Modulation scheme recognition using convolutional neural network

Qianwen Zhang¹, Zhan Xu¹ ², Peiyue Zhang¹

¹School of Information and Communication Engineering, Beijing Information Science & Technology University, Beijing 100192, People’s Republic of China
²E-mail: xuzhan@bistu.edu.cn

Abstract: Convolutional neural network (CNN) is an extremely powerful machine-learning tool, especially when dealing with computer vision problems. Here, the authors present a CNN-based modulation recognition model. In order to fully elaborate the powerful image feature extraction ability of CNN, the authors have created an image dataset of different complex signal spectrograms using short-time Fourier transform (STFT). In this case, the complex modulation recognition problem is converted to an image recognition problem. To study the accuracy of automatic recognition of signal spectrograms, the authors have applied two approaches recently developed for image classification. The first approach is to optimise activation functions. Experiments show that best performance can be achieved when using sigmoid as activation function. The second approach is using optimisation functions. At last, the authors compared the recognition accuracy under different signal-to-noise ratios (SNRs). The result shows that authors’ model achieves higher recognition accuracy under low SNR and stronger generalisation ability than other recognition methods.

1 Introduction

In the research field of non-cooperative communications, modulation recognition technology has always been an important branch. The modulation recognition refers to automatic identification of modulated signals with noise. It also can be used for the applications such as demodulation and feature extraction, which play an important role in military intelligence, electronic warfare, electronic reconnaissance, and other areas [1]. As of that, lots of studies have been accomplished and various algorithms have been introduced by using those techniques. Automatic recognition of analogue and digital modulation can use the artificial neural network (ANN) for classification [2]. In previous studies, higher accuracy in lower SNRs is one of the top priorities for researchers.

In recent years, deep learning algorithms have made a breakthrough in image classification. Convolutional neural network (CNN) has been recently identified as a powerful tool in image classification and voice signal processing [3]. Previous works show that CNN excels in feature extraction. Some researchers in the wireless communications field have started to apply deep neural networks to cognitive radio tasks with some success [4–6]. In this work, we propose a new modulation recognition method that refers to the idea of image recognition, that is, using a great number of spectrogram images of different modulation signals obtained by short-time Fourier transform (STFT) as the input of image recognition. Therefore, the difficult modulation recognition problem is converted to an image recognition problem which can be solved by CNN. As of that, complex feature extraction and data reconstruction in traditional recognition algorithms can be avoided.

The remainder of this paper is organised as follows: In Section 2, four different modulations will be introduced. Our CNN model will be presented in detail in Section 3. In Section 4, different activation functions and optimisation functions will be described. In Section 5, simulation results are provided and analysed, followed by a concluding summary in Section 6.

2 Signal model

2.1 Modulation scheme

Both analogue and digital modulations can be described as the following general expression:

\[ s(t) = a(t) \cos(\omega_c t + \phi(t)) \]  

where \( a(t) \) is the instantaneous amplitude, and \( \phi(t) \) is the generalised instantaneous phase, \( \omega_c \) is the carrier angle frequency.

The differences of \( a(t) \) and \( \phi(t) \) represent different modulations. The four modulations expressions for this experiment are shown in Table 1.

\( A_c = 1 \) represents DC bias quantity, \( m(t) = \cos \omega_m t \) is a modulated signal, and \( \omega_m \) is the modulation angle frequency, \( K_f \) is the FM index. For simplicity, we assume \( K_f = 1 \), both \( s(x) \) and \( s(y) \) are the unipolar pulse, and when \( s(x) \) is a positive pulse, \( s(x) \) is at zero level and vice versa. Here, the baseband signals of FSK and PSK are randomly generated with a sequence length of 10 symbols. \( \omega_o = 2 \pi f_o, \omega_r = 2 \pi f_r \) where \( f_o \) and \( f_r \) are the carrier frequencies. \( \phi_0 \) represents the absolute phase.

2.2 Dataset generation and pre-processing

\( x(n) \) represents a real sequence and \( X_o(mS, o) \) represents the output of STFT. The variable \( S \) is a positive integer, which represents the sampling period of \( X_o(n, o) \) in the variable \( n \). The analysis window used in STFT can be denoted as \( w(n) \). With little loss of generality, we assume that \( w(n) \) is real, \( L \) is the length of samples, and non-zero for \( 0 \leq n \leq L - 1 \). From the definition of the STFT, we have

\[ X_o(mS, o) = F \sum_{n=-\infty}^{\infty} x_o(n, mS, l) e^{-j \omega_o l} \]  

where

\[ x_o(mS, o) = \sum_{n=-\infty}^{\infty} x_o(n, mS, l) e^{-j \omega_o l} \]
Heisenberg uncertainty principle. The length of the window fluctuating, so a smoother spectrum can be obtained.

Classification task [11]. Categorical cross-entropy is a measure of categorical cross-entropy was selected for a modulation recognition task of modulation recognition. The network of our CNN model is 3 × 3. After convolutional layer, 2 × 2 max pooling is used for reducing complexity. The pooling process improves the generalisability of a model by merging multiple semantically similar features into one [10]. Local response normalisation (LRN) is used in norm layers, which shows a good performance in solving overfit problem. The objective function categorical cross-entropy was selected for a modulation recognition classification task [11]. Categorical cross-entropy is a measure of the difference between two probability distributions. For deep learning classification task, the probability distribution is usually a softmax of the classifier network which is then converted to one-hot encoding for classification purposes [12]. In our work, by using softmax, the category with the highest confidence value can be given as the output category.

3 Convolution neural network

CNN has a special structure for feature extraction [8]. The input of CNN is a 2-D image, and the classification result is given by a probability. CNN is not a fully connected neural network. Instead, there are some key technologies of CNN such as local connection and weight sharing which make the image recognition more compact and invariant. The idea of CNN is closer to the actual biological neural network. Weight sharing reduces the complexity of the network, especially when facing multidimensional input. Our model is based on AlexNet [9], which is a classic and powerful model based on CNN. Our design is modified and improved based on the task of modulation recognition. The network of our CNN model is shown as following: conv1--maxpool--norm1--conv2--norm2--max pool--local3--local4--softmax, in which local3 layer and local 4 layer are fully connected layers. The model structure is shown in Fig. 1.

The input is a 64 × 64 size normalised image. The size of kernels of our model is 3 × 3. After convolutional layer, 2 × 2 max pooling is used for reducing complexity. The pooling process improves the generalisability of a model by merging multiple semantically similar features into one [10]. Local response normalisation (LRN) is used in norm layers, which shows a good performance in solving overfit problem. The objective function categorical cross-entropy was selected for a modulation recognition classification task [11]. Categorical cross-entropy is a measure of the difference between two probability distributions. For deep learning classification task, the probability distribution is usually a softmax of the classifier network which is then converted to one-hot encoding for classification purposes [12]. In our work, by using softmax, the category with the highest confidence value can be given as the output category.

4 Model parameters

4.1 Activation functions

In CNN, the convolution operation is linear, and it includes only multiplication and addition. Each layer of the neural network can only do the linear transformation. However, the linear model does not have good performance, so it is necessary to add an activation function to introduce non-linear factors.

In the previous studies of neural networks, the effect of activation functions on the performance of neural networks has received considerable attention [13, 14]. Studies have shown that different activation functions lead to different performance. The activation functions usually have the following properties: non-linearity, differentiability, monotonicity, and the output value range can be limited or infinite. We have investigated different activation functions for performance comparison, which is shown in Table 2.

First of all, ReLU function [15] is linear, and the derivative of ReLU is always 1. As of that, the problem of gradient vanishing can be avoided, also, ReLU function can improve the computing speed. AlexNet's paper mentioned that for the same network structure, using ReLU as an activation function, the convergence speed is more than six times faster than using Tanh [13]. Apart from ReLU function, the advantage of Sigmoid function [16] is easy to derive, so the gradient descent algorithm can be implemented at a very low cost. Also, there is Tanh function [17] which is a variant of the Sigmoid function. The difference between Tanh function and Sigmoid function is that Tanh function is the average of 0. Therefore, in practical applications, Tanh function will be better than Sigmoid function. At last, the Softplus activation function [18] is often used for multi-category neural networks.

4.2 Optimisation functions

The optimisation of objective function is one of the core issues of machine learning and has always been the focus of attention in the academic area. Here, the most commonly used gradient descent algorithm [19], Adam [20], and RMSProp [21] algorithm are investigated to compare the recognition performance. These algorithms not only improve the global search ability and convergence speed but also reduce the influence of hyper-parameters (such as learning rate) on operating efficiency, simplifying the training process.

Gradient descent is simple, general and can be flexibly applied to any differential objective functions. RMSProp optimisation function accelerates the rate of gradient descent. As to Adam optimisation algorithm, it is an extension of the stochastic gradient descent algorithm which can iteratively update neural network weights based on training data. Recently, it is widely used in deep learning applications, especially tasks such as computer vision and

\[ F_i[x_i(mS,l)] = w(mS - l)x(l) \]  

\[ x_{l}(mS,l) = w(mS - l)x(l) \]  

Table 2 Activation functions

| Activation function | Formula |
|---------------------|---------|
| rectified linear units (ReLU) | \( f(x) = \max(0, x) \) |
| sigmoid | \( f(x) = \frac{1}{1 + e^{-x}} \) |
| hyperbolic tangent (Tanh) | \( f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \) |
| Softplus | \( f(x) = \log(1 + \exp(x)) \) |

![Convolutional neural network model](http://creativecommons.org/licenses/by/3.0/)
natural language processing. The common feature of RMSProp and Adam is that it can automatically adjust the learning rate.

5 Results

The result of our research will be presented in this section.

Our model was developed using TensorFlow, a well-known framework of neural network. Due to its strong modularity and friendly interface with Python, it greatly accelerates our work in building CNN model.

The recognition accuracy under different activation functions is shown in Fig. 2. Simulation results show that accuracy is close to 100% when SNR $\geq 12$ dB. The results show that using the spectrograms as the dataset trained on the CNN can achieve a great recognition accuracy. When SNR is lower than 4 dB, a better performance can be achieved by using Sigmoid function as activation. At SNR $= -10$ dB, the recognition performance using Sigmoid function is nearly 10% higher than that of using ReLU function. Accuracy and loss value changes during training process are shown in Figs. 3 and 4, respectively.

The total number of iterations is 80,000. In the first 20,000 iterations, the accuracy rate rapidly increases, and after that, it shows a slight upward trend. The recognition accuracy of signals under the Sigmoid function is bigger than 90% when the number of iterations exceeds 20,000.

During the training process, the loss also gradually converges with the increase of the number of iterations.

The simulation shows that the recognition accuracy of our model using different optimisation functions is over 80% when the SNR reaches $-6$ dB. When SNR $\geq 2$ dB, the recognition accuracy is close to 100%. The accuracy under three different optimisation functions is nearly equivalent, as shown in Fig. 5, which implies that all those methods can extract enough features reliably.

Fig. 2 Modulation recognition under 4 activation functions

Fig. 3 Change of accuracy during training

Fig. 4 Change of loss values during training

Fig. 5 Different recognition rates under different optimisation functions
8 References

[1] Rodomagoulakis, I., Maragos, P.: ‘On the improvement of modulation features using multi-microphone energy tracking for robust distant speech recognition’, 2017 25th European Signal Processing Conf. (EUSIPCO), Kos, Greece, 2017, pp. 558–562

[2] Liu, X., Yang, D., Guo, A.E.: ‘Deep neural network architectures for modulation classification’, arXiv preprint arXiv:1712.00443, 2017

[3] Nandi, A.K., Azizouz, E.E.: ‘Algorithms for automatic modulation recognition of communication signals’, IEEE Trans. Commun., 1998, 46, (4), pp. 431–436

[4] O’Shea, T.J., Corgan, J., Clancy, T.C.: ‘Convolutional radio modulation recognition networks’, Int. Conf. on Engineering Applications of Neural Networks, Aberdeen, UK, 2016, pp. 213–226

[5] O’Shea, T.J., West, N.: ‘Radio machine learning dataset generation with GNU radio’. Proc. of the GNU Radio Conf., Boulder, CO, USA, 2016, 1, (1)

[6] Migliori, B., Zeller-Townson, R., Grady, D., et al.: ‘Biologically inspired radio signal feature extraction with sparse denoising autoencoders’, arXiv preprint arXiv:1605.05239, 2016

[7] Nikunen, J., Virtanen, T.: ‘Source separation and reconstruction of spatial audio using spectrogram factorization’, in Pulkkki, V., Delikaris-Manias, S., Politi, A. (Eds.): ‘Parametric time-frequency domain spatial audio’ (Wiley-IEEE Press, Hoboken, NJ, USA, 2017), p. 215

[8] LeCun, Y., Huang, F.J., Bottou, L.: ‘Learning methods for generic object recognition with invariance to pose and lighting’. Proc. of the 2004 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2004, CVPR 2004, Washington, DC, USA, 2004, 2, p. II-104

[9] Cao, Y., Xu, R., Chen, T.: ‘Combining convolutional neural network and support vector machine for sentiment classification’. Chinese National Conf. on Social Media Processing, Guangzhou, China, 2015, pp. 144–155

[10] Bourau, Y.L., Ponce, J., LeCun, Y.: ‘A theoretical analysis of feature pooling in visual recognition’. Proc. of the 27th Int. Conf. on Machine Learning (ICML-10), Haifa, Israel, 2010, pp. 111–118

[11] West, N., O’Shea, T.: ‘Deep architectures for modulation recognition.’ 2017 IEEE Int. Symp. on Dynamic Spectrum Access Networks (DySPAN), Baltimore, MA, USA, 2017, pp. 1–6

[12] Heaton, J., Goodfellow, I., Bengio, Y., et al.: ‘Deep learning’, 2018, 19, (1–2), pp. 305–307

[13] Krizhevsky, A., Sutskever, I., Hinton, G.E.: ‘Imagenet classification with deep convolutional neural networks’, Advances in neural information processing systems, Lake Tahoe, NV, USA, December 2012, pp. 1097–1105

[14] Dahl, G.E., Sainath, T., Hinton, G.E.: ‘Improving LCRUS using rectified linear units and dropout’. 2013 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, Canada, 2013, pp. 8609–8613

[15] Li, Y., Yuan, Y.: ‘Convergence analysis of two-layer neural networks with ReLU activation’, Advances in neural information processing systems, Barcelona, Spain, 2017, pp. 597–607

[16] Sitti, P., Jones, S.A., Sudartha, P.: ‘Analysis of different activation functions using back propagation neural networks’. J. Theor. Appl. Inf. Technol., 2013, 47, (3), pp. 1264–1268

[17] Shi, W., Caballero, J., Huszár, F., et al.: ‘Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network’. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 1874–1883

[18] Huang, L., Wang, L.: ‘Accelerated Monte Carlo simulations with restricted Boltzmann machines’, Phys. Rev. B, 2017, 95, (3), p. 035105

[19] Carayanniss, N.B.: ‘Reformulated radial basis neural networks trained by gradient descent’. IEEE Trans. Neural Netw., 1999, 10, (3), pp. 657–671

[20] Kingma, D.P., Ba, J.: ‘Adam: a method for stochastic optimization’, arXiv preprint arXiv:1412.6980, 2014

[21] Mukkamala, M.C., Hein, M.: ‘Variants of RMSProp and adagrad with logarithmic regret bounds’, arXiv preprint arXiv:1706.05507, 2017, pp. 2545–2553

Fig. 6 Recognition rate of different methods

With the increase of SNR, the recognition accuracy improves as is shown in Fig. 6. The recognition accuracy of the red line (ReLU + Adam) and blue line (Sigmoid + Adam) proposed here outperforms green line (CNN) [3] and mauve line (CNN2) [11].

6 Conclusion

Here, we present a new method of modulation recognition based on CNN. We examined both individual and combined impacts on recognition accuracy with different activation functions and optimisation functions under different SNRs. The simulation shows that the proposed method can not only achieve a good performance but also has strong generalisation ability. More importantly, our model can also identify the modulations beside those four modulations mentioned above. At last, we believe, with all the potentials of our model, better accuracy can be achieved in the future.

7 Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (No. 61620106001), by Beijing Excellent Talent Support Program (No. 2016000026833ZK08), by Beijing Nova Program (No. Z161100004916086), by Construction PLAN of High Level Teachers in Beijing Municipal Universities (No. CIT&TCD201704065) and by Discipline construction project of Beijing Information Science and Technology University (no. 5121911006).