



Article

Mode Recognition of Orbital Angular Momentum Based on Attention Pyramid Convolutional Neural Network

Tan Qu ^{1,*} , Zhiming Zhao ¹, Yan Zhang ², Jiaji Wu ¹ and Zhensen Wu ³ ¹ School of Electronic Engineering, Xidian University, Xi'an 710071, China² School of Computer Science, Xi'an Shiyou University, Xi'an 710065, China³ School of Physics, Xidian University, Xi'an 710071, China

* Correspondence: tqu@xidian.edu.cn

Abstract: In an effort to address the problem of the insufficient accuracy of existing orbital angular momentum (OAM) detection systems for vortex optical communication, an OAM mode detection technology based on an attention pyramid convolution neural network (AP-CNN) is proposed. By introducing fine-grained image classification, the low-level detailed features of the similar light intensity distribution of vortex beam superposition and plane wave interferograms are fully utilized. Using ResNet18 as the backbone of AP-CNN, a dual path structure with an attention pyramid is adopted to detect subtle differences in the light intensity in images. Under different turbulence intensities and transmission distances, the detection accuracy and system bit error rate of basic CNN with three convolution layers and two full connection layers, i.e., ResNet18 and ResNet18, with a specified mapping relationship and AP-CNN, are numerically analyzed. Compared to ResNet18, AP-CNN achieves up to a 7% improvement of accuracy and a 3% reduction of incorrect mode identification in the confusion matrix of superimposed vortex modes. The accuracy of single OAM mode detection based on AP-CNN can be effectively improved by 5.5% compared with ResNet18 at a transmission distance of 2 km in strong atmospheric turbulence. The proposed OAM detection scheme may find important applications in optical communications and remote sensing.



Citation: Qu, T.; Zhao, Z.; Zhang, Y.; Wu, J.; Wu, Z. Mode Recognition of Orbital Angular Momentum Based on Attention Pyramid Convolutional Neural Network. *Remote Sens.* **2022**, *14*, 4618. <https://doi.org/10.3390/rs14184618>

Academic Editor: Carmine Serio

Received: 27 July 2022

Accepted: 13 September 2022

Published: 15 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: orbital angular momentum; mode detection; fine-grained image classification; attention pyramid; atmospheric turbulence

1. Introduction

Vortex beams with spiral phase structures have been used extensively in information transmission, radar imaging and rotational target detection, since Allen first investigated the Laguerre-Gaussian vortex beam and its orbital angular momentum (OAM) in 1992 [1]. Theoretically, there are infinite kinds of eigenstates; different eigenstates are orthogonal to each other, which is quite important in terms of improving communication capacity and imaging resolution in remote sensing. Multiplexing different OAM beams can effectively avoid the crosstalk between different modes in a channel, providing a new communication dimension that is no longer limited to amplitude, phase, frequency and polarization, thereby greatly improving the communication capacity [2,3]. In free-space OAM communication systems, the receiver needs to demodulate the OAM beam to recover the information sequence. Traditional OAM demodulation techniques, such as spatial light modulators, the diffraction method, the cylindrical lens method, plane wave interferometry and spherical wave interferometry, are based on optical hardware and have been researched extensively [4–7]. The OAM beam is pre-processed by optical hardware to obtain optical pattern features that can be distinguished by the naked eye [8]. However, on account of the high cost and the limited processing capability of optical hardware, high-performance transmission cannot be guaranteed with a cost-effective vortex optical communication system.

In recent years, with the rapid increase in computing power, OAM mode recognition based on deep learning has attracted growing attention. Some researchers have studied

OAM mode recognition based on neural networks. Krenn et al. proposed a self-organizing competitive neural network (SOM) based OAM mode recognition which verified the feasibility of machine learning in vortex optical communication systems for the first time [9]. They built a long-distance vortex optical communication system on the sea between the Canary Islands, achieving a recognition accuracy of 91.67%, which verified the possibility of the use of vortex beams for long-distance information transmission [10]. Deep neural network (DNN)-based recognition of different OAM modes is proposed in [11]. Convolutional neural networks (CNNs) are proposed in OAM mode recognition by Doster et al. [12]; that method achieved a mode recognition accuracy of up to 99%, which is far superior to the traditional methods. In addition, the CNN-based mode identification method is robust in terms of the influence of turbulence intensity, data size, sensor noise and pixels. Such research has paved the way for the application of CNN in OAM mode detection. Subsequently, many achievements in OAM mode recognition have been realized based on CNNs [13,14].

In 2017, Zhang et al. compared the performance of a K nearest neighbor neural network, a plain Bayesian classifier, a Back Projection artificial neural network and a CNN as OAM mode classifiers under different turbulence conditions. They observed that CNN yielded the best results [15]. The same authors improved the original LeNet-5 network and proposed a decoder scheme that could simultaneously implement OAM mode and turbulence intensity recognition [16]. OAM mode recognition technology combined with turbo channel coding is proposed in [17]; this approach effectively improved the recognition accuracy and reliability of communication transmission.

Similarly, many scholars have worked on niche applications of OAM mode detection. In 2018, Zhao et al. applied a CNN to learn a received OAM light intensity map under different tilt angles by adding a view-pooling layer. They also used a hybrid data collection technique to improve the performance [18]. Misaligned hyperfine OAM mode recognition was carried out in [19]. Machine learning based the recognition of fractional optical vortex modes in atmospheric environment was studied by Cao et al. [20]. When the marked data sample was insufficient, an OAM mode recognition method based on Conditional Generative Adversarial Networks (CGAN) was proposed to improve the recognition accuracy [21]. A Diffractive Deep Neural Network (D2NN) was utilized in OAM mode recognition in [22], eliminating the need for a CCD camera to capture images and pass them to a computer, making the communication rate independent of the hardware and neural network computation rate.

The above research was dedicated to training and identifying the OAM light distributions captured by CCD cameras; however, some other researchers have performed transformations on the vortex beam before training to highlight the characteristics of different modes. In 2018, Radon transform was introduced to preprocess a light intensity distribution map of an OAM beam to obtain more clearly distinguishing features [23]. A mode recognition technique based on coherent light interference at the receiver side to obtain more obvious recognition features was reported in [24]. A SVM-based single-mode recognition method was proposed in [25], using the relationship between the amount of OAM beam receiving the effect of atmospheric turbulence distortion and the topological charge number as an artificial feature of the design. A joint scheme combining the Gerchberg–Saxton (GS) algorithm and CNN (GS-CNN) to achieve the efficient recognition of the multiplexing LG beams was proposed in [26]. A technique to measure the OAM of light based on the petal interference patterns of modulated vortex beams and an unmodulated incident Gaussian beam reflected by a spatial light modulator was reported in [27].

In light of the aforementioned studies, it may be stated that most research has focused on OAM mode detection by neural networks, CNNs or CNN-based combination methods, preprocessing transform before network training, and OAM mode detection in misaligned or tilt angles special cases. However, in practical applications, there are many different multi-modes superpositions of OAM beams corresponding to quite similar light intensity maps, such as $\text{OAM} = \{-2, 3, -5\}$ and $\text{OAM} = \{1, -2, 3, -5\}$, $\text{OAM} = \{4, -4\}$ and $\text{OAM} = \{2, -6\}$, etc.

Additionally, for a single mode vortex beam with a large topological charge, the number of fringes in the interferogram is large. Because the area of the device that collects the image at the receiving end is certain, it is difficult to determine when the number of interference fringes is large, which will further affect the accuracy of OAM mode recognition.

To solve these problems, an OAM mode recognition technique based on an attention pyramid convolutional neural network (AP-CNN) is proposed in this paper. Fine-grained image classification [28] is introduced to make full use of low-level detailed information of a similar intensity in superimposed vortex beams and the dense fringes of a plane wave interferogram of a single-mode vortex beam with a large topological charge. A top-down feature path and a bottom-up attention path structure, combined with an attention pyramid, is adopted to improve the OAM mode recognition accuracy and reduce the bit error rate (BER) for indistinguishable intensity distributions.

The remainder of this paper is arranged as follows. The OAM mode recognition technical framework based on AP-CNN is described in Section 2. In Section 3, numerical results and discussions of different transmission conditions are presented to compare the recognition accuracy and bit error ratio. Section 4 is devoted to the conclusion.

2. Materials and Methods

2.1. Principle of AP-CNN

The principles of the AP-CNN [29,30] and the fine-grained image classification algorithm [28] used in this paper are shown in Figure 1. Figure 1a illustrates the dual-path algorithm structure, Figure 1b presents the attention pyramid, and Figure 1c illustrates the region of interest (ROI) pyramid. The blue border represents the feature map, and the orange border represents the channel/space attention.

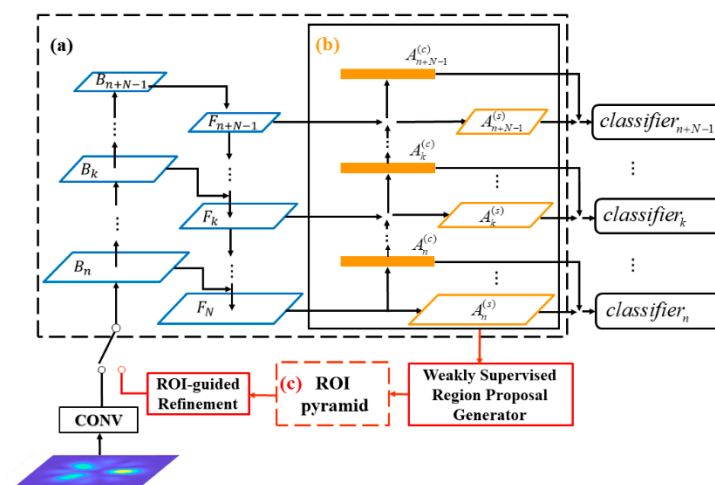


Figure 1. Structure of the AP-CNN for OAM mode recognition. (a) dual-path algorithm structure, (b) attention pyramid, (c) ROI pyramid.

First, the AP-CNN network takes an image as input and generates a feature pyramid network (FPN) and an attention pyramid [31] to enhance representations by improving on the CNN to obtain a dual-path algorithm structure, including a top-down feature path and a bottom-up attention path. The FPN [29] is used on the top-down path to extract features at different scales. Then, an additional attention hierarchy is introduced to further enhance the structure, including a spatial attention pyramid $\{A_n^{(s)}, A_{n+1}^{(s)}, \dots, A_{n+N-1}^{(s)}\}$ for locating discriminative regions at different scales, and a channel attention path $\{A_n^{(c)}, A_{n+1}^{(c)}, \dots, A_{n+N-1}^{(c)}\}$ for adding channel correlations in another bottom-up path and transferring local information from the lower pyramid level to the higher pyramid level.

For the spatial attention pyramid, each building block takes the feature map of the corresponding layer F_k as input and generates the spatial attention mask $A_k^{(s)}$. The feature

map F_k first passes through a 3×3 deconvolution layer with only one output channel to compress the spatial information. Each element of the spatial attention mask $A_k^{(s)}$ is normalized to be in the range of 0 to 1 using the sigmoid function, expressed as:

$$A_k^{(s)} = \sigma(v_c * F_k) \quad (1)$$

where σ denotes the sigmoid function, v_c denotes the convolution kernel, and $*$ denotes the deconvolution with the fixed convolution kernel. For the channel attention path, the channel attention can be obtained by passing the global average pooling layer and two fully connected layers in the corresponding feature layer of the feature pyramid. The channel attention mask formula is given by,

$$A_k^{(c)} = \sigma(W_2 \cdot \text{ReLu}(W_1 \cdot \text{GAP}(F_k))) \quad (2)$$

where GAP represents the global average pooling layer, W_1 and W_2 represent the weight matrices of the two fully connected layers. The learned attention is used to weight the feature F_k to obtain F'_k , which is used for classification as follows,

$$F'_k = F_k \cdot (A_k^{(s)} \oplus A_k^{(c)}) \quad (3)$$

where \oplus represents the broadcasting addition operation on semantics. Spatial attention tensor and channel attention tensor have different shapes, and the plus operator must be of broadcast type.

In the second step, after obtaining the spatial attention pyramid, the ROI pyramid continues to be generated by the region suggestion generator of adaptive non maximum suppression (NMS) [32] in a weakly supervised manner. The purpose of the Region Proposal Network (RPN) [33] is to select a frame that may contain a target. In essence, it is based on the unclassified target detector of the sliding window; it inputs an image of any scale and obtains a candidate frame with a predetermined size and scale. The general RPN network is mostly applied to the single- or multi-scale convolution network feature map. Multiple sizes and aspect ratios are preset to locate objects of different sizes and shapes. On the basis of RPN theory, AP-CNN uses a spatial attention mask as an anchor score and uses weak supervision to select the distinguishing area. According to the convolution receptive field of each pyramid layer, AP-CNN selects the corresponding recommended area with a preset size and aspect ratio for each pyramid layer, applies adaptive NMS to the selected area after calculating the score, reduces redundancy by eliminating overlapping, and maintains visual integrity by combining related areas. Figure 2 shows the workflow of the weak supervision area suggestion generator in the OAM mode detection task. Compared with the soft mechanism of setting the threshold on a feature map, the adaptive region suggestion generator based on the ROI can explicitly show distinguishable regions with high response values in the light intensity distribution of OAM modes.

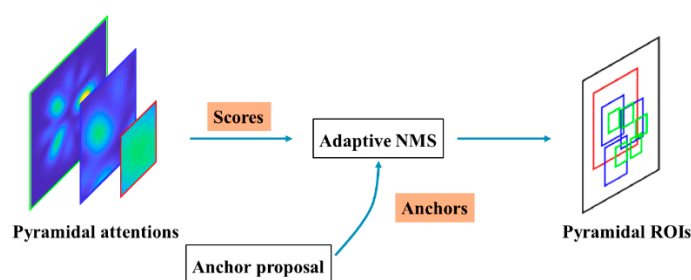


Figure 2. Workflow of the weakly supervised region proposal generator.

In the third step, for each layer of the pyramid, after selecting the ROI based on the region proposal generator and constructing the region pyramid $R_{\text{all}} = \{R_n, R_{n+1}, \dots, R_{n+N-1}\}$,

AP-CNN performs ROI-guided refinement of the feature map at the bottom of the pyramid B_n to improve the classification accuracy in the refinement stage. The first part is the ROI-guided drop block regularization [34], AP-CNN randomly selects an ROI joint R_s from the constructed N -layer region of the interesting pyramid $R_{all} = \{R_n, R_{n+1}, \dots, R_{n+N-1}\}$ based on the drop block selection probability of each layer $P_{all} = \{P_n, P_{n+1}, \dots, P_{n+N-1}\}$. Then the information region r_s is randomly selected with equal probability R_s and processed to the same sampling rate as the feature map at the bottom of the pyramid B_n to obtain the mask M by setting the activation of the information region to zero,

$$M(i, j) = \begin{cases} 0, & (i, j) \in r_s \\ 1, & \text{otherwise} \end{cases} \quad (4)$$

Apply the mask M on the low-level feature map B_n and normalize it to obtain the desired feature map D_n ,

$$D_n = B_n * M * \text{Count}(M) / \text{Count_ones}(M) \quad (5)$$

where $\text{Count}()$ and $\text{Count_ones}()$ represent the total number of elements and the total number of elements with the value of 1, respectively. The second part is the ROI-guided amplification operation, where the AP-CNN combines all ROI regions at the pyramid level to obtain the minimum enclosing rectangle of the input image in a weakly supervised manner,

$$\begin{aligned} t_{x1} &= \min(\forall x \in R_{all}), t_{y1} = \min(\forall y \in R_{all}) \\ t_{x2} &= \max(\forall x \in R_{all}), t_{y2} = \max(\forall y \in R_{all}) \end{aligned} \quad (6)$$

where t_{x1}, t_{y1} represent the minimum coordinates of the x - and y -axes of the merged bounding box and t_{x2}, t_{y2} represent the maximum coordinates of the x - and y -axes of the merged bounding box. The calibration area is then extracted from D_n and enlarged to the same size D_n to obtain the enlarged feature map.

Separate classifiers are set up for the original and refinement stages for their respective pyramids, and the final classification results are taken as the average of the predicted values in the original stage and the predicted values in the refinement stage.

2.2. Recognizing OAM Modes Based on AP-CNN

Figure 3 displays the light intensity distributions of two similar superposition mode vortex beams of $\text{OAM} = \{1, -2, 3, -5\}$ and $\text{OAM} = \{-2, 3, -5\}$. The ROI, localized from low to high level, are shown. This approach can be used to identify the ROI located on different pyramid levels, and more detailed information can be captured at the low levels to distinguish different OAM modes. Compared with high-level image semantic information, after thinning image features, this low-level information is very helpful to improve the accuracy of OAM mode detection.

The CCD camera at the receiving end captures the light intensity distribution of the vortex beam after atmospheric turbulence and inputs it into the AP-CNN to detect the OAM mode and retrieve the transmitted data. Here, we use the ResNet18 network as the backbone of the AP-CNN. The ResNet18 structure used in this paper differs slightly from the official ResNet18 [35] structure, as described below:

- (1) The size of the input light intensity map is $128 \times 128 \times 3$. To avoid the problems of the low resolution of the final feature map after multiple downsampling and the serious loss of semantic information, the maximum pooling layer of stage0 is removed.
- (2) To reduce the number of model parameters, the 7×7 convolutional kernel of stage0 is replaced by a 3×3 convolution kernel.

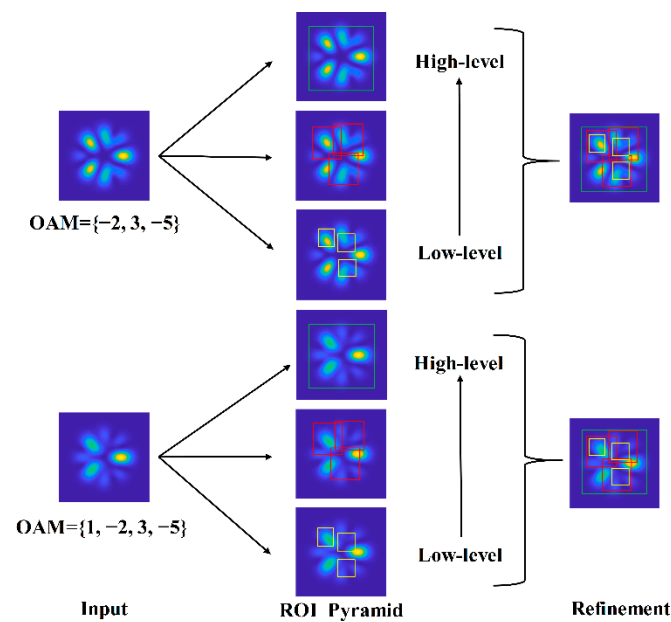


Figure 3. Regions of interest localized by AP-CNN at different levels for different OAM modes.

The structure of the modified ResNet18 is shown in Table 1.

Table 1. Modified ResNet18 network structure.

Network Layer	Output Feature Map Size	ResNet18
conv1	128×128	$3 \times 3, 64$
conv2_x	64×64	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
conv3_x	32×32	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$
conv4_x	16×16	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$
conv5_x	8×8	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$
	1×1	Average pooling + full connected + softmax

The OAM mode recognition algorithm AP-CNN consists of two parts: the backbone network ResNet18 and the refinement network, as shown in Figure 4. The size of the simulated light intensity distribution map of the OAM beam is set to 128×128 and is input into the ResNet18 network for training. Firstly, it is input into stage0 for pre-processing, and only the features are extracted (64 convolutional kernels with a size of 3×3 and a step size of 2), and the 128×128 feature map is output. Then, the feature map is fed into the next four layers of residual blocks, which reduces the size of the input feature map by half compared to the original size and doubles the number of channels. Next, the output feature maps of the third, fourth, and fifth layers of ResNet18 are denoted as B_3, B_4, B_5 , respectively, for subsequent building of the feature pyramid, as shown in Figure 4. Further refinement is carried out at the B_3 level of the pyramid. We respectively assign anchors with single scales of 18, 36, and 72 and a 1:1 ratio for each pyramidal level and choose the top 5, 3, and 1 anchors with the highest activation values as potential refinement candidates. For the adaptive NMS, the cutoff threshold is set to 0.05, the merge threshold to 0.9, and the drop block probability to {30%, 30%, 0%}.

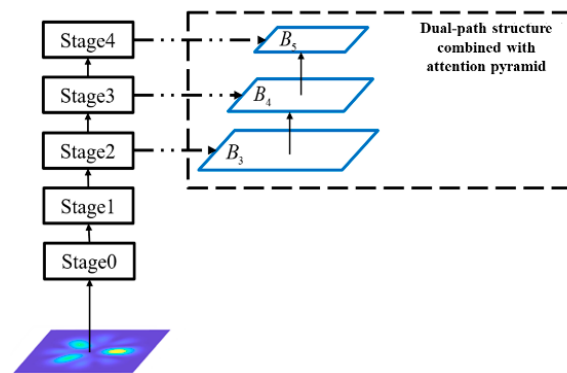


Figure 4. Connection between ResNet18 and the refinement network.

During AP-CNN training, the initial learning rate is set to 0.01, decreasing by 10% for every 20 iterations; as such, a total of 100 epochs are trained. The random gradient descent algorithm with a momentum coefficient of 0.9 and a minipatch of 16 is used for parameter optimization, and the weight attenuation is set to 5×10^{-4} . The experimental operating system is windows, the programming language of the algorithm part is python, and the deep learning framework is pytorch. The software versions are shown in Table 2. The graphics card we used was an RTX2060.

Table 2. Software used in experiments.

Software	Edition
Windows	Windows10 (21H2)
Python	3.7
Pytorch	1.7.1
Torchvision	0.8.2
CUDA	11.0.2
CuDNN	11.2

2.3. Performance Evaluation Index

The OAM mode detection performance of the network is evaluated by two indicators: detection accuracy and BER. The detection accuracy is defined as the ratio of the number of correct OAM mode detection samples to the total number of vortex light intensity distributions on the test set, determined using Equation (7):

$$\text{Accuracy} = \frac{\sum_{m=1}^M f(m)}{M} \quad (7)$$

where M represents the total number of light intensity distribution maps of vortex light and m represents the OAM mode. $f(m)$ is 1 when the identification is correct and 0 when it is wrong.

BER and the symbol error rate (SER) are commonly used to evaluate the probability of transmission errors in communication systems. The SER is defined as the probability of a symbol transmission error, i.e., the ratio of the number of erroneous symbols at the receiving end to the total number of transmitted symbols:

$$\text{SER} = \sum_{i=1}^M [p_i(1 - p(s_i|s_i))] \quad (8)$$

where M represents the number of symbol types, p_i denotes the probability of transmitting each symbol with the value $1/M$, and $p(s_i|s_i)$ represents the correct conditional probability density detected by the receiver in a certain OAM mode.

$$p(s_i|s_i) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \exp(h_T - \eta_{s_i s_i} G L_{path} \sqrt{\gamma_\beta}) \prod_{k=1, k \neq i}^M [1 - 0.5 \operatorname{erfc}(h_T - \eta_{s_j s_i} G L_{path} \sqrt{\gamma_\beta})] dh_T \quad (9)$$

where s_i represents the OAM mode and h_T , G , L_{path} are constants, representing the power detection threshold, the average gain of the receiver, and the path loss, respectively. The value of $p(s_i|s_i)$ is theoretically mainly determined by signal-to-noise ratio γ_β at the receiving end and the helical spectral distribution.

The relationship of BER with the SER is as follows:

$$\text{BER} = \text{SER} / (\log_2 M) \quad (10)$$

In our simulation experiments, the eight OAM superposition modes are one-to-one, corresponding to the octal symbols. At the receiving end, by identifying the light intensity distribution, the code word sequence is obtained by inversion. This is then compared with the theoretical code word sequence, while the ratio of the number of wrong code words to the total number of code words is the BER.

3. Results and Discussion

3.1. Simulation Data Set Construction

In order to verify the performance of the AP-CNN in OAM mode detection for similar distributions of multi-mode vortex beams, four pairs of OAM modes, namely, $\{1, -2\}$ and $\{1, -2, -5\}$, $\{1, -2, 3, -5\}$ and $\{-2, 3, -5\}$, $\{4, -4\}$ and $\{2, -6\}$, $\{6, -6\}$ and $\{9, -3\}$, are selected, as shown in Figure 5. The light intensity distributions in four columns are similar to each other. In addition, in order to test the detection performance of a single-mode vortex beam with a large topological charge interfering with the plane wave, a plane wave interferogram dataset of a single-mode vortex beam is constructed, choosing eight types of samples with large topological charges, i.e., $\pm 17, \pm 18, \pm 19, \pm 20$, as shown in Figure 6.

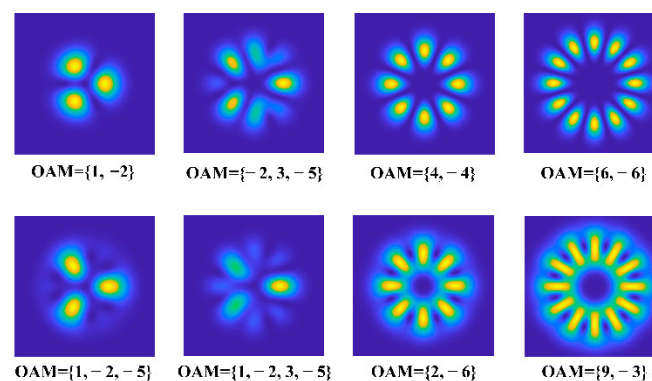


Figure 5. Light intensity distribution of a similar multi-mode OAM beam.

The wavelength of the OAM communication system is $0.6328 \mu\text{m}$ and the beam waist radius is 0.3 m . For the comparison experiments under different turbulent conditions, the transmission distance is fixed and six different atmospheric refractive index structure constants, C_n^2 , are selected: $1.0 \times 10^{-14} \text{m}^{-2/3}$, $3.0 \times 10^{-14} \text{m}^{-2/3}$, $5.0 \times 10^{-14} \text{m}^{-2/3}$, $1.0 \times 10^{-13} \text{m}^{-2/3}$, $3.0 \times 10^{-13} \text{m}^{-2/3}$ and $5.0 \times 10^{-13} \text{m}^{-2/3}$. For comparison experiments at different transmission distances, the C_n^2 is fixed and six different transmission distances are chosen: 500 m , 1000 m , 1500 m , 2000 m , 2500 m , and 3000 m . When simulating the atmospheric turbulence channel, the power spectrum inversion method is used to decimate the transmission distance in order to obtain ten phase screens with certain intervals. For each transmission condition, 2000 light

intensity maps are generated for each OAM mode. In this way, a total of 16,000 light intensity distribution maps are included in the hybrid dataset. This is then divided into a training and a test set at a ratio of 8:2 (12,800 images in the training set and 3200 images in the test set).

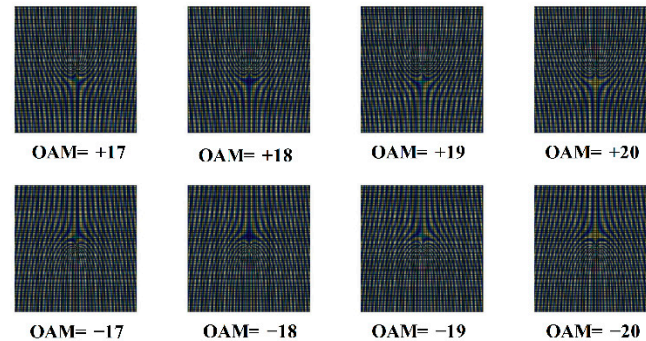


Figure 6. Plane wave interferogram of single-mode vortex beam with large topological charge.

3.2. Analysis of Multi-Mode OAM Mode Recognition Based on AP-CNN

In order to compare the performance of the AP-CNN with that of ResNet18, the OAM recognition accuracy, confusion matrix, and BER of both models are numerically analyzed in this section. Variations in OAM mode recognition accuracy are shown in Figures 7 and 8 under different turbulence intensities for transmission distances 2000 m and 3000 m, respectively.

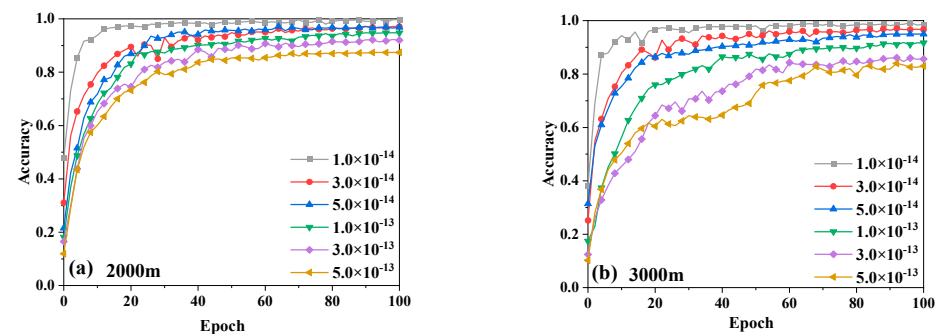


Figure 7. Accuracy of OAM mode recognition based on ResNet18 at different turbulence intensities (a) transmission distance: 2000 m (b) transmission distance: 3000 m.

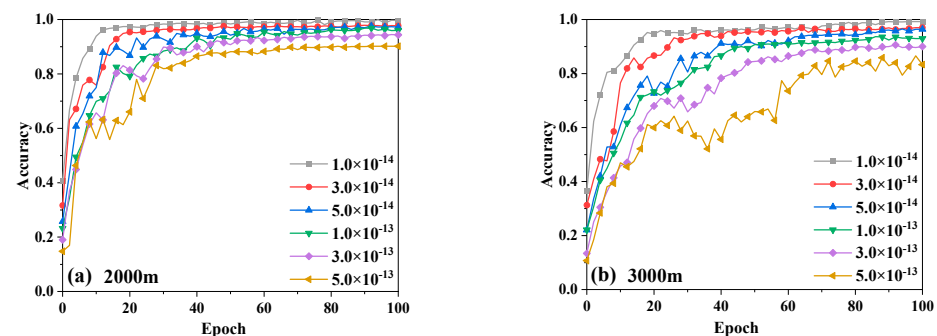


Figure 8. Accuracy of OAM mode recognition based on AP-CNN under different turbulence intensities. (a) transmission distance: 2000 m (b) transmission distance: 3000 m.

As shown, the detection accuracy increases gradually with an increase in the training epoch. Additionally, the comparisons show that the detection accuracy increases significantly more slowly in strong turbulence than in weak ones. When the turbulence is strong, the training process becomes more problematic because the image suffers more serious distortion. Taking a transmission distance of 2000 m as an example, as shown in Figure 7, after 100-times training in medium turbulence ($C_n^2 = 1.0 \times 10^{-14} \text{m}^{-2/3}$, $C_n^2 = 3.0 \times 10^{-14} \text{m}^{-2/3}$, $C_n^2 = 5.0 \times 10^{-14} \text{m}^{-2/3}$),

ResNet18 can still achieve an OAM mode detection accuracy of more than 96.9%, whereas under strong turbulence ($C_n^2 = 3.0 \times 10^{-13} \text{m}^{-2/3}$ and $C_n^2 = 5.0 \times 10^{-13} \text{m}^{-2/3}$), the detection accuracy decrease to 91.7% and 87.3%, respectively. The detection accuracy is further reduced from 87.3% to 83.2% when the transmission distance is increased to 3000 m under strong turbulence ($C_n^2 = 5.0 \times 10^{-13} \text{m}^{-2/3}$). In other words, the greater the turbulence intensity and the farther the transmission distance, the lower the detection accuracy.

A comparison of Figures 7 and 8 reveals that the accuracy of OAM mode recognition based on AP-CNN is superior to that of ResNet18. When the turbulence is weak, the optimization effect of the AP-CNN is minimal, while when the turbulence is strong, it is more obvious. For example, when the transmission distance is 2000 m for $C_n^2 = 3.0 \times 10^{-14} \text{m}^{-2/3}$, as shown in Figure 8, the accuracy of OAM mode recognition based on the AP-CNN can reach up to 98.1% after 100 training epochs, which is only a 0.6% improvement compared with ResNet18. Additionally, under these circumstances, the optimization effect is not obvious. Meanwhile, when $C_n^2 = 3.0 \times 10^{-13} \text{m}^{-2/3}$, the recognition accuracy based on ResNet18 is only 92.1%. The accuracy of OAM mode recognition based on AP-CNN, on the other hand, is significantly improved, and the recognition accuracy of the best model reaches 94.4%, showing an improvement of about 2.3%.

When the transmission distance is large, the improvement of the recognition accuracy based on AP-CNN is limited in a highly turbulent environment. For example, when $C_n^2 = 5.0 \times 10^{-13} \text{m}^{-2/3}$, the accuracy of OAM mode recognition based on the AP-CNN can reach up to 90.5% after 100 training epochs when the transmission distance is 2000 m, which is about a 3.1% improvement compared to ResNet18, as shown in Figure 8. However, when the transmission distance extends to 3000 m, the recognition accuracy of the best model can only reach 84.2% after AP-CNN, i.e., 1.2% higher than ResNet18. The reason for this is that in strong turbulence, when the distance is too great, the transmission of the OAM beam is greatly affected by the turbulent disturbance distortion. The light intensity distribution captured by the CCD camera at the receiving end is seriously distorted, and the image features used for recognition are compromised. Therefore, even if the AP-CNN introduces an attention mechanism to mine the underlying image details, it cannot significantly improve the accuracy of OAM mode recognition.

The confusion matrixes of superimposed vortex beams using ResNet18 and AP-CNN, with a transmission distance is 2000 m and when atmospheric refractive index structure constant $C_n^2 = 5.0 \times 10^{-13} \text{m}^{-2/3}$, are given in Figure 9. The results show that $\{1, -2, 3, -5\}$ has a 16% probability of being incorrectly identified as $\{-2, 3, -5\}$, $\{1, -2\}$ has a 10% probability of being incorrectly identified as $\{1, -2, -5\}$, and $\{2, -6\}$ has an 11% probability of being incorrectly identified as $\{4, -4\}$. In contrast, in the APP-CNN network, the accuracy of OAM mode detection improves by up to 7%, and the related incorrect identification rate is reduced by up to 3%, which confirms the necessity of designing similar OAM superposition mode datasets. The accuracy improvement and decrease of incorrect identifications may vary with the transmission conditions.

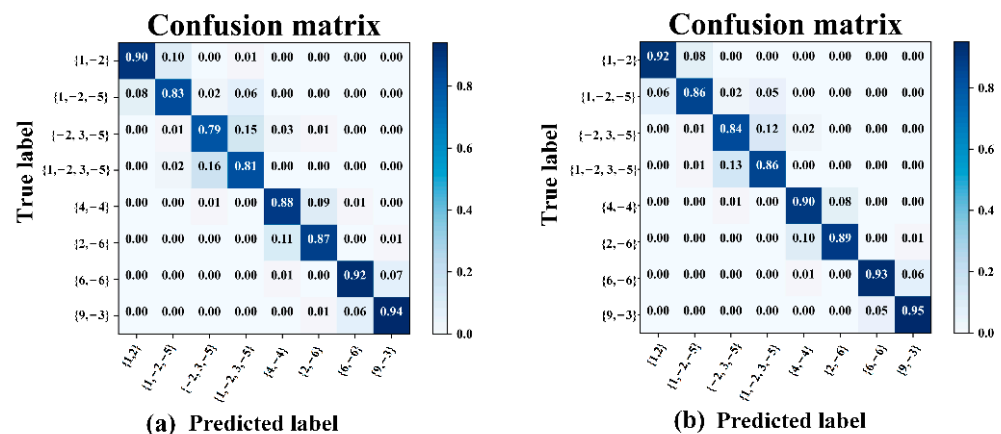


Figure 9. Confusion matrix (a) by ResNet18 (b) by AP-CNN.

Assuming the signal-to-noise ratio at the receiver side of the OAM communication system $\gamma_B = 10\text{dB}$, the system BER can be calculated based on the recognition accuracy. When the transmission distance is 2000 m, the demodulation performance of the CNN demodulator, ResNet18 demodulator, ResNet18 demodulator with a specified mapping relationship, and the AP-CNN demodulator under six different turbulence intensities C_n^2 and transmission distances, expressed as BER at the receiver end, are shown in Figure 10.

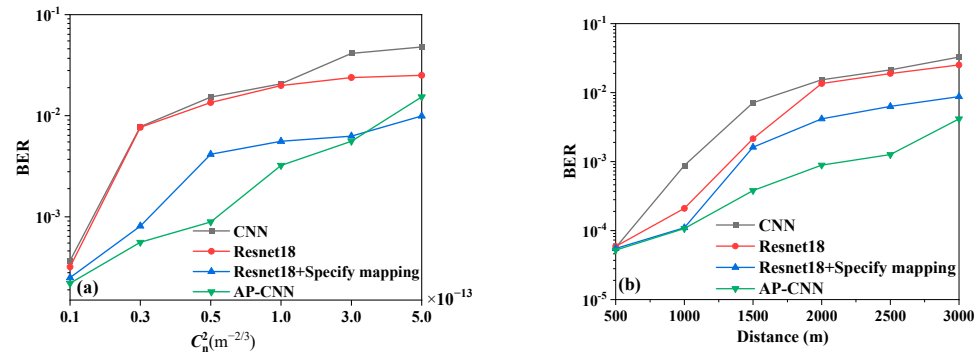


Figure 10. Performance comparison of four OAM demodulators with (a) atmospheric refractive index structure constant C_n^2 (b) transmission distance.

It can be seen from Figure 10 that when the turbulence intensity and transmission distance are certain, ResNet18 can mine more OAM intensity map information due to the presence of more convolution layers compared to CNN. Additionally, the BER of the OAM communication system is lower when using the ResNet18 demodulator compared to the CNN demodulator. The CNN structure used here consists of three convolution layers and two full connection layers. Each convolution network layer is composed of a convolutional layer, a batch normalization layer, and a maxpool layer. The layers are connected by a rectified linear unit (Relu), and each layer uses dropout. The dropout probability is set to 0.3. The convolutional layers of the first, second, and third convolution network layers contains 16 kernels of size 5×5 , 32 kernels of size 3×3 , and 64 kernels with size of 3×3 , respectively. The maxpool layer size of the three convolution network layers is 2×2 and the step size is 2; the difference is more obvious in a strong turbulence environment ($C_n^2 > 1.0 \times 10^{-13}\text{m}^{-2/3}$). Both the ResNet18 demodulator combined with specified mapping and the AP-CNN demodulator are optimized based on the ResNet18 demodulator, and both have lower BER than the ResNet18 demodulator. When $C_n^2 \leq 3.0 \times 10^{-13}\text{m}^{-2/3}$, the BER using the AP-CNN demodulator is lower than that using the ResNet18 demodulator combined with the specified mapping. However, when the turbulence is quite strong (e.g., $C_n^2 = 5.0 \times 10^{-13}\text{m}^{-2/3}$), the BER using the AP-CNN demodulator is higher than that of the ResNet18 demodulator with the specified mapping relationship. The reason for this is that the two optimization schemes (specifying the mapping relationship and introducing the attention pyramid) do not go in the same direction, as shown in Figure 11.

3.3. Analysis of Single-Mode OAM Mode Detection Based on AP-CNN

The detection accuracies of OAM mode with the ResNet18 network and AP-CNN are shown in Table 2 after 100 training cycles under different turbulence intensities and transmission distances. It can be concluded that the stronger the turbulence, the lower the accuracy of OAM detection. In addition, it can be seen from the data in Table 3 that no matter whether the transmission distance is 2000 m or 3000 m, under different turbulence intensities, AP-CNN has improved accuracy compared with ResNet18. In strong turbulence ($C_n^2 = 5.0 \times 10^{-13}\text{m}^{-2/3}$): the detection accuracy of AP-CNN is 85.2%, a 1.3% improvement compared with ResNet18, whereas in medium turbulence conditions ($C_n^2 = 5.0 \times 10^{-14}\text{m}^{-2/3}$), the detection accuracy of AP-CNN has a 3.4% improvement compared with ResNet18 at the transmission distance of 3000 m. When the transmission distance is reduced to 2000 m, the detection accuracy of AP-CNN shows 5.5% and 4.3% improvements compared with ResNet18 at $C_n^2 = 3.0 \times 10^{-13}\text{m}^{-2/3}$ and $C_n^2 = 5.0 \times 10^{-13}\text{m}^{-2/3}$, respectively. It should be emphasized that when the transmission

distance is long and there is strong turbulence, the light intensity distribution captured at the receiving end is damaged (because the vortex beam is greatly affected by the turbulent distortion during transmission), and the AP-CNN detection rate cannot be significantly improved.

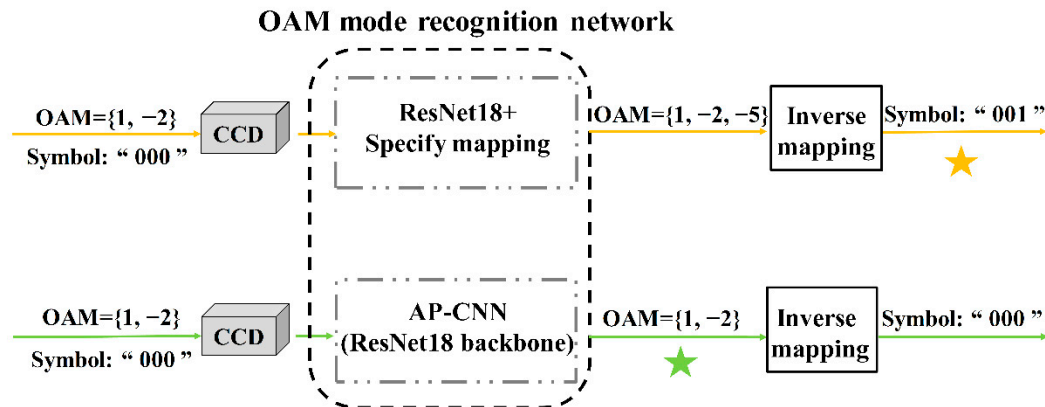


Figure 11. Comparison of the optimization direction of the AP-CNN and specified mapping relationship.

Table 3. Detection accuracy comparison of single mode OAM mode using ResNet18 and AP-CNN.

$C_n^2/(m^{-2/3})$	ResNet18		AP-CNN	
	2000 m	3000 m	2000 m	3000 m
$1.0 \times e^{-14}$	100.0%	100.0%	100.0%	100.0%
$3.0 \times e^{-14}$	98.9%	98.5%	99.2%	99.8%
$5.0 \times e^{-14}$	97.2%	92.4%	98.3%	95.8%
$1.0 \times e^{-13}$	93.5%	89.8%	96.6%	93.5%
$3.0 \times e^{-13}$	85.9%	84.2%	91.4%	87.1%
$5.0 \times e^{-13}$	84.6%	83.9%	88.9%	85.2%

A comparison of the demodulation performance of CNN, ResNet18, ResNet18 combined with plane wave interference, and AP-CNN combined with plane wave interference with a 2000 m transmission distance under six different turbulence intensities is shown in Figure 12a. The performance of the communication link under different transmission distances when $C_n^2 = 5.0 \times 10^{-14} m^{-2/3}$ is compared in Figure 12b.

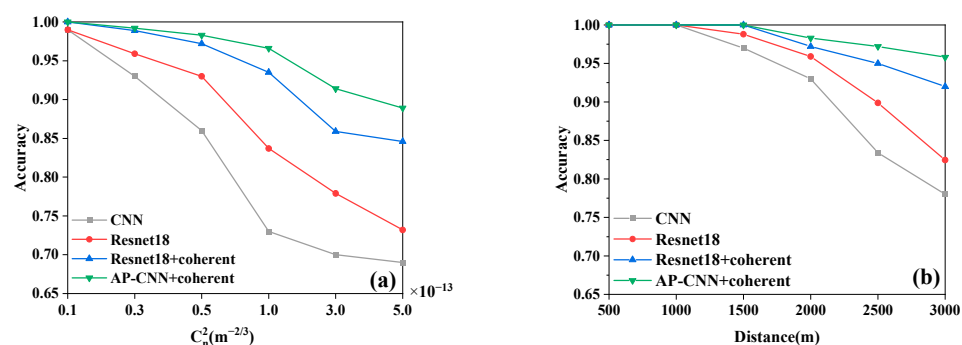


Figure 12. Performance of four OAM demodulators for detecting single-mode vortex light (a) under different turbulence intensities and (b) different transmission distances.

It can be seen from Figure 12 that no matter which demodulation scheme is adopted, when the transmission distance is fixed, the stronger the turbulence, the lower the accuracy of the OAM mode detection at the receiving end. When the turbulence intensity is constant, the longer the transmission distance, the lower the OAM mode detection rate. The detection accuracy of the CNN demodulator with three convolutional layers, a simple structure, and

no plane wave interference is low. The ResNet18 has a more complex network structure, including 17 layers of convolution, and has a stronger ability to mine light intensity information. However, it directly identifies the light intensity distribution of single-mode vortex light without interference. Although its detection accuracy is improved compared with that of CNN, it is still not satisfactory. After plane wave interference, the plane wave interferogram collected at the receiving end has more distinguishable characteristics, and the OAM mode detection accuracy of the ResNet18 demodulator is significantly improved compared with no interference. The AP-CNN adds a dual-path structure combined with an attention pyramid after the ResNet18 network. As such, the OAM mode detection accuracy is further improved compared to that of ResNet18. This reveals that the dual path with an attention pyramid is beneficial for single mode detection accuracy.

3.4. Discussions

In this work, an AP-CNN OAM mode detection method was described. By adding a dual path network with an attention pyramid to the backbone ResNet18, the AP-CNN network was constructed. The effects of turbulence intensity and transmission distance on the improvement of OAM mode detection accuracy were numerically analyzed.

We first studied the performance of the AP-CNN on multi-mode OAM mode detection with similar light intensity distributions. Comparisons of AP-CNN with ResNet18 under different degrees of atmospheric turbulence and transmission distances verified the improvement of the recognition accuracy due to the presence of a dual path structure with attention pyramid. The results reveal that the recognition accuracy increased to some extent with increasing turbulence intensity.

The demodulator performance of the CNN with three convolution layers and two full connection layers, i.e., ResNet18, ResNet18 + Specify mapping and AP-CNN, on a multi-mode vortex beam are studied. When $C_n^2 \leq 3.0 \times 10^{-13} \text{m}^{-2/3}$, the BER using AP-CNN demodulator was lower than with the other three methods.

In addition, OAM mode detection in single mode with a large topological charge was simulated under medium and strong turbulence intensities, i.e., C_n^2 ranges from 1.0×10^{-14} to 5.0×10^{-13} at 2000 m and 3000 m transmission distances. A comparison between 'ResNet18' and 'ResNet18 + coherent' verified the effect of plane wave interference. The single mode recognition by AP-CNN showed an accuracy improvement of up to 5.5% compared with ResNet18 when $C_n^2 = 3.0 \times 10^{-13} \text{m}^{-2/3}$ at a 2000 m transmission distance, indicating that the former detection method has strong detailed extraction and learning capabilities for dense interference fringes of single mode vortex beams with large topological charges.

4. Conclusions

In this paper, an OAM mode recognition technique based on AP-CNN is proposed. Utilizing ResNet18 as the backbone of the AP-CNN, a dual-path algorithm structure, including a top-down feature path and a bottom-up attention path, is added. Based on the dual-path algorithm structure combined with the attention pyramid, low-level detailed information of the similar light intensity map is fully utilized. In our simulated experiments, the size of the light intensity distribution map of the OAM beam was set to 128×128 and was input into the ResNet18 network for training. Then, the output feature maps of the third, fourth, and fifth layers of ResNet18 were selected to build a pyramidal hierarchy. After supervised training with a large OAM communication dataset with different turbulence conditions, the recognition accuracy and the BER were numerically determined. The simulation results showed that the AP-CNN achieved greatly improved OAM mode detection accuracy and demodulation performance compared with the ResNet18 network. When the turbulence was weak, the optimization effect of AP-CNN was not obvious, i.e., a 0.6% improvement, while when the turbulence was strong, the optimization was clear, with an improvement of about 2.3%. The OAM detection accuracy of the AP-CNN was up to 5.5% higher than that of ResNet18 at 2 km with strong turbulence. This technique

has significant applications in communication, target detection, and radar imaging. Due to the limitations of experimental conditions, our research was only based on a simulated intensity distribution dataset, and light intensity information was collected without phase information. The training and analysis of the real turbulence OAM communication data under different conditions, as well as the phase information, will be the focus of future work by our team.

Author Contributions: Conceptualization and methodology, T.Q. and Z.Z.; software, Z.Z.; validation, T.Q. and Y.Z.; formal analysis, Y.Z. and J.W.; editing, Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (62071359), Scientific Research Program Funded by Shaanxi Provincial Education Department (19JK0673), the Open Foundation of Laboratory of Pinghu, Pinghu, China, Postdoctoral Science Foundation in Shaanxi Province and the Fundamental Research Funds for the Central Universities.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Allen, L.; Beijersbergen, M.W.; Spreeuw, R.J.C.; Woerdman, J.P. Orbital angular momentum of light and the transformation of Laguerre-Gaussian laser modes. *Phys. Rev. A* **1992**, *45*, 8185–8189. [[CrossRef](#)] [[PubMed](#)]
- Djordjevic, I.B. Deep-space and near-Earth optical communications by coded orbital angular momentum (OAM) modulation. *Opt. Express* **2011**, *19*, 14277–14289. [[CrossRef](#)] [[PubMed](#)]
- Huang, H.; Milione, G.; Lavery, M.; Xie, G.; Ren, Y.; Cao, Y.; Ahmed, N.; Nguyen, T.; Daniel, A.; Li, M.; et al. Mode division multiplexing using an orbital angular momentum mode sorter and MIMO-DSP over a graded-index few-mode optical fiber. *Sci. Rep.* **2015**, *5*, 14931. [[CrossRef](#)] [[PubMed](#)]
- Mesquita, P.; Jesus-Silva, A.; Fonseca, E.; Hixjmann, N. Engineering a square truncated lattice with light's orbital angular momentum. *Opt. Express* **2011**, *19*, 20616–20621. [[CrossRef](#)]
- Dai, K.; Gao, C.; Zhong, L.; Na, Q.; Wang, Q. Measuring OAM states of light beams with gradually-changing-period gratings. *Opt. Lett.* **2015**, *40*, 562–565. [[CrossRef](#)]
- Denisenko, V.; Shvedov, V.; Desyatnikov, A.S.; Neshev, D.N.; Krolikowski, W.; Volyar, A.; Soskin, M.; Kivshar, Y.S. Determination of topological charges of polychromatic optical vortices. *Opt. Express* **2009**, *17*, 23374–23379. [[CrossRef](#)]
- Berkhout, G.; Lavery, R.; Courtial, R.; Beijersbergen, R.; Padgett, R. Efficient sorting of orbital angular momentum states of light. *Phys. Rev. Lett.* **2010**, *105*, 153601. [[CrossRef](#)]
- Guo, C.; Lu, L.; Wang, H. Characterizing topological charge of optical vortices by using an annular aperture. *Opt. Lett.* **2009**, *34*, 3686–3688. [[CrossRef](#)]
- Krenn, M.; Fickler, R.; Fink, M.; Handsteiner, J.; Malik, M.; Scheidl, T.; Ursin, R.; Zeilinger, A. Communication with spatially modulated light through turbulent air across Vienna. *New J. Phys.* **2014**, *16*, 113028. [[CrossRef](#)]
- Krenn, M.; Handsteiner, J.; Fink, M.; Fickler, R.; Ursin, R.; Malik, M.; Zeilinger, A. Twisted light transmission over 143 km. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 13648–13653. [[CrossRef](#)]
- Knutson, E.; Lohani, S.; Danaci, O.; Huver, S.; Glasser, R. Deep learning as a tool to distinguish between high orbital angular momentum optical modes. In Proceedings of the Optics and Photonics for Information Processing X, San Diego, CA, USA, 14 September 2016; Volume 9970. [[CrossRef](#)]
- Doster, T.; Watnik, A. Machine learning approach to OAM beam demultiplexing via convolutional neural networks. *Appl. Opt.* **2017**, *56*, 3386–3396. [[CrossRef](#)] [[PubMed](#)]
- Xiong, W.; Luo, Y.; Liu, J.; Huang, Z.; Fan, D. Convolutional neural network assisted optical orbital angular momentum identification of vortex beams. *IEEE Access* **2020**, *8*, 193801–193812. [[CrossRef](#)]
- Wang, Z.; Dedo, M.I.; Guo, K.; Zhou, K.; Shen, F.; Sun, Y.; Liu, S.; Guo, Z. Efficient recognition of the propagated orbital angular momentum modes in turbulences with the convolutional neural network. *IEEE Photonics J.* **2019**, *11*, 1–14. [[CrossRef](#)]
- Li, J.; Zhang, M.; Wang, D. Adaptive demodulator using machine learning for orbital angular momentum shift keying. *IEEE Photonics Technol. Lett.* **2017**, *29*, 1455–1458. [[CrossRef](#)]
- Li, J.; Zhang, M.; Wang, D.; Wu, S.; Zhan, Y. Joint atmospheric turbulence recognition and adaptive demodulation technique using the CNN for the OAM-FSO communication. *Opt. Express* **2018**, *26*, 10494–10508. [[CrossRef](#)]
- Tian, Q.; Li, Z.; Hu, K.; Zhu, L.; Pan, X.; Zhang, Q.; Wang, Y.; Tian, F.; Yin, X.; Xin, X. Turbo-coded 16-ary OAM shift keying FSO communication system combining the CNN-based adaptive demodulator. *Opt. Express* **2018**, *26*, 27849–27864. [[CrossRef](#)] [[PubMed](#)]
- Zhao, Q.; Hao, S.; Wang, Y.; Wang, L.; Wan, X.; Xu, C. Mode recognition of misaligned orbital angular momentum beams based on convolutional neural network. *Appl. Opt.* **2018**, *57*, 10152–10158. [[CrossRef](#)]

19. Wang, X.; Qian, Y.; Zhang, J.; Ma, G.; Zhao, S.; Liu, R.; Li, H.; Zhang, P.; Gao, H.; Huang, F. Learning to recognize misaligned hyperfine orbital angular momentum modes. *Photonics Res.* **2021**, *9*, I0001–I0006. [[CrossRef](#)]
20. Cao, M.; Yin, Y.; Zhou, J.; Tang, J.; Cao, L.; Xia, Y.; Yin, J. Machine learning based accurate recognition of fractional optical vortex modes in atmospheric environment. *Appl. Phys. Lett.* **2021**, *119*, 141103. [[CrossRef](#)]
21. Li, Z.; Tian, Q.; Zhang, Q.; Wang, K.; Xin, X. An improvement on the CNN-based OAM demodulator via conditional generative adversarial networks. In Proceedings of the 2019 18th International Conference on Optical Communications and Networks (ICOON), Huangshan, China, 5–8 August 2019; pp. 1–3. [[CrossRef](#)]
22. Zhao, Q.; Hao, S.; Wang, Y.; Wang, L.; Xu, C. Orbital angular momentum recognition based on diffractive deep neural network. *Opt. Commun.* **2019**, *443*, 245–249. [[CrossRef](#)]
23. Park, S.; Cattell, L.; Nichols, J.; Watnik, A.; Doster, T.; Rohde, G. De-multiplexing vortex modes in optical communications using transport-based pattern recognition. *Opt. Express.* **2018**, *26*, 4004–4022. [[CrossRef](#)] [[PubMed](#)]
24. Jiang, S.; Chi, H.; Yu, X.; Zheng, S.; Jin, X.; Zhang, X. Coherently demodulated orbital angular momentum shift keying system using a CNN-based image identifier as demodulator. *Opt. Commun.* **2019**, *435*, 367–373. [[CrossRef](#)]
25. Sun, R.; Guo, L.; Cheng, M.; Li, J.; Yan, X. Identifying orbital angular momentum modes in turbulence with high accuracy via machine learning. *J. Opt.* **2019**, *21*, 075703. [[CrossRef](#)]
26. Dedo, M.; Wang, Z.; Guo, K.; Guo, Z. OAM mode recognition based on joint scheme of combining the Gerchberg–Saxton (GS) algorithm and convolutional neural network (CNN). *Opt. Commun.* **2019**, *456*, 124696. [[CrossRef](#)]
27. Pan, S.; Pei, C.; Liu, S.; Wei, J.; Wu, D. Measuring orbital angular momentums of light based on petal interference patterns. *OSA Continuum.* **2018**, *1*, 451–461. [[CrossRef](#)]
28. Xiao, T.; Xu, Y.; Yang, K.; Zhang, J.; Peng, Y.; Zhang, Z. The application of two-level attention models in deep convolutional neural network for fine-grained image classification. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015), Boston, MA, USA, 7–12 June 2015; pp. 842–850. [[CrossRef](#)]
29. Ding, Y.; Ma, Z.; Wen, S.; Xie, J.; Chang, D.; Si, Z.; Wu, M.; Ling, H. AP-CNN: Weakly supervised attention pyramid convolutional neural network for fine-grained visual classification. *IEEE Trans. Image Process.* **2021**, *30*, 2826–2836. [[CrossRef](#)]
30. Lin, T.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object recognition. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125. [[CrossRef](#)]
31. Tian, Q.; Zhao, Y.; Li, Y.; Chen, J.; Qin, K. Multiscale building extraction with refined attention pyramid networks. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [[CrossRef](#)]
32. Liu, S.; Huang, D.; Wang, Y. Adaptive NMS: Refining detection in a crowd. In Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2019), Long Beach, CA, USA, 16–20 June 2019; pp. 6452–6461. [[CrossRef](#)]
33. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern. Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)]
34. Ghiasi, G.; Lin, T.; Le, Q. Dropblock: A regularization method for convolutional networks. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018), Salt Lake City, UT, USA, 18–22 June 2018; pp. 10727–10737. [[CrossRef](#)]
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]