Two-Stage Training Method for High-Quality Reconstruction in Single-Pixel Imaging

Hui Shao(邵辉)[1], He Huang(黄何)[1], Yu-Xiao Wei(魏宇霄)[2], Hui-Juan Zhang(张会娟)[2], Zhao-Hua Yang(杨照华)[3], and Yuan-Jin Yu(余远金)[2*]

[1]*College of Information Science and Engineering, Huaqiao University, Xiamen 361021, China*
[2]*School of Automation, Beijing Institute of Technology, Beijing 100081, China*
[3]*School of Instrumentation and Optoelectronics Engineering, Beihang University, Beijing 100191, China*

(Received 5 October 2024; accepted manuscript online 22 November 2024)

A two-stage training method is proposed to enhance imaging quality and reduce reconstruction time in data-driven single-pixel imaging (SPI) under undersampling conditions. This approach leverages a deep learning algorithm to simulate single-pixel detection and image reconstruction. During the initial training stage, an L_2 regularization constraint is imposed on convolution modulation patterns to determine the optimal initial network weights. In the subsequent stage, a coupled deep learning method integrating coded-aperture design and SPI is adopted, which utilizes backpropagation of the loss function to iteratively optimize both the binarized modulation patterns and imaging network parameters. By reducing the binarization errors introduced by the dithering algorithm, this approach improves the quality of data-driven SPI. Compared with traditional deep-learning SPI methods, the proposed method significantly reduces computational complexity, resulting in accelerated image reconstruction. Experimental and simulation results demonstrate the advantages of the method, including high imaging quality, short image reconstruction time, and simplified training. For an image size of 64×64 pixels and 10% sampling rate, the proposed method achieves a peak signal-to-noise ratio of 23.22 dB, structural similarity index of 0.76, and image reconstruction time of approximately 2.57×10^{-4} seconds.

DOI: 10.1088/0256-307X/41/12/124202

Single-pixel imaging (SPI)[1–4] is an emerging computational imaging method that has garnered significant attention owing to its low manufacturing cost, high sensitivity, and resistance to interference. SPI is widely applied in various domains, including optical encryption,[5] 3D imaging,[6] target classification,[7] and target tracking.[8] Additionally, a significant avenue for current research is the facilitation of single band detection for which focal plane arrays have not yet been fully developed, particularly in the mid-infrared[9,10] and terahertz ranges.[11,12] However, prolonged image reconstruction times have significantly limited the application of SPI in these fields. Therefore, achieving high-quality image reconstruction at low sampling rates is critical for advancing single-pixel imaging.

SPI relies on substantial number of samples to reconstruct an image, and the quality of the reconstructed image is strongly correlated with the number of illumination patterns. Initially, random patterns were employed; however, these resulted in inefficient image reconstruction.[13] Wang *et al.* utilized a "0-1" random matrix based on a Bernoulli distribution, to collect approximately 50% of the optical energy from the object to the detector in spectral imaging.[14] Subsequently, efficient orthogonal basis illumination patterns such as Hadamard basis patterns with different ordering methods have been investigated.[15–17] SPI has evolved from addressing nonlinear optimization problems[18] to leveraging correlation information between illumination patterns and measurements for reconstruction.[19] Owing to the limited reconstruction performance of non-iterative methods, iterative algorithms have been developed to optimize image reconstruction.[20] However, optimizing the uncertainty problem through iterative algorithms results in an increased reconstruction time and fails to meet real-time processing requirements. For uncertain optimization problems, deep learning (DL) has emerged as a promising solution.[21] The DL methods, which have been widely applied in fields such as computer vision and speech recognition, have also been integrated into SPI systems to achieve higher imaging quality at low sampling rates.[22–24] To address the problems of data-driven generalization and interpretability, DL algorithms have been combined with physical models of the reconstructed image, employing a reverse gradient constraint to enhance image quality.[25,26] However, these methods constrain imaging quality. To address this limitation, Tian *et al.* proposed an effective SPI method based on channel-attention convolutional neural networks,[27] wherein learned patterns were integrated with mean squared error (MSE) loss[28] and TV regularization loss[29] to ensure that the network had a strong generalization ability. However, the image reconstruction

*Corresponding author. Email: yuanjin.yu@bit.edu.cn
© 2024 Chinese Physical Society and IOP Publishing Ltd

time decreases as the complexity of the network increases. Additionally, most SPI currently employ DMDs as modulators. When employing grayscale patterns, a dithering algorithm is typically required to convert them into binary DMD patterns, which introduce practical errors.[26,30] Fu *et al.* proposed a two-stage method for the binary processing of optimized grayscale patterns. A symbol function was used to optimize the speckle, and the recognition network was fine-tuned to improve its accuracy.[31] To improve the correlation between the modulation matrix and the imaging network, Bacca *et al.* proposed a one-stage coupled deep learning method that integrates coded aperture design and SPI to constrain binarized patterns.[32] However, this approach fails to fully extract image information and is prone to local optima, resulting in suboptimal image quality. Therefore, the optimization of the correlation between binary modulation patterns and the imaging network remains a key challenge for DL-based SPI.

To solve these challenges, we propose a two-stage training method aimed at optimizing binary patterns and developing an efficient, streamlined imaging network. In contrast to traditional strategies, the first stage simulates the modulation of a DMD by utilizing a convolutional layer with dimensions matching those of the image. During training, the L_2 regularization constraint was applied to the modulation layer. In the second stage, a coupled deep learning approach was employed to integrate coded aperture design with SPI for imposing binarization constraints and optimizing the network parameters. The inclusion of non-binary patterns in this approach provides optimal initialization weights for the network, thereby improving its generalization performance.

The SPI imaging process is illustrated in Fig. 1. The image information $x \in \mathbb{R}^{MN}$ is reconstructed from the compressed measurement signals $y \in \mathbb{R}^{K}$, where K is the number of samples, and M, N are the width and height of the target image, such that $K \ll M \times N$. Assuming a series of illumination patterns $\boldsymbol{\Phi} \in \mathbb{R}^{K \times MN}$, the mathematical expression of SPI is given by

$$y = \boldsymbol{\Phi} x + d, \tag{1}$$

where d is the noise vector.

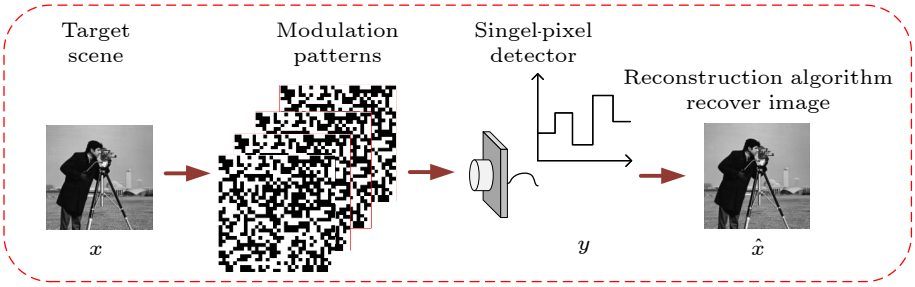

Fig. 1. Schematic of deep learning SPI.

The process described in Eq. (1) is known as the encoding, and different reconstruction algorithms can be employed to recover the image. In the context of SPI, the image reconstruction method based on DL can be divided into two parts: encoding and decoding, as expressed as

$$y = \mathrm{SPI}(x), \tag{2}$$

$$\hat{x} = f(y), \tag{3}$$

where $\mathrm{SPI}(\cdot)$ represents the simulated single-pixel detection, and $f(\cdot)$ denotes nonlinear decoding performed by the network.

For SPI networks, Gighham *et al.* first proposed a method that utilizes a fully connected layer to simulate modulated speckles and optimize gray speckles via network training.[22] However, in practical applications where a DMD is employed for modulation, the jitter algorithm is needed to approximate the optimized modulation patterns to the set $\{-1, +1\}$. Because the jitter algorithm incurs a certain degree of loss, Fu *et al.* proposed binarizing the optimized speckle using a symbolic function to enhance recognition accuracy.[31] Subsequent studies have demonstrated that speckle optimization can be binarized using a loss function. Therefore, to enhance the efficiency of both the encoder and decoder, this study proposes a novel two-stage training method, as illustrated in Fig. 2(a).

In the initial training stage, the total loss function includes an L_2 regularization constraint applied to the encoder along with the MSE loss function for the output result, as shown in Eq. (4). Optimized modulation patterns and imaging parameters were obtained using the gradient descent method.

$$L_{\mathrm{one}} = L_{\mathrm{mse}} + \lambda L_{l2}, \tag{4}$$

where L_{mse} denotes the MSE loss and λ is the weight factor of the L_2 constraint. Furthermore, L_{mse} and L_2 are expressed as:

$$L_{\mathrm{mse}} = \frac{1}{N}\sum_{i=1}^{N} \|\hat{x}_i - x_i\|^2, \tag{5}$$

$$L_{l2} = \frac{1}{2}\sum_{i=1}^{K} \|\boldsymbol{\varphi}^{[i]}\|^2, \tag{6}$$

where $\boldsymbol{\varphi}^{[i]}$ represents the i-th row of the modulation pattern $\boldsymbol{\Phi}$.

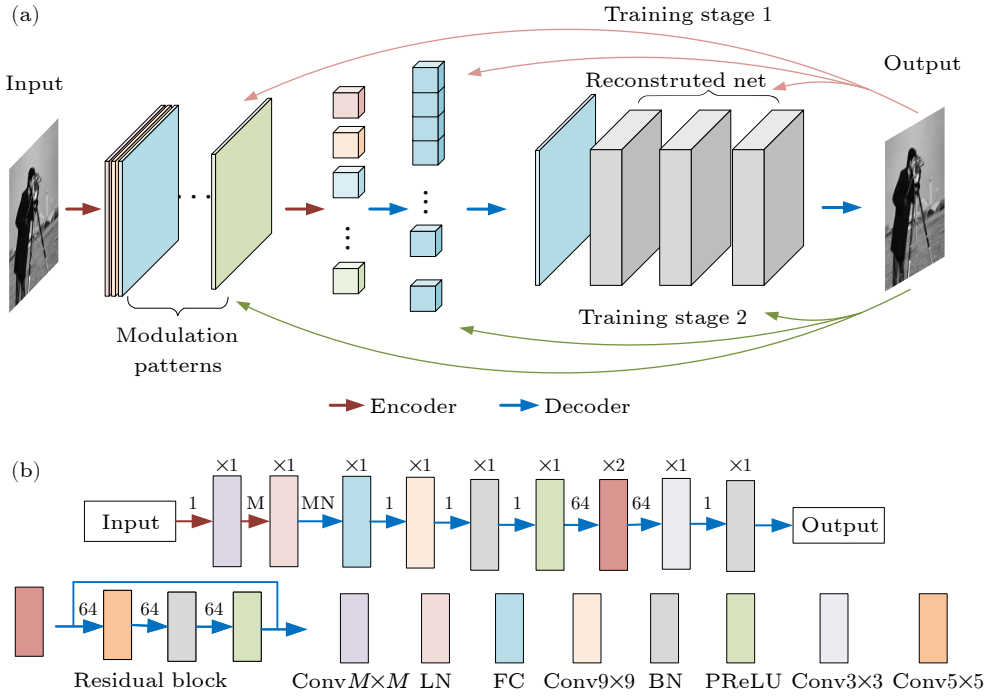

Fig. 2. Overview of the SPI framework: (a) two-step training strategy and (b) SPI network architecture. Conv $M \times M/9 \times 9/5 \times 5/3 \times 3$ represents the convolution kernel size and 1/2 represents the number of blocks per block. Numbers above the arrows indicate the number of channels.

The optimized modulation patterns can be obtained by iteratively updating the parameters; however, these patterns are not binary. Therefore, in the second stage, we adopted a coupled deep learning method that integrates coded-aperture design with SPI. This method leverages backpropagation of the loss function to iteratively binarize the modulation patterns and optimize the imaging network parameters. The obtained binary loss function is expressed as follows:

$$L_{\text{two}} = L_{\text{mse}} + \mu L_{\text{opt}}, \tag{7}$$

where μ is the binarized weight coefficient and L_{opt} represents the binarized constraint function, which is defined by the following equation:

$$L_{\text{opt}} = \sum_{k=1}^{K} \sum_{n=1}^{MN} (1 + \varphi_{k,n})^2 (1 - \varphi_{k,n})^2. \tag{8}$$

The designed network architecture is illustrated in Fig. 2(b). An unbiased convolution layer was employed as an encoder to simulate single-pixel modulation. The output of the modulation layer is first normalized to standardize the range of the actual acquired light intensity values. The output dimension $1 \times M$ feature vector is mapped to a dimension equal to the size of the target image $M \times N$ using a fully connected layer. This process simulates the solution of the linear pseudo-inverse in Eq. (3), thereby enabling the reconstruction of a pseudo-image. Inspired by large-scale convolution kernels, which facilitate broader coverage of pseudo-images and capture of rich and long-range information. Therefore, a convolution kernel of dimensions, 9×9 is used in the first layer. To further enhance image quality, we incorporated two smooth convolutional layers and two residual blocks at the backend for denoising. The k-th residual block is expressed using Eq. (9). The design of the decoder in this study effectively reduced computational complexity and accelerated the imaging process.

$$z_k = x_k + \text{PReLU}(\text{BN}(W x_k + b)), \tag{9}$$

where W is the convolution kernel weight, b is the bias, $\text{BN}(\cdot)$ is the batch normalization function, and $\text{PReLU}(\cdot)$ is the activation function, which can be expressed as

$$f(x_i) = \begin{cases} x_i, & \text{if } x_i > 0, \\ a_i x_i, & \text{if } x_i \leq 0, \end{cases} \tag{10}$$

where a_i is a learnable parameter and x_i is an input value.

To evaluate the performance of the proposed method, training was conducted on the STL-10 dataset, which consists of 100,000 unlabeled images. In this study, the images were converted to grayscale and resized to 64×64 pixels. The initial learning rate was set to 0.01, accompanied by an L_2 regularization coefficient of 0.01 and a binarization weight coefficient of 0.1. Each stage of the training process was iterated 100 times using an adaptive moment optimizer.[33] To emphasize the advantages of the two-stage modulated patterns, we initially compared them with different modulated pattern schemes including random patterns,[11] Hadamard patterns (GCS+S),[16] and one-stage binarization-optimized patterns.[32] Subsequently, we compared the proposed method with different deep learning-based SPI algorithms, including DCAN,[22] physics-enhanced methods,[25] and SPINet-unsupervised learning algorithms.[27] All the models were trained on

the STL-10 dataset to ensure a fair comparison. Additionally, to evaluate the quality of the reconstructed images, we employed two image quality assessment metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM).[34] These metrics facilitated a detailed comparison of the imaging reconstruction performance across different methods. The algorithms were implemented on a computer equipped with an Intel i9-13900K processor and NVIDIA GeForce RTX 4090 graphics card. Furthermore, to assess the generalization performance of the imaging algorithms, we tested them on an unfamiliar Set12 dataset. Table 1 lists the average reconstruction indices of the proposed algorithm for various modulation patterns. The optimal results are highlighted in bold.

Table 1. Comparison of average PSNR and SSIM with different modulated patterns in the integrated images of Set12 dataset.

Modulation patterns		**Sampling rates** 5%	8%	10%	20%
Random	PSNR	17.08	18.03	18.22	19.06
	SSIM	0.4367	0.5209	0.5742	0.6804
GCS+S	PSNR	18.06	19.15	19.31	19.86
	SSIM	0.5287	0.6463	0.6556	0.7597
One-Stage	PSNR	18.49	19.77	20.79	23.46
	SSIM	0.5608	0.6523	0.6974	0.7766
Ours	PSNR	**20.95**	**22.21**	**23.22**	**26.03**
	SSIM	**0.6889**	**0.7051**	**0.7660**	**0.8658**

Table 2. Comparison of the average PSNR and SSIM of different deep learning algorithms in the Set12 data integration image.

Algorithm		**Sampling rates** 5%	8%	10%	20%
DCAN	PSNR	18.39	19.66	20.43	22.25
	SSIM	0.5363	0.5718	0.6219	0.7340
Physics-enhanced	PSNR	18.64	19.76	20.61	22.73
	SSIM	0.5891	0.6544	0.7032	0.7507
SPINet-unsupervised	PSNR	**21.60**	**22.53**	23.21	25.91
	SSIM	0.6201	0.6809	0.7280	0.8394
Ours	PSNR	20.95	22.21	**23.22**	**26.03**
	SSIM	**0.6209**	**0.7051**	**0.7660**	**0.8658**

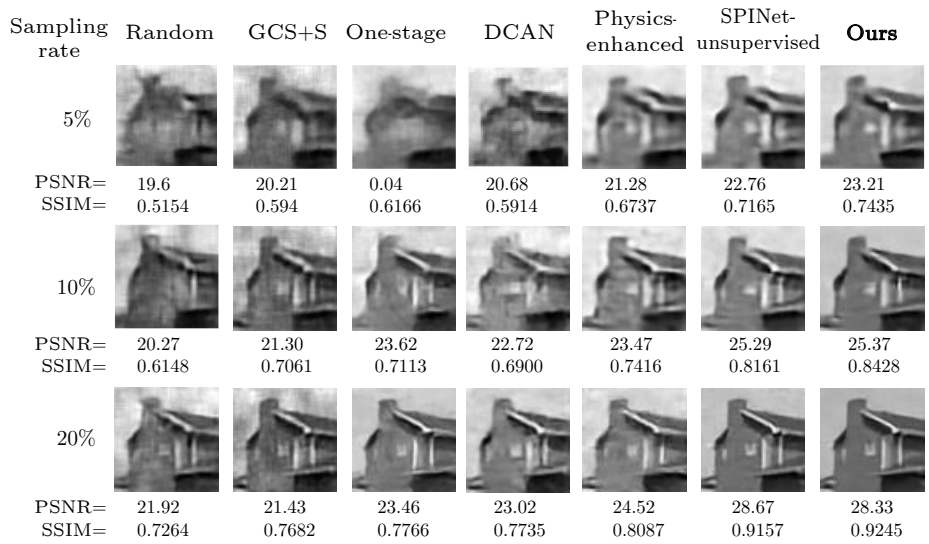

Fig. 3. Reconstructed imaged corresponding to each method at different sampling rates.

The results presented in Table 1 indicate that, under identical imaging network conditions, the proposed method achieves higher imaging quality, underscoring the effectiveness of the two-stage joint optimization method. Specifically, at a sampling rate of 5%, the proposed method achieved an average PSNR of 20.95 dB and an SSIM of 0.6209. When the sampling rate increased to 20%, the reconstructed images achieved a PSNR value of approxi-

mately 26.03 dB and an SSIM value approaching 0.8658. Table 2 presents a comparison of different image reconstruction algorithms, and Fig. 3 presents representative samples of the reconstructed images.

The results demonstrate that the proposed method achieves the highest SSIM value and ranks second in imaging PSNR at sampling rates of 5% and 8%, closely following the SPINet-unsupervised methods. Conversely, SPINet-unsupervised method employs an approximate binary approach that introduces certain errors in practical imaging applications. The results shown in Fig. 3 demonstrate that the proposed method can reconstruct additional image details.

To evaluate the robustness of the proposed method against noise, we compared its performance with three other DL methods. The PSNR and SSIM results of the proposed method, along with the DCAN, Physics-enhanced, and SPINet-unsupervised deep-learning algorithms at a 10% sampling rate under varying noise levels, are shown in Fig. 4.

Figure 4(a) shows the PSNR results of the four methods, where the measured single-pixel signal sequence exhibits Gaussian noise with signal-to-noise ratios varying from 15 to 30 dB. Figure 4(b) presents the SSIM results of the four methods. The results indicate that the proposed method maintains high-quality image reconstruction at different noise levels, demonstrating its robustness to noise.

To evaluate the image reconstruction time of the proposed two-stage approach, we compared the average imaging time for the four deep-learning algorithms when processing 100 images, as detailed in Table 3.

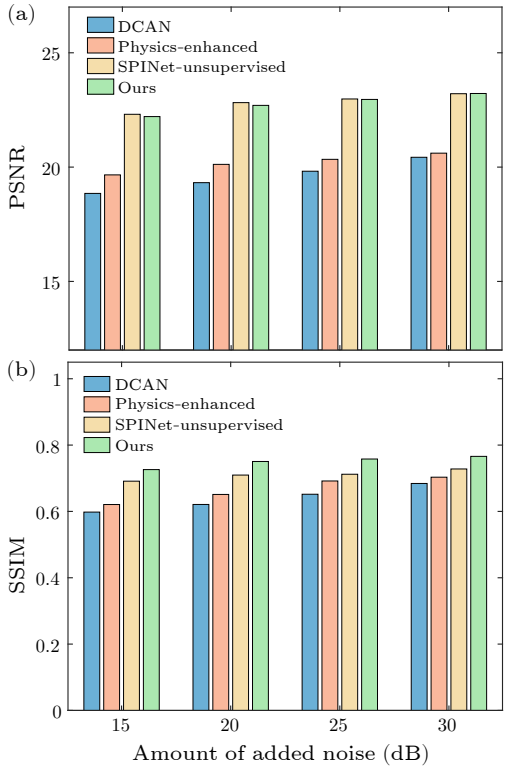

Fig. 4. Comparison with other DL methods at 10% sampling rate and different noise levels. (a) PSNR results at various noise levels (b) SSIM values at various noise levels.

Table 3. Average simulation time for 100 simulations.

Sampling rates	Average simulation time per method (ms)			
	DCAN	Physics-enhanced	SPINet-unsupervised	Ours
5%	0.204	1.07	4.91	0.226
8%	0.225	1.10	4.91	0.239
10%	0.248	1.15	4.92	0.257
20%	0.265	1.23	4.94	0.275

The solid underline represents the fastest time and the dashed underline indicates the second fastest time. The DCAN algorithm exhibited the fastest image reconstruction times across all sampling rates. Our method is ranked second, closely trailing the DCAN algorithm by a margin of less than 2×10^{-5} seconds. The effectiveness of the proposed method in practical applications is evaluated by constructing an SPI test platform.

The system employed an infinitely dimmable constant-current output light emitting diode (LED) light source to illuminate the target. The light reflected from the target was focused through a lens onto a DMD (F4320 SSD, 0.7XGA) for modulation. The modulated light from the DMD was subsequently converged by a lens and eventually detected using a photomultiplier tube. The initial experiment employed printed images of "photographer", "house" and "vegetable" as targets. The experimental setup is shown in Fig. 5, and the reconstructed images using the aforementioned methods at a sampling rate of 10% are shown in Fig. 6.

As shown in Fig. 6, the image reconstructed using the proposed method exhibited smoother contours and more sharply defined features.

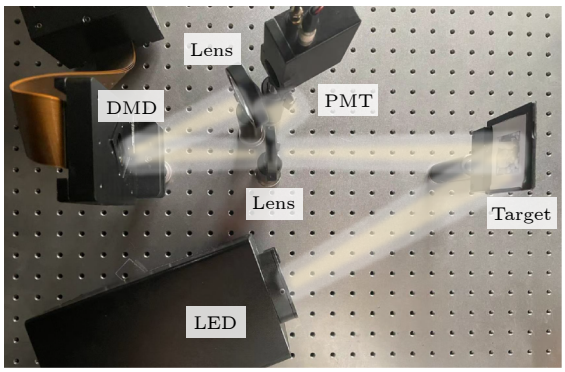

Fig. 5. SPI experimental setup.

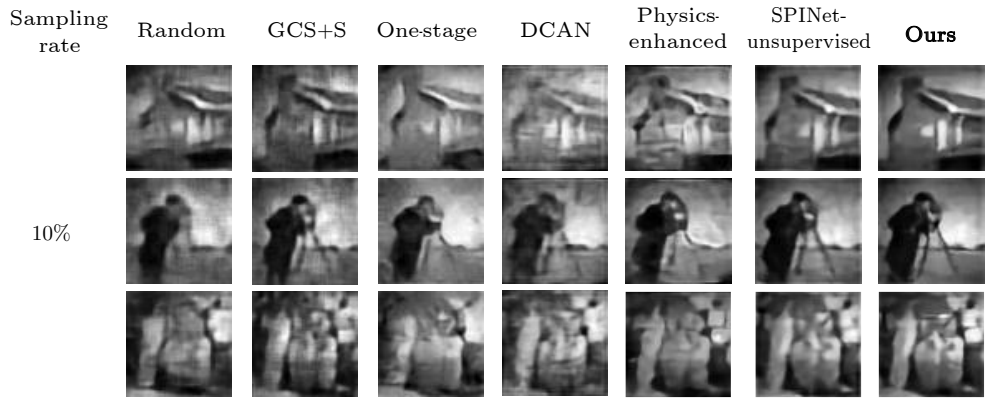

Fig. 6. Comparison of reconstructed results of printed targets at a 10% sampling rate.

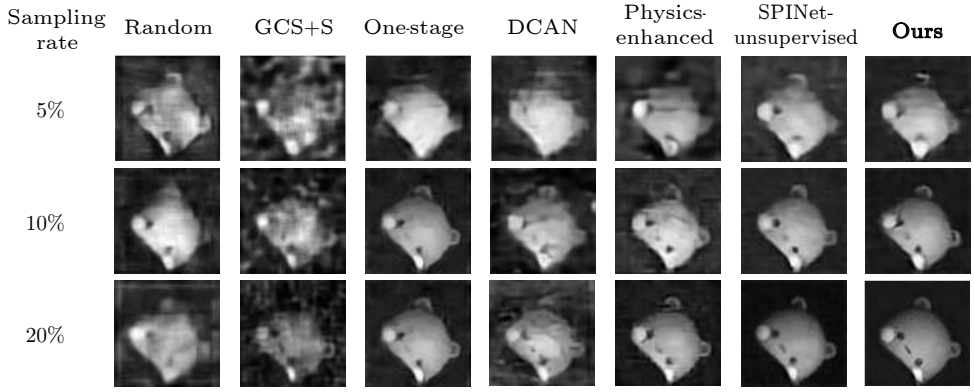

Fig. 7. Comparison of the reconstructed results of real-world targets at sampling rates of 5%, 10%, and 20%.

To further validate the effectiveness of the proposed method, an experiment was performed using a real-world doll as the target, and the reconstructed images are shown in Fig. 7.

Figure 7 shows a comparison of the imaging effects for the doll-bear head object at sampling rates of 5%, 10%, and 20%. It is evident that the proposed method exhibits strong adaptability to real-world scenarios. However, the experiment also indicated that uneven distribution of the light sources and ambient light interference contributed to a decline in image quality.

A two-stage coupled deep learning approach was proposed to train SPI illumination patterns and imaging reconstruction networks, demonstrating the effectiveness of this method for SPI imaging tasks. In the first stage, an L_2 regularization constraint was incorporated into the sensing operator to enhance the generalization ability, ensure convergence, and obtain optimized initial weights of the sensing network. In the second stage, the regularization constraints were transformed into binarization constraints. This SPI network design effectively reduced data redundancy and increased the image reconstruction time. We evaluated the performance of the method on the Set12 dataset. Simulation and experimental results indicated that the proposed approach achieved high perceptual quality, even at very low sampling rates. At a 10% sampling rate, the proposed method achieved an image reconstruction time of 0.257 ms. This efficient and streamlined method not only reduces hardware and software complexities but also offers a reliable approach for practical SPI applications.

References

[1] He Y H, Zhang A X, Yu W K, Chen L M, and Wu L A 2020 *Chin. Phys. Lett.* **37** 044208
[2] Zhu X, Li Y, Zhang Z, and Zhong J 2024 *Opt. Lett.* **49** 1065
[3] Chang S, Cai J, and Gong W 2024 *Opt. Express* **32** 10093
[4] Yue H, Guo J, Yin X, Zhang Y, Wen B, and Li C 2024 *IEEE Trans. Circuits Syst. Video Technol.* **34** 235
[5] Yu W K, Wang S F, and Shang K Q 2024 *Chin. Phys. Lett.* **41** 024201
[6] Liu Y, Yu P, Wu Y, Zhuang J, Wang Z, Li Y, Lai P, Liang J, and Gong L 2023 *Proc. Natl. Acad. Sci. USA* **120** e2304755120
[7] He Z and Dai S 2024 *AIP Advances* **14** 045316
[8] Peng L, Xie S, Qin T, Cao L, and Bian L 2023 *Opt. Lett.* **48** 2527
[9] Wang Y, Huang K, Fang J, Yan M, Wu E, and Zeng H 2023 *Nat. Commun.* **14** 1073
[10] Sun B, Huang K, Ma H, Fang J, Zheng T, Qin R, Chu Y, Guo H, Liang Y, and Zeng H 2024 *Laser & Photonics Reviews* **2024** 2401099
[11] Watts C M, Shrekenhamer D, Montoya J, Lipworth G, Hunt J, Sleasman T, Krishna S, Smith D R, and Padilla W J 2014 *Nature Photonics* **8** 605
[12] Chen X, Li J, Du L, Sun C, Zhai Z, Zhong S, Liu Q, Huang K, and Zhu L G 2024 *Opt. Express* **32** 35126
[13] Valencia A, Scarcelli G, D'Angelo M, and Shih Y 2005 *Phys. Rev. Lett.* **94** 063601

[14] Wang C, Liu X F, Yu W K, Yao X R, Zheng F, Dong Q, Lan R M, Sun Z B, Zhai G J, and Zhao Q 2017 *Chin. Phys. Lett.* **34** 104203
[15] Yu W K 2019 *Sensors* **19** 4122
[16] López-García L, Cruz-Santos W, García-Arellano A, Filio-Aguilar P, Cisneros-Martínez J A, and Ramos-García R 2022 *Opt. Express* **30** 13714
[17] Bacca J, Gelvez-Barrera T, and Arguello H 2021 *IEEE Trans. Comput. Imaging* **7** 1148
[18] Duarte M F, Davenport M A, Takhar D, Laska J N, Sun T, Kelly K F, and Baraniuk R G 2008 *IEEE Signal Process. Mag.* **25** 83
[19] Ferri F, Magatti D, Lugiato L A, and Gatti A 2010 *Phys. Rev. Lett.* **104** 253603
[20] Yang A Y, Sastry S S, Ganesh A, and Ma Y 2010 *17th IEEE International Conference on Image Processing* p. 1849
[21] LeCun Y, Bengio Y, and Hinton G 2015 *Nature* **521** 436
[22] Higham C F, Murray-Smith R, Padgett M J, and Edgar M P 2018 *Sci. Rep.* **8** 2369
[23] He Y, Wang G, Dong G, Zhu S, Chen H, Zhang A, and Xu Z 2018 *Sci. Rep.* **8** 6469
[24] Liansheng S, Liwen Z, Ying C, Zhaolin X, Ailing T, and Anand A 2021 *Opt. Commun.* **492** 126982
[25] Wang F, Wang C, Deng C, Han S, and Situ G 2022 *Photon. Res.* **10** 104
[26] Li Z, Huang J, Shi D, Chen Y, Yuan K, Hu S, and Wang Y 2023 *Optics & Laser Technology* **167** 109710
[27] Tian Y, Fu Y, and Zhang J 2023 *Optics & Laser Technology* **162** 109278
[28] Ren J, Zhang M, Yu C, and Liu Z 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* p. 7916
[29] Rodríguez P 2013 *J. Electr. Comput. Eng.* **2013** 217021
[30] Adler A, Elad M, and Zibulevsky M 2016 arXiv:1610.09615v1 [cs.CV]
[31] Fu H, Bian L, and Zhang J 2020 *Opt. Lett.* **45** 3111
[32] Bacca J, Galvis L, and Arguello H 2020 *Opt. Express* **28** 8528
[33] Kingma D P and Ba J 2017 arXiv:1412.6980v9 [cs.LG]
[34] Channappayya S S, Bovik A C, and Heath R W 2008 *IEEE Trans. Image Process.* **17** 1624