



OPEN

## Noise reduction in X-ray photon correlation spectroscopy with convolutional neural networks encoder–decoder models

Tatiana Konstantinova, Lutz Wiegart, Maksim Rakitin, Anthony M. DeGennaro<sup>✉</sup> & Andi M. Barbour<sup>✉</sup>

Like other experimental techniques, X-ray photon correlation spectroscopy is subject to various kinds of noise. Random and correlated fluctuations and heterogeneities can be present in a two-time correlation function and obscure the information about the intrinsic dynamics of a sample. Simultaneously addressing the disparate origins of noise in the experimental data is challenging. We propose a computational approach for improving the signal-to-noise ratio in two-time correlation functions that is based on convolutional neural network encoder–decoder (CNN-ED) models. Such models extract features from an image via convolutional layers, project them to a low dimensional space and then reconstruct a clean image from this reduced representation via transposed convolutional layers. Not only are ED models a general tool for random noise removal, but their application to low signal-to-noise data can enhance the data's quantitative usage since they are able to learn the functional form of the signal. We demonstrate that the CNN-ED models trained on real-world experimental data help to effectively extract equilibrium dynamics' parameters from two-time correlation functions, containing statistical noise and dynamic heterogeneities. Strategies for optimizing the models' performance and their applicability limits are discussed.

Noise reduction in experiments facilitates reliable extraction of useful information from a smaller amount of data. This allows for more efficient use of experimental and analytical resources as well as enables the study of systems with intrinsically limited measurement time, e.g. cases with sample damage or out-of-equilibrium dynamics. While instrumentation development and optimization of experimental protocols are crucial in noise reduction, there are situations where computational methods can advance the improvements even further.

X-ray photon correlation spectroscopy (XPCS)<sup>1–3</sup> is a statistics-based technique that extracts information about a sample's dynamics through spatial and temporal analysis of intensity correlations between sequential images (frames) of a speckled pattern collected from a coherent X-ray beam scattered from the sample. The two-time intensity-intensity correlation function<sup>4,5</sup> (2TCF) is a matrix calculated as:

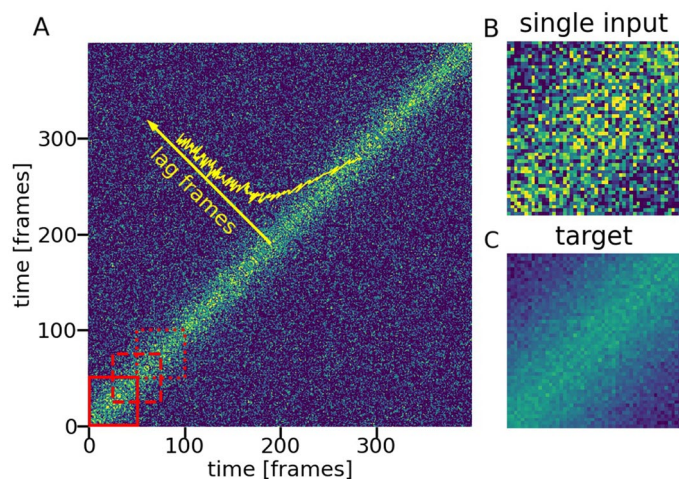
$$C2(\mathbf{q}, t_1, t_2) = \frac{\langle I(\mathbf{q}, t_1)I(\mathbf{q}, t_2) \rangle}{\langle I(\mathbf{q}, t_1) \rangle \langle I(\mathbf{q}, t_2) \rangle} \quad (1)$$

where  $I(\mathbf{q}, t)$  is the intensity of a detector pixel corresponding to the wave vector  $\mathbf{q}$  at time  $t$ . The average is taken over pixels with equivalent  $\mathbf{q}$  values. An example of a 2TCF is shown in Fig. 1. The dimensions of the matrix are  $N \times N$ , where  $N$  is a number of frames in the experimental series. The dynamics can be traced along the lag times  $\delta t = |t_1 - t_2|$ . In the case of equilibrium dynamics, information from a 2TCF can be 'condensed' to a single dimension by integrating along the  $(I, I)$  diagonal producing a time-averaged one-time photon correlation function (1TCF)<sup>6</sup>:

$$C1(\mathbf{q}, \delta t) = C_\infty + \beta |f(\mathbf{q}, \delta t)|^2 \quad (2)$$

where  $f(\mathbf{q}, \delta t)$  is the intermediate scattering function at lag time  $\delta t$ ,  $\beta$  is the optical contrast and  $C_\infty$  is the baseline that equals to 1 for ergodic samples. While 1TCF can be directly obtained from raw data<sup>7</sup>, calculating 2TCF as

Brookhaven National Laboratory, NSLS-II, Upton, NY 11973, USA. ✉email: adegennaro@bnl.gov; abarbour@bnl.gov



**Figure 1.** Data for the model. (A) 2TCF for an experimental series consisting of 400 frames. Red squares show examples of regions selected for the model training. Yellow arrow shows the temporal direction  $\delta t$  of the system's dynamics. Yellow solid line shows the 1TCF along  $\delta t$ , calculated from the 2TCF. (B) Example of  $50 \times 50$  2TCF, passed as an input to the model. (C) Example of the target data for the model, obtained by averaging multiple  $50 \times 50$  diagonal sections of the 2TCF. All images have the same intensity scale.

an intermediate step is beneficial even for presumably equilibrium cases. 2TCF contains time-resolved information about both samples' intrinsic dynamics and fluctuations of the experimental conditions, which enables one to determine between stationary and non-stationary dynamics and whether or not the time-averaged 1TCF is a valid representation of the scattering series. Investigation of 2TCF helps to identify single-frame events, such as cosmic rays detection, and beam-induced dynamics, where timescales might vary with the accumulation of X-ray dose absorbed by the sample during the acquisition of the dataset.

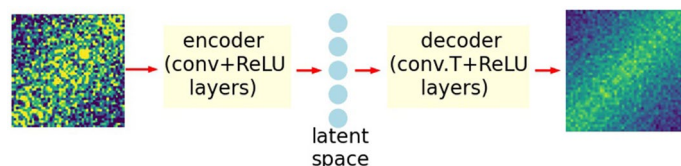
XPCS experiments can suffer from various sources of noise and artifacts: probabilistic nature of photon scattering, detector shot noise, and instrumental instabilities. Significant progress in reduction of the noise involved in photon detection and counting has been made by developing single-photon counting devices<sup>8,9</sup> and employment of the 'droplet' algorithm<sup>10</sup> or pixel binning<sup>11</sup>. Efforts have been dedicated to integrating feedback loops<sup>12,13</sup> into instrumental controls to reduce the impact of instabilities. Despite the current advances of experimental setup and methods for data analysis in reduction of noise and instability effects, achieving high signal-to-noise ratio is still a practical challenge in many XPCS experiments. The need to suppress the high-frequency fluctuations leads to extended data collection times—an approach that itself can introduce additional errors, for instance due to slow changes in experimental conditions. Limited experimental resources may not allow for multiple repeated measurements for systems with very slow dynamics. Besides, a sample's intrinsic properties can limit the time-range, within which the dynamics can be considered<sup>5</sup> as equilibrium and thus quantitatively evaluated with Eq. (2). A tool that helps to accurately extract parameters of the system's equilibrium dynamics from a limited amount of noisy data would be useful, but no generally applicable, out-of-the-box tool exists for XPCS results.

Solutions based on artificial neural networks are attractive candidates as they are broadly used for application-specific noise removal. Among such solutions are extensions of autoencoder models<sup>14</sup>, which are unsupervised algorithms for learning a condensed representation of an input signal. The principle behind an autoencoder is based on a common fact that the information about significant non-random variations in data is contained in a much smaller number of variables than the dimensionality of the data. An autoencoder model consists of two modules: encoder and decoder. The encoder transforms the input signal to a set of unique variables called latent space. The decoder part then attempts to transform the encoded variables back to the original input. As the number of components in the latent space is generally much smaller than the number of components in the original input, the nonessential information, i.e. random noise, is lost during such transformations. Thus, an autoencoder model on its own can be used as an effective noise reduction tool. However, in the scope of this work we employ a broader idea of noise. We treat all dynamic heterogeneities due to changes in a sample configuration caused by stress or diffusive processes, as well as correlated noise in 2TCF, as an unwanted signal. Such point of view can be preferred when one wants to quantify the average dynamics parameters with Eq. (3) or to separate the underlying (envelope) dynamics from stochastic heterogeneities. An autoencoder model can be modified to address the removal of a deterministic, application-specific noise by replacing its targets with 'noise-free' versions of the input signals. In the case of an image-like input, such as an XPCS 2TCF, convolutional neural networks (CNN) are the obvious choice for the encoder and decoder modules. CNN-based encoder-decoder (CNN-ED) models have been successfully implemented for noise removal and restoration of impaired signals in audio applications<sup>15,16</sup> and images<sup>17,18</sup>.

Here, we demonstrate an approach for noise reduction in 2TCFs by means of CNN-ED models. An ensemble of such models, trained on real experimental data, shows noticeable suppression of noise while preserving the functional form of system's equilibrium dynamics and the temporal resolution of the signal. Addressing noise

	Training	Validation
Unique inputs	12,236	5449
Unique targets	722	401

**Table 1.** Distribution of examples between the training and validation sets.



**Figure 2.** Architecture of the CNN-ED model. The input and the output images have the same intensity scale.

removal from 2TCF instead of the scattering signal at the detector makes the approach agnostic to the type of the registering device, the size of the selected area, the shape of the speckles, the intensity of the scattering signal and the exposure time, enabling the models' application to a wide range of XPCS experiments.

## Results

**Data processing.** The models are trained using data from the measurements of equilibrium dynamics of nanoparticle filled polymer systems conducted at the Coherent Hard X-ray Scattering (CHX) beamline at NSLS-II. For the nanoparticles' dynamics Eq. (2) can be approximated by the form<sup>5</sup>:

$$C1(\mathbf{q}, t) = C_{\infty} + \beta e^{-2(\Gamma t)^{\alpha}} \quad (3)$$

where  $\Gamma$  is the rate of the dynamics and  $\alpha$  is the compression constant. The baseline  $C_{\infty}$  is nearly 1 in the considered cases. Each experiment contains a series of 200–1000 frames. To augment the training data, additional 2TCFs are constructed using every second frame of the original series, which would be an equivalent to data collection with a twice longer lag period. Multiple regions of interest (ROI)—groups of pixels on the detector with equivalent wavevectors—are analyzed for each series and the 2TCFs are calculated for each ROI. For each model datum, or an “example”, the input image is obtained by cropping a  $50 \times 50$  pixels part from a 2TCF with the center on the  $(1,1)$  diagonal, starting at the lower left corner, as shown in Fig. 1A. Each next datum is obtained by shifting the center of the cropped image along the diagonal by 25 frames. The target image for each example is the average of all the cropped inputs extracted from the same 2TCF. Thus, groups of 3 to 39 inputs have the same target. While the target images still contain noise, its level is significantly reduced with respect to the noise of the input images. Here, the size of  $50 \times 50$  pixels is chosen as for the majority of the examples in the considered dataset the dynamics' parameters can be inferred from the first 50 frames. However, any size can be selected to train a model with little to no modification to its architecture if enough data are available.

The diagonal (lag = 0) 2TCF values of the raw data reflect the normalized variance of the photon count. Such values are vastly different between experiments and detector ROIs. They can by far exceed the values of photon correlation between frames (typically on a scale between 1 and 1.5) and are usually excluded from the traditional XPCS analysis. To prevent the influence of the high diagonal 2TCF values on the model cost function, the pixels along the diagonal are replaced with the values randomly drawn from the distribution of 2TCF values at lag = 1. In doing so, we avoid artificial discontinuities in the images.

For a proper model training process and to ensure its generalizability, we find that all the input data should be introduced to the model on the same scale. However, a commonly applied standard scaling is not suitable for the present case as the level of noise may affect the values of the essential parameters such as the baseline and the contrast. To bring all examples to a similar range, the estimated contrast for each series and each ROI is scaled to unity (see “Methods”). After processing, the data are split into the training and validation sets as shown in Table 1. The splitting is done in a way that no two inputs from different sets have the same target.

**Model training.** The ED model architecture used in this work is shown in Fig. 2. The encoder part consists of two convolutional layers with the kernel size  $1 \times 1$ . Training the model with larger kernel sizes did not improve the performance of the model. While kernels of size  $1 \times 1$  are used sometimes in CNN image applications<sup>19</sup> for creating non-linear activations, generally, they are not exclusively incorporated across the entire network. The reason for this is that the convolutional kernels are intended to catch distinctive edges, which form characteristic features of an image. To identify an edge, the distribution of intensities among the neighboring pixels is needed. However, the 2TCFs used in this work do not have sharp edges, which can partially explain the lack of improved learning with larger kernels. Besides, an equilibrium 2TCF has a unique structure, with symmetry along the diagonals. An equilibrium 2TCF and its modified copy with pixel values randomly shuffled along the diagonals would produce exactly the same 1TCF. This property is picked up by the model during compression of convolutional outputs to the latent space.

Both convolutional layers consist of 10 channels with rectified linear unit (*ReLU*) activation function applied to the output of each channel. We find that increasing the number of channels does not significantly change the performance of the model and the smaller number of channels gives poorer performance. No pooling layers<sup>20</sup> were introduced to prevent information loss<sup>21</sup> at the encoding stage. The output of the convolutional layers contains 25,000 features. A linear transformation is performed to convert them to the latent space of a much smaller dimension. While some ED image applications implement fully convolutional architectures<sup>16,18</sup>, we believe that the introduction of the linear layer for purpose of denoising equilibrium 2TCFs is beneficial. Not only does the bottleneck layer provide the regularization of the model, it also mixes the features derived by convolutional layers from different parts of the input image. The decoder part consists of two transposed convolutional layers, symmetrical to the encoder part, that convert the latent space back to a  $50 \times 50$  image. The *ReLU* function is applied only to the output of the first decoder layer.

The mean squared error (MSE) between the denoised output and the target is a natural choice of cost function for many image denoising applications. The MSE is shown to be useful for image denoising even in cases of some noise being present in the target<sup>22</sup>. Moreover, presence of noise in the input data puts a regularization on model weights, enforcing contractive property<sup>23</sup> on the reconstruction function of denoising EDs. The goal of the model presented here is to reduce the noise in 2TCF in such a way that the 1TCF, calculated from the model output, is as close to the target 1TCF as possible. Thus, the model's learning objective is modified by including the MSE between respective 1TCFs into the cost function.

We find that the regularization, which is enforced by the noise in both inputs and targets, in conjunction with the early stopping based on the cost function for the validation set, is sufficient for the model to avoid over-fitting. Introducing additional weight regularization reduced the accuracy of the model, especially for the cases of fast dynamics.

However, the cost function calculated for the validation set is not the only parameter to consider when selecting the optimal parameters for the model. When examining models trained for different latent space dimensions, the validation cost function (Fig. 3) does not have a pronounced minimum in the range of dimensions between 1 and 200. However, this metric may not reflect well the systematic errors in reconstructing the dynamics parameters, such as  $\beta$ ,  $\Gamma$ ,  $\alpha$  and  $C_\infty$ , which are essential to drawing scientific conclusions. An efficient model would precisely recover these parameters for a broad set of observations. Thus, the optimal dimension is selected based on how well the model output allows to recover those parameters for the validation data. Here, the rate of the dynamics,  $\Gamma$ , is the most important parameter to consider since the variation of  $\beta$  is taken care of by pre-processing normalization and the variations of  $\alpha$  and  $C_\infty$  are naturally very small in the considered examples.

To reduce the variance associated with the randomness of the initial weights initialization, ten models with different random initialization are trained for each latent space dimension. For each of the validation examples, the outputs of the ten models are converted to 1TCF, averaged and then fit to Eq. (3). The ground truth values, used for comparison, are obtained by fitting the 1TCF calculated from all (100–1000) frames in the same experiment and the same ROI as the input example is taken from.

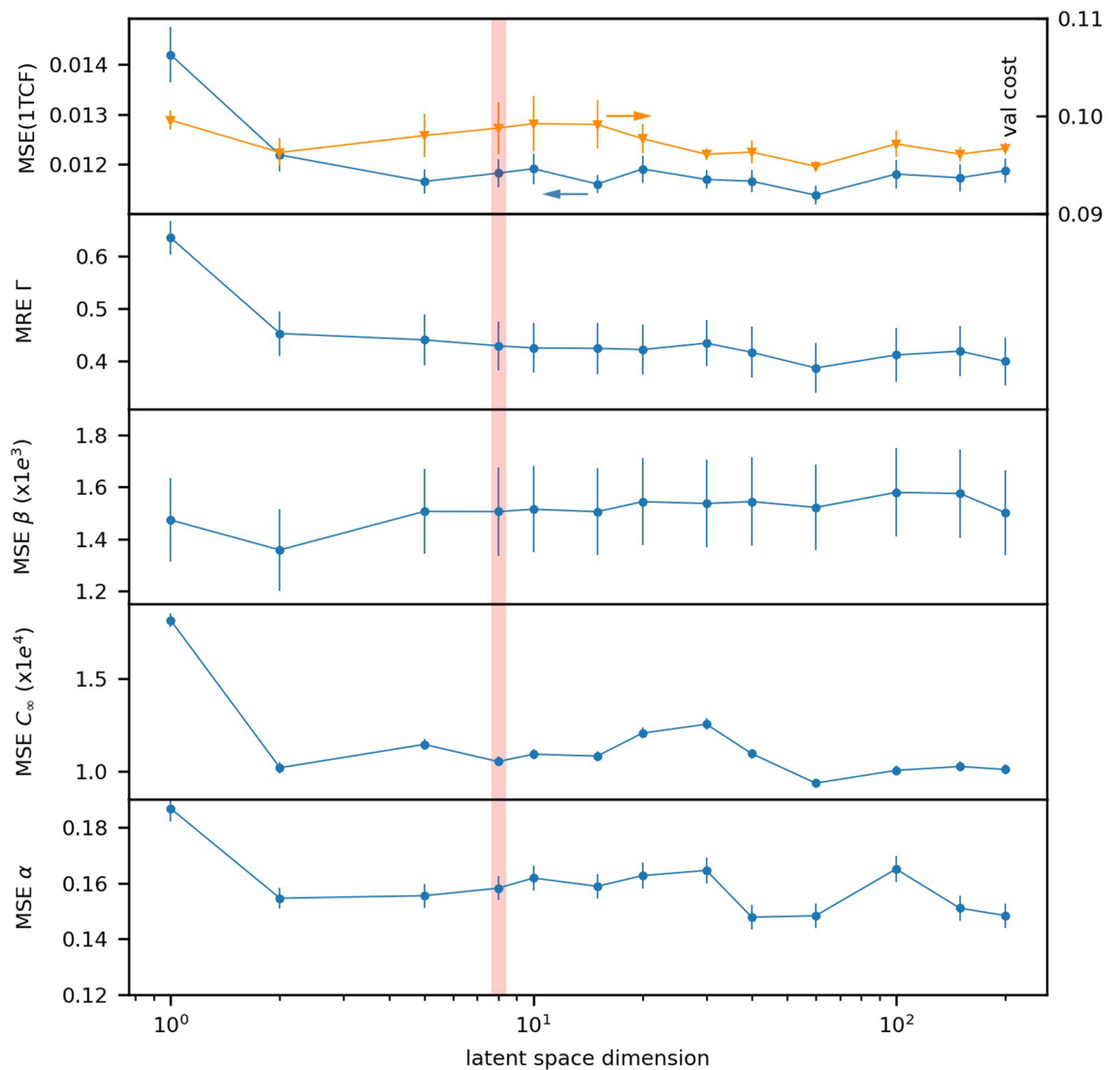
Since values of dynamics rate can be very close to zero, the mean absolute relative error (MRE) is considered for  $\Gamma$ . The MSE is calculated for other parameters. The accuracy of  $\Gamma$  keeps improving with increased number of hidden variables. But the rate of improvement slows down considerably above 5–8 variables. The same is observed for  $\alpha$  and  $C_\infty$ . This is in agreement with the MSE (1TCF) between the model output and the target, as shown in Fig. 3. The accuracy of  $\beta$  is relatively uniform across all the models. Based on the above, we select the models with eight latent variables for further consideration.

To address the variance of the selected CNN-ED, we train 100 such models with different random initialization and select among them the 10 best performing models based on the MSE (1TCF) for the validation set. Selecting only a limited number of the best performing models instead of combining all trained models also optimizes the use of storage memory and computational resources.

**Model testing.** The performance of the ensemble of models is evaluated through several tests. Firstly, we estimate the model applicability range by applying it to experiments that are similar to the ones used for training. An example of noise removal from a test datum is shown in Fig. 4. Reduction of the noise is especially important for larger lag times, where fewer scattering frames are available for calculating the correlations.

As mentioned above, despite the cost function working well for determining the optimal weights for a model, it is not sufficient to assess the reliability of the model output for quantitative analysis of the materials' dynamics. We assess the performance of the ensemble by comparing the fits with Eq. (3) for the 1TCFs calculated from the cropped  $50 \times 50$  pixels raw data (inputs), the corresponding denoised model outputs and the full-range raw data (ground truth target) (see the Supplemental Materials). From the results of the test set, the noise removal from the raw cropped 2TCFs with the CNN-ED ensemble noticeably improves the precision for the dynamics parameters in a wide range of cases with  $0.01 \text{ frames}^{-1} < \Gamma < 0.15 \text{ frames}^{-1}$  (i.e. the contrast drops by half within approximately the first 3–35 frames) in comparison with fitting the raw cropped 2TCFs. The application of the model enables reasonable estimates even in cases when the low signal-to-noise ratio of the raw cropped data prevents a convergent fit within the parameters' boundaries. In the region  $\Gamma > 0.15 \text{ frames}^{-1}$ , the results of the model are no longer more accurate than the raw data in general. Note that the precision of the model depends on the accuracy of identifying the optical contrast. Accurate measurements of optical contrast in XPCS experiments with fast dynamics can be challenging as they can involve data collection with reduced exposure or relying on averaged speckle visibility. Furthermore, a poor accuracy in identifying dynamics parameters is observed for inputs with very high noise levels (Fig. S7) and/or the presence of well pronounced dynamical heterogeneities.

If 100 or more frames are available for analysis, 2TCFs with slow dynamics can be reduced by considering every 2nd, 3rd, etc. frame, as it is done for augmenting the training data. This will effectively increase the exposure

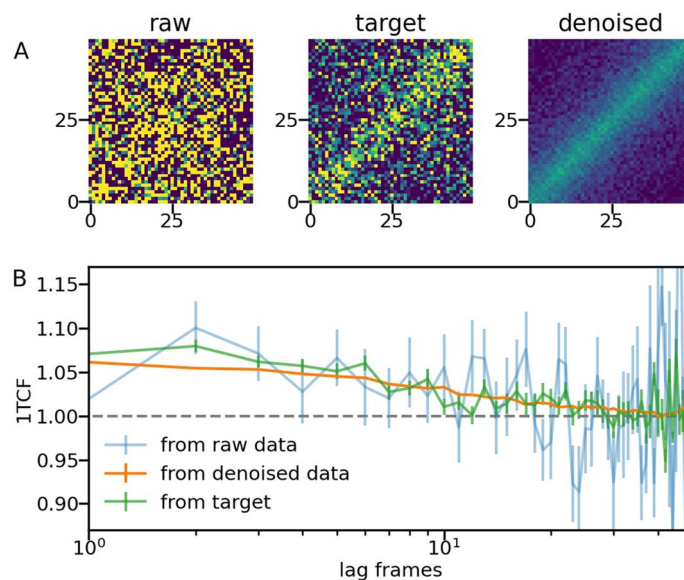


**Figure 3.** Selection of the latent space dimension for the model. From top to bottom, as the function of the latent space dimension: the MSE (1TCF) (blue circles) and the cost function (orange triangles) for the validation set; the MRE ( $\Gamma$ ); the MSE of  $\beta$ ,  $C_\infty$  and  $\alpha$ , extracted from the fit of the corresponding CNN-ED ensemble's outputs for the examples in the validation set to Eq. (3). The vertical line marks the choice of the latent space dimension for the model.

times and increase the  $\Gamma$  measured in  $frames^{-1}$ , making the model output more accurate. Alternatively, a model with a larger size of input 2TCF can be trained to handle cases of slow dynamics.

While it is clear from above how the model performs on average for individual independent 2TCFs, it is useful to see if application of the model can lead to reducing data collection in a typical XPCS experiment. We consider a single 700-frames series of scattering images among those used for creating the test set. The goal is to see if one can extract a sufficient information about the  $q$ -dependence of the dynamics rate  $\Gamma$  using only the first 50 frames with and without the model application. The target 2TCF for each of the concentric ROIs (shown in Fig. 5A) are calculated using all 700 frames. The first  $50 \times 50$  frame regions of the 2TCFs are considered and the ensemble CNN-ED model is applied to them. The visual comparison between the level of noise in the raw data and in the model output for an ROI with large  $q$  is shown in Fig. 5B. The 1TCFs, calculated from the raw cropped 2TCFs, from the model outputs and from the target 2TCFs for each ROI are fit to Eq. (3) with  $\alpha = 1$ . The results are shown in Fig. 5C–E. For the parameter  $\Gamma$ , at small  $q$ , where the signal-to-noise ratio is high, all three fits are close. However, as  $q$  grows and the noise level increases, the fit for the raw 50-frame 1TCFs starts to deviate from the target values more than the fit for the model outputs. In fact, the outcome of the model remains close to the actual values until the large  $q$  values (ROI# 16 and above). A similar tendency is observed for parameters  $\beta$  and  $C_\infty$ . This example demonstrates that application of the model can help to obtain sufficient information about the equilibrium system's dynamics from a smaller amount of data.

The applicability of the model to non-equilibrium data is also tested. Although the model is trained with the equilibrium examples, it still can be applied to quasi-equilibrium regions of a 2TCF with gradually varying dynamics parameters. Here, the model performance is demonstrated for a sample with ageing dynamics that



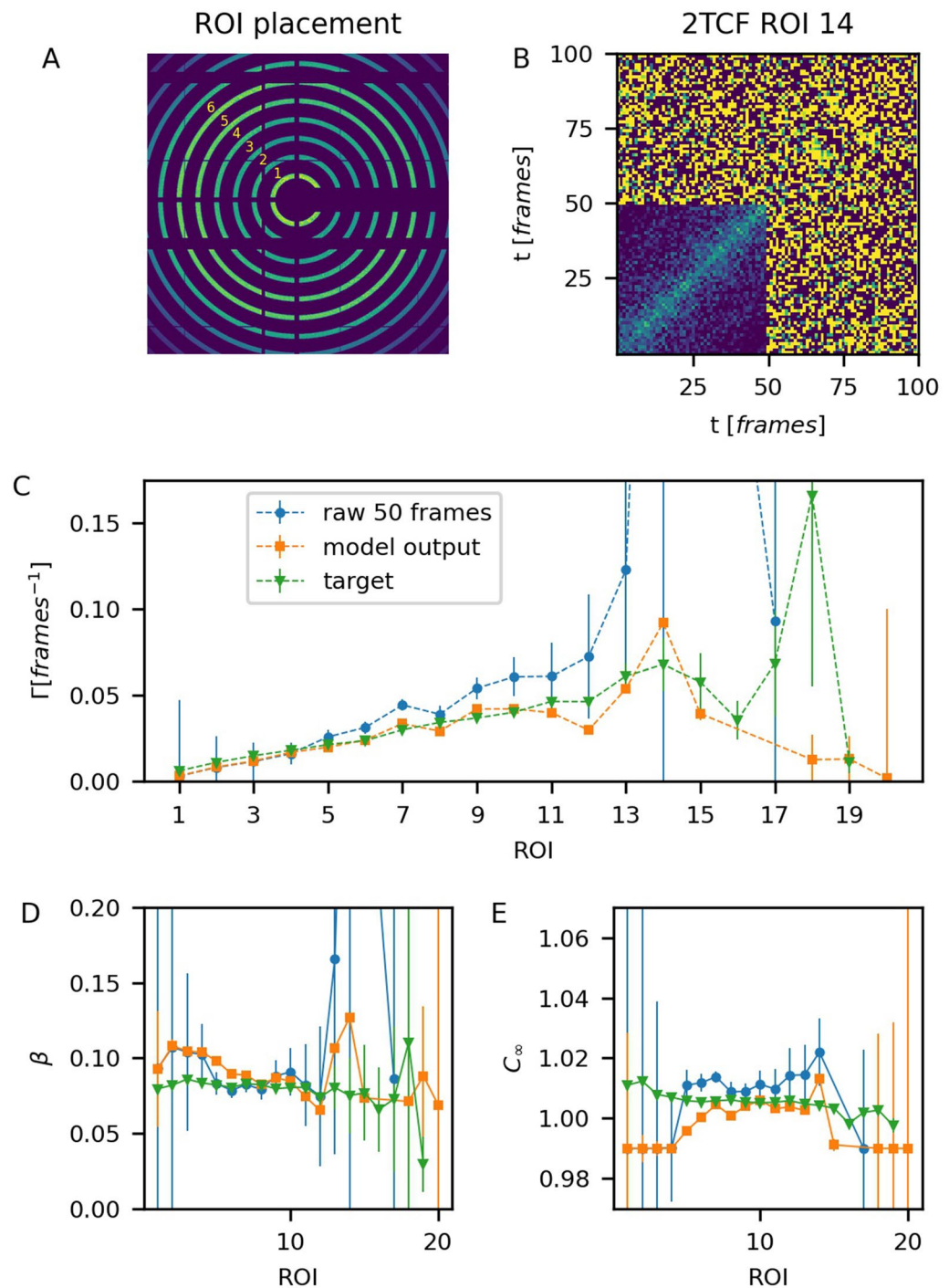
**Figure 4.** Example of 2TCF denoising with the CNN-ED models. **(A)** From left to right: the raw input 2TCF; the averaged target; the output of the ensemble of CNN-ED models. **(B)** 1TCF calculated from each 2TCF in **(A)**. The dashed line corresponding to a baseline  $C_\infty = 1$  is shown for convenience.

become slower with time. Since the target values cannot be obtained by averaging many frames for a such case, we calculate two 2TCFs with different noise levels, but carrying the same information, through sub-sampling pixels from the same ROI. The original ROI is a circle of small width with its center at  $q = 0$ . This ROI is used for calculating the target 2TCF. To calculate the test 2TCF with the reduced signal-to-noise ratio we randomly remove 74% of pixels from the original ROI. The model is applied along the  $(1,1)$  diagonal in a sliding window fashion (see “Methods”).

To compare the test 2TCF and the result of the model application, the cuts with width of 1 frame are taken perpendicular to the  $(1,1)$  diagonal and the resulting 1TCFs are fit to Eq. (3) in analogy to other XPCS analyses<sup>5,24</sup>. The target parameters are obtained by taking the cuts of 10 frames with the step of 10 frames from the target 2TCF and fitting the 1TCF, averaged over each cut, to the Eq. (3). Averaging is done for improving the accuracy of the target values. The contrast  $\beta$  is estimated as the mean of the respective raw 2TCF(lag = 1) at frames 250–300, where the dynamics are fairly slow, and is fixed during the fit. The results for  $\Gamma$ ,  $C_\infty$  and  $\alpha$  are shown in Fig. 6C–E. While the general trend of  $\Gamma(t)$  could be visually estimated from the raw test data, the output of the model gives much fewer outliers. Moreover, the temporal region, where  $\Gamma$  can be reasonably estimated is wider for the model output. The fit to the raw test data does not allow to estimate  $\Gamma$  in the first 30–40 frames, while the fit to the denoised data is close to the target in that region. The fit of the denoised data only shows a high uncertainty at the corners of the 2TCF, where the corresponding 1TCFs consist of less than 20 points. The variance of parameter  $C_\infty$  is also improved for the denoised data, but the most notable improvement in accuracy is observed for the parameter  $\alpha$ . The fits to the raw noisy data have high variance, which hides the upward trend of  $\alpha$ , in contrast to the fits to the denoised data.

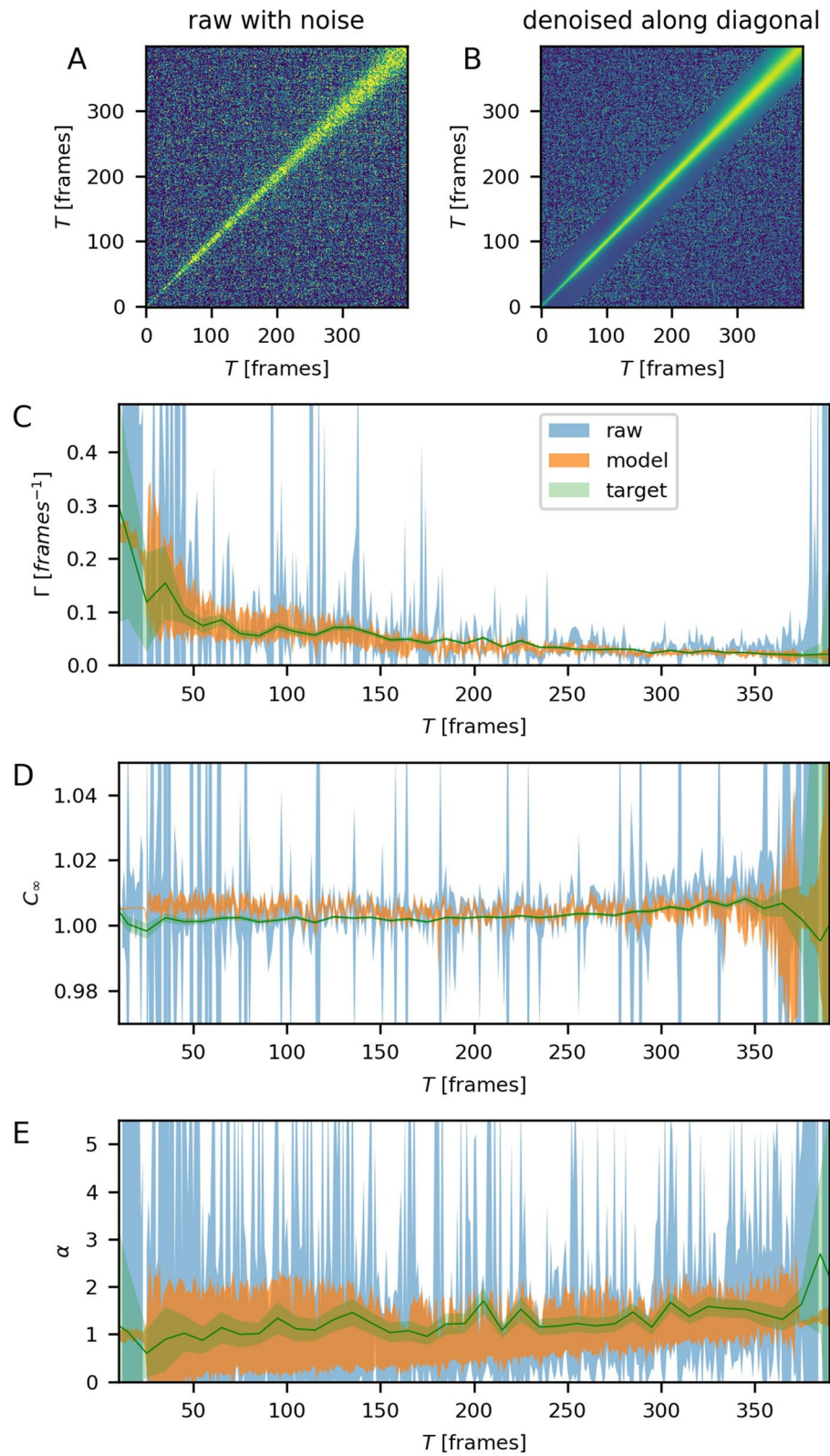
In a typical experiment, cuts with width of more than 1 frame are used for estimating the dynamics parameters achieving a better accuracy for the raw data than shown in Fig. 6. However, the selection of regions with quasi-equilibrium dynamics is not trivial. Since the fits to 1-frame-wide cuts from the denoised data have a low variance almost across the entire experiment, application of CNN-EDs makes the data more suitable for automated analysis and for visual inspection of the data when selecting the quasi-equilibrium regions.

**Comparison to other techniques.** We compare the performance of our approach to several of-the-shelf solutions for noise reduction in images: linear principle components-based, Gaussian, median and total variation denoising (Chambolles’ projection)<sup>25</sup> filters. The comparison of the application of these techniques to the same test example as in Fig. 4 is shown in Fig. 7. Principle components filters have the same idea as the ED model-preserving only the information from a few essential components of the original data. In fact, an autoencoder is a type of non-linear principle component generator. As one would expect, a filter based on linear principle components, trained on the same data as the CNN-EDs, under-performs comparing to the case of non-linear components due to a larger bias of the procedure for the components extraction. Gaussian and median filters are based on smoothing the intensity fluctuations between neighboring pixels and the total variation denoising is a regularized minimization of the additive normally distributed noise. While these approaches help to reduce pixel-to-pixel intensity variations, unlike the demonstrated here CNN-ED models, they do not learn the functional form of the equilibrium 2TCF images and cannot be improved by having a larger training set. By only considering local surrounding of individual pixels in a single image, such algorithms cannot recognize, for example, that the correlation function decays at larger lag times. Consequently, when an isolated high-intensity



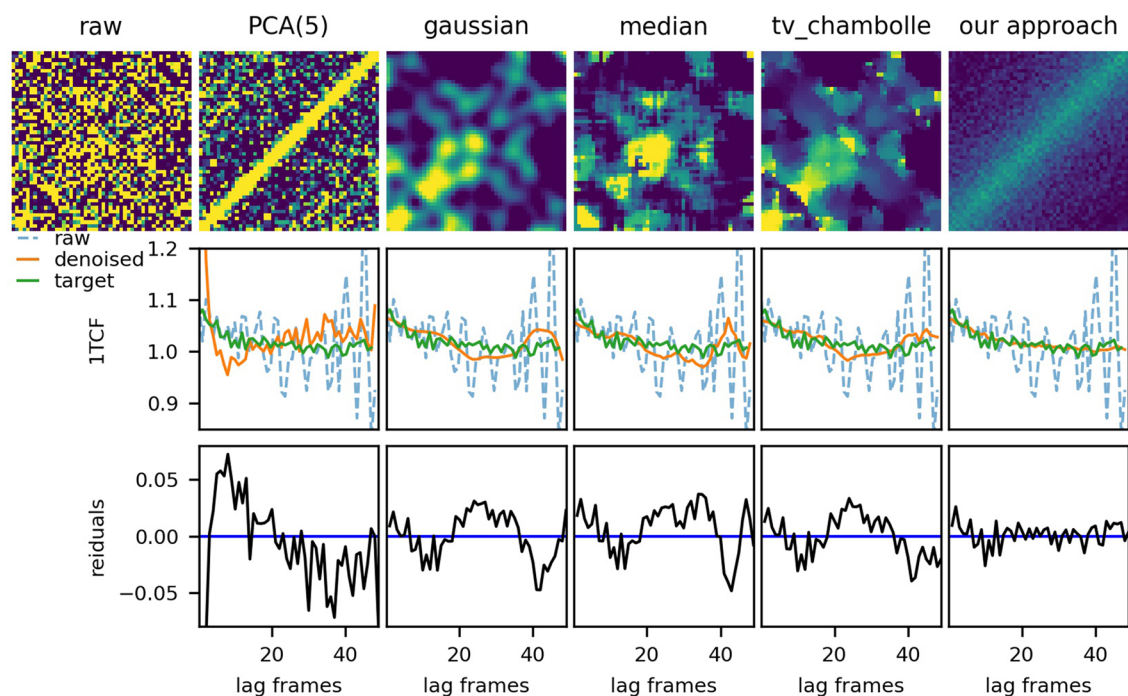
**Figure 5.** Model application for recovering the  $q$ -dependence of the dynamics parameters. **(A)** Scattering image of the sample with the ROI map on top of it. Dark blue corresponds to pixels excluded from the analysis. **(B)** A 100-frame fragment of 2TCF from ROI #14. The first 50-frame part is denoised with the model. Variation of  $\Gamma$  **(C)**,  $\beta$  **(D)** and  $C_\infty$  **(E)** parameters among ROIs with different  $q$ .

pixel (noise) is encountered in an image, an application of such filters leads to inflation of intensities in the surrounding pixels, highlighting the noise instead of correcting it. Thus, noise removal with the above filters can introduce false trends in 1TCF, which makes them unsuitable for quantitative XPCS data analysis. On the other hand, a CNN-ED, which is a regression model, learns from numerous examples the characteristic trends in the data and is less likely to introduce artifacts.



**Figure 6.** Model application for the case of non-equilibrium dynamics. (A) 2TCF calculated from reduced ROI. (B) The same 2TCF after the model is applied along its diagonal. Temporal evolution of  $\Gamma$  (C),  $C_\infty$  (D) and  $\alpha$  (E) parameters extracted from the noisy raw data, denoised data and the target 2TCF.





**Figure 7.** Comparison of various noise removal techniques applied to an example from the test set. Top row: results of applying filters to the raw 2TCF, middle row: 1TCFs calculated from the 2TCFs for the raw input (blue dashed line), the results of the respective filters (orange solid line) and the target (green solid line), bottom row: residuals of the 1TCFs calculated from the example after denoising with the respective filters.

## Discussion

The CNN-ED approach to noise removal in XPCS shows a reasonable improvement in the quality of the signal, allowing for quantification of a sample's dynamics from a limited amount of data, avoiding extensive data collection, accessing finite regions of reciprocal space and quasi-equilibrium intervals of non-equilibrium dynamics. The CNN-ED models go beyond and are superior to a simple smoothing of intensity between neighboring pixels since these models empirically learn the structural form of the 2TCF. The models are fast to train and do not require an extensive amount of training data. Their accuracy is pretty robust with respect to the choice of hyperparameters such as the number of channels in the hidden layers, the convolutional kernel size and the latent space dimension. The computational resources required for the application of the ensemble of 10 models are smaller than one needs to calculate 2TCFs for a typical number of frames required to achieve the same signal-to-noise ratio.

However, there are several limitations to keep in mind when applying CNN-ED models to a 2TCF. The testing results show that the models may not reliably remove the noise for the cases of very fast and very slow dynamics as well as from very noisy data (see the illustration in Supplementary Fig. S7). Some inaccuracy for the cases of fast dynamics comes from uncertainties in identifying the normalization factors (contrasts) for pre-processing of the inputs, which is also a challenge for traditional analysis. When the speckle visibility drops significantly within a single frame acquisition period, its estimation from the input data can have a high error. As it is seen from the model performance for the validation set and for the non-equilibrium test case in comparison to its performance for the equilibrium test set, a more accurate scaling of the inputs can improve the precision of the model for experiments with faster dynamics. Besides, one is advised to consider the context of extracted dynamics for a given material before relying solely on the information extracted from only a single 2TCF regardless of whether a CNN-ED model is applied. One benefits from a series of experiments on a single system, such as a temperature dependence or the demonstrated here  $q$ -dependence, to lend credibility to extracted dynamics for one particular experiment.

In this work, only equilibrium dynamics described by stretched exponents with the baseline close to 1 are used for training. Thus, the model learns to approximate any input with this type of dynamics. This can result in a loss of fine details, such as heterogeneities, oscillations or fast divergence of dynamics parameters in non-equilibrium cases. However, the demonstrated approach to the noise removal can be expanded to other types of dynamics with sufficient amount of data for training. Even in the absence of proper denoised target data, the autoencoder version of the model can significantly reduce the random noise. Furthermore, a CNN-ED model can be trained to correct for specific types of artifacts, such as impact of instrumentational instabilities or outlier frames, leading to a more efficient use of experimental user facilities<sup>26</sup>. Similarly to other fields<sup>27,28</sup>, the autoencoder models can be used for identifying unusual observations in the stream of XPCS data. Additionally, the encoded low-dimensional representation of the 2TCF can be used for classification, regression and clustering tasks, related to samples' dynamics. In the broader scope, the presented here CNN-ED models and their modifications have the

potential for application in automated high-rates XPCS data collection<sup>29</sup> and processing pipelines, reducing the reliance on the human-in-the-loop in decision making during experiments.

## Methods

**Training data.** The data for training and validation set contain experiments for 7 samples from 3 different material groups A (1 sample), B (2 samples) and C (4 samples). The experiments are taken at various exposures, acquisition rates and temperatures. Concentric ROIs with increasing  $q$  are used. Depending on the noise level and the dynamics duration, from 2 to 17 ROIs (median 10 ROIs) are considered for each experiment. The diversity of experimental conditions and regions in the reciprocal space allows one to obtain a realistic distribution of dynamics parameters and noise levels. We have not included the cases with very slow dynamics, for which only a small portion is complete within the 50 frames. To cut off the high noise data, we excluded the cases, where the fit to Eq. (3) did not converge for the full-range 1TCF. The distributions of dynamics parameters for the training and the validation set are shown in Fig. S1. For the model training purposes, all input data ( $2TCF_{noisy}$ ) are scaled as:

$$input = (2TCF_{noisy} - 1)/\beta^* + 1 \quad (4)$$

where  $\beta^*$  is the estimation of speckle visibility for the integration time of a single frame. It is obtained from fitting the equivalent pixels' intensity fluctuations with a negative binomial distribution<sup>6</sup>. For this, the speckle visibility, is calculated for each frame and is averaged among all the frames in the series. The target data are reversely scaled accordingly.

**Test data.** The test data are collected in a similar fashion to the training/validation data. Experiments for 5 different samples from the same material group (C) are considered. Experiments are performed for different temperatures, exposure times and acquisition rates. Concentric ROIs with increasing  $q$  are used. 10 ROIs with the smallest  $q$ -s are considered. However, no visual inspection of the data is done prior to model application and the ROIs with slow dynamics are not rejected. ROIs, where the full-range 1TCF fit to Eq. (3) does not converge, are not considered. Overall, 12,060 inputs (679 distinct targets) are considered in the test set. The distribution of the parameters from Eq. (3) is shown in Fig. S2. Unlike training/validation data, the contrast for normalization of test inputs is estimated from the 1TCF derived from the 2TCF) at lag = 1 frame instead of the speckle visibility. This is done to reduce the computation time and to test the model performance for the cases when only the  $50 \times 50$  2TCFs, and not the scattered images, are available. No adjustment is done to the baseline as the input data does not provide a good estimate for it. Thus, for each of the noisy  $2TCF_{noisy}$ , the model *input* is calculated as:

$$input = (2TCF_{noisy} - 1)/1TCF_{noisy}(lag = 1) + 1 \quad (5)$$

The denoised  $2TCF_{denoised}$  is then obtained from the output as:

$$2TCF_{denoised} = (output - 1) * 1TCF_{noisy}(lag = 1) + 1 \quad (6)$$

**Non-equilibrium test.** For the example of ageing dynamics considered in this work, the model is applied to each  $50 \times 50$  piece of the raw 2TCF along its (1,1) diagonal with the step size 5 frames, starting at the first frame. Prior the application of the model, each input is obtained from a raw  $2TCF_{noisy}$  as:

$$input = (2TCF_{noisy} - C_{\infty}^*)/\beta^*(lag = 1) + 1 \quad (7)$$

where  $\beta^*(lag = 1)$  is the estimation of contrast at lag=1 as the mean of  $2TCF_{noisy}(lag = 1)$  for frames 250–300 and  $C_{\infty}^*$  is the estimation of the baseline as the mean of  $2TCF_{noisy}$  at lags 270–300. The reverse transformation is applied to the model output. The overlapping model outputs between the current and the previous steps are averaged. The values outside of the  $50 \times 50$  diagonal sliding window remain unchanged. The same procedure is repeated with the model sliding window moving from the last frame towards the first frame. The two results are averaged to reduce the dominating influence of the later dynamics over the earlier dynamics and vice versa. The loss of the temporal resolution due to convolution between the model and the raw signal is not significant for the considered case of slowly-evolving sample dynamics.

**Model training details.** The cost function used for training the models is the sum of the Mean Squared Error (MSE) between the target 2TCF and the models' output and the MSE between the respective 1TCFs (with-out lag = 0):

$$cost = \frac{1}{2500 \cdot m} \sum_{k=1}^m \|x_k^{out} - x_k^{target}\|_2^2 + \frac{1}{49 \cdot m} \sum_{k=1}^m \|1TCF(x_k^{out}) - 1TCF(x_k^{target})\|_2^2 \quad (8)$$

where  $x_k^{out}$  is the model output for the  $k$ -th training example and  $x_k^{target}$  is the corresponding target's pixel,  $m$  is the number of examples,  $\|\cdot\|_2$  stands for 2-norm.

At every training epoch, batches of size 8 are processed. Adam<sup>30</sup> optimizer with initial learning rate from  $2.5e-6$  to  $4e-5$  is used. Learning rate is reduced by a factor of 0.9995 at every epoch. Initial weights in the

convolutional and linear layers are assigned according to Xavier uniform initialization<sup>31</sup>. The models are trained with Nvidia GPU accelerator GeForce RTX 2070 Super. For the selected CNN-ED configuration, the average training time is 27 seconds per epoch with 9–82 epochs necessary to train a model. Each input or target takes 30 kB of computer memory.

A model application does not require a GPU and, in fact, can be preformed faster without transferring the 2TCF data to a GPU. When using a CUDA accelerator, loading the model from a file, converting the 2TCF from numpy arrays to a CUDA Pytorch tensor, application of the model and converting the result back to a numpy array takes 2.3 ms on average with pure model computation taking 0.48 ms. Without using a CUDA accelerator, the corresponding times are 1.4 ms and 0.57 ms, respectively.

## Data availability

The code and the data used for the model training can be found at GitHub repository, <https://github.com/bnl/CNN-Encoder-Decoder> at the time of publication. DOI: <https://doi.org/10.11578/dc.20210704.1>.

Received: 3 February 2021; Accepted: 10 May 2021

Published online: 20 July 2021

## References

- Madsen, A., Fluorasu, A. & Ruta, B. *Structural Dynamics of Materials Probed by X-Ray Photon Correlation Spectroscopy*, 1617–1641 (Springer International Publishing, Switzerland, 2016).
- Shpyrko, O. G. X-ray photon correlation spectroscopy. *J. Synchrotron Radiat.* **21**, 1057–1064 (2014).
- Sinha, S. K., Jiang, Z. & Lurio, L. B. X-ray photon correlation spectroscopy studies of surfaces and thin films. *Adv. Mater.* **26**, 7764–7785 (2014).
- Brown, G., Rikvold, P. A., Sutton, M. & Grant, M. Speckle from phase-ordering systems. *Phys. Rev. E* **56**, 6601 (1997).
- Madsen, A., Leheny, R. L., Guo, H., Sprung, M. & Czakkel, O. Beyond simple exponential correlation functions and equilibrium dynamics in x-ray photon correlation spectroscopy. *New J. Phys.* **12**, 055001 (2010).
- Li, L. *et al.* Photon statistics and speckle visibility spectroscopy with partially coherent X-rays. *J. Synchrotron Radiat.* **21**, 1288–1295. <https://doi.org/10.1107/S1600577514015847> (2014).
- Lumma, D., Lurio, L. B., Mochrie, S. G. J. & Sutton, M. Area detector based photon correlation in the regime of short data batches: Data reduction for dynamic x-ray scattering. *Rev. Sci. Instrum.* **71**, 3274–3289. <https://doi.org/10.1063/1.1287637> (2000).
- Grybos, P., Kmon, P., Maj, P. & Szczygiel, R. 32k Channels readout IC for single photon counting detectors with 75  $\mu\text{m}$  pitch, ENC of 123 e- rms, 9 e- rms offset spread and 2% rms gain spread. In *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 1–4. <https://doi.org/10.1109/BioCAS.2015.7348438> (2015).
- Llopart, X., Campbell, M., Dinapoli, R., San Segundo, D. & Pernigotti, E. Medipix2: A 64-k pixel readout chip with 55-/spl mu/m square elements working in single photon counting mode. *IEEE Trans. Nucl. Sci.* **49**, 2279–2283 (2002).
- Livet, F. *et al.* Using direct illumination CCDs as high-resolution area detectors for x-ray scattering. *Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrom. Detect. Assoc. Equip.* **451**, 596–609 (2000).
- Falus, P., Lurio, L. & Mochrie, S. Optimizing the signal-to-noise ratio for x-ray photon correlation spectroscopy. *J. Synchrotron Radiat.* **13**, 253–259 (2006).
- Kongtawong, S. *et al.* Recent improvements in beam orbit feedback at NSLS-II. *Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrom. Detect. Assoc. Equip.* **164250**, 164250 (2020).
- Strocov, V. *et al.* High-resolution soft x-ray beamline ADDRESS at the swiss light source for resonant inelastic x-ray scattering and angle-resolved photoelectron spectroscopies. *J. Synchrotron Radiat.* **17**, 631–643 (2010).
- Kramer, M. A. Nonlinear principal component analysis using autoassociative neural networks. *AIChE J.* **37**, 233–243 (1991).
- Grais, E. M. & Plumbley, M. D. Single channel audio source separation using convolutional denoising autoencoders. In *2017 IEEE global conference on signal and information processing (GlobalSIP)*, 1265–1269 (IEEE, 2017).
- Park, S. R. & Lee, J. A fully convolutional neural network for speech enhancement. arXiv preprint [arXiv:1609.07132](https://arxiv.org/abs/1609.07132) (2016).
- Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T. & Efros, A. A. Context encoders: Feature learning by inpainting. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* **2536–2544**, (2016).
- Mao, X.-J., Shen, C. & Yang, Y.-B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. arXiv preprint [arXiv:1603.09056](https://arxiv.org/abs/1603.09056) (2016).
- Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014).
- Nagi, J. *et al.* Max-pooling convolutional neural networks for vision-based hand gesture recognition. In *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, 342–347 (IEEE, 2011).
- Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241 (Springer, 2015).
- Lehtinen, J. *et al.* Noise2noise: Learning image restoration without clean data. arXiv preprint [arXiv:1803.04189](https://arxiv.org/abs/1803.04189) (2018).
- Alain, G. & Bengio, Y. What regularized auto-encoders learn from the data-generating distribution. *J. Mach. Learn. Res.* **15**, 3563–3593 (2014).
- Malik, A. *et al.* Coherent x-ray study of fluctuations during domain coarsening. *Phys. Rev. Lett.* **81**, 5832 (1998).
- Duran, J., Coll, B. & Sbert, C. Chambolle's projection algorithm for total variation denoising. *Image Process. Line* **3**, 311–331 (2013).
- Campbell, S. *et al.* Outlook for artificial intelligence and machine learning at the NSLS-II. *Machine Learning: Science and Technology* (2020).
- Baur, C., Wiestler, B., Albarqouni, S. & Navab, N. Deep autoencoding models for unsupervised anomaly segmentation in brain MR images. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, 161–169 (Springer International Publishing, Cham, 2019).
- Chong, Y. S. & Tay, Y. H. Abnormal event detection in videos using spatiotemporal autoencoder. In *Advances in Neural Networks—ISNN 2017*, 189–196 (Springer International Publishing, Cham, 2017).
- Zhang, Q. *et al.* 20  $\mu\text{s}$ -resolved high-throughput x-ray photon correlation spectroscopy on a 500k pixel detector enabled by data-management workflow. *J. Synchrotron Radiat.* **28**, 259–265. <https://doi.org/10.1107/S1600577520014319> (2021).
- Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014).
- Glorot, X. & Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* 249–256, (2010).

## Acknowledgements

The authors thank A. Fluerasu and M. Fukuto for fruitful discussions. This research used CHX and CSX beamlines and resources of the National Synchrotron Light Source II, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory (BNL) under Contract No. DE-SC0012704 and under a BNL Laboratory Directed Research and Development (LDRD) project 20-038 “Machine Learning for Real-Time Data Fidelity, Healing, and Analysis for Coherent X-ray Synchrotron Data”.

## Author contributions

A.M.B, A.M.D, L.W and T.K. conceived the idea, L.W. performed the beamline experiments, generated the XPCS results, and identified individual scans for model development, T.K. processed the data for the model, A.M.D and T.K. wrote the code, M.R. provided technical consultation, T.K., L.W., A.M.D, A.M.B and M.R. analyzed the model performance, T.K. wrote the manuscript with contribution from all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-93747-y>.

**Correspondence** and requests for materials should be addressed to A.M.D. or A.M.B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2021