Modulation Recognition of Communication Signals Based on Deep Learning Joint Model

Zongyu Li

1 School of Information and Communication Engineering, Harbin Engineering University, Haerbin, Heilongjiang, 150001, China

*Corresponding author’s e-mail: zongyuli@hrbeu.edu.cn

Abstract. In modern communication, it is often required for a non-cooperative party to identify the modulation mode when no prior knowledge is given to facilitate the subsequent demodulation and analysis. However, the traditional modulation recognition process requires cumbersome and uncertain manual signal-feature extraction, making it inapplicable to the complex communication environment. In order to overcome this limitation, this paper proposes a communication-signal modulation recognition model based on the dense connection network (DenseNet) and residual connection network (ResNet). The convolutional block attention mechanism (CBAM) is introduced into the DenseNet and ResNet structures, significantly enhancing the modulation recognition accuracy of the proposed global network model. Besides DenseNet and ResNet, the long short-time memory (LSTM) network is also adopted. The experimental results show the performance of the proposed model.

1. Introduction

These days, science and technology have been rapidly developed, and mobile communication has also been achieved many innovative advancements. Wider adoption of 4G and the gradual emergence of 5G communication technologies show the advance of communication technology towards the faster transmission, stronger signal, less latency, and more coverage ability [1]. The complex and diverse environment of modern communication highlights the importance of signal recognition. However, in the real world, affected by various complex natural and human factors, identifying the individual communication signals is challenging without any prior knowledge. This also brings great resistance to the modulation and demodulation of non-cooperative communication. As an essential part of the non-cooperative communication process, signal modulation identification can guarantee non-cooperative communication under complex conditions, which plays an important role in both civil and military communication [2].

2. Network structure design

2.1. Residual network (ResNet)

Theoretically, the accuracy of a deep network is enhanced as its depth increases. However, many studies reported deep neural network degradation when the number of layers is increased, which has been known as deep network degeneration. To address this problem, He et al. [3] introduced a residual connection layer, effectively solving the degradation phenomenon. Based on the concept of residual connection, the ResNet was proposed. In the residual connection layer, the input feature is directly augmented to the output in addition to the standard convolutional layer. In this way, the output feature...
of the residual layer $H(\chi)$ can be represented as $\Gamma(\chi) + \chi$. Then, the residual error in the network is $\Gamma(\chi) = H(\chi) - \chi$. When $\Gamma(\chi) = 0$, the residual layer acts as the identity function. In most cases, the network can learn the updated features since the residual error is not 0, achieving better results. For different shapes of input and output features, the input feature shape is modified through parameter transformation. The residual connection layer is depicted in figure 1 and mathematically formulated as:

$$\gamma = \Gamma(\chi, \{W_i\}) + W_s \chi$$

(1)

where $\chi$ and $\gamma$ represent the input and output features, respectively. $W_s$ indicates a matrix for parameter transformation of the input feature, and $\{W_i\}$ refers to the network transfer weight matrix.

![Figure 1. The residual connection layer](image)

2.2. Dense network (DenseNet)

Inherited the concept of the residual connection of the ResNet, DenseNet adopted a dense connection layer. Instead of adding the input and the output features, DenseNet concatenated the directly-connected input features and the output features to deepen the number of layers and broaden the structure [4]. DenseNet significantly reduced network parameters through reusing features, avoiding redundant features being trained. Further, the DenseNet converged faster, and the possible gradient disappearance of a neural network can be alleviated. The feature reuse is realized in the DenseNet by channel-wise adding the output features from previous layers. Let define the $i$th layer output is $\Phi_i$. Then, the transfer function of a standard convolutional layer in CNN is represented as $\Phi_i = H(\Phi_{i-1})$, the residual layer in ResNet is represented as $\Phi_i = H(\Phi_{i-1}) + \Phi_{i-1}$, and the dense block in DenseNet can be represented as $\Phi_i = H([\Phi_0, \Phi_1, \ldots, \Phi_{i-1}])$. The schematic diagram of a dense connection block is depicted in figure 2. In addition to the dense connection block, a bottleneck layer was also adopted, where the concatenation (concat) operator connects feature maps from different previous layers. Since several feature maps are concatenated, the output feature becomes larger as the layer is deeper. In order to reduce the feature dimension, a 1x1 convolutional kernel is used, as shown in figure 3. The convolutional layer is then followed by the dropout layer. The DenseNet takes advantage of the dense connection blocks to achieve higher recognition ability yet efficiency.
2.3. Long short-term memory (LSTM) network

LSTM has been widely used for temporal sequences in speech, text, and video processing applications [12]. As a variant of a recursive neural network (RNN), the LSTM adopted the gate function to control historical data memory, as shown in figure 4. The forgotten gate, input gate, and output gate are defined as equation (2), (3), and (4), respectively.

\[ f_t = \sigma(W_f [H_{t-1}, L_t] + b_f) \]  
\[ i_t = \sigma(W_i [H_{t-1}, L_t] + b_i) \]  
\[ o_t = \sigma(W_o [H_{t-1}, L_t] + b_o) \]

where \( \sigma \) represents a sigmoid function, and \( L_t \) indicates the input to the current neuron. The temporary state and the state of the current neuron are defined as equation (5) and (6), respectively.

\[ C_{t-1} = \tanh([H_{t-1}, L_t] + b_c) \]  
\[ C_t = \tanh([H_{t-1}, L_t] + b_c) \]

where \( C_{t-1} \) and \( C_t \) are the temporary state and the state of the current neuron. \( C_{t-1} \) and \( H_{t-1} \) represent the state and the output of the last neuron. Then, the neuron output gate is operated as:

\[ H_t = o_t \ast \tanh(C_t) \]

where \( H_t \) is the output of the current neuron.

2.4. Convolutional block attention module (CBAM)

The CBAM module can improve the attention of effective features in both channel and spatial dimensions, ignoring unnecessary sample features. In this way, the recognition efficiency of the deep
network can be improved. The CBAM module is composed of the channel and spatial attention modules. The channel attention module is followed by the spatial attention module (figure 5). The objective of each sub-attention module is to focus on important information by increasing weights on those areas while suppressing irrelevant information. The detailed structures of the channel and spatial attention modules are depicted in figure 6 and figure 7, respectively.

![Figure 5. The overall structure of the CBAM attention mechanism](image5)

![Figure 6. Schematic diagram of the channel attention module](image6)

![Figure 7. Schematic diagram of the spatial attention module](image7)

2.5. The proposed DRL network

In the proposed overall framework, the ResBlock is used to extract new features from the earlier layers, while the DenseBlock is used to reuse the features that have been extracted in the previous layers. Figure 8 shows the overall structure of the proposed DenseNet+ResNet+LSTM (DRL) model. As shown in figure 8, the two network structures are connected in series to extract different degrees of features, and the CBAM is used to weigh the important features. Then, the serial network output passes through the 32*2*1 convolutional layer and converts them into 128*32-dimensional data through the reshape operation. The reshaped features, where 128 indicates the sequence length and 32 indicates sequence dimension, are fed into the LSTM layer. Finally, two fully connected layers followed by the softmax layer provide the classification score.

![Figure 8. The overall structure of the proposed DRL network](image8)

Residence structural block after adding the CBAM module is shown in figure 9a. The CBAM attention module is added after each convolution layer of a dense block, where the dual attention mechanism can make the network more efficient in feature utilization. The schematic diagram of DenseBlock added to the CBAM is depicted in figure 9b. In Table 1, the hierarchical connection structure of the DenseNet+ResNet series network is described in detail.

![Figure 9a](image9a)

![Figure 9b](image9b)
3. Simulation experiments

The dataset used in the experiments was divided into training (80%) and test (20%) sets. For each epoch, the performance of the trained model is validated using the test sets that were randomly selected from the whole sample. The used batch size was set as 64 for both training and testing, indicating that 2,750 epochs for traversing the entire training set. Once the entire training set is traversed, the training and testing samples are shuffled and randomly determined.

Because of the small amount of training data, the model could be over-fitted to find the relationship between input and output generated by noise. Such overfitting reduces the generalization of the trained network so that it cannot generally be applicable to other test sets. In order to prevent network overfitting, a dropout operation was adopted after each convolutional layer to inactivate randomly selected neurons. The dropout rates were set as 0.2 for DenseBlock and ResBlock and 0.4 for the fully connected layers. In addition, L2 regularization was adopted to reduce the overfitting, formulated as:

$$\Omega = \Omega_0 + \frac{\lambda}{2n} \sum w^2$$  \hspace{1cm} (8)
where $\Omega_0$ represents the primary loss function used during the training process. $n$ and $w$ represent the sample size and the weighting parameter, respectively. $\lambda$ is the L2 regularization coefficient, which is set as $e^{-8}$ here.

The quantitative comparison between the CNN+LSTM and the proposed DRL network is given in Table 2.

| Model          | Overall recognition rate(%) | Recognition rate above 0dB(%) |
|----------------|-----------------------------|------------------------------|
| CNN+LSTM       | 55.6                        | 82.5                         |
| DRL (CBAM)     | 62.2                        | 91.6                         |

Table 2 shows that the proposed DRL network provides superior performance over the CNN+LSTM network in the overall signal recognition rate and the recognition rate in the high SNR range above 0dB. The modulation recognition rate curves of the compared two networks for the signals ranged from -20dB to 18dB SNR are shown in figure 10.

As illustrated in figure 10, the recognition accuracy of the DRL network with CBAM is higher than the CNN+LSTM network for all SNR values. It can be found from the recognition rate curve that, in the SNR above -16dB, the proposed DRL network shows an increasing advantage in modulation recognition rate compared with the CNN+LSTM network.

Figure 11 and figure 12 show the modulation confusion matrices of the proposed DRL and the CNN+LSTM models for 0dB and 18dB SNR cases.

Figure 11. The confusion matrix for the proposed DRL network.
As shown in figure 11 and figure 12, the proposed DRL network has a good modulation recognition effect on communication signals. Under the condition of 0dB and 18dB SNR, 8PSK, AM-DSB, AM-SSB, BPSK, CPFSK, GFSK, PAM4, QPSK, QAM16, and QAM64 signals achieved a high recognition rate. Compared with the CNN+LSTM network, the recognition confusion matrix of the new network for QAM16 and QAM64 signals is clear diagonally. There is almost no ambiguity phenomenon, and the recognition effect is greatly improved. Overall, the proposed DRL network provides a low recognition rate for WBFM signal under the condition of medium-high SNR. This is due to the mute phenomenon of analog language signals, when only the carrier tone exists, making these samples challenging to distinguish [5].

4. Conclusion
This paper proposes a new recognition network for communication signal modulation based on the ResNet, DenseNet, and the CBAM module. The ResNet and DenseNet are used to extract diverse spatial features of input signals, and LSTM is adopted to utilize sequence information of feature maps. Further, the introduction of the CBAM module into ResNet and DenseNet further improved the overall recognition accuracy. The experimental results show that the recognition accuracy of the proposed DRL model is considerably increased, especially for QAM16 and QAM64. The results also indicate that the proposed DRL model obtains promising performance over the CNN+LSTM network for all SNRs. Especially, the recognition accuracy at medium and high SNR is significantly better than the CNN+LSTM network.

Acknowledgments
First of all, I would like to express my gratitude to my tutor, Guo Limin, for his instructive advice and consistent encouragement. Secondly, I would give my hearty thanks to all the other faculty members of School of Information and Communication Engineering in Harbin Engineering University for their patient instructions and precious suggestions for my research. Finally, I am indebted to my parents who tried their best to help me in my life.
Reference

[1] Li R, Li L, Yang S, et al. (2018) Robust Automated VHF Modulation Recognition Based on Deep Convolutional Neural Networks. IEEE Communications Letters, 22(05): 946-949.

[2] LeCun Y, Bengio Y, Hinton G. (2015) Deep learning. Nature, 521(7553): 436-444.

[3] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway. pp. 770-778.

[4] Huang G, Liu S, Laurens V D M, et al. CondenseNet: an efficient DenseNet using learned group convolutions. In: Conference on Computer Vision and Pattern Recognition. SaltLake City. pp. 2752-2761.

[5] Hinton G E, Osindero S, Teh Y W. (2006) A fast learning algorithm for deep belief nets. Neural computation, 18(7): 1527-1554.

[6] Felix A, Gers, Nicol N, Schraudolph, Jurgen Schmidhuber. (2002) Learning precise timing with LSTM recurrent network. Journal of Machine Learning Research, 3(1): 115-143.

[7] Nandi A K, Azzouz E E. (1996) Modulation Recognition Using Artificial Neural Networks. Signal Processing, 56(2): 165-175.

[8] E.E.Azzouz and A.K.Nandi. (1995) Automatic identification of digital modulations. Signal Processing, 47(1): 55-69

[9] Hinton G E, Osindero S, Teh Y W. (2006) A fast learning algorithm for deep belief nets. Neural computation, 18(7): 1527-1554.

[10] Felix A, Gers, Nicol N, Schraudolph, Jurgen Schmidhuber. (2002) Learning precise timing with LSTM recurrent network. Journal of Machine Learning Research, 3(1): 115-143.