Maximum Focal Inter-Class Angular Loss with Norm Constraint for Automatic Modulation Classification

[1]Sicheng Zhang, [1]Jiangzhi Fu, [1]Zherui Zhang, [2]Shui Yu, [3]Shiwen Mao, [1*]Yun Lin
[1]College of Information and Communication Engineering, Harbin Engineering University, Harbin, P. R. China
[2]School of Computer Science, University of Technology Sydney, Sydney, Australia
[3]Department of Electrical & Computer Engineering, Auburn University, Auburn, AL, USA
[*]Corresponding Author: linyun@hrbeu.edu.cn

GLOBECOM 2022 - 2022 IEEE Global Communications Conference | 978-1-6654-3540-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/GLOBECOM48099.2022.10001441

***Abstract*—Artificial intelligence (AI) has emerged as the most promising solution expected to overcome the high degree of abstraction of radio signals and achieve accurate automatic modulation classification (AMC). To further improve the classification performance of the AMC model and enhance its interpretability, the network output layer is modeled as a decision space into which the input data is projected. In this paper, we expand the inter-class angle between the classes with the largest confusion rate to increase the decision space. In addition, we extend the perspective to the softmax layer and evaluate the negative impact of the output distribution range on the confidence difference in the AMC problem. We further propose constraining the norm of the input data to the output layer in combination with prior knowledge of the distribution of modulation signal data. Combining the above two aspects, a Maximum Focal Inter-Class Angular Loss with Norm Constraint (MFICAL-NC) scheme is proposed. The experimental results show that the method can guide the model to obtain a better fitting state and a stronger generalization ability.**

***Index Terms*—Automatic modulation classification, maximum confusion class, inter-class angular, confidence difference, norm constraint.**

I. Introduction

Automatic modulation classification (AMC) is a critical component for applications such as spectrum monitoring, cognitive radio networks, and electromagnetic management, to enable efficient, reliable, and secure applications in future mobile and wireless communication networks [1]. Due to the openness of the electromagnetic space and the dynamic nature of channel propagation effects, the form of received electromagnetic signals is extremely abstract. The classification problem becomes particularly complex and difficult to model. Artificial intelligence (AI) technology has been recognized by both academia and industry as the most promising solution for many electromagnetic field problems, including spectrum sensing, radio resource assignment, and electromagnetic signal classification, due to its powerful feature extraction capabilities and low prior requirements [2].

AI has been widely adopted in many prior works to address the AMC problem. As the basis for applying AI techniques to the AMC problem, modulation signal datasets has been created [3], [4]. Drawing on the experience in the field of computer vision (CV), a method of preprocessing signals into images and classifying them with AI technology is proposed in [5]. Further, end-to-end AMC models for classifying time-domain waveforms has been studied [6]. In order to improve the applicability in different scenarios, few-shot [7], zero-shot [8], transfer learning [9], attention mechanism [10] and the adversarial attack problem [11], [12] have been studied. In addition, lightweight techniques such as network quantization [13] have proposed to reduce model complexity and make it more suitable for industrial applications.

In the field of CV, the earliest and fastest-growing field of AI, a variety of novel network structures are proposed. Additionally, various loss functions are designed and proposed to guide the network training process. For example, the center loss is proposed in [14], which improves the intra-class compactness by learning the center of deep features for each class and penalizing the distance between deep features and class centers. The Large-Margin softmax loss is proposed in [15], which defines an adjustable margin in the softmax loss function to expand the decision margin among classes. Under clear intuition and geometric interpretation, the weights of the last layer of the model are considered as the decision vector space. The Inter-class Angular loss (ICAL) function that increases the angle between classes to expand the decision space is proposed to improve the inter-class variability [16]. Further, Focal Inter-Class Angular loss (FICAL) is proposed in [17], which focuses more on the angle between two classes that have a higher confusion rate.

Inspired by the design concepts of ICAL and FICAL loss functions, this paper aims to identify their hidden defects and propose mitigation solutions by increasing the maximum focal angle. Further, the domain characteristic data distribution of the modulated signal in the AMC model is analyzed. The maximum Focal Inter-Class Angular Loss function with Norm Constraint (MFICAL-NC) is proposed to guide the model to converge towards a better state. We also evaluate the proposed MFICAL-NC with extensive experiments.

The remainder of this paper is organized as follows. In Section II, we present the system model. In Section III, we introduce the proposed algorithm, MFICAL-NC, in detail based on the two algorithms. In Section IV, we designed extensive experiments to prove the effectiveness of the proposed algorithm and verify the rationality. Finally, Section V concludes this paper with a discussion of future work.

II. System Model

The research perspective of this paper falls on the output layer of the deep learning model, which is composed of a fully connected layer and a $\mathrm{softmax}$ function. The input-output relationship of the output layer is given by:

$$\begin{aligned} \mathbf{Y} &= \mathbf{W}^{\mathrm{T}}\mathbf{X} + \mathbf{B} \\ \mathbf{O} &= \mathrm{softmax}(\mathbf{Y}), \end{aligned} \tag{1}$$

where, $\mathbf{X} \in \mathbf{R}^{M\times 1}$ represents the input to the layer; $\mathbf{w} \in \mathbf{R}^{M\times N}$ represents the weights; $\mathbf{B}$ represents the bias; M is the width of the input to the layer; N is the number of neurons, as well as the number of classes. To facilitate subsequent discussions, suppose $\mathbf{A} = \mathbf{W}^{\mathrm{T}}$. The output of the neuron corresponding to the i-th class, y_i, is given in (2), and which can be rewritten as in (3).

$$y_i = \mathbf{A}_{(i,)}\mathbf{X} = \sum_{j=1}^{m} a_{i,j}x_j \tag{2}$$

$$y_i = \left\|\mathbf{A}_{(i,)}\right\|_2 \|\mathbf{X}\|_2 \cos\left(\theta_i\right), \tag{3}$$

where, $\mathbf{A}_{(i,)}$ represents the i-th row of the transposed weight matrix; $a_{i,j}$ is the element in the i-th row and j-th column of the weight matrix; and $\|\cdot\|_2$ represents the norm of a vector. θ_i represents the angle between $\mathbf{A}_{(i,)}$ and $\mathbf{X}$. The diagram of the above calculation process is illustrated in Fig. 1.

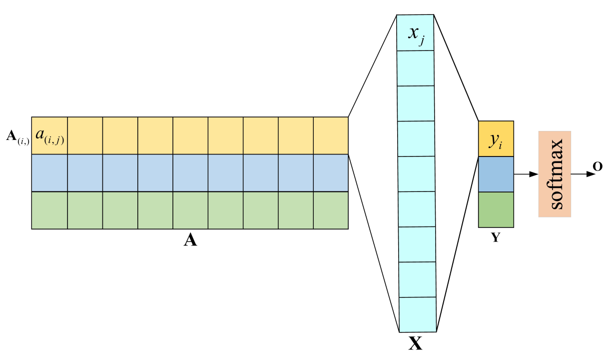


Fig. 1. Diagram illustration of the computations in output layer of the deep learning model.

Therefore, the classification process is to project the input vector $\mathbf{X}$ to the vector space composed of the weight vectors corresponding to different classes, and obtain the highest inner product value. Similarly, the inner product operation can also be performed between the weight vectors corresponding to the two classes, and the cosine value of the angle between them can be calculated as:

$$\cos\left(\theta_{ij}\right) = \frac{\mathbf{A}_{(i,)}^{\mathrm{T}}\mathbf{A}_{(j,)}}{\left\|\mathbf{A}_{(i,)}\right\|_2 \left\|\mathbf{A}_{(j,)}\right\|_2}. \tag{4}$$

Obviously, increasing the angle between the weight vectors corresponding to the two classes in the vector space can reduce the similarity of the two classes in the decision space, open up the decision space, and then improve the distinction between the two classes.

III. The Proposed Methodology

Based on the above model in Section II, ICAL [16] and its improved version FICAL [17] are proposed. The method in this paper is an improved and supplemental version to the existing methods after our detailed analysis.

A. ICAL

In the ICAL algorithm, the average sum of cosine similarity (termed AverageSim) loss function is proposed as an additional regularization term to the original loss function. The expression of the joint loss function is given by:

$$\begin{aligned} L_{\text{ICAL}} = &\frac{1}{B}\sum_{i=1}^{B} -\log\left(\frac{e^{y_i}}{\sum_{j=1}^{N} e^{y_j}}\right) \\ &+ \alpha\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\cos\left(\theta_{ij}\right), \end{aligned} \tag{5}$$

where the first term is the original loss function, B represents the number of samples in the training batch, and the second term is the AverageSim loss function is the regularization term. During the training process, the loss value gets smaller and the angle between the classes gets larger. In (5) α is the hyperparameter of the penalty strength of the regular term, which reflects the degree to which AverageSim is valued.

B. FICAL

In [17], the authors showed that the ICAL algorithm had the same expansion effect for the angle between all class pairs, without providing giving a stronger expansion for the angle between the two classes with higher confusion. To address this problem, the FICAL algorithm is proposed, in which a dynamic penalty strength parameter is designed according to the confusion rate between two classes. The dynamic penalty strength parameter d_{ij} between the i-th and j-th class is given as:

$$d_{ij} = \sqrt{c_{ij}c_{ji}}, \tag{6}$$

where $c_{ij} = n_{ij}/\sum_{k=1}^{N} n_{ik}$ is the confusion rate at which the i-th class of data is incorrectly classified as the j-th class, n_{ik} is the number of samples in the i-th class that is incorrectly classified as the k-th class. Therefore, the expression of the regular term L_{FICAL} in the FICAL algorithm is given as:

$$L_{\text{FICAL}} = \frac{1}{N^2}\sum_{i,j=1}^{N} d_{ij}^{\beta}\cos\left(\theta_{ij}\right), \tag{7}$$

where β is the control factor for the confusion rate.

C. MFICAL-NC

The FICAL algorithm makes greater efforts to expand the angle between classes with high confusion rates. However, it can be shown that the algorithms based on cosine value have a huge defect. The derivative function of the $\cos(x)$ function is $-\sin(x)$. The update direction of the optimization algorithm is the opposite direction of the gradient. The expansion strength coefficient of the inter-class angle is the $\sin(x)$ function as shown in Fig. 2. Therefore, these algorithms have inherently a limited expansion strength for smaller angles than larger angles.

Fig. 2. The trend and the gradient of the ICAL and FICAL algorithms.

To mitigate this defect, this paper proposes to find the angle between the current class and some class that the data in the current class is misclassified as the most. This class is called the maximum confusion class for the current class. We then find the largest confusion class of all classes and only expand these inter-class angles. This algorithm removes the presence of other larger inter-class angles in each calculation, which overcomes the hindrance of the expansion of smaller angles. The maximum focal inter-class loss function is given by:

$$L_{\text{MFICAL}} = \frac{1}{N}\sum_{i=1}^{N} d_{ij}^{\beta}\cos(\theta_{ij}), \quad s.t. \quad j = \max_{j;j\neq i} \mathrm{c}_{ij}. \quad (8)$$

Inaddition, this paper further extends the analysis from the output layer weight to the $softmax$ function. In order to explore the impact of the output layer decision space on the softmax layer, the statistical results of the norm value of the input data $\mathbf{X}$, the class weight $\mathbf{A}_{(i,)}$, and the value of the output y_i are shown in Table I. From the table, it can be seen that the norm value of the input data $\mathbf{X}$ is much larger than the class weight.

TABLE I
THE STATISTICAL RESULTS OF THE INPUT DATA, THE CLASS WEIGHT AND THE VALUE OF THE OUTPUT

Item	Maximum	Minimum	Mean	Variance
Input data $\mathbf{X}$	254.40	0.07	58.36	1829.31
Class weight $\mathbf{A}_{(i,)}$	1.01	0.75	0.82	0.01
Output y_i	31.59	-45.74	-2.72	66.68

Although this does not affect the output value weighting relationship, it does result in a wider distribution range or larger values of the output y_i. Due to the e^x function as shown in Fig. 3, confidence differences are magnified after the $softmax$ layer when y_i has a wide distribution or takes large values.

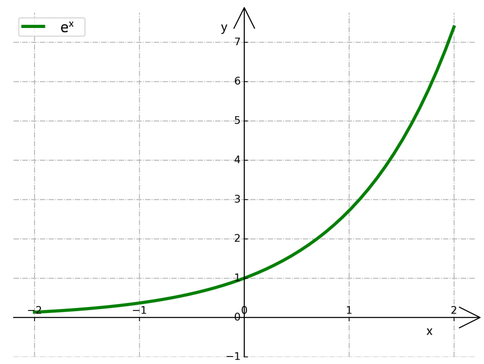


Fig. 3. The curve of the e^x function.

During model training, it is possible that some points have not been adequately fitted and are in an ambiguous situation. However, because $softmax$ magnifies the confidence difference, it presents the illusion of a good fit. As a result, the entire network will not be adequately fitted to the data. Aiming at this problem, this paper designs an algorithm of norm constraint on the input data $\mathbf{X}$, in order to narrow down the distribution range of the output data y_i, or reduce their value as given in (9). In this way, the distribution difference of the data passing through the e^x function will be reduced, thereby reducing the confidence difference and making the network a good fit. Because adding a constraint to the output data will affect the inter-class angles in the output layer, the norm constraint is applied to the input data, instead of the output data, as

$$L_{\text{NC}} = \sum_{i}^{B} \|\mathbf{X}_i\|_2 . \quad (9)$$

The performance of a network is dependent on multiple factors. This paper designs a combination of the maximum focal inter-class angle loss and the norm value constraint. Algorithm 1 presents the detailed steps of the algorithm.

IV. EXPERIMENT

In this paper, the AMC problem is chosen as the target application for evaluating the effectiveness of the proposed algorithm. The dataset RML2016.10a [3] is used, which contains 8 types of digital modulation and 3 types of analog modulation. The signal-to-noise ratio (SNR) of the data is from -20dB to 18dB in 2dB intervals. The dataset is divided into the training set, the cross-validation set, and the test set with a ratio of 8:1:1. The AMC model with the softmax loss is set as the baseline model, and the ICAL, FICAL, and MFICAL-NC loss functions will be used. The front end of the model is designed to be wide to extract more features. As the level

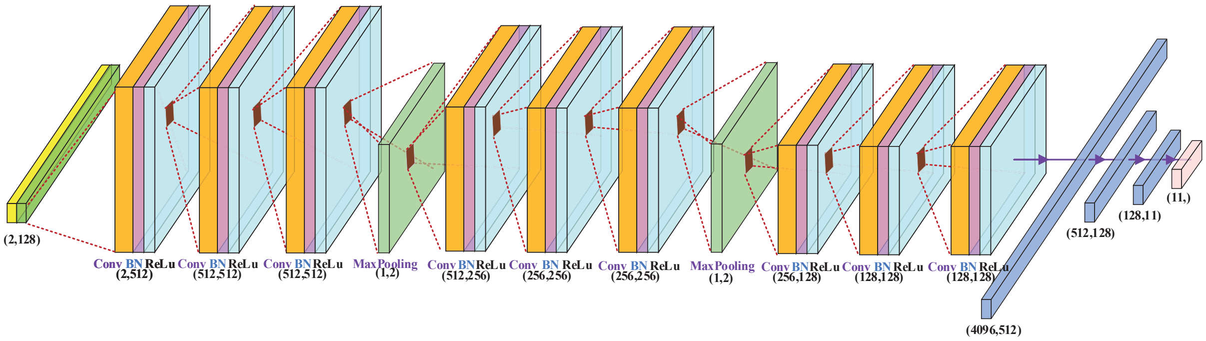


Fig. 4. The network structure diagram of the AMC model.

Algorithm 1 MFICAL-NC

Input: The AMC model $f_\theta(\cdot)$; validation dataset $(\mathbf{d}_v, \mathbf{l}_v)$; the original loss function L; class set $\mathbf{C}$; update function $U(\cdot,\cdot)$; the total number of iterations T;

Output: The AMC model $f_\theta(\cdot)$;

Initialize $f_\theta(\cdot)$;
for epoch t = 0 to $T-1$ **do**
  Input $\mathbf{d}_v$ into $f_\theta(\cdot)$, get the input of the last layer $\mathbf{x}$ and predict labels $\mathbf{l}_p$;
  Calculate the norm values of $\mathbf{x}$ and sum them up to L_{NC};
  Get the confusion matrix with $\mathbf{l}_p$ and $\mathbf{l}_v$;
  Set L_{MFICAL} = 0;
  for epoch c in $\mathbf{C}$ **do**
    Get The maximum confusion class m of c and calculate the $\cos(\theta_{cm})$:

$$\cos(\theta_{cm}) = \frac{\mathbf{A}_{(c,)}^{\mathrm{T}} \mathbf{A}_{(m,)}}{\left\|\mathbf{A}_{(c,)}\right\|_2 \left\|\mathbf{A}_{(m,)}\right\|_2}. \tag{10}$$

    Add $\cos(\theta_{cm})$ to L_{MFICAL};
  end for
  $L_{\mathrm{MFICAL\text{-}NC}} = L + L_{\mathrm{MFICAL}} + L_{\mathrm{NC}}$;
  Update the model: $f_\theta(\cdot) = U(f_\theta(\cdot), L_{\mathrm{MFICAL\text{-}NC}})$;
end for
return $f_\theta(\cdot)$;

of feature abstraction increases, the network is subsequently narrowed down to facilitate network fitting. Finally, stacked fully connected layers are used as classifiers. The network structure diagram of the AMC model is shown in Fig. 4. Next, the effectiveness of the algorithm will be analyzed from different aspects.

A. Comparison of classification accuracy

Classification accuracy is the most straightforward indicator to measure the effectiveness of the proposed algorithm. For a fair comparison, the learning rate scheme, loss function hyperparameters and optimization algorithms of the different algorithms evaluated in the experiment are consistent. The classification accuracies under different SNRs for different algorithms are shown in Fig. 5.

It can be seen from Fig. 5 that with the improvement of SNR, the classification accuracies of the four algorithms all exhibit an upward trend. When the SNR is lower than 14dB,

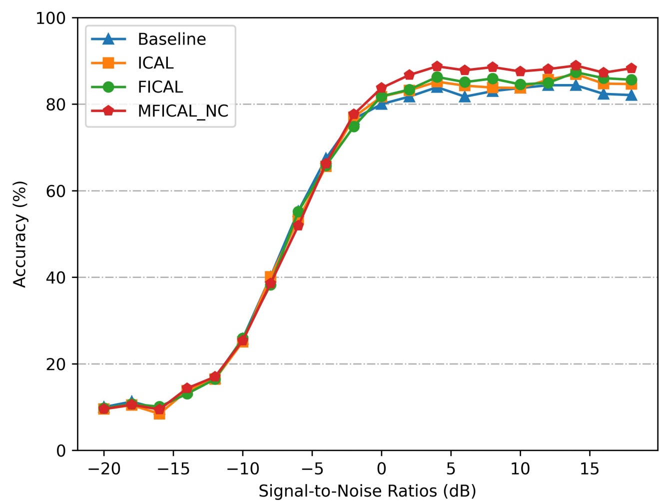


Fig. 5. Classification accuracy curves of the four algorithms under different SNRs.

none of the four models can effectively classify the modulation classes, and the results exhibits considerable randomness. When the SNR is increased to about -2dB, the proposed MFICAL-NC algorithm stands out from other algorithms and achieves the highest classification accuracy. When the SNR is increased to about 6dB, the classification accuracy tends to be stable. From the baseline network, to ICAL, to FICAL, and finally to the proposed MFICAL-NC algorithm, the classification accuracy increases in turn. A series of algorithms for expansion of the inter-class angle improves the classification accuracy, which implies the correlation between the classification accuracy and the inter-class angle, and the reasonable expansion of the inter-class angle helps to improve the model's classification performance. Especially, the result shows that MFICAL-NC can guide the model training process to achieve a better fit and generalization ability.

B. Comparison of Angle Expansion Effect

Further, the angle expansion effect, as a validation indicator, is used to verify whether the proposed MFICAL-NC algorithm can effectively expand the angle of the decision space. The confusion matrix is a commonly used indicator to present the degree of confusion between classes. Since the data at an SNR lower than -14dB is overwhelmed by noise and cannot be classified, we plot the confusion matrix for signals at an SNR from -14dB to 18dB as shown in Fig. 6. The confusion matrix will serve as a reference for analyzing the inter-class angle matrix of the four algorithms. The inter-class angle matrix of the four algorithms shown in Fig. 7.

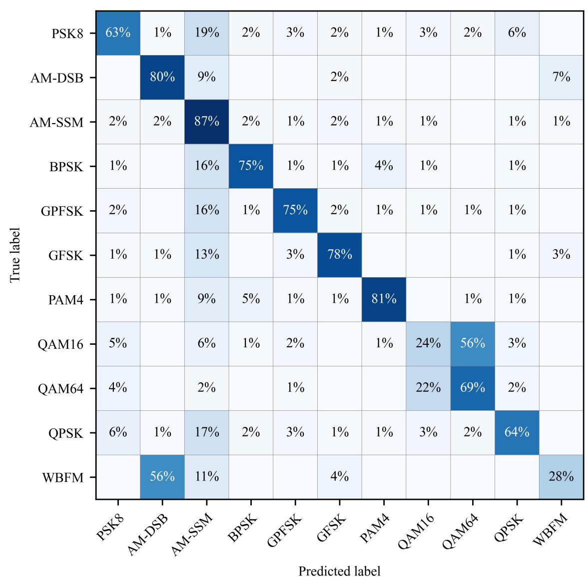
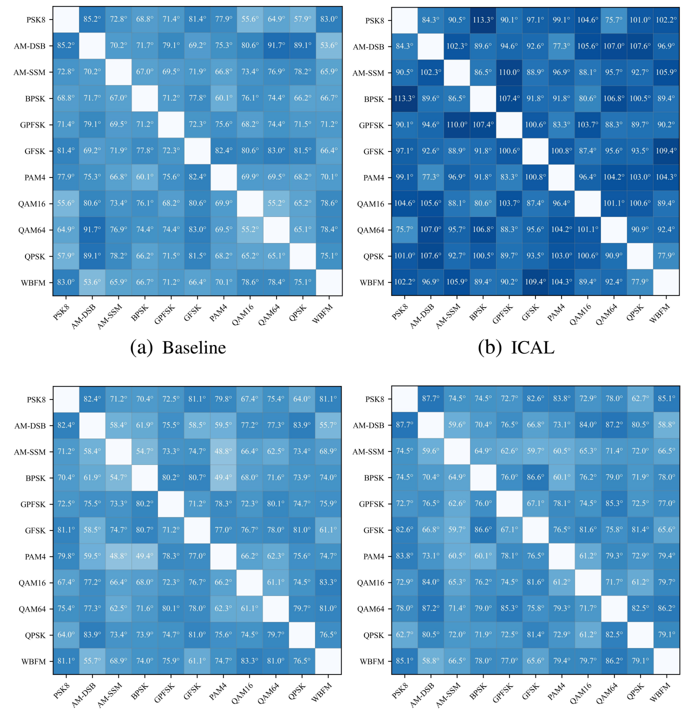


Fig. 6. Confusion matrix for data at SNR from -14dB to 18dB under obtained by the baseline model.

As can be seen in Fig. 6, the most severe confusion occurs between QAM16 and QAM64. Higher-order modulations are less distinguishable. QAM16 and QAM64 both belong to the same broad category, so it is reasonable to have a more serious confusion between the two. In addition, WBFM is misclassified as AM-DSB with a high probability, both of which are analog modulations. Further, modulation methods including PSK8, AM-DSB, BPSK, GPFSK, and QPSK, have also been misclassified as AM-SSM to a certain extent.

The inter-class angle matrix will be analyzed with the confusion matrix. From the inter-class angle matrix of the baseline model, it can be seen that the inter-class angle between the two classes where obvious confusion occurs is generally small, such as (QAM16, QAM64), (WBFM, AM-DSB) and (PSK8, AM-SSM). With the ICAL model, all the inter-class angles are generally improved, but the accuracy is significantly lower than that of the FICAL model and the MFICAL-NC model. Therefore, all the inter-class angles in a model are strongly correlated. Some inter-class angles are greatly expanded, which negatively affects some classes whose inter-class angles are slightly expanded. This phenomenon is due to the $\cos(x)$ function, which introduces unfairness. In the FICAL model and the MFICAL-NC model, since the expansion of angles is based on the inter-class confusion rate, the expansion behavior is more selective and targeted. The baseline model, FICAL model, and MFICAL-NC model are



(a) Baseline (b) ICAL (c) FICAL (d) MFICAL-NC

Fig. 7. The inter-class angle matrices obtained by the four algorithms.

continuously improved with respect to the expansion effect on the inter-class angles of (QAM16, QAM64), (WBFM, AM-DSB), and (PSK8, AM-SSM). Further more, MFICAL-NC narrows down the inter-class angle for some classes that are highly separable, unlike the ICAL algorithm, which is a desirable phenomenon. This implies that the model assigns redundant fitting power among other classes to more difficult-to-discriminate classes.

C. Comparison of Output Data Distribution

Finally, the output data distribution is also a verification indicator. It is used to verify whether constraining the norm value of the input data of the output layer can narrow down the distribution range or reduce the distribution location of the output data. We next collect the test set output y_i of the models under the four algorithms, and draw their data distribution in Fig. 8.

The location of the output data distribution of the baseline model and the ICAL model is larger than that of the FICAL model and the MFICAL-NC model. Due to the e^x function in the $softmax$ layer, the larger the data distribution position, the larger the confidence difference, which hides the truth that the model has not been well fitted. In this paper, by explicitly adding the norm constraint loss term to constrain the data distribution range, this problem is exposed and the network can be fully trained for a stronger generalization ability. It can be seen that the FICAL network without this loss term also shows a similar effect that the location of data distribution is smaller, which proves the correctness of the assumption made in this

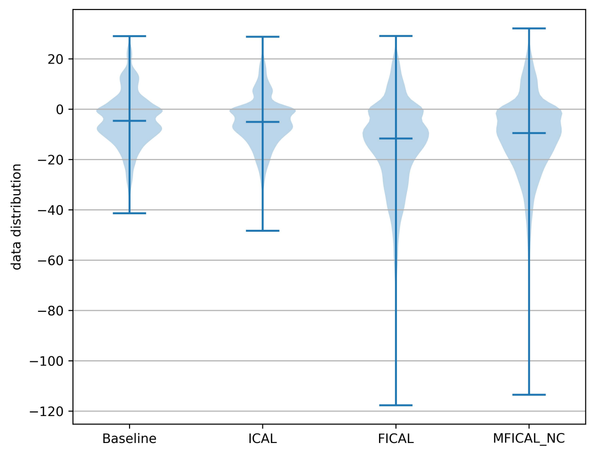


Fig. 8. The distribution of the output data obtained by different models.

paper. It is reasonable that the MFICAL-NC model obtains the best classification performance among the four algorithms under the joint action of the two loss terms, the maximum focal inter-class angle and the norm constraint.

V. Summary and Future Work

Automatic modulation classification based on artificial intelligence is playing an increasingly important role in electromagnetic applications such as spectrum monitoring, cognitive radio networks, and electromagnetic management. Inspired by the design concepts of ICAL and FICAL, this paper proposed a Maximum Focal Inter-Class Angular Loss function with Norm Constraint scheme, MFICAL-NC, to guide the model to converge towards a better state and generalization ability. Specifically, the output layer of the network is regarded as the decision space composed of the vectors corresponding to the classes. The classification process is viewed as a projection of the input data points of the output layer to the decision space. On one hand, the mathematical mechanisms underlying the inherent defect of the ICAL and FICAL algorithms that attempt to expand the decision space to obtain classification performance are explained. This paper addressed this problem by only expanding the angle between the classes with the largest confusion rate per epoch. On the other hand, this paper studied the influence of data distribution on the confidence difference after the softmax layer, and proposed a norm constraint, so that the network could be fully trained. The experimental results proved the effectiveness of the MFICAL-NC algorithm and verified the correctness of the assumption.

The AMC model based on deep learning is a black-box approach and uninterpretable, which is a hindrance to further improving the model performance for secure and trusted applications. Based on the prior knowledge of the data distribution of the modulated signal data in the model, this paper analyzed the model training and decision process, and designed a loss function to guide the model training. The rapid development of deep learning technology in the fields of computer vision and natural language processing is inseparable from the deep understanding of data characteristics and the combination of its prior knowledge. Therefore, combining prior knowledge, analyzing models, and designing constraints to guide model training will become an important way to improve model performance and interpretability.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (61771154). This work is also supported by the Key Laboratory of Advanced Marine Communication and Information Technology, the Ministry of Industry and Information Technology, and Harbin Engineering University, Harbin, China.

References

[1] S. Peng, S. Sun, and Y.-D. Yao, "A survey of modulation classification using deep learning: Signal representation and data preprocessing," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–19, 2021.

[2] M. Lin and Y. Zhao, "Artificial intelligence-empowered resource management for future wireless communications: A survey," *China Communications*, vol. 17, no. 3, pp. 58–77, Mar. 2020.

[3] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Engineering Applications of Neural Networks*, C. Jayne and L. Iliadis, Eds., Cham, 2016, pp. 213–226.

[4] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, Feb. 2018.

[5] Y. Lin, Y. Tu, Z. Dou, L. Chen, and S. Mao, "Contour stella image and deep learning for signal recognition in the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 1, Mar. 2021.

[6] Z. Ke and H. Vikalo, "Real-time radio technology and modulation classification via an LSTM auto-encoder," *IEEE Transactions on Wireless Communications*, vol. 21, no. 1, pp. 370–382, Jan. 2022.

[7] Q. Zhou, R. Zhang, J. Mu, H. Zhang, F. Zhang, and X. Jing, "AM-CRN: Few-shot learning for automatic modulation classification," *IEEE Communications Letters*, pp. 1–1, 2021.

[8] Y. Dong, X. Jiang, H. Zhou, Y. Lin, and Q. Shi, "SR2CNN: Zero-shot learning for signal recognition," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2316–2329, 2021.

[9] M. Wang, Y. Lin, Q. Tian, and G. Si, "Transfer learning promotes 6G wireless communications: Recent advances and future challenges," *IEEE Transactions on Reliability*, vol. 70, no. 2, pp. 790–807, 2021.

[10] Z. Liang, L. Wang, M. Tao, J. Xie, and X. Yang, "Attention mechanism based resnext network for automatic modulation classification," in *2021 IEEE Globecom Workshops (GC Wkshps)*, 2021, pp. 1–6.

[11] Z. Bao, Y. Lin, S. Zhang, Z. Li, and S. Mao, "Threat of adversarial attacks on dl-based iot device identification," *IEEE Internet of Things Journal*, pp. 1–1, 2021.

[12] H. Zhao, Y. Lin, S. Gao, and S. Yu, "Evaluating and improving adversarial attacks on dnn-based modulation recognition," in *Proc. IEEE GLOBECOM'20*, Taipei, Taiwan, Dec. 2020, pp. 1–5.

[13] S. ZHANG, Y. LIN, Y. TU, and S. Mao, "Electromagnetic signal modulation recognition technology based on lightweight deep neural network," *Journal on Communications*, vol. 41, no. 11, Nov. 2020.

[14] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 499–515.

[15] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-margin softmax loss for convolutional neural networks." in *ICML*, vol. 2, no. 3, Jun., p. 7.

[16] L. Hui, X. Li, C. Gong, M. Fang, J. T. Zhou, and J. Yang, "Inter-class angular loss for convolutional neural networks," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 3894–3901, Jul. 2019.

[17] X. Wei, D. Chang, J. Xie, Y. Zheng, C. Gong, C. Zhang, and Z. Ma, "Fical: Focal inter-class angular loss for image classification," in *2019 IEEE Visual Communications and Image Processing (VCIP)*, Dec. 2019, pp. 1–4.