Optics Letters

DNN-FZA camera: a deep learning approach toward broadband FZA lensless imaging

JIACHEN WU,[1] LIANGCAI CAO,[1,3] AND GEORGE BARBASTATHIS[2,4]

[1]State Key Laboratory of Precision Measurement Technology and Instruments, Department of Precision Instruments, Tsinghua University, Beijing 100084, China
[2]Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA
[3]e-mail: clc@tsinghua.edu.cn
[4]e-mail: gbarb@mit.edu

Received 7 October 2020; revised 2 December 2020; accepted 2 December 2020; posted 2 December 2020 (Doc. ID 411228); published 24 December 2020

In mask-based lensless imaging, iterative reconstruction methods based on the geometric optics model produce artifacts and are computationally expensive. We present a prototype of a lensless camera that uses a deep neural network (DNN) to realize rapid reconstruction for Fresnel zone aperture (FZA) imaging. A deep back-projection network (DBPN) is connected behind a U-Net providing an error feedback mechanism, which realizes the self-correction of features to recover the image detail. A diffraction model generates the training data under conditions of broadband incoherent imaging. In the reconstructed results, blur caused by diffraction is shown to have been ameliorated, while the computing time is 2 orders of magnitude faster than the traditional iterative image reconstruction algorithms. This strategy could drastically reduce the design and assembly costs of cameras, paving the way for integration of portable sensors and systems.

https://doi.org/10.1364/OL.411228

A lens-based camera adopts a series of lenses with different materials and thicknesses to focus light and correct aberrations. This architecture comes at the cost of increased system complexity and weight. Lensless imaging systems remove lenses by using alternative optical elements to encode the wavefront. The image is then recovered through computation, ideally incorporating the physical model of propagation. As an ultrathin, lightweight, low-cost, easy-to-build, and high degree of freedom imaging method, various approaches to lensless imaging have been proposed, e.g., light-field imaging [1], diffuse light-mediated three-dimensional (3D) imaging [2], and hyperspectral imaging [3].

Recently, various mask-based lensless cameras have been proposed for visible light imaging, such as the Fresnel zone aperture (FZA) camera [4,5], FlatCam [6], and DiffuserCam [2]. The common challenge in these methods is that the solution of the inverse problem is not unique and is unstable due to noise. To stabilize the solution, image priors are introduced to regularize the inverse problem. The optimal solution no longer has a closed-form, and the computation of the closed-form solution is infeasible based on the regularization term. Thus, optimization algorithms are adopted to search for the optimal solution but tend to be slow to converge; thus, it is difficult for real-time imaging. Moreover, manual adjustment is often required for both the models and the parameters in optimization algorithms, undermining the possibility of automatic operation for everyday applications.

Deep learning technology as a data-driven approach has been extensively applied for pattern recognition in the early days. Owing to the strong ability to model complex problems, deep neural networks (DNNs) have achieved success in computational imaging [7], such as digital holographic reconstruction [8–10], 3D particle field imaging [11,12], phase retrieval [13–15], and scattering imaging [16,17]. Sinha *et al.* successfully adopted a deep learning approach to solve the lensless imaging problem [18,19]. They were able to realize coherent diffraction imaging for a phase-only object. Monakhova *et al.* proposed the alternating direction method of multiplier (Le-ADMM) network to solve mask-based lensless imaging [20]. Sitzmann *et al.* utilized a phase plate with end-to-end optimization to achieve an achromatic extended depth of field [21]. Horisaki *et al.* proposed a method for jointly designing a coded aperture and U-Nets for lensless imaging, which adopted two U-Nets to generate the encoded image and reconstruct the original image despite the presence of strong scatter [22]. The aforementioned networks were successful for various lensless imaging systems in terms of reconstruction quality and speed, which shows a new direction for computational imaging.

In this Letter, we propose a DNN that consists of a U-Net and a deep back-projection network (DBPN) [23] to solve the image reconstruction for the FZA lensless camera, called the DNN-FZA camera, as shown in Fig. 1(a). The error feedback mechanism of the DBPN is effective in improving imaging quality. The straightforward structure of the DNN dispenses with iterations so that it could rapidly construct the underlying image. The strong feature extraction ability of DNNs makes

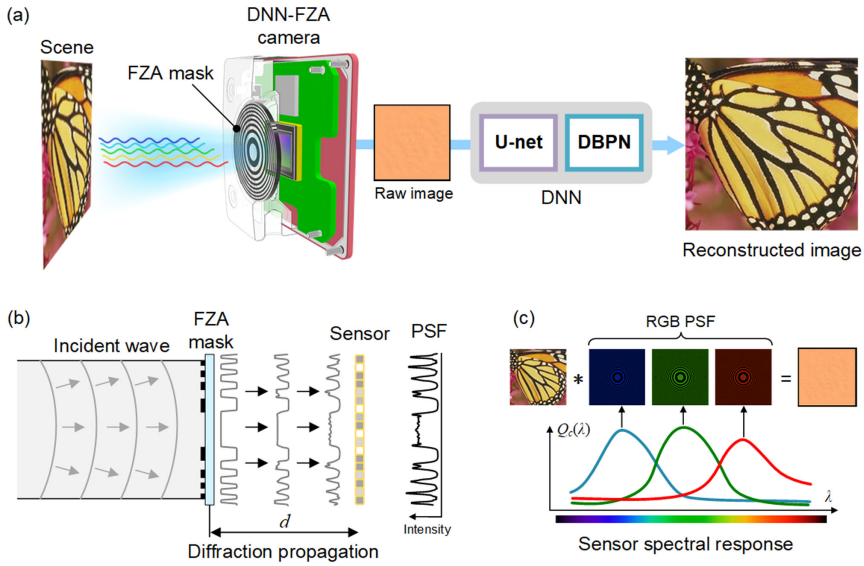

Fig. 1. Architecture of the proposed DNN-FZA camera. (a) Image acquisition pipeline and reconstruction for the DNN-FZA camera. (b) PSF simulation module. A light wave from a point source with a given wavelength λ is incident on the FZA mask and propagates forward to the sensor. (c) For RGB sensor, three channels are calculated, respectively. The training data could be generated by convolving the RGB PSF on the corresponding channel.

our approach naturally robust against noise. A diffraction model is developed to generate the training data, avoiding the tedious collection of real images and interferences from the environment.

The imaging model of a mask-based lensless imaging system is usually characterized by the convolution of the scene and the shadow of the mask, namely the point spread function (PSF). Because the convolution is a linear operation, the scene-to-sensor mapping can be described using a linear equation,

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

where Φ is the measurement matrix constructed by the PSF, x is the scene irradiance, y is the image formed on the sensor, and e is the measurement noise. To reconstruct x from y with a known Φ, the general image recovery approach for lensless imaging is to minimize an objective function that usually consists of a data fidelity term and a regularization term,

$$\underset{x}{\operatorname{argmin}} \|\Phi x - y\|_2^2 + \tau \Psi(x), \tag{2}$$

where $\|\Phi x - y\|_2^2$ quantifies the data fidelity, the regularization term $\Psi(x)$ imposes prior knowledge to mitigate the ill-posedness of the inverse problem, and the regularization parameter τ controls the relative weight of the two terms. We take diffraction into account and build an accurate forward model for the generation of a training dataset to realize broadband lensless imaging. For a single wavelength λ, the PSF is a pattern equivalent to diffraction propagation of the FZA pattern, as shown in Fig. 1(b), and is calculated by the angular spectrum method,

$$U(x, y; \lambda) = \left|\mathcal{F}^{-1}\{\mathcal{F}\{M(x, y)\} \cdot H(\xi, \eta; \lambda, d)\}\right|^2. \tag{3}$$

Here, $M(x, y)$ is the amplitude mask pattern, and $H(\xi, \eta; \lambda, d)$ is the angular spectrum transfer function. For a broadband light source, the PSF may be calculated by integrating the diffracted intensities of multiple wavelengths. Since the image sensor has a different sensitivity to different wavelengths of light, the integration should be weighted by the spectral responsivity $Q_c(\lambda)$,

$$I_c(x, y) = \int Q_c(\lambda) U(x, y; \lambda) \mathrm{d}\lambda. \tag{4}$$

For the red–green–blue (RGB) sensor, each channel has its own spectral responsivity and requires a separate calculation of the PSF, as shown in Fig. 1(c). Though the spectral distribution of the scene is usually unknown, assuming a uniform spectral distribution in the visible band is appropriate since the sensor responsivity largely determines the envelope of spectral response.

Our proposed DNN is a U-Net followed by a DBPN, as illustrated in Fig. 2(a). The DBPN at the end of the network exploits iterative up- and downsampling layers, providing an error feedback mechanism that realizes self-correcting of features, which effectively enhances the resolution of the output image. The size of the input image is twice of the output image so that scattered information can be collected into the network. The input image is normalized between 0 and 1 in each channel to remove the DC component. The encoder part of the U-Net consists of a series of downsampling blocks and residual blocks. The decoder part of the U-Net consists of upsampling and residual blocks, as shown in Fig. 2(b). Each residual block follows ResNet, which is composed of two sets of convolution layer with stride (1,1), a batch normalization (BN) layer, and a rectified linear unit (ReLU) layer stacked one above the other. The downsampling, upsampling, and dilated blocks are similar to the residual block, but replace the convolution layers with stride convolution, transposed convolution, and dilated convolution layers, respectively. Several skip connections connect encoder and decoder by dilated convolution layers. The dilation

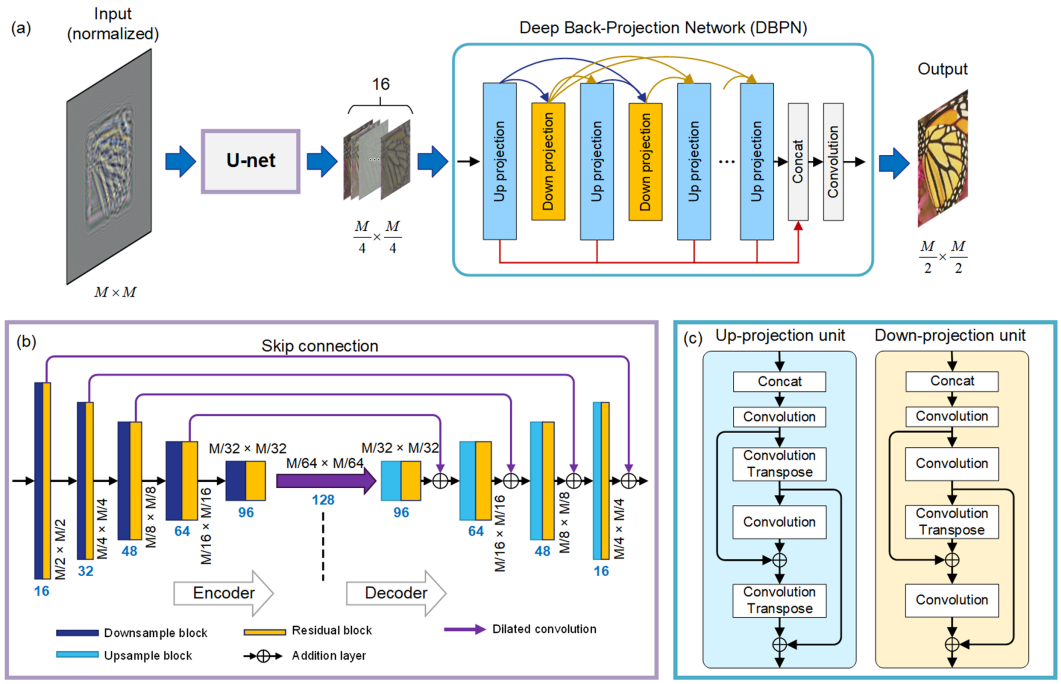

Fig. 2. Architecture of the proposed DNN. (a) The DNN is a U-Net followed by a DBPN. The DBPN provides an error feedback mechanism to realize the self-correction of features. (b) Architecture of the U-Net in (a). The numbers on the right and top of a block such as $M/2 \times M/2$ indicate the size of the feature map. The numbers on the bottom of the block indicate the feature depth. (c) Up- and down-projection unit in DBPN.

factor is adjusted according to the feature map size so that the feature map dimension could be downsized by 2. The up- and down-projection units of DBPN are shown in Fig. 2(c), where all convolution and transposed convolution layers are followed by ReLUs, which are not depicted in the figure.

The choice of loss function is important for any learning task. A generic one, such as the mean square error (MSE), may be poor because it operates pixel-wise and does not generalize well. In our case, the absolute pixel value is secondary; instead, for human visual perception it is the relative value between pixels that is important. Consistent with that observation, as loss function, we chose the negative Pearson correlation coefficient (NPCC), defined as

$$E_{\mathrm{NPCC}}(X, Y) = (-1) \times \frac{\sum_{i}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\left\{\sum_{i}^{n} \left(X_i - \bar{X}\right)^2 \sum_{i}^{n} \left(Y_i - \bar{Y}\right)^2\right\}^{1/2}}. \tag{5}$$

Network training and testing were performed on a workstation with Xeon CPU E5-2650 (2.20 GHz) and 128 GB of RAM, using NVIDIA Quadro GV100 GPU. The network was trained using images from DIV2K [24]. For each image, we cropped the four corners into four images with a size of 1024×1024 augmentation. We started the training with a learning rate of 0.001 and dropped it by multiplying a factor of 0.8 after every epoch. We trained the neural network for 20 epochs, at every epoch shuffling the training samples.

Our experimental apparatus is shown in Fig. 3(a). An LCD monitor was placed ~30 cm from the DNN-FZA camera. The test images displayed on the screen emitted broadband light and were captured by the DNN-FZA camera, which only consisted of a QHY163 CMOS sensor and an FZA mask attached to the sensor. The microscopic image of the FZA mask is shown in Fig. 3(b). The distance between the FZA pattern and sensor plane was 3 mm. Traditional optimization methods such alternating direction method of multipliers (ADMM) [2] have been proposed for inverse problem solving. Here an ADMM with 50 iterations was utilized as a reference. The input image was 2048×2048 pixels for both the ADMM and DNN methods. The reconstructions of the binary, gray scale, and color images are shown in Fig. 3(c). The obscure images in the first row demonstrate that the error caused by diffraction degrades the reconstruction. Using the diffracted PSF calculated using Eq. (4), the reconstructed image quality could be significantly improved, as shown in the second row. According to the resolution analysis in [5], the image resolution is limited to minimum feature size of the PSF. However, shrinking the FZA cannot limitlessly improve imaging resolution. The diffraction PSF reduces the contrast and broadens the feature size. The proposed U-Net+DBPN method achieved almost the same image quality as the ADMM method, as shown in the fourth row, but the average computing time was 0.6 s with GPU acceleration for the DNN, while the ADMM with 50 iterations required 50 s.

The peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) were used to evaluate the reconstructed image quality from the numerical and visual prospect, respectively. The ADMM method was also used for comparison. Varying degrees of Gaussian noise quantified by standard deviation σ were added to the input images to test the robustness of the methods. The regularization coefficient of the ADMM is adjusted to optimal for each noise level. The output images of both methods were normalized to [0,1] before calculating the evaluation index. Figure 3(d) shows that the overall PSNRs of the DNN method are higher than the ADMM

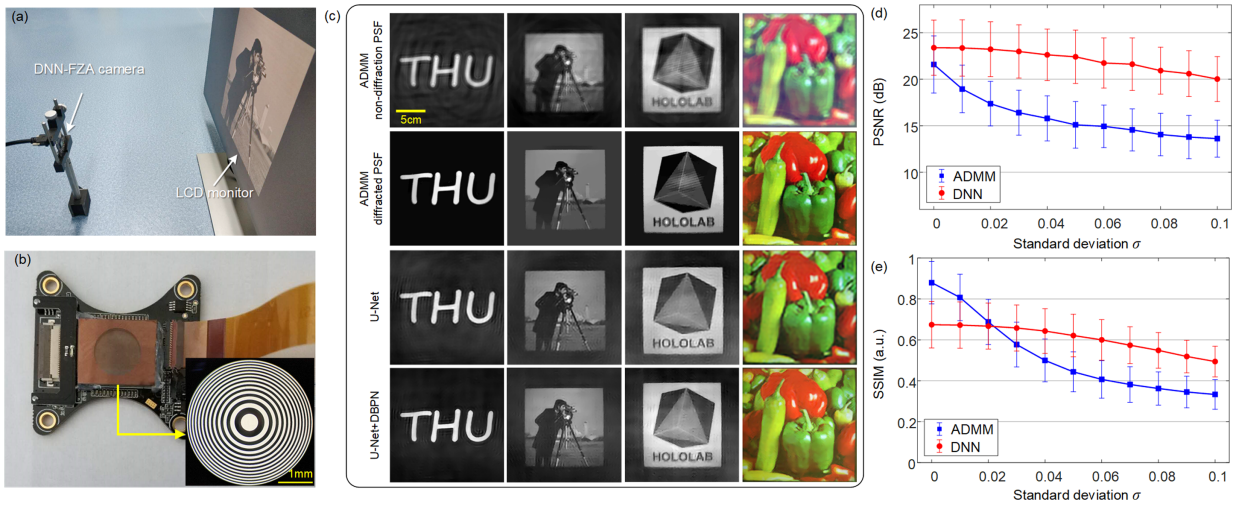

Fig. 3. Experimental setup, results comparison, and robustness analysis. (a) Demonstration of image capturing using the prototyped DNN-FZA camera. (b) Close-up of the DNN-FZA camera with the microscopic image of the FZA mask. (c) Reconstructed results for experimental data by ADMM and the DNN method. U-Net+DBPN method provides nearly the same reconstruction quality with the ADMM method while the computing speed is improved by 2 orders of magnitude. Image quality as a function of the standard deviation of Gaussian noise is evaluated by (d) PSNR and (e) SSIM. The test images were procured from the CSIQ image quality database [25].

method, and Fig. 3(e) shows that only when σ is less than 0.02, the SSIMs of the ADMM method are higher than the DNN method. Since the DNN is good at feature extraction, the noise that does not belong to the image features could be largely filtered out, which makes the DNN method naturally robust against noise.

To summarize, we proposed a DNN-FZA camera architecture that uses a DNN to realize broadband image reconstruction for FZA lensless imaging. The proposed DNN adopts a DBPN that is connected behind the U-Net to improve the network performance for the recovery of fine image details. The proposed imaging model involves a diffraction effect that eliminates the artifacts brought by diffraction. It does not have the ability of hyperspectral imaging, but it is sufficient for high-quality RGB reconstruction for a real scene. This easy-to-build and lightweight architecture has promising applications in miniaturized devices for surveillance and biomedicine.

Funding. National Natural Science Foundation of China (61827825).

Disclosures. The authors declare no conflicts of interest.

REFERENCES

1. Z. Cai, J. Chen, G. Pedrini, W. Osten, X. Liu, and X. Peng, Light Sci. Appl. **9**, 1 (2020).
2. N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, Optica **5**, 1 (2018).
3. K. Monakhova, K. Yanny, N. Aggarwal, and L. Waller, Optica **7**, 1298 (2020).
4. T. Shimano, Y. Nakamura, K. Tajima, M. Sao, and T. Hoshizawa, Appl. Opt. **57**, 2841 (2018).
5. J. Wu, H. Zhang, W. Zhang, G. Jin, L. Cao, and G. Barbastathis, Light Sci. Appl. **9**, 53 (2020).
6. M. S. Asif, A. Ayremlou, A. Sankaranarayanan, A. Veeraraghavan, and R. G. Baraniuk, IEEE Trans. Comput. Imaging **3**, 384 (2017).
7. G. Barbastathis, A. Ozcan, and G. Situ, Optica **6**, 921 (2019).
8. H. Wang, M. Lyu, and G. Situ, Opt. Express **26**, 22603 (2018).
9. K. Wang, J. Dou, Q. Kemao, J. Di, and J. Zhao, Opt. Lett. **44**, 4765 (2019).
10. Z. Ren, Z. Xu, and E. Y. M. Lam, Adv. Photon. **1**, 016004 (2019).
11. T. Shimobaba, T. Takahashi, Y. Yamamoto, Y. Endo, A. Shiraki, T. Nishitsuji, N. Hoshikawa, T. Kakue, and T. Ito, Appl. Opt. **58**, 1900 (2019).
12. S. Shao, K. Mallery, S. S. Kumar, and J. Hong, Opt. Express **28**, 2987 (2020).
13. A. Goy, K. Arthur, S. Li, and G. Barbastathis, Phys. Rev. Lett. **121**, 243902 (2018).
14. C. A. Metzler, P. Schniter, A. Veeraraghavan, and R. G. Baraniuk, "prDeep: robust phase retrieval with a flexible deep network," presented at the Proceedings of Machine Learning Research, Stockholm, Sweden, July 2018.
15. C. Bai, M. Zhou, J. Min, S. Dang, X. Yu, P. Zhang, T. Peng, and B. Yao, Opt. Lett. **44**, 5141 (2019).
16. S. Li, M. Deng, J. Lee, A. Sinha, and G. Barbastathis, Optica **5**, 803 (2018).
17. Y. Li, Y. Xue, and L. Tian, Optica **5**, 1181 (2018).
18. A. Sinha, J. Lee, S. Li, and G. Barbastathis, Optica **4**, 1117 (2017).
19. S. Li and G. Barbastathis, Opt. Express **26**, 29340 (2018).
20. K. Monakhova, J. Yurtsever, G. Kuo, N. Antipa, K. Yanny, and L. Waller, Opt. Express **27**, 28075 (2019).
21. V. Sitzmann, S. Diamond, Y. Peng, X. Dun, S. Boyd, W. Heidrich, F. Heide, and G. Wetzstein, ACM Trans. Graph. **37**, 28075 (2018).
22. R. Horisaki, Y. Okamoto, and J. Tanida, Opt. Lett. **45**, 3131 (2020).
23. M. Haris, G. Shakhnarovich, and N. Ukita, *IEEE Conference on Computer Vision and Pattern Recognition* (2018), p. 1664.
24. E. Agustsson and R. Timofte, *IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2017), p. 126.
25. E. C. Larson and D. M. Chandler, J. Electron. Imaging **19**, 011006 (2010).