All-Optical Nonlinear Diffractive Deep Network for Ultrafast Image Denoising

Xiaoling Zhou[1*], Zhemg Lee[2*], Wei Ye[1†], Rui Xie[1], Wenbo Zhang[2],
Guanju Peng[2], Zongze Li[3†], Shikun Zhang[1†]
[1]Peking University [2]Tianjin University [3]Pengcheng Laboratory
xiaolingzhou@stu.pku.edu.cn, zhemglee@tju.edu.cn

Abstract

Image denoising poses a significant challenge in image processing, aiming to remove noise and artifacts from input images. However, current denoising algorithms implemented on electronic chips frequently encounter latency issues and demand substantial computational resources. In this paper, we introduce an all-optical Nonlinear Diffractive Denoising Deep Network (N3DNet) for image denoising at the speed of light. Initially, we incorporate an image encoding and pre-denoising module into the Diffractive Deep Neural Network and integrate a nonlinear activation function, termed the phase exponential linear function, after each diffractive layer, thereby boosting the network's nonlinear modeling and denoising capabilities. Subsequently, we devise a new reinforcement learning algorithm called regularization-assisted deep Q-network to optimize N3DNet. Finally, leveraging 3D printing techniques, we fabricate N3DNet using the trained parameters and construct a physical experimental system for real-world applications. A new benchmark dataset, termed MIDD, is constructed for mode image denoising, comprising 120K pairs of noisy/noise-free images captured from real fiber communication systems across various transmission lengths. Through extensive simulation and real experiments, we validate that N3DNet outperforms both traditional and deep learning-based denoising approaches across various datasets. Remarkably, its processing speed is nearly 3,800 times faster than electronic chip-based methods.

1. Introduction

Image denoising is a crucial and challenging task across multiple fields, including signal processing [42, 46], computational imaging [5, 53, 58], and computer vision [50, 51, 60]. For instance, effective noise reduction in computational imaging is paramount to counteracting the detrimental effects of various noise sources and artifacts, such as those originating from image sensors, quantization, channel transmission, and environmental factors [38, 53]. Moreover, this task also holds considerable importance in fiber communication systems [4, 21]. As fiber communication links extend, the noise during mode communication inevitably degrades communication quality and increases the complexity of signal processing [14, 62].

Considering the pivotal role of image denoising, substantial research has been undertaken on noise reduction utilizing electronic chip-based devices. These methods can be broadly categorized into two groups: traditional denoisers such as Wiener filtering [17] and block-matching and 3D filtering (BM3D) [13], and deep neural network (DNN)-based methods [35, 39], such as denoising convolutional neural network (DnCNN) [54], real image denoising network (RIDNet) [3], and masked training (MT) [11]. Despite being promising, these methods are time-consuming and impose substantial computational burdens [24, 40]. Consequently, exploring more efficient and effective computational platforms such as optical computing and quantum computing presents an intriguing avenue. Diffractive Deep Neural Network (D^2NN) [33] is an all-optical diffractive neural network trained via deep learning, demonstrating impressive processing speed and performance. However, due to its oversight of noise filtering during training and its limited feature representation capability, the denoising performance of this architecture is considerably inferior to that of electronic chip-based methods [23, 36].

This paper introduces a novel all-optical Nonlinear Diffractive Denoising Deep Network (N3DNet) designed to filter out image noise at the speed of light. To construct N3DNet, we first incorporate an image encoding and pre-denoising module into the original D^2NN, with the goal of achieving preliminary noise suppression and signal enhancement. Furthermore, we integrate a novel nonlinear activation function, termed the phase exponential linear (PEL) function, following each phase modulation layer within the network, thereby enhancing its feature representation capability and flexibility. The entire structure of the proposed N3DNet is depicted in Fig. 1, consisting of a sequence

*Equal contribution. † Corresponding authors.

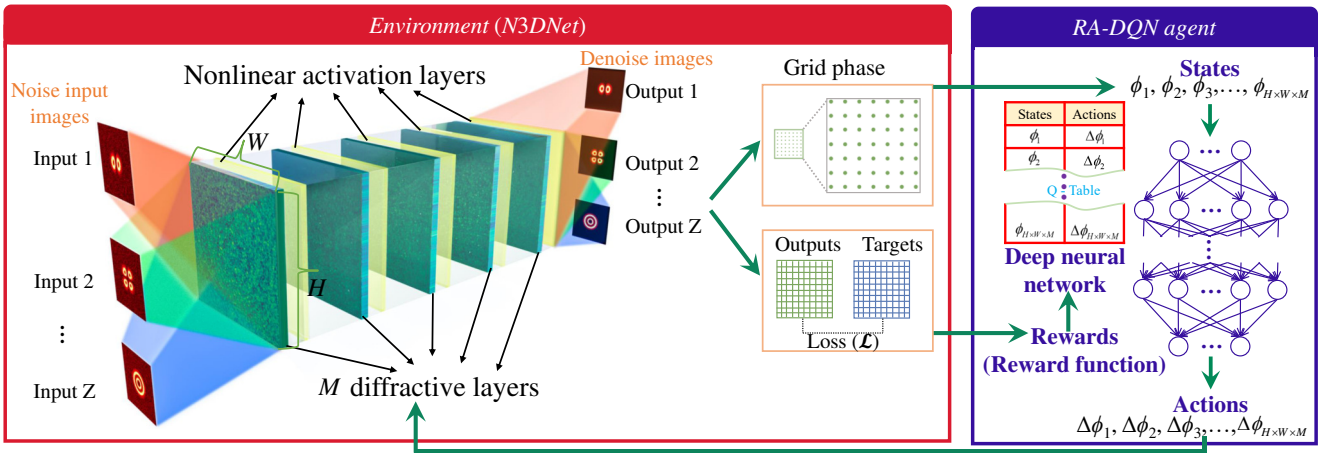

Figure 1. Schematic of the proposed N3DNet framework. We illustrate an example of utilizing N3DNet for mode image denoising, which is trained using the RA-DQN algorithm.

of consecutive nonlinear phase modulation planes (layers), all of which are transmissive. This innovative diffractive architecture can achieve ultrafast image denoising by utilizing optical diffraction and the pixel-level global attention mechanism to eliminate noise-related optical modes while preserving the features of interest. By selectively diffracting each cell in N3DNet and employing the nonlinear layer to alter phase values, exquisite manipulation of the output image is achieved. Consequently, after computer-based training, N3DNet generates denoised images at the speed of light in real-world applications, eliminating the need for any digital computation. To optimize N3DNet, we design a new reinforcement learning algorithm called regularization-assisted deep Q-network (RA-DQN), which significantly enhances N3DNet's convergence performance and improves the network's convergence speed. Ultimately, by leveraging the trained network parameters, we fabricate N3DNet using 3D printing and establish a physical experimental system to facilitate its practical application.

Extensive simulation and real experiments are conducted on both mode image and benchmark image processing datasets. Given that previous studies on mode processing in fiber communication systems have primarily relied on optical devices that necessitate only device design without the need for training [16, 37], accessible datasets in this domain are lacking. Taking the initiative, we introduce a dedicated mode image denoising dataset (MIDD) for six linearly polarized (LP) modes and three orbital angular momentum (OAM) modes that are widely used in scientific research and commercial fiber communication systems [30, 44]. This dataset consists of 100K noisy mode images for training and 20K noisy images for testing, all of which are captured from real-world fiber communication systems. In addition to MIDD, we conduct experiments on four synthetic noisy datasets and two real-world noisy datasets. The comprehensive experiments demonstrate that N3DNet significantly enhances denoising performance compared to previous optical computing methods while maintaining nearly identical processing speed. Specifically, it achieves an average performance improvement of 9.15 dB in PSNR, 0.248 in SSIM, and 0.202 in LPIPS compared to the best-performing optical computing methods. The processing time of N3DNet reaches the microsecond level, while its power consumption is in the nanojoule range. Moreover, in comparison with electronic chip-based noise reduction approaches, including both traditional and DNN-based methods, N3DNet not only achieves state-of-the-art (SOTA) performance but also boosts processing speed by nearly 3,800 times.

Overall, the main contributions are outlined as follows:

- We introduce N3DNet, an all-optical nonlinear diffractive deep neural network designed for image denoising at the speed of light.
- We propose an RA-DQN algorithm to optimize N3DNet, which is then fabricated using 3D printing to develop a physical experimental system for real-world applications.
- We introduce MIDD, a benchmark dataset for mode image denoising. It consists of 120K images captured in real scenarios, including single-mode, few-mode, and multi-mode fibers across various transmission distances.
- We conduct extensive simulated and real experiments across various datasets. The results demonstrate that N3DNet achieves SOTA performance with remarkable efficiency, operating approximately 3,800 times faster than methods relying on electronic chips.

2. Related Work

Image denoising is a critical aspect across various domains [9, 22, 50, 59], which aims to improve image qual-

ity by estimating clean images from noisy observations [15, 49, 61]. Over the years, researchers have proposed numerous denoising techniques based on electronic chips. Traditional methods include non-local means [8], total variation regularization [2], and BM3D [13]. More recently, DNN-based approaches, such as convolutional neural networks (CNNs) [31] and generative adversarial networks (GANs) [12], have shown superior performance in preserving fine details while effectively removing noise [32, 45, 54]. However, despite great potential, these methods frequently impose substantial computational burdens and latency issues, thereby limiting their scalability, especially in resource-constrained environments.

Diffractive deep neural network (D^2NN) [33] represents an all-optical diffractive neural network trained via deep learning [28], which comprises multiple transmitting and diffractive phase planes, each adorned with numerous neurons manipulating or guiding incident light through diffraction. In comparison with methods deployed on electronic devices [12, 19], D^2NN offers numerous advantages such as ultra-high speed, low power consumption, wavelength multiplexing, and polarization multiplexing [34, 36]. The phases or amplitudes of neurons serve as parameters awaiting training within D^2NN. Through iterative modulations of incident light across multiple diffractive surfaces, D^2NN can achieve the capability of all-optical inference [33, 40, 52]. Recently, researchers have proposed an analog diffractive image (ADI) denoiser [23] that leverages the D^2NN architecture for image denoising. However, this method falls short of achieving satisfactory results due to its negligence of nonlinear modeling capability and reliance on suboptimal optimization strategies.

3. Methodology

This section first elaborates on the propagation diagram of N3DNet (§ 3.1), as depicted in Fig. 2, followed by the introduction of the reinforcement learning algorithm (§ 3.2) aimed at optimizing N3DNet. Ultimately, the construction of the physical experimental system utilizing N3DNet for real applications (§ 3.3), as shown in Fig. 3, is delineated.

3.1. Nonlinear Diffractive Denoising Deep Network

The forward propagation of N3DNet, as shown in Fig. 2, consists of two components: the image encoding and pre-denoising module (I) and the all-optical diffractive propagation module (II). These components represent the encoding process of images and the transmission of optical signals in free space (prior to the input layer), as well as the propagation of optical signals across nonlinear diffractive layers, respectively.

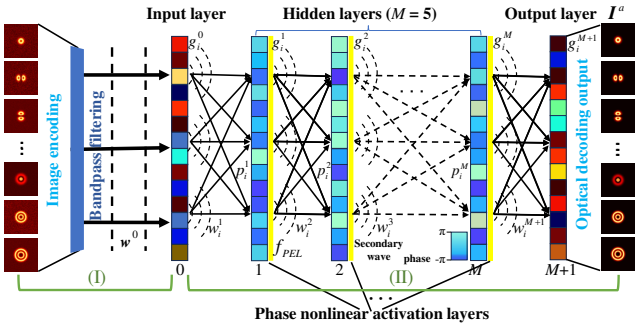

Figure 2. Forward propagation diagram of N3DNet.

3.1.1. Image Encoding and Pre-Denoising Module

The noisy input image is initially encoded onto an optical carrier through amplitude modulation [56, 57] in free space. This encoded signal can be represented as

$$\boldsymbol{E}(x, y, z) = \boldsymbol{A}_s(x, y) A_0 \exp\left[-j\boldsymbol{k} \cdot \boldsymbol{r}(x, y, z)\right], \quad (1)$$

where A_0 and $\boldsymbol{A}_s(x, y)$ refer to the amplitude of the carrier light and the image signal, respectively. Moreover, $\boldsymbol{r} = x\boldsymbol{e}_x + y\boldsymbol{e}_y + z\boldsymbol{e}_z$ denotes the position vector at point (x, y, z) and $\boldsymbol{k} = k_x\boldsymbol{e}_x + k_y\boldsymbol{e}_y + k_z\boldsymbol{e}_z$ is the wave vector, which delineates the wave's propagation direction. It satisfies $|\boldsymbol{k}| = 2\pi/\lambda$, where λ denotes the work wavelength. Additionally, $j = \sqrt{-1}$ represents the imaginary unit.

To achieve preliminary noise suppression and signal enhancement, we perform an optical Fourier transform on the encoded signal, followed by bandpass filtering. The filtered signal is represented as

$$\boldsymbol{h}(x, y, z) = \mathcal{F}^{-1}\{H(\tilde{u}, \tilde{v}) \cdot \mathcal{F}[\boldsymbol{E}(x, y, z)]\}, \quad (2)$$

where $H(\tilde{u}, \tilde{v})$ denotes a bandpass filter with center frequency $\tilde{u}$ and bandwidth $\tilde{v}$. $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ represent the Fourier transform function and its inverse, respectively.

Drawing from the Rayleigh-Sommerfeld diffraction equation [48], each neuron within the layers of N3DNet is regarded as a secondary source of a wave consisting of the following optical mode:

$$w_i^l(x, y, z) = \frac{z - z_i}{d^2}\left(\frac{1}{2\pi d} + \frac{1}{j\lambda}\right)\exp\left(\frac{j2\pi d}{\lambda}\right), \quad (3)$$

where l denotes the l-th layer of the network, i represents the i-th neuron located at (x_i, y_i, z_i) of layer l, and d is calculated as $d = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2}$. Across our diffractive neural network, all neurons adhere to the optical propagation process. The input pattern $h_i(x_i, y_i, z_i)$ of the i-th neuron in the input plane, is in general a complex-valued quantity and can carry information in its phase and amplitude channels. Accordingly, the resulting wave function g_i^0 on the input layer of N3DNet is generated through the diffraction of the plane-wave interacting

with the input:

$$g_i^0(x,y,z) = w_i^0(x,y,z)h_i(x_i,y_i,z_i), \tag{4}$$

which connects the input to the neurons of the first hidden layer, as shown in Fig. 2. Subsequently, the optical wave will propagate through the M hidden layers in N3DNet, as elaborated in the subsequent section.

3.1.2. All-Optical Diffractive Propagation Module

The amplitude and relative phase of the secondary wave are determined by the product of the input wave to the neuron and its transmission coefficient (p_i^l), both of which are complex-valued functions. Consequently, the output of the i-th neuron in the l-th layer, $g_i^l(x,y,z)$, is calculated as

$$g_i^l = f_{PEL}\{w_i^l(x,y,z)\, p_i^l(x_i,y_i,z_i) \sum\nolimits_k g_{k,i}^{l-1}\}, \tag{5}$$

where $g_{k,i}^{l-1}(x_i,y_i,z_i)$ denotes the output from the k-th neuron in the $(l-1)$-th diffractive layer incident on the i-th neuron in the l-th layer. Accordingly, the term $\sum_k g_{k,i}^{l-1}(x_i,y_i,z_i)$ represents the input wave to the i-th neuron in layer l. Besides, f_{PEL} refers to our proposed PEL activation function, which is

$$f_{PEL}(X) = \begin{cases} \beta X \exp(\alpha X), & \text{if } X > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

To obtain this activation function, we initially utilize metasurface [18] composed of Si_3N_4 and Er-doped TiO_2 to achieve phase modulation in the incident light. Then, we collect data pairs consisting of input and output phases, and fit these data pairs to obtain Eq. (6), with $\alpha = 0.5$ and $\beta = 0.2$. The secondary waves undergo diffraction between layers and interfere with each other to form a complex wave at the surface of the next layer, thereby supplying input to its neurons, as illustrated in Fig. 2. $p_i^l(x_i,y_i,z_i)$ in Eq. (5) refers to the transmission coefficient of the i-th neuron in layer l, expressed as a complex-valued function, and calculated as

$$p_i^l(x_i,y_i,z_i) = a_i^l(x_i,y_i,z_i)\exp\left[j\phi_i^l(x_i,y_i,z_i)\right], \tag{7}$$

where a_i^l indicates the amplitude term of the i-th neuron in the l-th layer, and ϕ_i^l is its phase term, with $0 \le a_i^l \le 1$ and $-\pi \le \phi_i^l \le \pi$.

From Eq. (5), the cells of the l-th phase plane perform complex modulation on the light, thereby modifying the original information of the optical wave. For a better understanding of these changes, Eq. (5) can be redefined with the relative amplitude of the secondary wave $\hat{A}_i^l$ and the additional phase delay $\Delta\Omega_i^l$ that the secondary wave encounters due to the input wave to the neuron and its transmission coefficient. Consequently, we have $g_i^l(x,y,z) = f_{PEL}\{w_i^l(x,y,z)\,\hat{A}_i^l \exp(j\Delta\Omega_i^l)\}$.

In summary, each hidden layer involves computations of Eqs. (5) and (7), culminating in its output $\boldsymbol{g}^l \in \mathbb{R}^{H\times W}$. Assuming N3DNet comprises M hidden layers (excluding the input and output planes), the input plane's output g_i^0 is propagated through these M layers. The intensity of the resulting optical field $\boldsymbol{I}^a$ detected at the output plane can be expressed as:

$$I_i^a = |g_i^{M+1}|^2, \tag{8}$$

where $g_i^{M+1} = \sum_k g_{k,i}^M(x_i,y_i,z_i)$ denotes the output of the i-th neuron in the output plane and $|g_i^{M+1}|$ represents its magnitude. Based on this forward propagation process, the results of the output plane $\boldsymbol{I}^a$ are compared with the intensity of the targets $\boldsymbol{I}^t$ and the resulting errors are backpropagated to iteratively update N3DNet.

3.2. Training Algorithm of N3DNet

To optimize N3DNet, we propose a novel reinforcement learning algorithm, named RA-DQN, which is adapted from DQN [41]. Its reward $\mathcal{R}$ equals the negative loss $\mathcal{L}$ of N3DNet, consisting of three components: Charbonnier loss [27] $\mathcal{L}_\mathcal{C}$, Fourier loss $\mathcal{L}_\mathcal{F}$, and FSIM loss [25] $\mathcal{L}_\mathcal{S}$. Among these, $\mathcal{L}_\mathcal{C}$ assesses the disparity between the pixel values of the network outputs and targets. To restore high-frequency information in images, $\mathcal{L}_\mathcal{F} = L_1(\mathcal{F}(\boldsymbol{I}^a), \mathcal{F}(\boldsymbol{I}^t))$ directs the model to learn frequency domain features and restore high-frequency information as effectively as possible, where $L_1(\cdot,\cdot)$ represents the L_1 loss. Additionally, $\mathcal{L}_\mathcal{S}$ evaluates image quality by comparing the structure and features. Accordingly, the loss function utilized to optimize N3DNet is calculated as

$$\mathcal{L} = \mathcal{L}_\mathcal{C}\left(\boldsymbol{I}^a, \boldsymbol{I}^t\right) + \hat{\alpha}\mathcal{L}_\mathcal{F}\left(\boldsymbol{I}^a, \boldsymbol{I}^t\right) + \hat{\beta}\mathcal{L}_\mathcal{S}\left(\boldsymbol{I}^a, \boldsymbol{I}^t\right), \tag{9}$$

where $\boldsymbol{I}^a$ and $\boldsymbol{I}^t$ represent the outputs and targets, respectively. Moreover, $\hat{\alpha}$ and $\hat{\beta}$ are two modulating hyperparameters used to adjust the relative strength of each loss term.

During the optimization process, the state $\mathcal{S}$ refers to the phase values of neurons in all diffractive layers, denoted as ϕ_i^l, where i ranges from 1 to $H\times W$ and l ranges from 1 to M, with $H\times W$ representing the number of cell grids in a phase plane. The action $\mathcal{A}$ represents the phase changes across all cells, denoted as $\Delta\phi_i^l$. At each step t, the RA-DQN agent receives the current state $\mathcal{S}_t$ and outputs an action $\mathcal{A}_t$. Subsequently, the environment (i.e., N3DNet) generates a new state ($\mathcal{S}_{t+1}$) and reward $\mathcal{R}_{t+1}$ based on $\mathcal{A}_t$. Consequently, the agent continually updates its parameters to enhance its decision-making abilities. The core of this process is the following value function iteration:

$$\begin{aligned} Q_{\boldsymbol{\theta}}(\mathcal{S}_t, \mathcal{A}_t) \leftarrow\ & Q_{\boldsymbol{\theta}}(\mathcal{S}_t, \mathcal{A}_t) + \\ & \gamma\left[\mathcal{R}_{t+1} + \epsilon \max\nolimits_\mathcal{A} \hat{Q}_{\hat{\boldsymbol{\theta}}}(\mathcal{S}_{t+1}, \mathcal{A}) - Q_{\boldsymbol{\theta}}(\mathcal{S}_t, \mathcal{A}_t)\right], \end{aligned} \tag{10}$$

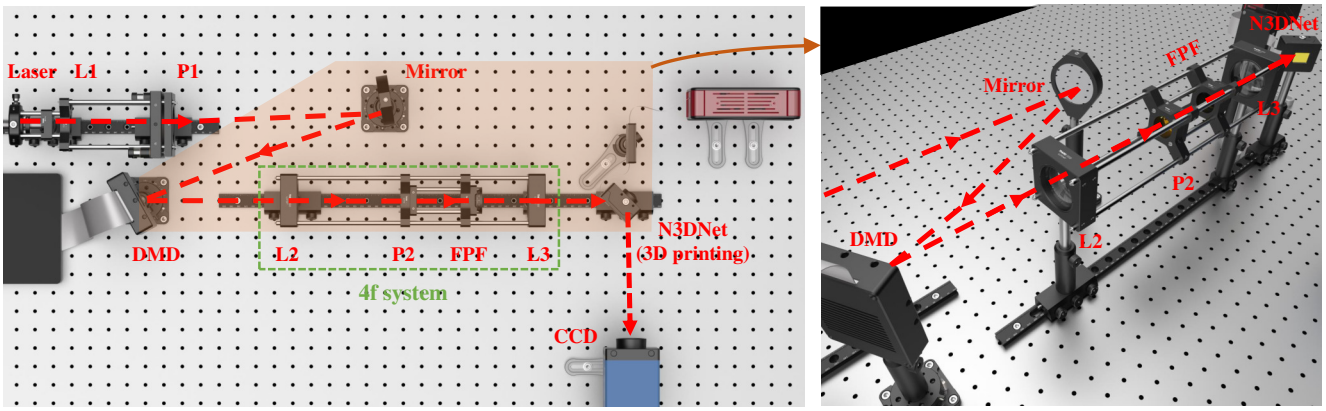

Figure 3. Physical experimental system for applying N3DNet. The image is first encoded by the DMD, then undergoes a Fourier transform, bandpass filtering, and an inverse Fourier transform through a 4f system. Subsequently, it is denoised using the diffractive phase planes of N3DNet and finally captured by the CCD for imaging.

where $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ refer to the parameters of the deep Q-network Q and the target network $\hat{Q}$, respectively. γ and ϵ represent the learning rate and discount factor, respectively, with settings following those of previous studies [20, 41, 47]. To enhance training stability, we apply a penalty to the disparity in Q-values between consecutive time steps when training the Q-network. Therefore, we minimize

$$\tilde{\mathcal{L}}(\boldsymbol{\theta}) = [\mathcal{R}_{t+1} + \epsilon \max_{\mathcal{A}} \hat{Q}_{\hat{\boldsymbol{\theta}}}(\mathcal{S}_{t+1}, \mathcal{A}) - Q_{\boldsymbol{\theta}}(\mathcal{S}_t, \mathcal{A}_t)]^2 + \kappa[Q_{\boldsymbol{\theta}}(\mathcal{S}_t, \mathcal{A}_t) - Q_{\boldsymbol{\theta}}(\mathcal{S}_{t-1}, \mathcal{A}_{t-1})]^2, \tag{11}$$

to optimize the deep Q-network, where κ denotes the regularization factor.

3.3. Physical Experimental System

The physical experimental system using N3DNet for image denoising is illustrated in Fig. 3. The energy source (optical carrier) is generated using a solid-state fiber laser operating at a wavelength of 1,550 nm. The coherent light wavefront is collimated using a relay lens, denoted as L1, followed by ensuring the single polarization property of the incident light using a polarizer, denoted as P1. Subsequently, the coherent light is reflected onto a digital micromirror device (DMD) at the correct incident angle utilizing a mirror. The DMD, which is mounted on a controller board, consists of 1,280 × 1,024 micro-mirrors with a pitch of 10.8 μm. Consequently, through the modulation of the DMD, images can be encoded onto the incident light.

The optical field from the DMD is then directed into a 4f optical system [10], which is composed of two relay lenses (L2 and L3), a linear polarizer (P2), and a Fabry–Perot filter (FPF). Among these, L2 performs the Fourier transform; P2 ensures the single polarization property; FPF positioned at the Fourier plane achieves bandpass filtering; and L3 conducts the inverse Fourier transform. Utilizing 3D printing techniques, the hidden layers of N3DNet are fabricated

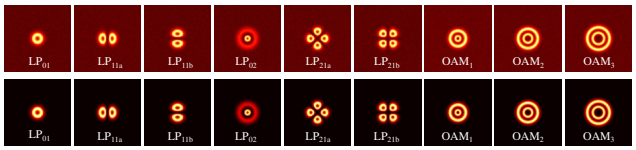

Figure 4. Illustration of the nine modes in the MIDD dataset. Top: Noise mode images. Below: Clean mode images.

with the trained parameters to modulate optical signals. The N3DNet consists of $5 \times 1{,}000 \times 1{,}000$ modulation elements, each with a size of 40 μm. After processing by the 4f optical system, the optical signal is modulated by the hidden layers of N3DNet, which are fabricated by a standard 3D-printing material (VeroBlackPlus RGD875) that have a high zero-order diffraction efficiency of 97% [33]. Ultimately, the modulated light from the N3DNet is captured by a sensor (located in an infrared-enhanced charge-coupled device, CCD) for optical decoding onto the CCD imaging plane.

4. Experimental Investigation

4.1. Configuration

Datasets. Currently, no existing dataset is available for mode image denoising in fiber communication systems. Therefore, we develop a benchmark MIDD dataset, which covers nine different modes, including six LP modes (LP_{01}, LP_{11a}, LP_{11b}, LP_{21a}, LP_{21b}, and LP_{02}) and three OAM modes (OAM_1, OAM_2, and OAM_3), as shown in Fig. 4.

The dataset is generated through transmissions and imaging using various types of optical fibers. A standard optical fiber communication system [30] was employed for data collection, comprising a laser source with a communication wavelength of 1,550 nm, a transmitter, an optical fiber, a lens, and an infrared-enhanced CCD. To construct the MIDD dataset, approximately 200 noisy images were

Dataset	σ/l	PSNR (↑)					SSIM (↑)					LPIPS (↓)				
		BM3D	DnCNN	RIDNet	MT	N3DNet	BM3D	DnCNN	RIDNet	MT	N3DNet	BM3D	DnCNN	RIDNet	MT	N3DNet
MIDD	20	28.89	29.07	29.10	28.95	**30.15**	0.835	0.817	0.855	0.849	**0.939**	0.207	0.204	0.185	0.190	**0.130**
	40	26.21	26.85	26.90	27.25	**28.12**	0.782	0.775	0.791	0.791	**0.914**	0.353	0.335	0.317	0.302	**0.197**
CSet9	25	29.95	31.65	**31.71**	31.70	31.68	0.941	0.955	0.955	0.957	**0.967**	0.162	0.155	0.170	0.165	**0.101**
	50	27.46	28.93	29.26	29.29	**30.01**	0.902	0.922	0.927	0.923	**0.930**	0.289	0.268	0.234	0.238	**0.175**
CBSD68	25	28.69	30.48	30.67	**30.73**	30.70	0.900	0.931	0.932	0.933	**0.941**	0.157	0.149	0.146	0.150	**0.095**
	50	25.79	27.35	27.65	27.68	**28.12**	0.836	0.870	0.878	**0.882**	0.880	0.328	0.300	0.245	0.242	**0.190**
Set12	25	27.59	29.92	30.01	29.96	**30.05**	0.777	0.842	0.850	0.845	**0.855**	0.220	0.159	**0.120**	0.162	0.159
	50	24.62	26.52	26.48	26.68	**26.80**	0.712	0.731	0.702	0.765	**0.780**	0.359	0.262	0.242	0.315	**0.195**
BSD68	25	28.58	29.00	29.31	29.55	**29.93**	0.800	0.817	0.801	0.785	**0.820**	0.262	0.221	0.197	0.231	**0.148**
	50	25.60	26.09	26.42	26.83	**27.35**	0.682	0.705	0.688	0.694	**0.715**	0.443	0.361	0.315	0.308	**0.251**

Table 1. Comparison of simulation results with advanced electronic chip-based denoising algorithms. ↑ (↓) denotes higher (lower) values are better. The best performance is highlighted in bold, while the second best is underlined.

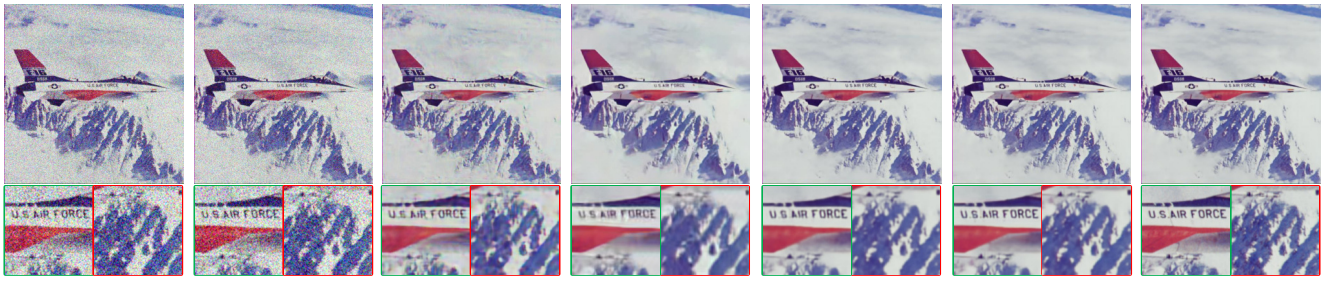

Figure 5. Visualization (simulation experiments) showcasing the denoising performance of various methods on an image from CSet9.

Dataset	BM3D	DnCNN	RIDNet	MT	N3DNet
Nam	37.30	35.55	39.09	39.11	**39.31**
SIDD	30.88	26.21	38.71	38.90	**38.97**

Table 2. The PSNR results for the two noisy real-world datasets.

collected for each of ten different types and manufacturers of fibers: single-mode, few-mode, and multi-mode, across 50 transmission lengths (l) ranging from 1 km to 50 km. This process yields over 100K mode images in total. Clean images were acquired with transmission distances of less than 1 m. Additionally, to assess the noise reduction performance, a total of 20K noisy mode images were collected for three additional optical fibers at transmission lengths of 5 km, 12 km, 20 km, 24 km, 36 km, 40 km, and 50 km. The data acquisition device was situated on a physics laboratory bench in the basement, designed to mitigate external interference from static electricity and vibrations.

Besides the MIDD data, we assess the performance of various denoising methods on four widely recognized image datasets: CSet9, CBSD68, Set12, and BSD68. The synthetic training data with Gaussian noise for these test sets are generated following the methodology outlined by Chen et al. [11]. Additionally, we incorporate two noisy real-world datasets, namely Nam [43] and SIDD [1], with the training and testing data constructed in accordance with the methodology outlined by Anwar et al. [3].

Evaluation Metrics. Building upon previous research [11, 45], we evaluate the performance of various methods using three image quality and perceptual metrics, including PSNR (dB) [26], SSIM [7], and LPIPS [55]. Furthermore, we compare the time and energy consumption of various denoising methods to assess their scalability.

Settings. In the simulation experiments, the 5 hidden layers (diffractive phase planes) in N3DNet operate at an incident wavelength (i.e., center frequency in the bandpass filter) of 1,550 nm. These planes are spaced 5 mm apart, each followed by a PEL activation layer. Moreover, every layer has dimensions of 40 mm $\times$ 40 mm and is divided into a grid of 1,000$\times$1,000 cells. The input and output layers have the same structure (including the number of units and plane spacing) as the hidden layer, except for omitting the PEL layer. Furthermore, the bandwidth of the bandpass filter is set to 500 GHz. The model is trained from scratch using the RA-DQN algorithm with an initial learning rate of 1×10^{-4} and a batch size of 32 for 2,000 iterations. The values of the three hyperparameters in the training process, $\hat{\alpha}$, $\hat{\beta}$, and κ, are set to 0.1, 0.5, and 5×10^{-3}, respectively. All computations are performed on a Linux x64 server equipped with a 20-core 2.40 GHz CPU processor, 128 GB of RAM, and an NVIDIA RTX 3090 GPU. The experimental setups for all comparative methods are consistent with those described in the original papers. For real experiments, all electronic chip-based methods, e.g., MT, are quantized to INT8 and

l	PSNR (↑)					SSIM (↑)					LPIPS (↓)				
	W/o	MPLC	MT	ADI	N3DNet	W/o	MPLC	MT	ADI	N3DNet	W/o	MPLC	MT	ADI	N3DNet
12	14.13	15.81	28.35	22.75	**29.87**	0.591	0.587	0.844	0.751	**0.934**	0.410	0.407	0.191	0.285	**0.132**
24	14.02	15.15	27.71	20.39	**28.93**	0.481	0.489	0.830	0.708	**0.929**	0.485	0.465	**0.195**	0.352	0.198
36	12.78	13.89	26.21	17.83	**27.45**	0.458	0.480	0.793	0.637	**0.899**	0.521	0.492	0.306	0.426	**0.205**
50	11.09	11.25	25.03	15.17	**26.47**	0.412	0.431	0.754	0.553	**0.877**	0.663	0.635	0.337	0.508	**0.230**

Table 3. Comparison of the results from real experiments with existing methods across various transmission lengths in the MIDD dataset. "W/o" denotes the results without denoising.

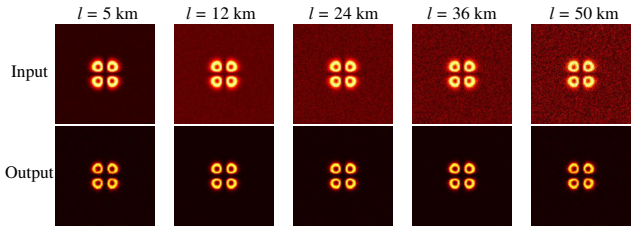

Figure 6. Visualization (real experiments) of N3DNet's denoising performance in the LP_{21b} mode under varying values of l.

deployed on the Snapdragon 8 Gen 3 high-end SoC.

4.2. Main Results

Simulation Experiments. We compare N3DNet with both traditional and advanced electronic chip-based denoising methods, including BM3D [13], DnCNN [54], RIDNet [3], and MT [11]. DnCNN and RIDNet are based on CNNs, while MT is based on visual transformers. Table 1 presents the comparative results for the MIDD and four synthetic noisy datasets, underscoring that N3DNet achieves superior denoising performance across all three metrics. Specifically, it outperforms the best-performing baselines by an average of 0.43 dB in PSNR, 0.032 in SSIM, and 0.066 in LPIPS. The effectiveness of N3DNet primarily stems from its reliance on pixel-level computations driven by the interference and diffraction of light. Because of interference and diffraction effects, it is possible to control both the intensity and phase of light at any point with a broader perspective. This method is significantly more complex than the multilayer structure in electronic chip-based neural networks, which only involve real-valued weights.

Moreover, Fig. 5 visually compares the denoising performance of various methods, where MPLC is a traditional mode processor. The results demonstrate that N3DNet achieves SOTA denoising performance. Table 2 demonstrates the effectiveness of our approach on real-world noisy datasets. It is clear that N3DNet outperforms RIDNet and MT by a margin of 0.24 dB and 0.14 dB, respectively.

Real Experiments. We utilize the constructed physical experimental system to perform real experiments on the MIDD dataset. Table 3 presents the comparison results among MPLC [6], MT [11], ADI [23], and N3DNet, where ADI directly utilizes the D^2NN architecture for image denoising. N3DNet consistently exhibits SOTA performance across a range of noise levels, underscoring its effectiveness in real-world denoising applications. Specifically, it demonstrates an average performance improvement of 9.15 dB in PSNR, 0.248 in SSIM, and 0.202 in LPIPS compared to the best-performing optical computing methods. Conversely, optical devices such as MPLC struggle with mode image denoising, resulting in poor performance. While ADI demonstrates improved performance compared to MPLC, its denoising capability is significantly inferior to that of N3DNet. This is primarily attributed to its suboptimal model structure and the oversight of the nonlinear modeling capabilities.

It is worth noting that the results of N3DNet's physical experiments are slightly inferior to the numerical simulation results. This discrepancy may be attributed to errors introduced by factors such as 3D printing, the optical system, and encoding. Nevertheless, the errors associated with N3DNet are comparable to or slightly smaller than those observed in algorithms, e.g., MT, deployed on electrical chips. Additionally, we visualize the denoising performance for mode images using N3DNet in Fig. 6. These images are captured by the CCD in the form of infrared imaging maps. As the level of noise increases, the proportion of energy in non-signal regions rises. Nevertheless, the results indicate that even in the presence of strong noise, N3DNet can achieve significant denoising performance.

4.3. Ablation Studies

Comparison of Optimization Algorithms. The loss variations of various optimization algorithms are illustrated in Fig. 7(a). The initial phase values of N3DNet for all algorithms are established using the same seed. In comparison to traditional gradient descent algorithms such as Adam and SGD, DQN-based methods demonstrate superior effectiveness in learning the state-action value function for N3DNet, as well as in enhancing learning efficiency through the implementation of experience replay and a target network. Moreover, by introducing a penalty for disparity in the Q-value, RA-DQN smooths the optimization path and reduces unnecessary fluctuations, leading to faster convergence and improved performance compared to DQN. Remarkably, af-

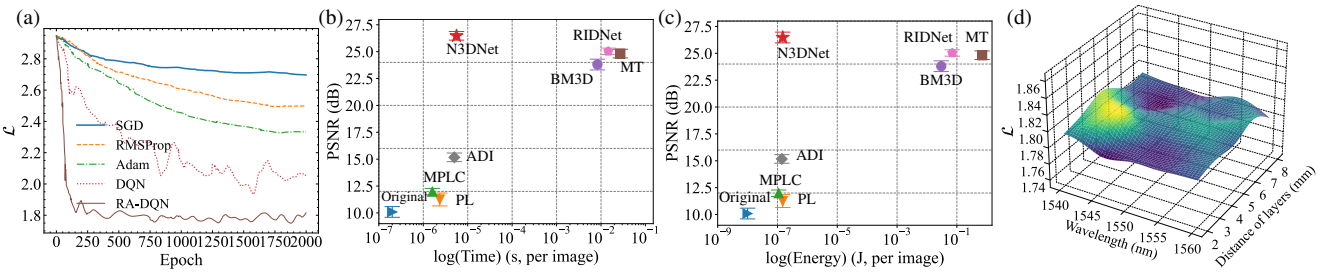

Figure 7. (a): Loss variation across epochs using different optimization algorithms on the MIDD dataset ($l = 50$). (b) and (c): Comparison of time and energy consumption across different methods. (d): 3D distribution of loss across various distances in layers and wavelengths.

$\mathcal{L}_\mathcal{C}$	$\mathcal{L}_\mathcal{F}$	$\mathcal{L}_\mathcal{S}$	**PSNR** (↑)	**SSIM** (↑)	**LPIPS** (↓)
✗	✗	✗	$14.89_{0.00}$	$0.484_{0.00}$	$0.479_{0.00}$
✓	✗	✗	$27.53_{0.39}$	$0.850_{0.04}$	$0.181_{0.05}$
✓	✓	✗	$\underline{29.32}_{0.45}$	$0.871_{0.06}$	$\underline{0.152}_{0.04}$
✓	✗	✓	$28.78_{0.37}$	$\underline{0.894}_{0.05}$	$0.178_{0.03}$
✗	✓	✓	$23.58_{0.42}$	$0.731_{0.03}$	$0.279_{0.05}$
✓	✓	✓	$\mathbf{30.56}_{0.37}$	$\mathbf{0.945}_{0.03}$	$\mathbf{0.103}_{0.04}$

Table 4. Ablation study on the configuration of the loss function. Both the average performance and standard deviation are provided. The first row shows the results without denoising.

ter training for less than 750 epochs, the N3DNet trained with the RA-DQN algorithm converges, whereas most traditional algorithms require more than 2,000 epochs. Additionally, the minimum loss achieved with the RA-DQN algorithm (1.78) is lower than those obtained with other algorithms (e.g., SGD, 2.70), indicating the superior convergence performance of RA-DQN.

Calculation Efficiency. Figs. 7(b) and (c) compare the latency and energy consumption of seven approaches deployed in real mode image denoising experiments. Here, PL [29] and MPLC [6] represent optical devices; ADI [23] is an analog diffractive image denoiser; and BM3D [13], RIDNet [3], and MT [11] are electronic chip-based methods. Unlike devices based on electrical neural networks, N3DNet operates without the need for electrical signal modulation, offering the benefits of faster processing and lower power consumption. Although both N3DNet and electricity-based methods achieve high PSNR values, significant disparities exist in terms of time and energy consumption. For 1,000 × 1,000 pixel images, our device's computation time per image is 4.4 μs, which is three orders of magnitude faster than electricity-based algorithms (16.7 ms, deployed on Qualcomm Snapdragon 8 Gen 3 SoC processor). Additionally, N3DNet consumes approximately six orders of magnitude less energy than electrical methods. Impressively, compared with optical computing approaches, N3DNet significantly enhances performance with nearly identical time and energy consumption.

Stability Analysis of Distance in Layers and Wavelength. Considering the inevitable errors in the distance between layers (5 mm) and the wavelength of incident light (1,550 nm) during the real experiments, we investigate the influence of disturbances in distance and wavelength on the loss in numerical experiments. As depicted in Fig. 7(d), when the distance between layers ranges from 2 mm to 8 mm and the wavelength varies from 1,540 nm to 1,560 nm, the loss consistently remains below 1.81, only approximately 1.69% higher than the optimal value of 1.78. This discrepancy falls entirely within the acceptable range of errors in the real-world experimental setup, which facilitates convenient processing and experimentation.

Analysis of Loss Function. We examine the influence of loss function configuration on the denoising performance of N3DNet. The average performance across various datasets with l = 12 km is reported in Table 4. Our adopted loss function demonstrates the highest efficacy. Moreover, all three loss terms are necessary and crucial for enhancing the denoising performance of N3DNet.

5. Conclusion

This study introduces N3DNet, an innovative all-optical nonlinear diffractive neural network, designed for image denoising at the speed of light. A newly proposed reinforcement learning algorithm is employed to optimize N3DNet, which is then fabricated using 3D printing with the trained parameters. Ultimately, a physical experimental system is constructed to apply N3DNet in real-world scenarios. Through extensive simulation and real experiments, we demonstrate that N3DNet not only outperforms existing methods in denoising performance but also significantly boosts processing speed, achieving a speedup of 3,800 times compared to electronic chip-based approaches. Given the ultrafast image processing speed, future research could explore extending this framework to address broader challenges in computer vision, such as video denoising, image super-resolution, and image restoration.

Acknowledgments

This work is partially supported by the Shenzhen Science and Technology Program (Grant No. RCBS20231211090517022).

References

[1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1692–1700, 2018. 6

[2] William K. Allard. Total variation regularization for image denoising, geometric theory. *SIAM Journal on Mathematical Analysis*, 39(4):1150–1190, 2008. 3

[3] Saeed Anwar and Nick Barnes. Real image denoising with feature attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3155–3164, 2019. 1, 6, 7, 8

[4] M. Arumugam. Optical fiber communication-an overview. *Pramana*, 57:849–869, 2001. 1

[5] Mario Bertero, Patrizia Boccacci, and Christine De Mol. *Introduction to inverse problems in imaging*. CRC press, 2021. 1

[6] Pauline Boucher, Arthur Goetschy, Giacomo Sorelli, Mattia Walschaers, and Nicolas Treps. Full characterization of the transmission properties of a multi-plane light converter. *Physical Review Research*, 3(2):023226, 2021. 7, 8

[7] Dominique Brunet, Edward R. Vrscay, and Zhou Wang. On the mathematical properties of the structural similarity index. *IEEE Transactions on Image Processing*, 21(4):1488–1499, 2011. 6

[8] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. Non-local means denoising. *Image Processing On Line*, 1: 208–212, 2011. 3

[9] Chakravarty R. Alla Chaitanya, Anton S. Kaplanyan, Christoph Schied, Marco Salvi, Aaron Lefohn, Derek Nowrouzezahrai, and Timo Aila. Interactive reconstruction of monte carlo image sequences using a recurrent denoising autoencoder. *ACM Transactions on Graphics*, 36(4):1–12, 2017. 2

[10] Julie Chang, Vincent Sitzmann, Xiong Dun, Wolfgang Heidrich, and Gordon Wetzstein. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Scientific Reports*, 8(1):12324, 2018. 5

[11] Haoyu Chen, Jinjin Gu, Yihao Liu, Salma Abdel Magid, Chao Dong, Qiong Wang, Hanspeter Pfister, and Lei Zhu. Masked image training for generalizable deep image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1692–1703, 2023. 1, 6, 7, 8

[12] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A. Bharath. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1):53–65, 2018. 3

[13] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising with block-matching and 3d filtering. In *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*, pages 354–365, 2006. 1, 3, 7, 8

[14] Alper Demir. Nonlinear phase noise in optical-fiber-communication systems. *Journal of Lightwave Technology*, 25(8):2002–2032, 2007. 1

[15] Linwei Fan, Fan Zhang, Hui Fan, and Caiming Zhang. Brief review of image denoising techniques. *Visual Computing for Industry, Biomedicine, and Art*, 2(1):7, 2019. 3

[16] Nicolas K. Fontaine, Joel Carpenter, Simon Gross, Sergio Leon-Saval, Yongmin Jung, David J. Richardson, and Rodrigo Amezcua-Correa. Photonic lanterns, 3-d waveguides, multiplane light conversion, and other components that enable space-division multiplexing. *Proceedings of the IEEE*, 110(11):1821–1834, 2022. 2

[17] Sandeep P. Ghael, Akbar M. Sayeed, and Richard G. Baraniuk. Improved wavelet denoising via empirical wiener filtering. In *Wavelet Applications in Signal and Image Processing*, pages 389–399, 1997. 1

[18] Yinghui Guo, Mingbo Pu, Fei Zhang, Mingfeng Xu, Xiong Li, Xiaoliang Ma, and Xiangang Luo. Classical and generalized geometric phase in electromagnetic metasurfaces. *Photonics Insights*, 1(1):R03, 2022. 4

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 3

[20] Zhang-Wei Hong, Shih-Yang Su, Tzu-Yun Shann, Yi-Hsiang Chang, and Chun-Yi Lee. A deep policy inference q-network for multi-agent systems. *arXiv preprint arXiv:1712.07893*, 2017. 5

[21] Rongqing Hui. *Introduction to fiber-optic communications*. Academic Press, 2019. 1

[22] Mustafa Işık, Krishna Mullia, Matthew Fisher, Jonathan Eisenmann, and Michaël Gharbi. Interactive monte carlo denoising using affinity of neural features. *ACM Transactions on Graphics*, 40(4):1–13, 2021. 2

[23] Çağatay Işıl, Tianyi Gan, Fazil Onuralp Ardic, Koray Mentesoglu, Jagrit Digani, Huseyin Karaca, Hanlong Chen, Jingxi Li, Deniz Mengu, Mona Jarrahi, et al. All-optical image denoising using a diffractive visual processor. *Light: Science & Applications*, 13(1):43, 2024. 1, 3, 7, 8

[24] Daniel Justus, John Brennan, Stephen Bonner, and Andrew Stephen McGough. Predicting the computational cost of deep learning models. In *IEEE International Conference on Big Data*, pages 3873–3882, 2018. 1

[25] Sergey Kastryulin, Jamil Zakirov, Denis Prokopenko, and Dmitry V. Dylov. Pytorch image quality: metrics for image quality assessment. *arXiv preprint arXiv:2208.14818*, 2022. 4

[26] Jari Korhonen and Junyong You. Peak signal-to-noise ratio revisited: Is simple beautiful? In *International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 37–38, 2012. 6

[27] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Fast and accurate image super-resolution with

deep laplacian pyramid networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(11):2599–2613, 2018. 4

[28] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. 3

[29] Sergio G. Leon-Saval, Alexander Argyros, and Joss Bland-Hawthorn. Photonic lanterns. *Nanophotonics*, 2(5-6):429–440, 2013. 8

[30] Guifang Li Li, Neng Bai, Ningbo Zhao, and Cen Xia. Space-division multiplexing: the next frontier in optical communication. *Advances in Optics and Photonics*, 6(4):413–487, 2014. 2, 5

[31] Zewen Li, Fan Liu, Wenjie Yang, Shouheng Peng, and Jun Zhou. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12):6999–7019, 2021. 3

[32] Kai Lin, Thomas H. Li, Shan Liu, and Ge Li. Real photographs denoising with noise domain adaptation and attentive generative adversarial network. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1717–1721, 2019. 3

[33] Xing Lin, Yair Rivenson, Nezih T. Yardimci, Muhammed Veli, Yi Luo, Mona Jarrahi, and Aydogan Ozcan. All-optical machine learning using diffractive deep neural networks. *Science*, 361(6406):1004–1008, 2018. 1, 3, 5

[34] Che Liu, Qian Ma, Zhang Jie Luo, Qiao Ru Hong, Qiang Xiao, Hao Chi Zhang, Long Miao, Wen Ming Yu, Qiang Cheng, Lianlin Li, et al. A programmable diffractive deep neural network based on a digital-coding metasurface array. *Nature Electronics*, 5(2):113–122, 2022. 3

[35] Alice Lucas, Michael Iliadis, Rafael Molina, and Aggelos K. Katsaggelos. Using deep neural networks for inverse problems in imaging: beyond analytical methods. *IEEE Signal Processing Magazine*, 35(1):20–36, 2018. 1

[36] Yi Luo, Yifan Zhao, Jingxi Li, Ege Çetintaş, Yair Rivenson, Mona Jarrahi, and Aydogan Ozcan. Computational imaging without a computer: seeing through random diffusers at the speed of light. *eLight*, 2(1):4, 2022. 1, 3

[37] Dan M. Marom, Yutaka Miyamoto, David T. Neilson, and Ioannis Tomkos. Optical switching in future fiber-optic networks utilizing spectral and spatial degrees of freedom. *Proceedings of the IEEE*, 110(11):1835–1852, 2022. 2

[38] Emil Martinec. Noise, dynamic range and bit depth in digital slrs. 2008. 1

[39] Michael T. McCann, Kyong Hwan Jin, and Michael Unser. Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine*, 34(6): 85–95, 2017. 1

[40] Deniz Mengu, Yi Luo, Yair Rivenson, and Aydogan Ozcan. Analysis of diffractive optical neural networks and their integration with electronic neural networks. *IEEE Journal of Selected Topics in Quantum Electronics*, 26:1–14, 2019. 1, 3

[41] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015. 4, 5

[42] Bernard Mulgrew, Peter Grant, and John Thompson. *Digital signal processing: concepts and applications*. Bloomsbury Publishing, 2002. 1

[43] Seonghyeon Nam, Youngbae Hwang, Yasuyuki Matsushita, and Seon Joo Kim. A holistic approach to cross-channel image noise modeling and its application to image denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1683–1691, 2016. 6

[44] Benjamin J. Puttnam, Georg Rademacher, and Ruben S. Luís. Space-division multiplexing for optical fiber communications. *Optica*, 8(9):1186–1203, 2021. 2

[45] Yuhui Quan, Mingqin Chen, Tongyao Pang, and Hui Ji. Self2self with dropout: Learning self-supervised denoising from single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1890–1898, 2020. 3, 6

[46] Richard A Roberts and Clifford T Mullis. *Digital signal processing*. Addison-Wesley Longman Publishing Co., Inc., 1987. 1

[47] Melrose Roderick, James MacGlashan, and Stefanie Tellex. Implementing the deep q-network. *arXiv preprint arXiv:1711.07478*, 2017. 5

[48] Fabin Shen and Anbo Wang. Fast-fourier-transform based numerical integration method for the rayleigh-sommerfeld diffraction formula. *Applied Optics*, 45(6):1102–1110, 2006. 3

[49] Chunwei Tian, Lunke Fei, Wenxian Zheng, Yong Xu, Wangmeng Zuo, and Chia-Wen Lin. Deep learning on image denoising: An overview. *Neural Networks*, 131:251–275, 2020. 3

[50] Koji Tsuda and Gunnar Rätsch. Image reconstruction by linear programming. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 57–64, 2004. 1, 2

[51] Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis, and Eftychios Protopapadakis. Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*, 2018. 1

[52] Peipei Wang, Wenjie Xiong, Zebin Huang, Yanliang He, Junmin Liu, Huapeng Ye, Jiangnan Xiao, Ying Li, Dianyuan Fan, and Shuqing Chen. Diffractive deep neural network for optical orbital angular momentum multiplexing and demultiplexing. *IEEE Journal of Selected Topics in Quantum Electronics*, 28(4):1–11, 2021. 3

[53] Heng Wu, Ruizhou Wang, Genping Zhao, Huapan Xiao, Jian Liang, Daodang Wang, Xiaobo Tian, Lianglun Cheng, and Xianmin Zhang. Deep-learning denoising computational ghost imaging. *Optics and Lasers in Engineering*, 134: 106183, 2020. 1

[54] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. 1, 3, 7

[55] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–595, 2018. 6

[56] Ziyang Zheng, Zhengyang Duan, Hang Chen, Rui Yang, Sheng Gao, Haiou Zhang, Hongkai Xiong, and Xing Lin. Dual adaptive training of photonic neural networks. *Nature Machine Intelligence*, 5(10):1119–1129, 2023. 3

[57] Jingshan Zhong, Zhong Wen, Quanzhi Li, Qilin Deng, and Qing Yang. Efficient reference-less transmission matrix retrieval for dmd amplitude modulation. In *Computational Optical Imaging and Artificial Intelligence in Biomedical Sciences*, pages 20–22, 2024. 3

[58] Xiaoling Zhou and Ou Wu. Which samples should be learned first: Easy or hard? *IEEE Transactions on Neural Networks and Learning Systems*, 2023. 1

[59] Xiaoling Zhou, Ou Wu, and Mengyang Li. Investigating the sample weighting mechanism using an interpretable weighting framework. *IEEE Transactions on Knowledge and Data Engineering*, 36(5):2041–2055, 2023. 2

[60] Xiaoling Zhou, Ou Wu, and Nan Yang. Adversarial training with anti-adversaries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 1

[61] Xiaoling Zhou, Wei Ye, Zhemg Lee, Rui Xie, and Shikun Zhang. Boosting model resilience via implicit adversarial data augmentation. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 5653–5661, 2024. 3

[62] John Zweck and Curtis R. Menyuk. Validity of the additive white gaussian noise model for quasi-linear long-haul return-to-zero optical fiber communications systems. *Journal of Lightwave Technology*, 27(16):3324–3335, 2009. 1