Optics Letters

Y-Net: a one-to-two deep learning framework for digital holographic reconstruction

Kaiqiang Wang,[1] Jiazhen Dou,[1] Qian Kemao,[2,3] Jianglei Di,[1,4] and Jianlin Zhao[1,*]

[1]MOE Key Laboratory of Material Physics and Chemistry under Extraordinary Conditions, Shaanxi Key Laboratory of Optical Information Technology, School of Science, Northwestern Polytechnical University, Xi'an 710072, China
[2]School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore
[3]e-mail: mkmqian@ntu.edu.sg
[4]e-mail: jiangleidi@nwpu.edu.cn
*Corresponding author: jlzhao@nwpu.edu.cn

Received 26 July 2019; revised 23 August 2019; accepted 29 August 2019; posted 30 August 2019 (Doc. ID 373772); published 23 September 2019

In this Letter, for the first time, to the best of our knowledge, we propose a digital holographic reconstruction method with a one-to-two deep learning framework (Y-Net). Perfectly fitting the holographic reconstruction process, the Y-Net can simultaneously reconstruct intensity and phase information from a single digital hologram. As a result, this compact network with reduced parameters brings higher performance than typical network variants. The experimental results of the mouse phagocytes demonstrate the advantages of the proposed Y-Net. © 2019 Optical Society of America

https://doi.org/10.1364/OL.44.004765

Human eyes and optical imaging sensors are sensitive to light intensity (amplitude) but cannot directly perceive phase information. Holography is a breakthrough technique for obtaining both intensity and phase of an object [1,2]. Digital holography (DH) records the interference pattern of object and reference waves as a hologram, and reconstructs the intensity and phase from the hologram by numerical methods such as the Fresnel method [3], convolution method [4] and angular spectrum method [5]. As a fast, non-destructive, non-invasive, high-resolution, and full-field method for quantitative amplitude and phase measurement, DH has been an important tool for microscopic imaging [6–8], 3D recognition [9,10], particle measurement [11–13], and flow field measurement [14,15].

Recently, deep learning has shown great potential in holography [16–27]. In addition to works using deep learning as an auxiliary link in the holographic reconstruction process [18–24], there are three so-called end-to-end methods to directly achieve digital holographic reconstruction by deep learning networks [25–27]. Sinha *et al.* pioneered the recovery of phase directly from a single diffraction intensity image by a deep learning network [25]. Wang *et al.* [26] and Ren *et al.* [27] used deep learning networks to reconstruct intensity or phase directly from a hologram. These three deep-learning-based holographic reconstruction methods show interesting possibilities such as no requirement of prior knowledge and shorter operation time and are simpler, faster, and more robust than the traditional methods [25–27].

However, in all three methods, the deep learning networks can be used only to reconstruct either intensity or phase. Thus, for the first time, we propose a one-to-two deep learning frame (Y-Net) to realize simultaneous reconstruction of both intensity and phase from a single hologram. In addition to inheriting the advantages of the above three deep-learning-based methods, the associated intensity and phase reconstruction process also offers the opportunity for higher reconstruction accuracy, shorter training and reconstruction time, and fewer parameters, making DH more convenient and powerful in applications.

The U-Net has been proven very effective in end-to-end image processing tasks [28]. An improved version, also called Y-Net, is to extend the U-Net's down-sampling path for joint segmentation and classification [29]. Our proposed Y-Net is designed as a fusion of two U-Nets, which has a more symmetrical network structure than [29] to serve different purposes. As illustrated in Fig. 1, our Y-Net consists of a down-sampling path (D) on the left side, a bridge path (B) in the middle, and two up-sampling paths (IU for intensity and PU for phase) on the right side. Both the down-sampling and up-sampling paths consist of four repeated stages. Each stage of the down-sampling path extracts three sets of feature maps by using single 3×3 convolution (cyan), double 3×3 convolutions (pink), and triple 3×3 convolutions (purple), respectively, followed by a batch normalization (BN) and a rectified linear unit (ReLU). These three sets of feature maps are concatenated, down-sampled using a 2×2 max pooling operation with stride 2, and then delivered to the next stage. The bridge path is similar to the down-sampling path but does not include the max pooling operation or single or triple 3×3 (cyan and purple) convolutions. The up-sampling path differs from the bridge path in that it has 3×3 up-convolution (transposed convolution) layers to up-sample the feature map. The up-sampled feature map is concatenated with the feature maps from the down-sampling paths by a skip connection. Three concatenated 3×3 (purple) convolutions are equivalent to a 7×7 convolution, and two concatenated 3×3 (pink) convolutions are equivalent to a 5×5 convolution [30]. Thus, the incorporation of single,

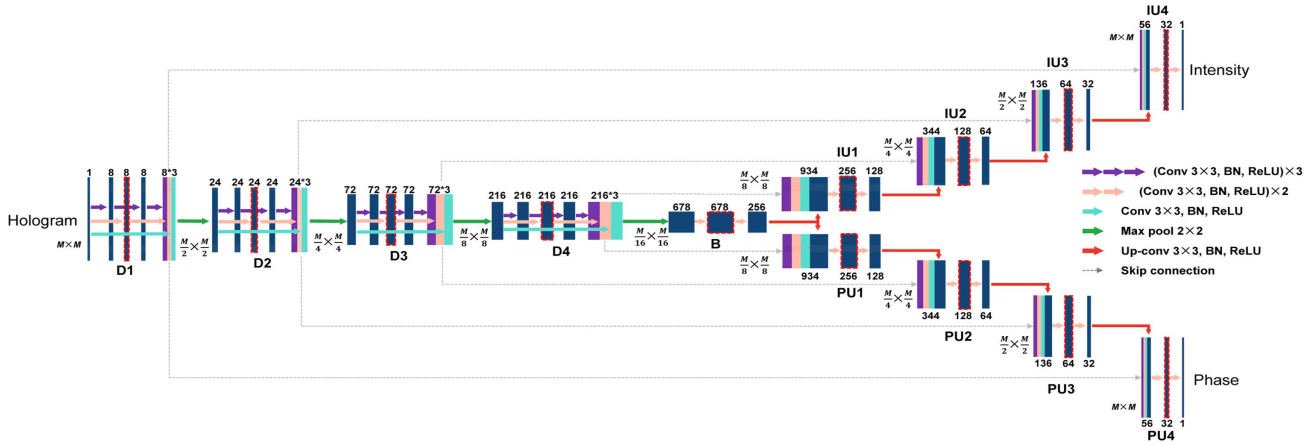

Fig. 1. Architecture of Y-Net. Each block represents a feature map extracted by convolution kernels. The numbers on the left of a block such as $M \times M$ indicate the size of the feature map, while the numbers on the top and bottom of the block indicate the feature depth. The operations are represented by arrows explained at the right. Red-dotted boxes are marked for visualization shown in Fig. 3.

double, and triple 3×3 convolutions into Y-Net is for more effective extraction of the features in different sizes [30] and the line-like features [31] (such as interference fringes in a hologram), and it is believed to be the first time for use in digital holographic reconstruction.

To train the Y-Net, the mean squared errors (MSEs) of intensity and phase are calculated as

$$L_I = \frac{1}{N}\sum_{i}^{N} \|\widehat{I}_i - I_i\|^2, \tag{1}$$

$$L_P = \frac{1}{N}\sum_{i}^{N} \|\widehat{P}_i - P_i\|^2, \tag{2}$$

where, I_i and $\widehat{I}_i$ are the true and reconstructed intensity maps, and P_i and $\widehat{P}_i$ are the true and reconstructed phase maps, respectively; i is the number of the training dataset, and N is the mini-batch size.

To associate intensity and phase, the following weighted sum of the loss functions is back-propagated through the network:

$$L_{IP} = \lambda L_I + L_P, \tag{3}$$

where the weight λ of 0.01 gives a good performance, hinting that the phase is more highlighted than the intensity in training the network, which is consistent with our experience. The Adam optimizer [32] with a learning rate starting from 0.001 and halved after every 10 epochs and the mini-batch size of 64 are adopted to update the network's parameters. To avoid over-fitting of the neural network, the training stops when the network performance on the validation dataset begins to decline.

The dataset is recorded by a typical digital holographic microscope with a 50× objective lens [33]. The specimens are mouse phagocytes maintained in a humidified incubator at 37°C and 5% CO_2 in Dulbecco's modified eagle medium (DMEM) supplemented with 10% heat-inactivated fetal bovine serum and 1% penicillin-streptomycin mixture. After attaching to the bottom of a culture dish, 1664 holograms are recorded by the digital holographic microscope. Then the intensity and phase maps of these cells are reconstructed by the convolution method used as ground truth. All the data we have are partitioned into 80% for training, 10% for validation, and 10% for testing.

After training, the network is randomly given a hologram from the testing dataset and rapidly and simultaneously outputs both intensity and phase maps. The Y-Net is implemented by Pytorch 1.0 based on Python 3.6.1. The network training and testing are performed on a PC with Core i7-8700K CPU (3.8 GHz) and 16 GB of RAM, using NVIDIA GeForce GTX 1080Ti GPU. The training process takes ∼2 h for 1332 pairs of images with the image size of 256×256 pixels and batch size of 64.

The proposed Y-Net is compared with the following three types of networks: (i) a typical U-Net (obtained by removing one up-sampling path, all purple and cyan convolutions from the Y-Net, and tripling the channels of the second pink convolution in each stage of the down-sampling path) with two output channels in the last layer for simultaneous intensity and phase reconstruction (U-Net 1), to show whether such a straightforward solution is adequate; (ii) two typical U-Nets with a single output channel for separate intensity reconstruction (U-Net 2.1) and phase reconstruction (U-Net 2.2), to show whether the association of intensity and phase is feasible; (iii) the same Y-Net but without the purple and cyan convolutions, and tripling the channels of the second pink convolution in each stage of the down-sampling path (Y-Net 1), to show the influence of such purple and cyan convolutions. In order to compare only the performance of the network structure, the hyperparameters of these four networks, including the learning rate, learning epoch, and batch size, are the same as those of the Y-Net.

We test all five networks by inputting four example holograms of different cells from the testing dataset, with the results shown in Fig. 2. The poor results of U-Net 1 show that it is not feasible to use the typical U-type network to learn the reconstruction of intensity and phase at the same time. On the other hand, all other networks perform very well. However, from the magnified view in cyan and orange boxes, the reconstruction results of the Y-Net are visually closer to the ground truth because of the addition of the purple and cyan convolution kernels.

To quantitatively evaluate the recovery accuracy of the networks, the following structural similarity (SSIM) index [34] is

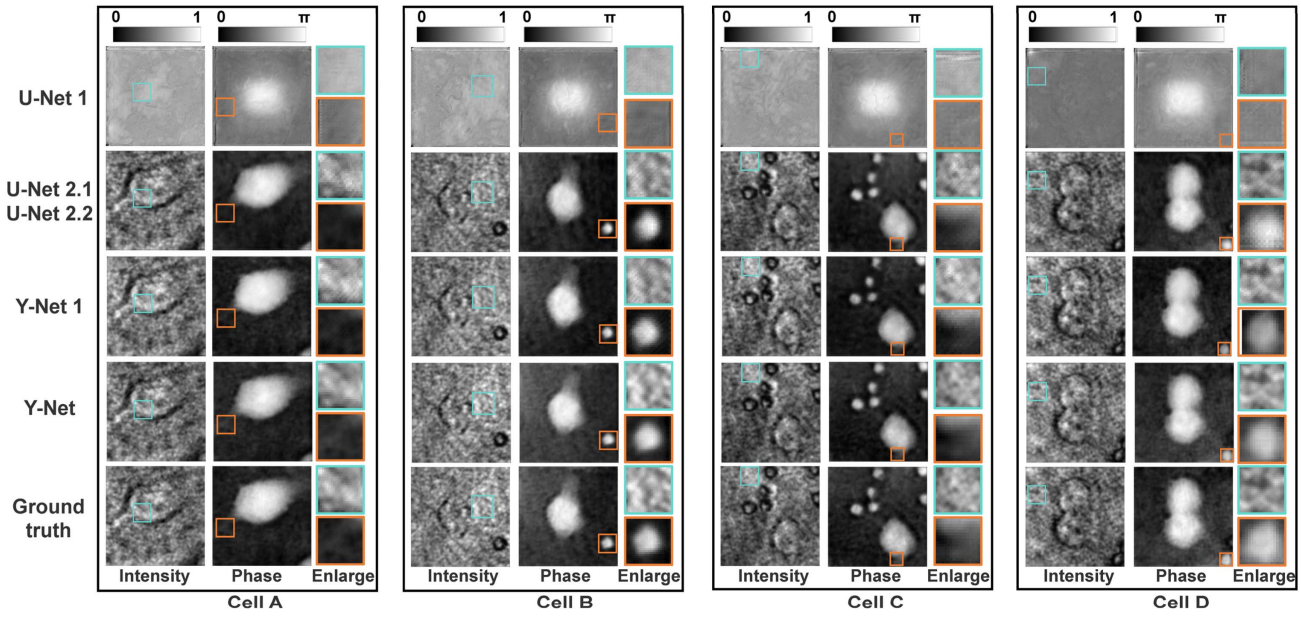

Fig. 2. Reconstructed results of U-Net 1, U-Net 2, Y-Net 1, and Y-Net.

Table 1. SSIM Indices of U-Net 1, U-Net 2, Y-Net 1, and Y-Net

		U-Net 1	U-Net 2.1, 2.2	Y-Net 1	Y-Net
Cell A	Intensity	0.2345	0.9264	0.9191	0.9643
	Phase	0.3034	0.8670	0.8720	0.9290
Cell B	Intensity	0.2143	0.9241	0.9274	0.9693
	Phase	0.3598	0.8166	0.8301	0.9158
Cell C	Intensity	0.2642	0.9377	0.9298	0.9644
	Phase	0.2758	0.8674	0.8661	0.9221
Cell D	Intensity	0.2458	0.9316	0.9288	0.9654
	Phase	0.3836	0.8476	0.8176	0.9343
Mean	**Intensity**	**0.2458**	**0.9243**	**0.9310**	**0.9672**
	Phase	**0.3275**	**0.8507**	**0.8504**	**0.9249**

used to measure the similarity between two arbitrary patterns, pattern 1 and pattern 2:

$$\mathrm{SSIM} = \frac{(2\mu_1\mu_2 + c_1)(2\sigma_{1,2} + c_2)}{(\mu_1^2 + \mu_2^2 + c_1)(\sigma_1^2 + \sigma_2^2 + c_2)}, \tag{4}$$

where μ_1 and μ_2 are their means, σ_1^2 and σ_2^2 are the variances, $\sigma_{1,2}$ is the covariance of two patterns, and $c_1 = (k_1 \times L)^2$ and $c_2 = (k_2 \times L)^2$ are regularization parameters with $L = 255$ (the range of pixel values), $k_1 = 0.01$, and $k_2 = 0.03$ [34]. In our case, two patterns are either $\{\widehat{I}, I\}$ or $\{\widehat{P}, P\}$. The SSIM indices for the results in Fig. 2 and the mean SSIM indices for the testing dataset are summarized in Table 1. The reconstruction results of the U-Net 1 are very poor, as the network does not converge. That is, the typical U-Net with two channels of outputs cannot learn the intensity and phase within a single up-sampling path. The SSIM indices of the U-Net 2

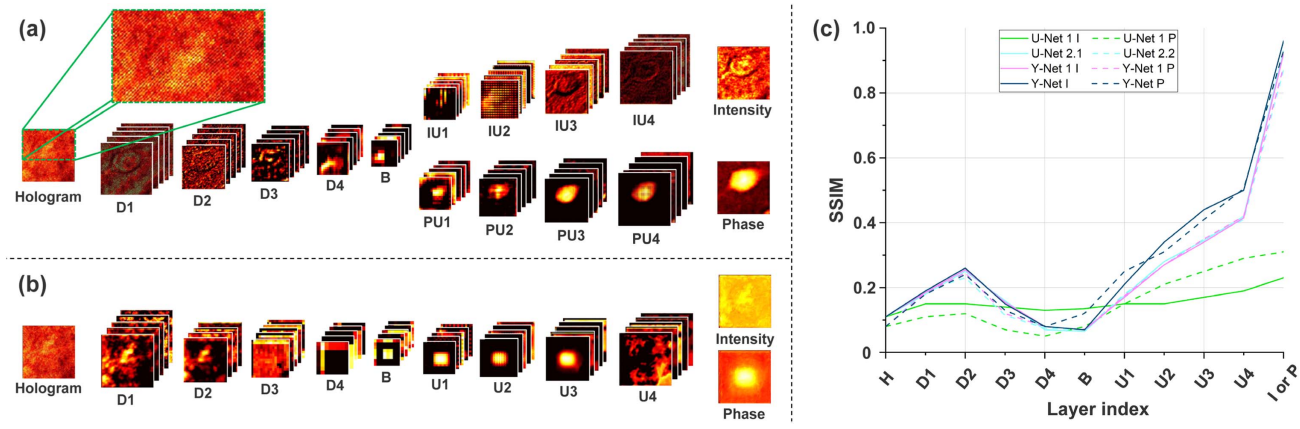

Fig. 3. Visualization of the Y-Net and U-Net 1 by inputting the hologram of cell A in Fig. 2. The number of channels in each layer is defined by the depths of each layer marked by the red-dotted boxes in Fig. 1. (a) Intermediate activation maps of the Y-Net. (b) Intermediate activation maps of the U-Net 1. (c) SSIM indices of the intermediate activation maps of the five networks with the ground truth (standard intensity and phase maps of the cell).

(2.1 and 2.2) and Y-Net 1 are almost equal, which verifies that the association of intensity and phase is feasible. Compared with the Y-Net 1, the Y-Net with addition of the purple and cyan convolution kernels has a higher-level reconstruction performance, about 4% for intensity and 8% for phase.

To further explore the internal mechanism of all five networks in the reconstruction process, we partly visualize their intermediate activation maps by inputting the hologram of cell A in Fig. 3. The channels are randomly selected from the layers marked by the red-dotted boxes in Fig. 1 and the corresponding positions of other four networks. The intermediate activation maps D1–D4, B, and U1–U4 (IU1–IU4 or PU1–PU4) belong to down-sampling, bridge, and up-sampling paths, respectively.

For the Y-Net, as shown in Fig. 3(a), we cannot see obvious feature information of the cell from its hologram, i.e., the intensity and phase information of the cell is implicitly hidden in the fringe pattern. After flowing into the network, D1 and D2 extract a distinct shape of the cell, and D3, D4, and B convert the feature information into a higher-order form. After further flowing from IU1 and PU1 to IU4 and PU4, the intensity and phase information is gradually reconstructed, respectively. The same phenomenon exists in the activation maps of the U-Net 2 and Y-Net 1. However, for the U-Net 1 shown in Fig. 3(b), the activation maps of all paths seem not informative or effective. More quantitatively, SSIM indices of the intermediate activation maps of all five networks with the ground truth (standard intensity and phase maps of cell A) are shown in Fig. 3(c). Consistent with Fig. 3(a), the SSIM indices of U-Net 2, Y-Net 1, and Y-Net increase in D1 and D2, decrease in D3, D4, and B, and then continuously increase from IU1 (PU1) to IU4 (PU4), while those of U-Net 1 remain at a low level throughout all activation maps.

Since the Y-Net saves a down-sampling path and a bridge path, it has 16,029,468 (16M) parameters, compared with 22,790,520 (23M) parameters in U-Net 2, leading to a reduction of 30%, which means that the Y-Net not only combines the functions of intensity and phase reconstruction, but also consumes less time in network training and holographic reconstruction, compared with using two typical U-Nets separately.

To summarize, in this Letter, we propose a unique one-to-two Y-Net specially designed for simultaneous reconstruction of intensity and phase information from a single digital hologram. This new method achieves excellent reconstruction results, with SSIM of over 0.96 for intensity and over 0.91 for phase, which defeats other network variants such as a typical U-Net with two output channels (U-Net 1), two typical U-Nets for intensity and phase reconstruction, respectively (U-Net 2), and a Y-Net without single or triple 3×3 convolutions (Y-Net 1). The number of parameters of Y-Net is lighter than the U-Net 2 (2.1 and 2.2) by 30%. Quantitatively, even with fewer parameters, the Y-Net is 4% and 8% better than the U-Net 2 in intensity and phase reconstruction, respectively. Finally, we believe that our one-to-two Y-Net has great potential in any physical processes where two physical quantities are extracted from one set of captured images, e.g., enabling fringe projection imaging to obtain the intensity distribution of an object while reconstructing its three-dimensional surface distribution.

Funding. National Natural Science Foundation of China (NSFC) (61927810); NSAF Joint Fund (U1730137); Fundamental Research Funds for the Central Universities (3102019ghxm018).

REFERENCES

1. D. Gabor, Nature **161**, 777 (1948).
2. J. W. Goodman and R. W. Lawrence, Appl. Phys. Lett. **11**, 77 (1967).
3. U. Schnars and W. Jüptner, Appl. Opt. **33**, 179 (1994).
4. U. Schnars and W. Jüptner, Meas. Sci. Technol. **13**, R85 (2002).
5. S. De Nicola, D. Alfieri, G. Pierattini, P. Ferraro, and D. Alfieri, Opt. Express **13**, 9935 (2005).
6. W. S. Haddad, D. Cullen, J. C. Solem, J. W. Longworth, A. Mcpherson, K. Boyer, and C. K. Rhodes, Appl. Opt. **31**, 4973 (1992).
7. B. Kemper and G. V. Bally, Appl. Opt. **47**, A52 (2008).
8. J. Di, Y. Li, M. Xie, J. Zhang, C. Ma, T. Xi, E. Li, and J. Zhao, Appl. Opt. **55**, 7287 (2016).
9. B. Javidi and E. Tajahuerce, Opt. Lett. **25**, 610 (2000).
10. M. DaneshPanah and B. Javidi, Opt. Express **15**, 10761 (2007).
11. S. I. Satake, H. Kanamori, T. Kunugi, K. Sato, T. Ito, and K. Yamamoto, Appl. Opt. **46**, 538 (2007).
12. G. Pan and H. Meng, Appl. Opt. **42**, 827 (2003).
13. W. Yang, A. B. Kostinski, and R. A. Shaw, Opt. Lett. **30**, 1303 (2005).
14. B. Wu, J. Zhao, J. Wang, J. Di, X. Chen, and J. Liu, J. Appl. Phys. **114**, 193103 (2013).
15. T. Xi, J. Di, Y. Li, S. Dai, C. Ma, and J. Zhao, Opt. Express **26**, 28497 (2018).
16. R. Horisaki, R. Takagi, and J. Tanida, Appl. Opt. **57**, 3859 (2018).
17. T. Shimobaba, T. Takahashi, Y. Yamamoto, Y. Endo, A. Shiraki, T. Nishitsuji, N. Hoshikawa, T. Kakue, and T. Ito, Appl. Opt. **58**, 1900 (2019).
18. T. Pitkäaho, A. Manninen, and T. J. Naughton, in *European Conference on Biomedical Optics* (Optical Society of America, 2017), paper 104140K.
19. Y. Rivenson, Y. Zhang, H. Günaydın, D. Teng, and A. Ozcan, Light Sci. Appl. **7**, 17141 (2018).
20. Z. Ren, Z. Xu, and E. Y. Lam, Optica **5**, 337 (2018).
21. G. Zhang, T. Guan, Z. Shen, X. Wang, T. Hu, D. Wang, Y. He, and N. Xie, Opt. Express **26**, 19388 (2018).
22. T. Pitkäaho, A. Manninen, and T. J. Naughton, Appl. Opt. **58**, A202 (2019).
23. K. Wang, Y. Li, K. Qian, J. Di, and J. Zhao, Opt. Express **27**, 15100 (2019).
24. Z. Luo, A. Yurt, R. Stahl, A. Lambrechts, V. Reumers, D. Braeken, and L. Lagae, Opt. Express **27**, 13581 (2019).
25. A. Sinha, J. Lee, S. Li, and G. Barbastathis, Optica **4**, 1117 (2017).
26. H. Wang, M. Lyu, and G. Situ, Opt. Express **26**, 22603 (2018).
27. Z. Ren, Z. Xu, and E. Y. Lam, Adv. Photon. **1**, 016004 (2019).
28. O. Ronneberger, F. Philipp, and B. Thomas, in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Lecture Notes in Computer Science, Vol. **9351** (Springer, 2015), pp. 234–241.
29. S. Mehta, E. Mercan, J. Bartlett, D. Weave, J. G. Elmore, and L. Shapiro, "Y-Net: joint segmentation and classification for diagnosis of breast biopsy images," arXiv: 1806.01313 (2018).
30. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2009), pp. 2818–2826.
31. L. Wang, Y. Li, and S. Wang, "DeepDeblur: fast one-step blurry face images restoration," arXiv: 1711.09515 (2017).
32. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv: 1412.6980 (2014).
33. J. Di, Y. Li, K. Wang, and J. Zhao, IEEE Photon. J. **10**, 6900510 (2018).
34. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, IEEE Trans. Image Process **13**, 600 (2004).

61.150.43.48 {ts '2019-09-19 21:24:56'}