Few-Shot Learning of Signal Modulation Recognition based on Attention Relation Network

1st Zilin Zhang
Beijing Institute of Technology
Beijing, China
zilinzhang@bit.edu.cn

2nd Yan Li
Beijing Institute of Technology Beijing, China
Peng Cheng Laboratory Shenzhen, China
liyan726@bit.edu.cn

3rd Meiguo Gao
Beijing Institute of Technology
Beijing, China
meiguo_g@bit.edu.cn

***Abstract*—Most of existing signal modulation recognition methods attempt to establish a machine learning mechanism by training with a large number of annotated samples, which is hardly applied to the real-world electronic reconnaissance scenario where only a few samples can be intercepted in advance. Few-Shot Learning (FSL) aims to learn from training classes with a lot of samples and transform the knowledge to support classes with only a few samples, thus realizing model generalization. In this paper, a novel FSL framework called Attention Relation Network (ARN) is proposed, which introduces channel and spatial attention respectively to learn a more effective feature representation of support samples. The experimental results show that the proposed method can achieve excellent performance for fine-grained signal modulation recognition even with only one support sample and is robust to low signal-to-noise-ratio conditions.**

***Index Terms*—Signal Modulation Recognition, Few-Shot Learning, Attention**

I. Introduction

Electronic reconnaissance aims to detect, identify and analyze the electromagnetic signals of radiation sources, and obtain information from them. An important part of electronic reconnaissance is to identify the intra-pulse modulation of the intercepted signals [1]. Traditional recognition methods are mainly based on decision theory. However, with the development of modern radar system, these methods appear inadequate because of low recognition accuracy as well as their dependence on prior information. Therefore, it is essential to develop a more intelligent and faster algorithm for jamming systems to detect and identify the intra-pulse modulation [2].

In recent years, deep neural networks (DNNs) have made a breakthrough in various machine learning tasks such as computer vision, natural language processing and speaker recognition. DNNs have also been successfully applied to the field of cognitive radio [3]–[7]. In [4], a DNN model based on a multi-constraint Boltzmann machine is designed for radar signal recognition. In [5], a diagram of a radar waveform classification system is presented which employs a supervised classifier based on statistical features of time-frequency distributions and cyclostationary spectral analysis. [6] proposes a convolutional neural network (CNN) based on lenet-5 to recognize the time-frequency images of radar signal. [2] and [7] transforms radar signals into time-frequency domain using Choi-Williams distribution (CWD) for feature extraction and image denoising. Then a CNN and an Elman neural network (ENN) are used to identify the modulation respectively. It is reported that this work can achieve reasonable accuracy at low signal-to-noise ratio (SNR) conditions. [3] discusses the influence of network depth on radio modulation recognition based on convolutional long short-term deep neural networks. Despite their powerful feature representation abilities, DNNs usually need to train with a large amount of labelled data and optimize their parameters in many iterations, which limits their applicability to rare categories where numerous annotated samples are hard to obtain. In contrast, human beings are very good at recognizing objects even with very few instances, i.e. Few-Shot Learning (FSL). For example, children have no problem generalizing the concept of "zebra" from a single picture in a book. Inspired by the Few-Shot cognition ability of human beings, there has been a recent resurgence of interest in FSL [8].

FSL aims to recognize novel visual categories from a small number of labelled examples. FSL methods often decompose training into auxiliary meta-learning stages, that is, learning the knowledge that can be transferred between training set and test set, such as initial condition, embedding space or learning rate. The target task of FSL is then conducted by fine-tuning with the learned optimization strategy or computing in a feed-forward pass without updating network weights. In each meta-learning task, FSL aims to train a classification model to identify categories in a query set, using only a small number of samples in a support set. Therefore, a kind of transfer learning mechanism is established through the knowledge learned between different meta-learning tasks. Model-agnostic meta-learning (MAML) approach [9] aims to find more transferable representations with sensitive initialization parameters. Meta-SGD [10] simultaneously initializes the model and updates its learning rate in a single meta learning process. Metric learning searches a high-dimensional embedding space to describe the similarity between samples. Siamese neural network [11] is a two-branch CNN architecture that computes similarity between an image pair. [12] provides a strong regularizer, containing an additional generator network and a deep residual network, to describe the similarity. Matching network [13] produces a weighted

This work was supported by the National Natural Science Foundation of China under Grant 61901032.

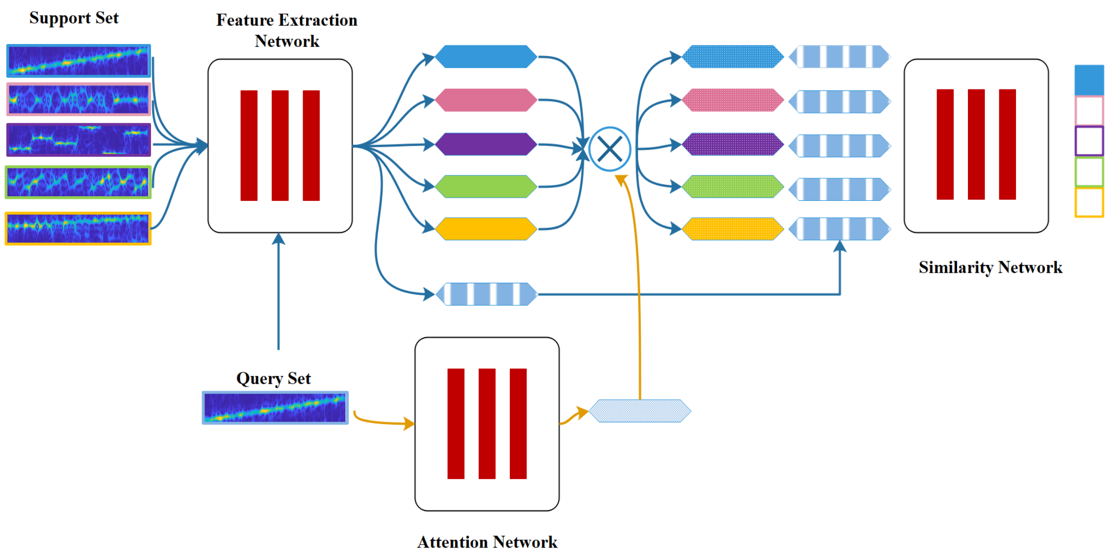


Fig. 1. The framework of ARN for a 5-way-1-shot task.

nearest neighbor classifier given the support set and adjusts feature embedding according to the performance on the query set. Prototypical network [14] computes class prototypes in the embedding space and find the nearest one for the embedded query points. Relation Network [8] learns a deep distance metric to compare a small number of images within episodes. [15] constructs an embedding local covariance representation to extract the second-order statistic information and defines a new deep covariance metric to measure the consistency based on Relation Network. However, these methods ignore the influence of query set samples during feature extraction of the support set. Metric-Agnostic Conditional (MACO) [16] network utilizes query samples to extract refined features for classification. In [17], an attention similarity network is constructed by obtaining the cross products of the extracted features to realize few-shot sound recognition. Deep Nearest Neighbor Neural Network (DN4) [18] also argues that a local descriptor based image-to-class measure, which shows that focusing on some stronger discriminating parts is efficient. In addition, previous methods of signal modulation identification mainly focus on distinguishing between different modulation types. However, different parameter settings of the same modulation type should also be considered as different classes. In this paper, two attention mechanisms are combined with the relation network to construct the FSL framework, and the features of the support set samples are generated by the query samples to extract more distinguishing features and fulfill the fine-grained classification of signal modulation.

The main contributions of this paper are as follows: 1) an Attention Relation Network (ARN) is proposed to identify the modulation patterns with only small samples; 2) a fine-grained classification is performed for LFM, Barker, Costas, Frank and T modulation types as well as their different parameter settings; 3) the experimental results show that the ARN can achieve high accuracies even in low SNR conditions (-10dB).

II. Methodology

A. Problem Definition

In this paper, for the above five modulation types (LFM, Baker, Costas, Frank, and T), different parameter combinations, such as frequency modulation slope, center frequency, code length and step length, are also considered to be different categories. CWD is utilized to transform the generated signals into time-frequency images. In the scenario of FSL, three datasets including a training set, a support set, and a query set are applied. The support set and query set share the same label space, which is disjoint from the training set. For a support set containing P unique classes each with K labelled instances, the target problem is called P-way-K-shot FSL. The training set is used to mimic the few-shot learning setting through task-based training to extract the transferable knowledge, so that the model can perform better on the support and query sets. As proposed in [8], [13], for each episode, the support set $S = (x_i, y_i), i = 1, 2, \ldots, P \times K$ is constructed by randomly selecting P classes from the training set with K labelled samples, and a fraction of the remain samples in the P classes form the query set $Q = (x_j, y_j), j = 1, 2 \ldots, N$. This split is designed to simulate the one which will be encountered at test procedure.

B. Architecture

As shown in Fig.1, an Attention Relation Network is proposed based on the Relation Network [8], where the feature extraction network consists of four convolution blocks. Each block contains a 64 channel 3×3 convolution, a batch normalization and a ReLU nonlinearity layer respectively. The input image x is fed into the feature extraction network and the feature maps $f_\varphi(x)$ are obtained, where φ is the weight of the feature extraction network. The query sample x_j generates the attention weight $a_\theta(x_j)$ through the attention network

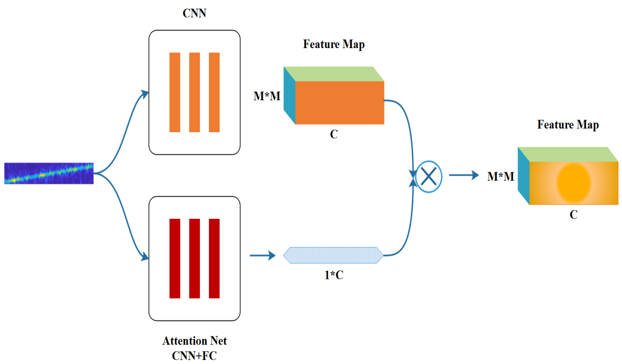


Fig. 2. The framework of channel ARN.

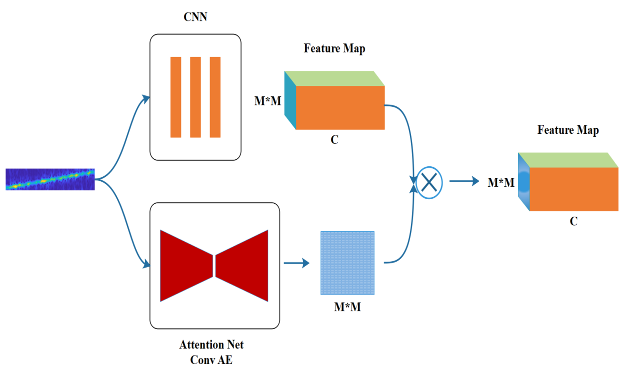


Fig. 3. The framework of spatial ARN.

alone, where θ is the weight of the attention network. Then the attention feature maps of support sample x_i under the condition of the query sample x_j is defined as:

$$g(x_i, x_j) = a_\theta(x_j) \times f_\varphi(x_i) \tag{1}$$

By concatenating the feature maps of support and query samples to generate support-query pairs $C(g(x_i, x_j), f\varphi(x_j))$, the similarity network learns the similarity metric using two convolution blocks and two fully connected layers. Relation score $r(x_i, x_j)$ of the final output defines the relation between x_i and x_j:

$$r_{i,j} = r(x_i, x_j) = h_\rho(C(g(x_i, x_j), f\varphi(x_j)), f\varphi(x_j)) \tag{2}$$

where $h_\rho(\bullet)$ represents the similarity network with weights ρ.

The mean square error (MSE) loss is used to train our model, quantizing the relation score rij into the binary similarity strategy, where the matched sample pairs have a similarity of 1 otherwise 0:

$$arg\ min_{\varphi,\theta,\rho} \sum_{i=1}^{P\times K} \sum_{j=1}^{N} (r_{i,j} - \delta(y_i - y_j)) \tag{3}$$

$$\delta(a) = \begin{cases} 1, a = 0 \\ 0, else \end{cases} \tag{4}$$

C. Attention Network

The attention mechanism originates from the study of human vision. In cognitive science, due to the bottleneck of information processing, human beings will selectively pay attention to only part of all the available visible information while ignoring the rest. For example, when people read, usually only a small number of words will be concerned and processed at a time. In neural networks, attention is usually served as an additional module to guide the network focus on some parts of inputs by assigning different weights. [19] and [20] introduce two different attention mechanisms: channel attention and spatial attention.

Channel attention operates at the channel level. After the original input is processed by the neural network, the feature maps of multiple channels are formed. As shown in Fig. 2, benefitting from the flexible channel attention, each channel is encoded with the calculated weights and then combined linearly. Suppose that the scale of the feature map obtained by the feature extraction network is $C \times M \times M$, where C is the channel number and M is the size of feature map. The channel attention network consists of three convolution layers and a full connection layer to generate a $1 \times C$ weight.

Different from channel attention, spatial attention takes the feature map as a whole and weighting features of all channels using the same set of weights, as Fig.3 shows. The attention network is composed of a convolutional autoencoder, which generates the $M \times M$ weights by convolution and deconvolution.

III. Experiment

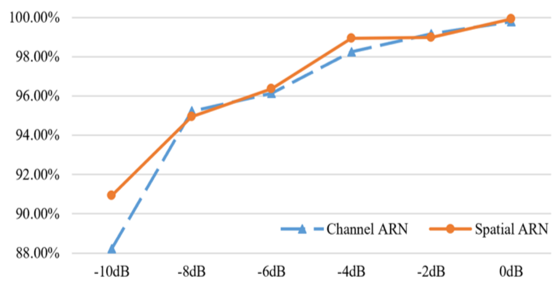
A. Setup

The proposed method is evaluated on simulated radar signal datasets. Five types of modulation are chosen, namely LFM, Barker, Costas, Frank and T codes. Different parameter combinations are used to form 50 classes of signals, including 32 classes for training, 8 classes for validation and 10 classes for testing. The sampling rate of the signal is 80 MHz and the pulse width is $10\mu s$. For each signal, time-frequency images are formed by CWD as samples, which are resized to 84×84. 50 samples are simulated for every 2 dB in each class with SNR ranging from -10dB to 0dB. The 50 samples of each category in train set are all used for training. The noise of each signal is introduced through additive Gaussian white noise.

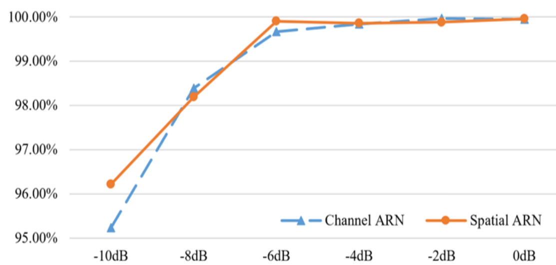
A three-layer convolution network and a fully connected network form the channel attention network. A six layer network with convolution and deconvolution constitutes the spatial attention network. All networks are trained using the Adam optimizer with a learning rate of 0.1. The number of iterations of training is 5000 and the model is trained end-to-end from scratch, with random initialization. Following the standard setting adopted by most existing few-shot learning works, this paper conducted 5-way-1-shot and 5-way-5-shot classification. In other words, the number of samples in support set is 1 and 5 respectively. The 5 classes are chosen randomly from the test set. All the results presented are the average of ten test procedures.

TABLE I
ACCURACIES OF DIFFERENT METHODS ON TWO KINDS OF FSL TASKS.

	5-way-1-shot	5-way-5-shot
MAML [9]	95.39%	99.64%
Prototypical Network [14]	95.75%	99.71%
Relation Network [8]	96.24%	99.70%
Channel ARN	96.13%	99.67%
Spatial ARN	**96.36%**	**99.90%**



a) accuracies of 5-way-1-shot FSL



b) accuracies of 5-way-5-shot FSL

Fig. 4. Accuracies of two attention mechanisms on different SNRs.

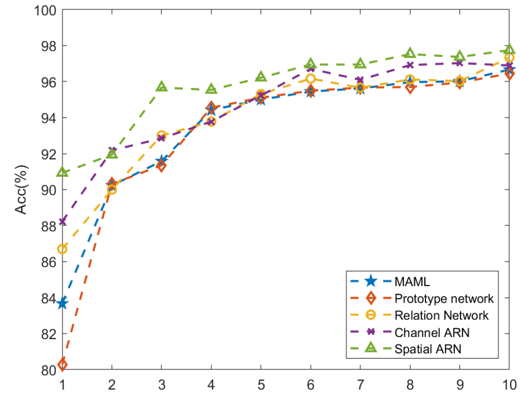
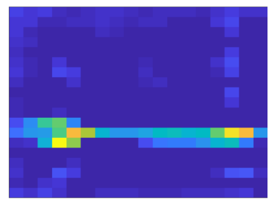
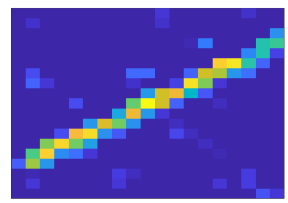


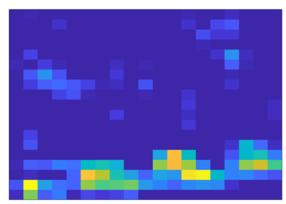
Fig. 5. Accuracies of different methods on sample numbers.



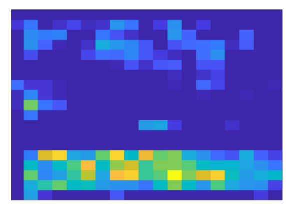
a) Support sample I



b) Support sample II



c) Query sample I



d) Spatial Attention

Fig. 6. The feature maps of different parts. a) and b) are the feature maps of two different classes in the support set, respectively; c) represents the feature map of the query sample; d) is the corresponding spatial attention weight.

B. Results

Table 1 shows the FSL recognition results of 5-way-1-shot and 5-way-5-shot at -6dB SNR. It can be seen that both channel and spatial ARN achieves high accuracies for the FSL task. In particular, spatial attention enhances performances of the traditional Relation Network, because it puts more emphasis on the regional characteristics of the input images while ignoring the background noises. On the other hand, channel attention may introduce small negative effects on the relation network. The reason is that the time-frequency images are relatively simple. For example, there is only one straight line in the images of LFM, so the extracted features of each channel are too similar to avoid overfitting.

Considering the influence of SNR on the clarity of time-frequency maps, Fig.4 shows the results of 5-way-1-shot and 5-way-5-shot under different SNR conditions. As expected, the accuracy is obviously improved with the increase of SNR. In particular, even if the SNR is as low as -10dB, the proposed spatial ARN still achieves an accuracy over 95%, indicating that the attention network is able to guide the model concentrate on the critical area of input images thus eliminating the interference of noise.

Fig.5 shows the results of five methods of different support sample numbers for SNR = -10dB. It can be seen that with the increase of the number of samples, the accuracies of all methods are increasing, and the spatial ARN always achieves the best performances while the channel version of ARN always takes second place. When the number of samples is more than 5, the increase of accuracy tends to be flatten out. Noting that when the number of samples is very small (only one), the performance improvement of attention mechanism is more obvious, which shows that attention mechanism can enhance the capability of the learning model in few-shot conditions.

To further illustrate the effect of the proposed attention mechanism, Fig.6 shows the output of support samples and query samples in the feature extraction and attention network respectively. It can be seen that spatial attention (Fig.6 d)) extends the characteristics of the query set (Fig.6 c)), allowing a better match of the corresponding support samples (Fig.6 a)), and distinction from other types of support samples (Fig.6 b)).

IV. Conclusion

In this paper, a novel Attention Relation Network is proposed to identify signal modulations with only a few samples. The proposed model utilizes channel or spatial attention to adjust the extracted features of support samples according to the query samples. The experimental results on simulated radar signal data set validate the robustness of the proposed model for different SNR levels. In future works, the authors plan to combine the two attention mechanisms into a unified model in order to make further improvements.

References

[1] M. Gupta, G. Hareesh, and A. K. Mahla, "Electronic warfare:issues and challenges for emitter classification," *Defence Science Journal*, vol. 61, no. 3, pp. 228–234, 2011.

[2] Z. Qu, X. Mao, and Z. Deng, "Radar signal intra-pulse modulation recognition based on convolutional neural network," *IEEE Access*, vol. 6, pp. 43 874–43 884, 2018.

[3] T. J. O. Nathan E West, "Deep architectures for modulation recognition," *IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, 2017.

[4] D. Zhou, X. Wang, Y. Tian, and R. Wang, "A novel radar signal recognition method based on a deep restricted boltzmann machine," *Engineering Review*, vol. 37, no. 2, pp. 165–171, 2017.

[5] J. Lunden and V. Koivunen, "Automatic radar waveform recognition," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 1, pp. 124–136, 2007.

[6] M. Zhang, M. Diao, and L. Guo, "Convolutional neural networks for automatic cognitive radio waveform recognition," *IEEE Access*, 2017.

[7] Z. Ming, L. Lutao, and D. Ming, "Lpi radar waveform recognition based on time-frequency distribution," *Sensors*, vol. 16, no. 10, p. 1682, 2016.

[8] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[9] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," *International Conference on Machine Learning*, p. 1126–1135, 2017.

[10] Z. Li, F. Zhou, F. Chen, and H. Li, "Meta-sgd: Learning to learn quickly for few-shot learning," *arXiv:1707.09835*, 2017.

[11] G. R. Koch, "Siamese neural networks for one-shot image recognition," in *International Conference on Machine Learning*, 2015.

[12] A. Mehrotra and A. Dukkipati, "Generative adversarial residual pairwise networks for one shot learning," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[13] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," *Advances in Neural Information Processing Systems*, p. 3630–3638, 2016.

[14] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," *Advances in Neural Information Processing Systems*, p. 4080–4090, 2017.

[15] W. li, J. Xu, J. Huo, L. Wang, Y. Gao, and J. Luo, "Distribution consistency based covariance metric networks for few-shot learning," in *The AAAI Conference on Artificial Intelligence*, 2019.

[16] N. Hilliard, L. Phillips, S. Howland, A. Yankov, C. D. Corley, and N. O. Hodas, "Few-shot learning with metric-agnostic conditional embeddings," *arXiv:1802.04376*, 2018.

[17] S.-Y. Chou, K.-H. Cheng, J.-S. R. Jang, and Y.-H. Yang, "Learning to match transient sound events using attentional similarity for few-shot sound recognition," *International Conference on Acoustics, Speech and Signal Processing*, 2019.

[18] W. Li, L. Wang, and J. Xu, "Revisiting local descriptor based image-to-class measure for few-shot learning," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[19] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[20] D. M. T. Feng Wang, "Survey on the attention based rnn model and its applications in computer vision," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.