PHOTONICS Research

Deep coded exposure: end-to-end co-optimization of flutter shutter and deblurring processing for general motion blur removal

ZHIHONG ZHANG,[1] [iD] KAIMING DONG,[1] JINLI SUO,[1,2,3,*] AND QIONGHAI DAI[1,2] [iD]

[1]*Department of Automation, Tsinghua University, Beijing 100084, China*
[2]*Institute for Brain and Cognitive Sciences, Tsinghua University, Beijing 100084, China*
[3]*Shanghai Artificial Intelligence Laboratory, Shanghai 200030, China*
*Corresponding author: jlsuo@tsinghua.edu.cn

Coded exposure photography is a promising computational imaging technique capable of addressing motion blur much better than using a conventional camera, via tailoring invertible blur kernels. However, existing methods suffer from restrictive assumptions, complicated preprocessing, and inferior performance. To address these issues, we proposed an end-to-end framework to handle general motion blurs with a unified deep neural network, and optimize the shutter's encoding pattern together with the deblurring processing to achieve high-quality sharp images. The framework incorporates a learnable flutter shutter sequence to capture coded exposure snapshots and a learning-based deblurring network to restore the sharp images from the blurry inputs. By co-optimizing the encoding and the deblurring modules jointly, our approach avoids exhaustively searching for encoding sequences and achieves an optimal overall deblurring performance. Compared with existing coded exposure based motion deblurring methods, the proposed framework eliminates tedious preprocessing steps such as foreground segmentation and blur kernel estimation, and extends coded exposure deblurring to more general blind and nonuniform cases. Both simulation and real-data experiments demonstrate the superior performance and flexibility of the proposed method. © 2023 Chinese Laser Press

https://doi.org/10.1364/PRJ.489989

1. INTRODUCTION

Due to the limited frame rate of imaging devices and the inevitable instability during capturing, motion blur has become a common problem in daily photography. It not only degrades the photo's visual quality, but also imposes a great challenge on subsequent high-level vision tasks such as image classification [1], object detection [2], and object tracking [3]. To cope with this problem, various postprocessing deblurring algorithms have been designed by the computer vision (CV) community in the past decades [4–6]. On the other hand, researchers from the computational imaging (CI) field have also proposed many approaches to tackle this problem by jointly considering the imaging and postprocessing processes [7–13]. Coded exposure photography [14] is one of the most representative methods among these approaches, and has received much attention since being proposed [15–20].

A. Coded Exposure Photography

Different from conventional photography keeping the camera's shutter open throughout the entire exposure elapse, the coded exposure technique flutters the shutter open and closed according to the designed binary sequence in the exposure duration.

In this manner, the captured blurry images can better preserve high-frequency details, thus facilitating a subsequent deblurring process [14]. For simplicity, we take one-dimensional (1D) motion as an instance to explain the underlying mathematical principle. It is known that the spatially uniform blur can be formulated as the convolution between the sharp image and the blur kernel, which is determined by the motion trajectory and the exposure pattern (i.e., the flutter shutter). The difference between conventional and coded exposure is demonstrated in Fig. 1.

From the spatial perspective, under conventional exposure the resulting blur kernel is a continuous line, and the corresponding blurry image features continuous blurry edges accordingly. By contrast, the blur kernel under coded exposure is an intermittent line and produces edge fringes along the motion trajectories, which is a superimposition of a sequence of sharp snapshots.

From the frequency perspective, the blurring process can be regarded as a frequency sampling or filtering operation since convolution in the spatial domain is equivalent to multiplication in the frequency domain [21]. As can be seen from Fig. 1, the spectrum of the blur kernel resulting from conventional
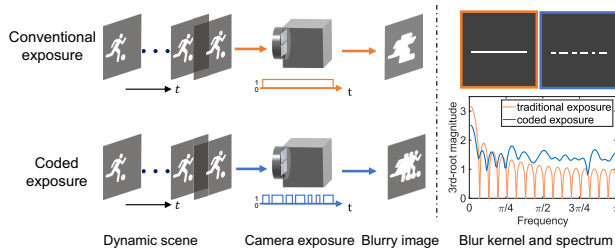
Fig. 1. Physical formation of blurring artifacts under conventional and coded exposure settings, and analysis in spatial and frequency domains.

exposure is a band-limited sinc function with periodic zeros and significant attenuation at higher frequencies, so the deblurring is strongly ill-posed. On the contrary, the spectrum of the blur kernel under coded exposure has no zeros and features a relatively flat magnitude across the whole spectrum. Therefore, by controlling the camera shutter with specially designed binary sequences, coded exposure photography can better preserve information of different frequencies and facilitate inverting the blur artifacts to obtain sharp images.

B. Coded Exposure Based Image Deblurring
Based on the fundamental principle of coded exposure deblurring, many works seeking to explore efficient exposure patterns (i.e., the coding sequence) have popped up. In the original paper of coded exposure photography [14], Raskar *et al.* performed an in-depth analysis of the blur kernels' invertibility and proposed to select coding sequences via maximizing the minimal value of the spectrum's magnitude and minimizing the variance of the spectrum values. Since one needs to estimate the blur kernel before deblurring process, Agrawal and Xu further took the blur kernel estimation process into consideration [15]. They found that the kernel of a smooth blur is easier to estimate via alpha matting and thus proposed to maximize the number of continuous ones and minimize the 0–1 transitions in the encoding sequence. Later, McCloskey *et al.* incorporated natural image statistics into encoding sequence design and achieved a significant reduction in reconstruction error [22]. Differently, Jeon *et al.* introduced the concept of low-autocorrelation binary sequences and a new measure for good shutter sequence [23]. Apart from the single-image deblurring methods, there are also some other works investigating deblurring from multiple coded exposure snapshots and corresponding code design [19,20].

Although various criteria for encoding sequence design have been proposed, and corresponding kernel estimation and deblurring algorithms are both improved, there still exist significant drawbacks for existing coded exposure deblurring methods. On the one hand, previous encoding sequence design is mainly based on handcrafted criteria and relies on random searching to find the optimal code. However, handcrafted criteria can hardly make extensive use of the natural image prior, which is important in image restoration. Besides, the random searching strategy is usually inapplicable in long sequence design due to its exponential expanding of search space. On the other hand, most coded exposure deblurring algorithms are still

limited to handling uniform (i.e., spatially invariant) blur and rely on tedious pipelines involving foreground segmentation, blur kernel estimation, deblurring, compositing, etc., to finish the deblurring task. These issues greatly limit their applications in motion-blur-free photography.

C. Deep Learning Based Image Deblurring
In recent years, deep learning has been popularized to cope with various CV problems and achieves significant performance promotion compared with conventional algorithms [24]. Benefiting from the powerful representation ability of deep neural networks (DNNs), novel learning-based deblurring methods have been developed and successfully applied to handling the tough spatially varying/nonuniform blur [6]. Besides, these methods also eliminate the preliminary step of blur kernel estimation involved in traditional deblurring algorithms and operate in an efficient end-to-end manner, which is called blind deblurring. Instead of formulating the blurring model as the convolution between a sharp image and the blur kernel, recent learning-based deblurring algorithms directly average consecutive video frames to synthesize the blurry image so as to simulate the more general cases. Nah *et al.* collected a widely used high-frame-rate video dataset called GoPro for blurry image simulation and proposed a deep multi-scale convolution neural network (CNN) to restore the latent sharp image [25]. Since then, the multi-scale "coarse-to-fine" scheme has become a widely used architecture for deep learning based deblurring [26–30]. Tao *et al.* further introduced the recurrent strategy into the multi-scale architecture and proposed the scale-recurrent network for image deblurring [26]. Differently, Zamir *et al.* [27] proposed a multi-stage progressive network and achieved excellent performance on various image restoration tasks including deblurring. Most recently, Cho *et al.* developed a multi-input multi-output U-net (MIMO-UNet) for single-image deblurring by revisiting the coarse-to-fine strategy [28], and Mao *et al.* extended their work by introducing a novel Fourier transformation based residual module [29]. Both methods achieved state-of-the-art (SOTA) performance in blind nonuniform single-image deblurring.

Although DNN-based deblurring algorithms have surpassed conventional optimization, which is of inferior performance, tedious pipelines, and significant limitations in practical applications, they have also encountered bottlenecks. On the one hand, the high frequencies lost in the blurry images or the intrinsic ill-posedness of the deblurring problem determine the performance upper-bound of learning-based algorithms. Even though some generative networks such as the variational autoencoder (VAE) [31] and generative adversarial networks (GANs) [32] could produce plausible results by imposing strong priors of nature scenes, they cannot reconstruct perfect results in image restoration problems. On the other hand, after years of progress in network design and hyper-parameter fine-tuning, the performance advances of deblurring neural networks have slowed down, but the practical applicability is still largely limited, especially in scenarios with complex motion or realistic noise. Fortunately, recent advances in high-level CV tasks such as semantic information retrieval have demonstrated that combining both CI and CV [33–35] to draw on each other's strengths can gain an overall performance promotion.

Inspired by this new trend, we revisit coded exposure deblurring and combine it with recent advances in deep learning to unlock insights to boost the deblurring performance and broaden its applicability.

D. Contributions of this Paper

Here we propose a novel single-image motion deblurring framework incorporating a coded exposure based imager and a learning-based deblurring network jointly. We co-optimize the imager's exposure pattern and the parameters of the deblurring network to achieve an optimal overall performance while avoiding the high-complexity exhaustive search of the encoding sequence and complex preprocessing steps. Besides, with the aid of deep learning's powerful representation ability, we loosen the strict assumptions of previous coded exposure based deblurring methods and can handle general motion blur blindly. To the best of our knowledge, this is the first work applying coded exposure photography in learning-based end-to-end blind nonuniform deblurring.

The contributions of this work can be summarized as follows.

• We successfully extend coded exposure deblurring to blind and nonuniform scenarios leveraging the recent advances in deep learning.
• We propose a novel data-driven encoding sequence design method by co-optimizing the optical encoder and the deblurring network.
• We build a coded exposure imaging prototype and demonstrate the high performance of the proposed method on both real and simulated data.
• The proposed approach is easy to use, widely applicable, and of advantageous performance, which pushes forward the applications of coded exposure photography.

2. ENCODER–DECODER CO-OPTIMIZATION FRAMEWORK

The overall flowchart of the proposed coded exposure blur encoding and learning-based deblurring co-optimization framework is demonstrated in Fig. 2. In the training stage, we model the physical imaging process as an optical blur encoder in accordance with the fundamental principle of coded exposure photography. Then, a CNN-based blur decoder is employed to estimate the latent sharp image from its coded blurry counterpart generated from the optical blur encoder. It is worth noting that both the optical blur encoder and the CNN-based blur decoder are optimized jointly during the training period. In this manner, the encoding sequence of the coded exposure and the parameters of the deblurring CNN will be updated simultaneously, allowing us to find a solution superior to the separate optimization. In the inference phase, the optical blur encoder is replaced with the real acquisition process, and the optimized encoding sequence is loaded to a camera as the shutter trigger signal accordingly. The trained CNN model will also be saved to perform the deblurring task for the real-captured coded blurry images.

In the following subsections, we will give a detailed description about the design and implementation of each module in the framework.

A. Learnable Optical Blur Encoder

To describe general nonuniform blurry images efficiently, we discard the convolution-based blur simulation method widely used in previous works. Instead, recalling that the digital camera records a blurry image by continuously accumulating light signals from a dynamic scene on the sensor during the exposure period, we thus regard the blurry snapshot as a summation of sharp images describing the scene in a continuous sequence of
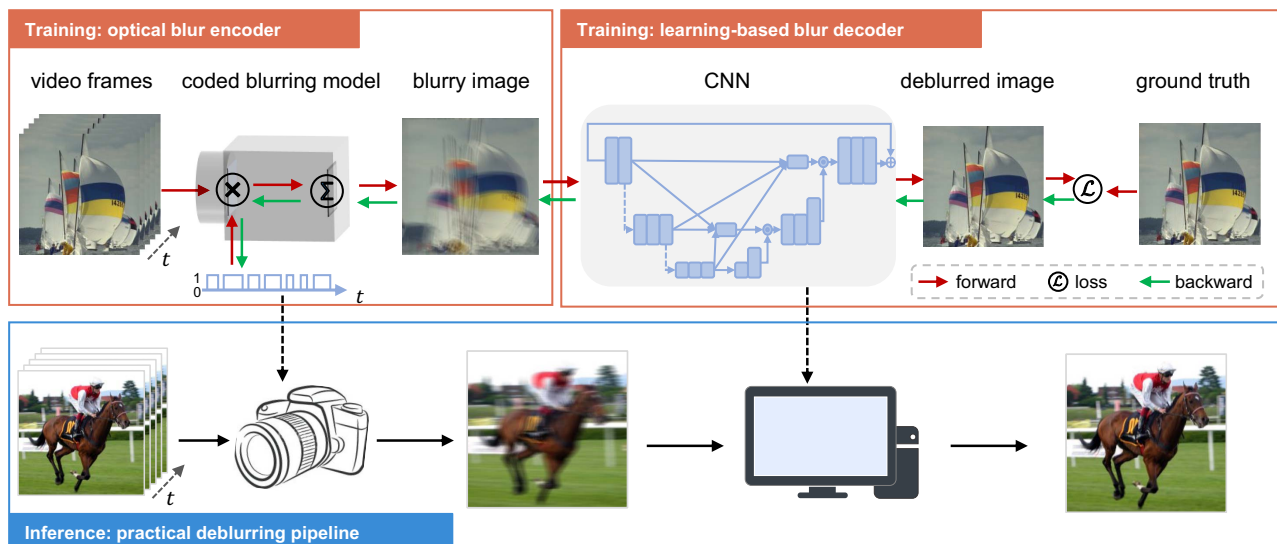

Fig. 2. Overall flowchart of the proposed framework. The coded exposure imaging system and the learning-based deblurring algorithm are respectively modeled with an optical blur encoder and a computational blur decoder, and together form an end-to-end differentiable forward model. In the training stage, the parameters of the whole model are optimized together through gradient descent until convergence. In the inference stage, the learned encoding sequence will be loaded to the controller of the camera shutter (or its equivalent), and the computational blur decoder will be employed to deblur the captured coded blurry images.

sufficiently short time slots. When taking the coded exposure into consideration, some of the exposure segments corresponding to the "close-shutter" state will be blocked out during the sensor integration. Therefore, the exposure encoding sequence actually serves as binary weights for these sharp images during the integration. Denoting the exposure duration as T, the scene's radiation at t as $\mathbf{S}(t)$, and the shutter's trigger signal as $\mathbf{e}(t)$, the coded blurry image $\mathbf{B}$ can be mathematically formulated as $\int_{t=0}^{T} \mathbf{S}(t) \cdot \mathbf{e}(t)\mathrm{d}t$, which can be further discretized into

$$\mathbf{B} = \sum_{i=1}^{M} \mathbf{S}[i] \cdot \mathbf{e}[i]. \tag{1}$$

Here M is the length of the encoding sequence and also the number of the short time intervals, $\mathbf{S}[i]$ and $\mathbf{e}[i]$ represent the ith short-exposure sharp image and corresponding binary encoding code, respectively. Note that we omit the camera's response function and postprocessing steps such as gamma transformation, which can be calibrated and compensated beforehand in real applications.

The selection of the encoding sequence is intrinsically a binary optimization problem, which is difficult for both conventional optimization and deep learning. Fortunately, there are mature solutions in deep learning to enable back-propagation training of binary parameters, and we employ a widely used technique called "straight-through-estimator" (STE) [36,37] in our work. Specifically, in the optical blur encoder, instead of directly defining a binary encoding sequence, we employ the reparameterization trick [31] by introducing a learnable parameter vector $b \in \mathbb{R}^M$ and deriving the desired binary sequence e via a sign function

$$\mathrm{e} = \frac{1}{2}(\mathrm{sign}(\mathrm{b}) + 1) \tag{2}$$

with

$$\mathrm{sign}(x) = \begin{cases} +1, & x > 0 \\ 0, & x = 0 . \\ -1, & x < 0 \end{cases} \tag{3}$$

Considering that the sign function is not differentiable at zero and its derivative vanishes (i.e., equals 0) at other points, the STE technique introduces a clip function to serve as the derivative of the sign function instead, which enables the back-propagation through gradient descent during network training. The clip function is formulated as follows:

$$\mathrm{clip}(x, -1, 1) = \max(-1, \min(1, x)). \tag{4}$$

B. Learning-Based Blur Decoder

With the flourishing development of deep learning, a number of deep deblurring neural networks have been proposed and achieved superior performance to traditional optimization-based methods [5,6]. In the proposed co-optimization deblurring framework, we employ a SOTA deblurring CNN called DeepRFT [29] to estimate the latent sharp image from its coded blurry counterpart. As an end-to-end blind deblurring method, DeepRFT eliminates the tedious blur kernel

estimation and pre-/post-processing steps required in conventional coded exposure deblurring methods.

The basic architecture of DeepRFT is shown in Fig. 3. It employs an MIMO-UNet [28] empowered by several specially designed feature extraction and fusion modules. Like many other deblurring networks, DeepRFT adopts the multi-scale strategy to facilitate deblurring by aggregating information from various spatial scales. Specifically, it first down-samples the blurry images to generate another two blurry images that are half and a quarter of the original spatial resolution, respectively. Then these three blurry images are sequentially input to the network and deblurred at different stages. Accordingly, multi-scale losses are employed during training to measure the distance between the outputs of different spatial scales and their respective ground truth. DeepRFT further replaces the vanilla convolution layer with the depth-wise over-parameterized convolutional layer (DO-Conv) [38] to achieve additional performance gains. DO-Conv is realized by enhancing the conventional convolution layer with an additional depth-wise convolution that convolves each input channel with a different two-dimensional (2D) convolution kernel, and has demonstrated superior performance in many vision tasks. The upper-most contribution of DeepRFT lies in proposing a novel Res-FFT-Conv Block, which augments the canonical ResBlock [39] with an extra frequency-domain convolution branch. The branch is implemented with 2D fast Fourier transformation (FFT), and helps to provide supplementary information from the frequency domain. In brief, the designed Res-FFT-Conv Block can effectively model the frequency discrepancies between the blurry and sharp image pairs, and can also capture both the long-term and short-term interactions to facilitate the deblurring process.

It is worth noting that although we choose DeepRFT as the coded blur decoder in the current implementation of the proposed framework thanks to its superior performance and low computation complexity, it can be flexibly switched to other learning-based deblurring networks to keep up with the latest advances in the CV deblurring field.

C. Loss Function

The loss function for model training consists of the following three terms.

Multi-scale Charbonnier loss [27] penalizes the deviation of the estimated sharp image from its ground-truth version at different spatial scales:
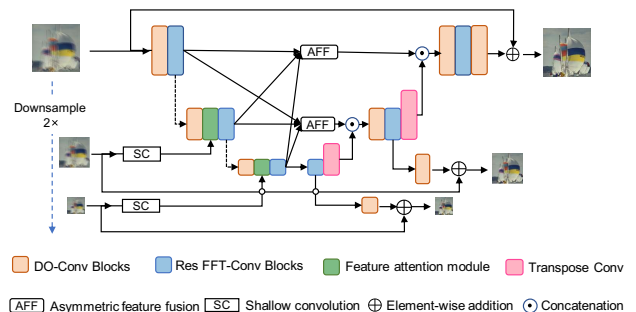

Fig. 3. Architecture of the deblurring neural network DeepRFT [29] in the proposed framework.

$$L_1 = \sum_{n=1}^{N} \frac{1}{p_n} \sqrt{\|\hat{\mathbf{I}}_n - \mathbf{I}_n\|^2 + \varepsilon^2}, \tag{5}$$

where N is the total number of spatial scales adopted by the multi-scale deblurring strategy of DeepRFT; p_n, $\hat{\mathbf{I}}_n$, and $\mathbf{I}_n$ represent the number of pixels, the estimated sharp image, and the corresponding ground-truth image at scale n, respectively; ε is a small constant to guarantee differentiability and is empirically set to be 10^{-3} (the same as below).

Multi-scale edge loss is defined as [27]

$$L_2 = \sum_{n=1}^{N} \frac{1}{p_n} \sqrt{\|\Delta\hat{\mathbf{I}}_n - \Delta\mathbf{I}_n\|^2 + \varepsilon^2}, \tag{6}$$

with Δ representing the Laplacian operator, which forces high consistency between the edges of the recovered sharp image and those of the ground truth.

Multi-scale frequency reconstruction loss [28] is also introduced to guide the prediction towards the latent sharp image but defined in the Fourier domain

$$L_3 = \sum_{n=1}^{N} \frac{1}{p_n} \|\mathcal{F}(\hat{\mathbf{I}}_n) - \mathcal{F}(\mathbf{I}_n)\|_1, \tag{7}$$

where $\mathcal{F}$ denotes the FFT.

The final loss function is defined as a weighted summation $L = L_1 + \gamma_1 L_2 + \gamma_2 L_3$, where the weighting coefficients γ_1 and γ_2 are empirically set to 0.05 and 0.01, respectively.

D. Prototype Building

Coded exposure imaging requires the camera to flutter its shutter according to the designed binary encoding sequence during the exposure period. Although not commonly used on commercial cameras, this feature can be implemented with photoelectric devices. One can either introduce an extra external shutter synchronized by a micro-controller [14] or directly employ the cameras supporting IEEE DCAM Trigger Mode 5 [15,19,22] to customize the exposure sequence. For simplicity and high compatibility with most commercial cameras, we
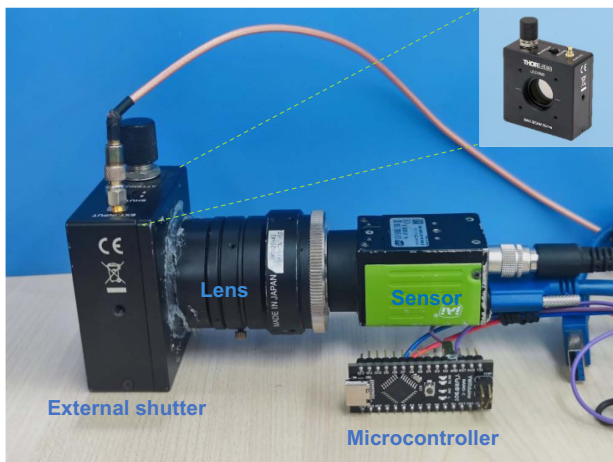

Fig. 4. Prototype system for coded exposure photography. It employs a liquid crystal element to serve as an external shutter for exposure encoding.

adopt the external shutter scheme to validate the proposed approach.

The prototype of our coded exposure imaging system is shown in Fig. 4. Apart from the conventional RGB camera (JAI GO-5000C-USB), it also consists of a liquid crystal optical shutter (Thorlabs LCC1620), a camera lens (KOWA, 12.5 mm/F1.4), and a micro-controller (Arduino Nano). The liquid crystal optical shutter is composed of a liquid crystal cell sandwiched between a pair of orthogonal polarizers. It has an average transmittance above 60% in the open state over the visible-light wavelength range and a contrast ratio (defined as the ratio of the transmittance in the open state to the transmittance in the close state) exceeding 8000:1. During acquisition, the micro-controller produces optimized binary voltage signals to control the shutter's open/close state by changing the liquid crystal's molecule orientation. Meanwhile, the micro-controller also functions to synchronize the camera with the shutter.

3. EXPERIMENTS

A. Implementation Details

In the following experiments, we employ the widely used high-frame-rate video dataset GoPro [25] to simulate coded blurry snapshots, train our network, and evaluate the proposed framework's performance. GoPro dataset is acquired using a GOPRO4 Hero Black camera at 240 frames per second (FPS). It contains approximately 35,000 sharp images in total, about two thirds of which are used for training and the rest for testing. Unless otherwise stated, the length of the exposure encoding sequence is set to 32 in the experiments; i.e., 32 sharp images from the dataset would be weighted by the binary encoding sequence and then collapsed to a single coded blurry image by the optical blur encoder. Besides, we also normalize the coded blurry images to [0,1] and add Gaussian noise with a standard deviation ranging from 0 to 0.02 to mimic the physical imaging process.

We implement the framework with PyTorch [40] and conduct the experiments on a workstation equipped with an AMD Ryzen Threadripper 3970X CPU and an NVIDIA GeForce RTX 3090 GPU. In the training phase, we adopt the Adam optimizer [41] to update the parameters and initialize the learning rate to 2×10^{-4}. The learning rate is steadily decayed to 1×10^{-6} using the cosine annealing strategy [42] after two rounds of warmup. For each training iteration, we randomly crop the video frames to 256×256 pixels and apply flip operations as the data augmentation trick to increase the dataset's diversity. During testing, the slicing window crop method with a stride of 256 pixels is employed to create 256×256 image patches. Patch-wise deblurring and patch merging are then performed to restore the latent sharp image. It should be noted that the patch-wise processing is not a necessary requirement, but rather a testing trick. Previous studies have shown that it can enhance performance slightly by ensuring constant patch size during training and testing [43].

It is worth noting that the STE technique might introduce instability during training, and thus we employ a multi-step training strategy to mitigate this issue. Specifically, the learning-based blur decoder (DeepRFT) is first pretrained for

150 epochs with the parameters of the optical blur encoder fixed, and then the whole framework is trained for another 450 epochs until converged.

B. Performance Evaluation and Analysis

To quantitatively evaluate the performance of the proposed co-optimization framework, we first compare them with the conventional noncoded deblurring method and SOTA coded exposure deblurring methods [14,15,23,44]. As mentioned above, most existing approaches focus on the design of exposure encoding sequences while employing traditional optimization-based deblurring algorithms and assuming uniform blur kernels. However, in this work, we aim for the more general blind and nonuniform deblurring problem. Therefore, for a fair comparison, we only change the encoding sequence in the optical encoder according to the competing methods, and adopt the same deblurring network architecture in the blur decoder to conduct blind nonuniform deblurring. Besides, for different encoding sequences, the deblurring network is separately trained from scratch on the same dataset.

The performance of different exposure codes is compared in terms of the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) on the GoPro dataset. The results are listed in Table 1, arranged by the time being proposed. Two deblurring experiments conducted on a spatially invariant blurry image (the "cars") and a spatially varying blurry image (the "flowers") are shown in Fig. 5 for visual comparison as well. It can be observed that the proposed co-optimization deblurring framework demonstrates a significant improvement over all competing methods, and its PSNR and SSIM gains compared with the second-best competitor by Cui *et al.* [44] are 1.44 dB and 0.0338, respectively. Qualitatively, as shown in Fig. 5, while all the methods demonstrate obvious blur-removing effects with the aid of the deep blur decoder's powerful image restoration ability, the proposed method provides more clear background and sharper structures across all parts of the restored image. On the contrary, the competitors suffer from ringing artifacts and fail to recover the sharp details in some regions. Overall, the performance promotion of the proposed framework in both the quantitative indices and qualitative visualization proves the effectiveness of the designed co-optimization framework and embodies the advantages of data-driven feature learning compared with hand-crafted criteria in encoding sequence design.

Table 1. Deblurring Performance with Different Encoding Sequences[a]

Methods	Noncoded	Raskar *et al.* [14]	Agrawal and Xu [15]	Jeon *et al.* [23]	Cui *et al.* [44]	Ours
Sequence (Hex)	FFFFFFFF	F1CD448D	7FFC2747	16A3809B	8076A061	**11CFF48C**
PSNR (dB)/SSIM	24.56/0.7695	24.37/0.7638	24.60/0.7695	25.47/0.8035	26.66/0.8289	**28.10/0.8627**

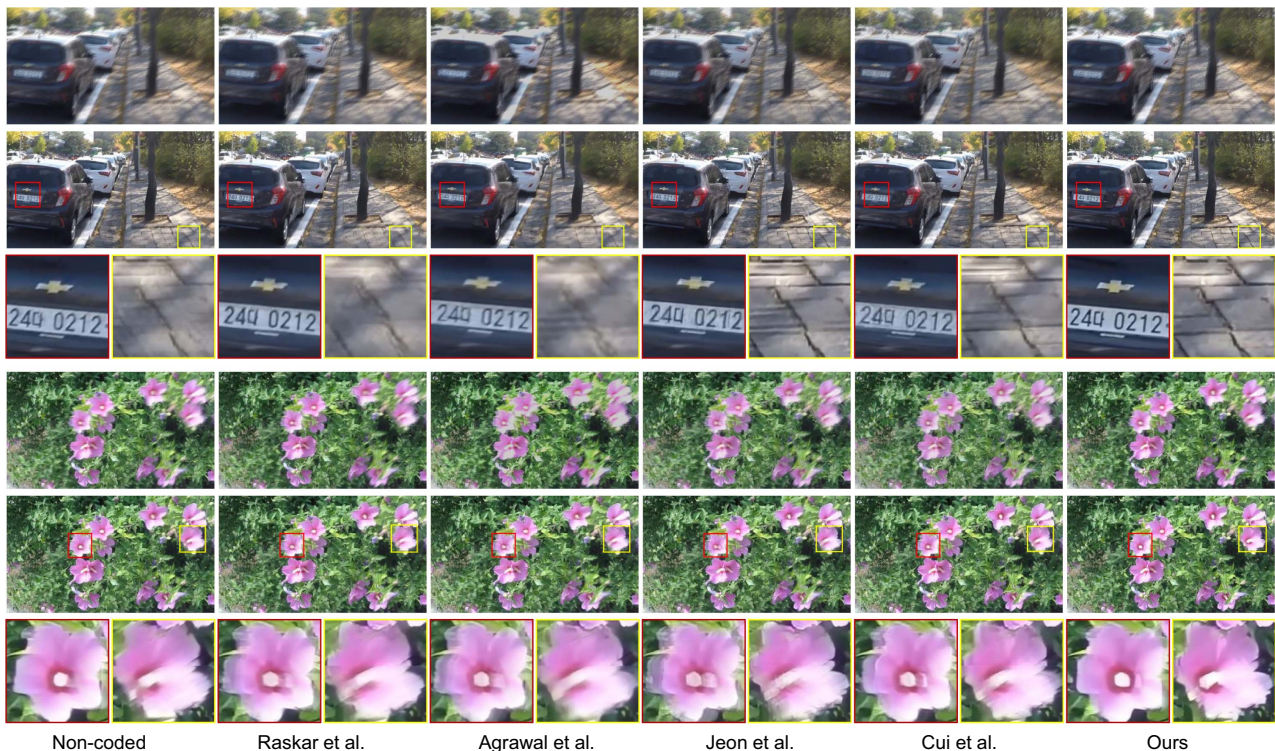[a]The encoding sequences are written in hexadecimal for simplicity.

Fig. 5. Synthesized blurry images under different exposure encoding settings and corresponding deblurring results. (Please zoom in for a better view.)

From Table 1, we can also observe that nearly all of the coded exposure deblurring methods achieve superior performance to the noncoded deblurring method, and their performance features a monotone increasing tendency, which validates the higher effectiveness of novel encoding sequence design criteria than the elder ones. Note that, different from other coded exposure deblurring methods, Raskar *et al.* [14] unexpectedly achieved inferior performance to the noncoded deblurring method. This is probably due to the trade-off between the encoding efficiency and signal-to-noise ratio (SNR) in coded exposure deblurring. To be specific, although coded exposure imaging can facilitate the preservation of image information at all frequencies, it will also sacrifice some light and cause a lower SNR hindering the deblurring process. Therefore, not all coded exposure deblurring methods but the ones with highly efficient encoding sequences could achieve better performance than the conventional noncoded deblurring method in practical applications. We further plot the frequency spectra of different encoding sequences in Fig. 6 to provide an intuitive visualization of their common properties. As can be seen from the figure, all the spectra of the coded exposure sequences share some common features including no zero points and relatively flat amplitude, which conform well with the fundamental theoretical analysis of coded exposure deblurring.

Next, we conduct another ablation experiment to study the influence of the sequence length on our deblurring performance. Learnable encoding sequences of 8, 16, 32, and 64 bits are tested in this experiment. Different from the previous evaluations, we use 64 consecutive sharp frames from the GoPro dataset to serve as the blur encoder's original input. Note that we keep the number of input frames constant throughout the experiment to simulate different coded blurry images taken under the same exposure time. To match the length of the encoding sequences with the number of input frames, we upsample the 8-bit, 16-bit, and 32-bit encoding sequences to 64-bit beforehand.

We report our experimental results in Fig. 7, from which one can observe that the deblurring performance of our framework increases monotonically with the increase of the sequence length. In other words, for a specific exposure time, a longer
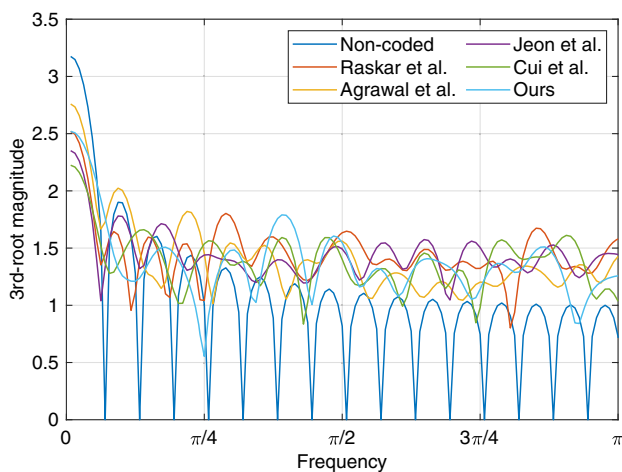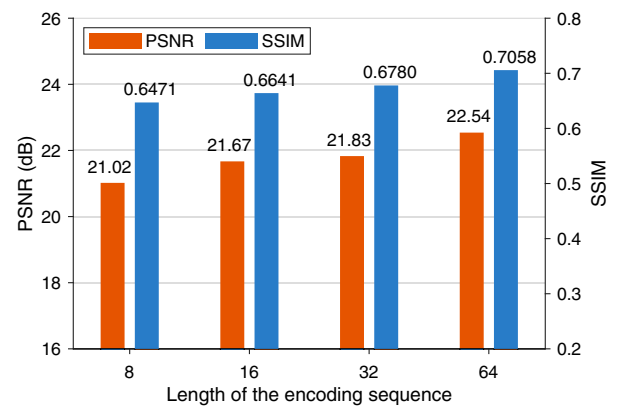

Fig. 7. Influence of the encoding sequence's length on the deblurring performance of the proposed framework.

encoding sequence will result in better deblurring performance. From the imaging perspective, given a certain duration of the total exposure, a longer encoding sequence corresponds to a shorter exposure elapse for each bit of the sequence, and the corresponding images will be sharper accordingly. From the optimization perspective, having more bits in the encoding sequences will result in a larger searching space, which facilitates finding a better solution. It is worth noting that, in practice, there are still some limitations in increasing the length of encoding sequences. On the one hand, in network training, a longer encoding sequence means more sharp frames are required to synthesize a coded blurry image. The huge data will impose much pressure on the memory consumption, making it the bottleneck for efficient training. On the other hand, in hardware implementation, the length of the encoding sequence in a single exposure will be restrained by the refresh rate of the external programmable shutter.

C. Qualitative Demonstration on Real Data

In order to validate the effectiveness of the proposed framework in real scenarios, we also use the built prototype system to capture encoded blurry snapshots of highly dynamic scenes and reconstruct their sharp versions computationally. In the experiment, we build an exemplar scene that contains a horizontally moving car and a swinging flower to generate spatially varying 1D and 2D motion blurs. The system is placed approximately 50 cm in front of the scene. During acquisition, the exposure time of the camera is set to 0.8 s, and thus each bit in the 32-bit encoding sequences corresponds to 25 ms. The captured blurry images are transferred from the sensor to the workstation through a universal serial bus (USB). Afterward, these images are deblurred with the corresponding pretrained deblurring models. Averagely, it takes about 0.52 s to handle one blurry image of 1280 × 1280 pixels with DeepRFT on our workstation.

Figure 8 shows the captured blurry images and corresponding deblurring results of our approach and other existing coded exposure photography methods. One can observe that the recovered sharp image of our co-optimization deblurring framework features sharper structures and clear background across different regions. By contrast, the noncoded deblurring method
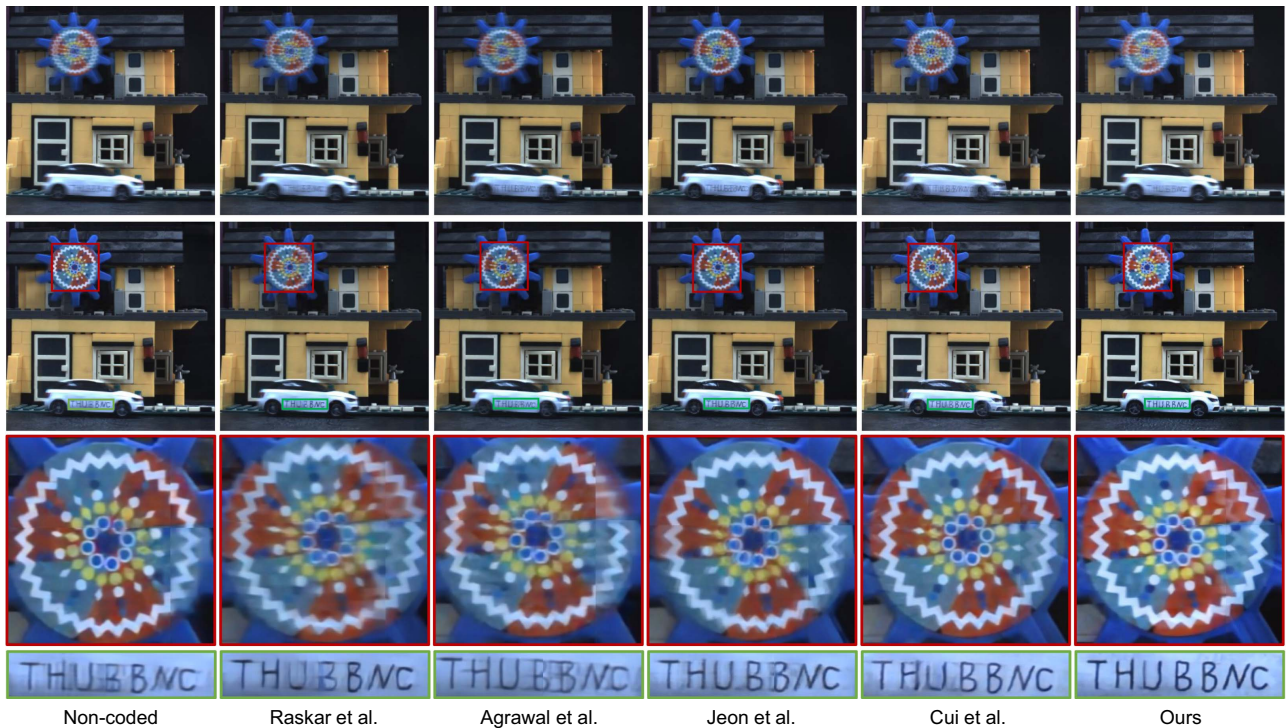

Fig. 6. Frequency spectra of different encoding sequences.

| Non-coded | Raskar et al. | Agrawal et al. | Jeon et al. | Cui et al. | Ours |

Fig. 8. Real-captured blurry images under exposure with different encoding sequences and corresponding deblurring results. (Please zoom in for a better view.)

and existing coded deblurring methods suffer from more artifacts and lower definition.

4. CONCLUSION AND DISCUSSION

In summary, we revisit the coded exposure technique and propose a novel co-optimization deblurring framework for simultaneous exposure pattern design and blind nonuniform image deblurring. To the best of our knowledge, this is the first work investigating the application of coded exposure photography in learning-based end-to-end blind nonuniform deblurring. By integrally modeling the whole process of blurry image formation and sharp image estimation in a differentiable manner, the proposed framework empowered by deep learning achieves superior performance to the separate optimization. Besides, compared with the previous exposure designs based on hand-crafted criteria and random searching, the proposed framework takes natural image prior into consideration via data-driven network training and solves the nondifferentiable issue in binary sequence optimization. Both the simulation experiments on standard datasets and the real-data experiments on our prototype have validated the effectiveness of the proposed approach.

Note that the objective of this work is to design a general deblurring framework combining CI and CV to bridge the gap between the coded exposure technique and recent advances in learning-based deblurring algorithms, rather than developing a specific SOTA network architecture for the image deblurring task. Therefore, the framework can flexibly incorporate other learning-based deblurring networks to keep up with the latest advances in the CV deblurring field. Besides, recent

development in novel binary neural networks could also provide possible variants for our network. We leave the extensions to future investigations.

In the future, this work can be further extended in the following two directions. From the algorithm perspective, the network architectures of the binary-modulation blur encoder and the coded blur decoder in the proposed framework could be investigated and improved to achieve higher deblurring quality and faster inference speed. From the application perspective, apart from the coded motion blur, other types of blur resulting from out-of-focus or lens aberration could also be involved in designing a more comprehensive blur encoder, making it possible for the framework to cope with more complex blur artifacts in an end-to-end manner. Furthermore, a more realistic noise model could also be employed in training data simulation to raise the deblurring network's robustness in low-light conditions.

REFERENCES

1. Y. Pei, Y. Huang, Q. Zou, X. Zhang, and S. Wang, "Effects of image degradation and degradation removal to CNN-based image classification," IEEE Trans. Pattern Anal. Mach. Intell. **43**, 1239–1253 (2021).

2. S. Zheng, Y. Wu, S. Jiang, C. Lu, and G. Gupta, "Deblur-YOLO: real-time object detection with efficient blind motion deblurring," in *International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2021), pp. 1–8.

3. Q. Guo, W. Feng, R. Gao, Y. Liu, and S. Wang, "Exploring the effects of blur and deblurring to visual object tracking," IEEE Trans. Image Process. **30**, 1812–1824 (2021).

4. R. Wang and D. Tao, "Recent progress in image deblurring," arXiv, arXiv:1409.6838 (2014).

5. J. Koh, J. Lee, and S. Yoon, "Single-image deblurring with neural networks: a comparative survey," Comput. Vision Image Understanding **203**, 103134 (2021).

6. K. Zhang, W. Ren, W. Luo, W.-S. Lai, B. Stenger, M.-H. Yang, and H. Li, "Deep image deblurring: a survey," Int. J. Comput. Vis. **130**, 2103–2130 (2022).

7. S. K. Nayar and M. Ben-Ezra, "Motion-based motion deblurring," IEEE Trans. Pattern Anal. Mach. Intell. **26**, 689–698 (2004).

8. A. Levin, P. Sand, T. S. Cho, F. Durand, and W. T. Freeman, "Motion-invariant photography," ACM Trans. Graph. **27**, 1–9 (2008).

9. S. McCloskey, "Temporally coded flash illumination for motion deblurring," in *International Conference on Computer Vision (ICCV)* (IEEE, 2011), pp. 683–690.

10. C. Ma, Z. Liu, L. Tian, Q. Dai, and L. Waller, "Motion deblurring with temporally coded illumination in an LED array microscope," Opt. Lett. **40**, 2281–2284 (2015).

11. S. Elmalem, R. Giryes, and E. Marom, "Motion deblurring using spatiotemporal phase aperture coding," Optica **7**, 1332–1340 (2020).

12. J. Lee and B. Jeon, "Multi-channel image deblurring using coded flashes," Proc. SPIE **11766**, 117660C (2021).

13. C. M. Nguyen, J. N. P. Martel, and G. Wetzstein, "Learning spatially varying pixel exposures for motion deblurring," in *IEEE International Conference on Computational Photography (ICCP)* (IEEE, 2022), pp. 1–11.

14. R. Raskar, A. Agrawal, and J. Tumblin, "Coded exposure photography: motion deblurring using fluttered shutter," ACM Trans. Graph. **25**, 795–804 (2006).

15. A. Agrawal and Yi Xu, "Coded exposure deblurring: optimized codes for PSF estimation and invertibility," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2009), pp. 2066–2073.

16. A. Agrawal and R. Raskar, "Optimal single image capture for motion deblurring," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2009), pp. 2560–2567.

17. S. McCloskey, "Velocity-dependent shutter sequences for motion deblurring," in *Computer Vision–ECCV* (Springer, 2010), pp. 309–322.

18. S. Harshavardhan, S. Gupta, and K. S. Venkatesh, "Flutter shutter based motion deblurring in complex scenes," in *Annual IEEE India Conference (INDICON)* (IEEE, 2013), pp. 1–6.

19. H.-G. Jeon, J.-Y. Lee, Y. Han, S. J. Kim, and I. S. Kweon, "Complementary sets of shutter sequences for motion deblurring," in *IEEE International Conference on Computer Vision (ICCV)* (IEEE, 2015), pp. 3541–3549.

20. G. Cui, X. Ye, J. Zhao, L. Zhu, and Y. Chen, "Multi-frame motion deblurring using coded exposure imaging with complementary fluttering sequences," Opt. Laser Technol. **126**, 106119 (2020).

21. R. Gonzalez and R. Woods, *Digital Image Processing*, 4th ed. (Pearson, 2017).

22. S. McCloskey, Y. Ding, and J. Yu, "Design and estimation of coded exposure point spread functions," IEEE Trans. Pattern Anal. Mach. Intell. **34**, 2071–2077 (2012).

23. H.-G. Jeon, J.-Y. Lee, Y. Han, S. J. Kim, and I. S. Kweon, "Generating fluttering patterns with low autocorrelation for coded exposure imaging," Int. J. Comput. Vis. **123**, 269–286 (2017).

24. J. Chai, H. Zeng, A. Li, and E. W. T. Ngai, "Deep learning in computer vision: a critical review of emerging techniques and application scenarios," Mach. Learn. Appl. **6**, 100134 (2021).

25. S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2017), pp. 257–265.

26. X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2018), pp. 8174–8182.

27. S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Multi-stage progressive image restoration," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2021), pp. 14816–14826.

28. S.-J. Cho, S.-W. Ji, J.-P. Hong, S.-W. Jung, and S.-J. Ko, "Rethinking coarse-to-fine approach in single image deblurring," in *IEEE/CVF International Conference on Computer Vision (ICCV)* (IEEE, 2021), pp. 4621–4630.

29. X. Mao, Y. Liu, W. Shen, Q. Li, and Y. Wang, "Deep residual Fourier transformation for single image deblurring," arXiv, arXiv:2111.11745v1 (2021).

30. K. Kim, S. Lee, and S. Cho, "MSSNet: multi-scale-stage network for single image deblurring," in *Computer Vision–ECCV* (Springer, 2023), Vol. **13802**, pp. 524–539.

31. D. P. Kingma and M. Welling, "Auto-encoding variational bayes," arXiv, arXiv:1312.6114 (2022).

32. A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: an overview," IEEE Signal Process. Mag. **35**, 53–65 (2018).

33. C. Hu, H. Huang, M. Chen, S. Yang, and H. Chen, "Video object detection from one single image through opto-electronic neural network," APL Photonics **6**, 046104 (2021).

34. Y. Liang, H. Huang, J. Li, X. Dong, M. Chen, S. Yang, and H. Chen, "Action recognition based on discrete cosine transform by optical pixel-wise encoding," APL Photonics **7**, 116101 (2022).

35. Z. Zhang, B. Zhang, X. Yuan, S. Zheng, X. Su, J. Suo, D. J. Brady, and Q. Dai, "From compressive sampling to compressive tasking: retrieving semantics in compressed domain with low bandwidth," PhotoniX **3**, 19 (2022).

36. M. Courbariaux, Y. Bengio, and J.-P. David, "BinaryConnect: training deep neural networks with binary weights during propagations," in *Advances in Neural Information Processing Systems (NeurIPS)* (Curran Associates, 2015), Vol. **28**.

37. I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio, "Binarized neural networks," in *Advances in Neural Information Processing Systems (NeurIPS)* (Curran Associates, 2016), Vol. **29**.

38. J. Cao, Y. Li, M. Sun, Y. Chen, D. Lischinski, D. Cohen-Or, B. Chen, and C. Tu, "DO-Conv: depthwise over-parameterized convolutional layer," IEEE Trans. Image Process. **31**, 3726–3736 (2022).

39. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2016), pp. 770–778.

40. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: an imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems (NeurIPS)* (Curran Associates, 2019), pp. 8024–8035.

41. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv, arXiv:1412.6980 (2017).

42. I. Loshchilov and F. Hutter, "SGDR: stochastic gradient descent with warm restarts," in *5th International Conference on Learning Representations (ICLR)* (OpenReview.net, 2017).

43. X. Chu, L. Chen, C. Chen, and X. Lu, "Improving image restoration by revisiting global information aggregation," in *Computer Vision–ECCV* (Springer, 2022), pp. 53–71.

44. G. Cui, X. Ye, J. Zhao, L. Zhu, Y. Chen, and Y. Zhang, "An effective coded exposure photography framework using optimal fluttering pattern generation," Opt. Lasers Eng. **139**, 106489 (2021).