DenseNet-ResNet-LSTM model for modulation recognition of communication signal

ZongYu Li, YanDong Zhang
School of Information and Communication Engineering, Harbin Engineering University, Haerbin 150001, China
lizongyu@hrbeu.edu.cn, 936812254@qq.com

Abstract. The traditional artificial neural network provides a low recognition rate in the modulation recognition of communication signals, suffering from the difficulty of feature extraction. Also, it requires a high signal-to-noise ratio (SNR). In order to solve these problems, this paper proposes the combined model: DenseNet-ResNet-LSTM. In the proposed model, DenseNet and ResNet extract different spatial features of samples, and then LSTM extracts the sequence of samples. Also, the attention mechanism is employed to improve the learning efficiency and ability to learn important features. Experimental results show that the proposed model achieves higher accuracy and better generalization ability over the CNN-LSTM network.

1. Introduction
The modulation mode recognition of the communication signal is the judgment under the premise that no relevant information is obtained except the sampled signal. It provides basic conditions for subsequent demodulation, monitoring, interface, and other behaviors in signal detection, electronic confrontation, interference recognition, and other non-cooperative communication applications.

The traditional modulation recognition method used the statistics of the extracted features from the communication signals. For example, Nandi and Azouz used the second moments of different definitions of the instantaneous envelope, instantaneous phase, and instantaneous frequency to classify analog and digital signals. Swami [1] employed the normalized fourth-order cumulants to classify MQAM, MPSK, and MASK signals using the symbol synchronous sampling rate. Those methods were based on the statistical moment characteristics of the signal. Also, several methods were proposed based on characteristics in the transform domain of the signal. Like et al. [2] extracted the relevant features of the signal cycle spectrum for classification based on the cyclic stationarity of the digital signal in nature. Weanver et al. [3] also used the feature in the Fourier transform domain to identify short-wave signals such as USB, LSB, CW, FSK, and MFSK. Zhao et al. [4] realized the effective identification of the modulation mode of a communication signal by combining the high-order cumulant and cyclic spectral density. However, all those methods require manual parameter tuning and feature extraction, which leads to poor accuracy and higher sensitivity to the noise.

In order to overcome the limitations of traditional recognition methods, machine learning methods have been widely used. Since AlexNet [5] won the ImageNet Large Scale Visual Recognition Challenge in 2012, deep learning networks, especially a convolutional neural network (CNN), have emerged. Inspired by a great success in the field of image classification, semantic segmentation, and object detection [6], deep networks have been studied for the communication signal modulation recognition. O'Shea et al. [7] proved that the neural network is more effective in the recognition of...
time-domain communication signals by comparing the hand-crafted and deep learned feature extraction methods. SAK et al. [8] adopted the combination of CNN and long short-term memory networks (LSTM) for speech sequence signals. Recently, Gao et al. [9] proposed DenseNet, which significantly reduces the parameters of a deep convolutional network while improving the accuracy.

2. The structural design of the proposed network

2.1. Residual neural network (ResNet)

The performance of a deep neural network, in theory, is proportionally increased as the network gets deeper. However, many studies have shown the performance degradation with the increased depth of the network, so-called the degeneration of deep neural networks. He et al. [10] proposed the ResNet to effectively solve the degeneration phenomenon of a network model based on the cross-layer connection principle of the neural layers, and the input $x$ is mapped and added to the output feature. Then, the output feature $H(x)$ is expressed as $H(x) = F(x) + x$. When the residual error $F(x)$ is 0, the output is the identity mapping of the input. In most cases, $F(x) > 0$, which allows the network to learn updated features to achieve better results. If the characteristic shape of output and input is different, the input shape needs to be changed through parameter transformation. The formula of the residual element is as follow:

$$y = F(x, \{W_i\}) + W_s x$$  \hspace{1cm} (1)

where $y$ is the output feature of the unit, $x$ is the input of the unit, $W_s$ is the parameter to transform the input shape, and $\{W_i\}$ is the network transfer weight matrix. The structure diagram of the residual element is as follows:

![Figure 1. Schematic diagram of ResBlock.](image)

2.2. Densely connected network (DenseNet)

DenseNet[11] was proposed inheriting the idea of a cross-layer connected feature map of ResNet. Unlike the traditional convolutional neural networks, the number and width of layers are not constrained in DenseNet. DenseNet significantly reduces network parameters and avoids many redundant features through the feature-reuse. The convergence speed is also improved with alleviated gradient disappearance. In DenseNet, the output feature maps from different layers are concatenated.
for reusing the learned features. Let the output of the ith layer be $X_i$, then the transfer function in a neural network can be expressed as $X_i = H(X_{i-1})$. The dense connection is expressed as $X_i = H([X_0, X_1, \cdots, X_{i-1}])$.

A dense block is the most crucial element of DenseNet, which is described in Fig. 2. In the bottleneck layer, Concat-layer in dense connections concatenates feature maps from different previous layers in the channel dimension. As a result, the number of channels in the feature map becomes larger and larger. In order to avoid the problem, the $1 \times 1$ convolution layer is employed so that the dimensionality of the features is reduced. The bottleneck structure is shown in Fig. 3.

Figure 2. Schematic diagram of DenseBlock

Figure 3. Structure of bottleneck
2.3. Long short-term memory networks (LSTM)

LSTM is a recursive neural network for sequences processing with temporal relationships, which has played an essential role in speech and text information processing [12]. LSTM controls the memory of historical data through the gate function, and its network structure is shown in Fig. 4.

\[ i_t = \sigma(W_i[H_{t-1}, X_t] + b_i) \]  \hspace{1cm} (2)  
\[ o_t = \sigma(W_o[H_{t-1}, X_t] + b_o) \]  \hspace{1cm} (3)  
\[ f_t = \sigma(W_f[H_{t-1}, X_t] + b_f) \]  \hspace{1cm} (4)  

where \( \sigma \) is a sigmoid function. \( X_t \) is the input to the current neuron, and \( H_{t-1} \) is the output of the previous neuron. The current temporary state of neurons \( (C_t) \) is defined as:

\[ (C_t) = \tanh([H_{t-1}, X_t] + b_c) \]  \hspace{1cm} (5)  

The state of the current neuron \( C_t \) is defined as:

\[ C_t = f_t \cdot C_{t-1} + i_t \cdot (C_t) \]  \hspace{1cm} (6)  

where \( C_{t-1} \) is the state of the previous neuron. Then, the final output of the output gate \( H_t \) is as follows:

\[ H_t = o_t \cdot \tanh(C_t) \]  \hspace{1cm} (7)  

2.4. Attention mechanism

Attention mechanisms enable neural networks to focus the ability to input (or feature) subset, namely selecting the specific input. The attention-based model originated from image recognition and has been widely used in regression problems. Inspired by the human visual system, the attention mechanism in deep network assigns more weights to the significant parts of the input sequence, improving the judgment accuracy [13]. The structure of the attention mechanism is displayed as follows:
In Fig. 5, $x_i$ represents the input sequence, and $x_i^1$ is a similar scalar obtained by simplifying the calculation of $x_i$. It is mapped to the interval of $[0,1]$ by a normalized exponential function, that is, the weight. Dot-product attention is the sum of $x_i^1$ to the weight. The principle expression is as follows:

$$f(x_i, y) = (W_1 * x_i, W_2 * y)$$

$$\text{Attention} = \sum_{i=1}^{n} \text{softmax}(f(x_i, y)) * x_i$$

(8)

2.5 Overall network structure

In this paper, the DenseBlock and ResBlock extract the features in parallel to learn the characteristics of the training set. The ResBlock reuses the features extracted from the previous layers, while the DenseBlock extracts new features from the previous layers. Different characteristics of features are extracted with the two blocks, and then the extracted features are combined with a convolutional layer. The temporal features of the samples are further extracted through the LSTM-Attention structure. Lastly, it is sorted through a fully connected Softmax layer. The overall structure of the network is shown in Fig. 6. The detailed information for each layer is summarized in Table 1.
Figure 6. The overall network structure of DenseNet-ResNet-LSTM
Table 1. Detailed structure information for the proposed DenseNet-ResNet-LSTM

| layer                  | Output size | Ksize | Stride | Connected | layer          |
|------------------------|-------------|-------|--------|-----------|----------------|
| Reshape1               | 2*128*1     |       |        |           | input          |
| Conv BN1/drop1         | 2*128*128   | 1*3   | 1      | drop1     | Reshape1       |
| Conv BN2/drop2         | 2*128*128   | 2*3   | 1      | drop2     |                |
| Conv BN3/drop3         | 2*128*128   | 1*3   | 1      | drop3     |                |
| Conv BN4/drop4         | 2*128*128   | 1*3   | 1      | drop4     |                |
| ADD                    | 2*128*128   |       |        |           |                |
| BN5/Relu5              |             |       |        |           | ADD            |
| Bottleneck1/Conv1/drop5| 2*128*128   | 1*3   | 1      | reshape   |                |
| Concat1                | 2*128*129   |       |        |           | Input          |
| Bottleneck2/Conv2/drop6| 2*128*128   | 2*3   | 1      | Concat1   | Drop5          |
| Concat2                | 2*128*257   |       |        |           |                |
| Bottleneck3/Conv3/drop7| 2*128*256   | 1*3   | 1      | Concat2   |                |
| Concat3                | 2*128*513   |       |        |           | Concat2        |
| Concat4                | 2*128*641   |       |        |           | Concat3        |
| Conv0                  | 1*128*32    | 2*1   | 1      | Concat4   |                |
| Reshape2               | 128*32      |       |        |           | Conv0          |
| LSTM                   | 128         |       |        |           | Reshape2       |
| Attention              |             |       |        |           | LSTM           |
| Flatten                |             |       |        |           | Attention      |
| Softmax                | 11          |       |        |           | Flatten        |

Figure 7. The structure of CN
3. Experiments

3.1. Data sets and learning frameworks
The proposed method was implemented in Keras (framework) and TensorFlow (backend). The training and testing were conducted on NVIDIA GeForce GTX 1660Ti. The communication signal dataset is employed, which was presented by O’Shea at the 6th GNU Radio (Open Source Software Radio) Conference. The dataset includes sampled communication signals in 11 different modulation modes, including 8PSK, AM-DSB, AM-SSB, BPSK, CPFSK, GFSK, PAM4, QAM16, QAM64, QPSK, and WBFM. There are 220,000 signal samples, with a sampling SNR of -20dB to 18dB and an interval of 2dB. The number of samples for each signal under a specific SNR is 1000, and the number of sampling points for each sample is 128. This data set simulates practical factors such as central frequency offset, channel fading, and additive white Gaussian noise, making the generated signal very close to the real communication signal. The I/Q waveforms of different types of modulated signals under 18dB are shown in Fig. 8.

![Figure 8. The I/Q waveforms of signals from various data sets at 18DB noise](image)

3.2. Hyperparameters selection
The most widely used Adam-optimizer was used in this study. Compared with SGD, Adam can converge fast during the training and achieve high recognition accuracy under an appropriate learning rate. The initial learning rate of the network was set as 0.0001. When the verification accuracy cannot be improved in the short term, the learning rate is reduced to 0.1 times. The network converged when the training duration was about 20 epochs.

Due to the limited number of training datasets, the network may be overfitted to the data, decreasing the generalization ability of the network. The dropout layer was added after each convolutional layer to avoid the overfitting, which randomly inactivates some neurons during the training process. The drop rates of DenseBlock and ResBlock were 0.2, and the drop rate of the fully connected layer was 0.5. Also, L2 regularization is employed, which is defined as follows:

$$ C = C_0 + \frac{\lambda}{2n} \sum w^2 $$

where $C_0$ is the network primitive cost function, w is the weighting parameter, and n is the sample size. The L2 regularization coefficient $\lambda$ was set as $e^{-8}$ in this study.

3.3. Analysis and comparison of the results
In this experiment, 80% of the data set is used as the training data set, and the remaining 20% is used as the verification sample set. The batch size of the training set and the tests is set to 64, and it takes 2,750 times to traverse a training set. After each iteration of the training set, the sample sets were randomly shuffled to ensure a new order of sample is used in the next epoch. As shown in Figs. 9 –
11, the proposed network was converged faster and provided higher accuracy than the compared CNN+LSTM model [14]. As summarized in Table 3, the proposed model, DRL, outperforms the CNN+LSTM model in terms of recognition rates and convergence speed.

Figure 9. The accuracy of CNN+LSTM according to epochs

Figure 10. The loss of CNN+LSTM according to epochs

Figure 11. The accuracy of DRL according to epochs
Figure 12. The loss of DRL according to epochs

Table 2. The comparison of training performance between CNN+LSTM and ResNet-DenseNet-LSTM (DRL)

| Model                  | Overall recognition rate | Recognition rate above 0dB | Convergence epochs |
|------------------------|--------------------------|----------------------------|--------------------|
| CNN+LSTM               | 0.556                    | 0.825                      | 21                 |
| DenseNet+ResNet+LSTM   | 0.592                    | 0.901                      | 23                 |

(a) 8PSK

(b) AM-DSB
(c) AM-SSB

(d) BPSK

(e) CPFSK

(f) GFSK
(g) PAM4

(h) QAM16

(i) QAM64

(j) QPSK
Figure 13. The recognition accuracy of Dense-ResNet-LSTM and CNN+LSTM for various signals.

Fig. 13 shows that the proposed DRL significantly improves the recognition accuracy of QAM16, QAM64, WBFM, and other signals compared with CNN+LSTM. Moreover, for BPSK, CPFSK, GFSK, PAM4, QPSK, AM-SSB signals, the recognition rate is higher when the SNR is above 0dB.

Figure 14. The signal recognition confusion matrix of CNN+LSTM for SNR=0dB

Figure 15. The signal recognition confusion matrix of DRL for SNR=0dB

Figs. 14 and 15 show the signal recognition confusion matrix of CNN+LSTM and DRL for SNR=0dB. For BPSK, CPFSK, PAM4, GFSK, and QPSK signals, the recognition rate can reach more than 90% or even close to 100% under the condition of 0dB. The proposed DRL network shows a
slightly worse result for AM-DSB signals. However, the DRL network shows the outperforming overall recognition ability, especially for the QAM16, QAM64, WBFM signals, with strong generalization ability.

4. Conclusion
This paper proposes to combine ResNet, DenseNet, LSTM, and the attention mechanism for addressing the recognition of communication signals. ResNet and DenseNet extract different characteristics of spatial features in parallel structure. Furthermore, the LSTM-Attention structure is used further to extract the temporal information of the extracted features. The experimental results on the simulated communication signals show that the proposed network provides improved recognition accuracy, outperforming the compared CNN+LSTM model. Especially, the proposed DRL network significantly improved the recognition rate for QAM16, QAM64, and WBFM signals under various SNRs. Also, the overall recognition rate of the network above 0dB is much higher than that of the CNN+LSTM structure.

References
[1] Swami A, Sadler BM. Hierarchical digital modulation using cumulants. IEEE Trans. Commun.2000,48(3):416-429.
[2] Li E, Chakravarthy VD, Ratazi P, et al. Signal classification in fading channels using cyclic spectral analysis. EURASIP Journal on Wireless Communications and Networking, 2009:14.
[3] Weaver CS, Cole C A, Krumland RB, et al. The automatic classification of modulation types by pattern recognition. California: Stanford University,1969.
[4] Zhao XW, Guo CX, Li JC. A hybrid recognition algorithm for signal modulation based on high order cumulant and cyclic spectrum. Journal of Electronics Information Technology, 2016,38(3):674-680.
[5] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. Neural Information Processing Systems Conference, 2012:1097-1105.
[6] Qiang Z, Li Z, Li J, et al. Vehicle color recognition using multiple layer feature representations of lightweight convolutional neural network. Signal Processing, 2018,147:146-153.
[7] O’SHEA T J, CORGAN J, CLANCY T C. Convolutional Radio Modulation Recognition Networks[C]//International Conference on Engineering Applications of Neural Networks. Springer, Cham, 2016.
[8] SAK H, SENIOR A, BEUFAYS F, Long Short-term Memory Recurrent Neural Network Architectures for LargeScale Acoustic Modeling[C]//Fifteenth Annual Conference of The International Speech Communication Association, 2014.
[9] Gao H, Zhuang L, Laurens VDM, et al. Densely Connected Convolutional Networks[C]// CVPR. 2017.
[10] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016:770-778.
[11] Huang G, Liu S, Laurens V D M, et al. CondenseNet: an efficient DenseNet using learned group convolutions[C]//Conference on Computer Vision an Pattern Recognition. SaltLake City, 2018:2752-2761.
[12] Felix A. Gers, Nicol N. Schraudolph, Jurgen Schmidhuber. Learning percise timing with LSTM recurrent network. Journal of Machine Learning Research, 2002, 3(1):115-143.
[13] LIU Hao, ZENG Guangping, ZHANG Qingchuan, et al. Part-of-speech and position attention mechanism based BLSTM for question answering system[P]. Other Conferences, 2018.
[14] Wu YL. Research on modulation Recognition Algorithm based on deep learning [D]. Chengdu: University of Electronic Science and Technology, 2018.
[15] Nandi A K, Azzouz E E. Modulation Recognition Using Artificial Neural Networks[J]. Signal Processing,1996,56(2):165-175.
[16] Nandi A K, Azzouz E E. Automatic Modulation Recognition of Communication Signal[M]. Springer-Science+Business Media, B.V. 1996.

[17] O.A.Dobre, A.Abdi,Y and Bar-Ness et al. Survey of automatic modulation classification techniques: classical approaches and new trends[J]. IET Commun, 2007, 1(2).

[18] Bouchert M L, Mortensen I, Favaro H, etc. A New Algorithm for Nonlinear Estimation of PSK-Modulated Carrier Phase[C]. Satellite Communications-Eesc-3, 1993, 3rd European Conference, on1993.

[19] Nandi A K, Azouz EE. Algorithms for Automatic Modulation Recognition of Communication Signals. IEEE Trans.Com, 1998.

[20] Zhang LM. The model and application of artificial neural network [M]. Shanghai Fudan University Press, 1994.

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