Automatic Modulation Recognition: A Few-Shot Learning Method Based on the Capsule Network

Lixin Li, Junsheng Huang, Qianqian Cheng, Hongying Meng, Zhu Han

Abstract—**Recent convolutional neural network (CNN) has been proved to be applied to automatic modulation recognition (AMR) with high classification accuracy. However, in the case of few samples, due to the lack of characteristic information of modulation signals, CNN does not achieve the desired accuracy requirements. In this paper, inspired by the capsule network (CapsNet), we propose a new network structure named AMR-CapsNet to achieve higher classification accuracy of modulation signals with few samples, and analyze the influence of digital capsule (DigitCaps) dimension on classification accuracy. The simulation results demonstrate that when 3% of the samples are used to train and the signal-to-noise ratio (SNR) is greater than 2dB, the overall classification accuracy of the AMR-CapsNet is greater than 80%. Compared with CNN, the classification accuracy is improved by 20%.**

Index Terms—**Convolutional neural network (CNN), Automatic modulation recognition (AMR), Capsule network (CapsNet), Few-Shot learning, Deep learning.**

I. INTRODUCTION

Automatic modulation recognition (AMR) can accurately determine the modulation type of signals when the modulation information is unknown [1]. Classicial AMR methods are mostly based on likelihood [2] or feature based methods [3]–[5]. Likelihood based methods mostly calculate the optimal solution of the likelihood ratio or average likelihood ratio. Amuru et al. estimated the complex channel state and different noise distribution by the Gibbs algorithm, and then selected the appropriate classification algorithm to distinguish the signals [2]. However, the computational complexity of the likelihood based methods is high, and simplified calculation often leads to the lack of classification information, which causes the degradation of classification performance.

In order to overcome the shortcomings of the likelihood based methods, the feature-based methods are mostly combined with machine learning (ML) algorithms to achieve higher classification accuracy. Under multiple input and multiple output (MIMO) channels, Kharbech et al. compared the computational complexity and accuracy of artificial neural network (ANN), support vector machine (SVM), and k-nearest neighbours in blind signal classification [3].

Recently, deep learning (DL) has been applied to AMR [4], [5], which reduces human participation in the process and reduce the complexity of operations. As an important forward neural network in DL, the convolutional neural network (CNN) reduces the complexity of the model through weight sharing and sparse connection, and automatically extracts feature information from the input signal, which achieves the classification of modulation signals. O'Shea et al. proposed the use of CNN to distinguish among different modulation signals [4].

However, in order to obtain high classification accuracy, DL requires a large number of training samples. In most cases, we are unable to obtain a sufficient number of samples, resulting in poor classification performance of DL neural networks in the case of few samples. Therefore, the concept of few-shot learning is proposed [6]. A new network structure-capsule network (CapsNet) is proposed to distinguish different handwritten digits, and obtained better results than CNN on the MNIST dataset [7].

In this letter, inspired by the CapsNet, we design the CapsNet-based network structure to study the influence on the classification accuracy of modulation signals. Specifically, the characteristic information of the signals are extracted through the CapsNet, and classified different modulation signals under a fewer samples. The contributions of the letter are summarized as follows:

- The new network structure named AMR-CapsNet is proposed to improve the classification accuracy of the modulation signals with few samples.
- In order to reduce the training parameters of the proposed network, we explore the influence of the dimension of the capsules in digital capsule (DigitCaps) on the classification accuracy.
- Simulation results show that compared with other network structures, when the training samples account for 3% of the total samples and the SNR is greater than 2dB, the overall classification accuracy of AMR-CapsNet is higher than 80%.

The rest of the letter is organized as follows. Section II introduces the CapsNet-based network structure applied to AMR. The simulation results are discussed in Section III. Finally, the conclusion is drawn in Section IV.

II. AMR-CAPSNET:THE CAPSNET-BASED NETWORK STRUCTURE OF AMR

In this section, inspired by the CapsNet, we propose a CapsNet-based network structure AMR-CapsNet to improve

L. Li, J. Huang, Q. Cheng are with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, 710129, China (e-mails: lilixin@nwpu.edu.cn, {1047863620, chengqianqian}@mail.nwpu.edu.cn.

H. Meng is with Department of Electronic and Computer Engineering, Brunel University London, Uxbridge, U.K..

Z. Han is with the Electrical and Computer Engineering Department, University of Houston, Houston, TX 77005 USA (e-mails: zhan2@uh.edu).

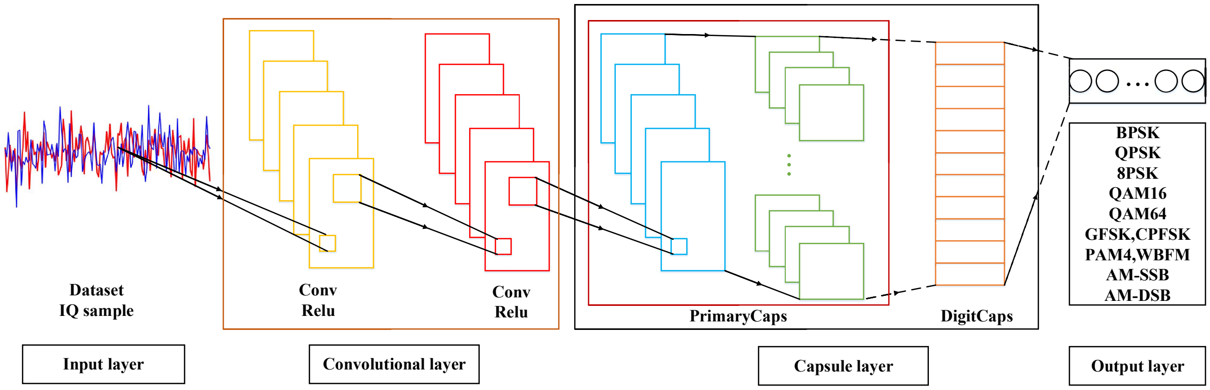


Fig. 1. The structure of AMR-CapsNet.

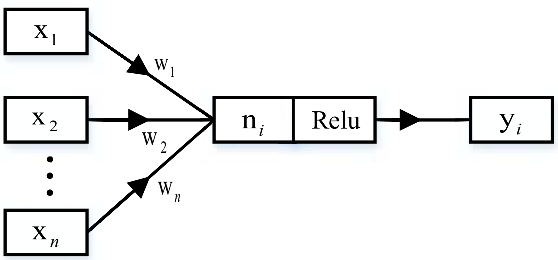


Fig. 2. The flow chart of a single convolutional layer.

TABLE I
THE CORRESPONDING DIMENSION OF EACH LAYER IN THE AMR-CAPSNET

layer	Output dimension
Input	$1 \times 2 \times 128$
Conv(filers 64,size(2,9))+Relu	$64 \times 1 \times 60$
Conv(filers 64,size(1,5))+Relu	$64 \times 1 \times 28$
PrimaryCaps	$16 \times 4 \times 1 \times 28$
DigitCaps	16×11
Classification	11

the classification accuracy of modulation signals in the case of few samples.

As shown in Fig. 1, the overall structure of AMR-CapsNet consists of four parts, including the input layer, convolutional layer, capsule layer, and output layer. The corresponding dimensions of each layer are shown in Table I. Each part is described in detail below.

1) Input layer: The modulation signals are used as the input of the AMR-CapsNet for training the network model. The input layer simply passes the input to the convolution layer, which is a one-way transmission.

2) Convolutional layer: As shown in Fig. 2, the convolutional layer consists of two parts, one is convolution operation (Conv), and the input of each neuron is connected to the local receiving domain of the previous layer to extract the local feature information. $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_n$ are the input of the convolutional layer, and $\mathbf{w}_1$, $\mathbf{w}_2$ and $\mathbf{w}_n$ represent the corresponding weight. $\mathbf{n}_i$ represents the output of the corresponding convolution operation. The other part is non-linear mapping. An non-linear mapping from low-level features to high-level features is achieved by selecting a suitable activation function. In order to effectively avoid gradient explosion and gradient disappearance, rectified linear unit (Relu) is used as the activation function. $\mathbf{y}_i$ represents the output of the nonlinear mapping.

$$\mathbf{n}_i = \sum_{j=1}^{n} \mathbf{x}_j \cdot \mathbf{w}_j, \tag{1}$$

$$\mathbf{y}_i = \mathrm{Relu}(\mathbf{n}_i). \tag{2}$$

In computer vision applications, 3×3 or 5×5 convolution kernels are mostly used to extract feature information. In this paper, however, Due to the limited size of input signals, 64 convolution kernels of size 2×9 and 64 convolution kernels of size 1×5 are respectively used to extract the feature information. In the convolution operation, the padding is set to ”valid” and the strides is set to 2. Finally, the output tensor is 64×1×28.

3) Capsule layer: The capsule layer consists of two parts, one is primary capsule (PrimaryCaps), and the other is Digitcaps. The two parts are described in detail below.

a) PrimaryCaps: Next, 64 convolution kernels of size 1×6 are used to convolve the input vector. In the convolution operation, the padding is set to ”same” and the step size is set to 1. The dimension of the final output vector is 64 ×1×28. The concept of vector is introduced here, i.e. 4 scalars of 1×1 form 1 vector of 1×4. Operate the output results: divide 4 channels into 1 group and divide into 16 groups. The final output can be regarded as 16 groups, each group has 1×28 vectors.

b) DigitCaps: In the CapsNet, output $\boldsymbol{a}_i$ of the ith capsule in the PrimaryCaps and input $\boldsymbol{s}_j$ of the jth capsule in DigitCaps are vectors. The DigitCaps contains 11 capsules, in which each capsule is represented by a 1×16 vector. Each capsule corresponds to a modulation type. Inside the capsule, each input vector maps a 4-dimensional input space to a 16-dimensional output space through the weight matrix $\mathbf{Z}$ of 4×16, and transform $\boldsymbol{a}_i$ into prediction vector $\boldsymbol{b}_i$ by transformation matrix. Different from the fully connected neural network, the CapsNet adds coupling coefficient $\mathbf{C}$ in the process of linear summation. According to the coupling coefficient, input $\boldsymbol{s}_j$ is obtained. Finally, the squashing activation function is used to get output $\boldsymbol{v}_j$, i.e.,

$$\boldsymbol{b}_i = \mathbf{Z} \cdot \boldsymbol{a}_i, \tag{3}$$

$$\boldsymbol{s}_j = \sum_{i=1} \mathbf{C} \cdot \boldsymbol{b}_i, \tag{4}$$

$$\boldsymbol{v}_j = \frac{\|\boldsymbol{s}_j\|^2}{1+\|\boldsymbol{s}_j\|^2} \cdot \frac{\boldsymbol{s}_j}{\|\boldsymbol{s}_j\|}, \tag{5}$$

where $\boldsymbol{v}_j$ is the output vector of capsule the *j*th in the DigitCaps. This activation function both preserves the direction of the input vector and compresses the modulus of the input vector between 0 and 1.

4) Output layer: According to the output $\boldsymbol{v}_j$ of the DigitCaps, the modulation type of modulation signals is classified.

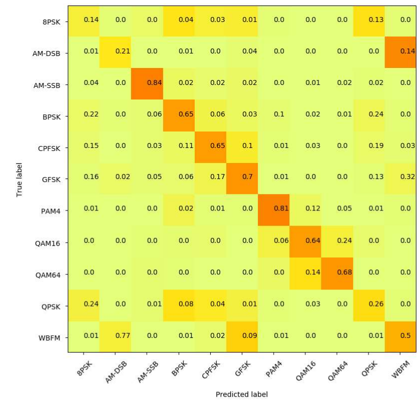
III. Simulations and Discussions

In this section, the experimental simulation is divided into three parts. Firstly, the feasibility and effectiveness of the AMR-CapsNet is discussed. Secondly, the influence of dimension of DigitCaps on modulation signals classification is discussed. Finally, the classification performance of the different network structures with few samples is analyzed.

The RML2016.04C is used as the dataset in our experiments to train and evaluate the proposed network structure [7]. There are 162,060 samples in the dataset, including 11 type of modulated signals. Among them, 8 kinds of diaital signals are composed of BPSK, QPSK, 8 phase shift keying (8PSK), QAM16, QAM64, gauss frequency shift keying (GFSK), continuous phase frequency shift keying (CPFSK), pulse amplitude modulation 4 (PAM4), and 3 kinds of analog signals are composed of wide band frequency modulation (WBFM), amplitude modulation-single side band (AM-SSB), amplitude modulation-double side band (AM-DSB). The SNR ranges ranges from -6dB to 12dB with an interval of 2dB.

A. The classification accuracy of AMR-CapsNet

In this subsection, 5% dataset is used to train the AMR-CapsNet, and the remaining dataset is used to verify the feasibility of the model. The confusion matrix is used to analyze the classification results of modulation signals under different SNRs. As can be seen from Fig. 3, when the SNR is at the low degree, such as -4 dB, due to the influence of noise, AMR-CapsNet cannot accurately distinguish the effective characteristics of the modulated signals, so there is a deviation in determining the modulation type of signal. With the increase of SNR, although the classification accuracy is improved, it cannot distinguish PSK accurately. PSK carrier phase modulation by the digital signals obtained. In QPSK, the carrier phase is used to represent the four different input digital signals, respectively, carrier phase $45°$, $135°$, $225°$, $315°$.



(a) SNR = -4 dB

(b) SNR = 12 dB

Fig. 3. When 5% of the dataset is used for training, the classification accuracy of 11 modulation signals by AMR-CapsNet.

TABLE II
Training parameters corresponding to different dimension of capsules in the DigitCaps

Dimensional change	Parameters
4	212,800
8	393,024
16	753,472
32	1,474,368

8PSK has 8 different phase differences: $0°$, $45°$, $90°$, $135°$, $180°$, $225°$, $270°$, $315°$. 8PSK and QPSK phase difference overlap, resulting in AMR-CapsNet not accurately distinguish PSK.

B. Impact of capsule dimension on classifaication accuracy in DigitCaps

In this subsection, the capsule dimension in the DigitCaps is set to 4, 8, 16, and 32 for experiments. As shown in Table II, when the dimension of DigitCaps is set to 4, the network training parameter is 212,800. As the capsule dimension increases, the network training parameters also gradually increase. When the dimension of DigitCaps is set to 32, the network training

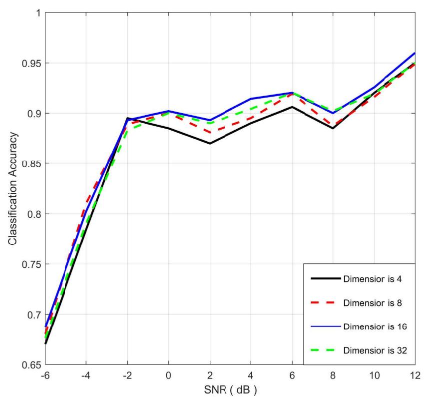
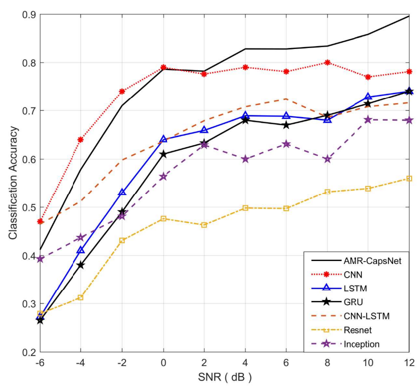
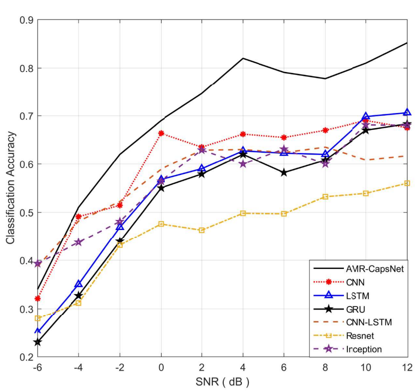


Fig. 4. The influence of DigitCaps dimension on classification accuracy.



(a) The training samples account for 5% of the total samples



(b) The training samples account for 3% of the total samples

Fig. 5. Influence of the number of training samples on classification accuracy under different network structures.

parameter increases from 212,800 to 1,474,368, almost seven times, which leads to the increase of network training time.

As shown in Fig. 4, when the SNR is between -2dB and 8dB, as the dimension increases from 4 to 16, the network training parameters increase, and the classification accuracy increases gradually. When the dimension is set to 32, the classification accuracy does not improve. On the whole, the change of the dimension of the DigitCaps has no significant effect on the signal classification performance.

C. Classification accuracy of different network structures with few samples

In this subsection, in addition to AMR-CapsNet, we further explore the classification accuracy of other networks, namely CNN, long short-term memory (LSTM), CNN-LSTM, gated recurrent unit (GRU), residual network (ResNet) and Inception network in different number of training samples.

As shown in Fig. 5, On the whole, when 5% of the samples are used for training, for the modulation signals, using AMR-CapsNet and CNN can achieve high classification accuracy. When the SNR is greater than 0dB, the classification accuracy of CNN fluctuates around 80%, while the classification accuracy of AMR-CapsNet continues to increase as the SNR increases. When 3% of the samples are used for training, as the SNR increases, the classification accuracy of most network structures for modulated signals fluctuates between 60% and 70%. In contrast , the classification accuracy of AMR-CapsNet fluctuates around 80%. When the SNR is 12dB, the classification accuracy is close to 85%.

IV. Conclusion

In this letter, inspired by the CapsNet, we design the CapsNet-based architecture for AMR and verify it through the dataset RML2016.04C. Then, the effect of capsule dimension in DigitCaps on classification accuracy is discussed. Simulation results show that when the training samples account for 5% of the total samples and the SNR is greater than 0dB, the overall classification accuracy of the AMR-CapsNet is higher than 80%. When 3% of the samples are used for training, and SNR is 12dB, the classification accuracy is close to 85%, and the accuracy is improved by 20% compared with CNN, which verifies the effectiveness of the AMR-CapsNet in AMR.

References

[1] L. Xie and Q. Wan, "Cyclic Feature-Based Modulation Recognition Using Compressive Sensing," in IEEE Wireless Communications Letters, vol. 6, no. 3, pp. 402-405, Jun. 2017.

[2] S. Amuru and C. R. C. M. da Silva, "A Blind Preprocessor for Modulation Classification Applications in Frequency-Selective Non-Gaussian Channels," in *IEEE Transactions on Communications*, vol. 63, no. 1, pp. 156-169, Jan. 2015.

[3] S. Kharbech, I. Dayoub, M. Zwingelstein-Colin and E. P. Simon, "On classifiers for blind feature-based automatic modulation classification over multiple-input-multiple-output channels," in *IET Communications*, vol. 10, no. 7, pp. 790-795, May 2016.

[4] T. J. O'Shea, J. Corgan, and T. Charles Clancy, "Convolutional Radio Modulation Recognition Networks," 2016, *arXiv:1602.04105*. [Online]. Available:https://arxiv.org/abs/1602.04105

[5] Y. Zeng, M. Zhang, F. Han, Y. Gong and J. Zhang, "Spectrum Analysis and Convolutional Neural Network for Automatic Modulation Recognition," in *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 929-932, Jun. 2019.

[6] S. Rahman, S. Khan and F. Porikli, "A Unified Approach for Conventional Zero-Shot, Generalized Zero-Shot, and Few-Shot Learning," in *IEEE Transactions on Image Processing*, vol. 27, no. 11, pp. 5652-5667, Nov. 2018.

[7] S. Sabour and N. Frosst and G. E. Hinton, "Dynamic Routing Between Capsules," 2017, *arXiv:1710.09829*. [Online]. Available:https://arxiv.org/abs/1710.09829