applied optics

# Free-space 16-ary orbital angular momentum coded optical communication system based on chaotic interleaving and convolutional neural networks

Shimaa A. El-Meadawy,[1,*] Hossam M. H. Shalaby,[2] Nabil A. Ismail,[3] Fathi E. Abd El-Samie,[1,4] and Ahmed E. A. Farghal[5]

[1]Department of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf 32952, Egypt
[2]Electrical Engineering Department, Faculty of Engineering, Alexandria University, Alexandria 21544, Egypt
[3]Department of Computer Science and Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf 32952, Egypt
[4]Department of Information Technology, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh 84428, Saudi Arabia
[5]Electrical Engineering Department, Faculty of Engineering, Sohag University, 82524 Sohag, Egypt
*Corresponding author: shimaaelmeadawy@yahoo.com

Received 20 February 2020; revised 18 May 2020; accepted 29 May 2020; posted 1 June 2020 (Doc. ID 390931); published 5 August 2020

**Recently, orbital angular momentum (OAM) rays passing through free space have attracted the attention of researchers in the field of free-space optical communication systems. Throughout free space, the OAM states are subject to atmospheric turbulence (AT) distortion leading to crosstalk and power discrepancies between states. In this paper, a novel chaotic interleaver is used with low-density parity-check coded OAM-shift keying through an AT channel. Moreover, a convolutional neural network (CNN) is used as an adaptive demodulator to enhance the performance of the wireless optical communication system. The detection process with the conjugate light field method in the presence of chaotic interleaving has a better performance compared to that without chaotic interleaving for different values of propagation distance. Also, the viability of the proposed system is verified by conveying a digital image in the presence of distinctive turbulence conditions with different error correction codes. The impacts of turbulence strength, transmission distance, signal-to-noise ratio (SNR), and CNN parameters and hyperparameters are investigated and taken into consideration. The proposed CNN is chosen with the optimal parameter and hyperparameter values that yield the highest accuracy, utmost mean average precision (MAP), and the largest value of area under curve (AUC) for the different optimizers. The simulation results affirm that the proposed system can achieve better peak SNR values and lower mean square error values in the presence of different AT conditions. By computing accuracy, MAP, and AUC of the proposed system, we realize that the stochastic gradient descent with momentum and the adaptive moment estimation optimizers have better performance compared to the root mean square propagation optimizer.** 

https://doi.org/10.1364/AO.390931

## 1. INTRODUCTION

Attaining excessive data transmission capacity and overcoming the crunch problem of the unresolved bandwidth are the utmost crucial concerns of the photonics community [1]. An exemplary model for enhancing both the transmission capacity and the spectral efficiency of lightwave systems depends on multiplexing miscellaneous autonomous data channels. The different data channels can be localized on a diversity of polarizations, wavelengths, or spatial channels, congruous to different categories of division multiplexing [2]. Recently, a great deal of curiosity has been given to space-division multiplexing (SDM) for capacity augmentation in optical systems along with the existing multiplexing techniques. Mode-division multiplexing (MDM) is a distinctive SDM case, where every mode can convey an autonomous data channel [3]. Orbital angular momentum (OAM), exposed and verified by Allen *et al.* in 1992 [4], is a prospective candidate for an MDM system with an orthogonal mode elementary set. OAM is the circumstance of the spatial dispersal of the electric field around the beam axis, resulting in a helical phase front. Laguerre–Gaussian (LG) beams are considered as an extraordinary subcategory among all beams of OAM, and

their fundamental distribution is counted from the actuality that they are paraxial eigen-solutions of the wave equation in both free-space and cylindrical coordinates [5]. To meet the ever-increasing demands of wireless communication systems, alternative technologies are recommended by the implementation of OAM. There are three types of the implementation of free-space optical (FSO) communication systems concentrating on the nature of OAM, including OAM-division multiplexing (OAM-DM) [6], OAM multicasting [7] and OAM shift keying (OAM-SK) [5].

Recently, to cope with the rapid growth of the machine intelligence (MI), deep learning (DL) methods have been successfully employed in several applications such as image classification and speaker recognition [8]. The utilization of convolutional neural networks (CNNs) has achieved great success in the field of computer vision applications [9]. The utilization of CNNs in different fields has motivated us to use them in adaptive demodulation. The major task of the OAM adaptive demodulator is to categorize the received OAM beam images to get the original information of the OAM modes. When OAM demodulation is performed with a neural network, a high recognition rate is required, which means that a high-quality training set and a more complex network structure are required to train an excellent model [10]. Although the recognition rate may reach extreme levels, the bit error rate (BER) may not preserve the realistic communication process.

Forward error correction (FEC) coding is an efficient tool that can be used to enhance the reliability of data communication and get a minimal value of BER, and this is done by adding redundant bits before transmitting the data. The FEC provides the receiver with the ability to correct the errors with no supplementary channel to ask for data retransmission. To enhance the performance of OAM FSO systems, some codes such as Reed–Solomon (RS) codes [11], low-density parity-check (LDPC) codes [12], and Turbo codes [10] have been introduced.

The major objective of this study is to present an OAM FSO system that adopts a novel CNN-based technique for modulation and coding classification with the aid of a chaotic interleaver to reduce the BER. The proposed system consists of three parts. The first part is the chaotic interleavering with LDPC coding and OAM modulation. The second part is the free-space atmospheric turbulence (AT) channel. The third part is chaotic deinterleavering and OAM demodulation. The major objectives of this system are decreasing the BER of the communication process and increasing the peak SNR (PSNR) for image communication in the presence of different turbulence parameters. This is accomplished through successful demodulation and decoding processes. The adaptive moment estimation (ADAM), root mean square propagation (RMSProp), and stochastic gradient descent with momentum (SGDM) equalizers are used and compared to get the optimal parameter and hyperparameter values of the CNN for the three optimizers in order to achieve the highest demodulation accuracy.

The main original contribution of this study is presenting a novel coded OAM-SK–FSO system with chaotic interleaving. In addition, the paper presents an alternative CNN architecture that is designed based on the optimal values of both the parameters and hyperparameters of the network using the ADAM, RMSProp, or SGDM optimizers. The proposed CNN model achieves the minimum loss and the highest value of accuracy for different batch sizes, epochs, and learning rates based on optimization. The rest of this paper is organized as follows. In Section 2, we present related work on OAM in the presence of AT. In Section 3, we describe the components of the proposed system. In Section 4, we demonstrate the process of image transmission with the proposed system. In Section 5, we explain the CNN architecture and show all simulation results with different parameters and hyperparameters based on optimization.

## 2. RELATED WORK

In [13], a deep CNN model based on compensation of turbulence was proposed for the purpose of modifying the distorted vortex beam and enhancing the performance of OAM multiplexing systems by increasing the accuracy from 39.52% to 98.34% for strong turbulence strengths. In [14], a trade-off was made between the system's computational complexity and the recognition efficiency by introducing a particularly designed CNN architecture to professionally realize the OAM mode. Numerical simulation indicates that the recognition accuracy is increased to 96.25% even with elongated distance and high turbulence. In [15], the decoding accuracy of 16-ary-OAM-SK based on CNNs in an underwater optical communication system was experimentally demonstrated; the results showed that the decoder accuracy was more than 99% in clean water, and it needed more pixels to reach an accuracy of more than 99% in turbid water.

In [16], the simulation results showed that the CNN decoders achieved an excellent performance of nearly 100% within dozens of meters or in the presence of weak to moderate turbulence, and 93% with strong turbulence or at a distance of 60 m through an oceanic turbulence channel. The average BER (ABER) performance of the CNN-based demodulator outperforms that of the conventional conjugate demodulator by several orders of magnitude. The ABER of a noisy system approaches the saturation level when the instantaneous SNR is about 26 dB greater than the pass loss, as in [17]. An adaptive demodulator based on machine learning for optical beams transferring OAM over free-space turbulence channels was presented, and the value of the CNN demodulator error rate (DER) was 0.86% in the case of 1000 m for an 8-OAM system in the presence of strong turbulence, as shown in [18]. In [19], by combining five bidirectional recurrent neural networks (B-RNNs) into one model, the multiple time interval feature-learning network (MTIFLN) becomes strongly able to extract the long-term traffic characteristics at different time intervals. In addition, the MTIFLN stacked architecture helps to diminish prediction errors by a resampling mechanism.

## 3. PROPOSED COMMUNICATION SYSTEM

Figure 1 demonstrates the implementation of the coded OAM-SK–FSO system with chaotic interleavering and adaptive CNN-based demodulation and decoding. The transmitter sequence is ordered as follows. First, input data are translated into bit streams and coded via Turbo or LDPC code. Then, the encoded data are interleaved using a chaotic interleaver. Subsequently, an enhanced mapping scheme is applied on the

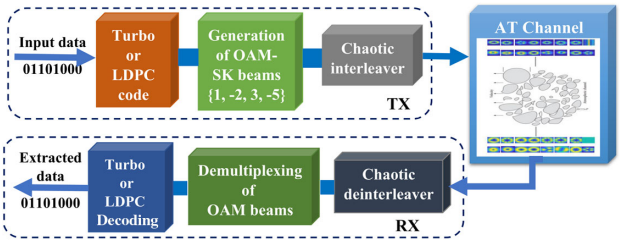


**Fig. 1.** 16-ary OAM coded FSO communication system based on chaotic interleaving.

interleaved coded data through the use of a spatial light modulator (SLM) to yield super-imposed OAM states. Through the mapping process, each quadruple bit is transformed into one of 16 states that are obtained from the superposition of four OAM basic states due to OAM beams' orthogonality. Those orthogonal beams conveying the raw binary interleaved coded data are passed through the AT channel to the receiver. At the receiver, the corrupted images shown in Fig. 1 are passed through the charge coupled device's (CCD) image sensor. After that, the received data are passed directly to the CNN, which can be used as a switch with two options: a demodulator or a classifier to identify OAM states. The conjugate mode sorting technique is used to determine the OAM mode of a detected beam. The demapping scheme is implemented by multiplying the transmitted coded data with the conjugates of the original beams. Finally, we apply the chaotic deinterleaver, and then perform decoding via LDPC or Turbo decoding to recover the original data.

The OAM-SK image recognition (IR) capability is restricted by the response times of the SLM and the CCD sensor. Due to the low scanning speed and the low frame rate of CCD-based cameras, long data acquisition times affect the performance. Despite these deficiencies, SLM and CCD cameras achieve better performance in many applications of IR-based OAM systems. They are used in OAM holography to achieve ultra-high capacity and high security [20]. CCD cameras have been used in stable OAM mode fiber laser systems with a polarization-maintaining fiber (PMF) structure to create an OAM mode fiber laser that is resistant to environmental disturbances without any polarization controller [21].

CCD cameras have been used in OAM propagation with cylindrical vector beams in an annular core photonic crystal fiber (AC-PCF). This system has many potential applications in SDM, optical sensing, and trapping [22]. These cameras have also been implemented in OAM-SK underwater wireless optical communication (UWOC) systems in oceanic turbulence channels [17]. In low-frequency heterodyne interferometry, CCD cameras are used to measure the wave front distortion of optical beams induced by AT [23].

### A. OAM Transmission through AT Channel

One of the common formulas of electromagnetic beams, which transfers OAM modulation, is known as the LG beam. The Gaussian beam expression is used to represent the LG beam. The expression for the LG beam that conveys OAM is given by [5,17]

$$G_{\mathrm{LG}(l,p)}(r,\theta,z) = \frac{D}{\omega(z)} \cdot e^{-\frac{ik_o r^2 z}{4\omega^2(z)}} \cdot e^{-\frac{r^2}{\omega^2(z)}} \cdot \left(\frac{r\sqrt{2}}{\omega(z)}\right)^{|L|} \times l_p^{|L|} l_p^{l}(2r^2/w^2(z)) \cdot e^{-il\theta} \times e^{-i(2p+|L|+1)\tan^{-1}\left(\frac{z}{z_R}\right)}, \tag{1}$$

where $D$ is a normalization constant, $r$ is the radial distance from $z$, $L$ is the topological charge and the number of twists in the helix wavefront, $p$ is the radial index, $(r,\theta,z)$ are the cylindrical coordinates, $k_o = 2\pi/\lambda$ is the wave number, $\lambda$ is the optical wavelength, and $l_p^{|L|}$ is the generalized Laguerre polynomial. The beam radius of the fundamental Gaussian beam at distance $z$ is given by

$$\omega(z) = \omega_0\sqrt{1+\left(\frac{z}{z_R}\right)^2}. \tag{2}$$

The Rayleigh range $z_R$ is given by

$$z_R = (\pi\omega_0^2)/\lambda, \tag{3}$$

where $\omega_0$ is the beam waist at $z=0$. From the previous equations, we can get an approximation for the beam radius by putting $\omega_0$ as a constant and setting $\lambda = 1550$ nm. The term $z_R$ will be constant, and the beam radius satisfies the relation

$$\omega(z) \propto \sqrt{1+z^2}. \tag{4}$$

Depending on radial indices $p$ and the orthogonality property of LG beams, we can use Eq. (1) to get the 16-ary OAM states using different values of $L$, and then adopt these states to map the coded binary sequence to these states. After mapping, we transmit the data through weak, moderate, and strong AT channels with different propagation distances. The equation that controls the transmitted data through the turbulence channel is [24]

$$\psi(h) = \frac{2(\alpha\beta)^{\frac{\alpha+\beta}{2}}}{\Gamma(\alpha)\Gamma(\beta)} \times h^{\frac{\alpha+\beta}{2}-1} \times K_{\alpha-\beta}\left(2\sqrt{\alpha\beta h}\right), \tag{5}$$

where $\psi(h)$ refers to the probability density function (pdf) of AT $h$, $\alpha$ represents the effective number of large-scale eddies of the scattering process, and $\beta$ represents the effective number of small-scale eddies of the scattering process. The values of $\alpha$ and $\beta$ are calculated with the help of the following equations:

$$\alpha = \left[\exp\left(0.49\rho_v^2/(1+1.11\rho_v^{12/5})^{7/6}\right)-1\right]^{-1},$$
$$\beta = \left[\exp\left(0.51\rho_v^2/\left(1+0.69\rho_v^{12/5}\right)^{5/6}\right)-1\right]^{-1}, \tag{6}$$

where $\rho_v^2 = 1.23c_n^2 k_o^{\frac{7}{6}} l_p^{\frac{11}{6}}$ is the variance of the irradiance fluctuations, $I$ is the normalized received irradiance, $K_n(\cdot)$ is the modified Bessel function of the second kind of order $n$, $\Gamma(\cdot)$ represents the Gamma function, $c_n^2$ is the AT strength, and $l_p$ is the propagation distance. Due to the existence of the modified Bessel function in Eq. (5), a significant mathematical complexity in dealing directly with the $\psi(h)$ of the

channel exists. Consequently, we use the Meijer G-function to express the modified Bessel function and turbulence channel as follows [24]:

$$K_{\alpha-\beta}\left(\sqrt{\alpha\beta h}\right)=\frac{1}{2}G_{0,2}^{2,0}\left(\alpha\beta h\,\middle|\,\tfrac{\alpha-\beta}{2},\,\tfrac{\beta-\alpha}{2}\right),$$

$$\psi(h)=\frac{(\alpha\beta)^{\frac{\alpha+\beta}{2}}}{\Gamma(\alpha)\Gamma(\beta)}h^{\frac{\alpha+\beta}{2}-1}G_{0,2}^{2,0}\left(\alpha\beta h\,\middle|\,\tfrac{\alpha-\beta}{2},\,\tfrac{\beta-\alpha}{2}\right). \tag{7}$$

After traveling through the turbulence channel, the distorted data arrive at the receiver. At the receiver, the detection is performed by multiplying the received data with the conjugates of the originally used OAM states. The result of this product is estimated according to one of the two following decisions.

1. If the real and the imaginary parts equal zero, then the present mode number is 16.
2. If only the imaginary part equals zero, then it is the correct state, and the mode number is detected successfully.

We demap all received data to different OAM states, and then decode the data to get the original bit sequence:

$$\begin{aligned}&\langle G_{\mathrm{LG}(l_m,p)}(r,\theta,z)_m\times G^{*}_{\mathrm{LG}(l_k,p)}(r,\theta,z)_k\rangle\\&=\int G_{\mathrm{LG}(l_m,p)}(r,\theta,z)_m\times G^{*}_{\mathrm{LG}(l_k,p)}(r,\theta,z)_k r\,\mathrm{d}r\,\mathrm{d}\theta\\&=\begin{cases}\int\left|G_{\mathrm{LG}(l_m,p)}\right|^2 r\,\mathrm{d}r\,\mathrm{d}\theta; & \text{for } m=k,\\ 0; & \text{for } k=16.\end{cases}\end{aligned} \tag{8}$$

Using the conjugate light field detection method, the estimation of the value of $l$ is performed based on the orthogonality property of OAM states. Assuming that the transmitted LG beam is $G_{\mathrm{LG}(l_m,p)}(r,\theta,z)_m$, we can compute the product between the selected LG beam and each of the modes $G^{*}_{\mathrm{LG}(l_k,p)}(r,\theta,z)_k$, where $m$ and $k$ characterize the $m$th and $k$th beams, respectively:

$$G_{\mathrm{Detection}}(r,\theta,z)_m=G_{\mathrm{LG}(l_m,p)}(r,\theta,z)_m G^{*}_{\mathrm{LG}(l_k,p)}(r,\theta,z)_k. \tag{9}$$

At the receiver, the arriving photonic beam is developed by the product of the vortex field with all the available OAM beams, and the result is given in the following equation:

$$\begin{aligned}&G_{\mathrm{Detection}}(r,\theta,z)_m\\&=G_{\mathrm{LG}(l_m,p)}(r,\theta,z)_m\sum_k G^{*}_{\mathrm{LG}(l_k,p)}(r,\theta,z)_k.\end{aligned} \tag{10}$$

The total error rate is evaluated by multiplying the calculated error rate of the proposed model with the error rate of the turbulence channel. Then, the error rate value is calculated by taking the average value as follows:

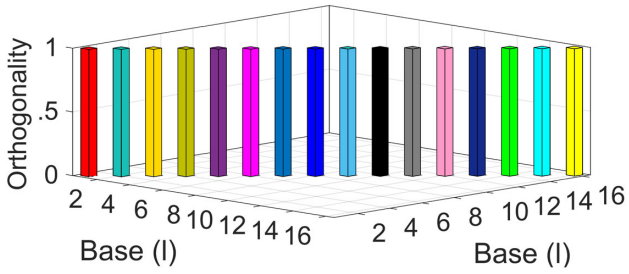


**Fig. 2.** Orthogonality between the bases designed by OAM beams with a varity of $l$ values.

$$\begin{aligned}\mathrm{BER}&=G_{\mathrm{LG}(l_m,p)}(r,\theta,z)_m\sum G^{*}_{\mathrm{LG}(l_k,p)}(r,\theta,z)_k\\&\quad-G_{\mathrm{LG}(l_m,p)}(r,\theta,z)_m.\end{aligned} \tag{11}$$

Figure 2 reveals this principle through computing the orthogonality between OAM beams with different values of $l$ by taking the inner product of the two equivalent optical fields [5,17].

### B. Coding with LDPC and Turbo Codes

Coding is one of the essential techniques that make near-capacity operation conceivable. By encoding and decoding of data, error detection and correction can be realized. The LDPC and Turbo codes are common types of near-capacity codes that allow the noise threshold to be very close to the theoretical Shannon limit for a symmetric memoryless channel. Turbo codes are error-correcting codes that are used to enhance the reliability of communication. They achieve an impressive efficiency by encoding and decoding algorithms with relatively low complexity. Turbo coding is adopted to reduce continuous errors in the transmission process for more effective retrieval of information. The configuration of the encoder depends on parallel concatenation of two convolutional coders separated by an interleaver, while the decoding process is performed according to an iterative procedure to decode the received data from the channel. LDPC codes are structured by exploiting a generator matrix $\mathbf{H}$ with slight non-zero elements [25]. In the case of binary codes, $\mathbf{H}$ has a large number of zeros and few ones. To be consistent with the Gallager demarcation, the following conditions must be satisfied.

1. In each row and column, the matrix $\mathbf{H}$ has $R$ and $V$ ones, respectively.
2. To avoid the cycles in each dual rows (or columns), the occurrence of ones should not be more than one location.
3. The $R$ and $V$ should be as small as possible compared to the codeword length $L$, and the LDPC code is defined as $C_{\mathrm{LDPC}}(L,V,R)$.

For a control matrix with $m$ columns and $m-k$ rows, the following condition must be satisfied:

$$V\times m=R\times(m-k), \tag{12}$$

and the code rate (CR) is given by

$$\mathrm{CR}=1-\frac{V}{R}. \tag{13}$$

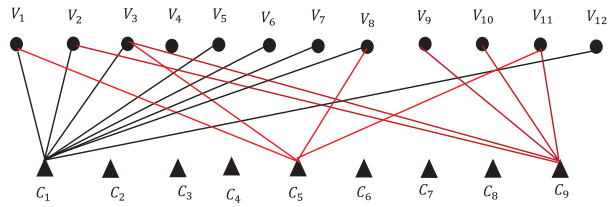


**Fig. 3.** Tanner graph of an LDPC code.

The Tanner graph shown in Fig. 3 is a bipartite graph revealing the relationship between two nodes, namely, variable bit nodes $V$ that signify the launched symbols, and the parity nodes $C$ that represent the emanating symbols.

For every bit node, a code symbol exists, and for every parity node, a parity equation exists. A straight line is connected between the bit node and the parity node only if that bit is involved in the parity equation. Some parity node equations are given by

$$C_1 = V_1 + V_2 + V_3 + V_5 + V_6 + V_7 + V_8 + V_{12},$$
$$C_9 = V_2 + V_3 + V_9 + V_{10} + V_{11}, \quad \textbf{(14)}$$

where $V$ represents the launched nodes, and $C$ represents the emanating nodes. The LDPC decoding may be "soft decision" or "hard decision" decoding, and it is used to reconstruct the primary message in the absence of cycles in the bipartite graph. If $\mathbf{q}$ is an obtained vector, the resultant equation is $\mathbf{S} = \mathbf{q} \times \mathbf{H}^T$.

### C. Chaotic Interleaver

To protect the transmission of data over turbulence channels, the necessity for error correcting codes exists to correct the errors due to AT. Due to the burst nature of communication channels, interleaving is utilized to reorganize the transmitted coded data and let errors propagate over various codewords. Initially, the block interleaver was the simplest and most popular one. Due to its deficiency in dealing with two-dimensional (2D) error bursts, a progressive interleaver was used. With this interleaver, the bits are arranged in a 2D format and after that, the randomization is performed with a chaotic Baker map. The spatial dimensions of chaotic maps could vary from one-dimensional to higher dimensions. With one-dimensional maps, the randomization operation is performed on a single axis, while with higher-dimension ones, this unique axis is not sufficient. The chaotic interleaver is an effectual tool to randomize the items in a rectangular matrix and yield a permuted version of the data with less correlation between items. It has a simple scrambling algorithm with small delay. Due to these benefits, it adds a high degree of encryption to the coded transmitted data and achieves better BER performance [26]. The chaotic permutation is accomplished as indicated in [27].

Let $D_{(a_1,\ldots,a_n)}$ signify the discretized chaotic Baker map, and the vector $(a_1, \ldots, a_n)$ represent the secret key ($\mathrm{SE_{key}}$). The secret key is selected such that each integer $m_i$ divides $M$ and $a_1 + \ldots + a_k = M$. The data item at the indices $(u, v)$ after interleaving is moved to the indices given by the following equation:

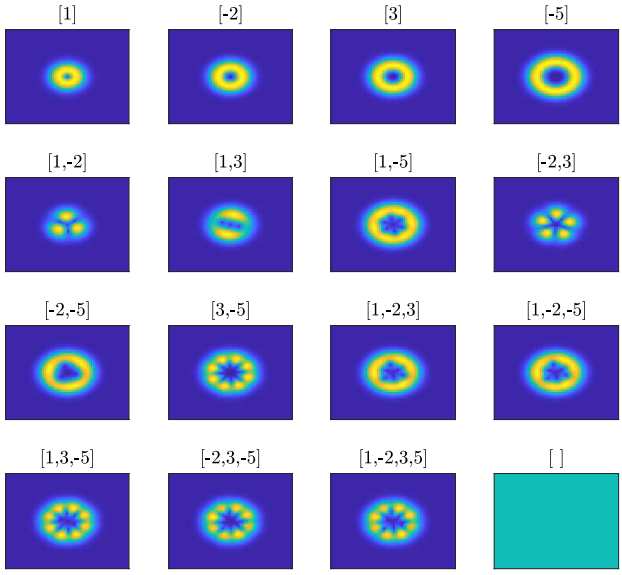


**Fig. 4.** Images for the states of OAM of $\{1, -2, 3, -5\}$.

$$D_{(a_1,\ldots,a_n)}(u, v) = \left[\left(\frac{M}{m_i}\right)(u - M_i) + \mathrm{mod}\left(v, \frac{M}{m_i}\right)\right.$$
$$\left.\frac{m_i}{M}\left(v - \mathrm{mod}\left(v, \frac{M}{m_i}\right) + M_i\right)\right], \quad \textbf{(15)}$$

where $M_i \le u \le M_i + m_i$, $0 \le v \le M$, $M_1 = 0$, and $\mathrm{mod}(x, y)$ represents the remainder of $x/y$. Here, $M$ is the number of elements in a single row.

### D. Enhanced Mapping Scheme for OAM

Due to the miscellaneous impacts of the free-space channel on the different OAM states, the choice of the OAM fundamental states has an excessive influence on the performance of the OAM-SK–FSO system. Because the CNN can be used as a good classifier for the characteristic features of OAM beams, we use quaternary states to form 16-ary OAM. Figure 4 indicates the 16-ary format tensors that characterize the code alphabet in the form of {(0,0,0,0), (0,0,1,0), (0,0,0,1), (0,0,1,1), (0,1,0,1), (0,1,0,0), (0,1,1,0), (1,0,0,0), (0,1,1,1), (1,0,0,1), (1,0,1,1), (1,0,1,0), (1,1,0,0), (1,1,0,1), (1,1,1,1), (1,1,1,0)}.

## 4. IMAGE BROADCASTING WITH THE PROPOSED SYSTEM

In this section, we investigate digital image transmission with the proposed system with and without chaotic interleaving in different scenarios. We study the BER performance of the proposed system for different SNR values, AT strengths, and propagation ranges. We also investigate the designed CNN architecture that yields the highest recognition rate of OAM symbols and the highest MAP for ADAM, RMSProp, and SGDM with different parameters.

Figure 5 shows the image transmission process through the coded OAM-SK–FSO system with chaotic interleaving. First, the digital image is encoded using Turbo or LDPC code. After that, the encoded image is mapped using OAM modulation,

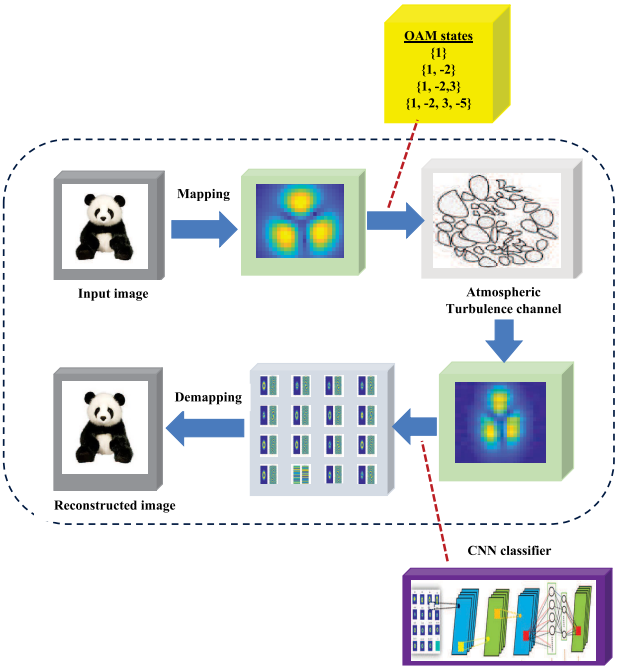


**Fig. 5.** Complete image transmission process with the proposed system.

and this is done by mapping each set of four bits into a single state of the 16-ary OAM states. After that, we apply chaotic interleaving on the modulated data. Then, the modulated data are transmitted over the AT channel to the receiver. At the receiver, the demodulation is performed after applying chaotic deinterleaving to demap each OAM state to four bits in order to get the original coded data. Finally, we apply Turbo or LDPC decoding to get the original image. The PSNR value of the received image is used as a quality metric.

## 5. SIMULATION RESULTS

### A. Design Methodology of the CNN Model

The use of high-cost optical devices can be diminished with the assistance of an adaptive demodulator based on a CNN, and this proficiently enhances the recognition rate of OAM states in a turbulent atmosphere. Figure 6 demonstrates the construction of the proposed CNN model. This model is designed using trial and error, and it depends on numerous considerations to obtain the optimal parameters and hyperparameters to yield the highest accuracy and the maximum MAP. This is achieved by:

1. changing the number of layers (convolution and pooling) and/or the number of neurons per layer;
2. modifying the CNN parameters such as filter size, pooling size, stride, and number of kernels;
3. changing the hyperparameters of the network such as optimization algorithm, batch size, learning rate, and number of epochs.

Now, we can construct the proposed CNN model using Eqs. (16) and (17), and it contains a single input layer, four convolution layers, one pooling layer, three dropout layers, three batch normalization layers, one additional layer, a single fully

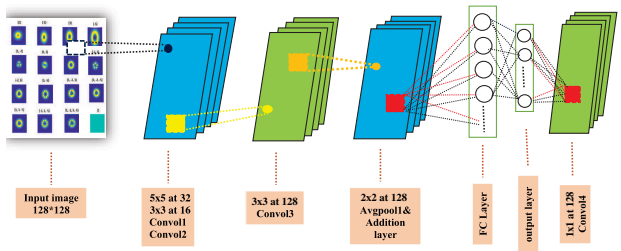


**Fig. 6.** Structure of the proposed CNN demodulator.

connected layer, and one output layer. In each convolution layer, filters are convolved with the input image to get a number of convolution outputs or activations, based on the sizes of the kernels, the distance between accessible fields (stride), and padding. Initially, resizing of the input images to $128 \times 128$ is performed before the input layer. Then, convolution is applied on the input image in the first convolution layer using 32 kernels each of size $5 \times 5$ to acquire the feature maps of the image. In the second and third convolution layers, the same stages are performed but with a filter of size $3 \times 3$. After that, in the average pooling layer, we perform pooling with a size of $2 \times 2$, and stride of 2 to get feature maps of $128 \times 6 \times 6$ elements with the average pooling algorithm that diminishes the computational cost. Subsequently, 16 nodes in the fully connected layer are correlated with the nodes in the first average pooling layer. Finally, the detection of the OAM states is performed with 16 nodes through the SoftMax classifier. After that, we add a convolution layer to the used layers to modify the network performance with a filter size of $1 \times 1$ to get feature maps of $128 \times 4 \times 4$ elements, as indicated in Table 1.

The size of the convolution layer output image is given by [28]

$$O_c = \frac{I + 2p - k}{s_d} + 1, \tag{16}$$

where $O_c$ is the size of output image, $I$ is the size of input image, $k$ is the size of kernels in the convolution layer, $s_d$ is the stride of the convolution operation, and $p$ is the padding size The size of the output image of the pooling layer is given by

$$O_P = \frac{I - p_s}{s_d} + 1, \tag{17}$$

**Table 1. Different Parameters of Each Layer in the Proposed CNN**

| Layer Name | No. of Filters | Filter Size | Stride | Pad Size |
|---|---|---|---|---|
| **Conv1** | 32 | $5 \times 5$ | 2 | 2 |
| **Dropout** | | 0.5 dropout | | |
| **Conv2** | 16 | $3 \times 3$ | 2 | 1 |
| **Conv3** | 128 | $3 \times 3$ | 3 | 1 |
| **Average pooling** | 128 | 2 | — | 2 |
| **Fully connected** | | 16 fully connected layers | | |
| **Conv4** | 128 | $1 \times 1$ | 3 | 2 |

**Table 2. Platform Specifications**

| System | Specifications |
|---|---|
| **Type** | 64 bit Windows 10 |
| **Processor** | Intel Core i7-6700 CPU at 2.6 Ghz |
| **Graphics card** | NVIDIA Geforce GTX 1070 compatible |
| **Installed memory (RAM)** | 16 GB memory |

where $O_P$ is the size of the output pooling image, and $p_s$ is the pooling size.

In the proposed model, the nonlinear rectified linear unit (ReLU) activation function is utilized with the convolution layers to allow the other layers to contribute to the learning task, and the dropout layer is used to reduce the probability of overfitting of the network. Finally, a mini-batch value is used with the batch normalization layer to normalize each input channel. The usage of this layer accelerates the CNN training and makes it less sensitive to network initialization. Then, the SoftMax layer is utilized to discriminate between the OAM mode patterns. In the training steps of the proposed network, the SGDM, RMSProp, and ADAM optimizers are tested to update the overall weights during 100 epochs, and the loss cross-entropy is used to compute the loss of the network. Also, we use different values of both parameters and hyperparameters of the network to get the optimal conditions that give the highest accuracy and the lowest cost. We have 16,000 images for the 16-ary OAM states in the training process. The input images are reshaped from the size of 875 × 656 to 128 × 128 to diminish the whole system cost. All results have been obtained using the platform specifications displayed in Table 2.

In Table 3, we present a comparison between the accuracy of the model of [10] and the proposed model through the use of different hyperparameters such as batch size, learning rate, number of epochs, and regularization parameter to select the optimal values for the three used optimizers. The letters S, R, and A refer to SGDM, RMSProp, and ADAM optimizers, respectively. By making a trade-off between the accuracy and training time for obtaining the optimum hyperparameter values, it is realized that these values are: 128, 0.001, 2150, and 0.0001 for batch size, learning rate, number of iterations, and regularization parameter, respectively.

The comparison between different machine learning algorithms such as deep neural network (DNN) and distributed random forest (DRF) is performed by measuring the area under curve (AUC) using the ADAM optimizer, and it reveals the following conclusions.

1. From [29], it is clear that the classification accuracy for forest DNN (FDNN) is from 77% to 98% according to the selected hyperparameters, and it is superior to those of the DNN and RFs.
2. For the proposed model, it is found that the classification accuracy for the CNN ranges from 96% to 99% according to the hyper parameters values.

Based on the obtained optimal values, we measure the performance of the proposed deep CNN model as shown in Table 4 by calculating the precision, recall, specificity, $F_{\text{score}}$, AUC, and negative predictive value (NPV) using Eqs. (18)–(25).

**Table 3. Effect of Different Parameters on the Accuracy of the System with Different Optimizers (%)**

| | | Proposed Model | | | Previous Model | | |
|---|---|---|---|---|---|---|---|
| **Parameter** | **Values** | **S** | **R** | **A** | **S** | **R** | **A** |
| **Batch size** | **32** | 97.9 | 94 | 97.8 | 95.5 | 92.9 | 94.5 |
| | **64** | 97.8 | 97.2 | 97.8 | 95.9 | 97.2 | 96.7 |
| | **128** | 97.9 | 97.7 | 97.8 | 95.9 | 95.9 | 95.2 |
| | **256** | 97.9 | 97.9 | 97.9 | 96.7 | 97 | 94.3 |
| **Learning rate** | **0.1** | 6.3 | 31.3 | 69.7 | 94.8 | 92 | 78.9 |
| | **0.01** | 97.9 | 44.2 | 89.2 | 96.8 | 88.9 | 97.1 |
| | **0.001** | 97.8 | 97 | 97.9 | 95.9 | 97.2 | 96.7 |
| | **0.0001** | 97.9 | 93.3 | 97.9 | 97 | 96.5 | 94.1 |
| **Number of iterations** | **215** | 97.9 | 94.3 | 97.9 | 93.8 | 97.8 | 96.3 |
| | **430** | 97.9 | 96.8 | 97.9 | 93.8 | 97.9 | 97.5 |
| | **860** | 97.9 | 96.3 | 97.8 | 94.9 | 97.8 | 97.5 |
| | **2150** | 97.8 | 97 | 97.9 | 95.9 | 96. | 96.66 |

**Table 4. Effect of Different Parameters on the Performance of the System with Different Optimizers (%)**

| | | Iterations | | Batch Size | | Learning Rate | |
|---|---|---|---|---|---|---|---|
| **Hyperparameters** | | **430** | **2150** | **64** | **128** | **0.001** | **0.0001** |
| **Precision** | **S** | 97.9 | 97.8 | 97.8 | 97.9 | 97.8 | 97.9 |
| | **R** | 96.3 | 96.8 | 96.8 | 97.8 | 96.8 | 95.8 |
| | **A** | 97.9 | 97.9 | 97.9 | 97.9 | 97.9 | 97.8 |
| **Recall** | **S** | 98.1 | 98.1 | 98.1 | 98.1 | 98.1 | 98.1 |
| | **R** | 96.8 | 97 | 97 | 98 | 97 | 98.1 |
| | **A** | 98.1 | 98.1 | 98.1 | 98.1 | 98.1 | 98.2 |
| **SP (TNR)** | **S** | 99.9 | 99.9 | 99.9 | 99.9 | 99.9 | 99.9 |
| | **R** | 99.8 | 99.8 | 99.8 | 99.9 | 99.8 | 99.9 |
| | **A** | 99.9 | 99.8 | 99.8 | 99.9 | 99.8 | 99.9 |
| $F_{\text{score}}$ | **S** | 97.8 | 97.8 | 97.8 | 97.8 | 97.9 | 97.9 |
| | **R** | 95.9 | 96.6 | 96.6 | 97.8 | 96.6 | 97.8 |
| | **A** | 97.8 | 97.8 | 97.8 | 97.8 | 97.8 | 97.9 |
| **NPV** | **S** | 99.9 | 99.9 | 99.9 | 99.9 | 99.9 | 100 |
| | **R** | 99.8 | 99.8 | 99.8 | 99.9 | 99.8 | 99.9 |
| | **A** | 99.8 | 99.8 | 99.8 | 99.8 | 99.8 | 99.9 |
| **AUC** | **S** | 99.7 | 99 | 99 | 99 | 99 | 96.2 |
| | **R** | 98.3 | 98.3 | 98.3 | 98.9 | 98.3 | 95.1 |
| | **A** | 98.7 | 99 | 99 | 99 | 99 | 96.2 |

These metrics are computed from the confusion matrix in different cases with the help of the following parameters.

1. True positive (TP), defined as the number of accurately categorized cases that belong to the class.
2. True negative (TN), defined as the number of accurately categorized cases that do not belong to the class.
3. False positive (FP), defined as the cases that are erroneously categorized as belonging to the class.
4. False negative (FN), defined as the cases that are not categorized as class cases.

The equations used to measure the above-mentioned metrics are [30] as follows.

Recall or TP rate (TPR):

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{18}$$

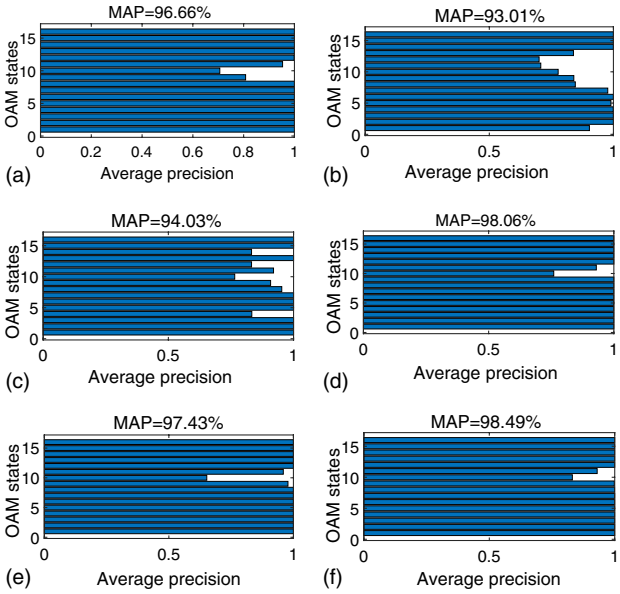


**Fig. 7.** MAP for different models using different optimizers: (a), (d) SGDM; (b), (e) RMSProp; (c), (f) ADAM.

Specificity or TN rate (TNR):

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}. \tag{19}$$

Accuracy:

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}. \tag{20}$$

Precision or positive predictive value (PPV):

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}. \tag{21}$$

NPV:

$$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{FN} + \mathrm{TN}}. \tag{22}$$

To achieve a balance between precision and recall, we measure the value of $F_{\mathrm{score}}$ according to the following equation:

$$F_{\mathrm{score}} = \frac{2 \times \mathrm{TP}}{(2 \times \mathrm{TP}) + \mathrm{FP} + \mathrm{FN}}. \tag{23}$$

The AUC is given by

$$\mathrm{AUC} = 0.5 \times \left( \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} + \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \right). \tag{24}$$

The MAP, defined as the average value of PPVs that are calculated for all classes, is used to assess the model:

$$\mathrm{MAP} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{PPV}_i, \tag{25}$$

where $N$ is the number of classes.

Figure 7 demonstrates a comparison between the proposed model and the model in [10] using different optimizers. In this figure, we present the MAP values for the two models to be compared and notice that the proposed model has the best detection performance compared to the previous model. The ADAM optimizer has the ability to detect the classes of the model better than the other optimizers, which agrees with the state of the art.

Figure 8 shows the plots of the receiver operating characteristic (ROC) curves for different optimizers using different hyperparameters for the two models. The ROC curve is used to check the quality of the classifier depending on the values of the TPR and the FP ratio (FPR). We notice in the figure that the proposed model is superior to the previous model in all cases. The utilization of 10 epochs and different learning rates makes the SGDM optimizer better than the other two optimizers, while 50 epochs make the ADAM optimizer better. Changing the learning rate and batch size makes the proposed model better than the previous models. The ADAM optimizer has the best performance compared to the other optimizers with different batch sizes.

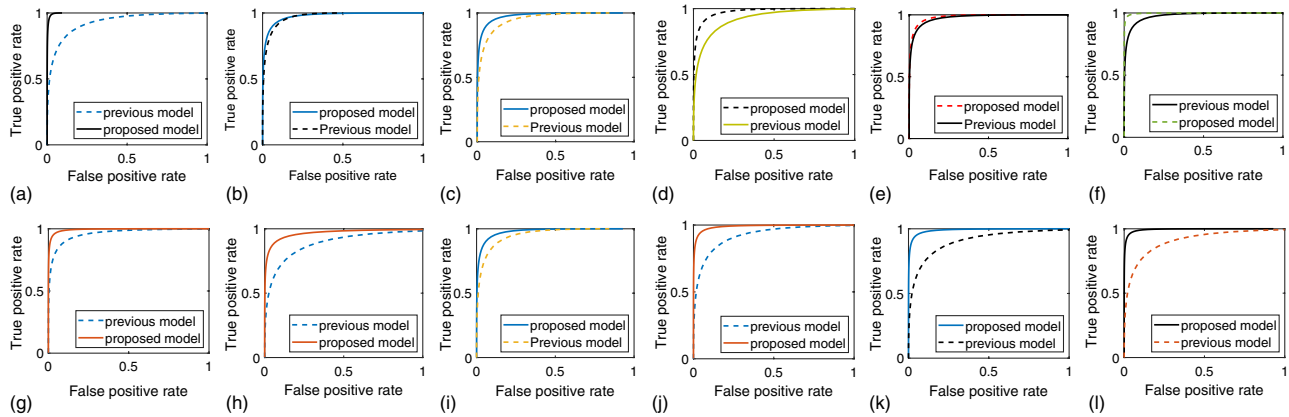


**Fig. 8.** ROC curves for different models using different optimizers for different hyper-parameters: (a) SGDM with 430 iterations, (b) RMSProp with 430 iterations, (c) ADAM with 430 iterations, (d) SGDM with 2150 iterations, (e) RMSProp with 2150 iterations, (f) ADAM with 2150 iterations, (g) SGDM with 0.0001 LR, (h) RMSProp with 0.0001 LR, (i) ADAM with 0.0001 LR, (j) SGDM with 256 batch size, (k) RMSProp with 256 batch size, and (l) ADAM with 256 batch size.

### B. Performance of the Proposed System in Different Scenarios

Figure 9 demonstrates the effect of using different codes on the quality of the system with and without chaotic interleaving. In the figure, we notice that the value of PSNR decreases with an increase in turbulence strength. For weak, moderate, and strong turbulence strengths, the LDPC code gives better results than those of the Turbo code with and without interleavering. At $10^{-14}$, the chaotic interleaver improves the performance by about 2 dB and 4 dB for Turbo and LDPC codes, respectively. Figure 10 shows a comparison between the Turbo and LDPC codes with chaotic interleaving. The figure indicates that increasing the propagation distance or turbulence strength makes the LDPC code have a better BER performavce than that of the Turbo code in cases of interleaving and no interleaving.

In Fig. 11, we plot the measured values of PSNR for diverse propagation distances to evaluate the performance of the system. As long as the value of PSNR is high, the quality of the reconstructed image will be high. In this figure, we find that the PSNR value decreases with the increase in propagation distance with and without chaotic interleaving. For small propagation distances, we find that the Turbo code is better than the LDPC code. However, increasing the value of the propagation distance makes the LDPC code better than the Turbo code with and without a chaotic interleaver. Figure 12 demonstrates a PSNR comparison between the utilization of Turbo and LDPC codes with chaotic interleavering. The figure indicates that increasing the values of the propagation distances and turbulence strength makes the LDPC code have a greater value of PSNR than the Turbo code in the two cases by about 7 dB. In Fig. 13, we introduce a comparison between the original and reconstructed images with chaotic interleaving and differnt codes for different turbulence strengths and propagation distances.

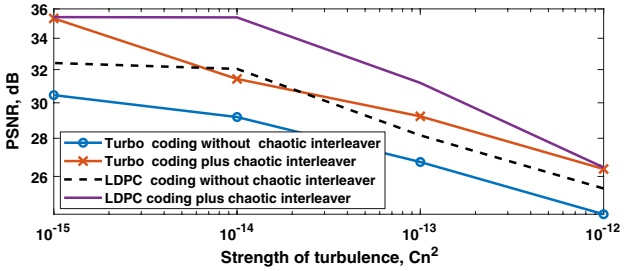


**Fig. 9.** PSNR versus turbulence strength with and without chaotic interleaving using different codes.

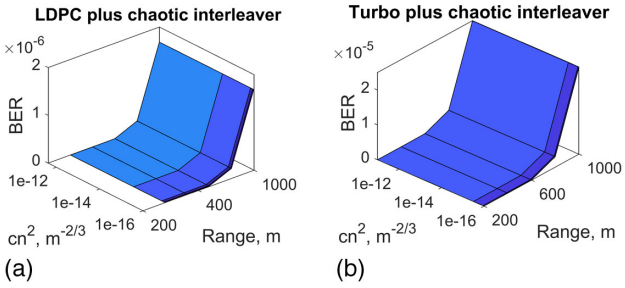


**Fig. 10.** BER comparison between Turbo and LDPC codes.

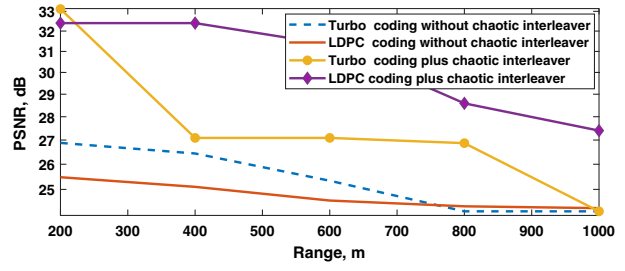


**Fig. 11.** PSNR versus propagation ranges with and without chaotic interleaving using different codes.

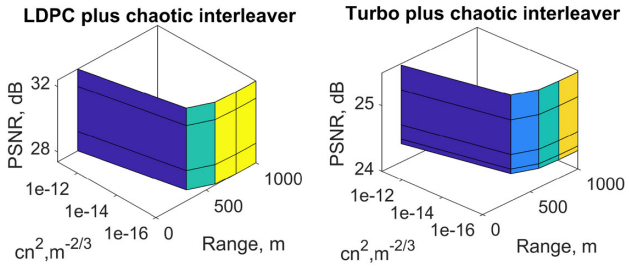


**Fig. 12.** PSNR comparison between Turbo and LDPC codes.

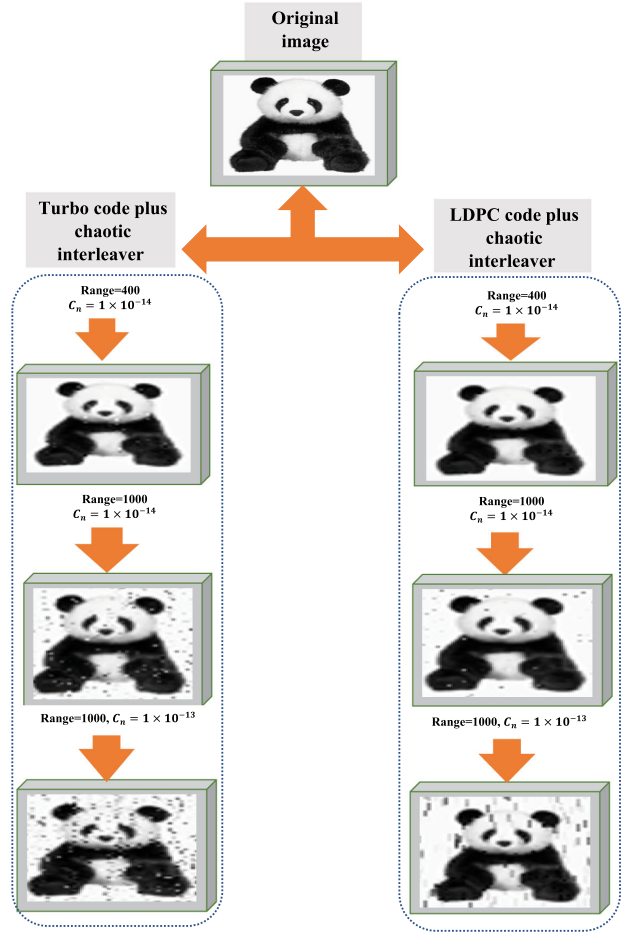


**Fig. 13.** Original image, and recovered images with chaotic interleaving with different coding schemes.

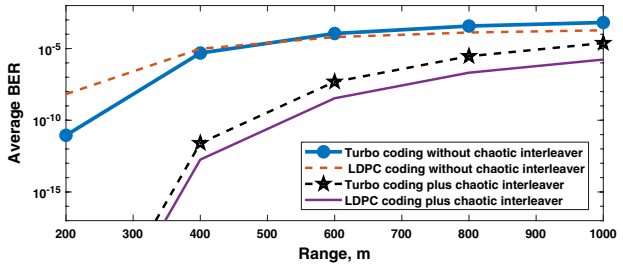


**Fig. 14.** BER versus different values of propagation distances with LDPC code and Turbo code.

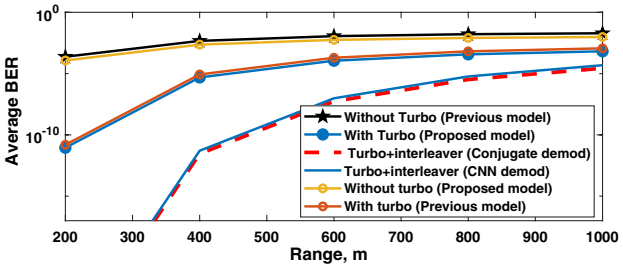


**Fig. 15.** BER versus propagation range with and without chaotic interleaving and different coding schemes.

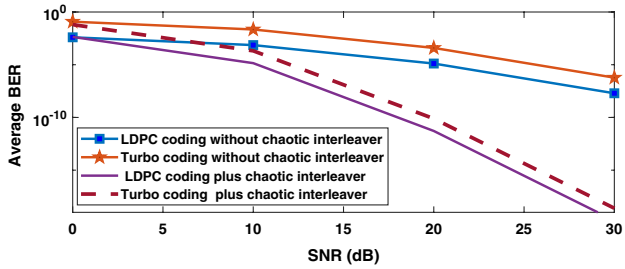


**Fig. 16.** BER versus SNR with and without chaotic interleaving with both codes.

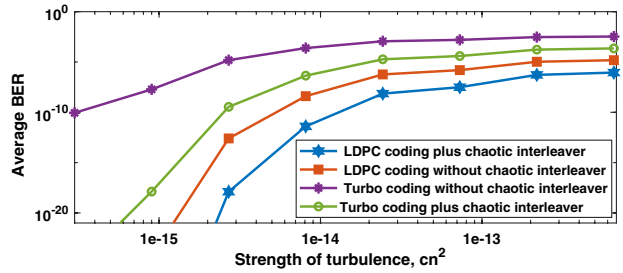


**Fig. 17.** BER versus turbulence strength with and without chaotic interleaving and different codes.

The different propagation distances are 400 m at PSNR value of 29.09 dB for the Turbo code and 31.34 dB for the LDPC code, 1000 m, $1 \times 10^{-14}$ at PSNR value of 26.87 dB for the Turbo code and 30.15 dB for the LDPC code, and 1000 m, $1 \times 10^{-13}$ at PSNR value of 24.16 dB for the Turbo code and 27.36 dB for the LDPC code. In Fig. 13, we find that the LDPC code with a moderate turbulence strength and 400 m propagation distance reduces the BER between the original and received images to approximately zero. In this case, the CNN network achieves a recognition accuracy of about 100%.

In Fig. 14, we notice that the BER increases with an increase in propagation distance for the two codes. We note also that with the interleaver, the LDPC code is better for all values of propagation distance. Without interleavering, the Turbo code is preferred for small propagation distances, and the LDPC code is preferred for large propagation distances.

In Fig. 15, we present a comparison between the model in [12] and the proposed model using two OAM demodulation techniques: CNN and conjugate light field. It is clear in the figure that both techniques have nearly the same performance for different cases. With the chaotic interleaver, the performance of the proposed model is better than that of the previous model for large values of propagation distance. At a distance of 1000 m, the value of the BER decreases from $1^{-2}$ to nearly $1^{-5}$, while at a distance of 600 m, the value of the BER reduces from $1e^{-2}$ to approximately $1e^{-7}$.

Figure 16 demonstrates the BERs for different values of SNR. We find that the value of BER decreases with the increase in SNR for both codes. The figure shows that the LDPC code always gives lower BERs than those of the Turbo code with and without chaotic interleavering. In Fig. 17, we find that increasing the turbulence strength increases the BER for small turbulence strengths, while for strong strengths, the value of BER is nearly constant. With and without a chaotic interleaver, the LDPC code achieves lower BER values than those of the Turbo code for all turbulence strengths: weak, moderate, and strong.

## 6. CONCLUSION

In this paper, we have proposed a novel 16-ary OAM-SK–FSO communication system based on coding, chaotic interleaving, and CNN-based adaptive demodulation to accommodate for strong turbulence strengths. At the transmitter, we first adopt an encoder structure, and a progressive mapping scheme to diminish the BER for different parameters and hyperparameters of the CNN demodulator. At the receiving side, OAM demodulation is performed with conjugate or CNN detection methods. The LDPC coding is recommended more than Turbo coding in the proposed system. It is superior by about 3 dB without interleaving and by 6 dB with interleaving. Training of the CNN has been performed with different classifiers to yield the optimum values that achieve the highest detection accuracy through a diversity of parameters and hyperparameters. The SGDM and ADAM optimizers have nearly the same accuracy, which is greater than that of the RMSProp optimizer by about 3.8%. Variation of the batch size and learning rate allows the ADAM and SGDM optimizers to have high accuracy. By averaging the precision value through all classes, we find that the ADAM optimizer has the best MAP value compared to the RMSProp and SGDM optimizers by about 1% and 4%, respectively. To measure the model performance, we have estimated the ROC curves, which revealed that the SGDM and ADAM optimizers have the best performance. The AUC is a measure of the degree of separability, and it reveals the ability of the model to discriminate

between classes. From this study, we find that the ADAM and SGDM optimizers have almost the same performance, and the selection between them will be consistent with the application of interest.

**Disclosures.** The authors declare no conflicts of interest.

## REFERENCES


1. T. Richter, E. Palushani, C. Schmidt-Langhorst, R. Ludwig, L. Molle, M. Nölle, and C. Schubert, "Transmission of single-channel 16-QAM data signals at terabaud symbol rates," J. Lightwave Technol. **30**, 504–511 (2012).
2. A. H. Gnauck, P. J. Winzer, S. Chandrasekhar, X. Liu, B. Zhu, and D. W. Peckham, "Spectrally efficient long-haul WDM transmission using 224-Gb/s polarization-multiplexed 16-QAM," J. Lightwave Technol. **29**, 373–377 (2011).
3. W. Zhang, S. Zheng, X. Hui, R. Dong, X. Jin, H. Chi, and X. Zhang, "Mode division multiplexing communication using microwave orbital angular momentum: an experimental study," IEEE Trans. Wireless Commun. **16**, 1308–1318 (2017).
4. S. M. Mohammadi, L. K. S. Daldorff, J. E. S. Bergman, R. L. Karlsson, B. Thide, K. Forozesh, T. D. Carozzi, and B. Isham, "Orbital angular momentum in radio—a system study," IEEE Trans. Antennas Propag. **58**, 565–572 (2010).
5. Z. Guo, Z. Wang, M. I. Dedo, and K. Guo, "The orbital angular momentum encoding system with radial indices of Laguerre–Gaussian beam," IEEE Photon. J. **10**, 7906511 (2018).
6. Q. Tian, L. Zhu, Y. Wang, Q. Zhang, B. Liu, and X. Xin, "The propagation properties of a longitudinal orbital angular momentum multiplexing system in atmospheric turbulence," IEEE Photon. J. **10**, 7900416 (2018).
7. G. Gibson, J. Courtial, M. J. Padgett, M. Vasnetsov, V. Pas'ko, S. M. Barnett, and S. Franke-Arnold, "Free-space information transfer using light beams carrying orbital angular momentum," Opt. Express **12**, 5448–5456 (2004).
8. G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups," IEEE Signal Process. Mag. **29**(6), 82–97 (2012).
9. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), pp. 1–9.
10. Q. Tian, Z. Li, K. Hu, L. Zhu, X. Pan, Q. Zhang, Y. Wang, F. Tian, X. Yin, and X. Xin, "Turbo-coded 16-ary OAM shift keying FSO communication system combining the CNN-based adaptive demodulator," Opt. Express **26**, 27849–27864 (2018).
11. S. Zhao, B. Wang, L. Gong, Y. Sheng, W. Cheng, X. Dong, and B. Zheng, "Improving the atmosphere turbulence tolerance in holographic ghost imaging system by channel coding," J. Lightwave Technol. **31**, 2823–2828 (2013).
12. I. B. Djordjevic and M. Arabaci, "LDPC-coded orbital angular momentum (OAM) modulation for free-space optical communication," Opt. Express **18**, 24722–24728 (2010).
13. J. Liu, P. Wang, X. Zhang, Y. He, X. Zhou, H. Ye, Y. Li, S. Xu, S. Chen, and D. Fan, "Deep learning based atmospheric turbulence compensation for orbital angular momentum beam distortion and communication," Opt. Express **27**, 16671–16688 (2019).
14. Z. Wang, M. I. Dedo, K. Guo, K. Zhou, F. Shen, Y. Sun, S. Liu, and Z. Guo, "Efficient recognition of the propagated orbital angular momentum modes in turbulences with the convolutional neural network," IEEE Photon. J. **11**, 7903614 (2019).
15. X. Cui, X. Yin, H. Chang, H. Liao, X. Chen, X. Xin, and Y. Wang, "Experimental study of machine-learning-based orbital angular momentum shift keying decoders in optical underwater channels," Opt. Commun. **452**, 116–123 (2019).
16. X. Cui, X. Yin, H. Chang, Y. Guo, Z. Zheng, Z. Sun, G. Liu, and Y. Wang, "Analysis of an adaptive orbital angular momentum shift keying decoder based on machine learning under oceanic turbulence channels," Opt. Commun. **429**, 138–143 (2018).
17. W. Wang, P. Wang, L. Guo, W. Pang, W. Chen, A. Li, and M. Han, "Performance investigation of OAMSK modulated wireless optical system over turbulent ocean using convolutional neural networks," J. Lightwave Technol. **38**, 1753–1765 (2020).
18. J. Li, M. Zhang, and D. Wang, "Adaptive demodulator using machine learning for orbital angular momentum shift keying," IEEE Photon. Technol. Lett. **29**, 1455–1458 (2017).
19. A. Yu, H. Yang, T. Xu, B. Yu, Q. Yao, Y. Li, T. Peng, H. Guo, J. Li, and J. Zhang, "Long-term traffic scheduling based on stacked bidirectional recurrent neural networks in inter-datacenter optical networks," IEEE Access **7**, 182296 (2019).
20. X. Fang, H. Ren, and M. Gu, "Orbital angular momentum holography for high-security encryption," Nat. Photonics **14**, 102–108 (2020).
21. Z. Dong, Y. Zhang, H. Li, R. Tao, C. Gu, P. Yao, Q. Zhan, Q. Zhan, and L. Xu, "Generation of stable orbital angular momentum beams with an all-polarization-maintaining fiber structure," Opt. Express **28**, 9888–9995 (2020).
22. M. Sharma, F. Amirkhan, S. K. Mishra, D. Sengupta, Y. Messaddeq, F. Blanchard, and B. Ung, "Transmission of orbital angular momentum and cylindrical vector beams in a large-bandwidth annular core photonic crystal fiber," Fibers **8**, 22 (2020).
23. X. Ding, G. Feng, and S. Zhou, "Detection of phase distribution of vortex beams based on low frequency heterodyne interferometry with a common commercial CCD camera," Appl. Phys. Lett. **116**, 031106 (2020).
24. Z. Ghassemlooy, W. Popoola, and S. Rajbhandari, *Optical Wireless Communications: System and Channel Modelling with MATLAB* (CRC Press, 2019).
25. P. Ivaniš and D. Drajić, *Information Theory and Coding-Solved Problems* (Springer, 2017).
26. Y. Xiao, J. Cao, Z. Wang, C. Long, Y. Liu, and J. He, "Polar coded optical OFDM system with chaotic encryption for physical-layer security," Opt. Commun. **433**, 231–235 (2019).
27. E. S. Hassan, X. Zhu, S. E. El-Khamy, M. I. Dessouky, S. A. El-Dolil, and F. E. A. El-Samie, "A chaotic interleaving scheme for the continuous phase modulation based single-carrier frequency-domain equalization system," Wireless. Pers. Commun. **62**, 183–199 (2012).
28. J. Wu, *Introduction to Convolutional Neural Networks* (National Key Lab for Novel Software Technology, Nanjing University, 2017), Vol. **5**, p. 23.
29. Y. Kong and T. Yu, "A deep neural network model using random forest to extract feature representation for gene expression data classification," Sci. Rep. **8**, 16477 (2018).
30. F. Idrees, M. Rajarajan, M. Conti, T. M. Chen, and Y. Rahulamathavan, "PIndroid: a novel android malware detection system using ensemble learning methods," Comput. Secur. **68**, 36–46 (2017).