



Research on the X-ray polarization deconstruction method based on hexagonal convolutional neural network

Ya-Nan Li^{1,2} · Jia-Huan Zhu³ · Huai-Zhong Gao^{1,2} · Hong Li³ · Ji-Rong Cang⁴ · Zhi Zeng² · Hua Feng³ · Ming Zeng^{1,2}

Received: 14 November 2023 / Revised: 21 April 2024 / Accepted: 23 April 2024 / Published online: 12 January 2025

© The Author(s), under exclusive licence to China Science Publishing & Media Ltd. (Science Press), Shanghai Institute of Applied Physics, the Chinese Academy of Sciences, Chinese Nuclear Society 2024

Abstract

Track reconstruction algorithms are critical for polarization measurements. Convolutional neural networks (CNNs) are a promising alternative to traditional moment-based track reconstruction approaches. However, the hexagonal grid track images obtained using gas pixel detectors (GPDs) for better anisotropy do not match the classical rectangle-based CNN, and converting the track images from hexagonal to square results in a loss of information.

We developed a new hexagonal CNN algorithm for track reconstruction and polarization estimation in X-ray polarimeters, which was used to extract the emission angles and absorption points from photoelectron track images and predict the uncertainty of the predicted emission angles. The simulated data from the PolarLight test were used to train and test the hexagonal CNN models. For individual energies, the hexagonal CNN algorithm produced 15%–30% improvements in the modulation factor compared to the moment analysis method for 100% polarized data, and its performance was comparable to that of the rectangle-based CNN algorithm that was recently developed by the Imaging X-ray Polarimetry Explorer team, but at a lower computational and storage cost for preprocessing.

Keywords X-ray polarization · Track reconstruction · Deep learning · Hexagonal conventional neural network

1 Introduction

Astronomical X-ray polarimetry is a powerful tool for probing the magnetic fields, geometries, and emission physics of high-energy astrophysical sources [1, 2]. Astronomical X-ray polarization measurements originated in the 1960s to detect the soft X-ray polarization of the crab nebula,

Scorpius X-1, and other objects using Bragg diffraction and Thomson scattering polarimeters [3]. However, the limited sensitivity of the polarimeter has stalled astronomical X-ray polarization measurements for more than 40 years since the experiments on the OSO-8 satellite in 1968.

The photoelectric effect dominates light–matter interactions in the energy range of a few kiloelectron volts. The differential cross section of photoelectrons is proportional to $\cos^2\varphi$; here, $\varphi = \varphi_e - \varphi_0$, where φ_e is the azimuthal angle of the photoelectron and φ_0 is the electric vector position angle (EVPA) of the X-ray [4]. Therefore, the polarization fraction and polarization angle, EVPA, of the X-ray source can be obtained by measuring the emission angles of numerous photoelectrons. With the development of micropattern gas detectors, polarimetry based on the photoelectric effect has become possible by measuring the emission angles of photoelectrons, which has greatly improved the polarization sensitivity [5]. The PolarLight CubeSat test [6–8], Imaging X-ray Polarimetry Explorer (IXPE) mission [9], the enhanced X-ray Timing and Polarimetry (eXTP) mission [10, 11], and the Cosmic X-ray Polarization Detection

This work was supported by the National Natural Science Foundation of China (No. 12025301) and the Tsinghua University Initiative Scientific Research Program.

✉ Ming Zeng
zengming@tsinghua.edu.cn

¹ Key Laboratory of Particle & Radiation Imaging, Tsinghua University, Ministry of Education, Beijing 100084, China

² Department of Engineering Physics, Tsinghua University, Beijing 100084, China

³ Department of Astronomy, Tsinghua University, Beijing 100084, China

⁴ StarDetect Co., Ltd, Beijing 100084, China

(CXPD) [12, 13] all use gas pixel detectors (GPDs) to detect polarization.

However, owing to Coulomb scattering, transverse diffusion during drift, and electronic noise, the reconstruction of emission angles from photoelectron tracks is complicated. The performance of the photoelectron track reconstruction algorithm significantly affects the polarimeter sensitivity.

Recently, there have been two types of track reconstruction methods: traditional algorithms, which include the moment analysis method [14], adaptive cut method [15], and graph-based method [16], and reconstruction methods based on convolutional neural networks (CNNs) [17–19], which demonstrate great advantages in track reconstruction owing to their powerful image processing capabilities. However, major space polarimetry missions use a hexagonal-pixel application specific integrated circuit (ASIC) to read track images for better isotropy [6–11], resulting in photoelectron track images with a hexagonal-pixel structure (Fig. 1). Existing CNN-based methods use classical rectangle-based CNNs with an additional step to convert hexagonal-pixel track images into approximate square-pixel track images, which results in the loss of information in the photoelectron image.

Therefore, the development of CNN methods that match the hexagonal-pixel track structure is a worthwhile research direction with good scientific significance and promising performance. Hexagonal CNNs are deep learning models based on hexagonal-pixel structures. In hexagonal CNNs, hexagonal convolutional kernels are used instead of the rectangular convolutional kernels used in classical CNNs to better capture the spatial context information in hexagonal-pixel images. The use of hexagonal CNNs to process hexagonal-pixel photoelectron track images in GPD is expected to achieve better polarization reconstruction.

In this study, we proposed a new X-ray polarization reconstruction method based on hexagonal CNNs. The remainder of this paper is organized as follows. Section 2 briefly introduces hexagonal CNNs and uncertainty quantification in

deep learning. Section 3 describes the training procedure for the hexagonal CNNs for photoelectron track reconstruction. Section 4 presents the prediction and reconstruction results of the hexagonal CNN method. Finally, Sect. 5 concludes the paper and presents prospects for future development.

2 Hexagonal CNNs and uncertainty quantification in deep learning

2.1 Hexagonal CNNs

CNNs have received considerable attention in recent years owing to their excellent performance in computer vision and big data applications [20–24]. With the increase in their application fields, classical CNNs based on a Cartesian architecture can no longer meet the demands of complex problems. Many studies have made significant advances in the design of network architectures and convolutional operations, generalizing CNNs for multi-view applications [25], non-Euclidean spaces [26], and other domains.

Typically, images are acquired using square sensor arrays. However, square grids are not the best solution for planar segmentation. Compared with square grids, hexagonal grids have many advantages such as sixfold rotational symmetry, a smaller edge-to-area ratio, and equidistant neighbors. Hexagonal grids are widely used in cosmological, astrophysical, and visual systems.

Hexagonal CNNs are a class of deep learning networks based on hexagonal grids, in which a hexagonal convolution kernel is used to replace the rectangular convolution kernel in classical CNNs. The differences between the two types of convolution kernels are illustrated in Fig. 2. Compared with classical CNNs, hexagonal CNNs have better symmetry and exhibit unique advantages for aerial scenes and geospatial information [27].

Despite the abovementioned advantages over classical CNNs, hexagonal CNNs have a higher computational complexity and are generally more difficult to train. The existing research on hexagonal CNNs involves two main approaches.

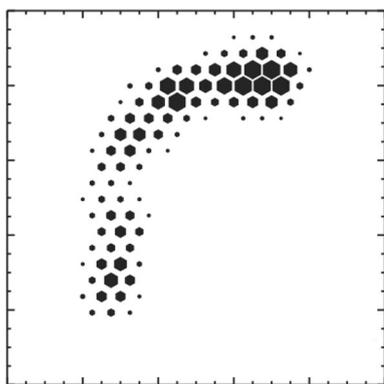


Fig. 1 Typical photoelectron track image in GPD

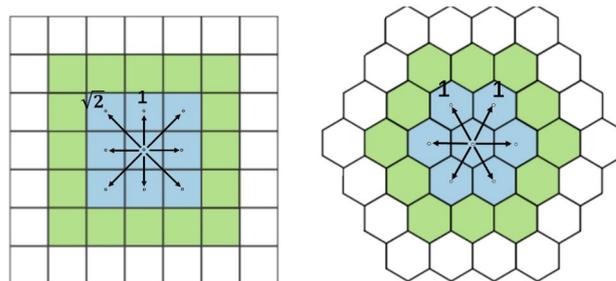


Fig. 2 (Color online) Rectangular convolution kernel (left) and hexagonal convolution (right)

One is to implement hexagonal convolution by reusing existing highly optimized rectangle-shaped convolution routines, such as HexagDLY [28] and HexagonNet [29], whereas other studies have focused on native hexagonal CNN architectures that can implement hexagonal convolutional operations directly, such as HexCNN [30]. Although native hexagonal CNNs have tremendous advantages in terms of their training time and memory space cost, they cannot yet be implemented on GPUs to exploit their efficient parallel computation to accelerate model training and inference [30]. HexagDLY is a Python library that performs convolution and pooling operations on hexagonal-pixel data. Figure 3 shows a convolutional implementation with a hexagonal kernel size of 1 in HexagDLY as an example [29]. We constructed a hexagonal CNN architecture for track reconstruction using HexagDLY, considering its flexibility and user-friendliness. The detailed hexagonal CNN architecture for photoelectron track reconstruction is discussed in Sect. 3.

2.2 Uncertainty quantification

Deep learning uncertainty estimation is also a popular research direction, which allows a neural network to output not only prediction \hat{y} for input x , but also predictive uncertainty $\sigma_{\hat{y}}$, greatly expanding the application fields of deep learning. Uncertainty estimation is also helpful for X-ray polarization reconstruction, as confirmed in [18, 19].

There are two main types of uncertainties in deep learning: aleatoric uncertainty (also known as data uncertainty) and epistemic uncertainty (also known as model uncertainty) [31]. Aleatoric uncertainty is used to assess the data

uncertainty that arises because of class overlap or inherent noise in the data and cannot be reduced by collecting more data. Epistemic uncertainty is used to assess the model uncertainty caused by a lack of cognition regarding the distribution of the data or an inadequate model structure. Theoretically, epistemic uncertainty can be reduced using more complex models, expanding the data, or using regularization techniques [32].

The aleatoric uncertainty can be modeled by augmenting the loss function. For example, assuming that the noise of the data obeys a Gaussian distribution (i.e., $\epsilon \sim N(0, \sigma^2)$), the predicted output distribution of the model for a given input, x , is $N(\hat{y}, \sigma^2)$. The loss function that predicts aleatoric uncertainty σ_a can be obtained by minimizing the negative log-likelihood (NLL) loss function for all the training data (as seen in Eq. 1).

$$L(y_i|x_i) = \frac{\log(\hat{\sigma}_a^2(x_i))}{2} + \frac{\|y_i - \hat{y}(x_i)\|_2^2}{2\hat{\sigma}_a^2(x_i)} \tag{1}$$

Epistemic uncertainty is significantly more difficult to quantify than aleatoric uncertainty, although many methods to do so have been proposed, including Bayesian neural networks (BNNs), deep ensembles, and evidential deep regression (EDR).

BNNs introduce priori assumptions to model epistemic uncertainty by setting a prior distribution, ω , upon the weight parameters of a neural network and using dataset D to derive the posterior distribution, $P(\omega|D)$, of ω . However, BNNs are difficult to apply in practice because posterior

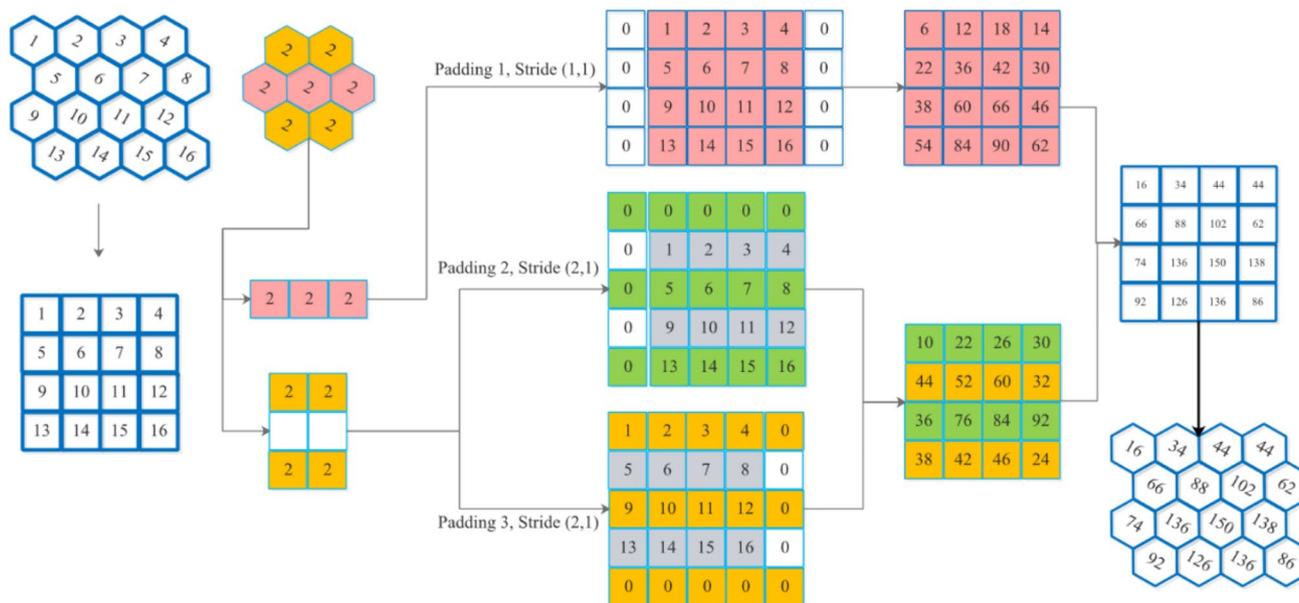


Fig. 3 (Color online) Example of hexagonal convolution with a kernel size of 1 in HexagDLY

distribution $P(\omega|D)$ is usually intractable. Approximation algorithms for BNNs have been proposed to estimate the epistemic uncertainty, such as variational inference and Monte Carlo dropouts.

Deep ensembles are another powerful approach for modeling epistemic uncertainty and have been widely used in many applications. A deep ensemble uses several base models and can generate multiple predictions, $\{\hat{y}_j\}_{j=1}^M$, for the same input, x , where M is the number of models in the deep ensemble. The variance in the predictions can be used as the epistemic uncertainty. Deep ensembles are easy to implement and can achieve as good or better uncertainty estimation than BNN approximation algorithms.

Furthermore, EDR directly learns the higher-order distribution of the neural network output. It uses a deterministic network to learn both the aleatoric and epistemic uncertainties by placing evidential priors over the original loss function that predicts the aleatoric uncertainty. EDR has achieved satisfactory results in many applications. However, a recent study found that EDR has theoretical shortcomings in terms of its mathematical foundations [33].

Taking all these considerations into account, we chose to estimate the aleatoric uncertainty by augmenting the loss function and estimate the epistemic uncertainty using deep ensembles, which are more stable and easier to implement.

3 Hexagonal CNN model training for photoelectron track reconstruction

X-ray polarization reconstruction algorithms are typically divided into two steps. First, the track features are extracted from individual photoelectron track images (also called track reconstruction), which typically include the photoelectron emission angles (φ), absorption points (x, y), and photoelectron energies (E). Subsequently, the polarization parameters (polarization fraction and EVPA) of the X-ray source are estimated based on the emission angles extracted from a large number of photoelectrons. Among these, the extraction of track features from blurred photoelectron track images is the key to polarization reconstruction. The following section describes the implementation of the hexagonal CNN-based photoelectron track feature reconstruction algorithm in detail.

3.1 Dataset

Supervised learning was used for the photoelectron track reconstruction. Because the true emission angle, φ , in the experimental data is unknown following a distribution of $\cos 2\varphi$, the dataset used for track reconstruction had to be generated by simulation.

PolarLight is a small X-ray GPD onboard a CubeSat that performs on-orbit scientific observations. An ASIC designed by the INFN-Pisa group with a pixel matrix of 352×300 (105 k pixels) and hexagonal pixel with a pitch of $50 \mu\text{m}$ is used for track readout in PolarLight [34]. The PolarLight test utilized a Monte Carlo Geant4/Garfield simulation, and the consistency was validated using experimental data. A simulation algorithm was used to generate the photoelectron track dataset.

The photoelectron track features included the emission angles, absorption points, and photoelectron energies. Because the reconstruction of photoelectron energies is relatively simple and can be done well using non-CNN algorithms, we only reconstructed (φ, x, y) in this study.

The dataset used for CNN model training should be uniformly distributed; otherwise, the model may suffer from overfitting, low prediction accuracy, or biased prediction results. Hence, the parameters of the incident X-rays in the simulation algorithms were set as follows.

1) To ensure a uniform distribution of emission angles, the polarization of the incident X-rays was set to zero. In other words, the incident X-rays were unpolarized.

2) To ensure that the hexagonal CNN model performed well in the tracking feature extraction for the entire detector plane, the coordinates (x, y) of the incident X-rays were uniformly distributed.

3) Because the effective energy range of PolarLight is 2–8 keV, the incident X-rays were uniformly distributed in the range of 2–9 keV to ensure that the hexagonal CNN model was adequately trained for the data at the edge of the energy interval. Because low-energy photoelectron track images are noisy, and it is difficult to extract track features from them, the dataset was not expanded to include lower energies.

We simulated 870,050 photoelectron tracks with uniform distributions of emission angles, absorption points, and energies and then split them into a training set (90%), validation set (5%), and test set (5%).

It is important to note that photoelectron tracks are generated not only by photons interacting with the gas in the GPD, but also by photons interacting with the detector components outside the gas volume (e.g., the beryllium window and gas electron multiplier (GEM)), in which case the photoelectrons lose some of their energy and produce a low-energy tail in the energy histogram [18, 19]. It is often difficult to recover emission angles from these tracks. Our study focused on reconstructing the photoelectron tracks generated within the GPD gas volume; therefore, tail tracks were removed from the dataset. In addition, photoelectron tracks, particularly those of high-energy photoelectrons, may pass through the detector without completely depositing their energy. These tracks were removed.

Because the HexagDLY used in this study was implemented based on a rectangle-shaped convolution, image preprocessing was required to convert the hexagonal grid tracks into square grid images. Figure 4 shows an example of the image preprocessing. In order to take full advantage of the sixfold rotational symmetry of the hexagonal grid images, each photoelectron track image was converted into three input images, including those that were unrotated, rotated by +60°, and rotated by -60°. Furthermore, considering the photoelectron track length and pixel size of the readout ASIC, we set the photoelectron track image size to 64 × 64.

3.2 Loss function

Because track reconstruction is a multitasking problem, it was necessary to separately establish loss functions for the emission angles and absorption points.

The emission angles of photoelectrons are periodic and their distribution is more consistent with a von Mises (VM) distribution, which is a continuous probability distribution with a range of 0–2π and is the circular analog of the normal distribution on a line. To predict the epistemic uncertainty, the NLL of the von Mises distribution is a better choice than the Gaussian NLL described in Sect. 2.2. The loss function of the emission angles based on the von Mises distribution of a single hexagonal CNN model is given by Eq. (2), with a detailed description in [19]:

$$L_\varphi(v|x) = -\hat{\kappa}^a (\hat{v}_2 \cdot v_2) + \log I_0(\hat{\kappa}^a) \tag{2}$$

where I_0 is a modified Bessel function of the first kind with order 0, $v_2 = (\cos 2\varphi, \sin 2\varphi)$ considering that the X-ray polarization is associated with 2φ , and non-negative $\hat{\kappa}^a$ is the predicted aleatoric VM uncertainty parameter of the emission angle. The circular variance of the VM distribution can

be derived from $\hat{\kappa}^a$ using Eq. (3), where I_1 is a modified Bessel function of the first kind on the order of one.

$$\sigma_a^2 = 1 - \frac{I_1(\hat{\kappa}^a)}{I_0(\hat{\kappa}^a)} \tag{3}$$

Assuming that the epistemic VM uncertainty, κ^e , also follows a von Mises distribution, $VM(0, \kappa^e)$, $\hat{\kappa}^e$ can be estimated from the emission angles of a deep ensemble of M hexagonal CNN models [19].

$$\bar{R}^2 = \left(\frac{1}{M} \sum_{j=1}^M \cos 2\hat{\varphi}_j \right)^2 + \left(\frac{1}{M} \sum_{j=1}^M \sin 2\hat{\varphi}_j \right)^2 \tag{4}$$

$$\frac{I_1(\hat{\kappa}^e)}{I_0(\hat{\kappa}^e)} = \bar{R} \tag{5}$$

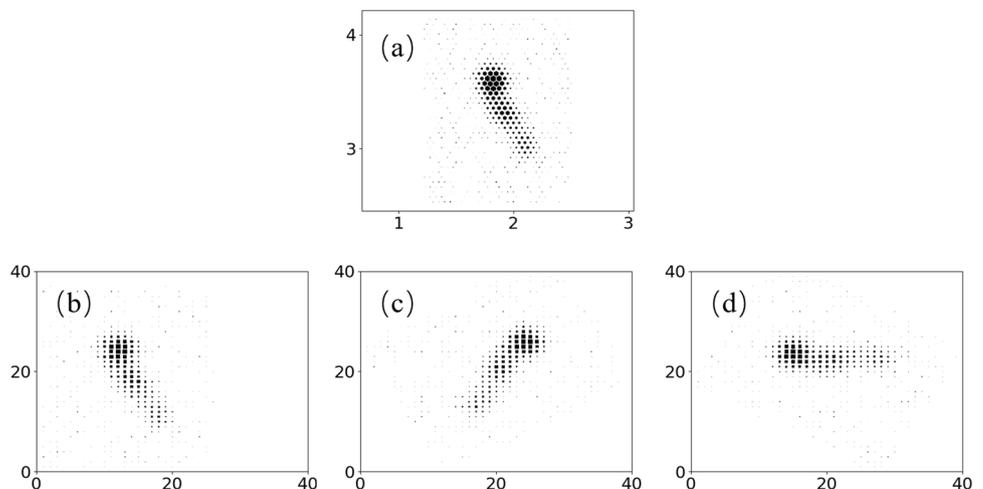
The total uncertainty variance, σ^2 , of emission angle φ can be obtained by summing aleatoric uncertainty variance σ_a^2 and epistemic uncertainty variance σ_e^2 . The total error, σ_i , of emission angle prediction φ_i for track x_i is then given by Eq. (6):

$$\sigma_i = \frac{1}{2} \sqrt{\frac{1}{M} \sum_{j=1}^M \left(1 - \frac{I_1(\hat{\kappa}_{ij}^a)}{I_0(\hat{\kappa}_{ij}^a)} \right) + \left(1 - \frac{I_1(\hat{\kappa}_i^e)}{I_0(\hat{\kappa}_i^e)} \right)}, \tag{6}$$

where a factor of 1/2 is used to transform the errors from $2\varphi_i$ to φ_i .

The loss function of the absorption points is the L_2 loss function (Eq. 7), which is commonly used in CNNs. The uncertainty in the absorption points is not the focus of this study and can be obtained using Eq. (1) combined with a deep ensemble if needed.

Fig. 4 Track image preprocessing: **a** original track image generated by the simulation algorithms; **b–d** input images of hexagonal CNN after preprocessing, including images that were unrotated (left), rotated by +60° (middle), and rotated by -60° (right)



$$L_{xy}(x_0, y_0|x) = \frac{1}{2} \|(x_0, y_0) - (\hat{x}(x), \hat{y}(x))\|_2^2 \tag{7}$$

The total loss function of a single hexagonal CNN model for track reconstruction is found as follows:

$$L = L_\varphi + \alpha \|v_1 - \hat{v}_1\|_2^2 + \beta L_{xy} + \sigma \|\omega\|_2, \tag{8}$$

where $\|v_1 - \hat{v}_1\|_2^2$ is added to allow the hexagonal CNN model to predict emission angles in a range of 2π , and the last term, $\sigma \|\omega\|_2$, is used to prevent training overfitting.

These individual loss functions are connected by three hyperparameter weights: α, β, σ . An individual hexagonal CNN model will predict a five-dimensional vector $(\cos\varphi, \sin\varphi, \kappa, x, y)$ for a given track image input, x .

3.3 Hexagonal CNN architecture

The reconstruction of the emission angles and absorption points is complex, and a simple architecture with three or four convolution layers cannot satisfy the demand for track reconstruction. Considering the photoelectron track features, we built hexagonal CNNs for track reconstruction based on the ResNet-18 architecture [35].

A residual block is the basic unit of a residual network. The hexagonal residual block (Fig. 5) used in this study was constructed using the hexagonal convolution operation provided by HexagDLY. It has a hexagonal convolution layer with a kernel size of one defined by HexagDLY, a batch normalization layer, and an ReLU activation function, followed by another hexagonal convolution layer

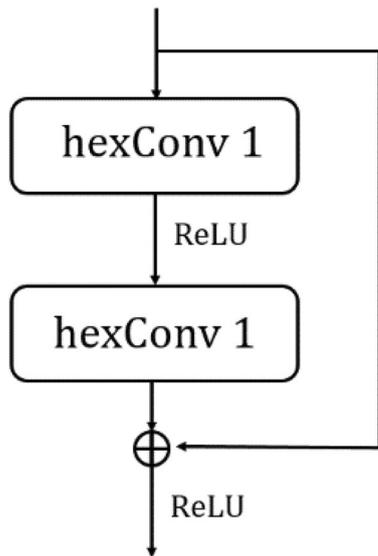


Fig. 5 Residual block structure based on hexagonal convolutional layers

with a kernel size of one and a batch normalization layer. Subsequently, the skip connection skips these layers and directly adds a rectified linear unit (ReLU) activation function. These hexagonal residual blocks are repeated to form a complete hexagonal CNN architecture for track reconstruction.

In this hexagonal CNN architecture (Table 1), the conv1 layer used a hexagonal convolution layer (kernel size = 1) and hexagonal maximum pooling layer (kernel size = 1, stride = 2) to extract track features. Then, the conv2–conv5 layers formed by the hexagonal residual blocks were used to extract deeper track features. Finally, feature maps generated by conv5 were converted into a five-dimensional vector $(\cos\varphi, \sin\varphi, \kappa, x, y)$ using an average pooling layer and a fully connected layer.

3.4 Training

A standardization operation was applied to the training data before training to prevent vanishing and exploding gradients and to accelerate convergence:

$$x_{\text{norm}} = \frac{x - \mu}{\sigma}, \tag{9}$$

where μ is the pixel mean and σ is the pixel standard deviation calculated over the entire training set of track images.

The hexagonal CNN model was optimized using stochastic gradient descent with momentum (SGD), which is a typical optimization algorithm used in deep learning. The learning rate was decreased in steps starting at 0.005. The model parameters were randomly initialized before training to provide a different start for training and to generate five different initialized hexagonal CNN models for deep ensembles. Considering the memory consumption of the hexagonal CNN model, batch sizes of 512 and 1024 were selected. The hexagonal CNN model training lasted for 150 epochs, and the hyperparameters in the loss function were $\alpha = 0.3, \beta = 0.2, \sigma = 5 \times 10^{-5}$.

Table 1 Hexagonal CNN for photoelectron track reconstruction

Layer name	Layer	Output size	Feature map
conv1	hexconv 1 hexMaxPool2d	32 × 32	64
conv2	Residual block × 2	16 × 16	64
conv3	Residual block × 2	8 × 8	128
conv4	Residual block × 2	4 × 4	256
conv5	Residual block × 2	2 × 2	512
avgpool	Average pooling	1 × 1	512
fc	Fully connected	5	

4 Results

This section reports the performance of the hexagonal CNN method for track and polarization reconstruction using simulated PolarLight track images.

4.1 Emission angle reconstruction and uncertainty estimation

The reconstruction of photoelectron emission angles is the basis of X-ray polarization reconstruction. More accurate emission angle reconstruction can significantly improve the performance of a polarization reconstruction algorithm.

The hexagonal CNN method reported here used $(\cos\varphi, \sin\varphi)$ to predict emission angle φ . As described in Sect. 3, image-rotation augmentation was used to convert the track image into three input images. Therefore, each hexagonal CNN model output three sets of predicted vectors for each track image. Because a deep ensemble method was used to obtain the epistemic uncertainty by estimating the predictions of M ($M=5$ in this study) hexagonal CNN models, there were $3M$ predictions for each track image. The emission angle for a single track predicted using the hexagonal CNN method is calculated using Eq. 10:

$$\bar{\varphi}_i = \arctan2\left(\frac{1}{3M} \sum_j^{3M} \sin\varphi_{ij}, \frac{1}{3M} \sum_j^{3M} \cos\varphi_{ij}\right). \tag{10}$$

Figure 6 shows the results of the emission angle reconstructions using the moment analysis and hexagonal CNN methods with photoelectron energies of 3 and 9 keV. Because of the inability to accurately distinguish between the beginning and end of a photoelectron track, especially at lower energies, there was a 180° confusion in the emission angle reconstruction, as shown by the two sub-bright lines parallel to the central bright line in Fig. 6. Notably, this 180° confusion did not affect the polarization reconstruction, where the EVPA ranged from $-\pi/2$ to $+\pi/2$. It can be seen that compared to the moment analysis method, the hexagonal CNN method distinguished the beginning and end of the track with higher accuracy and had a higher emission angle reconstruction accuracy.

The root-mean-square error (RMSE) is calculated using Eq. 11 to evaluate the accuracy of the emission angle reconstruction. Figure 7 shows the RMSE of the emission angles as a function of the incident X-ray energy for both the moment analysis and hexagonal CNN methods on an unpolarized PolarLight dataset:

$$RMSE = \sqrt{\frac{1}{N} \sum_i^N (\hat{\varphi}_i - \varphi_i)^2}, \tag{11}$$

where $\hat{\varphi} - \varphi$ is collapsed into the range $-\frac{1}{2\pi} \sim \frac{1}{2\pi}$ because the polarization reconstruction only depends on 2φ .

The hexagonal CNN method did not significantly improve the reconstruction accuracy of the emission angles

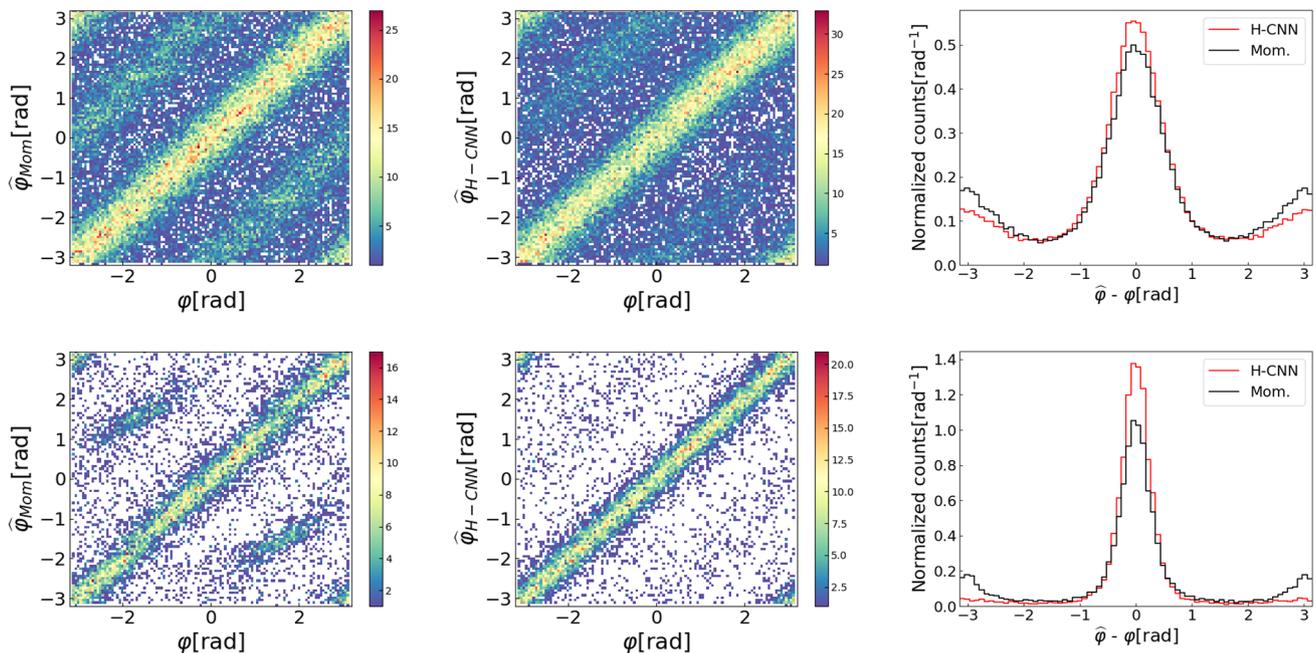


Fig. 6 (Color online) Emission angle reconstruction using the moment analysis (left) and hexagonal CNN methods (middle), along with histograms (right) of the differences between predicted emission

angle $\hat{\varphi}$ and true emission angle φ for the moment analysis (Mom., black) and hexagonal CNN methods (H-CNN, red) with photoelectrons energies of 3 keV (top) and 9 keV (bottom)

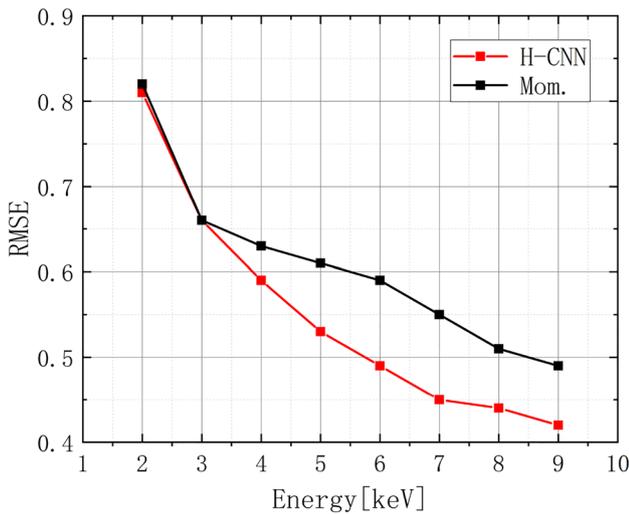


Fig. 7 RMSE of emission angle reconstruction for both the moment analysis (black) and hexagonal CNN methods (red)

compared to the moment analysis method for low-energy photoelectrons with short and noisy tracks. As the X-ray energy increased, the photoelectron tracks became longer, with clearer initial segments. Both reconstruction algorithms provided high accuracy for emission angle reconstruction, but the hexagonal CNN method was significantly better than the moment analysis method for these complex tracks.

Another important prediction for the emission angle is the predicted error, σ_i , which can be calculated using output κ of the hexagonal CNN method (Eq. 6). The predicted uncertainty in the emission angle can help screen out events with poor emission angle reconstruction, thus improving the effectiveness of X-ray polarization estimation. Because it was difficult to quantitatively determine whether the predicted error of the hexagonal CNN method was correct, we performed a basic validation of its reasonableness by comparing the predicted error with the factors that influence the effectiveness of the emission angle reconstruction.

The difficulty in reconstructing the emission angle of the photoelectron track is related to the degree of transverse electron diffusion during drift, degree of track shortening due to the projection of the 3D photoelectron track onto a 2D readout plane, and photoelectron energy. To facilitate this discussion, a coordinate system was established for the effective sensitive volume of the PolarLight GPD. The effective sensitive volume of PolarLight is 15 mm × 15 mm × 10 mm. We defined the x - y plane as the plane of the readout, the z -axis direction as the direction of the electric field, and the coordinates of the center point of the effective sensitive volume as (0,0,0).

The distributions of predicted error σ_i given by the hexagonal CNN as functions of these factors are shown in Fig. 8. Figure 8a shows the distribution of the predicted error as a function of drift distance d , which is defined as the distance between the absorption point of the photon and the GEM in the GPD for photoelectrons of 7 keV. The degree of transversal electron diffusion during drift is related to d , with standard deviation $\sigma_{\text{drift}} = \sigma_f \sqrt{d}$, where σ_f is the diffusion coefficient of the GPD gas. As the drift distance increased, the transverse diffusion became more severe and the predicted error of the emission angles increased. Figure 8b shows the distribution of the predicted error as a function of photoelectron scattering angle θ , which is defined as the angle between the directions of the incident X-rays and photoelectrons, for photoelectrons of 3 keV. When the photoelectron direction was more parallel to the readout plane, the track projected onto the readout plane was longer, resulting in a smaller predicted error in the emission angle. Figure 8c shows the distribution of the predicted error as a function of the photoelectron energy. As the photoelectron energy increased, the photoelectron track became longer, with clearer beginning segments, and the predicted error of the emission angle decreased.

Taken together, the predicted error of the emission angle obtained using the hexagonal CNN method was

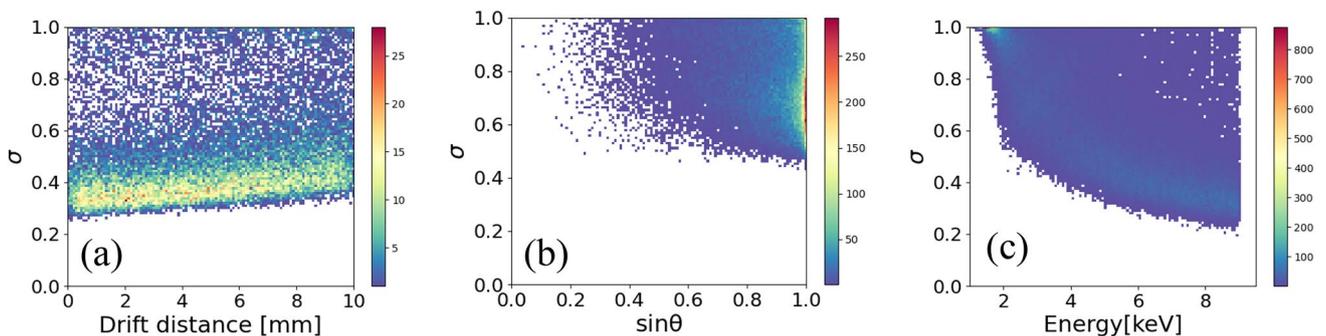


Fig. 8 (Color online) Distributions of predicted error as functions of the drift distance (left), photoelectron scattering angle (middle), and photoelectron energy (right)

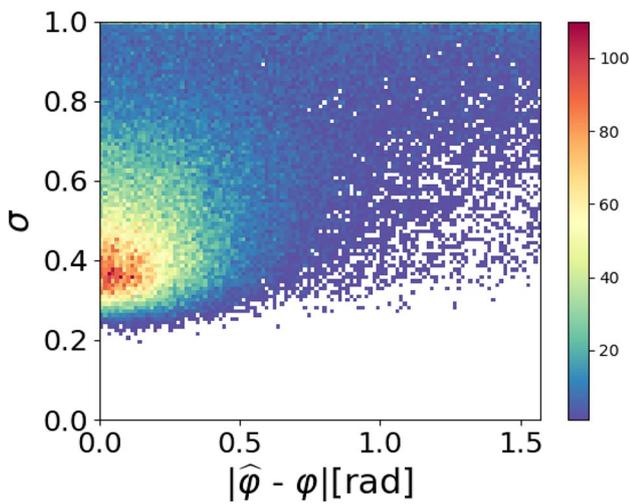


Fig. 9 (Color online) Relationship between the real emission angle reconstruction error, $|\hat{\varphi} - \varphi|$, and predicted error

reasonable, reflecting the degree of blurring of the photoelectron tracks and the difficulty in reconstructing the emission angle.

Furthermore, the relationship between the real emission angle reconstruction error, $|\hat{\varphi} - \varphi|$, and predicted error is shown in Fig. 9. It can be seen that, as expected, a larger real emission angle reconstruction error led to greater uncertainty in the predicted emission angle.

4.2 Absorption point reconstruction

The reconstruction of the absorption points is important for improving the spatial resolution of a polarimeter. The absorption point accuracy can be evaluated using the half-power diameter (HPD), which is a commonly used parameter in X-ray imaging that is defined as the diameter of a circle that can cover exactly 50% of the reconstructed absorption point, taking the true absorption point as the center of the circle. Therefore, a larger HPD indicates a worse reconstruction of the photoelectric absorption point. Figure 10 compares the absorption point accuracies of the moment analysis and hexagonal CNN methods at different energies. The hexagonal CNN method was superior to the moment analysis method, particularly for highly complex photoelectron tracks at high energies.

4.3 Polarization estimation

The polarization reconstruction performance of a polarization estimation algorithm directly affects the sensitivity of the polarimeter. We analyzed the polarization reconstruction performance of the hexagonal CNN method and compared

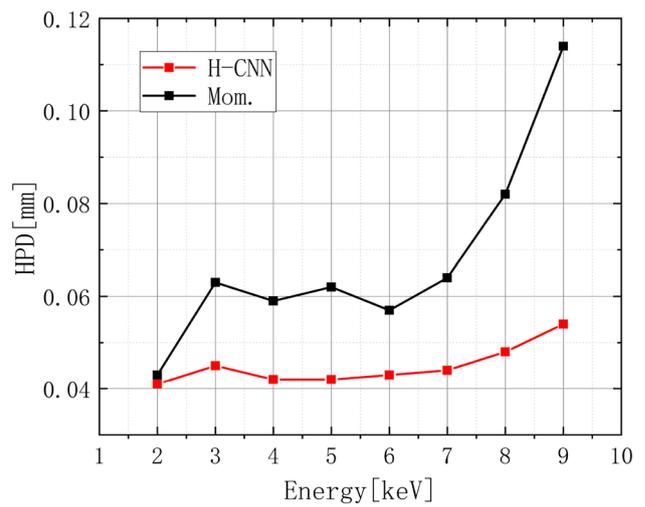


Fig. 10 Absorption point reconstruction with the moment analysis (black) and hexagonal CNN methods (red)

it with those of the moment analysis method and rectangle-based CNN method developed by the IXPE team.

Polarized and unpolarized simulation tracks were generated using PolarLight simulation algorithms for the polarization reconstruction analysis. Similar to the training data, the photoelectric tracks from incomplete energy deposition in the gas of the GPD or from interactions with the detector components outside the gas volume were removed.

The binned modulation curves created using the predicted emission angles for the unpolarized and 100% polarized simulated data are shown in Fig. 11. The residual systematic modulation curve of the hexagonal CNN method was as flat as that of the moment analysis method, indicating that the hexagonal CNN method did not introduce redundant systematic errors. In addition, the hexagonal CNN method recovered significantly more of the modulation of the polarized data compared with the moment analysis method. An unbinned polarization estimation algorithm based on the Stokes parameters was used to estimate the polarization fraction and EVPA from a set of predicted track angles. Figure 12 shows the recovered modulation response on the simulated PolarLight dataset for the moment analysis method, the rectangle-based CNN method developed by the IXPE team, and our hexagonal CNN method. It can be seen that our hexagonal CNN method performed better than the moment analysis method, with 15%–30% improvements in the modulation factor for individual energies.

Compared to the CNN method developed by the IXPE team based on classical rectangular convolution, our hexagonal CNN method had a similar performance in polarization reconstruction, although the hexagonal convolutional structure of the hexagonal CNN was better matched with hexagonal grid tracks. This may be because the

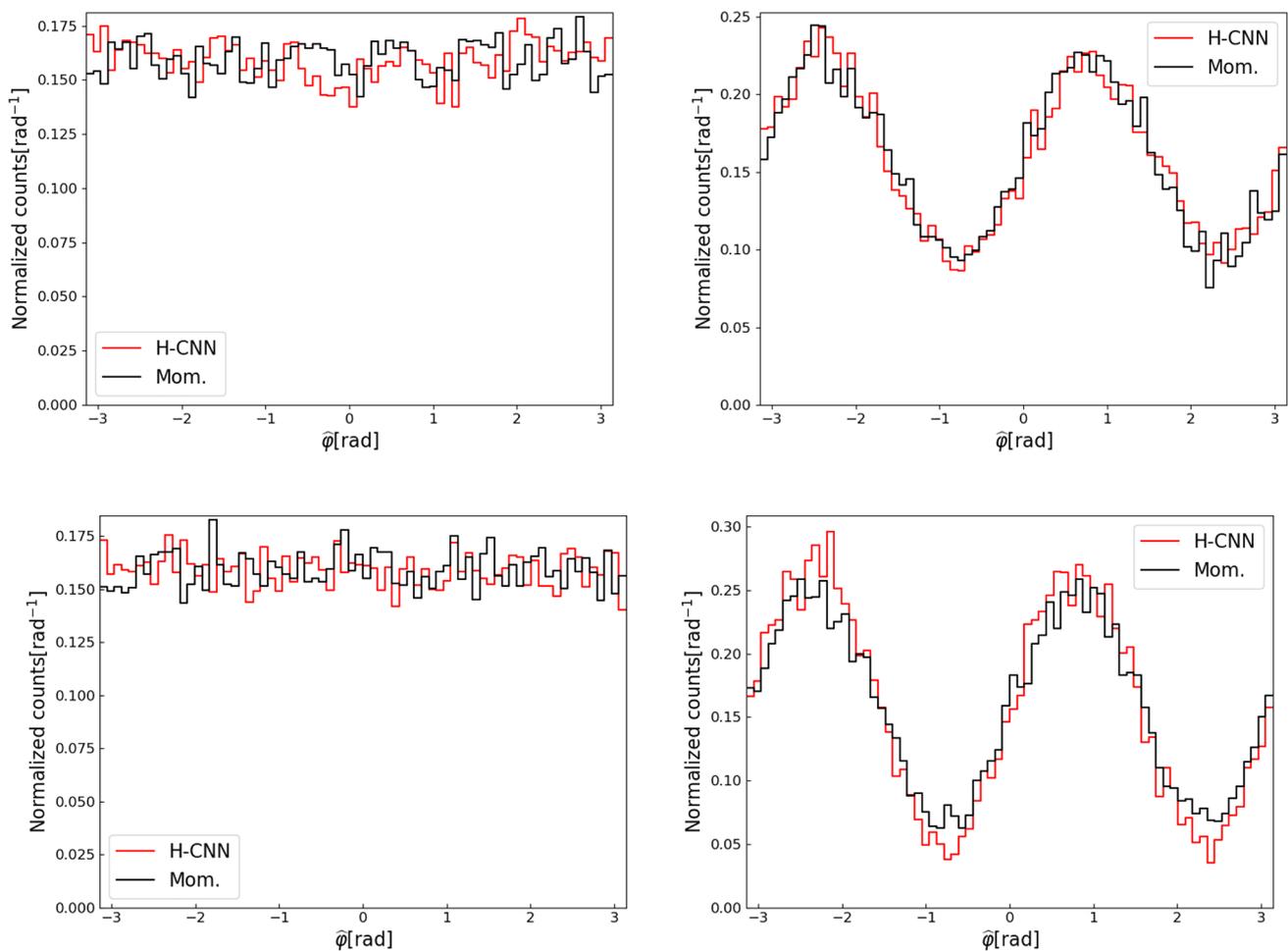


Fig. 11 Track angle reconstruction for unpolarized (left column) and polarized (right column) simulated data for 3 keV (top) and 7 keV (bottom) with moment analysis method (black) and hexagonal CNN method (red)

double-channel input track images of the rectangle-based CNN method compensated for the loss during the conversion from hexagonal images to square images, or because the existing neural network method is already close to the upper limit of polarization reconstruction owing to the blurring of the photoelectron tracks, which is difficult to improve with a better CNN architecture.

Our hexagonal CNN method takes a step toward a more straightforward implementation of hexagonal grid track processing. The hexagonal CNN method halves the amount of input data and has lower computational and storage costs during preprocessing because each track image is converted into three single-channel input images for prediction, whereas the rectangle-based CNN method developed by the IXPE team converts an image into three double-channel input images. However, the existing hexagonal convolution is mainly implemented based on rectangular convolution; therefore, the memory consumption

of the hexagonal CNN is high, which can be improved using native hexagonal CNN architectures.

5 Conclusion

We developed a track reconstruction and polarization estimation algorithm based on hexagonal CNNs to match the hexagonal grid tracks in a GPD for X-ray polarization measurements. The emission angles, absorption points, and uncertainties in the emission angles of X-ray photoelectron tracks were predicted using the hexagonal CNN method developed in this study. The predicted absorption points were used for image reconstruction, and the predicted emission angles and uncertainties were used to estimate the polarization of the X-ray source. We tested the proposed hexagonal CNN method using simulated PolarLight data. The results showed that the performance

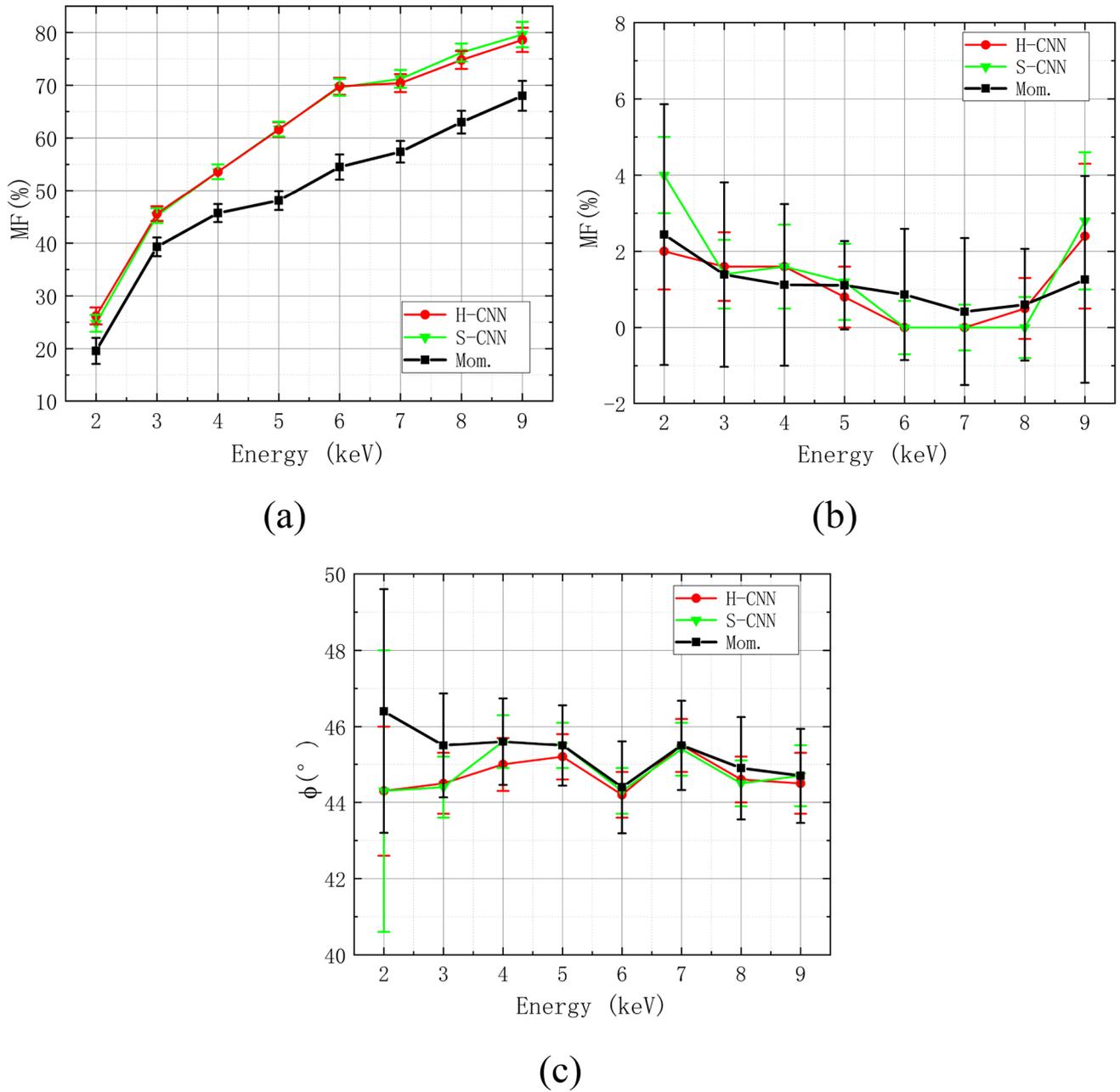


Fig. 12 Modulation responses of the moment analysis method (black), rectangle-based CNN method developed by the IXPE team (S-CNN, green), and our hexagonal CNN method (red). **a** Response

for 100% polarized data. **b** Response for unpolarized data. **(c)** Recovered EVPA for 100% polarized data

of the absorption point reconstruction in an HPD using the hexagonal CNN method was better than that of the moment analysis method, and the modulation factor of the hexagonal CNN method produced improvements of 15%–30% compared to the moment analysis method. The performance of our hexagonal CNN method is comparable to that of the CNN method developed by the IXPE team based on classical rectangular convolution, but it has lower computational and storage costs for preprocessing.

Our hexagonal CNN method also provides a good research basis for the development of polarization reconstruction algorithms for eXTP missions.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Ya-Nan Li, Jia-Huan Zhu and Huai-Zhong Gao. The first draft of the manuscript was written by Ya-Nan Li, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability The data that support the findings of this study are openly available in Science Data Bank at <https://cstr.cn/31253.11.sciencedb.j00186.00246> and <https://doi.org/10.57760/sciencedb.j00186.00246>.

Declarations

Conflict of interest Ming Zeng and Zhi Zeng are the editorial board members for Nuclear Science and Techniques and were not involved in the editorial review, or the decision to publish this article. All authors declare that there are no competing interests.

References

1. T. Kallman, Astrophysical motivation for X-ray polarimetry. *Adv. Space Res.* **34**, 2673–2677 (2004). <https://doi.org/10.1016/j.asr.2003.03.059>
2. P. Meszaros, R. Novick, A. Szentgyorgyi et al., Astrophysical implications and observational prospects of X-ray polarimetry. *Astrophys. J.* **324**, 1056 (1988). <https://doi.org/10.1086/165962>
3. R. Giacconi, H. Gursky, F.R. Paolini et al., Evidence for x-rays from sources outside the solar system. *Phys. Rev. Lett.* **9**, 439–443 (1962). <https://doi.org/10.1103/physrevlett.9.439>
4. M. Gavrila, Relativistic K-shell photoeffect. *Phys. Rev.* **113**, 514–526 (1959). <https://doi.org/10.1103/physrev.113.514>
5. E. Costa, P. Soffitta, R. Bellazzini et al., An efficient photoelectric X-ray polarimeter for the study of black holes and neutron stars. *Nature* **411**, 662–665 (2001). <https://doi.org/10.1038/35079508>
6. H. Feng, H. Li, X.Y. Long et al., Re-detection and a possible time variation of soft X-ray polarization from the crab. *Nat. Astron.* **4**, 511–516 (2020). <https://doi.org/10.1038/s41550-020-1088-1>
7. X.Y. Long, H. Feng, H. Li et al., A significant detection of X-ray polarization in Sco X-1 with polarlight and constraints on the corona geometry. *Astrophys. J. Lett.* **924**, L13 (2021). <https://doi.org/10.3847/2041-8213/ac4673>
8. H. Feng, H. Li, L.D. Kong et al., X-ray polarimetry of the accreting pulsar 1A 0535+262 in the supercritical state with polarlight. *Astrophys. J.* **950**, 76 (2023). <https://doi.org/10.3847/1538-4357/acd0af>
9. M.C. Weisskopf, B. Ramsey, S.L. O'Dell et al., The imaging X-ray polarimetry explorer (IXPE). *Results Phys.* **6**, 1179–1180 (2016). <https://doi.org/10.1016/j.rinp.2016.10.021>
10. S.N. Zhang, M. Feroci, A. Santangelo et al., eXTP: enhanced X-ray timing and polarimetry mission. *Proceedings Volume 9905, Space Telescopes and Instrumentation 2016: Ultraviolet to Gamma Ray; 99051Q* (2017). <https://doi.org/10.1117/12.2232034>
11. S.N. Zhang, A. Santangelo, M. Feroci et al., The enhanced X-ray timing and polarimetry mission—eXTP. *Sci. China Phys. Mech. Astron.* **62**, 29502 (2018). <https://doi.org/10.1007/s11433-018-9309-2>
12. H. Wang, D. Wang, R. Chen et al., Electronics system for the cosmic X-ray polarization detector. *Nucl. Sci. Tech.* **34**, 64 (2023). <https://doi.org/10.1007/s41365-023-01221-3>
13. Z.W. Fan, H.B. Liu, H.B. Feng et al., Front-end electronics of CXPD for measuring transient X-ray sources. *IEEE Trans. Nucl. Sci.* **70**, 1507–1513 (2023). <https://doi.org/10.1109/TNS.2023.3269091>
14. R. Bellazzini, F. Angelini, L. Baldini et al., Novel gaseous X-ray polarimeter: data analysis and simulation, in *Proc. SPIE 4843, Polarimetry in Astronomy* (2003). <https://doi.org/10.1117/12.459381>
15. T. Kitaguchi, K. Black, T. Enoto et al., An optimized photoelectron track reconstruction method for photoelectric X-ray polarimeters. *Nucl. Instrum. Meth. Phys. Res. Sect. A* **880**, 188–193 (2018). <https://doi.org/10.1016/j.nima.2017.10.070>
16. T.L. Li, M. Zeng, H. Feng et al., Electron track reconstruction and improved modulation for photoelectric X-ray polarimetry. *Nucl. Instrum. Meth. Phys. Res. Sect. A* **858**, 62–68 (2017). <https://doi.org/10.1016/j.nima.2017.03.050>
17. T. Kitaguchi, K. Black, T. Enoto et al., A convolutional neural network approach for reconstructing polarization information of photoelectric X-ray polarimeters. *Nucl. Instrum. Meth. Phys. Res. Sect. A* **942**, 162389 (2019). <https://doi.org/10.1016/j.nima.2019.162389>
18. A.L. Peirson, R.W. Romani, H.L. Marshall et al., Deep ensemble analysis for Imaging X-ray Polarimetry. *Nucl. Instrum. Meth. Phys. Res. Sect. A* **986**, 164740 (2021). <https://doi.org/10.1016/j.nima.2020.164740>
19. A.L. Peirson, Neural network analysis of X-ray polarimeter data, in *Handbook of X-ray and Gamma-ray Astrophysics*, ed. by C. Bambi, A. Santangelo (Springer, Singapore, 2024), pp.5781–5828. https://doi.org/10.1007/978-981-19-6960-7_144
20. C. Szegedy, W. Liu, Y.Q. Jia et al., Going deeper with convolutions, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA. IEEE, pp. 1–9 (2015)
21. Y.H. Chen, W. Li, C. Sakaridis et al., Domain adaptive faster R-CNN for object detection in the wild, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA. IEEE, 3339–3348 (2018)
22. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA. IEEE, pp. 3431–3440 (2015)
23. Y.D. Zeng, J. Wang, R. Zhao et al., Decomposition of fissile isotope antineutrino spectra using convolutional neural network. *Nucl. Sci. Tech.* **34**, 79 (2023). <https://doi.org/10.1007/s41365-023-01229-9>
24. F.D. Qin, H.Y. Luo, Z.Z. He et al., Counting of alpha particle tracks on imaging plate based on a convolutional neural network. *Nucl. Sci. Tech.* **34**, 37 (2023). <https://doi.org/10.1007/s41365-023-01190-7>
25. Y.F. Feng, Z.Z. Zhang, X.B. Zhao et al., GVCNN: group-view convolutional neural networks for 3D shape recognition, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA. IEEE, pp. 264–272 (2018)
26. M.M. Bronstein, J. Bruna, Y. LeCun et al., Geometric deep learning: going beyond euclidean data. *IEEE Signal Process. Mag.* **34**, 18–42 (2017). <https://doi.org/10.1109/MSP.2017.2693418>
27. J.T. Ke, H. Yang, H.Y. Zheng et al., Hexagon-based convolutional neural network for supply-demand forecasting of ride-sourcing services. *IEEE Trans. Intell. Transp. Syst.* **20**, 4160–4173 (2019). <https://doi.org/10.1109/TITS.2018.2882861>
28. C. Steppa, T.L. Holch, HexagDLy—processing hexagonally sampled data with CNNs in PyTorch. *SoftwareX* **9**, 193–198 (2019). <https://doi.org/10.1016/j.softx.2019.02.010>
29. J.R. Luo, W.P. Zhang, J.M. Su et al., Hexagonal convolutional neural networks for hexagonal grids. *IEEE Access* **7**, 142738–142749 (2019). <https://doi.org/10.1109/ACCESS.2019.2944766>
30. Y.X. Zhao, Q.H. Ke, F. Korn et al., He_xCNN: a framework for native hexagonal convolutional neural networks, in *2020 IEEE International Conference on Data Mining (ICDM)*. Sorrento, Italy. IEEE, 1424–1429 (2020)
31. M. Abdar, F. Pourpanah, S. Hussain et al., A review of uncertainty quantification in deep learning: techniques, applications and

- challenges. *Inf. Fusion* **76**, 243–297 (2021). <https://doi.org/10.1016/j.inffus.2021.05.008>
32. K. Zou, Z.H. Chen, X.D. Yuan et al., A review of uncertainty estimation and its application in medical imaging. *Meta-Radiol.* **1**, 100003 (2023). <https://doi.org/10.1016/j.metrad.2023.100003>
33. N. Meinert, J. Gawlikowski, A. Lavin, The unreasonable effectiveness of deep evidential regression. *Proc. AAAI Conf. Artif. Intell.* **37**, 9134–9142 (2023). <https://doi.org/10.1609/aaai.v37i8.26096>
34. R. Bellazzini, G. Spandre, M. Minuti et al., Direct reading of charge multipliers with a self-triggering CMOS analog chip with 105 k pixels at 50 μm pitch. *Nucl. Instrum. Meth. Phys. Res. Sect. A* **566**, 552–562 (2006). <https://doi.org/10.1016/j.nima.2006.07.036>
35. K.M. He, X.Y. Zhang, S.Q. Ren et al., Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA. IEEE, pp. 770–778 (2016)

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.