Deep Learning Based Classification of Radar Spectral Maps

Tong Lin, Xin Chen, Xiao Tang, Ling He, Song He, and Qiaolin Hu
Air Force Early Warning Academy, Wuhan, China
Email: ltong9463@gmail.com; cxin917@126.com; tangxiao@whu.edu.cn; {1006012426; hesong_1988; 28772844}@qq.com

***Abstract*—This paper discusses the use of deep convolutional neural networks for radar target classification. In this paper, three parts of the work are carried out: firstly, effective data enhancement methods are used to augment the dataset and address unbalanced datasets. Second, using deep learning techniques, we explore an effective framework for classifying and identifying targets based on radar spectral map data. By using data enhancement and the framework, we achieved an overall classification accuracy of 0.946. In the end, we researched the automatic annotation of image ROI (region of interest). By adjusting the model, we obtained a 93% accuracy in automatic labeling and classification of targets for both car and cyclist categories.**

***Index Terms*—Radar spectral, deep learning, target recognition**

I. Introduction

Radar is capable of detecting long-range targets day and night because of its all-weather, round-the-clock nature. It is unobstructed by fog, clouds and rain and has some ability to penetrate. Therefore, it has an extremely broad range of applications in socio-economic development and scientific research [1]. For example, Autonomous driving technology will have great application value in the future. As one of the key sensing means of automatic driving, radar has the advantage of all-weather and all-weather performance and can be accurately used in unmanned vehicles. Therefore, it plays an important role in identifying the surrounding environment.

Along with the radar data collection ability to continuously improve, the automatic interpretation of radar data has received widespread attention [2], the radar target identification is one of the important research directions. Because synthetic aperture radar can achieve high-resolution imaging. Most of the researches on radar target identification are carried out on SAR images [3]. Most radars, however, use traditional signal processing methods to achieve target information extraction. The accurate identification of targets is always difficult to enhance due to various factors.

In recent years, with the continuous development of deep learning theory, research based on it has made breakthrough progress in many fields [4]. Convolutional neural networks (CNNs) are most commonly used model for learning and has become one of the main methods in the field of image classification.

This paper explores a framework for radar target classification using deep learning techniques, combined with current big data research and high-performance computing techniques. By using data augmentation techniques and a suitable classification network, this framework achieves an overall classification accuracy of 0.946 on our dataset.

This work contributes three aspects to the study of radar target recognition. i) We use appropriate data enhancement methods to augment the dataset and address unbalanced datasets. ii) Training the model with the effective CNN (convolutional neural network) model to achieve the classification of radar spectral picture data. We determine the appropriate convolutional neural network for model training based on the characteristics of the spectrum picture. Transfer learning strategy is adopted to improve the training rate and efficiency. Through many attempts to train the model, we adjust the network weight parameters and train the model by combining the literature experience, and finally get a more effective Identify the model. (iii) The YOLO algorithm was used to complete the region of interest (ROI) of the automatically labeled spectrum picture, and the model classification performance was greatly improved by the ROI data.

II. Methodology

A. Data Enhancement

The advantages of deep learning techniques in solving the problem of radar image target recognition are mainly in their automatic learning to extract valid features from the data. The mechanism of the method. Due to the complex scattering mechanism of radar images, it is difficult to manually design effective features for radar image target recognition. The features extracted by the deep learning method are expected to greatly exceed the manually designed features in classification performance [5].

Deep learning models must rely on large amounts of data, and radar spectrogram data is often difficult to obtain. Direct training of deep neural networks under

Manuscript received June 22, 2020; revised August 1, 2020; accepted August 22, 2020.

Corresponding author: Tong Lin (email: ltong9463@gmail.com).

doi: 10.18178/ijeetc.10.2.99-104

small sample conditions can produce serious overfitting problems, resulting in an extreme reduction in the generalization capability of the network. Insufficient sample size has become a major factor limiting the application of deep learning methods in radar image target recognition. Therefore, it is necessary to perform data enhancement on the dataset [6].

We tested several common data enhancement methods [7], and found several data enhancement methods, such as image panning and brightness contrast adjustment, that do not change the spectral map image characteristics [8]. We use image processing functions in the OpenCV library for data enhancement. In the image flipping part, we use the Image.FLIP_TOP_BOTTOM function. Image panning was performed using cv2.warpAffine(img, M, (cols, rows)) function, where **M** is the variation matrix, set to [[1, 0, 10], [0, 1, −10]]. The cols and rows are the translational distances in both vertical and horizontal directions. Brightness contrast adjustment is calculated using $g(x)=\alpha f(x)+\beta$, where α (>0) and β are called gain and offset values, which control the contrast and brightness of the image, respectively. Contrast needs to be controlled by α and β together, and brightness is controlled by β. In fine tuning $\alpha = 1.1$ and $\beta = 5$.

B. Convolutional Neural Network Architecture

The classification network is based on the ResNet model. It is an excellent performing CNN model. The core idea of ResNet is an identity shortcut connection that skips one or more layers directly [9].

The ResNet network is a reference to the VGG19 [10] network, modified from it, with residuals added through a short-circuit mechanism unit, as shown in Fig. 1. The main changes are that ResNet does downsampling directly with the stride=2 convolution. An important design principle of ResNet is that when the number of feature maps doubles when the feature map size is halved, which keeps the complexity of the network layers. As can be seen in Fig. 2, ResNet adds a short-circuit mechanism between every two layers compared to a normal network, which results in residual learning [11].

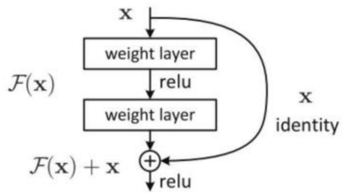


Fig. 1. Residual learning: a building block. Photograph source: [arXiv:1512.03385v1 [cs.CV] 10 Dec 2015]

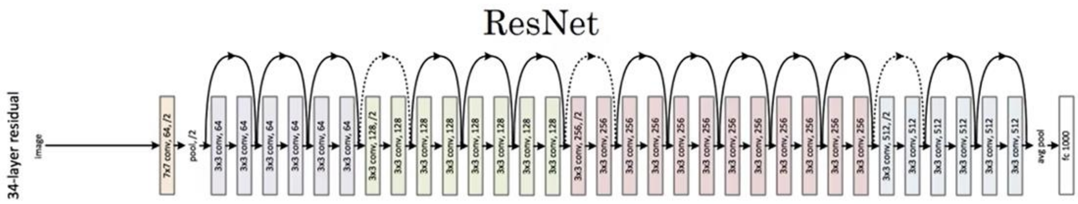


Fig. 2. Structure of ResNet. Photograph source: [arXiv:1512.03385v1 [cs.CV] 10 Dec 2015].

These models were trained by using the transfer learning strategy [12]. Transfer learning is an approach in the field of deep learning. Using pre-trained models as a starting point for new models. Transfer learning allows you to transfer powerful skills that have been acquired to related problems.

In this paper, pre-trained models, vgg16 and resnet18, from the Pytorch were used. These models are trained on the large dataset ImageNet. We used transfer learning strategy to fine-tune the model.

We retain the convolutional part of the model, while for the full-connection layer, it is adjusted to the actual situation of the training set. The full connection layer of the original model is too large for the dataset in this paper. For the dataset in this paper, the fully connected layer is partially modified into 3 layers with 512 nodes in the first layer and 256 nodes in the second layer. The number of nodes in the last layer is equal to the number of categories. Depending on the size of the data put into training, the number of nodes and layers can be reduced by an appropriate amount. Use Relu as the activation function for the first two fully connected layers to get a better fit. And use the dropout method to adjust the parameters according to the training effect to enhance the generalization performance of the model. During training, we use Adam's algorithm as a gradient descent algorithm to converge the model faster while avoiding model Performance Oscillation. The learning rate is exponentially decayed and the initial learning rate is set to 0.002.

C. YOLO Architecture

YOLO (you only look once) model is an object detector that uses features learned by a deep convolutional neural network to detect an object, as shown in Fig. 3.

YOLO model frames object detection as a regression problem to spatially separated bounding boxes and associated class probabilities.

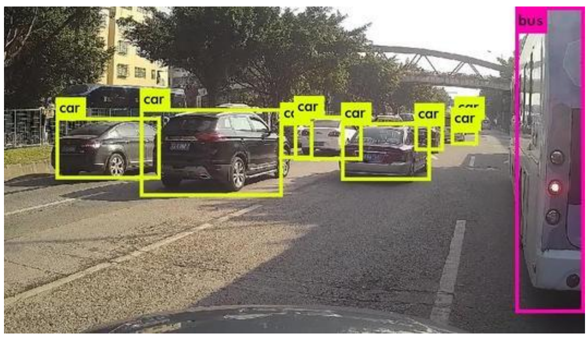


Fig. 3. Object detect based YOLO model. Photograph source: https://zhuanlan.zhihu.com/p/31227909

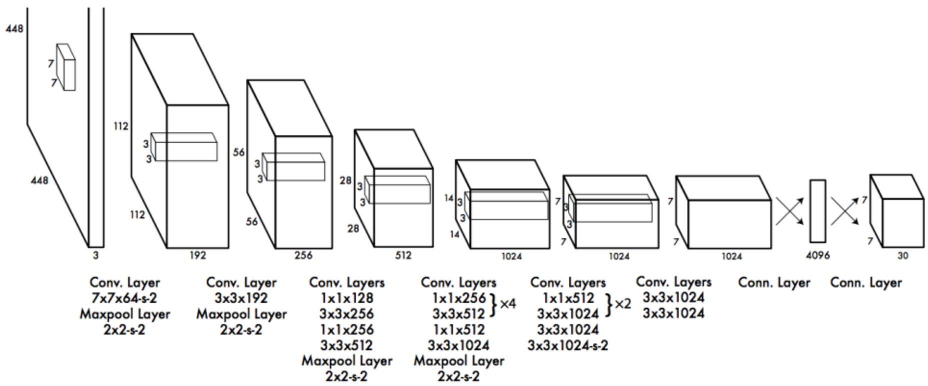


Fig. 4. Structure of YOLO model [13].

YOLO makes use of only convolutional layers, making it a fully convolutional network (FCN). It has 75 convolutional layers, with skip connections and up-sampling layers. No form of pooling is used, and a convolutional layer with stride 2 is used to down-sample the feature maps. This helps in preventing loss of low-level features often attributed to pooling [13], as shown in Fig. 4. We will use PyTorch to implement an object detector based on YOLO v3.

YOLO v3 is the enhanced model from YOLO and YOLOv2. The convolutional layers are based on the darknet53. It borrows from ResNet by setting up shortcut connections between some layers. Instead of using softmax when predicting object classes, it uses logistic output for prediction. This enables support for multi-labeled objects (e.g. a car has both truck and car labels).

In summary, YOLO3 draws on the residual network structure to form deeper network layers and multi-scale detection to improve small object detection. These properties make the YOLO3 model more suitable for the radar spectrogram ROI annotation task.

III. Experiments and Analysis

A. Dataset

The training of deep convolutional neural network models relies on a large amount of sample data, and all experiments in this paper make use of the 77GHz FM continuous-wave radar spectrogram dataset provided by Pérez Rodrigo at the Technical University of Munich [14].

To perform the measurements and acquire the data a test vehicle was used. The vehicle is a BMW 5 Series equipped with a variety of sensors. For this work only the radar, the front lidar and the front camera are of relevance. The radar system was fitted within the right kidney grill using a specially designed case. The front camera, mounted in place of the rear-view mirror, is exclusively used as an aid to label the radar data. The data was gathered by driving in the surrounding area of the Technical University of Munich. This resulted in a varied number of scenarios in real urban settings, i.e. subjects from all classes in a varied range of directions (both lateral and longitudinal with respect to the radar's orientation). All measurements were performed during daytime with no precipitation present. In order to process the dataset the ROIs given by the lidar object lists were labeled semi-automatically and all frames were controlled and, if necessary, corrected manually. Since the classification approach produces one prediction per ROI, only tracks with one target present or with a clear dominant target within the ROI were selected.

The spectral map data are divided into four main categories: pedestrians, cars, cyclists, and noise (no targets). And the cars are divided into two categories: cars and trucks. There are numerous targets in each category of data, and each target is labeled as a target ID. The radar picks up several frames of each target to dozens of frames ranging from slices of data. Therefore, to make it easier to track the effect of model training, each image is named as TargetID_Frame where the TargetID represents the target label and the Frame represents the frame of this target. The raw spectrogram image is processed as a three-channel jpg with a width of 515 pixels and a height of 176 pixels. And the ROI spectrogram image is processed as a three-channel jpg with a width of 48 pixels and a height of 32 pixels.

The dataset has two kinds of data, raw data and ROI data, which will be made into a raw spectrogram dataset and a ROI spectrogram dataset, respectively, and the pictures of these two datasets correspond to each other, as shown in Fig. 5.

In order to make more efficient use of the data and to facilitate the adjustment of model parameters and the evaluation of model effects, the raw spectral map dataset, and the ROI dataset were divided into three categories, respectively: training set (for training neural network model parameters), validation set (for tuning and selecting models during training), and test set (for evaluating model effects), as shown in Table I.

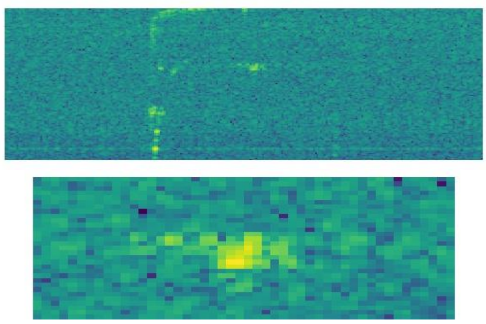
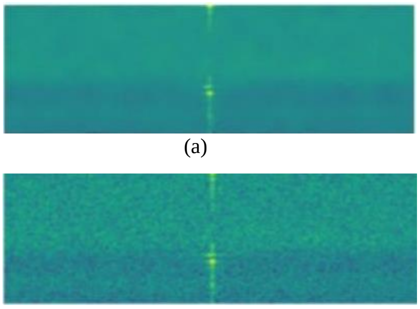


Fig. 5. Raw data (top) and ROI data (bottom): both first cyclist data with ID=1.

TABLE I: DATA SET SAMPLE DISTRIBUTION

Class	Number of validation samples	Number of train samples	Number of test samples
Pedestrian	264	740	264
Cyclist	746	1943	746
Car	1834	2616	1834
Noise	388	1473	744


(a)

(b)

Fig. 6. Data enhancement: (a) Median filtering image and (b) Gaussian filtering image.

It is clear to see that the dataset is very long and unbalanced. The proportion of the training set occupied by pedestrian and car data is 11% and 38%, respectively. This results in the model learning more about the features of the car class and ignoring the features of the pedestrian data when training the neural network. In order to improve the model performance and reduce the unevenness of the dataset, the pedestrian data is augmented.

Data enhancement of the raw data set, combined with image flipping, panning, changing the brightness contrast, and filtering, as shown in Fig. 6. The training set was expanded to 5008 images of cars, 3886 images of cyclists, 3512 images of noise, and 4440 images of pedestrians, and only Gaussian filtering was done on the test set.

B. Classification Results

Based on the size of the dataset in this paper, combined with previous work done by Pérez Rodrigo, a simple 6-layers neural network, ResNet18, and ResNet34 (while using dropout) were trained separately to determine a more appropriate network structure [15]. Using the raw spectral map dataset, a round of four-classification (pedestrian, cyclist, car, noise) testing was conducted. With the main objective of this round of testing being to determine the network structure and therefore do not pre-process the dataset.

Based on the experimental results (as shown in Table II), it can be seen that the 6-layer neural network is underfitted to the data and has a low-test set accuracy, while ResNet34 has a severe overfitting performance, and ResNet18 has the best training performance between the three model. So, it can be determined that the model complexity should be more compatible with the data complexity when the depth of the model is similar to the network depth of ResNet18.

TABLE II: THREE MODEL TEST SET ACCURACY

Model	6-layer neural network	ResNet18	ResNet34
test set accuracy	0.56	0.7	0.68

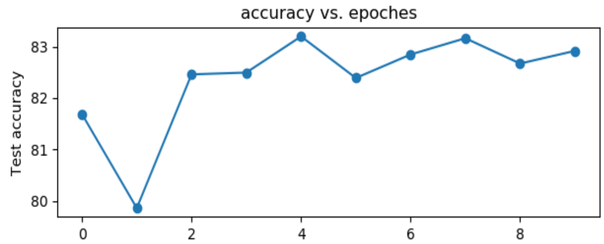


Fig. 7. Five-category task accuracy.

Using Resnet18 for 4-classification, the accuracy of 0.76 was 7 percentage points higher than before the expansion. After using various data enhancement methods, the classification accuracy reached 0.76, but we believe the performance is still insufficient. To further improve the performance, we train the network using ROI data. The ROI picture, which is based on the original spectrum picture, is the target presence region is extracted as a separate dataset. This results in less redundant information for the images and the model learns more about the features of the target, reducing the interference of invalid information. After data enhancement of the ROI dataset, the model is trained. Finally, an overall classification accuracy of 0.946 is obtained.

Finally, we attempt to classify trucks and cars. In the car sample, it is subdivided into trucks and cars, which have smaller differences in the data spectrum picture and are more difficult to classify. Using ResNet18 to classify the car and truck ROI datasets, in general, the lower the number of classifications, the more accurate the classifications are higher. However, in terms of the classification effect, the accuracy of the classification of cars and trucks is only 73%, compared to a classification accuracy of cars and pedestrians of 0.98. It is believed that the model performance is still inadequate for data with more similar characteristics. And then the trucks in the ROI data were isolated from the car sample as a new class. A five-category task (cars, trucks, pedestrians, cyclists, noise) was performed on the network using the model. The dataset is augmented while the samples are balanced. The training was performed using the ResNet18 model, as shown in Fig. 7.

C. Automatic Annotation

In experiments, it has been found that ROI data can greatly improve model performance, and ROI can reduce redundant information in the data, allowing neural networks to focus on the valuable part of the picture. Therefore, we try to implement automatic annotation using the ROI on the spectrum image, as shown in Fig. 8. We used the YOLO3 model [16]. Two types of data, car and cyclist, are used to train the network.

Both LIDAR and cameras were used to assist in the data collection process during dataset production. Using LIDAR and camera data, the ROI of raw data is manually labeled. We find the ROI coordinates in the corresponding raw data by using the target presence area provided by the ROI data. Generate the appropriate label file for the dataset. For each image, it is labeled as "Class_id x_center y_center width height"

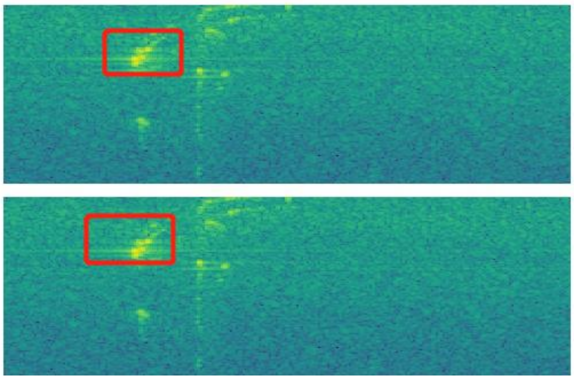


Fig. 8. From top to bottom, two cases are corresponding to the original image and auto-annotated image.

The input to the YOLO network is a 416×416 square image. Therefore, we had to adjust the size of the spectrogram to 416×416. In the configuration file, set the class_number to 2.

For each image, YOLO predicts multiple possible ROIs and gives the probability of the category to which each region belongs. This leads to a lower overall classification accuracy, as we only have one target per image. Therefore, we use the highest probability of the ROI in an image as the image's ROI.

The experimental results show that the YOLO3 [17] model can [18] accurately label the ROI region on the original spectral map and the Images are sorted with a classification accuracy of 0.93. At the same time, we also found an exciting result. On the whole, YOLOv3 automatically labeled ROIs are typically smaller in area than manually labeled ROIs. This means that redundant information for images is further reduced, which can further improve model classification performance.

IV. Conclusion and Outlook

This paper explores a model that can effectively use spectral map data to classify and identify radar targets using deep learning neural network techniques. Using appropriate data enhancement methods to augment the dataset and address unbalanced datasets. Transfer learning was carried out on the ResNet, with data enhancement, dynamic adjustment of the learning rate, parameter normalization, and methods such as dropout further improved the accuracy of the model, resulting in a final dichotomous classification accuracy of 0.98 and overall classification with an accuracy of 0.946. Finally, we attempted to automatically label the ROI of the spectral map using the YOLO algorithm to obtain a classification accuracy of 0.93. ROI with YOLO automatic annotation can be achieved with higher quality than manual annotation.

The model proposed in this paper still has a lot of room for improvement. From the current experimental results, the classification performance of the model still needs to be improved for classification tasks with more categories. In the multi-category task, how to extract target features more effectively, especially to distinguish between two types of targets with more similar data features, is the next step. In addition, ROI data can effectively improve classification performance, and this paper has done a little preliminary research in this regard. In practical applications, how to automatically and accurately label the multi-target ROI in a raw spectrum picture is also a worthy direction of research.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Throughout the study, Xin Chen, Xiao Tang and Tong Lin completed the construction of the classification network; Xiao Tang and Tong Lin conducted data enhancement experiments, and Song He implemented an automated labeled experiment; Tong Lin wrote the paper; Xin Chen, Ling He and Qiaolin Hu provided guidance on the research process; all authors had approved the final version.

Acknowledgment

Support from colleagues, friends, and experts from Cloud Computing Lab is gratefully acknowledged. Also, we would like to extend special thanks to Mr. Rodrigo Pérez at the Technical University of Munich, Germany, for sharing the spectrogram dataset, which provided important support for the work in this paper.

References

[1] A. Bhattacharya and R. Vaughan, "Deep learning radar design for breathing and fall detection," *IEEE Sensors Journal*, vol. 20, no. 9, pp. 5072-5085, May 2020.

[2] J. Dickmann, J. Klappstein, and M. Hahn *et al*., "Automotive radar the key technology for autonomous driving: From detection and ranging to environmental understanding," in *Proc. IEEE Radar Conf. (RadarConf)*, May 2016, pp. 1–6.

[3] W. B. Wang, Y. S. Wen, X. Dong, *et al.*, "Sea ice classification of SAR image based on wavelet transform and gray level co-occurrence matrix," presented at 2015 Fifth Int. Conf. on Instrumentation & Measurement, Computer, Communication and Contro, 2015.

[4] J. Y. Zhou and Z. Yanming, "Application of convolution neural network in image classification and object detection," *Computer Engineering and Applications*, vol. 53, no. 13, pp. 34-41, 2017.

[5] L. David, "Data enhancement," *Practitioners Guide to Data Quality Improvement*, 2011.

[6] F. d. P. Calmon, D. Wei, B. Vinzamuri, K. N. Ramamurthy, and K. R. Varshney, "Data pre-processing for discrimination prevention: information-theoretic optimization and analysis," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 5, pp. 1106-1119, Oct. 2018.

[7] P. D. Kaware, "Pre-processing approach for discrimination prevention in data mining," *Int. Journal of Engineering Sciences & Research Technology*, vol. 3, no. 4, pp. 51-53, 2014.

[8] S. R. William, J. Heidi, M. P Collin, *et al.*, "Optimization of RS-fMRI pre-processing for enhanced signal-noise separation, test-retest reliability, and group discrimination," *Neuroimage*, vol. 117, no. 15, pp. 67-79, Sep. 2015.

[9] K. M. He, X. Y. Zhang, S. Q. Ren, *et al*., *Deep Residual Learning for Image Recognition*, CVPR, 2015.

[10] X. Li, L. Ding, L. Wang and F. Cao, "FPGA accelerates deep residual learning for image recognition," in *Proc. IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conf.*, Chengdu, 2017, pp. 837-840.

[11] Sengupta, Abhronil, Y. Yuting, W. Robert, *et al*., "Going deeper in spiking neural networks: VGG and residual architectures," *Frontiers in Neuroscience*, vol. 13, no. 2, pp. 95-96, 2019.

[12] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345-1359, Oct. 2010.

[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, 2016, pp. 779-788.

[14] R. Pérez, F. Schubert, R. Rasshofer, *et al*., "A machine learning joint lidar and radar classification system in urban automotive scenarios," *Advances in Radio Science*, vol. 17, pp. 129-136, Sep. 2019.

[15] J. Wu and Y. Wang, "Complexity and accuracy analysis of common artificial neural networks on pedestrian detection," in *Proc. MATEC Web of Conferences*, 2018, pp. 232-236.

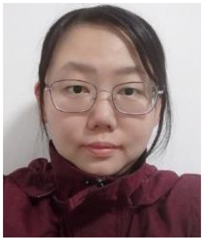
[16] A Yrapetov, D. Mitriy, and B. Buyanov, "Analysis of work of YOLO v.3 AND YOLO v.2 neural networks," presented at International Conference Technology & Entrepreneurship in Digital Society, 2019.

[17] M. J. Shafiee, B. Chywl, F. Li, and A. Wong, "Fast YOLO: A fast you only look once system for real-time embedded object detection in video," ArXiv abs/1709.05943, Sep. 2017.

[18] J. Redmon and A. Farhadi. "YOLOv3: An incremental improvement," ArXiv, abs/1804.02767, Apr. 2018.

Tong Lin was born in 1998, received the bachelor's degree from the School of Air Force Early Warning Academy, Wuhan, in 2020. His research interests include radar signal and data processing, deep learning, and artificial intelligence.

Xin Chen was born in 1981, received the Ph.D. degree from National University of Defense Technology, Changsha, China. He is currently an associate professor in Air Force Early Warning Academy. His major research interests include big data and deep learning.



Xiao Tang is currently a lecturer in Air Force Early Warning Academy. Her major research interests include Artificial Intelligence, Data Analysis

Ling He is a lecturer in Air Force Early Warning Academy. Her major research interests include Artificial Intelligence, Data Analysis

Song He is a lecturer in Air Force Early Warning Academy. His major research interests include Deep Learning

Qiaolin Hu is a lecturer in Air Force Early Warning Academy. His major research interests include IOT, network virtualization, Big data.