Analyzing OAM mode purity in optical fibers with CNN-based deep learning

Tianying Lin (林天营)[1,2], Ang Liu (刘 昂)[1,2], Xiaopei Zhang (张笑裴)[1,2], He Li (李 鹤)[1,2], Liping Wang (汪利平)[1,2], Hailong Han (韩海龙)[1,2], Ze Chen (陈 泽)[1,2], Xiaoping Liu (刘晓平)[1,2,*], and Haibin Lü (吕海斌)[1,2,]**

[1]*National Laboratory of Solid State Microstructures and College of Engineering and Applied Sciences, Nanjing University, Nanjing 210093, China*

[2]*Collaborative Innovation Center of Advanced Microstructures, Nanjing University, Nanjing 210093, China*

**Corresponding author: xpliu@nju.edu.cn; **corresponding author: lvhaibin203@163.com*

Received March 22, 2019; accepted August 22, 2019; posted online September 16, 2019

Inspired by recent rapid deep learning development, we present a convolutional-neural-network (CNN)-based algorithm to predict orbital angular momentum (OAM) mode purity in optical fibers using far-field patterns. It is found that this image-processing-based technique has an excellent ability in predicting the OAM mode purity, potentially eliminating the need of using bulk optic devices to project light into different polarization states in traditional methods. The excellent performance of our algorithm can be characterized by a prediction accuracy of 99.8% and correlation coefficient of 0.99994. Furthermore, the robustness of this technique against different sizes of testing sets and different phases between different fiber modes is also verified. Hence, such a technique has a great potential in simplifying the measuring process of OAM purity.

OCIS codes: 060.2310, 060.2330, 060.2350.

doi: 10.3788/COL201917.100603.

To increase information capacity of an optical fiber communication channel, the amplitude, phase, and wavelength of light are widely used in multiplexing methods[1–3]. Lately, due to mutual orthogonality and infinite topological charges of orbital angular momentum (OAM) modes, OAM has attracted much attention and become a widely investigated dimension for increasing the information capacity[4–7]. OAM generation has been realized via a variety of ways such as grating fibers[8–12], metasurfaces[13,14], and holograms[15]. Based on this, the data transmission employing OAM multiplexing has also been realized[4,6,7]. However, in the detection side, quantitative measurements typically require complicated or custom apparatuses for OAM demultiplexing, such as holograms[16], Shack–Hartmann wavefront sensors[17], or Dove prism arrays[18]. Among them, it is worth mentioning that in 2012, Ramachandran *et al.* presented a technique to measure the OAM mode purity in two-mode fibers[19]. This method uses the vortex basis set to analyze the OAM modes and requires a complicated set of bulk optic devices. In 2017, Ren *et al.* proposed a scalar intensity analysis method (SIAM) to determine the purity of OAM modes in optical fibers[20]. This method uses the amplitude and phase spectrum obtained from the filtered electric-field intensity to analyze OAM modes, which has made great progress and could be implemented without additional apparatuses in certain experiments and applications. However, this method requires modulating the polarizer in order to project the OAM beams and may be ineffective when the purity of the OAM mode is low. Besides, there is a growing interest in using mode-sorter-based OAM mode (de)multiplexing techniques[18,21] to simplify the measurement process and improve the measurement accuracy. It should be pointed out that these previous OAM detection methods typically suffer from unwanted additional losses, require precise alignment of bulk optic devices, and even fail in the case of low OAM purity.

Recently, deep learning, as an emerging powerful interdisciplinary science field, has attracted tremendous attention. It has key fundamental differences from other techniques[22], such as a support vector machine (SVM)[23], k-means[24], and random forest[25,26], which all require prior knowledge to design a future extractor in order to transform raw data into appropriate representations. Deep learning is more intelligent and requires no substantial domain experiences and engineering skills. It has an intrinsic ability to complete feature extraction and self-learning from raw data[22,27]. Owing to this specific advantage, it has been extensively employed in many applications such as microscopy[28,29], laser machining[30], and hologram[31]. It is worth noting that deep learning has also gained considerable popularity in OAM analysis[32–35], due to the overwhelming advantages over traditional methods[32–35], including (1) largely simplified experimental measuring setup, and (2) possibly very high precision predication results at extremely fast speeds. However, so far, previous studies have mainly focused on OAM mode classification rather than quantitative prediction of OAM mode purity, which can reveal more detailed information about the optical transmission channel and is necessary for characterizing the (de)multiplexing devices.

In this Letter, we propose a convolutional-neural-network (CNN)-based deep learning algorithm to predict the OAM purity in optical fibers. Due to the ability of correlating the far-field diffraction intensity patterns of a superposition of multiple fiber modes with its modal power distribution, the trained CNN is capable of predicting the OAM purity with very high accuracy. Furthermore, this technique's generalization is demonstrated by enlarging the testing set space, while the robustness against the phase noise is verified by adding random phase factors to each fiber mode.

To demonstrate the ability of the CNN-based deep learning algorithm in predicting the OAM mode purity in optical fibers, we take a few-mode step-index fiber (FMF) as an example [see Fig. 1(a)]. The radius of the core and cladding is assumed to be 6 μm and 62.5 μm, respectively, while the refractive index is set to be 1.449 and 1.444, accordingly. This fiber can support six modes in total at a 1.55 μm wavelength, including two degenerate fundamental modes HE_{11}^x, HE_{11}^y and four high-order modes TE_{01}, TM_{01}, HE_{21}^{even}, HE_{21}^{odd}. The latter two modes are degenerate, and they are commonly used to construct OAM modes by combining them with an equal amplitude and a $\pi/2$ phase difference, i.e., $OAM_{\pm1}^{\pm} = HE_{21}^{even} \pm iHE_{21}^{odd}$[10], where the superscript of $OAM_{\pm1}^{\pm}$ denotes the spin angular momentum corresponding to the circular polarization states of optical fiber modes, and the subscript denotes the OAM corresponding to the helicity of the transverse wave front. Notice that TE_{01} and TM_{01} can in principle also be used to construct OAM modes, but the resulted OAM modes are unstable due to the non-degenerate nature of the TE_{01} and TM_{01} modes and the resulted phase walk-off[8,36]. Hence, these unstable OAM modes are of no particular interest here and are not considered. A simple fiber system, which utilizes co-propagation coupling to excite OAM modes, can generate only an OAM mode with one particular helicity at a time[8–10] because of the phase-matching condition in the coupling scheme. In this regard, it is safe to assume that only the purity of one OAM mode in the output field needs to be measured or predicted. As shown in Fig. 1(b), we pick OAM_{+1}^{+} mode as our object of study and mix it with other four modes (HE_{11}^x, HE_{11}^y, TE_{01}, TM_{01}). Thus, the total

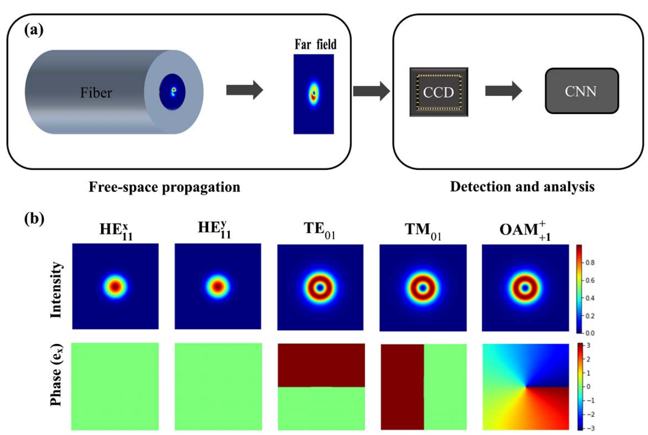

Fig. 1. (a) Schematic with a superimposed mode at the exit of the optical fiber and the corresponding far-field pattern that can be easily recorded by an imaging device (e.g., CCD) and analyzed by CNN. (b) Mode profiles including the electric intensity profiles and phase profiles of the x-direction field component for the modes of interest in this work.

electric field in the output of fiber is $\boldsymbol{E}(x,y) = \sum_i a_i \boldsymbol{e}_i(x,y)$, $i = 1, 2, \ldots, 5$, where a_i and $\boldsymbol{e}_i(x,y)$ are the normalized mode-field amplitudes and corresponding mode-field profiles, respectively. They satisfy $\sum_i |a_i|^2 = 1$ and $\iint \boldsymbol{e}_i(x,y) \cdot \boldsymbol{e}_j(x,y) \mathrm{d}x\mathrm{d}y = \delta_{ij}$[37]. In the paraxial approximation, the far-field intensity distribution can be calculated as $I(x', y') \propto |\boldsymbol{E}_{ff}(x', y')|^2$, where $\boldsymbol{E}_{ff}(x', y') \propto \iint \sum_i a_i \boldsymbol{e}_i(x,y) \exp[-j2\pi(x'x + y'y)/\lambda z] \mathrm{d}x\mathrm{d}y$[38]. In this work, we numerically generate far-field intensity patterns using the above far-field diffraction model with a random sampling approach for the modal amplitudes. These far-field intensity patterns are pixelated into images with dimensions of [108, 108, 1], where 108, 108, and 1 represent the number of pixels along the vertical direction and horizontal direction and the number of color channels, respectively. In total, 6000 groups of $[a_1, a_2, a_3, a_4, a_5]$, subject to $\sum_i |a_i|^2 = 1$, are randomly generated without considering the phase differences between them, and the purity of a mode can be herein defined by the percentage of the corresponding modal power; in particular, $|a_5|^2$ is regarded as the OAM mode purity, which is used as labels in training and testing. Then, 5000 samples are randomly picked as the training set, and the remaining 1000 samples are used as the testing set to validate the reliability of the trained neural network.

CNN is a type of artificial neural network, applicable for any grid-like data such as videos, skeleton animations, and images[22,27,39–41]. Compared to other neural networks, CNN has the characteristic of translation invariance, owing to its sparse-connected and parameter-shared structure. CNN generally consists of three stages: convolutional stage, detection stage, and pooling stage. The convolutional stage uses the convolutional kernels to do the affine transformation for the input and then outputs feature maps. The detection stage completes the nonlinear operation with the rectified linear function, sigmoid function, or other activation functions. After that, the pooling stage is used to merge similar features to extract valuable information. As mentioned above, CNN has been employed to realize classifications of OAM modes with different topological charges[32–35]. In this work, we propose a CNN to calculate the OAM purity, and the architecture is composed of three convolutional layers, denoted as Conv1, Conv2, and Conv3, and five fully connected layers, denoted as FC1, FC2, FC3, FC4, and FC5, as shown in Fig. 2. The three convolutional layers use filters with the stride of [1, 1] and the same padding and with shapes of [5, 5], [5, 5], and [3, 3], respectively. Besides, all convolutional layers use the rectified linear unit (ReLU)[42] as an activation function and use the same max pooling (MP) with a shape of [3, 3] and a stride of [2, 2]. The number of filters of Conv1, Conv2, and Conv3 is 32, 64, and 128, separately. After a series of convolutional layers, the output will be fed into a series of fully connected layers. The number of neurons used for FC1, FC2, FC3, FC4, and FC5 is 1024, 1024, 512, 128, and 1. The final layer does not use the activation function, and other fully connected

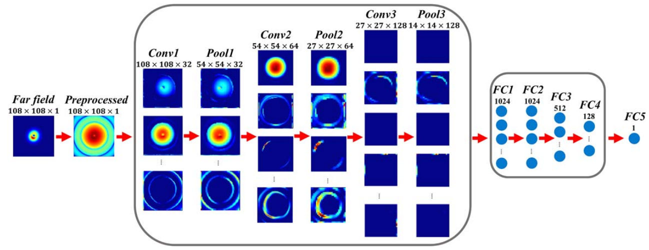

Fig. 2. Diagram of our proposed CNN architecture to predict the OAM mode purity. The CNN architecture consists of eight layers with three convolutional layers (Conv1, convolutional layer 1; Pool1, pool layer 1; etc.) and five fully connected layers (FC1, fully connected layer 1; etc.).

layers use the ReLU activation function. The final layer, with only one node, directly outputs the predicted OAM purity of a given far-field intensity pattern.

The mean square error between the output of FC5, namely the predicted OAM mode purities, and labels of training samples is employed as our loss function J, while the count percentage P of the testing samples predicted precisely within a predetermined absolute error (AE) tolerance is regarded as the criteria for assessment of the prediction performance[43]. The mathematical definitions of J, AE, and P are as follows:

$$J = \frac{1}{n}\sum_{i=1}^{n}(y_p^i - y_l^i)^2, \tag{1}$$

$$\mathrm{AE}(j) = \left|y_p^j - y_l^j\right|, \tag{2}$$

$$P = \frac{\mathrm{num}[\mathrm{AE}(j) < \mathrm{tol}]}{m}. \tag{3}$$

Here, n is the size of the training set. y_p^i and y_l^i denote the predicted and labeled OAM mode purity of the i^{th} training sample, respectively. y_p^j and y_l^j denote the predicted and labeled OAM mode purity of the j^{th} testing sample, respectively. num() is the number of testing samples that satisfy a certain relationship. m is the size of the testing set. tol denotes the predetermined AE tolerance. In an epoch, or a pool of training data, minibatches with a size of 64 out of 5000 samples in the training set are fed to the network in turn until the entire set is traversed. The training set is then shuffled and used in a new epoch. The process would be repeated 6000 times to make the network converge. It should be pointed out that it is preferable to use preprocessed data rather than raw data as input to efficiently train the neural network. The data preprocessing starts with taking the logarithm of the raw data in order to reduce the contrast ratio. Then, the statistical mean and variance of the training set are calculated and used to scale the data, including the training and testing data, obtained from the previous step by subtracting the mean and subsequently dividing them by variance. After the preprocessing, the data have zero mean, unit variance, and reasonable contrast ratio. The preprocessed data are very beneficial in characteristic learning and in boosting the optimization speed during the training phase. Notice that the Adam optimization[44] is used through this work to train our designed neural network.

The network training was accomplished using the TensorFlow[45] software library and a Nvidia GeForce GTX 1080Ti GPU. The training time costs approximately 5 h, and the result is illustrated in Fig. 3(a), where the performance of our neural network converges well, and the final loss function is below 10^{-6}. The testing set is then used to test the prediction performance of this well-trained CNN, which costs approximately 0.25 s. A typical histogram for the actual distribution of AE is shown in Fig. 3(b), and a zoom-in view plot for the AE larger than 0.01 is illustrated in the inset for clarity. Clearly, the prediction accuracy here can be as high as 99.3% when the tolerance is set to $\mathrm{tol} = 0.01$, and, if the tolerance is set to $\mathrm{tol} = 0.02$, the accuracy will be 99.8%, indicating extraordinary performance of this trained CNN in predicting the OAM mode purity.

In order to further investigate the performance of this well-trained CNN, the correlation coefficient between the predicted and labeled values of the OAM mode purity is calculated as[46,47]

$$r = \frac{\sum_{i=1}^{N}(y_p^i - \overline{y}_p)(y_l^i - \overline{y}_l)}{\sqrt{\sum_{i=1}^{N}(y_p^i - \overline{y}_p)^2 \cdot \sum_{i=1}^{N}(y_l^i - \overline{y}_l)^2}}. \tag{4}$$

Here, N is the size of the data set, and $\overline{y}_p$ and $\overline{y}_l$ denote the mean of the predicted and labeled OAM mode purity of the data set, respectively. As shown in Fig. 4, the data points mainly concentrate on the diagonal line, and the correlation coefficient is 0.99999 for the training set and 0.99994 for the testing set, strongly indicating that the predicted OAM mode purities agree very well with the labels. The reason why the CNN architecture has such a kind of extraordinary performance can be explained briefly as follows. When the layer goes deeper and deeper, as shown in Fig. 2, the images become more and more elusive to human eyes, but the extracted features of the pictures become progressively distinct to the network with diverse representations at higher and more abstract levels[48]. This powerful feature extraction capability, inherent to the CNN-based prediction model, is the key

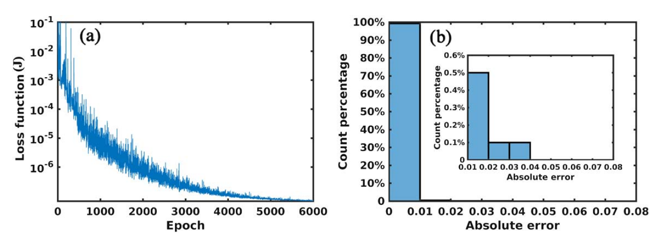

Fig. 3. (a) Training progress of our proposed CNN-based prediction model. (b) Histogram of AE distribution of 1000 testing samples. Inset: histogram with AE larger than 0.01.

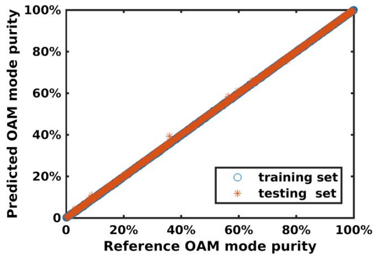

Fig. 4. Predicted OAM mode purities versus the labeled ones for both the training set and the testing set.

for this type of application, and assures the model's generality and robustness against other degrees of freedom, as we will discuss below.

As presented above, the deep learning algorithm is good at extracting features and self-learning, which means a well-trained neural network could be robust against the size of the testing set. To demonstrate this, testing sets containing more samples are utilized to test the generality of our proposed model. As shown in Fig. 5, the size of the testing sets ranges from 2000 to 20,000, and the obtained predication accuracy still maintains more than 99.7% when the tolerance is set to $\mathrm{tol} = 0.02$. Moreover, it is noteworthy that the accuracy fluctuates slightly rather than decreases monotonically as the size of the testing sets increases, suggesting that the network has received sufficient training. These results strongly illustrate that the proposed CNN has great generality.

The results discussed above have ignored possible phase differences of the modes due to the existence of mode dispersion and various phase perturbations in a practical optical system. It is thus indispensable to consider its impact on the prediction performance of the CNN-based model. To study such an effect, 100,000 groups of $[a_1, a_2, a_3, a_4, a_5]$ and $[\varphi_1, \varphi_2, \varphi_3, \varphi_4]$ are first generated using the random sampling approach, where φ_i $(i = 1, 2, 3, 4)$ represents the phase difference between the i^{th} mode and the OAM mode, whose value is in the range of $[0, 2\pi)$. The superposition of the modes is then given by $\boldsymbol{E}(x, y) = \sum_i a_i \boldsymbol{e}_i(x, y) \exp(j\varphi_i)$ with φ_5 always equaling zero. Then, the same CNN architecture as that shown in Fig. 2 is retrained with 90,000 groups of data as the training set and another 10,000 groups as the testing set. It should be pointed out that due to the introduced phase degree of freedom, a much bigger training set is

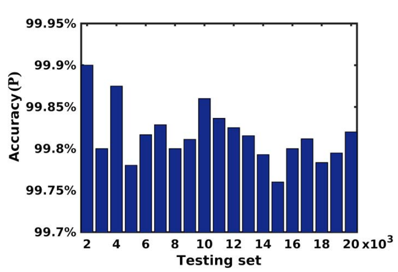

Fig. 5. Prediction accuracy as a function of the size of the testing set with the tolerance of 0.02.

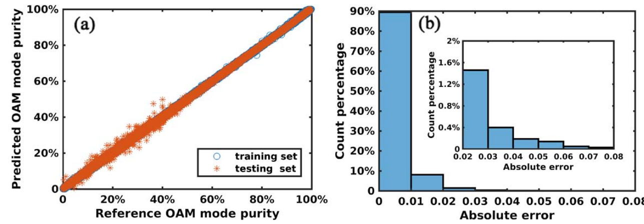

Fig. 6. (a) Predicted OAM mode purities versus the labels for both the training set and the testing set when phase differences exist between different modes in the optical fiber. (b) Histogram of AE distribution of 10,000 testing samples. Inset: histogram with AE larger than 0.02.

necessary in order to prevent overfitting of the neural network[49]. Besides, batch normalization (BN) is used before the activation function of each layer to exert the regularization effect and speed up the convergence[50].

The training and testing take approximately 87.7 h and 3.5 s, respectively. The correlation distribution of the predicted OAM mode purities relative to the labeled ones is shown in Fig. 6(a). The results presented here show data with a certain degree of deviation, especially for those with relatively low OAM purities. However, most of the 10,000 testing data are distributed diagonally. Based on Eq. (4), the overall correlation coefficients for the training set and testing set are very satisfactory with values of 0.99993 and 0.99966, respectively. The prediction accuracy still reaches 97.7% when the tolerance is set to $\mathrm{tol} = 0.02$ [see Fig. 6(b)]. Notice that the prediction performance for the data with relatively low OAM purities can be improved further by introducing more targeted training data with low OAM purities. Nevertheless, these findings suggest that our proposed CNN has great ability to deal with general cases when both the amplitude and phase degrees of freedom are considered.

Here, it is important to point out that because these three high-order modes (TE_{01}, TM_{01}, OAM^{+}_{+1}) have the same annular intensity distribution, the CNN cannot distinguish OAM^{+}_{+1} from the other high-order modes if there is no fundamental mode in the fiber. We are going to study the interference characteristics of the fundamental modes with the high-order modes. For the sake of simplicity, we only show the coherent interference patterns between fundamental modes and the high-order modes with the same intensity in Fig. 7. It is shown that these interference

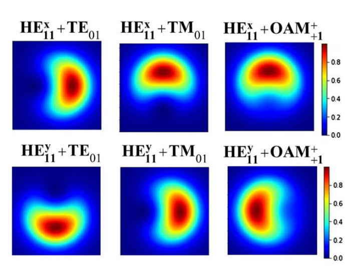

Fig. 7. Interferograms of the fundamental modes with the high-order modes.

patterns are different from the doughnut intensity profile, as shown in Fig. 1(b). Besides, it is easy for the human eyes to distinguish the interferograms of the fundamental modes with OAM_{+1}^{+} from the others. Thus, the existence of the fundamental modes makes OAM_{+1}^{+} have a unique interference pattern. Therefore, in theory, it is possible to extract the information of the OAM_{+1}^{+} purity based on the interference pattern in the presence of the fundamental modes.

To further illustrate the importance of the fundamental modes for the high-accuracy result of CNN in analyzing the OAM purity, we would test the performance of the trained CNN in the presence and absence of fundamental modes in the fiber, respectively. Without loss of generality, we test the second well-trained network in this Letter, which takes into account the phase differences among different modes. Since the power ratio of each fiber mode is generated using the random sampling approach, the probability that the power of fundamental modes is zero is so small that it can be ignored. The results of Figs. 6(a) and 6(b) demonstrate that the accuracy of the OAM purities predicted by the CNN is very high in the presence of fundamental modes. For comparison, 10,000 testing data are generated based on the same method: $\boldsymbol{E}(x,y)=\sum_i a_i \boldsymbol{e}_i(x,y)\exp(j\varphi_i)$, $i=1,2,\ldots,5$. The difference is that the magnitudes of the two fundamental modes are set up to be zero (i.e., $a_1=a_2=0$). The correlation distribution of the predicted OAM mode purities relative to the labeled ones is shown in Fig. 8, indicating the bad performance of the CNN in the absence of the fundamental modes. To sum up, the existence of the fundamental models is crucial for the extraordinary performance of the CNN in predicting the OAM mode purity.

In order to make the algorithm applicable for all situations even when no fundamental modes exist in the under test fiber, we can let the light of an under test fiber interfere with a coherent reference light, which is output from an auxiliary fiber containing only the fundamental modes. Then, the OAM_{+1}^{+} mode purity could be deduced according to the following formula:

$$S_r=\frac{(W_a+W_b)S_p}{W_b}, \tag{5}$$

where W_a is the power of the auxiliary fundamental modes, W_b is the power of all modes in the under test fiber, S_p is the OAM_{+1}^{+} mode purity of all modes including the auxiliary fundamental modes, and S_r is the deduced OAM_{+1}^{+} mode purity.

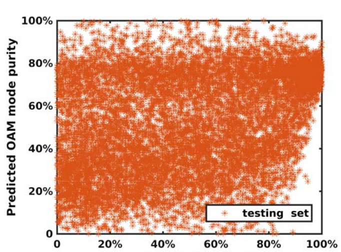

Fig. 8. Predicted OAM mode purities versus the labels for the testing set when no fundamental modes exist in the optical fiber.

In summary, a CNN-based deep learning technique to predict the OAM mode purity in optical fibers has been proposed, and its performance has been evaluated using synthetic data. A specific neural network composed of three convolutional layers and five fully connected layers is trained to determine the OAM purity with preprocessed far-field intensity patterns. The trained CNN has performed excellently in predicting OAM mode purity with very high accuracy of >99%. Besides, the generality of this technique is demonstrated by enlarging the testing set space, while the robustness is verified by adding random phase factors to the modes of optical fibers. It is worth noting that the architecture of our proposed CNN may also be tuned to deal with other types of OAM fibers besides the FMF, which we used as an example in this work. Contrast to traditional evaluation methods, which typically require many bulk optic devices and careful alignment, this method could, in principle, dramatically simplify the process of measuring OAM purity. We believe that our image-processing-based method holds great promise for potential applications in OAM-related spatial-division-multiplexing technology and will be an important future research direction for other optical communication technologies.

This work was supported by the National Basic Research Program of China (No. 2015CB659400), the Natural Science Foundation of Jiangsu Province (No. BK20150057), and the Fundamental Research Funds for the Central Universities (No. 021314380100).

References

1. D. J. Richardson, J. M. Fini, and L. E. Nelson, Nat. Photon. **7**, 354 (2013).
2. J. Sakaguchi, B. J. Puttnam, W. Klaus, Y. Awaji, N. Wada, A. Kanno, T. Kawanishi, K. Imamura, H. Inaba, K. Mukasa, R. Sugizaki, T. Kobayashi, and M. Watanabe, J. Lightwave Technol. **31**, 554 (2013).
3. G. Li, N. Bai, N. Zhao, and C. Xia, Adv. Opt. Photon. **6**, 413 (2014).
4. A. E. Willner, H. Huang, Y. Yan, Y. Ren, N. Ahmed, G. Xie, C. Bao, L. Li, Y. Cao, Z. Zhao, J. Wang, M. P. J. Lavery, M. Tur, S. Ramachandran, A. F. Molisch, N. Ashrafi, and S. Ashrafi, Adv. Opt. Photon. **7**, 66 (2015).
5. S. Ramachandran and P. Kristensen, Nanophotonics **2**, 455 (2013).
6. N. Bozinovic, Y. Yue, Y. Ren, M. Tur, P. Kristensen, H. Huang, A. E. Willner, and S. Ramachandran, Science **340**, 1545 (2013).
7. S. Li and J. Wang, IEEE Photon. J. **5**, 7101007 (2013).
8. H. Xu and L. Yang, Opt. Lett. **38**, 1978 (2013).
9. S. Ramachandran, P. Kristensen, and M. F. Yan, Opt. Lett. **34**, 2525 (2009).
10. R. D. Niederriter, M. E. Siemens, and J. T. Gopinath, Opt. Lett. **41**, 3213 (2016).
11. J. Xing, J. Wen, J. Wang, F. Pang, Z. Chen, Y. Liu, and T. Wang, Chin. Opt. Lett. **16**, 100604 (2018).
12. Z. Zhang, W. Wei, L. Tang, J. Yang, J. Guo, L. Ding, and Y. Li, Chin. Opt. Lett. **16**, 110501 (2018).
13. H. Zhang, L. Yao, Y. Pang, and J. Xia, Chin. Opt. Lett. **16**, 092601 (2018).

14. Z. Zhao, J. Wang, S. Li, and A. E. Willner, Opt. Lett. **38**, 932 (2013).
15. S. Li and Z. Wang, Appl. Phys. Lett. **103**, 141110 (2013).
16. S. J. Parkin, T. A. Nieminen, N. R. Heckenberg, and H. Rubinsztein-Dunlop, Phys. Rev. A **70**, 690 (2004).
17. J. Leach, S. Keen, M. J. Padgett, C. Saunter, and G. D. Love, Opt. Express **14**, 11919 (2006).
18. J. Leach, M. J. Padgett, S. M. Barnett, S. Franke-Arnold, and J. Courtial, Phys. Rev. Lett. **88**, 257901 (2002).
19. N. Bozinovic, S. Golowich, P. Kristensen, and S. Ramachandran, Opt. Lett. **37**, 2451 (2012).
20. Y. Jiang, G. Ren, H. Li, M. Tang, Y. Liu, Y. Wu, W. Jian, and S. Jian, Appl. Opt. **56**, 1990 (2017).
21. Y. Wen, I. Chremmos, Y. Chen, J. Zhu, Y. Zhang, and S. Yu, Phys. Rev. Lett. **120**, 193904 (2018).
22. Y. LeCun, Y. Bengio, and G. Hinton, Nature **521**, 436 (2015).
23. C. J. C. Burges, Data Min. Knowl. Disc. **2**, 121 (1998).
24. T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, IEEE Trans. Pattern Anal. **24**, 881 (2002).
25. V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, J. Chem. Inf. Comput. Sci. **43**, 1947 (2003).
26. J. Ham, C. Yangchi, M. M. Crawford, and J. Ghosh, IEEE Trans. Geosci. Remote Sens. **43**, 492 (2005).
27. J. Schmidhuber, Neural Networks **61**, 85 (2015).
28. Y. Rivenson, Z. Gorocs, H. Gunaydin, Y. B. Zhang, H. D. Wang, and A. Ozcan, Optica **4**, 1437 (2017).
29. E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, Optica **5**, 458 (2018).
30. B. Mills, D. J. Heath, J. A. Grant-Jacob, and R. W. Eason, Opt. Express **26**, 17245 (2018).
31. M. D. Hannel, A. Abdulali, M. O'Brien, and D. G. Grier, Opt. Express **26**, 15221 (2018).
32. J. Li, M. Zhang, and D. Wang, IEEE Photon. Technol. Lett. **29**, 1455 (2017).
33. S. R. Park, L. Cattell, J. M. Nichols, A. Watnik, T. Doster, and G. K. Rohde, Opt. Express **26**, 4004 (2018).
34. J. Li, M. Zhang, D. Wang, S. Wu, and Y. Zhan, Opt. Express **26**, 10494 (2018).
35. T. Doster and A. T. Watnik, Appl. Opt. **56**, 3386 (2017).
36. P. Z. Dashti, F. Alhassen, and H. P. Lee, Phys. Rev. Lett. **96**, 043604 (2006).
37. R. Kashyap, *Fiber Bragg Gratings* (Academic, 2009).
38. E. Hecht, *Optics* (Addison-Wesley, 2002).
39. M. A. Arbib, *The Handbook of Brain Theory and Neural Networks* (MIT, 1998).
40. I. S. A. Krizhevsky and G. E. Hinton, in *Advances in Neural Information Processing Systems* (2012), p. 1097.
41. A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and F. F. Li, in *IEEE Conference on Computer Vision and Pattern Recognition* (2014), p. 1725.
42. G. E. Dahl, T. N. Sainath, and G. E. Hinton, in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2013), p. 8609.
43. A. Liu, T. Lin, H. Han, X. Zhang, Z. Chen, F. Gan, H. Lv, and X. Liu, Opt. Express **26**, 22100 (2018).
44. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv:1412.6980 (2014).
45. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, in *12th USENIX Symposium on Operating Systems Design and Implementation* (2016), p. 265.
46. J. Shao, *Mathematical Statistics* (Springer, 2003).
47. D. C. Elton, Z. Boukouvalas, M. S. Butrico, M. D. Fuge, and P. W. Chung, Sci. Rep. **8**, 9059 (2018).
48. C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006).
49. I. V. Tetko, D. J. Livingstone, and A. I. Luik, J. Chem. Inf. Comp. Sci. **35**, 826 (1995).
50. S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," arXiv:1502.03167 (2015).