Deep learning for estimation of Kirkpatrick–Baez mirror alignment errors

Authors: Jianan Xie, Hui Jiang, Aiguo Li, Naxi Tian, Shuai Yan, Dongxu Liang, Jun Hu, Hui Jiang

Date: 2023-07-09T00:00:00+00:00

Abstract

A deep learning-based automated Kirkpatrick–Baez mirror alignment method is proposed for synchrotron radiation. We trained a convolutional neural network (CNN) on simulated and experimental imaging data of a focusing system. Instead of learning directly from bypass images, we use a scatterer for X-ray modulation and speckle generation for image feature enhancement. The smallest normalized root mean square error on the validation set was 4%. Compared with conventional alignment methods based on motor scanning and analyzer setups, the present method simplified the optical layout and estimated alignment errors using a single-exposure experiment. Single-shot misalignment error estimation only took 0.13 s, significantly outperforming conventional methods. We also demonstrated the effects of the beam quality and pretraining using experimental data. The proposed method exhibited strong robustness, can handle high-precision focusing systems with complex or dynamic wavefront errors, and provides an important basis for intelligent control of future synchrotron radiation beamlines.

Full Text

Preamble

Deep Learning for Estimation of Kirkpatrick–Baez Mirror Alignment Errors

Jianan Xie[1,2], Hui Jiang[1,3,4,*], Aiguo Li[1,2,3,4,**], Naxi Tian[1], Shuai Yan[1], Dongxu Liang[1], Jun Hu[2]

[1]Shanghai Synchrotron Radiation Facility, Shanghai Advanced Research Institute, Chinese Academy of Sciences, 239 Zhangheng Road, Pudong District, Shanghai 201204, China

[2]ShanghaiTech University, 393 Middle Huaxia Road, Pudong District, Shanghai 201210, China

[3]Shanghai Institute of Applied Physics, Chinese Academy of Sciences, 2019 Jialuo Road, Jiading District, Shanghai 201800, China

[4]University of Chinese Academy of Sciences, Beijing 100049, China

*Hui Jiang. E-mail address: jiangh@sari.ac.cn

**Aiguo Li. E-mail address: liag@sari.ac.cn

Abstract: A deep learning-based automated Kirkpatrick–Baez mirror alignment method is proposed for synchrotron radiation applications. We trained a convolutional neural network (CNN) on simulated and experimental imaging data of a focusing system. Instead of learning directly from bypass images, we employed a scatterer for X-ray modulation and speckle generation to enhance image features. The smallest normalized root mean square error on the validation set was 4%. Compared with conventional alignment methods based on motor scanning and analyzer setups, the present method simplifies the optical layout and estimates alignment errors using a single-exposure experiment. Single-shot misalignment error estimation required only 0.13 s, significantly outperforming conventional methods. We also demonstrated the effects of beam quality and pretraining using experimental data. The proposed method exhibited strong robustness, can handle high-precision focusing systems with complex or dynamic wavefront errors, and provides an important basis for intelligent control of future synchrotron radiation beamlines.

Keywords: Deep learning; Synchrotron radiation; Optics alignment

1. Introduction

Synchrotron micro- and nano-focused X-rays have been widely used to explore the microstructures of materials in energy research, materials science, and life sciences. For a coherent source, smaller focusing spots yield higher coherent fluxes, improving the resolution of reconstructed images. To meet the high demands placed on X-ray focusing systems, aberration-free focusing is desired for perfect wavefronts. The Kirkpatrick–Baez (KB) mirror pair is a typical X-ray focusing system consisting of two perpendicular mirrors with elliptical or parabolic shapes. To achieve an ideal focusing spot, in addition to high-precision surface figure, KB mirrors must be placed accurately along the optical axis.

Conventionally, KB mirrors are aligned by trial and error. A knife-edge scan is the simplest method for measuring the focusing spot size; however, it can only provide one-dimensional intensity profile information per scan. The Hartmann wavefront sensor can provide two-dimensional wavefront information directly in one shot, but the device is expensive and has relatively low spatial resolution. Grating-based shearing interferometry is another approach for obtaining two-dimensional wavefront information, but these measurements require accurate alignment of interferometers and nearly perfect fabrication of gratings. Recently, a near-field speckle-tracking method based on digital image correlation was developed for extracting wavefront information, and relevant characteristics such as the effect of diffuser grain size and correlation subset size were carefully explored. A speckle-based scanning method was also proposed to directly provide information on pitch angle errors; however, the scanning process is time-consuming, and the acquired alignment information is limited to pitch angle and wavefront curvature. Another speckle analysis method based on coherent X-ray beams was developed for obtaining alignment error-related information in one shot based on the inverse relationship between speckle and focusing spot sizes, though a few more shots were required to attenuate noise.

In recent years, the rapid development of deep learning has extended to the field of optical systems. A convolutional neural network was trained for estimating Zernike coefficients of wavefront aberration using a preconditioner to increase the number of informative pixels in the detected image. A deep residual wavefront learning method was also proposed to extend the usable range of Lyot-based low-order wavefront sensors. Instead of using images as network input, Tchebichef moments were introduced to extract features of point-spread functions and passed to a multiple neural network. In the telescope field, a CNN was used for determining initial estimates of Zernike coefficients from PSF images for further iterative refinement. In microscopic systems, deep learning has been used for autofocusing and denoising. In the field of X-rays, neural networks have been applied to screening macromolecular crystallographic diffraction images, classification of crystal structures, tomographic denoising, and Bragg peak analysis. Machine learning-based methods have also been applied to the diagnosis and tuning of accelerators. However, relevant applications of deep learning in X-ray optics alignment and metrology remain lacking.

In this study, we propose a single-exposure deep learning-based KB mirror alignment error estimation method. We use detected images to train a CNN for estimating alignment errors. To improve the performance of the deep learning methods, we place a thin scatterer in the focal plane to modulate the direct beam and generate speckles in the detector plane, as these speckles can carry wavefront error-related information from the focal plane. We first train the CNN using simulated data, then fine-tune the network using experimental data. The influence of beam quality and pretraining with simulated data on misalignment error estimation is also discussed.

2. Method

As shown in [Figure 1: see original paper], a thin scatterer was placed in the focal plane to generate speckles in the direct-beam images on the detector as a modulation preconditioner for training the CNN. Unlike other image-based machine learning methods, we used both direct beam and speckle-modulated images for training the network to extract features from the detected images, and we compared the network's performance on these two types of images.

For a narrowband incident light wave $u(P,t)$, the mutual intensity can be defined as $J_i(P_1,P_2) = \langle u(P_1,t)u^*(P_2,t)\rangle$. After passing through a thin scatterer with a complex transmittance function $t(P)$, the exit mutual intensity becomes $J_t(P_1,P_2) = t(P_1)t^*(P_2)J_i(P_1,P_2)$. The propagation of the mutual intensity from the scatterer surface S_1 to the detector surface S_2 can be written as

$$J_t(Q_1,Q_2) = \int_{S_1}\int_{S_1} J_i(P_1,P_2)\exp\left[j\frac{2\pi}{\lambda}(r_1-r_2)\right]\frac{\chi(\theta_1)\chi(\theta_2)}{\lambda^2 r_1 r_2}\,d\sigma_1\,d\sigma_2$$

where λ represents the center wavelength of the narrowband beam, r_i is the distance between P_i and Q_i, and $\chi(\theta_i)$ is the obliquity factor at position P_i.

Because only intensity information can be detected, we consider $I(Q) = J(Q,Q)$ in the detector plane. Let (x,y) denote position Q in the detector plane, and let (α,β) denote position P in the scatterer plane. By assuming a Schell-model field $J_i(\alpha_1,\beta_1,\alpha_2,\beta_2) = A(\alpha_1,\beta_1)A(\alpha_2,\beta_2)\mu(\alpha_1-\alpha_2,\beta_1-\beta_2)$ and using the Fresnel approximation, we obtain the following simplified expression:

$$I(x,y) = \int_{S_1}\int_{S_1} t(\alpha_1,\beta_1)t^*(\alpha_2,\beta_2)A(\alpha_1,\beta_1)A(\alpha_2,\beta_2)\mu(\alpha_1-\alpha_2,\beta_1-\beta_2)\exp\left[j\frac{2\pi}{\lambda z}\left((\alpha_1^2-\alpha_2^2)+(\beta_1^2-\beta_2^2)\right)\right]\exp\Big[$$

As indicated by the above equation, the scatterer perturbs the detected intensity. Using deep learning methods, we can extract the speckle difference due to the scatterer by learning from the detected images with and without scattering. Deep learning methods use backpropagation and gradient descent to optimize the parameters of artificial neural networks. A typical artificial neural network contains many layers of neurons. Each layer receives the output of the previous layer as input and feeds processed input to the next layer; the last layer outputs the model predictions. The gradient of the loss function capturing the difference between model predictions and ground truth is propagated backward through the network starting from the output layer, according to the backpropagation algorithm. Specifically, for a layer of neurons, the forward and backward propagation rules are as follows:

Forward propagation: $z = \varphi(w^T x + b)$

Backward propagation: $\frac{\partial L}{\partial x} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial x}$

where z indicates the output of a given layer, w and b represent the weight and bias parameters of individual neurons in this layer, respectively, x is the input to this layer (also the output from the previous layer), L is the loss function reflecting the optimization objective, and constructs such as $\partial L/\partial x$ are backpropagation gradients formulated as Jacobian matrices.

For image processing, CNNs have become much more popular than vanilla multilayer neural networks, according to recent advances in deep learning. Compared with conventional multilayer neural networks, CNNs leverage the ideas of local connectivity, parameter sharing, and pooling of hidden units. Thus, they are more computationally efficient, exploit the 2D topology of image pixels, and account for translation invariance. The backbone CNN in the present study was ResNet50 [Figure 2: see original paper], which demonstrated good performance on various vision tasks and has been a popular backbone network for many vision-related tasks. We added a dropout layer with a drop ratio of 0.2 to mitigate overfitting and a fully connected layer for regression.

3. Simulation Results

Considering the sensitivity of different alignment errors associated with KB mirrors, we focused on six main degrees of freedom: (1,2) vertical and horizontal pitch angles, (3,4) vertical and horizontal curvatures, (5) astigmatism, and (6) defocus. Single-micron focusing was pursued based on the experimental environment.

[Figure 3: see original paper] shows example simulated optical elements, including a scatterer composed of Cu particles and simulated figure error of the KB mirror. First, a simulation was performed to verify the feasibility of the proposed method. The optical layout and photon energy of 10 keV were the same as in the experiment at beamline BL15U (more detailed information is provided in Section 4). The simulation was conducted in 1D for horizontal and vertical focusing first, then combined to form a 2D focus by shearing and matrix multiplication, and finally propagated to the detector plane to reduce computational complexity. The source was simulated using an array of point sources with random phases sampled from a uniform distribution within a preset range. Based on the diffraction fringes generated by the source and the relationship between fringe contrast and coherence length of the Gauss-Schell model, the transverse coherence length values at the source position were measured as less than 3 m horizontally and 42 m vertically in simulations (averages over one thousand replicates).

A thin scatterer was simulated using five layers of randomly distributed 500-nm Cu particles, as shown in [FIGURE:3(a)]. Particle sizes were drawn from a Gaussian distribution (mean: 500 nm; standard deviation: 0.25 of the mean). The tangential figure of the KB mirrors was simulated using an elliptical curve with randomly added high- and low-frequency errors. High-frequency figure errors were drawn from a Gaussian distribution (root mean square: 0.2 nm; peak-to-valley: 1 nm), while low-frequency figure errors were drawn using a sine function of a complete cycle with 30 nm peak-to-valley and a random initial phase. Medium- or low-frequency figure errors yielded coherent stripes in far-field images, different from lower-frequency alignment errors. To overcome possible alignment error-related estimation bias introduced by figure errors, the mirror figure was not fixed in the simulations; for each generated figure, the simulation was repeated 10 times, and a new figure was generated and used. An example error curve is shown in [FIGURE:3(b)]. The lengths of the horizontal focusing mirror (HFM) and vertical focusing mirror (VFM) were both 200 mm to simulate the KB mirrors used in experiments. The ideal focus size without alignment errors was approximately 1 m in both directions.

The tolerances for the six main misalignment parameters were estimated using the Strehl ratio based on the Marechal criterion. The Strehl ratio was averaged over 100 simulations with a random source field sampled from a uniform distribution, and the corresponding results are listed in .

For better reliability and robustness, we also added deviations of other degrees of freedom to the simulations as experimental noise to avoid overfitting to an ideal environment. These errors included detector position error, roll angle error, and beam in/out-axis error (the distance that the mirror moved from the optical axis in the tangential plane). We did not include roll angle error in the estimation because the simulation program was unable to simulate comprehensive 2D interaction of the beam and mirrors, which led to focus insensitivity to roll error, and the proposed experimental platform was unable to adjust the roll angle between mirrors. This perpendicularity error can be easily included in the estimation by adding a dimension to the output of the last fully connected layer if required.

The simulation was implemented according to the optical layout shown in [Figure 1: see original paper], and 8000 samples were generated using the simulation program for machine learning, one of which is shown in [Figure 4: see original paper]. Owing to the 1D simulation of the KB focusing process, the vertical and horizontal beam propagations were decoupled, and there was an obvious correlation between detected images for different directional points due to matrix multiplication of the focus field. The alignment parameters and error ranges are listed in , where errors were randomly drawn from a uniform distribution. The range of estimated errors was chosen to enlarge the focal spot to approximately 10 m; the range of errors acting as noise was chosen to not visibly affect the size of the focal spots.

We used the Adam optimizer for training our CNN because it generally converges faster than stochastic gradient descent. The learning rate was 10^{-4}, with a rate decrease factor of 0.978 per epoch. The batch size was 10, and training was performed for 80 epochs. Gaussian noise with a signal-to-noise ratio of 50 dB was added to images to simulate a practical noise environment for further robustness. Misalignment errors were normalized to the $[-1, 1]$ range as target values to avoid gradient explosion or imbalanced weights among estimated errors. The loss function was mean square error, commonly used in regression tasks. The method was implemented in PyTorch and executed on an NVIDIA A100 GPU with 40 GB of VRAM.

[Figure 4: see original paper] shows example simulated speckle and direct beam detector images. Training was performed using three different datasets: speckles, direct beams, and both. We combined speckle and direct-beam images in different channels for training. The results are listed in , and training results for the speckle dataset are shown in [Figure 5: see original paper]. The RMSEs between ground truth and estimated alignment errors for test datasets were 0.271 ± 0.126 for the combined dataset of speckle and direct-beam images, 0.265 ± 0.122 for the speckle image dataset, and 0.273 ± 0.126 for the direct-beam image dataset, while corresponding RMSEs for training datasets were 0.05, 0.045, and 0.047, respectively. Overfitting occurred early during training, and the RMSE of the test dataset decreased slowly from approximately epoch 10, as shown in [FIGURE:5(c). While training error RMSE continued decreasing to 0.045 over the entire training process, validation error RMSE stopped at approximately 0.3 and began oscillating with small amplitude. This occurred because randomness in the incident beam status and mirror figure strongly affected features of detected images; for one set of misalignment error parameters, detected images generated in simulations varied strongly with different incident beam and mirror figure statuses. Increasing data volume or using constant phase for incident beams and mirror figures may relieve this problem.

The best performance was obtained for the speckle dataset. Combining images from speckle and direct-beam datasets did not utilize advantages of both datasets, and resultant performance was similar to that obtained using only direct-beam images. Among the six parameters, astigmatism and defocus were the main contributors to RMSE of alignment error. Because these two parameters spanned small ranges compared with the large depth of focus, detected images had low sensitivity to changes.

Compared with estimation of Zernike wavefront coefficients, the proposed method directly obtains information about KB mirror misalignment rather than inferring misalignment from wavefront phase information provided by Zernike coefficients and wavefront amplitude information calculated from intensity information of detected images. By using an end-to-end misalignment estimation model, this method introduces fewer calculation errors.

4. Experimental Results and Discussion

The experiment was conducted at beamline BL15U of the Shanghai Synchrotron Radiation Facility using 10 keV photons. The secondary source aperture was 200×30 m^2 for improving beam coherence, and coherence lengths at the aperture were 4.25 m in the horizontal direction and 66.55 m in the vertical direction. Bendable KB mirrors (active length: 200 mm) were pre-aligned with a silicon substrate at the beamline. The slope error along the optical axis of the KB mirrors was less than 0.5 rad, and surface roughness of Pt and Rh coatings was at least 0.3 nm RMS. Mirror ends were equipped with two bending rods above and below the surfaces for applying bending force, and two fixed rods supported and stabilized the mirrors. KB mirror curvatures were adjusted using actuators attached to the benders. A thin sandpaper scatterer was placed at the focus. A microscope objective lens system (Optique Peter) coupled to a complementary metal-oxide semiconductor camera (Hamamatsu) was placed 2075 mm downstream from the focus. Detector pixel size was 1.625 m with 2048×2048 pixels. To pass values to our CNN, we cropped the image from the beam center to reduce background space and downsampled it to 500×500 pixels, considering network computational capacity. The position of minimal spot size based on knife-edge scan results was considered the zero-error alignment position.

Although the theoretical focus area was 1.1×1.6 m (H $\times$ V) calculated by ray tracing using the Shadow VUI simulation program, due to optical degradation from upstream components, experimental noise, and knife-edge vibration, the measured focus spot was approximately 4 m in both directions—much larger than the size used in simulations. Two sets of focus spots measured using the knife-edge scan method at different foci and pitch angles are shown in [Figure 6: see original paper].

The network trained on simulated data was used as the first guess for training on experimental data. To examine the extent to which simulations helped estimate experimental alignment errors, we also attempted to fine-tune the network trained on simulated data using experimental data, with backbone CNN and parameters determined from simulated data. We attempted to freeze parameters of the full ResNet50 network (including only convolution layers) and unfreeze parameters of the last convolution layer of ResNet50. Because simulation results differed slightly from experimental results due to simulation limitations and environmental differences between source and optical characteristics of experimental and simulation setups, estimating experimental alignment errors using the CNN trained on simulated data directly could yield relatively large estimation errors.

The learning rate was 10^{-4} and batch size was 10. Training was performed for 80 epochs on an NVIDIA A100 GPU with 40 GB of VRAM, with one epoch taking on average less than 1 s. During training, the model with the lowest RMSE on the verification set was selected. In the validation process, one estimation of misalignment error took on average 0.13 s.

The experiment was performed twice under different beam-quality conditions, generating two datasets. Overall, 1762 images and their corresponding alignment parameters (selected from a regular grid) were collected; images were divided into training and validation sets at an 8:2 ratio. As shown in [FIGURE:7(a)] and [FIGURE:7(b)], images obtained in the first experiment had many more stripes than those from the second experiment, indicating that the beam underwent serious phase modulation due to the beryllium window. Before passing data to the CNN, alignment errors were normalized; their ranges are listed in . RMSEs of training results for the two experiments are listed in , indicating that normalized RMSE achieved best accuracy of approximately 4%. Example detected images from the validation set are shown in [FIGURE:7(a–d)], and estimation results of the model trained with the full network corresponding to [FIGURE:7(d)] are shown in [FIGURE:7(e). For direct beam data, the defocus error term was zero because no sample at the focal spot enabled defocus detection. The astigmatism error was fixed because the distance between HFM and VFM at BL15U was constant. Owing to differences between image characteristics of simulated and experimental data, training convolution layers was necessary for improving estimation precision, and training the last convolution block could substantially improve performance. For training with the full trainable network, overfitting was still remarkable for data from the first experiment, while this phenomenon was much less likely for training using data from the second experiment, as seen in [FIGURE:7(f–g). Comparing learning processes between the two experiments, beam quality significantly affected the generalization capability of the neural network. Although it provided more features, phase modulation noise made it more difficult for the network to recognize useful information for alignment error estimation and caused the network to confuse different focus states; more training data is required to address this problem.

A line chart of different errors estimated by the network trained on simulation data as the first estimate is shown in [Figure 8: see original paper]. According to normalization, for validation results of the model trained using speckle data from the first and second experiments, RMSEs of VFM pitch error estimation are listed in . Compared with defocus errors, pitch and curvature errors are more sensitive because changes in pitch and curvature induce more internal structural changes in images and are easier to extract and recognize as features. With respect to simulation data, figure errors of mirrors were randomly generated to alleviate their influence on misalignment error estimation. With respect to experimental data, because the estimation model was specifically trained for a KB mirror focusing system in a fixed optical layout, figure error was not considered separately because it would not change for a given KB mirror. For further figure errors arising from in-situ elements, curvature and pitch angle errors caused by figure errors may even be compensated by estimated misalignment errors.

Misalignment error-estimation models trained on data from the first experiment performed even better than models trained on data from the second experiment when compared in terms of RMSEs of non-normalized misalignment errors. This indicates that the deep learning method has potential to overcome effects of beam noise such as stripes. In addition to the explanation that more features were introduced by the noisy beam, another possible reason for this unusual performance reversal is the effect of normalization range on estimation accuracy. For larger ranges of misalignment errors, there are larger intervals between different misalignment error data, meaning that much smaller estimation errors of normalized misalignment errors can cause large estimation errors of non-normalized misalignment errors. This also shows that the ranges of misalignment errors we chose in experiments were still far from the resolution limit of the neural network model. For smaller ranges, the network may still provide relatively accurate estimation of normalized misalignment errors, indicating that the network has potential for more accurate estimation.

Considering the overfitting problem, we assumed that the neural network model would perform better on larger datasets. We also attempted to train the network using experimental data without pretraining on simulated data. After a sufficiently long training process, the network achieved performance comparable to that of the previously trained network. However, with a pretrained network, the training process was significantly shorter.

To explore the training process of the neural network model for KB mirror misalignment error estimation more clearly, we extracted feature maps from one detected image, shown in [Figure 9: see original paper]. The first convolution layer was responsible for extracting related information from the input image and standardizing the format of feature images, focusing mainly on high-intensity loci. The first ResNet layer extracted texture-related information from the image, distinguishing speckles surrounding the center beam image and different texture patterns in the beam image. The second layer further extracted texture features, while the third layer summarized features. As shown in [FIGURE:9(d)], both localized and global features were present in feature maps of the third layer. The last convolution layer of ResNet extracted highly abstract information with a large field of view, yielding coarse-grained structural features instead of fine-textured features.

The saliency map of network attention is shown in [Figure 10: see original paper]. The network extracted information from both the detected beam and speckles surrounding the beam. It successfully avoided positions with saturated intensity at the lower and right edges of the beam that contained no information. Using this technique with a well-trained neural network that can make accurate estimations, we can trace which features were captured by the network and how they were processed to obtain results, helping verify whether the network learned the mechanics correctly rather than simply fitting data by some tricky means. By carefully examining and analyzing the feature-processing and attention capabilities of the model, we can understand the structural relationship between input images and target values, and may find directions to explain the estimation process and establish relationships between detected phenomena and underlying factors.

5. Conclusion

A machine learning-based method was presented for estimating KB mirror alignment errors. Direct-beam and speckle-modulated images were captured by a detector for a CNN to estimate alignment errors. In this study, we used simulations to predict different performances of detected images under various alignment errors and scatterer conditions, and generated training data for deep learning. We verified this method experimentally and estimated the effect of beam quality. Both experiments exhibited good estimation accuracy, proving the repeatability of this method. The results demonstrate the applicability of the proposed approach.

The proposed method can provide fast and relatively accurate alignment error estimation based on a single-exposure experiment, even under noisy beam conditions. Compared with existing methods, the proposed method is much faster and more robust, achieving best normalized RMSE accuracy of 4% on average in 0.13 s. Similar approaches can be applied to other machine learning-based beam diagnosis problems. By combining simulations and multiple experiments with related visualization technology, we aimed to provide a reliable, trustworthy, and traceable deep learning-based optical metrology approach. However, network error estimation relies heavily on beamline layout and calibration accuracy. Estimation becomes inaccurate if the beamline is modified because the network cannot learn the intrinsic relationship between alignment error and beam propagation. In addition, different networks for different beamline layouts require large amounts of data and time to train. In future studies, we will optimize the light source model and elucidate the role of mirror figure errors in beam propagation. A more accurate model will significantly improve the accuracy of the proposed method. For further applications, the robustness, reliability, and generalization of machine learning methods in optical tasks should be improved. This framework can be applied to a wide range of optical imaging systems involving alignment of optical elements. With improvements in analytical ability of neural networks and development of adaptive optical techniques, we are hopeful that this method will eliminate dependency on specific scenarios and enable learning from a generalized beam-generating scheme, avoiding repetitive training on different optical layouts while providing better estimations. This will help advance toward fully intelligent beamline control.

Funding. The National Key Research and Development Program (2021YFA1601000), National Natural Science Foundation of China (12175294), and Natural Science Foundation of Shanghai, China (21ZR1471500).

Acknowledgments. The authors thank Yanan Fu and Guohao Du from Beamline 13HB for their assistance in testing the focusing system and verifying the method described in this paper.

References

[1] P. Schöppe, C.S. Schnohr, M. Oertel, et al., Improved Ga grading of sequentially produced Cu(In, Ga)Se_2 solar cells studied by high resolution X-ray fluorescence. *Appl Phys Lett.* 106, 013909 (2015). https://doi.org/10.1063/1.4905347

[2] C. Sanchez-Cano, D. Gianolio, I. Romero-Canelon, et al., Nanofocused synchrotron X-ray absorption studies of the intracellular redox state of an organometallic complex in cancer cells. *Chem Commun.* 55, 7065-7068 (2019). https://doi.org/10.1039/C9CC01675A

[3] P. Kirkpatrick, A.V. Baez, Formation of optical images by X-rays. *J Opt Soc Am.* 38, 766-774 (1948). https://doi.org/10.1364/JOSA.38.000766

[4] D.-C. Zhu, J.-H. Yue, Y.-F. Sui, et al., Performance of beam size moni-

tor based on Kirkpatrick–Baez mirror at SSRF. *Nucl Sci Tech.* 29 (2018). https://doi.org/10.1007/s41365-018-0477-y

[5] S. Handa, T. Kimura, H. Mimura, et al., Extended knife-edge method for characterizing sub-10-nm X-ray beams. *Nucl Instrum Methods Phys Res, Sect A.* 616, 246-250 (2010). https://doi.org/10.1016/j.nima.2009.10.131

[6] B.C. Platt, R. Shack, History and principles of Shack-Hartmann wavefront sensing. *J Refract Surg.* 17, S573-S577 (2001). https://doi.org/10.3928/1081-597X-20010901-13

[7] H. Wang, S. Berujon, K. Sawhney, Development of at-wavelength metrology using grating-based shearing interferometry at Diamond Light Source. *J Phys Conf Ser.* 425, 052021 (2013). https://doi.org/10.1088/1742-6596/425/5/052021

[8] T. Weitkamp, B. Nöhammer, A. Diaz, et al., X-ray wavefront analysis and optics characterization with a grating interferometer. *Appl Phys Lett.* 86, 054101 (2005). https://doi.org/10.1063/1.1857066

[9] S. Bérujon, E. Ziegler, R. Cerbino, et al., Two-dimensional X-ray beam phase sensing. *Phys Rev Lett.* 108, 158102 (2012). https://doi.org/10.1103/PhysRevLett.108.158102

[10] N. Tian, H. Jiang, A. Li, et al., Influence of diffuser grain size on the speckle tracking technique. *J Synchrotron Radiat.* 27, 146-157 (2020). https://doi.org/10.1107/S1600577519015200

[11] N. Tian, H. Jiang, A. Li, et al., High-precision speckle-tracking X-ray imaging with adaptive subset size choices. *Sci Rep.* 10, 1-12 (2020). https://doi.org/10.1038/s41598-020-71158-9

[12] T. Zhou, H. Wang, O. Fox, et al., Auto-alignment of X-ray focusing mirrors with speckle-based at-wavelength metrology. *Opt Express.* 26, 26961-26970 (2018). https://doi.org/10.1364/OE.26.026961

[13] T. Inoue, S. Matsuyama, J. Yamada, et al., Generation of an X-ray nanobeam of a free-electron laser using reflective optics with speckle interferometry. *J Synchrotron Radiat.* 27, 883-889 (2020). https://doi.org/10.1107/S1600577520006980

[14] Y. Nishizaki, M. Valdivia, R. Horisaki, et al., Deep learning wavefront sensing. *Opt Express.* 27, 240-251 (2019). https://doi.org/10.1364/OE.27.000240

[15] G. Allan, I. Kang, E.S. Douglas, et al., Deep residual learning for low-order wavefront sensing in high-contrast imaging systems. *Opt Express.* 28, 37103-37113 (2020). https://doi.org/10.1364/OE.397790

[16] G. Ju, X. Qi, H. Ma, et al., Feature-based phase retrieval wavefront sensing approach using machine learning. *Opt Express.* 26, 31767-31783 (2018). https://doi.org/10.1364/OE.26.031767

[17] S.W. Paine, J.R. Fienup, Machine learning for improved image-based wavefront sensing. *Opt Lett.* 43, 1235-1238 (2018). https://doi.org/10.1364/OL.43.001235

[18] H. Ding, F. Li, Z. Meng, et al., Auto-focusing and quantitative phase imaging using deep learning for the incoherent illumination microscopy system. *Opt Express.* 29, 26385-26403 (2021). https://doi.org/10.1364/OE.434014

[19] J. Liao, X. Chen, G. Ding, et al., Deep learning-based single-shot autofocus method for digital microscopy. *Biomed Opt Express.* 13, 314-327 (2022). https://doi.org/10.1364/BOE.446928

[20] S. Montresor, M. Tahon, P. Picart, Review of deep learning based de-noising algorithms for phase imaging and applications to high-speed coherent imaging, in *OSA Imaging and Applied Optics Congress 2021* (3D, COSI, DH, ISA, pcAOP), Washington, DC, 19 July 2021 (Optica Publishing Group).

[21] T.-W. Ke, A.S. Brewster, S.X. Yu, et al., A convolutional neural network-based screening tool for X-ray serial crystallography. *J Synchrotron Radiat.* 25, 1432-1440 (2018). https://doi.org/10.1107/S1600577518004873

[22] S. Lolla, H. Liang, A.G. Kusne, et al., A semi-supervised deep-learning approach for automatic crystal structure classification. *J Appl Crystallogr.* 55, 908-918 (2022). https://doi.org/10.1107/S1600576722006069

[23] Y.-J. Ma, Y. Ren, P. Feng, et al., Sinogram denoising via attention residual dense convolutional neural network for low-dose computed tomography. *Nucl Sci Tech.* 32 (2021). https://doi.org/10.1007/s41365-021-00874-2

[24] Z. Liu, H. Sharma, J.-S. Park, et al., BraggNN: fast X-ray Bragg peak analysis using deep learning. *IUCrJ.* 9, 104-113 (2022). https://doi.org/10.1107/S2052252521011258

[25] L.-Y. Zhou, H. Zha, J.-R. Shi, et al., A non-invasive diagnostic method of cavity detuning based on a convolutional neural network. *Nucl Sci Tech.* 33 (2022). https://doi.org/10.1007/s41365-022-01069-z

[26] Y.-B. Yu, G.-F. Liu, W. Xu, et al., Research on tune feedback of the Hefei Light Source II based on machine learning. *Nucl Sci Tech.* 33 (2022). https://doi.org/10.1007/s41365-022-01018-w

[27] J.W. Goodman, *Statistical Optics* (John Wiley & Sons, 2015).

[28] K. He, X. Zhang, S. Ren, et al., Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[29] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in *3rd International Conference on Learning Representations*, San Diego, CA, USA, 7-9 May 2015.

[30] H. Wang, S. Yan, F. Yan, et al., Research on spatial coherence of undulator source in Shanghai Synchrotron Radiation Facility. *Acta Phys Sin.* 61, 144102 (2012). https://doi.org/10.7498/aps.61.144102

[31] Z. Lili, Y. Shuai, J. Sheng, et al., Hard X-ray micro-focusing beamline at SSRF. *Nucl Sci Tech.* 26 (2015). https://doi.org/10.13538/j.1001-8042/nst.26.060101

Note: Figure translations are in progress. See original paper for figures.