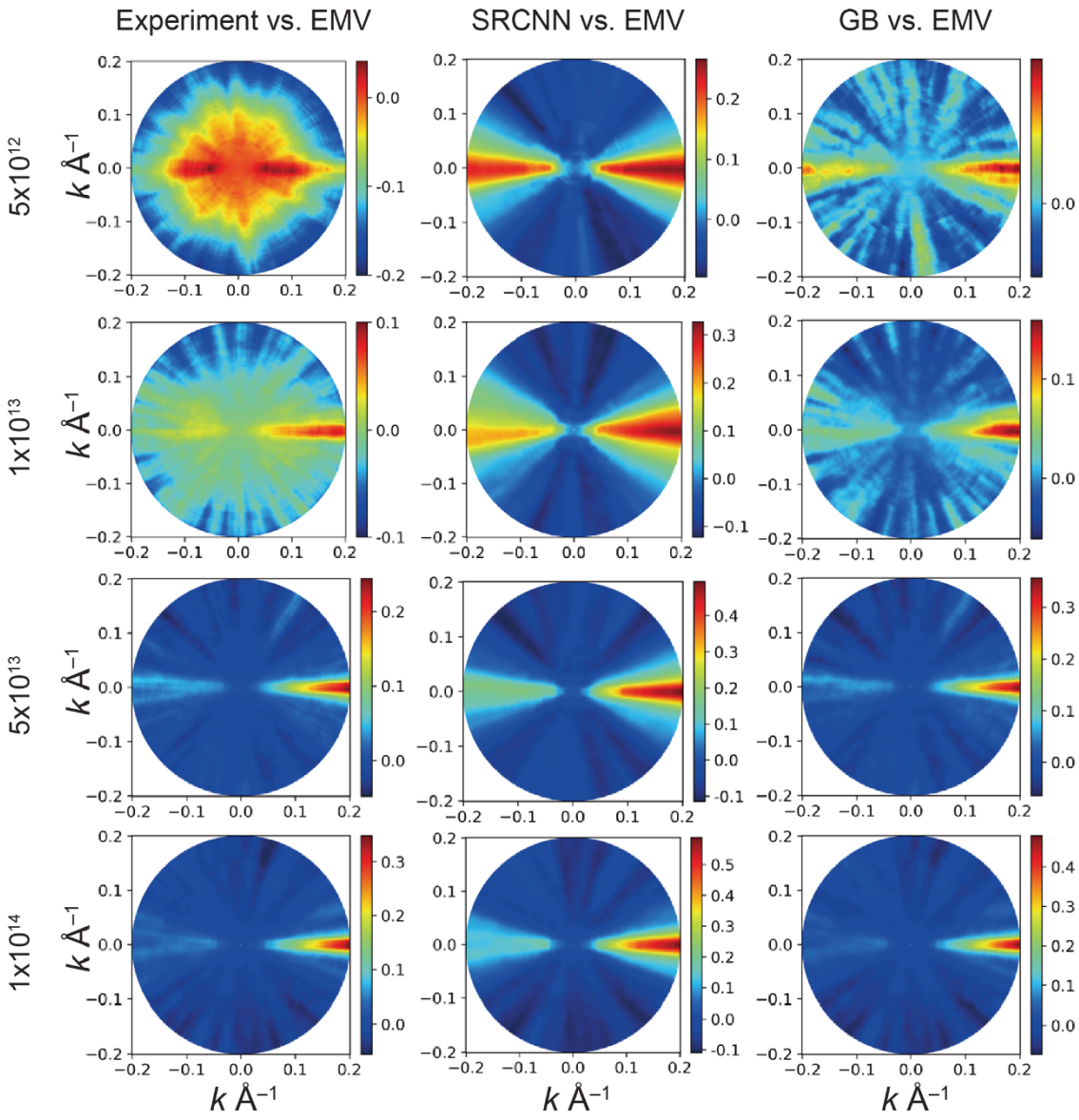

Figure 4. Integral correlation patterns at various incident X-ray intensities, from those above 5×10^{12}, 1×10^{13}, 5×10^{13}, and 1×10^{14} photons pulse^{-1} μm^{-2}, and in different molecular orientations between diffraction images. The experiment vs. EMV, SRCNN vs. EMV, and GB vs. EMV images are shown from left to right.

Next, to investigate the molecular orientation dependency of the image improvement,

a test was performed on diffraction images of 320 orientations for each intensity using the trained SRCNN model. Figure 5 presents the results for four different incident X-ray intensities (A: 5×10^{12}, C: 1×10^{13}, E: 5×10^{13}, and F: 1×10^{14} photons pulse^{-1} μm^{-2}, corresponding to datasets 1, 3, 5, and 7, respectively, in Table 2). Figure 5-A shows the R_c scores of the SRCNN images were all higher than those of the experiment images. In addition, the R_c scores of the SRCNN images were higher than those of the GB images (Figure 5-B). From these results, it was confirmed that the present SRCNN model was effective for improving diffraction images with various molecular orientations.

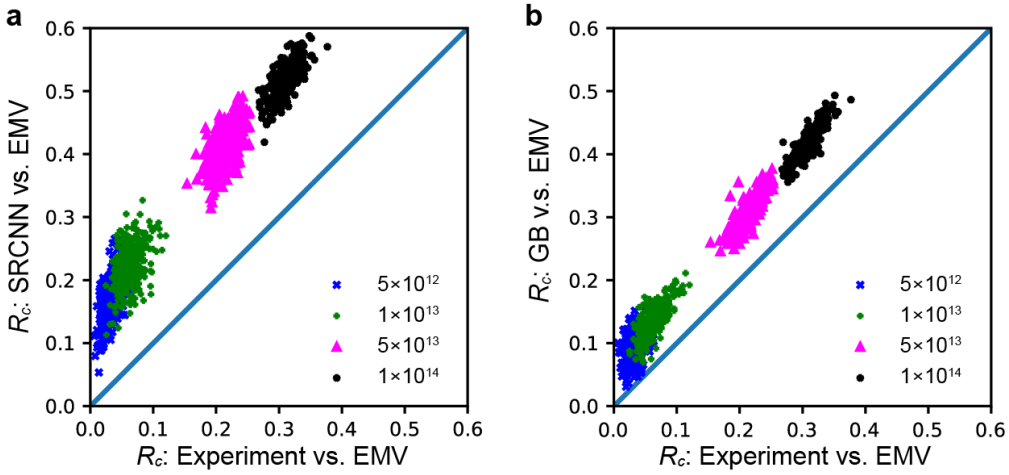

Figure 5. (A) Similarity improvement by SRCNN model. X: R_c score of experiment. Y: R_c score of SRCNN. (B) Similarity improvement by GB model. X: R_c score of experiment. Y: R_c score of experiment.

3.3. Effects of structural and incident X-ray intensity fluctuations on diffraction image improvement

We investigated the effects of the structural fluctuations of the sample biomolecules and incident X-ray intensity fluctuations, both of which should be considered in single-particle experiments, on the diffraction image improvement achieved using the SRCNN. First, we studied the effects of structural fluctuations on the diffraction image improvement. Structural-mixed training was performed using the diffraction image datasets (datasets 2, 4, 6, and 8 in Table 2) that involved two types of structures that differed from the reference structure by approximately 5 and 10 Å in root-mean-square deviation (RMSD). Subsequently, a test was performed using diffraction images of the reference structure (RMSD = 0 Å) that were not used in the training. Figures 5-A and 5-B show the results of the structurally mixed training at incident X-ray intensities of 5×10^{12} photons pulse^{-1} μm^{-2} and 1×10^{13} photons pulse^{-1} μm^{-2}. Structurally mixed training was conducted at all intensities, and the R_c score of the SRCNN was higher than that of the experiment. Thus, it was concluded that the constructed SRCNN model was tolerant of structural fluctuations to some extent at a resolution of 5 Å.

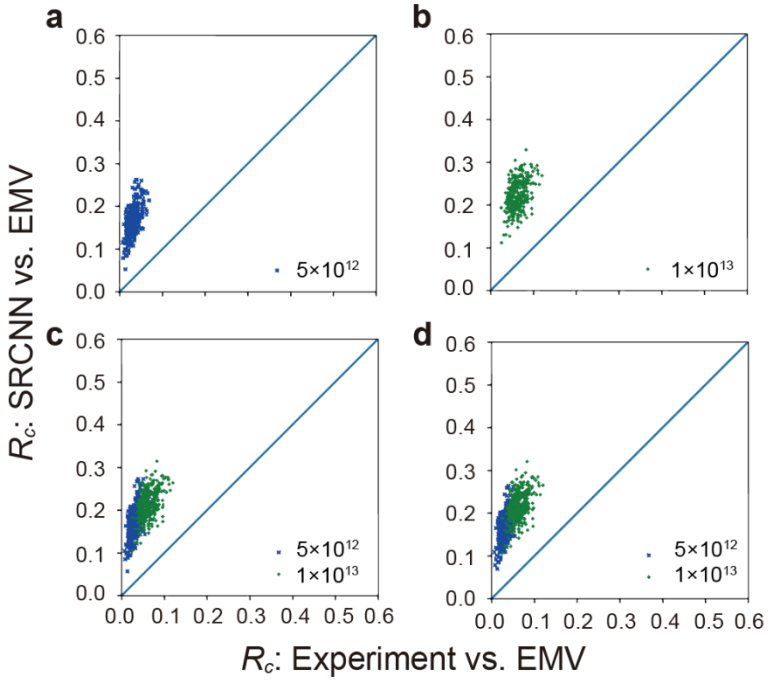

Figure 6. Results of learning considering structural and incident X-ray intensity fluctuations. X: R_c score of experiment. Y: R_c score of SRCNN. (A) Test of structural fluctuation using dataset 2, which contained mixed data of RMSD = 5 and 10 Å to improve the diffraction image of RMSD = 0 Å at 5×10^{12} photons pulse^{-1} μm^{-2}. (B) Test of structural fluctuation using dataset 4, which contained mixed data having RMSD = 5 and 10 Å, to improve the diffraction image of RMSD = 0 Å at 5×10^{13} photons pulse^{-1} μm^{-2}. (C) Test of incident X-ray intensity fluctuation using dataset 13, which was a mixed dataset of diffraction images at 5×10^{12} and 1×10^{13} photons pulse^{-1} μm^{-2}. (D) Test of structural and intensity fluctuations using dataset 14, which was a mixed dataset of RMSD = 5 and 10 Å to improve the diffraction image of an RMSD to 0 Å at 5×10^{12} photons

pulse^{-1} μm^{-2} and 5×10^{13} photons pulse^{-1} μm^{-2}.

Next, we investigated the effects of the incident X-ray intensity fluctuations on the diffraction image improvement. X-ray intensity-mixed training was performed using the diffraction image datasets (datasets 9, 11, 13, and 15 in Table 2) that differed in incident X-ray intensity by at most three times. Figure 6-C presents the results of tests performed with two different intensities using a mixed-intensity SRCNN training model. At each intensity, the R_c scores with all 320 different orientations were improved. The same tendency was observed in the training of mixed intensity data at other intensities (datasets 9, 11, and 15 in Table 2). The above results indicated that the SRCNN model constructed in the presence of various incident X-ray intensities was effective for improving images regardless of the molecular orientation.

Finally, we investigated the effects of a more realistic experiment situation in which both structural and incident X-ray intensity fluctuations existed simultaneously. The structural/incident X-ray intensity fluctuation mixing diffraction image datasets (datasets 10, 12, 14, and 16 in Table 2) were used to perform the structural/incident X-ray intensity-mixing training. Figure 6-D depicted the results of testing for two different intensities using a mixed training model. At all intensities, the R_c scores were improved at all molecular orientations of the reference structure, which was not used in the training.

In summary, an effective single-image SRCNN model was constructed in the presence of structural and incident X-ray intensity fluctuations by using simulation data that reflect the actual situation of single-particle experiments.

3.4. Improvement of intermediate-intensity diffraction images not included in training

Because the incident X-ray intensity continuously changed by up to three orders of magnitude in an actual experiment, this section addresses whether the SRCNN model constructed for an intermediate-intensity diffraction image not included in the training data was effective. The intermediate-intensity diffraction image improvement was performed using the incident X-ray mixed-intensity dataset that did not include some intensity data (dataset 18 in Table 2), which mimicked the actual XFEL experimental situation. The training data included intensities A: 5×10^{12}, C: 1×10^{13}, and E: 5×10^{13} photons pulse^{-1} μm^{-2} but did not include intensities B: 8×10^{12}, D: 3×10^{13}, and F: 1×10^{14} photons pulse^{-1} μm^{-2}. Figure 7 shows the results of testing all incident X-ray intensities using a training model in which missing mixed-intensity data training was performed. The images with intensities B, D, and E, which were not involved in the training data, were improved by the trained SRCNN model to the same extent as those with intensities A, C, and E. Thus, the constructed SRCNN model was effective for improving intermediate-intensity diffraction images not included in the training. We demonstrated

that the SRCNN model constructed in this manner was effective for X-ray diffraction image improvement under actual experimental conditions.

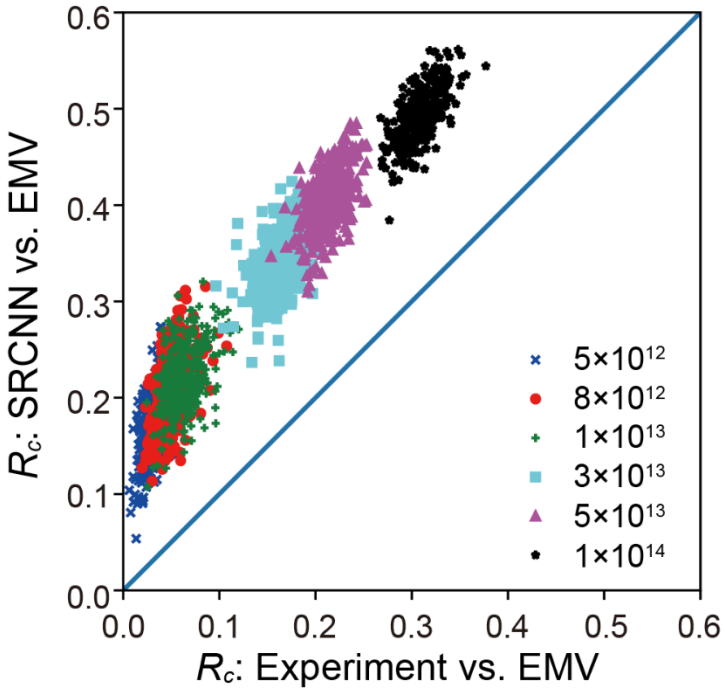

Figure 7 Similarity improvement of the intermediate X-ray intensity data set not included in the training. X: R_c score of experiment. Y: R_c score of SRCNN. The X-ray incident intensity contained in the training data set is 5×10^{12}, 1×10^{13}, and 5×10^{13} photons pulse^{-1} μm^{-2}, whereas that not contained in the training data set is 8×10^{12}, 3×10^{13}, and 1×10^{14} photons pulse^{-1} μm^{-2}.

3.5. Conversion of diffraction image improvement into incident X-ray intensity

The trained SRCNN model's performance was evaluated by converting the diffraction image improvement into a comparable incident X-ray intensity. Figure 7 shows the

constructed SRCNN model increased R_c for various incident X-ray intensities. The increase in R_c indicated that the noise in the observed diffraction images was reduced by the SRCNN and that the improved images were closer to the corresponding EMV images. The degree of noise in the observed diffraction images depended on the diffraction image intensity, which in turn depended on the molecular size and incident X-ray intensity.[52] The following analysis was conducted to convert the increase in R_c into an increase in incident X-ray intensity. The averages and variances of the R_c scores of the experiment images were calculated for 320 different molecular orientations at various incident X-ray intensities. Then, the R_c versus incident X-ray intensity conversion curve was constructed (see Figure 8-A), which also included the standard deviation with error bars centered on the mean R_c score of the experiment. For the R_c versus incident X-ray intensity conversion curve, the regression curve was obtained by fourth-order polynomial fitting using the function $f(x) = ax^4 + bx^3 + cx^2 + dx + e$. The fitting parameters were obtained as $a = -0.011959$, $b = 0.64769$, $c = -13.042$, $d = 115.95$, and $e = -384.65$.

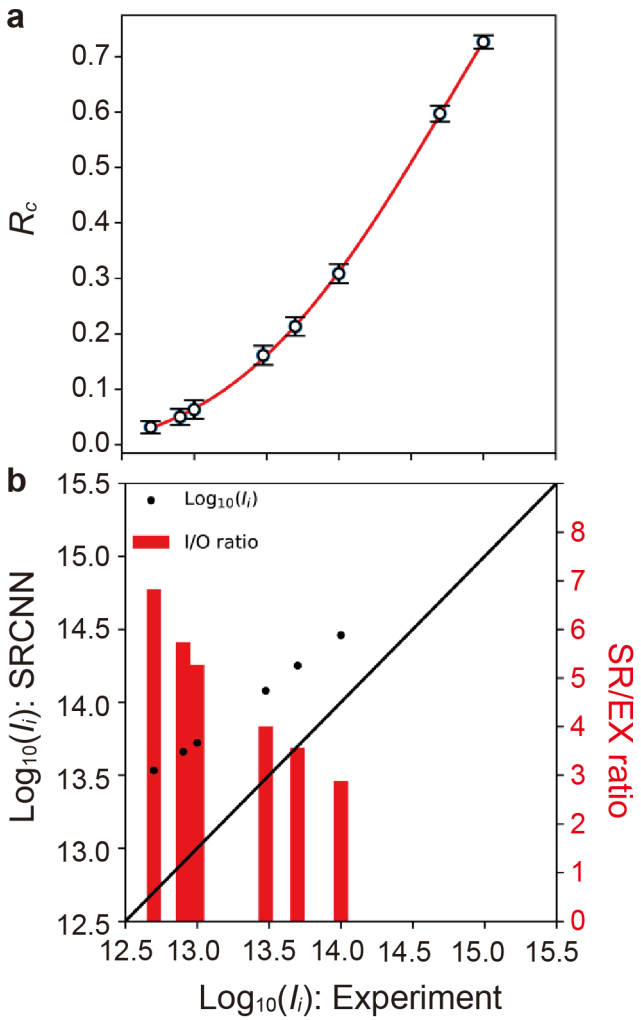

Figure 8. (A) Conversion between R_c vs. incident X-ray intensity with I_i on a log scale. X: $\mathrm{Log}_{10}(I_i)$. Y: R_c score of SRCNN. (B) Improvement ratio of the incident X-ray intensity between the SRCNN and Experimental results. X: $\mathrm{Log}_{10}(I_i)$: Experiment; Y_1: $\mathrm{Log}_{10}(I_i)$: SRCNN; Y_2: SR/EX ratio, which is I_i: SRCNN/I_i: Experiment.

Using the X-ray mixed-intensity training results (see Figure 7), $\bar{R}_c$, which was the average R_c obtained using the SRCNN, was calculated for each intensity. Subsequently, $\bar{R}_c$ was converted into the incident X-ray intensity using the R_c versus incident X-ray

intensity conversion curve. Figure 8-B presents the resulting improvement ratios with black dots and I_i of SRCNN/I_i of the experiment = SR/EX shown by the red bar as a function of the X-ray intensity of the experiment image. According to SR/EX, the lower the incident X-ray intensity, the higher the improvement rate. Focusing on the intensities of 5×10^{12} to 1×10^{13} photons pulse^{-1} μm^{-2} (three red columns from the left in Figure 8-B), which could be obtained in the current XFEL facilities, the improvement rate was comparable to an incident X-ray intensity of five to seven times that observed in the context of experimental observations, which was approximately an order of magnitude higher. The constructed SRCNN model proved capable of improving image quality to the extent that the obtained images were comparable in quality to diffraction images with an intensity approximately an order of magnitude higher than that of the original images. Thus, we succeeded in demonstrating the effectiveness of the proposed method on the similarity detection between diffraction images for nanoscale flexible biomolecules, whose analysis was at the limit of what was achievable through XFEL experiments currently. By using the predicted SRCNN image in the template matching method, it was expected that the more accurate 3D structural model was estimated.

3.6. Improvement of the oversampling ratio and real image analysis

We investigated whether the oversampling ratios of diffraction images could be

improved by the SRCNN model and whether the real images could be recovered using the HIO algorithm for phase retrieval. In single-particle coherent diffraction imaging, a real image was obtained by conducting oversampling and phase-retrieval calculations. However, a major issue in single-particle structure analysis of nanoscale flexible biomolecules was the insufficient diffracted X-ray intensity per pixel solid angle. Oversampling was the process of sampling at a frequency that was higher than the Nyquist frequency in the diffraction image. To perform phase retrieval calculations for an experiment diffraction image, it was necessary to measure the diffraction image with a high oversampling ratio (linear oversampling ratio $\sigma = 4$), and this rigorous experiment condition accelerated the lack of diffraction intensity per pixel solid angle. If a finer diffraction image that satisfied the oversampling condition was obtained by the SRCNN model from an experiment diffraction image at a low oversampling ratio, i.e., $\sigma = 1$, it would dramatically relax the strict experiment conditions. The SRCNN model was trained by using the experiment diffraction images at $\sigma = 2$, which were obtained by upscaling the size of the experiment diffraction images at $\sigma = 1$ twice, which did not satisfy the phase retrieval condition, using the nearest neighborhood interpolation while retaining the patterns of the original images. The EMV images as GTHR images at $\sigma = 2$, which satisfied the phase retrieval condition, were used as supervisory data.

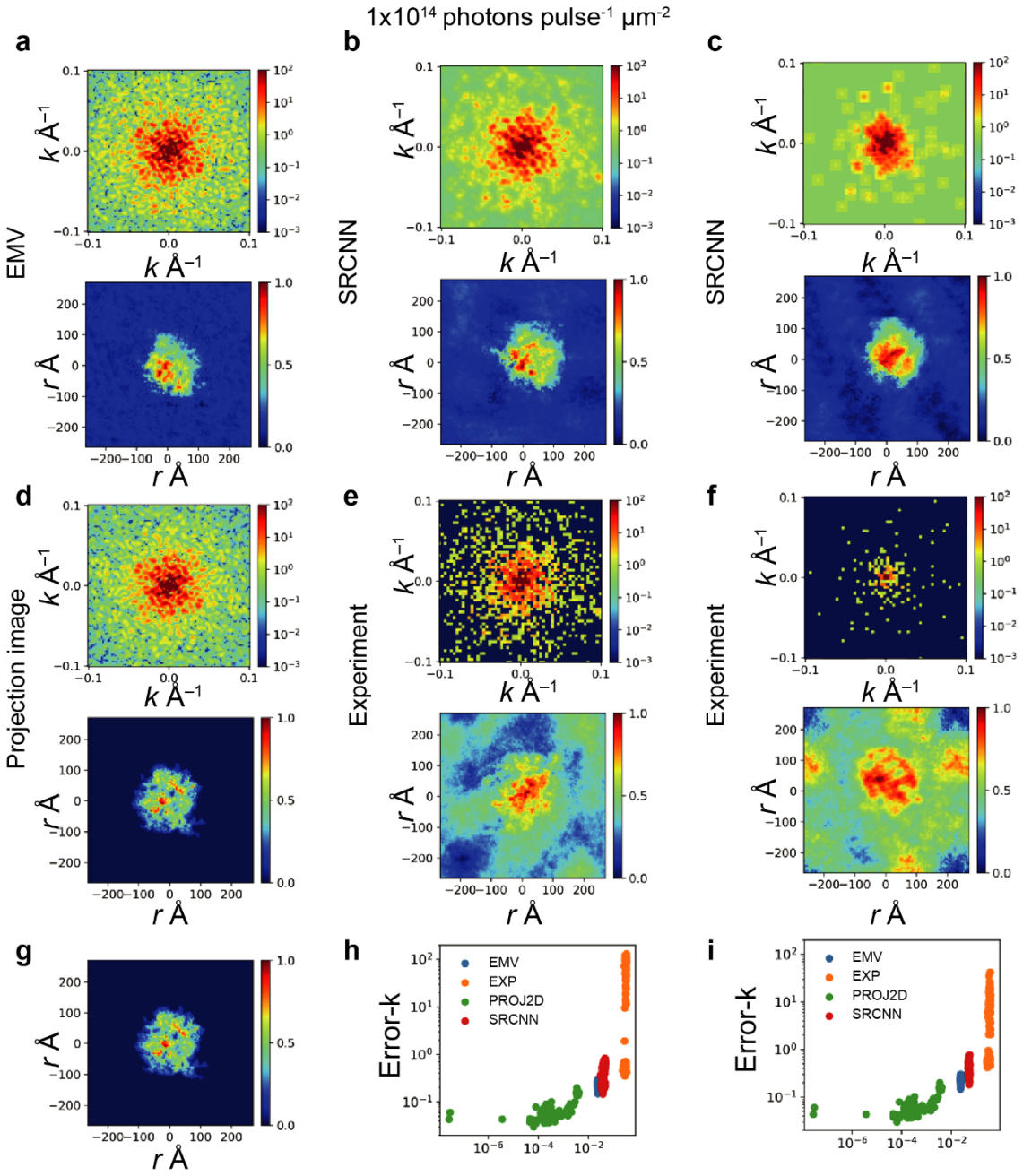

Figure 9. Upper row of A-F: diffraction images used as inputs for the phase retrieval calculations in the EMV, SRCNN, Projection image, and experiment approach, from upper left to lower right. Lower row of A-F: real-space images obtained by phase retrieval calculation, which have the maximum real-space C. values compared with G of the

electron density projection image, as shown in the electron density plots. The error values in k-space and real-space obtained by HIO method are also shown for the cases of (H) 1×10^{14} photons pulse^{-1} μm^{-2} and (I) 5×10^{12} photons pulse^{-1} μm^{-2}.

Figure 9 shows the results of the phase retrieval calculations using four types of diffraction images, the EMV image, the SRCNN image, the experiment image and the projection image, which was FT image of the 2D electron density map. The projection image in *k*-space had perfect point symmetry due to the lack of curvature of Ewald sphere (Figure 9-D). The phase retrieval analysis was performed 100 times using the HIO method with different initial phase, each consisting of 10,000 iterations. The lower parts in Figures 9-A, B, C, D, E, and F present the best real-space images obtained by the phase-retrieval calculation for each case. In the case of the projection image (see Figure 9-D), we obtained the reconstructed real-space image, which was almost equal to the electron density plots (Figure 9-G). However, in the case of the EMV image, the reconstructed real-space image was blurred (see Figure 9-A) compared with the electron density image if there was no noise. These results indicated that it was inferred that the effect of the curvature of the Ewald sphere affected the phase retrieval calculation in the EMV, SRCNN, and experiment in this condition. Moreover, this tendency was confirmed from the error values obtained by HIO calculations (see Figures 9-H and I).

We look at the results of the SRCNN images and the experiment images, keeping in mind that it included the effects of the curvature of the Ewald sphere on the HIO method. Figure 9-B, E and 9-C, F show the results of the phase retrieval calculations in the cases of 1×10^{14} and 5×10^{12} photons pulse^{-1} μm^{-2}, respectively. In Figures 9-E and 9-F, the case of upscaled experiment diffraction images with $\sigma = 1$ clearly did not work well because it did not satisfy the oversampling conditions, and the molecular shape was not retrieved correctly. In contrast, the real-space images obtained from the SRCNN images (Figure 9-B and 9-C) were similar to the electron density projection image, indicating that the phase retrieval calculation worked well.

It was concluded that the constructed SRCNN model improved diffraction images to the extent that phase retrieval calculation was performed, and finer images could be recovered. These results indicated that it was possible to reconstruct real images by using the improved SRCNN image from the observed diffraction image, which did not satisfy the phase retrieval conditions. Consequently, it became possible to conduct experimental measurements with smaller oversampling ratios, which could increase the number of measured photons per pixel. We successfully demonstrated that it was possible to obtain an actual image from an observed diffraction image under the condition that $\sigma = 1$ was satisfied with an incident X-ray intensity of 5×10^{12} photons pulse^{-1} μm^{-2} through SRCNN

analysis in the case of ribosomes.

We succeeded in demonstrating that a plausible real image could be recovered by phase retrieval for an experiment diffraction image that did not satisfy the phase retrieval conditions by applying the constructed SRCNN model to the experiment diffraction image to estimate a finer pattern. Our proposed scheme for X-ray diffraction image improvement was advantageous in that the oversampling ratio could be improved because we used the GTHR images for supervised training of SRCNN model in silico in various molecular orientations and conformations.

3.7 Generalization achievable by the SRCNN model with respect to molecular species

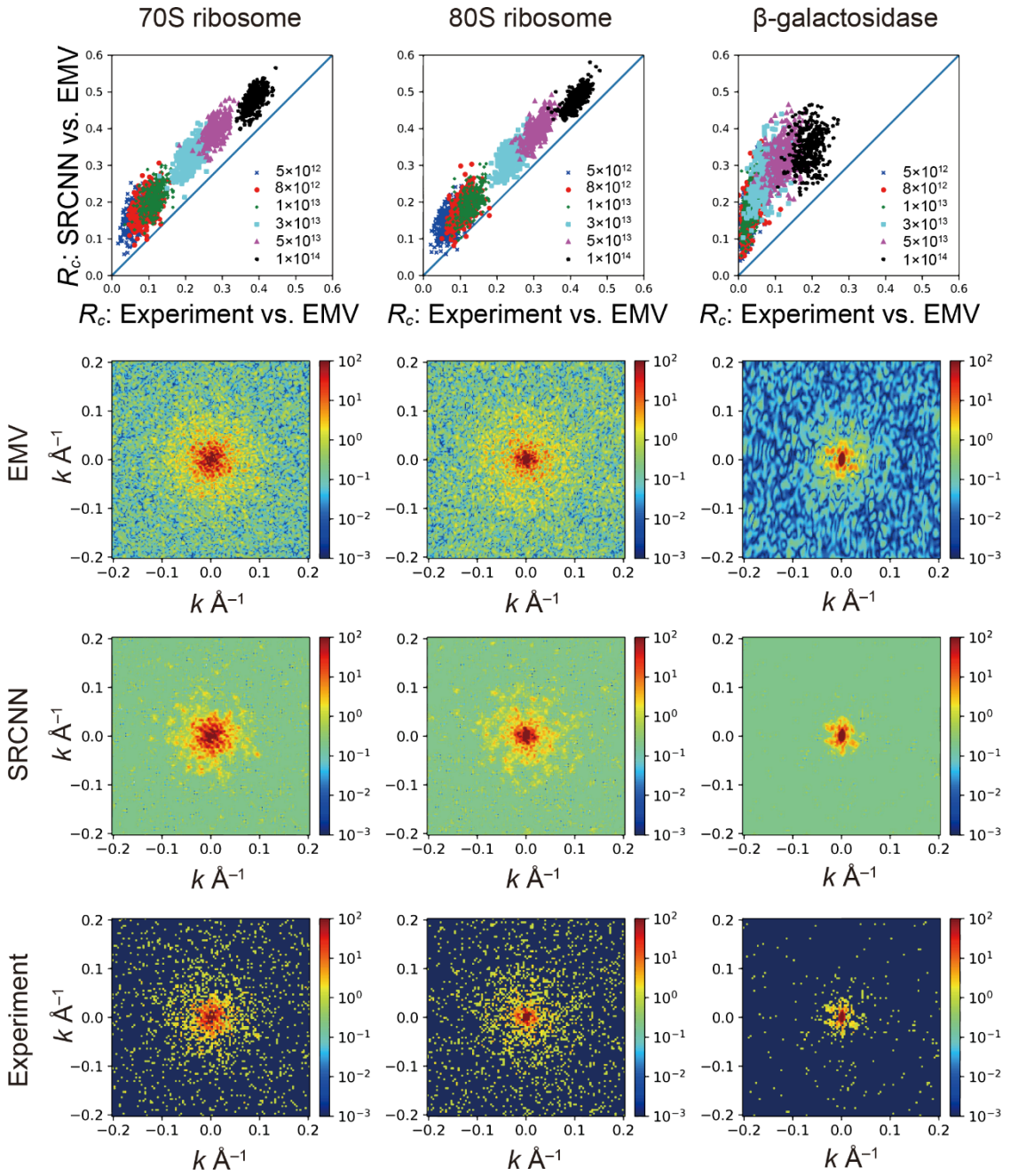

Figure 10. Improvement of the diffraction images of different molecular species by the constructed SRCNN model. From left to right: 70S ribosome, 80S ribosome, and *β*-galactosidase. Top row: similarity improvement. X: R_c score of experiment; Y: R_c score of SRCNN. The EMV, SRCNN, and experiment diffraction images are also shown, from

top to bottom.

Finally, we tested the constructed SRCNN model for different molecular species. The SRCNN model described in Section 2.5 was trained using diffraction images of the 70S ribosome. Then, we applied the same model to the experiment diffraction images of other molecular species, namely, the β-galactosidase and 80S ribosomes. Figure 10 shows the diffraction images were improved for all molecular species at all incident X-ray intensities with almost all molecular orientations. In particular, the application to the 80S ribosome resulted in 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_c 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 learn 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.

4. Conclusion

Here, we constructed a practical SRCNN model by data augmentation in a virtual space by using simulation data assuming actual experimental conditions. We showed that the constructed SRCNN model could be applied to diffraction images of all molecular orientations with three-fold X-ray intensity fluctuations and 10 Å structural fluctuations. The constructed SRCNN model improved the quality of single noisy X-ray diffraction image under actual experimental conditions, i.e., if the sample molecule had various conformational states and the incident X-ray intensity changed with each measurement. The diffraction image improvement rate was quantitatively evaluated using the similarity with the EMV diffraction image as a standard. We showed that the diffraction image improvement by the present SRCNN model corresponded to an increase in the incident X-ray intensity by three to seven times in the context of experimental observations.

We also confirmed that the SRCNN model not only contributes to noise reduction, but is also effective at improving the oversampling ratio. The real-space image analysis shows that the proposed SRCNN model not only amplifies the intensity, similar to the GB approach, but is also effective at reproducing the fine pattern of the diffraction image. It is expected that improving the oversampling ratio is a powerful approach because it

enables relaxation of the experimental conditions by the Nyquist frequency observation. Because this method alleviated the harsh experimental conditions required by the phase retrieval algorithm, it facilitated elucidation of the structures and dynamics of nanoscale flexible biomolecules and accelerated drug discovery.

In conclusion, single-image SR was an effective method for improving X-ray coherent diffraction image quality while maintaining the individuality of diffraction images that reflected the multi-conformational states of biomolecules. Because the proposed method enabled the measurement of low oversampling ratios, it relaxed the rigorous experiment conditions and expanded the applicability of SPA to nanoscale flexible biomolecules. By introducing the SRCNN model at the beginning of the structural analysis pipeline (see Figure 1), it was expected that the applicability of the existing SPA methods will be greatly expanded.

We confirmed that the results of two key techniques, the similarity detection between pair of diffraction images and the phase recovery, improved its accuracy by using the single improved 2D images. The improvement of the similarity value led to improving the estimation accuracy of the plausible 3D structure by the template matching method. Moreover, the improved oversampling ratio allowed observations at the Nyquist frequency and relaxed the harsh experimental conditions. The improvement of the

diffraction image by SRCNN dramatically enhanced the effectiveness of one diffraction image in which the individuality of biomolecules appeared.

Regarding the limitations of this work, the constructed SRCNN model depended on the training datasets because of supervised learning. We believed that we needed to further augment our training data considering a variety of molecular species and realistically-simulated experimental data. Although the current SRCNN model was not sufficient, we expected that the pre-trained SRCNN model conducting supervised learning in silico could be applied to real experimental data by combining it with real-world data. For example, we believe it was necessary to incorporate the beam stop region of the detector, Gaussian-like electronics noise of the detector, and fluctuations in the geometry of the detector and the sample into the learning process. Also, it would be necessary to examine the difference in performance of the methods other than SRCNN, including unsupervised methods, to see which methods were effective in improving XFEL images. Additionally, it was necessary to develop a workflow in which various structural analysis methods were organically combined to construct a structural analysis pipeline for flexible 100-nm biomolecules. Therefore, high-performance computing could be used with parallel processing for an artificial intelligence-driven big data analysis system linked with single-particle experimental measurements.

Supporting Information

The Supporting Information is available free of charge. The contents are "Hyperparameter optimization results", "learning curve of single intensity training", and "Protocol for the MD simulation of 70s ribosome".

Data and Software Availability

The software GROMACS is publicly available under academic license for research (https://www.gromacs.org). Keras is publicly available under open source software (https://github.com/keras-team/keras/releases/tag/2.4.0). The molecular operating environment is available as paid software (https://www.chemcomp.com/Products.htm) All relevant data are shown in figures, listed in tables, or included in the Supporting Information. The datasets of 18 in table2 and the constructing model that is presented in Figure7 can also be accessed at https://github.com/TokuhisaAtsushi/Improvement-of-X-ray-diffraction-image-using-SRCNN.

Author Information

Corresponding Author

* Correspondence: tokuhisa@riken.jp, okuno.yasushi.4c@kyoto-u.ac.jp.

ORCID

Atsushi Tokuhisa: 0000-0002-9584-1819

Kei Terayama: 0000-0003-3914-248X

Author Contributions

A.T. and Y.O. designed the study. A.T. and Y.A. performed the diffraction matching and phase retravel calculations and constructed the SRCNN model. K.T. constructed the SRCNN model. All authors reviewed the final manuscript.

Notes

The authors declare no competing financial interests.

Funding Sources

This work was supported in part by JSPS KAKENHI (grant number JP26870852); the FOCUS Establishing Supercomputing Center of Excellence project; and MEXT, as "Priority Issue 1 on Post-K computer" (Building Innovative Drug Discovery Infrastructure through Functional Control of Biomolecular Systems). This research used the computational resources of supercomputers Fugaku and Camphor2 of CRAY XC40

provided by Kyoto University, through the High Performance Computing Infrastructure System Research Project (Project IDs: hp140121, hp170036, hp180011, hp180123, hp190105, hp200053, hp200178 and hp210126).

Acknowledgments

We would like to thank Dr. Y. Joti (RIKEN) for help with the XFEL experimental conditions and Professor Y. Ishikawa (RIKEN), Dr. A. Hori (RIKEN), Mr. T. Kameyama (RIKEN), and Dr. M. Matsuda (RIKEN) for the parallelization of programs used for matching. We would also like to thank Editage (www.editage.jp) for English language editing.

REFERENCES

[1] Emma, P.; Akre, R.; Arthur, J.; Bionta, R.; Bostedt, C.; Bozek, J.; Brachmann, A.; Bucksbaum, P.; Coffee, R.; Decker, F.J.; Ding, Y. First Lasing and Operation of an Ångstrom-Wavelength Free-Electron Laser. *Nat. Photonics*. **2010**, *4* (9), 641–647. DOI: 10.1038/nphoton.2010.176

[2] Ishikawa, T.; Aoyagi, H.; Asaka, T.; Asano, Y.; Azumi, N.; Bizen, T.; Ego, H.; Fukami, K.; Fukui, T.; Furukawa, Y.; Goto, S. A Compact X-Ray Free-Electron Laser Emitting in the Sub-Ångström Region. *Nat. Photonics.* **2012**, *6* (8), 540–544. DOI: 10.1038/nphoton.2012.141

[3] Neutze, R.; Wouts, R.; Van der Spoel, D.; Weckert, E.; Hajdu, J. Potential for Biomolecular Imaging with Femtosecond X-Ray Pulses. *Nature*. **2000**, *406* (6797), 752–757. DOI: https://doi.org/10.1038/35021099

[4] Solem, J. C. Imaging Biological Specimens with High-Intensity Soft X Rays. *J. Opt. Soc. Am. B.* **1986**, *3* (11), 1551. DOI: 10.1364/JOSAB.3.001551

[5] Barends, T. R. M.; Foucar, L.; Botha, S.; Doak, R. B.; Shoeman, R. L.; Nass, K.; Koglin, J. E.; Williams, G. J.; Boutet, S.; Messerschmidt, M.; Schlichting, I. De Novo Protein Crystal Structure Determination from X-Ray Free-Electron Laser Data. *Nature*. **2014**, *505* (7482), 244–247. DOI: 10.1038/nature12773

[6] Chapman, H. N.; Fromme, P.; Barty, A.; White, T. A.; Kirian, R. A.; Aquila, A.; Hunter, M. S.; Schulz, J.; DePonte, D. P.; Weierstall, U.; Doak, R. B. Femtosecond X-Ray Protein Nanocrystallography. *Nature*. **2011**, *470* (7332), 73–77. DOI: 10.1038/nature09750

[7] Suga, M.; Akita, F.; Sugahara, M.; Kubo, M.; Nakajima, Y.; Nakane, T.; Yamashita, K.; Umena, Y.; Nakabayashi, M.; Yamane, T.; & Nakano, T. Light-Induced Structural Changes and the Site of O=O Bond Formation in PSII Caught by XFEL. *Nature*. **2017**, *543* (7643), 131–135. DOI: 10.1038/nature21400

[8] Aquila, A.; Barty, A.; Bostedt, C.; Boutet, S.; Carini, G.; DePonte, D.; Drell, P.; Doniach, S.; Downing, K. H.; Earnest, T.; Elmlund, H.; Elser, V.; Gühr, M.; Hajdu, J.; Hastings, J.; Hau-Riege, S. P.; Huang, Z.; Lattman, E. E.; Maia, F. R. N. C.; Marchesini, S.; Ourmazd, A.; Pellegrini, R.; Santra, R.; Schlichting, I.; Schroer, C.; Spence, J. C. H.; Vartanyants, I. V.; Wakatsuki, S.; Weis, W. I.; Williams, G. J. The Linac Coherent Light Source Single Particle Imaging Road Map. *Struct. Dyn*. **2015**, *2* (4), 041701. DOI: 10.1063/1.4918726

[9] Fortmann-Grote, C.; Buzmakov, A.; Jurek, Z.; Loh, N. T. D.; Samoylova, L.; Santra, R.; Schneidmiller, E. A.; Tschentscher, T.; Yakubov, S.; Yoon, C. H.; Yurkov, M. V.; Ziaja-Motykac, B.; Mancusoa, A. P. Start-to-End Simulation of Single-Particle Imaging using Ultra-Short Pulses at the European X-Ray Free-Electron Laser. *IUCrJ*. **2017**, *4* (5), 560–568. DOI: 10.1107/S2052252517009496

[10] Giewekemeyer, K., Aquila, A., Loh, N. T., Chushkin, Y., Shanks, K. S., Weiss, J. T., Tate, M. W., Philipp, H. T., Stern, S., Vagovic, P., Mehrjoo, M., Teo, C., Barthelmess, M., Zontone, F., Chang, C., Tiberio, R. C., Sakdinawat, A., Williams, G. J., Gruner, S. M., Mancusoa, A. P. Experimental 3D Coherent Diffractive Imaging from Photon-Sparse Random Projections. *IUCrJ*. **2019**, *6* (3), 357–365. DOI: 10.1107/S2052252519002781

[11] Maia, F. R. The Coherent X-Ray Imaging Data Bank. *Nat. Meth.* **2012**, *9* (9), 854–855. DOI: 10.1038/nmeth.2110

[12] Maia, F. R.; White, T. A.; Loh, N. T. D.; Hajdu, J. CCP-FEL: A Collection of Computer Programs for Free-Electron Laser Research. *J. Appl. Crystallog*. **2016**, *49* (4), 1117–1120. DOI: 10.1107/S1600576716011134

[13] Ourmazd, A. Cryo-EM, XFELs and the Structure Conundrum in Structural Biology. *Nat. Meth*. **2019**, *16* (10), 941–944. DOI: 10.1038/s41592-019-0587-4

[14] Spence, J. C. H. XFELs for Structure and Dynamics in Biology. *IUCrJ*. **2017**, *4* (4), 322–339. DOI: 10.1107/S2052252517005760

[15] Sun, Z.; Fan, J.; Li, H.; Jiang, H. Current Status of Single Particle Imaging with X-Ray Lasers. *Appl. Sci.* **2018**, *8* (1), 132. DOI: 10.3390/app8010132

[16] Yoon, C. H.; Yurkov, M. V.; Schneidmiller, E. A.; Samoylova, L.; Buzmakov, A.; Jurek, Z.; Ziaja, B.; Santra, R.; Loh, N. T. D.; Tschentscher, T.; Mancuso, A. P. A. A Comprehensive Simulation Framework for Imaging Single Particles and Biomolecules at the European X-Ray Free-Electron Laser. *Sci. Rep.* **2016**, *6* (1), 24791. DOI: 10.1038/srep24791

[17] Mimura, H.; Yumoto, H.; Matsuyama, S.; Koyama, T.; Tono, K.; Inubushi, Y.; Togashi, T.; Sato, T.; Kim, J.; Fukui, R.; Sano, Y. Generation of 10^{20} W cm^{-2} Hard X-Ray Laser Pulses with Two-Stage Reflective Focusing System. *Nat. Commun*. **2014**, *5* (1), 1–5. DOI: 10.1038/ncomms4539

[18] Ekeberg, T.; Svenda, M.; Abergel, C.; Maia, F. R. N. C.; Seltzer, V.; Claverie, J.M.; Hantke, M.; Jönsson, O.; Nettelblad, C.; Van der Schot, G.; Liang, M. Three-Dimensional Reconstruction of the Giant Mimivirus Particle with an X-Ray Free-Electron Laser. *Phys. Rev. Lett*. **2015**, *114* (9), 098102. DOI: 10.1103/PhysRevLett.114.098102

[19] Lundholm, I. V.; Sellberg, J. A.; Ekeberg, T.; Hantke, M. F.; Okamoto, K.; Van der Schot, G.; Andreasson, J.; Barty, A.; Bielecki, J.; Bruza, P.; Bucher, M. Considerations for Three-Dimensional Image Reconstruction from Experimental

Data in Coherent Diffractive Imaging. *IUCrJ*. **2018**, *5* (5), 531–541. DOI: 10.1107/S2052252518010047

[20] Rose, M.; Bobkov, S.; Ayyer, K.; Kurta, R. P.; Dzhigaev, D.; Kim, Y. Y.; Morgan, A. J.; Yoon, C. H.; Westphal, D.; Bielecki, J.; Sellberg, J. A. Single-Particle Imaging Without Symmetry Constraints at an X-Ray Free-Electron Laser. *IUCrJ*. **2018**, *5* (6), 727–736. DOI: 10.1107/S205225251801120X

[21] Fienup, J. R. Phase Retrieval Algorithms: A Comparison. *Appl. Opt.* **1982**, *21* (15), 2758–2769. DOI: 10.1364/AO.21.002758

[22] Huldt, G., Szoke, A., Hajdu, J. J. Diffraction Imaging of Single Particles and Biomolecules. *J. Struct. Biol.* **2003**, *144* (1–2), 219–227. DOI: 10.1016/j.jsb.2003.09.025

[23] Freeman, W. T.; Jones, T. R.; Pasztor, E. C. Example-Based Super-Resolution. *IEEE Comput. Graphics Appl.* **2002**, *22* (2), 56–65. DOI: 10.1109/38.988747

[24] Dong, C.; Loy, C. C.; He, K.; Tang, X. Image Super-Resolution using Deep Convolutional Networks. *Euro. Conf. Comput. Vis.* **2014**, *38* (2), 184–199. DOI: 10.1109/TPAMI.2015.2439281

[25] Kim, J.; Lee, J. K.; Lee, K. M. Terahertz Image Super-Resolution based on a Deep Convolutional Neural Network. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* **2016a**, 1646–1654. DOI: 10.1364/AO.58.002731

[26] Kim, J.; Lee, J. K.; Lee, K. M. Deeply-Recursive Convolutional Network for Image Super-Resolution. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*. **2016b**, 1637–1645. https://openaccess.thecvf.com/content_cvpr_2016/html/Kim_Deeply-Recursive_Convolutional_Network_CVPR_2016_paper.html

[27] Shi, W.; Caballero, J.; Huszar, F.; Totz, J.; Aitken, A. P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-Time Single Image and Video Super-Resolution using an Efficient Sub-Pixel Convolutional Neural Network. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*. **2016**, 1874–1883. https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Shi_Real-Time_Single_Image_CVPR_2016_paper.html

[28] *Lecture Notes in Computer Science*; Johnson, J., Alahi, A., Fei-Fei, L., Eds.; Springer, 2016.

[29] Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; Shi, W. Photo-Realistic Single Image

Super-Resolution using a Generative Adversarial Network. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* **2017**, 4681–4690. https://openaccess.thecvf.com/content_cvpr_2017/html/Ledig_Photo-Realistic_Single_Image_CVPR_2017_paper.html

[30] Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. Esrgan: Enhanced Super-Resolution Generative Adversarial Networks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. 2018, 0–0. https://openaccess.thecvf.com/content_eccv_2018_workshops/w25/html/Wang_ESRGAN_Enhanced_Super-Resolution_Generative_Adversarial_Networks_ECCVW_2018_paper.html

[31] Yang, W.; Zhang, X.; Tian, Y.; Wang, W.; Xue, J.H.; Liao, Q. Deep Learning for Single Image Super-Resolution: A Brief Review. *IEEE Trans Multimedia*. **2019**, *21* (12), 3106–3121. DOI: 10.1109/TMM.2019.2919431

[32] Wang, Z.; Chen, J.; Hoi, S.C. ConceptExplorer: Visual Analysis of Concept Drifts in Multi-Source Time-Series Data. *EEE Trans. Pattern Anal. Mach. Intell*. **2020**, 1–11 DOI: 10.1109/VAST50239.2020.00006

[33] Yuan, Y.; Liu, S.; Zhang, J.; Zhang, Y.; Dong, C.; Lin, L. Unsupervised Image Super-Resolution using Cycle-in-Cycle Generative Adversarial Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018, 701–710. https://openaccess.thecvf.com/content_cvpr_2018_workshops/w13/html/Yuan_Unsupervised_Image_Super-Resolution_CVPR_2018_paper.html

[34] Lee J, Park J, Lee K, Min J, Kim G, Lee B, Ku B, Han DK, Ko H. FBRNN: Feedback Recurrent Neural Network for Extreme Image Super-Resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.* 2020, 456–457. https://openaccess.thecvf.com/content_CVPRW_2020/html/w31/Lee_FBRNN_Feedback_Recurrent_Neural_Network_for_Extreme_Image_Super-Resolution_CVPRW_2020_paper.html

[35] Shocher, A.; Cohen, N.; Irani, M. “Zero-Shot” Super-Resolution using Deep Internal Learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, 3118–3126. https://openaccess.thecvf.com/content_cvpr_2018/html/Shocher_Zero-Shot_Super-Resolution_Using_CVPR_2018_paper.html

[36] Ayyer, K.; Lan, T. Y.; Elser, V.; Loh, N. T. D. J. Dragonfly: An Implementation of the Expand–Maximize–Compress Algorithm for Single-Particle Imaging. *Appl. Crystallog*. **2016**, *49* (4), 1320–1335. DOI: 10.1107/S1600576716008165

[37] Dragonfly. https://github.com/duaneloh/Dragonfly (accessed on 2021-11-25).

[38] Loh, N. T. D.; Elser, V. Reconstruction Algorithm for Single-Particle Diffraction Imaging Experiments. *Phys. Rev. E.* **2009**, *80* (2), 026705. DOI: 10.1103/PhysRevE.80.026705

[39] Jin, Q.; Miyashita, O.; Tama, F.; Yang, J.; Jonic, S. Poisson Image Denoising by Piecewise Principal Component Analysis and its Application in Single-Particle X-Ray Diffraction Imaging. *IET Image Process*. **2018**, *12* (12), 2264–2274. DOI: 10.1049/iet-ipr.2018.5145

[40] Chen, C., Seff, A., Kornhauser, A., Xiao, J. Learning Affordance for Direct Perception in Autonomous Driving. In *Proceedings of the IEEE International Conference on Computer Vision*, 2015, 2722– 2730. DOI:10.1007/978-3-319-46475-6_43.

[41] Gupta, A.; Vedaldi, A.; Zisserman, A. Synthetic Data for Text Localisation in Natural Images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, 2315–2324.

https://openaccess.thecvf.com/content_cvpr_2016/html/Gupta_Synthetic_Data_for_CVPR_2016_paper.html

[42] Kortylewski, A.; Schneider, A.; Gerig, T.; Egger, B.; Morel-Forster, A.; Vetter, T. Training Deep Face Recognition Systems with Synthetic Data. *arXiv*. **2018**, preprint arXiv:1802.05891.

[43] Tokuhisa, A.; Jonic, S.; Tama, F.; Miyashita, O. Hybrid Approach for Structural Modeling of Biological Systems from X-Ray Free Electron Laser Diffraction Patterns. *J. Struct. Biol.* **2016**, *194* (3), 325–336.

[44] Tokuhisa, A.; Kanada R.; Chiba, S.; Terayama, K.; Isaka, Y.; Ma, B.; Kamiya, N.; Okuno, Y. Coarse-Grained Diffraction Template Matching Model to Retrieve Multiconformational Models for Biomolecule Structures from Noisy Diffraction Patterns. *J. Chem. Inf. Model.* **2020**, *60* (6), 2803–2818. DOI: 10.1021/acs.jcim.0c00131

[45] Daurer, B. J.; Okamoto, K.; Bielecki, J.; Maia, F. R.; Mühlig, K.; Siebert, M. M.; Hantke, M. F.; Nettelblad, C.; Benner, W. H.; Svenda, M.; Timneanu, N. Experimental Strategies for Imaging Bioparticles with Femtosecond Hard X-Ray Pulses. *IUCrJ*. **2017**, *4* (3), 251–262. DOI: 10.1107/S2052252517003591

[46] Munke, A.; Andreasson, J.; Aquila, A.; Awel, S.; Ayyer, K.; Barty, A.; Bean, R. J.; Berntsen, P.; Bielecki, J. Boutet, S.; Bucher, M.; Chapman, H. N.; Daurer, B. J.; DeMirci, H.; Elser, V.; Fromme, P.; Hajdu, J.; Mantke, M. F.; Higashiura, A.; Hogue, B. G.; Hosseinizadeh, A.; Kim, Y.; Kirian, R. A.; Reddy, H. K. N.; Lan, T. Y.; Larsson, D. S. D.; Liu, H.; Loh, N. D.; Maia, F. R. N. C.; Mancuso, A. P.; Muhlig, K.; Nakagawa, A.; Nam, D.; Nelson, G.; Nettelblad, C.; Okamoto, K.; Ourmazd, A.; Rose, M.; Van der Schot, G.; Schwander, P.; Seibert, M. M.; Sellberg, J. A.; Sierra, R. G.; Song, C.; Svenda, M.; Timneanu, N.; Vartanyants, I. A.; Westphal, D.; Wiedorn, M. O.; Williams, G. J.; Xavier, P. L.; Yoon, C. H.; Zook, J. Coherent Diffraction of Single Rice Dwarf Virus Particles using Hard X-Rays at the Linac Coherent Light Source. *Sci. Data.* **2016**, *3* (1), 160064. DOI: 10.1038/sdata.2016.64

[47] Chollet, F. *Deep Learning with Python*, 2nd ed.; Manning, 2015. https://keras.io (accessed 2021-11-25).

[48] Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv*. **2014**, preprint arXiv:1412.6980.

[49] Tokuhisa, A.; Taka, J.; Kono, H.; Go, N. Classifying and Assembling Two-Dimensional X-Ray Laser Diffraction Patterns of a Single Particle to

Reconstruct the Three-Dimensional Diffraction Intensity Function: Resolution Limit due to the Quantum Noise. *Acta Crystallogr. A*. **2012**, *68* (3), 366–381. DOI: 10.1107/S010876731200493X

[50] Tokuhisa, A.; Arai, J.; Joti, Y.; Ohno, Y.; Kameyama, T.; Yamamoto, K.; Hatanaka, M.; Gerofi, B.; Shimada, A.; Kurokawa, M.; Shoji, F. High-Speed Classification of Coherent X-Ray Diffraction Patterns on the K Computer for High-Resolution Single Biomolecule Imaging. *J. Synchrotron Radiat.* **2013**, *20* (6), 899–904. DOI: 10.1107/S0909049513022152

[51] Bortel, G.; Faigel, G. Classification of Continuous Diffraction Patterns: A Numerical Study. *J. Str. Biol*. **2007**, *158* (1), 10–18. DOI: 10.1016/j.jsb.2006.10.018

[52] Tokuhisa, A. Characterization of X-Ray Diffraction Intensity Function from a Biological Molecule for Single Particle Imaging. *Biophys. Physicobiol.* **2019**, *16* 430–443. DOI: 10.2142/biophysico.16.0_430