photonics

MDPI

Article

VLCMnet-Based Modulation Format Recognition for Indoor Visible Light Communication Systems

Xin Zheng, Ying He *, Chong Zhang and Pu Miao

College of Electronics and Information, Qingdao University, Qingdao 266071, China; 2021020649@qdu.edu.cn (X.Z.); zhangchong@qdu.edu.cn (C.Z.); mpvae@qdu.edu.cn (P.M.)
* Correspondence: yinghe@qdu.edu.cn

Abstract: In indoor visible light communication (VLC), the received signals are subject to severe interference due to factors such as high-brightness backgrounds, long-distance transmissions, and indoor obstructions. This results in an increase in misclassification for modulation format recognition. We propose a novel model called VLCMnet. Within this model, a temporal convolutional network and a long short-term memory (TCN-LSTM) module are utilized for direct channel equalization, effectively enhancing the quality of the constellation diagrams for modulated signals. A multi-mixed attention network (MMAnet) module integrates single- and mixed-attention mechanisms within a convolutional neural network (CNN) framework specifically for constellation image classification. This allows the model to capture fine-grained spatial structure features and channel features within constellation diagrams, particularly those associated with high-order modulation signals. Experimental results obtained demonstrate that, compared to a CNN model without attention mechanisms, the proposed model increases the recognition accuracy by 19.2%. Under severe channel distortion conditions, our proposed model exhibits robustness and maintains a high level of accuracy.

Keywords: modulation format recognition; convolutional neural network; visible light communication; attention mechanism; channel equalization

Citation: Zheng, X.; He, Y.; Zhang, C.; Miao, P. VLCMnet-Based Modulation Format Recognition for Indoor Visible Light Communication Systems. *Photonics* **2024**, *11*, 403. https://doi.org/10.3390/photonics11050403

Received: 31 March 2024
Revised: 21 April 2024
Accepted: 24 April 2024
Published: 26 April 2024

1. Introduction

Direct current biased optical orthogonal frequency division multiplexing (DCO-OFDM) is a mature modulation technique widely adopted in indoor visible light communication (VLC) systems. To ensure optimal transmission performance and reliability, adaptive modulation should be carried out according to different channel conditions; hence, the receiver must autonomously detect and identify the modulation format of the received signals. However, challenges, such as high-brightness environments, long-distance transmission, and indoor obstructions, significantly reduce the signal-to-noise ratio (SNR), which, in turn, affects modulation format recognition (MFR) performance. Therefore, automatic modulation recognition methods under low SNR conditions have gained increasing attention.

Modulation format recognition techniques are generally classified into two categories: likelihood-based methods and feature-based methods [1]. Likelihood-based approaches, although typically characterized by higher accuracy, often require substantial amounts of prior information and higher computational complexity. Feature-based methods rely on extracting key features from the signals, such as higher-order accumulations [2], sequence features [3], and image features [4]. These methods generally exhibit higher computational efficiency, and their recognition results tend to be more interpretable and analyzable. However, improper feature selection can lead to decreased recognition performance.

Initially, the modulation format recognition technique rapidly developed in the field of radio communications [5–7]. In recent years, numerous machine learning algorithms have also made strides in the field of MFR in optical communications [8–13]. While inputting time-series signals can capture the complete dynamic characteristics of signals

over time, severe noise interference may cause modulation features to become blurred, thus increasing the difficulty of recognition. Constellation diagrams, by mapping signals onto discrete points, present unique patterns for different modulation formats; hence, utilizing constellation diagrams for modulation format recognition can mitigate the impact of noise.

Zhen. Z et al. employed an improved AlexNet [14] to classify constellation diagrams of the same six modulation formats, achieving a recognition accuracy of 77.45% across the SNR range of 0 dB to 15 dB. The incorporation of data augmentation (DA) operations was able to further improve the recognition accuracy to 88%, showcasing the potential of advanced image processing techniques in optical communication. Ma et al. [15] investigated a blind modulation format recognition method based on constellation diagrams derived from channel estimation in an OFDM system. They employed a combination of a peak density clustering algorithm and K-nearest neighbor (KNN) regression. At an OSNR of 30 dB and above, high-order modulation signals achieved a recognition accuracy of 80%. However, studies regarding low OSNR environments were not conducted.

Some research efforts are focusing on the preprocessing of constellation diagram features. W. Liu [16] proposed a novel modulation classification scheme for optical communications. Within an SNR range of 0 dB to 15 dB, the original constellation diagrams are preprocessed into density constellation maps, which are used for classifying six different modulation formats. Due to the reduced Euclidean distance in normalized density constellation diagrams, the accuracy rate at an SNR of 0 dB is merely 54%. In addition, a fan-beam projection algorithm to handle constellation diagrams was presented in [17]. Wei et al. [18] utilized probabilistic constellation shaping to modify the distribution of signal constellation points, aiming to enhance the quality of constellation diagrams and enhance recognition accuracy.

The complexity and nonlinearity of VLC channels result in the received signal's constellation diagram deviating from a standard lattice structure with each constellation point appearing stretched inward toward the center [19]. Pre-distortion [20–23] is a valid equalization scheme for compensating the memory nonlinearity. Nevertheless, some additional physical circuits must be devised to assist the pre-distortion at the transmitter. Unlike the fomer method, nonlinear post-equalization (NPE) is a superior and cost-effective technique as it can address the comprehensive nonlinearity of the LED and optical channel cascade. H. Chen et al. [24] introduced a collaborative time-frequency deep neural network (TFDNet) to mitigate nonlinear distortions in VLC systems. Miao et al. [25] also presented a model-driven deep learning (DL) equalization scheme aimed at resolving severe channel impairment issues. In [26], a CRNN-based equalizer employing a combined method of CNN and LSTM was proposed, which achieved an accuracy rate of 90% for identifying quadrature phase shift keying (QPSK) signals when the SNR exceeded 10 dB. However, the symbol error rate rose above 30% when the SNR fell below 0 dB. To simplify the model, Costa et al. [27] proposed a channel equalization scheme based on a one-dimensional CNN, which eliminated the need for channel estimation interpolation operations. But the bit error rate (BER) increased to 1% at 4 dB. A novel equalizer based on a gated recurrent unit (GRU) neural network was introduced for the first time to mitigate both linear and nonlinear distortions in carrier-less amplitude phase (CAP) band-limited VLC systems [28]. A hybrid temporal CNN with frequency domain assistance was employed to enhance the nonlinear compensation efficiency of the CAP underwater VLC system [29]. Nonetheless, the training datasets utilized in these research works consisted of real-numbered signals that were not down-sampled, thus lacking direct relevance to the efficiency of complex-valued VLC systems, leading to limited generalization capacity. Moreover, the up-sampling factor further adds to the computational complexity.

In summary, we choose constellation diagrams as the primary feature for addressing the MFR problem in indoor visible light communication, and employ channel equalization to restore the constellation diagrams. This paper introduces a novel model, termed the visible light communication modulation recognition network (VLCMnet), which consists of TCN-LSTM and MMAnet modules. It leads to increased robustness of the modulation

format recognition model, enabling accurate signal identification even under harsh channel conditions. The major contributions of this paper are as follows:

(1) A novel channel equalization model is proposed, named TCN-LSTM. Firstly, a temporal convolutional network (TCN) [30] is used to extract deep local features from the signal, followed by a two-layer LSTM network that screens critical information and suppresses noise. The model significantly improves the quality of constellation diagrams at low SNR.

(2) At the recognition stage, a constellation diagram classifier based on a multi-mixed attention network (MMAnet) is proposed. It integrates shallow feature extraction, single attention, and hybrid attention mechanisms, possesses multi-scale feature fusion capabilities, and effectively captures crucial spatial structure and channel features within constellation diagrams, thereby significantly boosting the model's recognition performance.

The rest of this paper is organized as follows: Section 2 introduces the indoor visible light communication system based on DCO-OFDM. Section 3 provides a detailed description of the principles behind the TCN-LSTM and MMAnet models. Section 4 discusses the performance advantages and disadvantages of our proposed MFR model from multiple perspectives. Finally, we conclude the full article in Section 5.

2. DCO-OFDM System

The DCO-OFDM system comprises a transmitter, intensity modulation, a direct detection (IM/DD) channel, and a receiver. A common modulation scheme is M-ary quadrature amplitude modulation (M-QAM), as shown in Figure 1. The VLCMnet model proposed in this paper serves as an intelligent channel equalization and modulation format identification module within the receiver. The modulation formats studied in this paper include M-QAM (where M = 4, 8, 16, 32, 64).

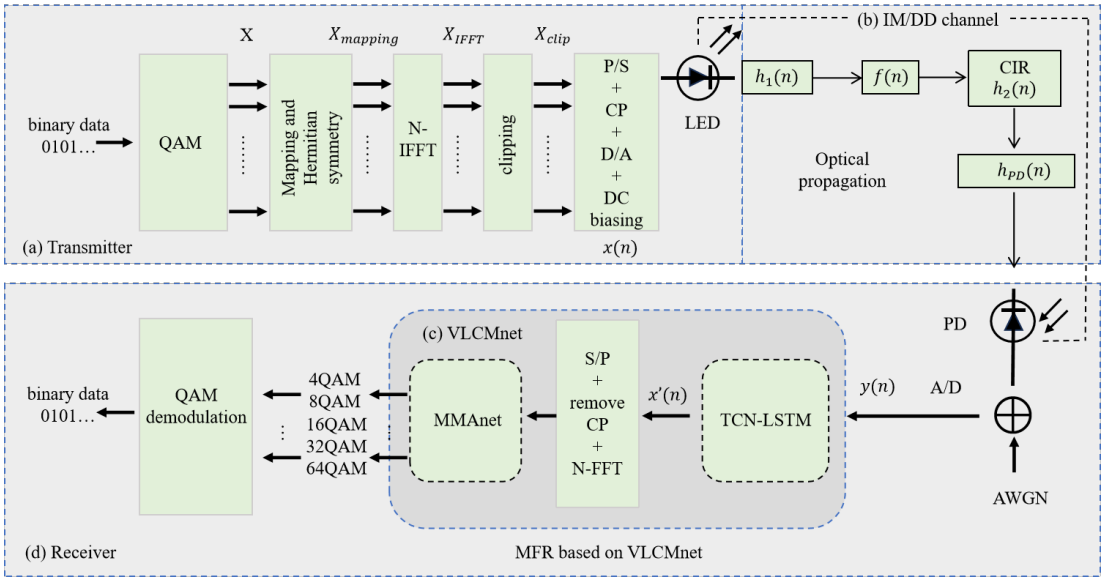


Figure 1. VLCMnet-based modulation format recognition framework for DCO-OFDM. (**a**) Transmitter side of the visible system. (**b**) IM/DD channel. (**c**) MFR based on VLCMnet. (**d**) Receiver.

At the transmitter end, firstly, M-QAM and constellation mapping are adopted to generate OFDM symbols. The information bits are mapped into real-valued symbols and the inserted cyclic prefix. Then, the bipolar baseband signal $x(n)$ is transformed into a unipolar signal by adding a direct current (DC) bias. Finally, the light-emitting diodes (LEDs) are used to generate modulated unipolar light signals. The IM/DD channel consists of a modulator at the transmitting end, a transmission medium, and a photodetector at the

receiving end. During transmission along the optical link, the signal is subject to additive white Gaussian noise (AWGN) interference. Taking all these factors into account, the output $y(n)$ of the channel can be expressed as:

$$y(n) = R_{PD} \times f\{(x(n) + I_{DC}) * h_1(n)\} * h_2(n) + \varepsilon(n), \quad (1)$$

where $h_1(n)$ represents the memory effect nonlinearity compensation of the LED, and $h_2(n)$ denotes the channel impulse response (CIR) function of the optical propagation link. R_{PD} stands for the photoelectric conversion coefficient of the photodiode (PD), and I_{DC} signifies the DC component. $\varepsilon(n)$ represents the Gaussian distribution channel noise. The $*$ indicates convolution. A range of diverse input current values were chosen to measure the real output brightness of the LEDs corresponding to each input current value. The errors between the measured data and the curves were reduced to a minimum using the least squares method. This enabled the identification of the polynomial function that most accurately matched the data points. The coefficients of the third-order polynomial were established as 0.2855–1.0886, 2.0565, and -0.0003. The function $f(\cdot)$ adopted to describe the memoryless nonlinear response of the LED can be calculated by:

$$f(x) = 0.2855x^3 - 1.0886x^2 + 2.0565x - 0.0003. \quad (2)$$

As the modulation frequency increases, the modulation efficiency of the LED gradually decreases, exhibiting a low-pass filter effect. The pulse response $h_1(n)$ can be represented as:

$$h_1(n) = \exp(-2\pi n f_0), \quad (3)$$

where f_0 denotes the 3 dB cut-off frequency. Considering that light travels through multiple reflections and scatterings before reaching the receiver, a multi-path propagation model for VLC channels is established using ray tracing methods. The channel impulse response $h_2(n)$ is related to the path length and number of reflections for each individual ray and can be calculated as follows:

$$h_2(n) = \sum_{i=1}^{N_r} p_i \delta(n - \tau_i), \quad (4)$$

where N_r represents the number of rays received by the detector, p_i denotes the optical power of the ith ray, and τ_i signifies the delay of the ith ray. The impulse response of a photodetector can be approximately modeled as the product of a Dirac delta function and the photoelectric conversion coefficient R_{PD}, given by the following formula:

$$h_{PD}(n) = R_{PD}\delta(n). \quad (5)$$

The output signal $y(n)$ is affected by the memoryless nonlinearity of the light-emitting diodes and multipath effects in the optical transmission link, causing the constellation points to deviate from their intended grid locations, resulting in degraded constellation diagram quality. Therefore, channel equalization techniques are needed to improve the distribution of points on the constellation diagram, bringing it closer to the ideal state.

In this paper, the VLCMnet model first employs an integrated neural network model to accomplish the task of channel equalization, following which it utilizes constellation diagrams along with the neural network model for modulation format recognition. The system unmaps the signal based on the category information outputted by the model to recover the original information bits.

3. Proposed Method

This section outlines the architecture and components of VLCMnet, a novel approach for modulation format recognition in VLC systems. We detail the overarching structure of VLCMnet, followed by specific discussions on the roles and functionalities of the TCN-LSTM and MMAnet modules within our system.

3.1. Overall Architecture of VLCMnet

As illustrated in Figure 1c, VLCMnet's architecture integrates three primary components: a TCN-LSTM module signal equalization module, a signal preprocessing module, and a MMAnet module. Firstly, the TCN-LSTM module reconstructs $y(n)$ into $x'(n)$. Subsequently, preprocessing operations, such as removing the CP and generating constellation diagrams, are performed. Finally, through the extraction of constellation diagram features by MMAnet, the modulation format is recognized.

3.2. TCN-LSTM for Channel Equalization

In the channel, noise and multipath fading cause diffusion and displacement of points on the constellation diagram. To mitigate these adverse effects, we propose the TCN-LSTM model to achieve channel equalization, as shown in Figure 2. This is because LSTM, although powerful, is not particularly adept at capturing local features in both space and time for transient distortions in communication channels. The TCN network complements LSTM's limitations in this regard.

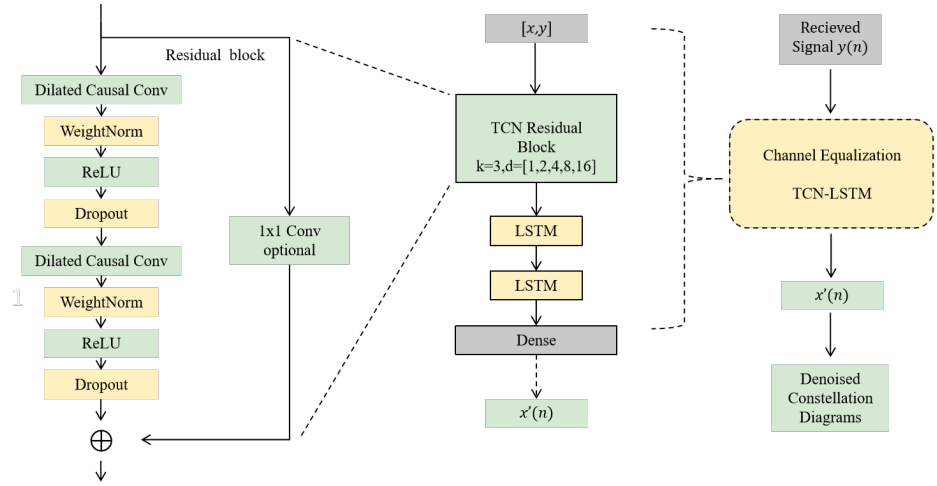


Figure 2. The hierarchical structures of TCN-LSTM.

The model training process is as follows: Using a sliding window to segment the input signal, the sequence $y(n)$ is fed into five residual modules in TCN to be computed. Each module consists of one-dimensional dilated convolutions, causal convolutions, weighted normalization, and one-dimensional convolutional residual connections. By expanding the receptive field and alleviating gradient vanishing, noise-induced point spread and constellation map distortion can be recognized and mitigated. The output from the TCN network further undergoes processing by two layers of LSTMs. The first LSTM layer captures short-term and local dependencies within the signal sequence, removing easily discernible high-frequency noise or periodic disturbances. The second LSTM layer more effectively captures long-range temporal dependencies, demonstrating increased robustness when dealing with non-stationary noise and abrupt signal changes. The model employs the mean squared error (MMSE) as its loss function. Through training, the fully connected layer outputs predict results $x'(n)$, aiming to restore the original transmitted data sequence $x(n)$ as closely as possible.

3.3. MMAnet for Modulation Format Recognition

Traditional convolutional neural networks, relying solely on local convolution operations, are unable to fully capture the global spatial layout and structural relationships in constellation diagrams. A single attention mechanism can only focus on one aspect of the features, whereas multiple attention mechanisms can extract features from multiple perspectives, constructing a more complex nonlinear model that enhances the model's expressive capability. We propose the MMAnet model, aimed at enhancing the accuracy

of MFR in severely distorted channel environments. It comprises three main modules: a shallow feature extraction (SFE) module, a multi-channel mixed attention (MCMA) module, and a feature fusion classification (FFC) module, as shown in Figure 3.

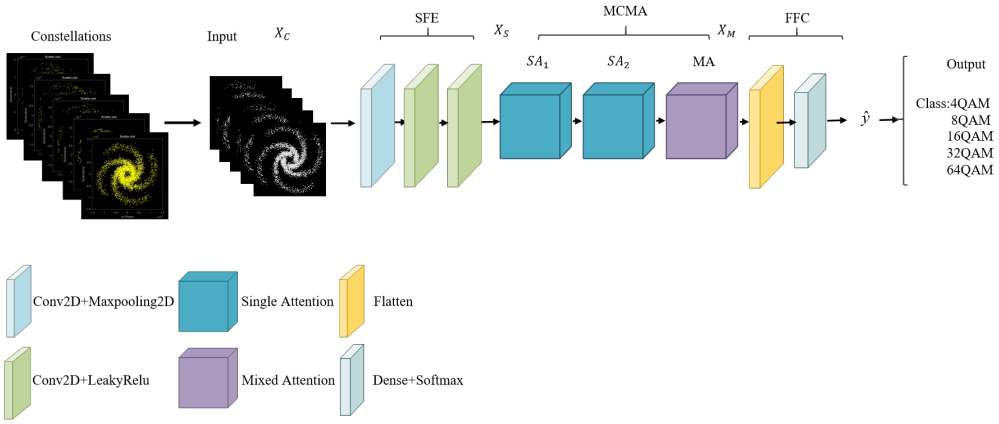


Figure 3. The overall network framework of our proposed MMAnet.

The input data consist of constellation diagrams after preprocessing steps such as channel equalization and CP removal. The training samples are denoted as X_C, which are composed of vectors $[x, y]$, where x represents a grayscale matrix of the constellation diagram with dimensions 64 × 64, and y indicates the corresponding modulation format type.

The SFE module is composed of three convolutional layers. Through this module, the raw constellation diagrams undergo a progressive extraction and transformation of features from lower to higher levels, culminating in the output vector X_S.

The MCMA submodule is the key component of the MMAnet module. It consists of two single attention (SA) submodules and one mixed attention (MA) submodule, as shown in Figure 4. After the input image X_S undergoes processing through the SA modules, crucial features X_{SA2} are extracted from critical regions on the constellation diagram. Subsequently, the MA submodule focuses channel attention and spatial attention onto feature responses across different channels, as well as the positions and adjacency relationships among the constellation points. Ultimately, this enables the classification model to enhance its understanding of the modulation signal characteristics, outputting effective features X_M.

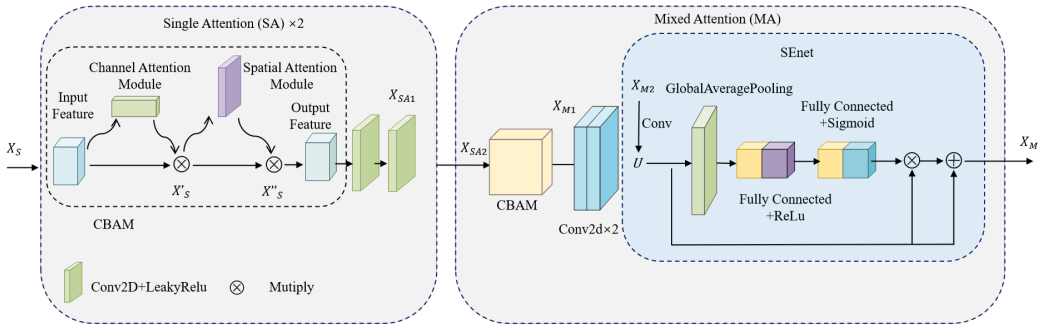


Figure 4. The architecture of the proposed MCMA module.

The SA module employs a single attention mechanism to process constellation diagrams. It consists of a convolutional block attention module (CBAM) and two convolutional layers. To enhance the model's ability to learn complex and subtle features in constellation charts, here, we use two consecutive single-attention modules in series with the aim to realize multi-level feature extraction and dynamically adjust the focus on different features. Input feature X_S first passes through the channel attention module (CAM) in CBAM, using average pooling and max pooling techniques to compress the spatial dimensions of the

input features. Subsequently, the compressed feature vectors are processed through two fully connected layers. Finally, the generated channel attention coefficients vector $M_c(X_S)$ is element-wise multiplied with the original input feature map X_S to produce a weighted feature map X'_S. The computation formula is as follows:

$$\begin{aligned} M_c(X_S) &= \sigma[MLP(AvgPool(X_S))] + MLP[MaxPool(X_S)] \\ &= \sigma(W_1(W_0(X_{S_{avg}}^c))) + W_1(W_0(X_{S_{max}}^c)), \end{aligned} \tag{6}$$

$$X'_S = M_c(X_S) \otimes X_S, \tag{7}$$

where W_0 and W_1 are the weight parameters of the two fully connected layers, respectively.

After filtering out redundant information, the spatial attention module (SAM) focuses on the spatial layout of feature maps pertaining to different modulation formats. Firstly, by performing global average pooling and global maximum pooling on the input feature maps, spatial feature statistics information at different scales is obtained. Finally, two fully connected layers are employed to generate the attention weight map. The formula for the output X''_S is:

$$M_s(X'_S) = \sigma(f^{7\times7}([AvgPool(X'_S); MaxPool(X'_S)])) \tag{8}$$

$$X''_S = M_s(X'_S) \otimes X'_S. \tag{9}$$

After obtaining the attention weight maps, further processing is carried out through two subsequent convolutional layers to maintain the depth of the entire network and refine and transform the features guided by attention. The output vector from the first-stage SA module is denoted as X_{SA1}. This process is then repeated, culminating in the final output referred to as X_{SA2}.

The feature maps are fed into the MA module, which is capable of integrating feature information across different scales, thereby enabling the model to adapt to variations in constellation diagrams under varying channel conditions. The MA module consists of a CBAM as well as a squeeze-and-excitation networks (SE) submodule. The input vector X_{SA2} undergoes processing by the CBAM to generate an attention weight map X_{M1}. Following this, the output vector X_{M2} is produced through two convolutional layers. The SE module enhances the dynamic selection and amplification of features within the constellation diagram of the input modulated signal. Initially, after the convolutional operations, an output vector U is obtained, consisting of C feature maps, each with dimensions $H \times W$. Each u_c represents the cth two-dimensional matrix within U. Subsequently, the squeeze operation, achieved through global average pooling, maps the output dimension from $H \times W \times C$ to a $1 \times 1 \times C$ matrix, which is the statistical vector of global information for each channel. Finally, the excitation operation carried out by the two fully connected layers performs nonlinear transformations to model and reweight the importance of each channel. The formula for calculating each channel attention weight vector s is:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2\delta(W_1 z)), \tag{10}$$

where F_{ex} represents the dynamic weighting function for channel features. z denotes the input feature map, W_1 signifies the weight matrix of the reduction layer, and W_2 represents the weight matrix of the expansion layer. δ denotes the rectified linear unit (ReLU) activation function, and σ represents the sigmoid activation function. F_{scale} denotes a scale adjustment operation function. s_c refers to the cth element of the attention weight vector, while u_c represents the feature of the cth channel in the input feature map. The final output X_M is derived as:

$$X_M = F_{scale}(u_c, s_c) = s_c u_c. \tag{11}$$

The FFC module consists of a flattening layer and a dense layer. The constellation map features X_M extracted by the preceding module undergo nonlinear transformations

within the FFC, followed by the computation of matching probabilities through the Softmax activation function to generate the output $\hat{y}$. The computation formula is as follows:

$$\hat{y}_i = soft\max(o) = \frac{\exp(o_i)}{\sum_k \exp(o_k)}. \quad (12)$$

The output of training samples is represented using one-hot encoding, denoted as $y = [y_1, y_2, ..., y_5]^T$. The highest predicted value in $\hat{y}_i$ is taken as the output category y_p, which represents the category of the modulation format.

4. Results and Discussion

4.1. Experiment Details

The VLC experimental setup encompasses an array of equipment, including a source computer, an arbitrary function generator, an amplifier, a bias-T, an LED, a photodiode, and a mixed-domain oscilloscope. The parameters for these components are meticulously outlined in Table 1.

Table 1. Devices and parameters of the VLC system prototype.

Device/Parameter	Value
Arbitrary function generator	Tektronix AFG3152C
Amplifier	Mini-Circuits ZHL-6A-S+
Bias-T	SHWBT-006000-SFFF
PD	PDA10A-EC
Responsivity of PD	0.44 A/W at 750 nm
Mixed-domain oscilloscope	Tektronix MDO
Power of LED	7.35 W

The entire model training involves TCN-LSTM for channel equalization on received signals, as well as MMAnet for classifying the generated constellation diagrams. Table 2 lists the configurations of the training parameters. It also presents the training parameters of the original 9-layer CNN before the algorithmic improvements.

Table 2. Network training parameters.

Parameters	TCN-LSTM	CNN	MMAnet
Optimizer	Adam	Adam	Adam
Learning rate	0.0001	0.00001	0.00001
Epochs	2000	200	200
Datatype	Sequence	Image	Image
Loss fuction	mse	categorical_crossentropy	categorical_crossentropy
Activation function	tanh	softmax	softmax
Batch size	64	32	32

The training dataset consists of 21,000 constellation diagrams, while the testing dataset includes 5250 constellation diagrams. The data were collected from received signals under 21 different SNR conditions ranging from −10 dB to 30 dB with an interval of 2 dB. Both the original CNN and the MMAnet models have their input data undergo an image preprocessing procedure before training, including cropping, grayscale conversion, and normalization. First, the CV2 library is utilized to resize the images to a dimension of 64×64. Following this, transformation is applied to obtain grayscale images of the constellation diagrams. Lastly, each pixel value is divided by 255. The experiments were executed using Python 3.9 and Keras 2.6.0 on an NVIDIA GeForce RTX 3060 GPU (NVIDIA, Santa Clara, CA, USA).

4.2. Evaluation Metrics

To quantitatively assess the performance of our model, we employed standard evaluation metrics, including the loss value, accuracy, recall, F1-score, kappa coefficient, and mean average precision (mAP). The kappa coefficient measures the consistency between model predictions and true labels (values range from 0 to 1, with higher values indicating superior consistency). The mAP represents class-wide average precision, which is calculated by a weighted average, combining the recall and precision of each class. The formulas for these metrics are as follows:

$$mAP = \frac{1}{n}\sum_{1}^{n}(AP)_i, \tag{13}$$

$$kappa = (p_0 - p_e)/(1 - p_e) = (\sum_{i=1}^{n} x_{ii}/N - \sum_{i=1}^{n} x_{i+}x_{+i}/N^2)/(1 - \sum_{i=1}^{n} x_{i+}x_{+i}/N^2), \tag{14}$$

where x_{ii} denotes diagonal elements in the confusion matrix, x_{i+}, x_{+i} represent the sum of the row and column elements for i, respectively, and N is the total element count. p_0 denotes the agreement observed, while p_e denotes the agreement expected by chance.

4.3. Channel Equalization Effects on Classification Performance

We trained a 9-layer CNN model as our baseline for evaluating VLC modulated signal recognition capabilities. The detailed structural parameters of this benchmark model are shown in Table 3. The classification performance of the benchmark model was compared with classical image classification networks, including Googlenet, Vgg16, Alexnet, and Resnet.

Table 3. Parameter settings of CNN.

Layer	Filters	Kernel Size	Layer	Filters	Kernel Size
Conv2D-1	64	7×7	MaxPooling	\	2×2
MaxPooling	\	2×2	Conv2D-6	256	3×3
Conv2D-2	128	3×3	Conv2D-7	256	3×3
Conv2D-3	128	3×3	MaxPooling	\	2×2
Conv2D-4	64	3×3	Conv2D-8	64	5×5
Conv2D-5	64	3×3	Conv2D-9	64	5×5

The experiments were performed with five prevalent convolutional neural network models for modulation recognition without channel equalization, respectively. Figure 5 presents a comparative analysis of their test accuracies. There is a uniform improvement in recognition accuracy across all models as SNR increases. The CNN model performs best, followed by Googlenet. When the SNR is higher than 26 dB, all models can achieve an accuracy of 100%. The CNN model achieves a modulation recognition accuracy of 80% when SNR is 14 dB, which represents an improvement of about 2 dB to 8 dB over the other four models. But at SNRs below 10 dB, the recognition accuracies are below 60%. Due to the possibility of noise masking the unique time-frequency characteristics inherent in different modulation signals, the challenge of improving recognition accuracy persists. Consequently, we deploy a TCN-LSTM model at the receiver end with the objective of enhancing the quality of the constellation diagrams.

After adding the TCN-LSTM module, both the CNN and Googlenet models achieved an average accuracy improvement of over 6%, as depicted in Figure 6. In the range from 0 dB to 10 dB, the CNN model exhibited significant performance enhancement. At an SNR of 6 dB, the CNN model reached an accuracy close to 100%. The threshold of the signal-to-noise ratio required for accurate signal recognition by the CNN model dropped by 14 dB compared to Googlenet.

The changes in constellation diagrams before and after equalization can be observed in Figure 7. The first row of this figure shows the constellation diagrams of the five modulated signals without channel equalization. It visually illustrates that noise-induced

point spreading diminishes the clarity between adjacent constellation points. After channel equalization by TCN-LSTM, the amplitude and phase distortions introduced by the channel are significantly mitigated. The constellation points are more clustered, the distribution becomes more regular, and the edge clarity improves. Therefore, it becomes easier to distinguish the modulated signals, especially the 8QAM signals.

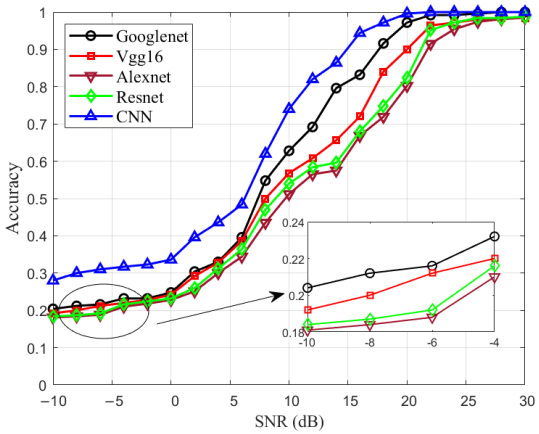


Figure 5. The recognition accuracy of five models on test samples at various SNRs.

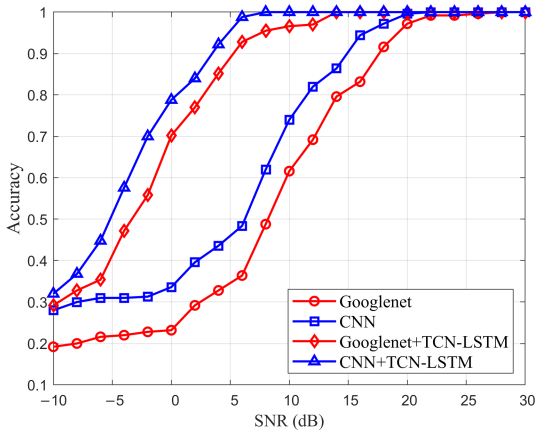


Figure 6. A comparison of modulation recognition accuracy before and after incorporating the TCN-LSTM module into both the CNN and GoogleNet models.

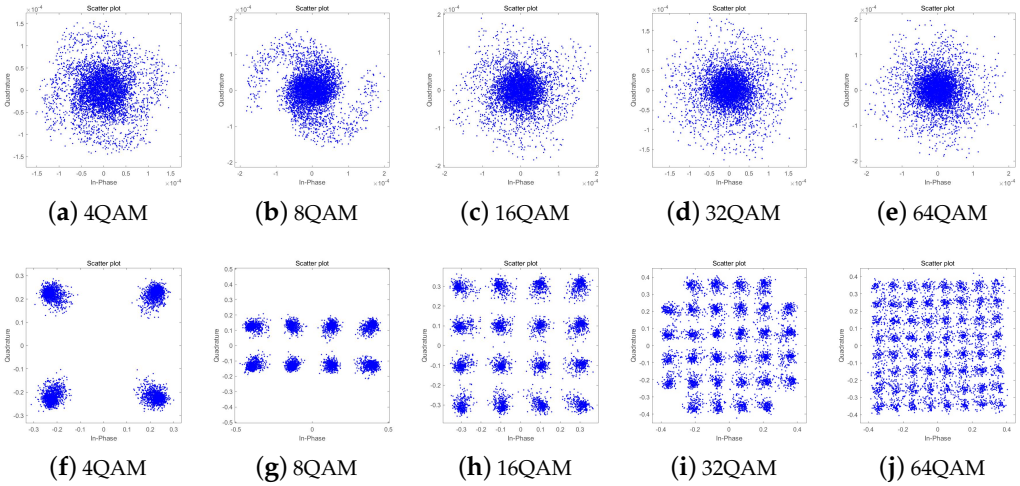


Figure 7. Comparing the constellation diagram quality for modulated signals under two conditions. (**a**–**e**) Without TCN-LSTM. (**f**–**j**) With TCN-LSTM.

The following experiments illustrate the effect of our proposed channel equalization model on the overall recogniton accuracy of the different modulated signals. Figure 8 presents a detailed comparison of the performance differences before and after combining the TCN-LSTM model with the CNN model, leading to the following conclusions:

(1) After adding a TCN-LSTM module for channel equalization prior to modulation recognition, the modulation recognition accuracy of 4QAM, 8QAM, 16QAM, 32QAM, and 64QAM increased by 4.2%, 1.7%, 8.1%, 6.8%, and 6.1%, respectively. The TCN model efficiently extracts both local and global temporal features, which contributes to the attenuation of high-frequency noise. The LSTM, through its gating mechanism, captures time-contextual information within the signal. Two models are cascaded to achieve complementary advantages.

(2) There are differences in recognition accuracy among the various signal types. Lower-order QAM modulation formats generally exhibit higher recognition accuracy. Higher-order modulation signals tend to suffer from more misclassifications or missed detections compared to lower-order modulation. The main reason is that lower-order QAM constellation diagrams have simpler distributions and larger point spacings. Due to interference, such as burst noise and multipath fading, the constellation diagrams of 16QAM, 32QAM, and 64QAM become more alike due to blurring.

(3) There are two noteworthy recognition results. After channel equalization, there is only a marginal improvement in the recognition accuracy for 8QAM. The recognition rate for 32QAM remains higher than that of 16QAM and 64QAM. The underlying reason is that the constellation pattern of 8QAM distinctly differs from those of other signals. Even under poor channel conditions that cause blurriness in the constellation diagrams, 8QAM is relatively easier to discern. For 16QAM and 64QAM, their constellation diagrams are arranged in a square lattice pattern, making them quite similar in shape. The 32QAM scheme is composed of two orthogonally superimposed 16QAM constellation diagrams. It possesses unique boundary shapes at the points of maximum and minimum amplitudes and thereby has relatively higher distinctions.

Although the use of the TCN-LSTM model proposed in this paper to accomplish channel equalization results in more time-consuming training of the entire modulation recognition model, the prediction execution time of the TCN-LSTM + CNN model during testing is only 0.35 s longer than that of the CNN model alone. In summary, while the introduction of the TCN-LSTM model has indeed led to an increase in the overall average accuracy of the CNN model from 0.795 to 0.838, its performance remains unsatisfactory when dealing with the recognition of higher-order modulation formats. Therefore, deeper mining of constellation diagram features for higher-order modulation formats is needed.

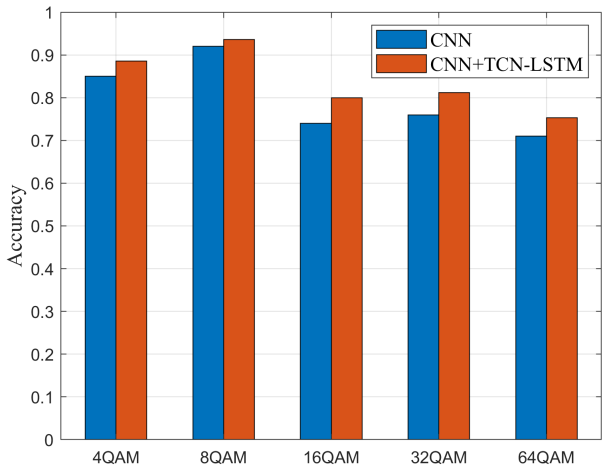


Figure 8. Prediction accuracy for different modulation formats across all SNRs.

4.4. MFR Performance Analysis Based on VLCMnet

In this section, the modulation format recognition model employs the sub-module MMAnet of our proposed VLCMnet model, which introduces a multi-attention mechanism into the original 9-layer CNN model. Based on the experimental results, we initially assessed the models using the loss function, followed by conducting a comparative analysis of the recognition accuracy of different models, and ultimately analyzed the differences in constellation diagram features extracted by the convolutional layers before and after the attention mechanism is integrated. The specific network structure and corresponding output dimensions are presented in Table 4.

Table 4. The structural parameters of MMAnet.

	Layer Name	Output Shape
SFE	Conv2d-1	(64, 64, 64)
	MaxPooling2D	(32, 32, 64)
	Conv2D-2	(32, 32, 128)
	Conv2D-3	(32, 32, 128)
SA1	cbam_block1	(32, 32, 128)
	Conv2D-4	(32, 32, 64)
	Conv2D-5	(32, 32, 64)
SA2	cbam_block2	(32, 32, 64)
	Conv2D-6	(32, 32, 256)
	Conv2D-7	(32, 32, 256)
MA	cbam_block3	(32, 32, 256)
	Conv2D-8	(32, 32, 64)
	Conv2D-9	(32, 32, 64)
	Se_block	(32, 32, 64)
FEC	Flatten	(65, 536)
	Dense	(5)

The following compares the convergence speed and fitting degree of the CNN and MMAnet models. The curves showing the change in training and validation losses are displayed in Figure 9. For both models, the validation losses do not notably exceed the training losses, indicating no signs of overfitting. MMAnet reaches a loss value of 0.05 at the 50th generation and converges after the 100th generation, suggesting a faster convergence rate compared to CNN. The validation loss curve for CNN exhibits slight oscillations, and its validation loss is on average 0.1 higher than that of MMAnet, indicating that CNN is less capable of learning from data with substantial noise compared to MMAnet.

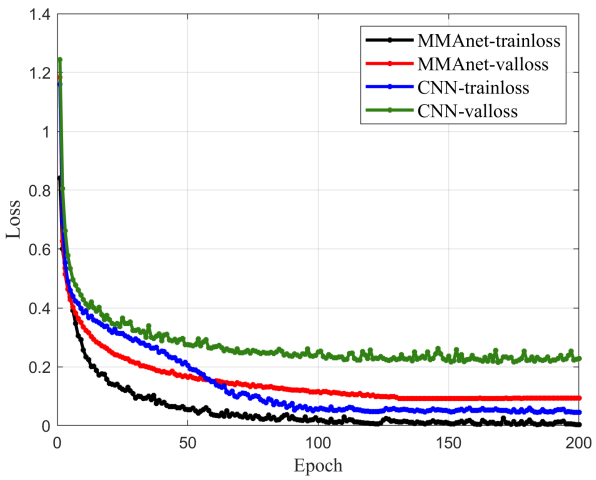


Figure 9. Comparison of CNN and MMAnet loss function curves.

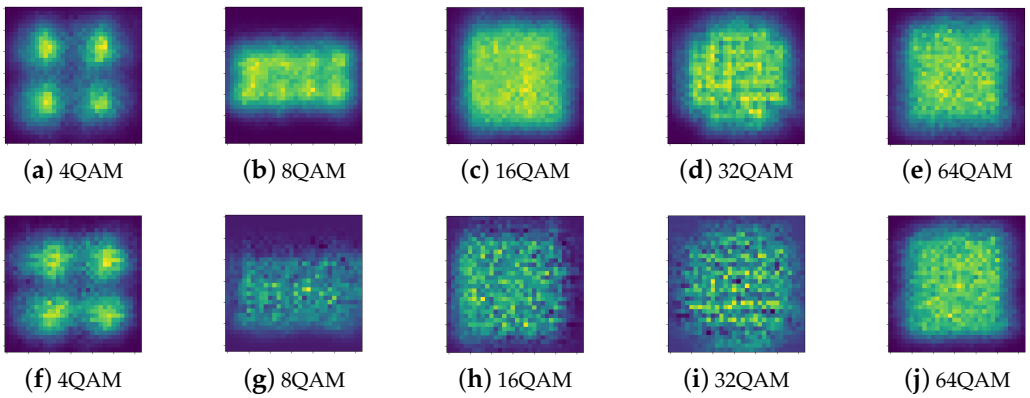
Our proposed algorithm can enhance the recognition accuracy of modulation formats in indoor visible light communication under severe channel distortion. Table 5 presents the accuracy and mAP values of six models across a range of SNRs from 0 dB to 12 dB. The experimental results demonstrate that, under extremely adverse channel conditions (SNRs of 0 dB and 4 dB), the VLCMnet model achieves recognition rates exceeding those of other models by more than 7%, and also shows an improvement of over 6% in mAP for various categories of modulation formats. At an SNR of 12 dB, all six models are virtually capable of correctly identifying all sample categories in the test set.

Table 5. Comparison of test accuracies for six models under different SNRs.

	SNR (dB)	0	4	8	12	Average
ACC	TCN-LSTM + Alexnet	0.648	0.724	0.916	0.984	0.782
	TCN-LSTM + Resnet	0.684	0.802	0.968	0.996	0.799
	TCN-LSTM + Vgg16	0.728	0.824	0.985	1	0.814
	TCN-LSTM + Googlenet	0.732	0.902	0.988	1	0.822
	TCN-LSTM + CNN	0.788	0.922	1	1	0.838
	VLCMnet	0.933	0.992	1	1	0.948
mAP	TCN-LSTM + Alexnet	0.674	0.743	0.934	0.994	0.791
	TCN-LSTM + Resnet	0.712	0.825	0.984	1	0.812
	TCN-LSTM + Vgg16	0.752	0.847	1	1	0.826
	TCN-LSTM + Googlenet	0.754	0.914	1	1	0.833
	TCN-LSTM + CNN	0.822	0.941	1	1	0.847
	VLCMnet	0.936	1	1	1	0.955

The TCN-LSTM channel equalization algorithm plays a pivotal role here. Concurrently, it is noted that MMAnet adopts a hybrid attention mechanism, dynamically allocating attention weights to the most critical channels and spatial locations, enabling the model to better focus on key spatial regions that determine the position and distribution of constellation points, thereby ignoring background noise or other non-critical spatial information. So this model improves its fine-grained parsing capability of constellation diagrams, subsequently boosting both recognition accuracy and stability.

The VLCMnet model demonstrates a more accurate ability to identify multi-level QAM modulation formats even under severe channel distortion, indicating that the MMAnet submodule enhances the feature representation of the constellation diagram. We perform feature map visualization on the convolutional layers of both the TCN-LSTM + CNN model and the VLCMnet model. The experimental analysis examines the capacity of both models to capture fine-grained constellation diagram features at an SNR of 4 dB, as shown in Figure 10.



(**a**) 4QAM (**b**) 8QAM (**c**) 16QAM (**d**) 32QAM (**e**) 64QAM

(**f**) 4QAM (**g**) 8QAM (**h**) 16QAM (**i**) 32QAM (**j**) 64QAM

Figure 10. (**a**–**e**) Feature maps of the constellation diagrams extracted by CNN. (**f**–**j**) Feature maps of the constellation diagrams extracted by VLCMnet.

The feature maps of CNN have distinct features in 4QAM and 8QAM, and the detailed features in 16QAM, 32QAM, and 64QAM are not obvious. In contrast, there are numerous regions with varying color depths in VLCMnet. The internal features of the 16QAM, 32QAM, and 64QAM constellation diagrams are assigned different weights. It indicates that after incorporating multiple attention mechanisms, the convolutional layers focus more intently on the image areas that are most crucial for the recognition task, resulting in clearer and more pronounced feature outlines and structures within the feature maps, which are related to the arrangement patterns and distribution shapes of the constellations for different signals. Consequently, the VLCMnet model successfully captures intricate edge and interior features in the constellation diagrams, thereby enhancing the model's generalization capability.

4.5. The Impact of Multi-Attention Mechanisms on VLCMnet Performance

In this section, we discuss the impact of multiple attention mechanisms on the modulation format recognition performance of the VLCMnet model. Here, the model without any attention mechanism is denoted as (Base), the model using only a single attention mechanism is referred to as (W/O MA), the model employing only a mixed attention mechanism is termed (W/O SA), and the model using multi-attention mechanisms is denoted as (With All).

As shown in Figure 11, when different attention mechanisms are incorporated into the classification model, the accuracy for recognition consistently surpasses that of models with no attention mechanisms. Within the SNR range from −10 dB to 0 dB, the recognition accuracy can be improved by more than 15%.

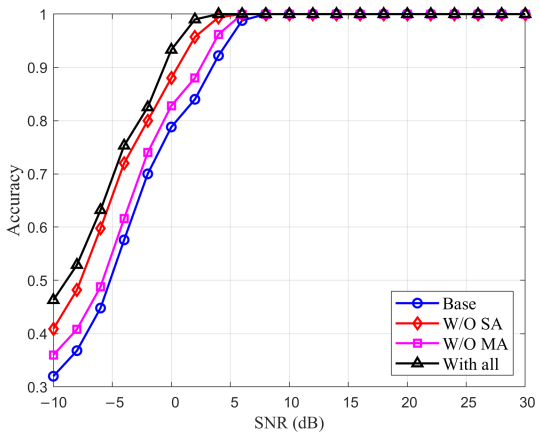


Figure 11. An analysis of ablation experiments on the modulation recognition model with the inclusion of different attention mechanisms.

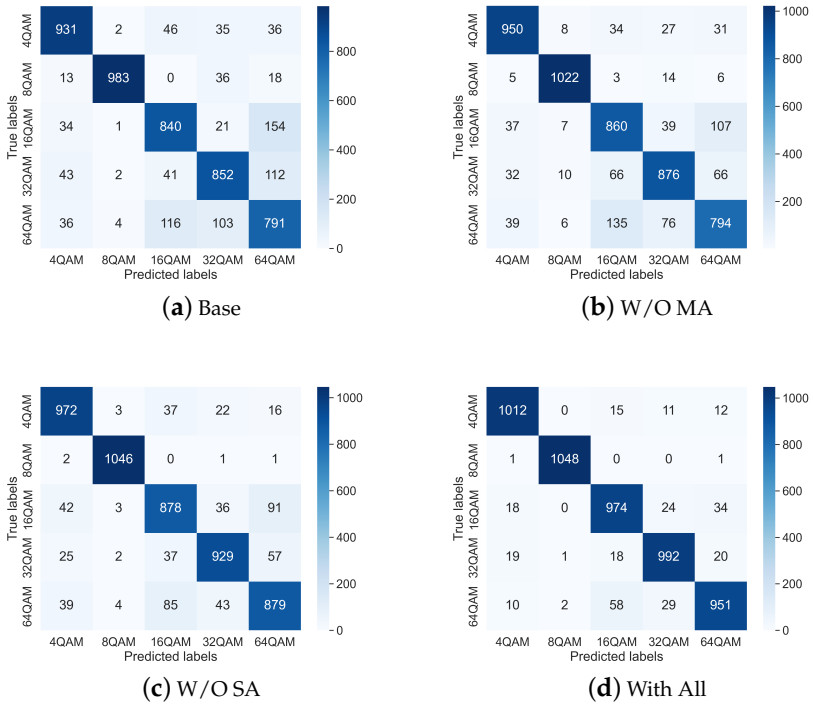
To evaluate the effect of different attention mechanisms on the recognition performance for each modulation format, an ablation study was conducted on 5250 test samples. The classification confusion matrix is shown in Figure 12.

(1) As shown in Figure 12a, without attention mechanisms, the features extracted from the convolutional layers tend to be susceptible to noise and irrelevant feature interference. While the model demonstrates relatively high accuracy in recognizing lower-order modulations, particularly 8QAM, there exist numerous errors in recognizing higher-order modulations. For instance, 154 instances of 16QAM are misclassified as 64QAM.

(2) Figure 12b shows that the recognition performance improves a little if only the SA module is introduced into VLCMnet. The number of misclassifications for higher-order modulations decreases but not significantly. The recognition accuracy of 64QAM is almost unchanged. It is easily confusedly recognized as 16QAM and 32QAM, possibly due to inadequate extraction of spatial structure features within the distribution of constellation points.

(3) Figure 12c shows that when solely the MA module is incorporated, VLCMnet exhibits a better enhancement in recognition performance compared to when only the SA module is added. The recognition accuracy for all five modulation formats improves. In particular, the recognition accuracy of 64QAM is improved by 11%. The proportion of 64QAM being misclassified as 32QAM notably declines.

(4) Figure 12d demonstrates that the inclusion of both SA and MA modules in VLCMnet is more helpful in improving the model performance. The misclassification between 16QAM and 64QAM is further reduced. The recognition accuracy for higher-order modulations increases by more than 13% compared to the Base model.



(**a**) Base (**b**) W/O MA

(**c**) W/O SA (**d**) With All

Figure 12. Confusion matrixes at Base, W/O MA, W/O SA, and With All.

In summary, VLCMnet is able to concentrate on critical point-position features via its SA module and simultaneously attend to key spatial structures in the distribution of constellation points through its MA module, thereby enhancing the model's recognition accuracy and robustness, particularly for higher-order QAM modulation signals.

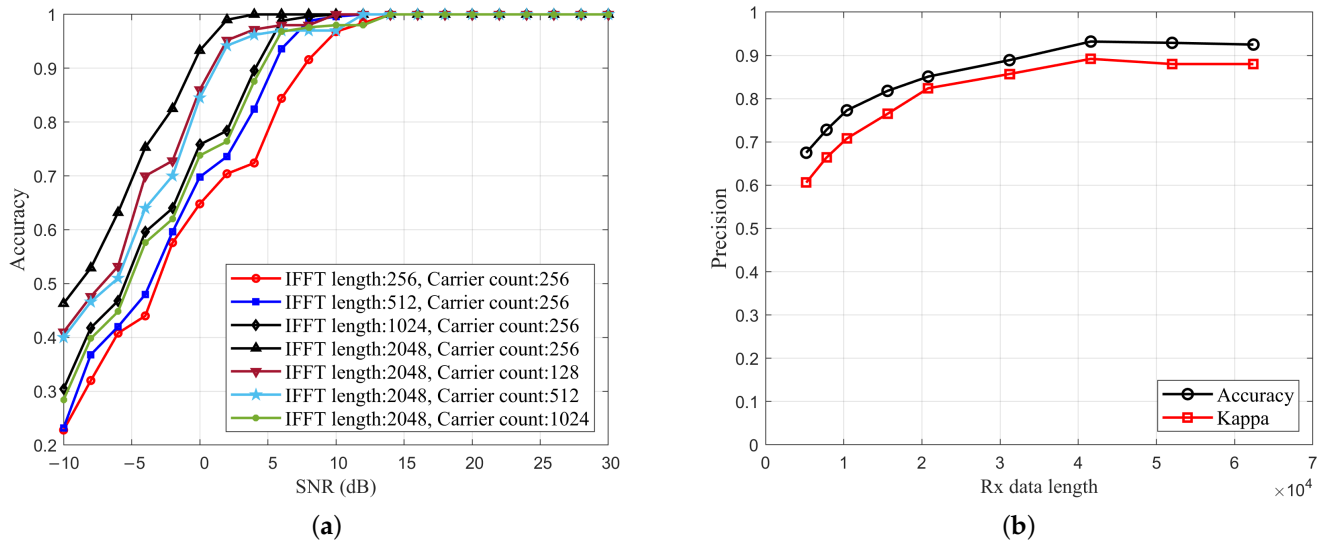
4.6. Comprehensive Evaluation of Model Performance

This section first discusses the impact of system transmission data volume indicators on model performance and subsequently conducts a comparative analysis of the generalization capability and network complexity among six models. The IFFT length and the carrier count represent the number of subcarriers used by the transmitter to generate OFDM symbols and actually carry the data payload.

As shown in Figure 13a, the recognition accuracy of the model increases with growth in the IFFT length. When the carrier count is 256 and the IFFT length is 2048, the model performance outperforms other combinations of carrier counts. However, both parameter values should not be excessively large, as this can lead to oversampling, causing severe point overlapping in the constellation diagram, which, in turn, reduces the recognition accuracy.

The Rx data length refers to the length of the effective data payload at the receiver end. As illustrated in Figure 13b, when the Rx data length ranges from 20,000 to 60,000, the model accuracy generally exhibits an upward trend as the data length increases. The model performs optimally at a length of 41,600. However, excessive increase in the Rx data length

can, conversely, introduce more noise and redundant information, reducing the clarity of the constellation diagram and recognition accuracy.



(**a**) (**b**)

Figure 13. (**a**) Comparison of recognition accuracy for different channel parameters. (**b**) Comparison of recognition precision for different Rx data length.

In the following, the MMAnet module within the VLCMnet model is replaced with five alternative convolutional neural networks, and a comprehensive comparison of the classification performance and complexity differences among these models is presented, as shown in Table 6. The model complexity is assessed using million floating-point operations per second (MFLOPs) and the number of parameters. The generalization ability of the model is evaluated using the F1-score, recall, and accuracy metrics.

Table 6. Comparison of the comprehensive performance of the models.

Method	MFLOPs	Parameters (M)	Recall	F1	Accuracy
Alexnet	43.19	21.59	0.783	0.781	0.782
Resnet	42.56	21.28	0.801	0.800	0.799
Vgg16	33.75	16.88	0.816	0.815	0.814
Googlenet	12.01	6.01	0.823	0.822	0.822
CNN	3.2	1.6	0.839	0.838	0.838
MMAnet	3.9	1.9	0.949	0.948	0.948

The experimental results indicate that MMAnet achieves higher precision with relatively lower computational requirements. Although CNN has the lowest complexity, its accuracy is 11% lower than that of MMAnet. The precision and complexity of Googlenet are moderate. AlexNet and Vgg16 are both complex models and show inferior precision. Therefore, the proposed model in this paper possesses good generalization capabilities, fast training speed, and convenient model deployment.

In Section 4.3, it is reported that the average accuracy achieved by a 9-layer CNN model for modulation recognition was only 0.795, whereas our proposed VLCMnet reaches an accuracy of 0.948, representing a 19.2% improvement over the CNN model. This indicates that the model effectively mitigates channel-induced interference and accurately captures and discriminates key features of different modulation formats.

5. Conclusions

In this paper, we propose the VLCMnet model for DCO-OFDM systems, an integrated deep learning-based network model aimed at addressing the problem of low accuracy in

modulation format recognition under severe channel distortion conditions. We analyze the model characteristics from several aspects, including the changes in model performance before and after channel equalization, the effects of different attention mechanisms on model performance, the model complexity, and the effects of the system input parameters. The experimental results show that the TCN-LSTM module effectively extracts temporal and frequency local features of the signals, and performs repair on impaired signals. The MMAnet module introduces a multi-attention mechanism in the CNN, which captures diverse feature information from distinct subspace perspectives, and contributes to improving recognition accuracy. In summary, the modulation format recognition model presented in this paper can effectively reduce the confusion and misclassifications for higher-order modulation formats. It exhibits strong robustness, computational efficiency, and effectiveness. This model plays a significant role in promoting the task of modulation recognition under complex channel environments.

Author Contributions: Conceptualization, X.Z. and Y.H.; methodology, X.Z. and P.M.; software, X.Z. and C.Z.; validation, X.Z., Y.H., and C.Z.; formal analysis, X.Z.; investigation, X.Z. and C.Z.; resources, Y.H. and P.M.; writing—original draft preparation, X.Z.; writing—review and editing, Y.H. and X.Z.; funding acquisition, Y.H. and P.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (NSFC) under grants 61801257 and 61701271, and by the Qingdao University Research Program through project (ID: RH2000004239).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available from the first author upon request.

Acknowledgments: The authors express their gratitude to the editors and the anonymous reviewers for their insightful suggestions and general assistance.

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

References

1. Xiao, W.; Luo, Z.; Hu, Q. A Review of Research on Signal Modulation Recognition Based on Deep Learning. *Electronics* **2022**, *11*, 2764. [CrossRef]
2. Liu, K.; Li, F. Automatic modulation recognition based on a multiscale network with statistical features. *Phys. Commun.* **2023**, *58*, 102052. [CrossRef]
3. Hou, S.; Fan, Y.; Han, B.; Li, Y.; Fang, S. Signal Modulation Recognition Algorithm Based on Improved Spatiotemporal Multi-Channel Network. *Electronics* **2023**, *12*, 422. [CrossRef]
4. Leblebici, M.; Çalhan, A.; Cicioğlu, M. Deep learning-based modulation recognition with constellation diagram: A case study. *Phys. Commun.* **2024**, *63*, 102285. [CrossRef]
5. Peng, Y.; Guo, L.; Yan, J.; Tao, M.; Fu, X.; Lin, Y.; Gui, G. Automatic Modulation Classification Using Deep Residual Neural Network with Masked Modeling for Wireless Communications. *Drones* **2023**, *7*, 390. [CrossRef]
6. Marey, A.; Marey, M.; Mostafa, H. Novel Deep-Learning Modulation Recognition Algorithm Using 2D Histograms over Wireless Communications Channels. *Micromachines* **2022**, *13*, 1533. [CrossRef]
7. Ali, A.K.; Erçelebi, E. Modulation Format Identification Using Supervised Learning and High-Dimensional Features. *Arab. J. Sci. Eng.* **2023**, *48*, 1461–1486. [CrossRef]
8. He, J.; Zhou, Y.; Shi, J.; Tang, Q. Modulation Classification Method Based on Clustering and Gaussian Model Analysis for VLC System. *IEEE Photonics Technol. Lett.* **2020**, *32*, 651–654. [CrossRef]
9. Ağır, T.T.; Sönmez, M. The modulation classification methods in PPM–VLC systems. *Opt. Quantum Electron.* **2023**, *55*, 223. [CrossRef]
10. Zhou, Z.; Guan, W.; Wen, S. Recognition and evaluation of constellation diagram using deep learning based on underwater wireless optical communication. *arXiv* **2020**, arXiv:2007.05890.
11. Gu, Y.; Wu, Z.; Li, X.; Tian, R.; Ma, S.; Jia, T. Modulation Format Identification in a Satellite to Ground Optical Wireless Communication Systems Using a Convolution Neural Network. *Appl. Sci.* **2022**, *12*, 3331. [CrossRef]

12. Mohamed, S.E.D.N.; Mortada, B.; Ali, A.M.; El-Shafai, W.; Khalaf, A.A.M.; Zahran, O.; Dessouky, M.I.; El-Rabaie, E.S.M.; El-Samie, F.E.A. Modulation format recognition using CNN-based transfer learning models. *Opt. Quantum Electron.* **2023**, *55*, 343. [CrossRef]
13. Li, J.; Ma, J.; Liu, J.; Lu, J.; Zeng, X.; Luo, M. Modulation Format Identification and OSNR Monitoring Based on Multi-Feature Fusion Network. *Photonics* **2023**, *10*, 373. [CrossRef]
14. Zhao, Z.; Khan, F.N.; Li, Y.; Wang, Z.; Zhang, Y.; Fu, H. Application and comparison of active and transfer learning approaches for modulation format classification in visible light communication systems. *Opt. Express* **2022**, *30*, 16351–16361. [CrossRef] [PubMed]
15. Ma, Y.; Gao, M.; Zhang, J.; Ye, Y.; Chen, W.; Ren, H.; Yan, Y. Modulation format identification based on constellation diagrams in adaptive optical OFDM systems. *Opt. Commun.* **2019**, *452*, 203–210. [CrossRef]
16. Liu, W.; Li, X.; Yang, C.; Luo, M. Modulation classification based on deep learning for DMT subcarriers in VLC system. In Proceedings of the 2020 Optical Fiber Communications Conference and Exhibition (OFC), San Diego, CA, USA, 8–12 March 2020; pp. 1–3.
17. Mortada, B.; El-Shafai, W.; Zahran, O.; El-Rabaie, E.S.; Abd El-Samie, F. Fan-beam projection based modulation classification for optical systems with phase noise effect. *J. Opt.* **2023**, *52*, 1–15. [CrossRef]
18. Wei, Z.; Zhang, J.; Li, W.; Plant, D.V. Active Learning-Aided CNN-Based Entropy-Tunable Automatic Modulation Identification for Rate-Flexible Coherent Optical System. *J. Light. Technol.* **2023**, *41*, 4598–4608. [CrossRef]
19. Xu, C.; Jin, R.; Gao, W.; Chi, N. Efficient Modulation Classification Based on Complementary Folding Algorithm in UVLC System. *IEEE Photonics J.* **2022**, *14*, 1–6. [CrossRef]
20. Chen, X.; Wang, Y.; Liu, X.; Wang, Z.; Zhang, X.; Zhou, F.; Ng, D.W.K. A Hybrid Active-Passive Single-Order Equalizer for Visible Light Communication Systems. *IEEE Photonics Technol. Lett.* **2023**, *35*, 1395–1398. [CrossRef]
21. Sun, W. Research on the receiving and equalization of visible light communication system based on the light-emitting-diode emission process and mechanism. *Opt. Eng.* **2023**, *62*, 098102. [CrossRef]
22. Bostanoglu, M.; Dalveren, Y.; Catak, F.O.; Kara, A. Modelling and Design of Pre-Equalizers for a Fully Operational Visible Light Communication System. *Sensors* **2023**, *23*, 5584. [CrossRef] [PubMed]
23. Chen, C.; Nie, Y.; Liu, M.; Du, Y.; Liu, R.; Wei, Z.; Fu, H.Y.; Zhu, B. Digital Pre-Equalization for OFDM-Based VLC Systems: Centralized or Distributed? *IEEE Photonics Technol. Lett.* **2021**, *33*, 1081–1084. [CrossRef]
24. Chen, H.; Zhao, Y.; Hu, F.; Chi, N. Nonlinear Resilient Learning Method Based on Joint Time-Frequency Image Analysis in Underwater Visible Light Communication. *IEEE Photonics J.* **2020**, *12*, 1–10. [CrossRef]
25. Miao, P.; Yin, W.; Peng, H.; Yao, Y. Study of the Performance of Deep Learning-Based Channel Equalization for Indoor Visible Light Communication Systems. *Photonics* **2021**, *8*, 453. [CrossRef]
26. Li, Z.; Hu, F.; Li, G.; Zou, P.; Wang, C.; Chi, N. Convolution-Enhanced LSTM Neural Network Post-Equalizer used in Probabilistic Shaped Underwater VLC System. In Proceedings of the 2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Macau, China, 21–24 August 2020; pp. 1–5. [CrossRef]
27. Costa, W.S.; Samatelo, J.L.; Rocha, H.R.; Segatto, M.E.; Silva, J.A. Direct Equalization with Convolutional Neural Networks in OFDM based VLC Systems. In Proceedings of the 2019 IEEE Latin-American Conference on Communications (LATINCOM), Salvador, Brazil, 11–13 November 2019; pp. 1–6. [CrossRef]
28. Li, S.; Zou, Y.; Shi, Z.; Tian, J.; Li, W. Performance enhancement of CAP-VLC system utilizing GRU neural network based equalizer. *Opt. Commun.* **2023**, *528*, 129062. [CrossRef]
29. Chen, H.; Jia, J.; Niu, W.; Zhao, Y.; Chi, N. Hybrid frequency domain aided temporal convolutional neural network with low network complexity utilized in UVLC system. *Opt. Express* **2021**, *29*, 3296–3308. [CrossRef]
30. Bai, S.; Kolter, J.Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv* **2018**, arXiv:1803.01271.