ISSN: 2414-1895
DOI: 10.6919/ICJE.202507_11(7).0013

Research on Communication Signal Modulation Recognition Method based on Optimized VGG16 Network

Hongqing Ma*, Yanwen Wang, Tianchen Long, Feng Liu, Chang Cai

School of Electronic Information, Xijing University, Xi'an, China

*Corresponding author

Abstract

With the advancement of communication technology, Multicarrier Modulation (MCM) has emerged as a crucial technique for mitigating interference and achieving high-rate data transmission. The first step in demodulating signals during communication is to identify the modulation scheme. Cooperative communication utilizes pilot sequences for this purpose; however, non-cooperative environments rely on techniques such as modulation scheme recognition, channel estimation, and carrier frequency estimation. Leveraging the powerful feature extraction capabilities of deep learning, modulation recognition technology based on deep learning can automatically extract signal features to determine the modulation scheme. By employing the VGG16 convolutional neural network and utilizing a constructed dataset for the recognition of OFDM signals, a recognition accuracy of 91.9% was achieved on the dataset presented in this paper. An optimized version of the VGG16 neural network, referred to as VGG16R, is proposed. This network architecture treats convolutional layers as convolutional groups, adds a Batch Normalization (BN) layer after each convolutional layer, and introduces branches that sum the input image with the results of the convolutional groups, thereby forming a new feature set.

Keywords

OFDM; Modulation Recognition; Deep Learning; Convolutional Neural Network; VGG16.

1. Introduction

With the advancement of communication technology, the initial single-carrier modulation signals exhibited sensitivity to noise and issues related to insufficient resource utilization. To meet the growing demand for high-bit-rate data transmission, Multi-Carrier Modulation (MCM) technology has emerged as a prominent solution, characterized by its robust anti-interference capabilities. MCM technology has been widely applied across various fields, including drone image transmission systems, 4G/5G mobile networks, digital video broadcasting, and many others. Among these, Orthogonal Frequency Division Multiplexing (OFDM) is extensively utilized in 4G and 5G communication systems due to its high transmission efficiency, ease of implementation through Fourier transforms, and compatibility with Multi-Input Multi-Output (MIMO) systems[1].

OFDM is a multi-carrier modulation technique that can also be viewed as a multiplexing method. In multi-carrier transmission, the data stream is decomposed into multiple sub-bit streams, each with a bit rate significantly lower than that of the original stream. These low-rate multi-state symbols are then modulated and combined with their corresponding sub-carriers, forming a transmission system that sends multiple low-rate symbols in parallel. OFDM represents a special case of orthogonal frequency division multiplexing, distinguished by the orthogonality of its sub-carrier signals. This

Volume 11 Issue 7, 2025
DOI: 10.6919/ICJE.202507_11(7).0013

orthogonality allows the expanded spectrum to overlap, thereby not only reducing interference between sub-carriers but also significantly enhancing spectral efficiency.

In an OFDM system, each symbol is composed of multiple mutually orthogonal subcarrier signals. If we denote the number of subcarriers as N, then OFDM modulation transforms the binary bit stream input by the user into N parallel data streams, modulating them with carriers of different frequencies. At the receiving end, the corresponding carriers are utilized for demodulation.

1.1 Modulation Recognition

Modulation recognition is a technique used to detect and identify the modulation scheme present in a signal[2]. In communication systems, information is typically transmitted by altering the modulation scheme of the signal to accommodate various transmission media and communication requirements. Modulation is the process of converting digital or analog signals into a form suitable for transmission, while modulation recognition is the reverse process[3], which involves determining the modulation scheme used from the received signal.

The objective of modulation recognition is to ascertain the modulation scheme employed based on the characteristics of the received signal, enabling accurate demodulation and restoration of the original information. Modulation recognition finds extensive applications in fields such as wireless communication, radar systems, television broadcasting, and radio listening. By accurately identifying the modulation scheme of the received signal, it is possible to enhance the processing and decoding of the signal, thereby improving the performance and reliability of communication systems. With the advancement of deep learning, researchers have begun to leverage deep learning techniques for signal modulation recognition. Currently, the most advanced modulation recognition technologies fundamentally rely on deep learning to identify signals.

Traditional supervised recognition methods require a substantial amount of annotated samples for algorithm training, which is a critical prerequisite. These samples are essential as they facilitate a comprehensive exploration of known instances, ultimately aiding in the classification of unfamiliar samples. In contrast, traditional unsupervised recognition methods are rooted in the operational framework of clustering algorithms. Initially, these methods employ an initial model to cluster various features within the spatial distribution. Subsequently, they iteratively refine these clusters by redefining centroids until all samples within the spatial domain are appropriately classified and subsequently aggregated into clusters.

In the context of signal modulation recognition, preprocessing is applied to specify the signals intended for identification. Feature extraction techniques are then utilized to effectively extract relevant features, ultimately leading to the classification and recognition of the modulation schemes employed in the signals. Signal preprocessing is crucial for optimizing the signals, making them more suitable for subsequent feature extraction and direct information retrieval. In the domain of recognizing modulation signal patterns, common preprocessing techniques include filtering, wavelet transforms, noise estimation, time-frequency representation, and carrier frequency estimation, with the choice of preprocessing methods depending on the unique characteristics of the communication environment.

Traditional manual feature extraction methods necessitate skilled operators to extract feature parameters through multiple experiments, often resulting in significant errors. As researchers leverage the advantages of deep learning technologies, they recognize the capability of neural networks to autonomously extract and select relevant sample features. Consequently, deep learning techniques have gained widespread application in feature extraction and classification, facilitating more accurate signal modulation pattern recognition through the extraction of meaningful features.

1.2 Convolutional Neural Networks

Deep learning has emerged as a focal point in modern computer science, evolving over several decades from theoretical exploration to practical industrial applications. Notably, deep learning demonstrates significant advantages in addressing the complexities of model construction. Compared

ISSN: 2414-1895 DOI: 10.6919/ICJE.202507_11(7).0013

to traditional machine learning algorithms, it excels at recognizing the inherent complexities of data features. Consequently, deep learning has found extensive applications across various fields, including medical research, natural language processing, and image recognition.

Convolutional Neural Networks (CNNs) represent a quintessential algorithm within the realm of deep learning, characterized by their robust capability for data representation learning. Unlike traditional algorithms, CNNs autonomously adjust parameters and effectively extract the most salient features. Furthermore, while traditional algorithms necessitate preprocessing of data prior to feature extraction, CNNs can directly access the raw information of images, thereby reducing processing time. Typically, CNNs commence with data preprocessing at the input layer to facilitate subsequent operations. Feature extraction is then achieved through convolutional and activation layers, while pooling layers are responsible for data compression. Ultimately, classification is performed via fully connected layers.

2. VGG16 Convolutional Neural Network

VGGNet, developed by the Visual Geometry Group (VGG) at the University of Oxford, is a deep convolutional neural network named after the initials of the team[4]. VGGNet serves as a quintessential classification network, demonstrating that increasing the depth of the network by stacking layers enhances feature extraction capabilities, thereby improving model performance. VGGNet encompasses various architectures, with differences primarily arising from the varying network depths. Each VGGNet model is constructed by stacking block structures composed of convolutional and pooling layers. The convolutional layers utilize convolutional kernels to map local regions of the input feature maps, employing weight sharing to extract features. As the network deepens, the number of channels increases, resulting in richer feature information. Conversely, the fully connected layers play a crucial role in feature aggregation to accomplish classification tasks.

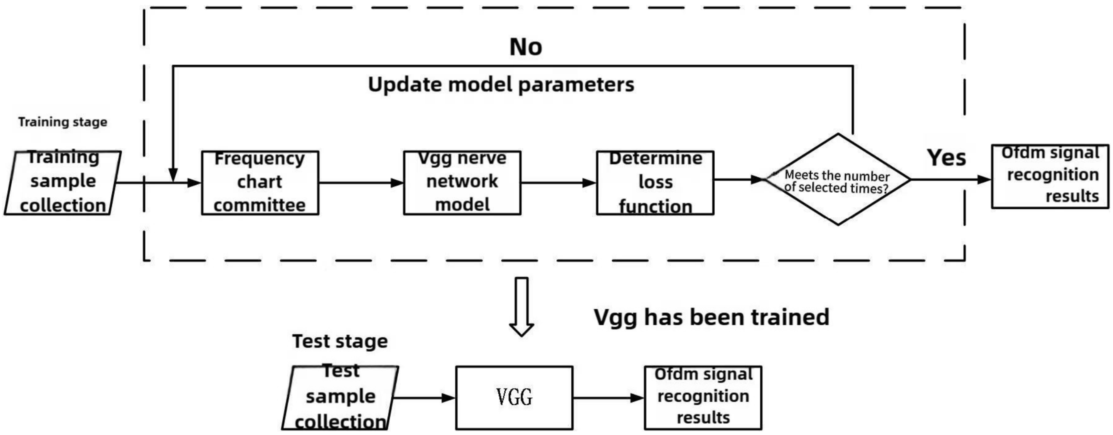


Figure 1. Flowchart of the OFDM signal recognition algorithm based on the VGG16 network.

As illustrated in Figure 1, during the training phase, the preprocessed signals are first extracted for their time-frequency spectral features, which serve as input characteristics to the VGG16 neural network model[5]. The loss function is then established, followed by an evaluation of the number of training iterations. If the criteria are not met, the model parameters are updated continuously until the iteration requirements are satisfied, at which point training is halted, yielding the recognition results.

As shown in Table 1, the VGG16 network comprises one input layer, thirteen convolutional layers, five pooling layers, three fully connected layers, and one output layer. It features five sets of convolution operations with channel counts of 64, 128, 256, 512, and 512, respectively. The nonlinear activation function employed is the ReLU function, and each set of convolution operations can be regarded as a block structure, facilitating feature extraction. The pooling layer has a parameter size

ISSN: 2414-1895 DOI: 10.6919/ICJE.202507_11(7).0013

of 2x2 and utilizes the max pooling method. The three fully connected layers have channel counts of 4096, 4096, and 1000, respectively. Finally, the output layer includes a Softmax classifier.

Table 1. VGG16 Network Structure Parameters.

Network Layer Name	Convolution Kernel Size	Pooling Method	Number of Channels	Classifier
Input				
Conv1, Conv2	3×3		64	
Pool1		max-pooling		
Conv3, Conv4	3×3		128	
Pool2		max-pooling		
Conv5, Conv6, Conv7	3×3		256	
Pool3		max-pooling		
Conv8, Conv9, Conv10	3×3		512	
Pool4		max-pooling		
Conv11, Conv12, Conv13	3×3		512	
Pool5		max-pooling		
FC1			4096	
FC2			4096	
FC3			1000	
Output				Softmax

3. Optimization Improvements based on the VGG16 Network Model

3.1 Incorporation of Batch Normalization Layers

Batch Normalization (BN) is a pivotal technique in deep neural networks, aimed at addressing issues such as vanishing and exploding gradients during the training process[6]. It normalizes the inputs of each layer to ensure that their mean is close to zero and their standard deviation is near one. Subsequently, the normalized values are rescaled and shifted using learnable scaling and shifting parameters, thereby preserving the network's expressive capability. The advantages of BN include accelerated training, enhanced gradient flow, regularization effects, and robustness against adversarial perturbations. It has become a standard technique in deep learning, widely applied across various neural network architectures and tasks, resulting in significant performance improvements. The computational process is as follows:

First, calculate the mean μ and variance σ of m samples:

$$\mu = \frac{1}{m}\sum_{i=1}^{n} x_i \tag{1}$$

$$\sigma = \frac{1}{m}\sum_{n}^{1}(x_i - \mu) \tag{2}$$

Volume 11 Issue 7, 2025
DOI: 10.6919/ICJE.202507_11(7).0013

Furthermore, calculate the average value:

$$\hat{x} = \frac{x_i - \mu}{\sqrt{\sigma^2 + \varepsilon}} \# \tag{3}$$

In this context, ε is a constant introduced to prevent the denominator from becoming zero. The output value yi is then reconstructed using the scaling factor γi and the shifting factor βi:

$$y_i = \gamma_i x_i + \beta_i \tag{4}$$

$$\gamma_i = \sqrt{Var[x_i]} \tag{5}$$

$$\beta_i = E[x_i] \tag{6}$$

3.2 Configuration of Basic Network Units

The VGG16 network model is relatively deep, encompassing a substantial number of parameters and computational demands, which consequently extends the training duration. During the convolution operations, there may be instances of feature information becoming blurred or lost, thereby impacting the accuracy of the experimental results. In this section, building upon the VGG16 network, we have incorporated cross-layer connections inspired by the ResNet architecture to develop a new model, referred to as VGG16R. The new structural unit is illustrated in Figure 2 below.

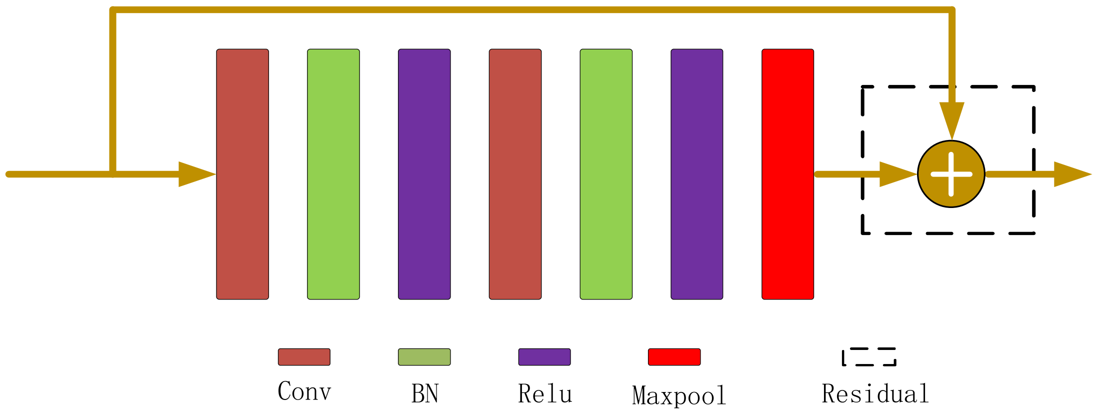


Figure 2. New Unit Structure.

The new architectural unit treats consecutive convolutional layers as a single convolutional group. Within each convolutional group, a batch normalization (BN) layer is added following each convolutional layer. Additionally, an extra branch is introduced at the input end of each convolutional group. This branch directly transmits the input image to subsequent layers, combining its output with the results obtained from the convolutional group to form a new set of features. During the fusion process, discrepancies may arise due to variations in kernel size or feature map dimensions. To address this issue, a residual module is implemented. When the dimensions and sizes of the results from the two branches match, a simple addition operation is performed. In cases where direct addition of the results from the two branches is not feasible, the features in the new branch are adjusted to match the dimensions or sizes of the convolutional group branch before the addition operation is executed.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202507_11(7).0013

The optimization changes to the network architecture unit have led to a more comprehensive extraction and retention of feature information within the convolutional layers. Consequently, the amount of feature information passed to the fully connected layers has increased, thereby enhancing the recognition accuracy of the neural network.

3.3 Overall Network Architecture

After the design of the new unit structure is completed, the network architecture is adjusted without altering the overarching framework of VGG16. The new basic units of the network are stacked into five groups. A new set of features, formed by adding two branches, is subsequently passed to the fully connected layer. The structural optimization is finalized, and this configuration is designated as the VGG16R model. The overall architecture is illustrated in Figure 3.

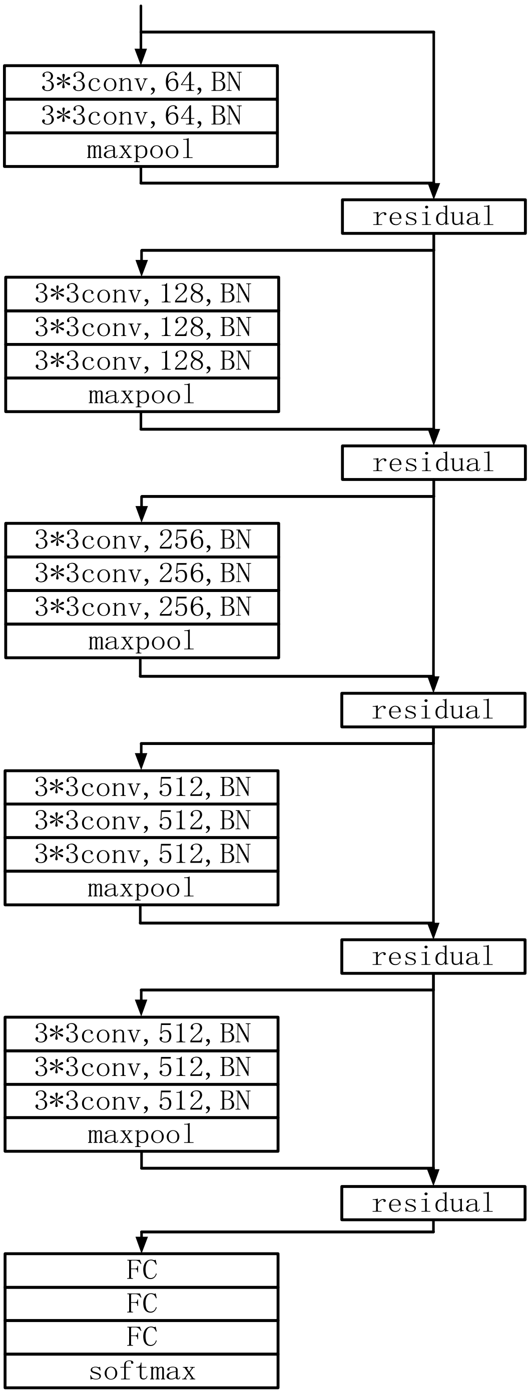


Figure 3. Overall Structure of VGG16R.

ISSN: 2414-1895
DOI: 10.6919/ICJE.202507_11(7).0013

In the network architecture diagram, "3*3conv,64" denotes a convolutional layer with a kernel size of 3 and 64 filters, indicating that this layer performs convolution operations with 64 kernels of size 3×3. The remaining convolutional layers in the diagram carry the same meaning, while "FC" signifies a fully connected layer. Despite optimizations, no modifications were made to the network parameters. The VGG16R network model still accepts RGB images of size 224×224 as input. After passing through five convolutional groups, the model extracts image features with dimensions of 7×7×512. These features are then flattened into one-dimensional data and fed into the fully connected layer, ultimately utilizing the SoftMax function for classification.

4. Experiments

This study primarily focuses on modulation recognition of six types of OFDM signals with subcarrier modulation schemes including 2PSK, 8PSK, 4QAM, 16QAM, 64QAM, and 256QAM. All signals are generated using the MATLAB software platform, with an AWGN channel serving as the simulated environment. For each modulation signal, the signal-to-noise ratio (SNR) ranges from -10dB to 15dB in increments of 5dB. At each SNR level, 1000 signal samples are generated, with 700 allocated for training and the remaining 300 for testing. In total, 36,000 signal samples are collected, with 25,200 designated for training and 10,800 for testing. The relevant simulation parameters for the OFDM signals are presented in Table 2.

Table 2. Simulation Parameters for OFDM Signals.

Signal Parameters	Values
Sampling Frequency	100kHz
Symbol Rate	40kbit/s
Number of Symbols	12
Subcarrier Spacing	15kHz
Bits per Symbol	4
IFFT Size	512
Cyclic Prefix	128
Symbol Period	66.49μs

Based on the signal simulation parameters outlined in Table 2, the generated signals underwent SPWVD transformation, resulting in time-frequency representations for various signal types. The time-frequency images of six different types of modulated signals were produced under varying signal-to-noise ratios. The processed time-frequency images of the OFDM signals were then input into the VGG16 and VGG16R models, where the changes in the loss function curves were recorded, along with the accuracy metrics for both the training and validation datasets.

4.1 Experimental Environment

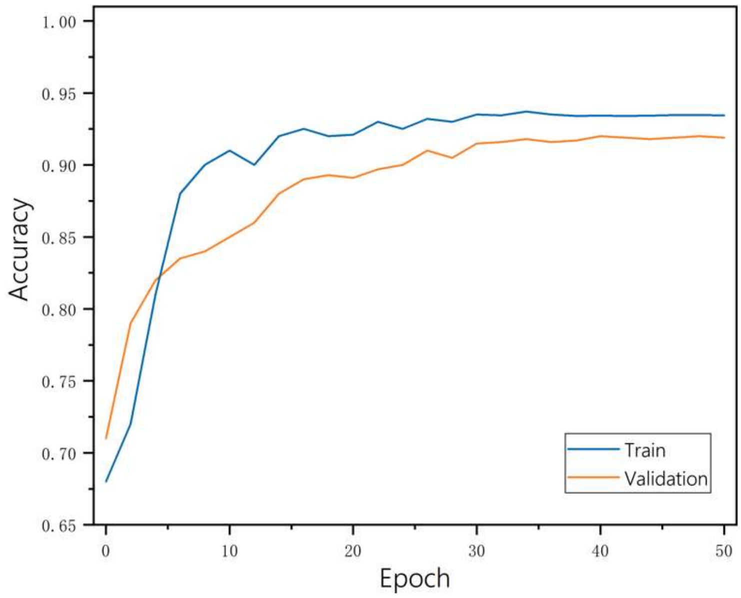
This experiment utilizes the deep learning framework TensorFlow 2.3, with the application programming interface provided by Keras 2.3, and the programming language employed is Python. The coding environment is PyCharm, and the experimental setup operates on the Windows 11 operating system. The hardware configuration includes a CPU: AMD R7 5800H, 8GB of video memory, GPU: RTX 3060, and 16GB of RAM.

4.2 Results and Analysis

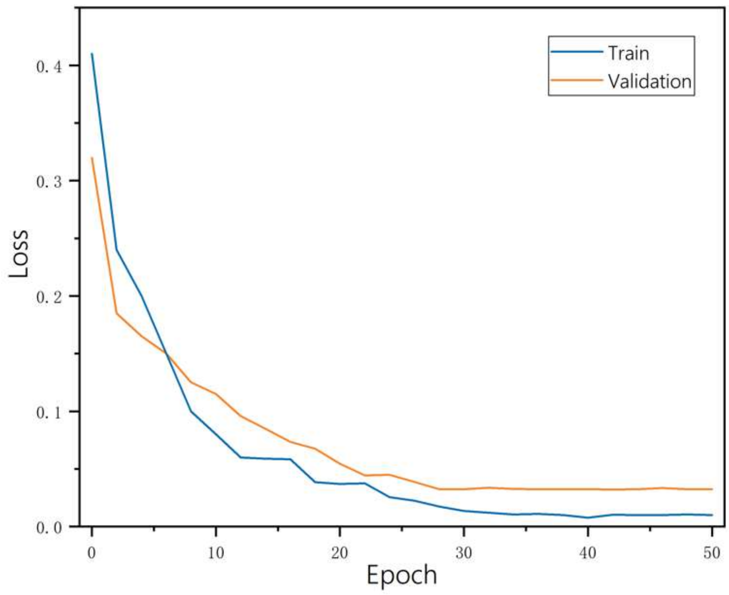
The classification performance of the original VGG16 network on the time-frequency representation dataset of OFDM signals is illustrated in Figure 4. In panel (a), the VGG16 model, after undergoing

ISSN: 2414-1895 DOI: 10.6919/ICJE.202507_11(7).0013

50 training epochs, shows that the accuracy curve of the validation set stabilizes after approximately 28 epochs, with reduced fluctuations, achieving an accuracy of 93.4% on the training set. The accuracy curve for the validation set indicates that stability is reached after 30 training epochs, with the maximum recognition accuracy of the model recorded at 91.9%. In panel (b), the training loss curve for the VGG16 model converges to a lower value, indicating effective training of the model. Analyzing the convergence speed, the model reaches convergence after 30 training epochs, with the loss value on the training set approaching 0.01 and the loss value on the validation set nearing 0.03.



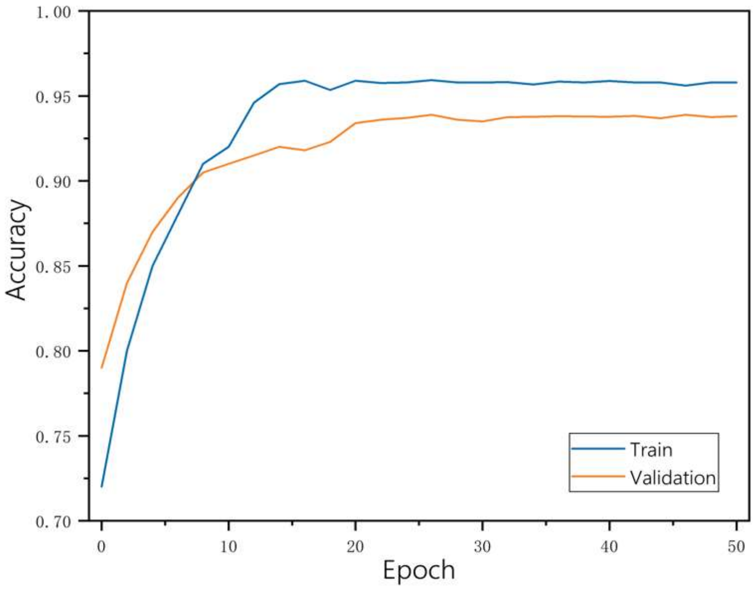
(a)Accuracy graph of the VGG16 network.



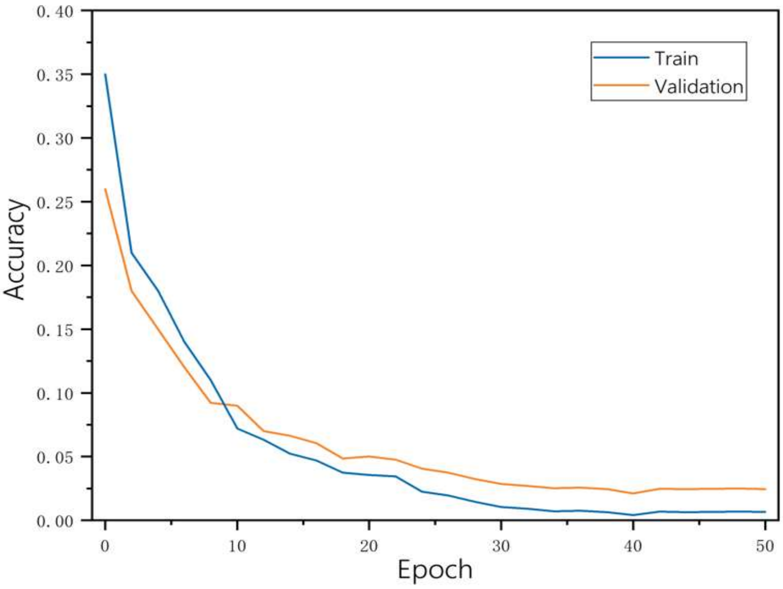
(b) Loss value graph of the VGG16 network.

Figure 4. Experimental Results of the VGG16 Network.

The experimental results of the improved VGG16R network model on the OFDM signal time-frequency image dataset are illustrated in Figure 5. In panel (a), the accuracy curve of the validation set stabilizes after approximately 18 training epochs, following 50 training cycles, with the training set accuracy of the VGG16R model approaching 95.8%. The accuracy curve for the validation set indicates that after 30 training epochs, the accuracy stabilizes, achieving a maximum recognition accuracy of 93.8%. In panel (b), the training set loss curve for the VGG16R model converges to a lower value, indicating effective training of the model. Analyzing the convergence speed, the model reaches convergence after 30 training cycles, with the training set loss value approaching 0.025 and the validation set loss value nearing 0.006.



(a)Accuracy plot of the VGG16R experiment.



(b)Loss plot of the VGG16R experiment.

Figure 5. Experimental results of the VGG16R network.

The data obtained from the experiments of the two models are presented in Table 3, which compares the experimental results of the two convolutional neural networks. It was observed that the improved VGG16R network achieved a recognition accuracy that is 1.9% higher than that of the modified VGG16 network, with a lower loss value compared to the VGG16 network. This improvement can primarily be attributed to the utilization of residual learning units, which enhanced the model's feature extraction capabilities. Additionally, the incorporation of batch normalization layers effectively mitigated the model's overfitting.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202507_11(7).0013

Table 3. Experimental Results of VGG16 and VGG16R Networks.

Model Name	Loss Value	Accurac
VGG16	0.03	91.9%
VGG16R	0.02	93.8%

It is evident that the recognition accuracy of signals is higher when utilizing the improved VGG16R network on the same OFDM signal dataset.

5. Conclusion

The VGG16R network introduces a novel architectural unit that treats consecutive convolutional layers as a convolutional group, comprising a total of five such groups. Within each convolutional group, a batch normalization (BN) layer is added following each convolutional layer. Furthermore, an additional branch is introduced at the input of each convolutional group, which directly transmits the input image to subsequent layers. The results from this branch are then combined with the outcomes of the convolutional group, resulting in a new set of feature collections. By employing deep learning techniques for signal recognition, the constructed OFDM signal dataset is utilized to train, test, and classify the CNN model. The model achieves a recognition accuracy of 93.8% on the developed dataset, representing a 1.9% improvement over the original VGG16 model. This demonstrates that the enhanced model provides greater accuracy in dataset recognition.

References

[1] Goldsmith G. Eireless Communication [M]. London:Cambridge Umiversity Press, 2005.

[2] Wu Tian Research on Digital Signal demodulation and recognition based on machine learning [D]. Beijing University of Posts and Telecommunications, 2019

[3] K. Karra, S. Kuzdeba and J. Petersen, "Modulation recognition using hierarchical deep neural networks[C]" 2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Piscataway, NJ, 2017, pp. 1-3.

[4] Li Bingzhen, Liu Ke, Gu Jiaojiao, etc A Review of Convolutional Neural Network Research [J]. Computer Age, 2021 (04): 8-12+17

[5] Xu Zhaohong, Liu Yu, Quan Jicheng, etc Remote sensing image building semantic segmentation based on VGG16 precoding [J]. Science, Technology and Engineering, 2019,19 (17): 250-255

[6] Yang Hanyu, Zhao Xiaoyong, Wang Lei Overview of Data Normalization Methods [J]. Computer Engineering and Applications, 2023, 59 (03): 13-22