Automatic Target Recognition of Millimeter-Wave Radar Based on Deep Learning

Chen Xi Wang\textsuperscript{1}, Xin Chen \textsuperscript{1*}, Han Yu Zou \textsuperscript{1}, Song He\textsuperscript{1} and Xiao Tang\textsuperscript{1}

\textsuperscript{1} Air Force Early Warning Academy, Wu Han, Hu Bei, 430014, China
\textsuperscript{*}Corresponding author's e-mail: chenxinaihappy@dingtalk.com

Abstract. All-day, all-weather wide-area search discovery target capability makes radar become a key piece of equipment in many military and civilian fields, and plays an indispensable role in tasks such as identification and positioning. However, at the same time, the current radar equipment generally relies on complex signal processing systems, and designing artificial feature extraction algorithms based on prior knowledge is both difficult and time-consuming, and it is difficult to fine-grained identify the target. In recent years, deep learning has been widely used in the radar field with its strong adaptability and strong self-learning ability. The paper presents a radar Doppler image based on YOLO v3 algorithm. With Millimeter-Wave radar distance-doppler reflectance image as input and darknet-53 as the feature extraction network, the automatic detection of vehicle millimeter wave radar cars, bicycles, pedestrians and trucks was realized, and the average detection accuracy reached 84.3\%, providing new ideas and technical support for the development of intelligent radar and radar target detection.

1. Introduction
Radar automatic target recognition technology (ATR) is the automatic identification of radar and its auxiliary equipment for the attributes of detected number and categories. The development of automatic target identification technology has an important role in promoting the development of radar technology. The current radar automatic target identification technology method can be divided into three levels. The first level is to judge the number of targets. The second level is the coarse classification identification of the object's category attributes, size, etc. The third level is to identify the specific model of the target, and its micro-Doppler characteristics.

The millimeter-wave radar is the radar that works in the millimeter-wave band. Millimeter-wave radar is classified according to the way of radiation and electromagnetic waves, which can be divided into pulse Doppler system radar and continuous wave system radar. Because the pulse system radar separates receiving and transmitting on the time scale, it can integrate sending and receiving. \cite{1} The radar system is becoming more and more miniaturization and integrated in the development, so the radar rapidly develops from the foundation to on-board, airborne, carrier, satellite and other platforms. However, in close detection, continuous wave radar relies on its ease of use, low economic cost and high distance resolution, which has also been applied in specific fields. Due to the characteristics of on-board radar mission and environment, pulse radar on-vehicle millimeter wave radar is mainly continuous wave system radar. This paper focuses on the automatic target identification technology of vehicle millimeter wave radar. The application scene of on-board radar is generally unmanned, so the detection performance of radar on road targets is the focus. At present, the mainstream data processing method is mainly based on the target trace and conducts data statistics according to the prior target
movement characteristics. If the alarm rules meet, the alarm is triggered. This method requires the signal processing system with strong real-time performance and limited feature acquisition ability, so it is difficult to achieve three levels of automatic objective recognition.

Currently, the most important problem in radar target identification is the selection of target characteristics, correct, effective features can greatly reduce the difficulty of identification, however, how to obtain high recognition target feature data from a finite data sample is a difficult problem. Therefore, the significance of introducing deep learning is that through a strong adaptive learning ability of deep learning technology, we can find a feature extraction model that can be widely used in various radar targets and scenarios, and combine feature extraction with target recognition through a multi-layer network to obtain a complete learning classification model. Aiming at the automatic target recognition of millimeter-wave radar, YOLO v3 [1] is proposed to identify radar distance Doppler image by spectrum map. Introduce the advantages of deep learning parameter sharing and strong adaptability, and realize the detection and identification of pedestrians, vehicles, trucks and bicycles.

2. Radar spectrum map recognition framework based on YOLO algorithm

2.1. Method and overall architecture

The Doppler image recognition task for the millimeter-wave radar is to detect the position of the target on the radar reflectivity Distance-Doppler spectral image and determine the specific model. The proposed method of deep learning can be divided into training and test stages. The overall architecture of the system method is shown in Figure 1.

![Figure 1 System Method Diagram](image)

The training stage is on the basis of the data processing method mentioned in the second part, divide the standard data set into training set, verification set and test set and corresponding annotation files, obtain input data applicable to the neural network, then input the image into the YOLO network model, through feature extraction, grid generation prediction box, probability prediction, non-maximum suppression, and filter the location coordinates and type of the final output target.

2.2. Network models

YOLO v3[1], a single-stage object detection network proposed in 2018, is improved based on YOLO v1[1] and YOLO v2[2], which mainly solves the problem of poor detection of small targets, adding the characteristic pyramid structure into the network structure to achieve multi-scale prediction. The network structure of YOLO v3 is as shown:
Yolov3 target detection algorithm can be divided into 4 steps.

2.2.1. Extract image features using a feature extraction network
In this paper, the radar Doppler spectrum image of size $416 \times 156$ is extracted using the Darknet-53 network. Darknet-53 is improved by Darknet-9, which is similar to the ResNet, adds fast connection after the convolutional structure, and uses a convolutional layer replacing the pooling layer of 2, and the radar Doppler map spectrum output the feature map of $13 \times 13$ after feature extraction.

2.2.2. Generate Grids
According to the size of the network output feature diagram, such as the feature diagram of $13 \times 13$ will be divided into $13 \times 13$ grids, each grid corresponding to the region of the image $32 \times 32$ pixels, and which grid is determined by the location of the center of the target calibration box.

2.2.3. Generate the prediction box and the category probability
In the YOLO v3 network, each grid generates three prediction boxes, and the algorithm takes the detection box with the area of the largest overlap with the target dimension box as the target, and each prediction box contains $x, y, w, h, \text{confidence}$ 5 prediction values. $x, y$ refers to the coordinate offset value of the grid relative to the center of the current grid prediction box. $w$ and $h$ are the wide height of the prediction box after normalization, and confidence is the confidence of the current prediction box, while each prediction box predicts the $C$ category probability, i.e. the probability for which category the target belongs.

2.2.4. Filter the forecast box
After the final prediction is completed, all the prediction boxes are sorted by confidence, and the final target detection results are screened using the non-maximum suppression (NMS) algorithm.

2.3. Feature Pyramid Network
Feature Pyramid Network (FPN)[3] is a pyramid network structure proposed in 2018, mainly used for feature extraction. Different layers of the pyramid can be used for different detection of target scales, to be detected for simple targets with only low-level features, and to detect for complex targets. The low-level features contain less semantic information, but provide more accurate location information, while the high-level features instead bias the high-level features through multiple drop sampling and up sampling operations. FPN will superposition the processed low-level features and high-level features, integrate multi-layer feature information, and effectively construct multi-scale feature maps. YOLO v3 adds the feature pyramid in the network, enhancing the feature expression ability,
effectively improves the detection performance of small targets, and makes the model more robustness.

2.4. Loss function
The centre point loss mainly includes the horizontal and longitudinal coordinate loss, and both adopt the binary cross-entropy loss function, with the centre point loss such as the formula (1):

\[ L_{xy} = -\frac{1}{N} \sum_{\lambda \in \{x,y\}} \sum_{i=1}^{N} \lambda^{(i)} \log \hat{\lambda}^{(i)} + (1 - \lambda^{(i)}) \log (1 - \hat{\lambda}^{(i)}) \]  

(1)

In formula (1), \( \lambda^{(i)} \) represents the center point coordinate of article i data and \( \hat{\lambda}^{(i)} \) represents the centre point prediction coordinate of article i data.

Square loss function for width and height loss is calculated as formula (2):

\[ L_{wh} = \frac{1}{2} \sum_{\lambda \in \{w,h\}} \sum_{i=1}^{N} (\mu^{(i)} - \hat{\mu}^{(i)})^2 \]  

(2)

In formula (2), \( \mu^{(i)} \) represents the width height of the i prediction box and the \( \hat{\mu}^{(i)} \) represents relative width height of the i prediction box.

Prediction box confidence loss adopts a binary cross-entropy loss function, such as formula (3):

\[ L_{confi} = -\frac{1}{N} \sum_{i=1}^{N} \theta^{(i)} \log \hat{\theta}^{(i)} + (1 - \theta^{(i)}) \log (1 - \hat{\theta}^{(i)}) \]  

(3)

In formula (3), \( \theta^{(i)} \) represents the true label of the i prediction box indicates, \( \hat{\theta}^{(i)} \) represents the confidence of the i prediction box.

Category confidence loss adopts a binary cross-entropy loss function, such as formula (4):

\[ L_{class} = -\frac{1}{N} \sum_{i=1}^{N} \omega^{(i)} \log \hat{\omega}^{(i)} + (1 - \omega^{(i)}) \log (1 - \hat{\omega}^{(i)}) \]  

(4)

In formula (4), \( \omega^{(i)} \) represents the category label of the i prediction box, \( \hat{\omega}^{(i)} \) represents the prediction category of the i prediction box.

Therefore, the loss function of YOLO v3 is shown in formula (5):

\[ Loss = L_{xy} + L_{wh} + L_{confi} + L_{class} \]  

(5)

In formula (5), \( L_{xy} \) represents the central point loss, \( L_{wh} \) represents the wide loss, \( L_{confi} \) represents the prediction box confidence loss, \( L_{class} \) represents the category confidence loss.[1]

3. Experiment and Analysis

3.1. Data Preparation
This article uses the data used in [4]. Radar spectrum data were measured in urban scenarios around the Munich University of Technology using the 77GHz FM continuous wave radar. Collect data from various targets through cars driving in a real urban environment. With the assistance of a lidar and front-facing camera associated with 77GHz FM continuous wave radar, three data of pedestrians, bicycles and cars were collected as slices of radar distance-Doppler spectrum. Use the Laser Radar to semi-automatically mark the target presence areas and make manual correction if necessary. Spectrum map data is divided into four categories: pedestrians, cars, bicycles and noise (no target), among which automobile is divided into two categories: cars and trucks. Each category of data has many targets, marking each target as the target ID. Radar collected data slices ranging from frames to dozens of frames for each target. Therefore, to facilitate tracking model training, each picture is named targetID_Frame, where targetID represents the target serial number, and Frame indicates that the picture is the few frames of the target. [4]
The dataset contains 7757 images, including 3636 car, 975 pedestrians, 551 truck, and bicycles 2595. The initial resolution of the spectral graph image is uniformly processed as a three-channel jpg format image with a width of 496 pixels and a height of 156 pixels. The data distribution diagram is shown in Figure 4:

3.2. Experimental environment configuration
The experimental machine is configured as the Intel (R) Core (TM) i7-8700K CPU, one-card NVIDIA GeForce GTX 1080ti GPU and 32G RAM, CUDA10.1, cudnn7.6.1, networks based on the Python3.7.3, PyTorch1.7.1 deep learning framework. The network uses the method of using migration learning on a pretrained model on the MS COCO dataset and then fine-tuning.

3.3. Experimental Results
Multi-classes of Target Average Precision (AP) and mean Average Precision (mAP) are used to evaluate the performance of the radar Doppler image recognition model. For target detection problems, the samples can be divided into TP according to the detection results, indicating the correct number of detected samples; negative samples (FP) indicate the number of wrong detected samples; and FN, indicates the number of missed samples. As shown by the formula 6,7:

\[
Recall = \frac{TP}{TP + FN}
\]
\[ \text{Precision} = \frac{TP}{TP + FP} \quad (7) \]

In formula Precision rate represents the accuracy and Recall rate the recall. AP measures the average accuracy of a category, as shown in formula 8:

\[ AP(C) = \int \text{Precision rate}(c) \text{dRecall rate}(c) \quad (8) \]

mAP measures the average accuracy of all categories of detection across the entire dataset, as shown in formula 9:

\[ \text{mAP} = \frac{1}{c} \sum_{c \in C} AP(c) \quad (9) \]

3.4. Experimental result

The target detection model based on the YOLO v3 algorithm after 30 rounds of training can recall of 91.7%, detection precision of 86.2%, and 84.3% for the mAP of four targets at intersection ratio (IOU) of 0.5.

The model recall curve and precision curve are as shown Figure 5:

![Figure 5 recall curve and precision curve](image)

The model mAP curve is as shown Figure 6:
4. Conclusion
By introducing deep learning technology, this paper provides an effective idea of the long-term difficult problem of radar automatic target recognition from the perspective of radar spectrum map, puts forward a radar object detection method based on Doppler features and deep learning, and realizes the automatic radar Doppler image detection model based on YOLOv3. Compared to the traditional recognition methods, the deep learning-based radar Doppler image recognition algorithm not only improves the robustness of the target recognition system, but also reduces the human participation, automatically learns the target features, and gives the identification conclusions. Both the detection speed and accuracy have certain advantages, which is of certain practical significance to the automatic entry of radar targets, especially the development of intelligent unmanned radar equipment. In the future research, the automatic target recognition technology of radar should pay more attention to the integration of information. The radar Doppler spectrum map recognition can be used as a branch of information fusion, and use the detection results for auxiliary discrimination.

References
[1] Xin, B.. (2020). Millimeter wave radar technology. Redmon, J., & Farhadi, A.. (2018). Yolov3: an incremental improvement. arXiv e-prints. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A.. (2016). You only look once: unified, real-time object detection. IEEE.
[2] Redmon, J., & Farhadi, A.. (2017). YOLO9000: Better, Faster, Stronger. IEEE Conference on Computer Vision & Pattern Recognition (pp.6517-6525). IEEE.
[3] Lin, T. Y., Dollar, P., Girshick, R., He, K., Hariharan, B., & Belongie, S.. (2017). Feature pyramid networks for object detection. IEEE Computer Society.
[4] R Pérez, et al. (2019) A machine learning joint lidar and radar classification system in urban automotive scenarios. Advances in Radio Science,17:129-136.
[5] J Liu, & Wang, X.. (2021). Plant diseases and pests detection based on deep learning: a review. Plant Methods, 17(1).
[6] He, K., Zhang, X., Ren, S., & Sun, J.. (2016). Deep residual learning for image recognition. IEEE. pp. 770-778