

## Computational ghost imaging using deep learning

Tomoyoshi Shimobaba<sup>a,\*</sup>, Yutaka Endo<sup>b</sup>, Takashi Nishitsuji<sup>c</sup>, Takayuki Takahashi<sup>a</sup>,  
Yuki Nagahama<sup>a</sup>, Satoki Hasegawa<sup>a</sup>, Marie Sano<sup>a</sup>, Ryuji Hirayama<sup>a</sup>, Takashi Kakue<sup>a</sup>,  
Atsushi Shiraki<sup>a</sup>, Tomoyoshi Ito<sup>a</sup>

<sup>a</sup> Chiba University, Graduate School of Engineering, 1-33 Yayoi-cho, Inage-ku, Chiba, 263-8522, Japan

<sup>b</sup> Kanazawa University, Institute of Science and Engineering, Kakuma-machi, Kanazawa, Ishikawa 920-1192, Japan

<sup>c</sup> Mitsubishi Electric Corporation, Information Technology R&D Center, 5-1-1 Ofuna, Kamakura, Kanagawa, 247-8501, Japan

### ARTICLE INFO

#### Keywords:

Computational ghost imaging  
Ghost imaging  
Deep learning

### ABSTRACT

Computational ghost imaging (CGI) is a single-pixel imaging technique that exploits the correlation between known random patterns and the measured intensity of light transmitted (or reflected) by an object. Although CGI can obtain two- or three-dimensional images with a single or a few bucket detectors, the quality of the reconstructed images is reduced by noise due to the reconstruction of images from random patterns. In this study, we improve the quality of CGI images using deep learning. A deep neural network is used to automatically learn the features of noise-contaminated CGI images. After training, the network is able to predict low-noise images from new noise-contaminated CGI images.

© 2017 Elsevier B.V. All rights reserved.

## 1. Introduction

Computational ghost imaging (CGI) [1] has garnered attention in recent years as a promising single-pixel imaging method. In CGI, we project several known random patterns onto the object to be imaged and then use a lens to collect the light transmitted an object or reflected by an object. The light intensities are measured by a bucket detector, such as a photodiode. An image of the object is then created by calculating the correlations between the known random patterns and the measured light intensities. CGI can image objects even in noisy environments.

Originally, CGI only measured the light intensity of objects, but methods have also been devised for measuring its phase [2,3]. The acquisition time for CGI schemes is long as they require a large number of illuminating random patterns to objects. Recently, the situation has been improved by using high-speed random pattern illumination [4,5]. In addition, three-dimensional [6] and multi-spectrum CGI [7] have been developed.

Since random patterns are used to create the object images, the reconstructed images are contaminated by noise. To improve the quality of CGI images, improved correlation calculation methods have been devised, such as differential [8] and normalized CGI [9]. Iterative optimization schemes based on the Gerchberg–Saxton algorithm [10] as well as compressed sensing [7,11] have also been applied to CGI.

In this study, we propose an approach to improve CGI image quality by using deep learning [12] and confirm our technique's effectiveness through simulations. Deep neural networks (DNNs) can learn features for the noisy images reconstructed by CGI schemes automatically. We used a dataset of 15,000 images and their CGI reconstructions to train a network. After training, the network could predict lower-noise images from new noisy CGI images that were not included in the training set. In Section 2, we describe our DNN-based CGI scheme. Section 3 presents the simulation results and demonstrates the effectiveness of the proposed method. Finally, Section 4 presents the conclusions of this study.

## 2. Proposed method

In this section, we first outline the CGI scheme used and then we describe the architecture of the DNN.

### 2.1. Computational ghost imaging

We use a differential CGI [8] scheme because its image quality is superior to that of traditional CGI [1]. The optical setup required for differential CGI is shown in Fig. 1.

\* Corresponding author.

E-mail address: [shimobaba@faculty.chiba-u.jp](mailto:shimobaba@faculty.chiba-u.jp) (T. Shimobaba).

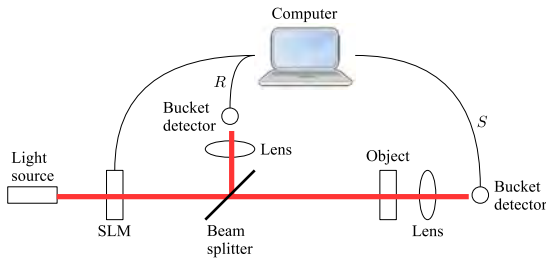


Fig. 1. Optical setup for differential CGI.

In this scheme, a sequence of random patterns is shown on a spatial light modulator (SLM). We denote the  $i$ th random pattern as  $I_i(x, y)$ . The light transmitted by the SLM is divided into two beams by a beam splitter. One beam then irradiated the object to be imaged, and the light transmitted by the object is collected by a lens, and its intensity  $S_i$  is measured by a bucket detector for each  $I_i(x, y)$ . The other beam is immediately focused by a lens, and its intensity  $R_i$  is measured by another bucket detector for each  $I_i(x, y)$ . The final image  $O(x, y)$  that is reconstructed by differential CGI is then calculated as follows:

$$O(x, y) = \langle O_i(x, y) \rangle_N, \tag{1}$$

where  $\langle a_i \rangle_N = \frac{1}{N} \sum_i^N a_i$  denotes the ensemble average over all  $N$  random patterns. The  $O_i(x, y)$  are calculated as follows:

$$O_i(x, y) = \left( \frac{S_i}{R_i} - \frac{\langle S_i \rangle_N}{\langle R_i \rangle_N} \right) (I_i(x, y) - \langle I_i(x, y) \rangle_N). \tag{2}$$

As can be seen from Eq. (2), the reconstructed image is expressed as a superposition of the random patterns; thus, the resulting image is noisy. Fig. 2 shows a series of example images that are reconstructed by differential CGI. The images are arranged from left to right in such a manner that the original image is followed by images that are reconstructed using  $N = 1,000, 2,000, 5,000,$  and  $10,000$  patterns. As the number of random pattern  $N$  increases, the image quality gradually improves. However, it increases the processing and measurement time of differential CGI.

### 2.2. Improving image quality using a deep neural network

In this study, we use a DNN to improve the quality of CGI images. Fig. 3 shows the proposed network structure which is called U-Net [13]. This network was originally used for image segmentation, but it can also be used for image restoration [14].

The network consists of the following two paths: a contracting path and expansive path. These paths include convolution, max-pooling, and up-sampling layers denoted as ‘‘Conv’’, ‘‘MaxPooling’’ and ‘‘UpSampling’’, respectively. The convolution layers generate feature maps for

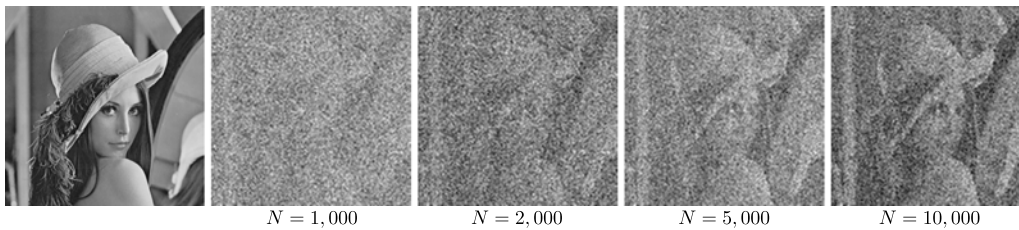


Fig. 2. Example images reconstructed by differential CGI. From left to right, these are the original image that is followed by images reconstructed using  $N = 1,000, 2,000, 5,000,$  and  $10,000$  patterns.

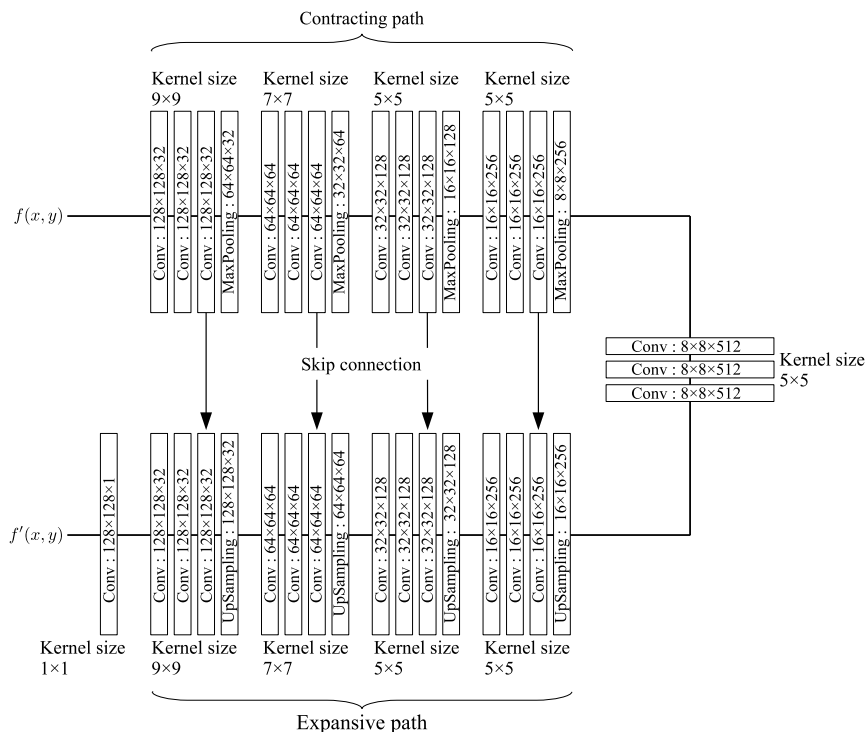


Fig. 3. Our network structure [13]. This network was originally used for image segmentation, but, it can be also used for image restoration [14].

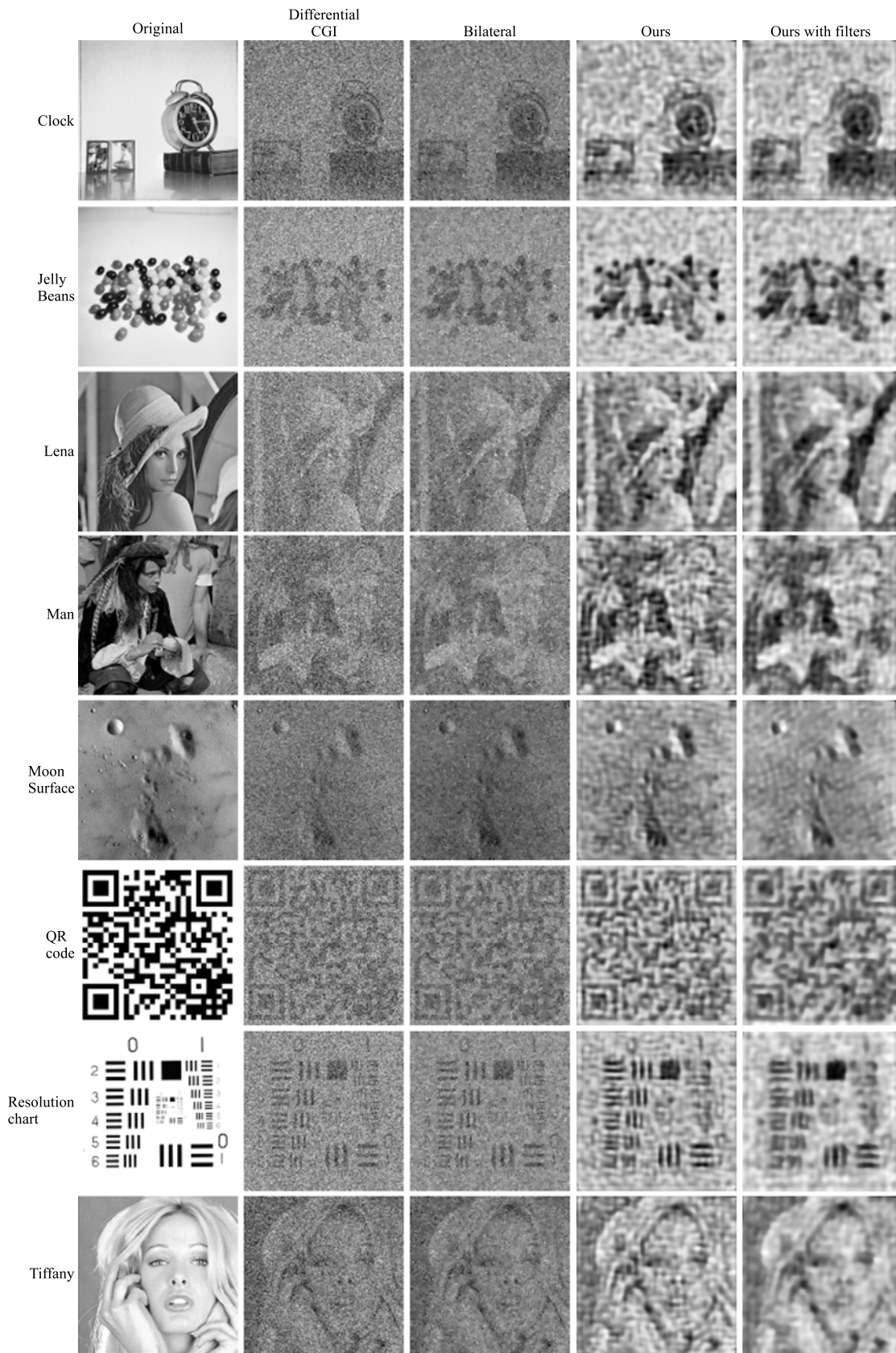
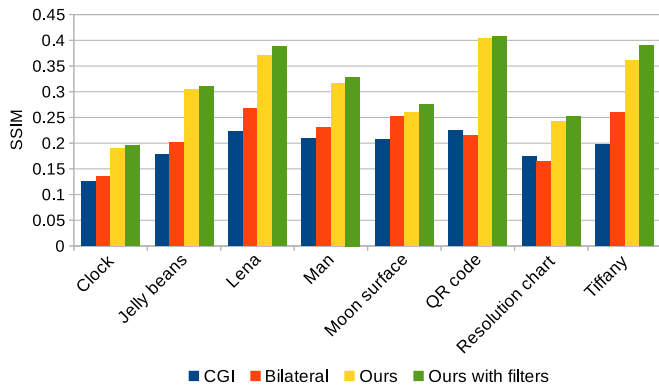


Fig. 4. Comparison of the proposed methods and other methods for eight new images.

**Table 1**  
Average SSIMs for all methods.

	CGI	Bilateral	Ours	Ours with filters
Averaged SSIM	0.19	0.22	0.31	0.32



**Fig. 5.** Comparison of the SSIM values for the proposed methods and the other methods for the eight new images.

the input images using convolution operations, which are frequently used in image processing. For example, the first convolution layer is denoted as “ $128 \times 128 \times 32$ ”, which means that it outputs 32 feature maps that have an output size of  $128 \times 128$  pixels each. The first three convolution layers have convolution filters with a kernel size of  $9 \times 9$ , and the kernel weights are learned from the training dataset. Generally, convolution reduces the number of pixels in the output (feature map). To avoid reducing the number of pixels, we use zero padding in each convolution layer.

The output of each convolution layer except the last uses a ReLU activation function. The last convolution layer, which has a kernel size of  $1 \times 1$ , outputs a predicted image that is the same size as the input image and uses a sigmoid activation function.

The max-pooling layers down-sample the input data to reduce the influence of changes in position and size. We used a down-sample rate of  $2 \times 2$ ; this means that for input data of size  $M \times M$ , the output size is  $M/2 \times M/2$ . The up-sampling layers then up-sample the input data

again at a rate of  $2 \times 2$ ; this means that for input data of size  $M/2 \times M/2$ , the output size is  $M \times M$ .

Although the max-pooling layers are important for robustness against changes in position and size in the input images, in the deeper max-pooling layers much of the input data resolution has been lost, this has resulted in them behaving like low-pass filters. To address this drawback, skip connections have been added to the network to forward the feature maps generated by the contracting path directly to the expansive path.

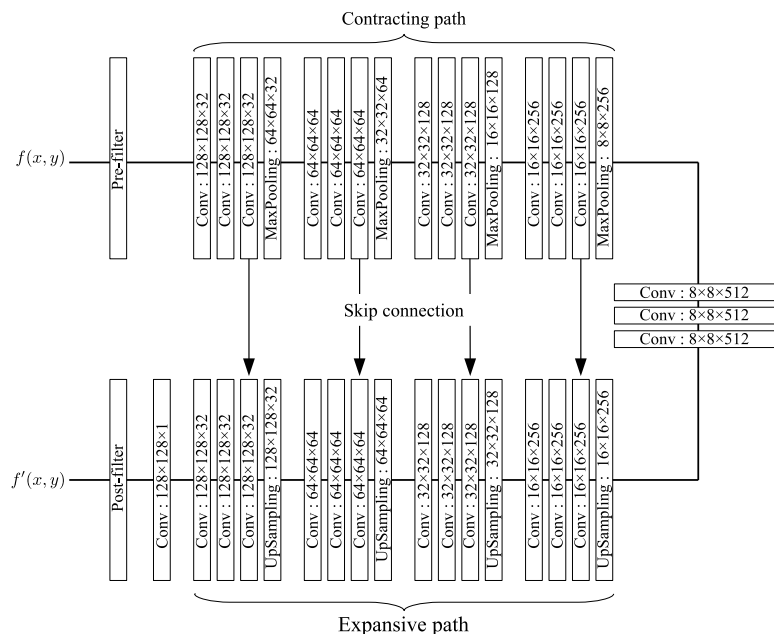
To optimize the kernel weights and other network parameters, the network is trained by minimizing the mean squared error (MSE) between the noisy images  $f'(x, y)$  that are reconstructed by differential CGI and the original images  $f(x, y)$ . The reconstructed images are calculated using Eq. (1). We used Adam optimizer [15] to minimize the MSE using stochastic gradient descent (SGD). In SGD, part of datasets is randomly selected. The size  $B$  of the partial dataset is referred to as the batch size, which was 50 in this study. The number of epochs, i.e., the number of iterations used to optimize the network parameters, was 3. In addition, we used the Dropout technique [16] to prevent over-fitting in the network. Dropout randomly disables  $d$  percent of the units in a layer during the training process, and we used  $d = 80\%$ .

### 3. Results

To train the network, we needed to prepare a large dataset comprising pairs of original and reconstructed images. Here, we used the Caltech-256 [17] dataset, which includes  $\sim 30,000$  general images with different resolutions. In this study, the objects measured by differential CGI are  $128 \times 128$  pixels in size; therefore, we randomly selected 15,000 images from Caltech-256 and resized them to  $128 \times 128$  pixels. These images were then reconstructed by differential CGI, using Eq. (1) with 5,000 random  $128 \times 128$ -pixel patterns. The reconstructed images were generated by incoherent light simulation using our numerical optics library [18].

Fig. 4 shows example images that are produced using the proposed method that is based on eight original images that were not included in the training dataset. We compared the proposed method with a bilateral filter [19] to demonstrate its effectiveness. Bilateral filters are a well-known type of noise-reduction filter that preserves edges.

The first column shows the eight original images, while the second shows the images reconstructed by differential CGI, the third shows



**Fig. 6.** DNN with additional pre- and post-processing filters.

the results obtained by the bilateral filter, and the fourth shows results produced by the proposed method. Subjectively, the images obtained by the proposed method show improved noise and contrast compared with the images from the other methods.

The image quality improvement achieved by the proposed method was evaluated in terms of the structural similarity (SSIM) index [20]. Fig. 5 compares the SSIM values for the proposed method with those for the other methods. SSIM values can evaluate image quality more accurately than the peak-signal-noise ratio (PSNR), and larger SSIMs indicate better image quality. Overall, the SSIMs for the proposed method were better than those for the other methods for all the original images.

To improve the proposed method further, we also added pre- and post-processing filters to the DNN as shown in Fig. 6, using a bilateral filter in both cases. The bilateral filters reduced the noise in the training dataset, thus allowing the DNN to learn the filtered dataset. The reconstructed images and SSIM values for the proposed method with filters are also shown in Figs. 4 and 5, and these results show a small additional improvement while using the filters. For example, the filters increase the SSIM for the “Tiffany” image from 0.36 to 0.39. Table 1 shows the average SSIM values for the different methods that are averaged over the eight images. This again shows that our proposed methods can produce better results than the other methods. The prediction time of the DNN running on a graphics processing unit took only 5 ms.

#### 4. Conclusions

In this study, we have proposed using a DNN to improve the quality of images produced with CGI and have presented results from simulations where a DNN was trained using a dataset of 15,000 images. We compared the images reconstructed by the proposed method with those obtained by differential CGI and bilateral denoising. While testing with eight images that were not included in the training dataset, the average SSIM of the proposed method was over 0.3 compared with only around 0.2 for the differential CGI and bilateral denoising methods. In our next study, we will improve the structure of the DNN to further improve the image quality.

#### Acknowledgment

This work was partially supported by JSPS KAKENHI, Grant Number 16K00151.

#### References

- [1] J.H. Shapiro, Computational ghost imaging, *Phys. Rev. A* 78 (6) (2008) 061802.
- [2] T. Shirai, T. Setälä, A.T. Friberg, Ghost imaging of phase objects with classical incoherent light, *Phys. Rev. A* 84 (4) (2011) 041801.
- [3] P. Clemente, V. Durán, E. Tajahuerce, J. Lancis, Single-pixel digital ghost holography, *Phys. Rev. A* 86 (4) (2012) 041803.
- [4] M.P. Edgar, G.M. Gibson, R.W. Bowman, B. Sun, N. Radwell, K.J. Mitchell, S.S. Welsh, M.J. Padgett, Simultaneous real-time visible and infrared video with single-pixel detectors, *Sci. Rep.* 5 (2015) srepl0669.
- [5] Y. Wang, Y. Liu, J. Suo, G. Situ, C. Qiao, Q. Dai, High speed computational ghost imaging via spatial sweeping, *Sci. Rep.* 7 (2017) 45325.
- [6] B. Sun, M.P. Edgar, R. Bowman, L.E. Vittert, S. Welsh, A. Bowman, M.J. Padgett, 3D computational imaging with single-pixel detectors, *Science* 340 (6134) (2013) 844–847.
- [7] S.S. Welsh, M.P. Edgar, R. Bowman, P. Jonathan, B. Sun, M.J. Padgett, Fast full-color computational imaging with single-pixel detectors, *Opt. Express* 21 (20) (2013) 23068–23074.
- [8] F. Ferri, D. Magatti, L.A. Lugiato, A. Gatti, Differential ghost imaging, *Phys. Rev. Lett.* 104 (25) (2010) 253603.
- [9] B. Sun, S.S. Welsh, M.P. Edgar, J.H. Shapiro, M.J. Padgett, Normalized ghost imaging, *Opt. Express* 20 (15) (2012) 16892–16901.
- [10] W. Wang, X. Hu, J. Liu, S. Zhang, J. Suo, G. Situ, Gerchberg-saxton-like ghost imaging, *Opt. Express* 23 (22) (2015) 28416–28422.
- [11] C. Zhao, W. Gong, M. Chen, E. Li, H. Wang, W. Xu, S. Han, Ghost imaging lidar via sparsity constraints, *Appl. Phys. Lett.* 101 (14) (2012) 141123.
- [12] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016. <http://www.deeplearningbook.org>.
- [13] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2015, pp. 234–241.
- [14] K.H. Jin, M.T. McCann, E. Froustey, M. Unser, Deep convolutional neural network for inverse problems in imaging, *IEEE Trans. Image Process.* 26 (9) (2017) 4509–4522.
- [15] D. Kingma, J. Ba, Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [16] N. Srivastava, G.E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, *J. Mach. Learn. Res.* 15 (1) (2014) 1929–1958.
- [17] G. Griffin, A. Holub, P. Perona, Caltech-256 object category dataset. 2007.
- [18] T. Shimobaba, J. Weng, T. Sakurai, N. Okada, T. Nishitsuji, N. Takada, A. Shiraki, N. Masuda, T. Ito, Computational wave optics library for C++: Cwo++ library, *Comput. Phys. Comm.* 183 (5) (2012) 1124–1138.
- [19] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: *Computer Vision, 1998 Sixth International Conference on*, IEEE, 1998, pp. 839–846.
- [20] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: From error visibility to structural similarity, *IEEE Trans. Image Process.* 13 (4) (2004) 600–612.