Military Target Detection Method Based on Improved YOLOv3 Network

Yu Bowen, Zhang Jie

Abstract— **Improving the performance and accuracy of image target detection technology is an effective means to improve the generation and analysis ability of battlefield situational awareness. A training data set is constructed for complex battlefield environment, which contains relevant data conforming to the conditions of small targets, occlusion and relatively dense, etc., and can provide a test environment for various target detection algorithms. A military target detection method based on convolutional neural network RDBN-YOLOv3 algorithm is proposed to improve the efficiency and accuracy of military target detection in complex environments. Based on the characteristics of residual network and dense connection network, the residual dense connection structure is proposed and RDBN-YOLOv3 network structure is designed. The dense residual connection structure improves the fusion and reuse capability of the original YOLOv3 network for feature information at all levels. By combining local residual learning, global residual learning and global feature fusion strategy, the transmission of image feature information is optimized, and the detection performance of small targets, occlusion and relatively dense military targets is improved. Finally, experiments are carried out in the data set constructed in this paper. The experimental results show that compared with the original YOLOv3 algorithm, the average accuracy is 4.82% higher, which can provide effective technical support for battlefield situation generation and analysis.**

***Index Terms*—target detection, deep learning, YOLOv3, complex scenes, situational awareness, residual network, dense connection network**

I. INTRODUCTION

Battlefield situational awareness is a process of real-time perception of troop deployment, weapons and equipment and battlefield environment (terrain, meteorology, hydrology, etc.) of combat troops and supporting troops, including traditional reconnaissance, surveillance, intelligence, navigation beacon indication, damage assessment, information interception and information resource management and control. In the present stage of multi-arms joint operation, battlefield situation is used to analyze and predict the combat form, and then control the whole battle process, and finally obtain the battle advantage of the whole war. In the future information warfare, improving battlefield situational awareness can effectively improve the overall control ability of war, and all military powers are strengthening the development and research of relevant technologies [1]. At present, the key technology affecting the battlefield situational awareness system is the recognition and positioning of military targets [2].

Deep learning is a multi-layer neural network model, a nonlinear expression, multilayer characteristics extracted the characteristics of learning and independent, in recent years, scholars have gradually will be based on the deep learning algorithm is applied to battlefield situational awareness [3, 4], can well solve the complex battlefield in the future battlefield generating and analyzing problems. How to improve the timeliness and accuracy of military target detection is of great significance to the generation and analysis of battlefield situation. Therefore, the application of target detection model based on deep learning to battlefield military target identification and location can provide effective technical support for battlefield situation awareness system [5]. Bao Zhuangzhuang et al. designed a Deformable-ScratchNet detection network model based on ResNet34 network structure for small target features, and enhanced the network's fusion ability of deep features and shallow features through deconvolution, thus achieving high accuracy in small target detection tasks [6]. Dai Wenjun et al. proposed a multi-scale deformation target detection method, which improved the detection ability of deformation and multi-scale targets through resnet-101-Deformable network based on deformable convolution and deformable ROI pooling [7]. Du Zexing et al. adopted the dense connection network structure to ensure the effectiveness of the features of the shallow network, and fused the shallow features and deep features through the expansion block with large receptive field and deconvolution network structure, which greatly reduced the network reasoning time and enhanced the detection ability of multi-scale targets [8]. Yang Chuandong et al. solved the problem of poor detection accuracy of missile-borne image targets in bad weather, target deformation, precision and speed tradeoff and other problems through data enhancement, scale transformation and other processing methods [9]. Aiming at the problem that complex environment has a great influence on static target detection, Wang Zhi et al. proposed a method of region combination for end-to-end alternating training, which has high accuracy and robustness for single and multiple target detection in complex environment [10].

As deep learning has made breakthroughs in the field of target detection, domestic and foreign scholars have proposed a variety of deep learning-based target detection methods. At present, there are two main research directions for deep learning-based target detection: Multi-stage target detection algorithm (RCNN series [11-13]). This kind of method firstly generates candidate Region Proposal according to image information, then extracts features through convolution operation, and finally classifies data samples based on extracted feature information. Single-stage target detection algorithm (YOLO series [14-16]). This kind of method deals with target detection as a regression problem and directly outputs the positions of all objects in sample images and the

Yu Bowen, Department of Automation, Nanjing University of Science and Technology, Nanjing, China..

confidence of their categories. Compared with the latter, the former has a certain advantage in target detection accuracy, but is slightly inferior to the latter in operational efficiency. Scholars usually deepen the depth of the network to improve the detection accuracy of YOLO series algorithms, and the increase of the network depth increases the computational complexity, and at the same time, some shallow image features will be lost, and the problems of network degradation and gradient disappearance will occur [17]. In the process of feature extraction, the original feature information is directly transmitted to the backbone Network layer through identity mapping to ensure that the accumulation layer can learn new features on the basis of basic features, thus improving the training speed and accuracy of the Network. At the same time, the problem of degradation caused by network hierarchy is effectively solved [18]. The Dense Network (DenseNet) improves the feature fusion ability of the whole Network by establishing dense connections between the shallow Network and the deep Network, which makes the deep features include the shallow features, strengthens the feature reuse ability of the network, and improves the accuracy of the network model for sample image classification. The effect of network depth on gradient disappearance is eliminated and the learning speed is accelerated [19].

Because battlefield situation is highly dynamic, the real-time detection of military targets is highly required. In addition, the battlefield image information obtained through the perception system has a wide description range, and military targets occupy fewer pixels in the image, resulting in local occlusion, which makes the military target detection method inefficient in complex environment. Therefore, how to quickly and accurately identify enemy military targets in complex environment is an important factor affecting battlefield situational awareness. Based on this, this paper will start from the basic problem of battlefield situation awareness, positioning and identification of military targets, detect military targets through deep learning, and propose a military target identification method based on RDBN-YOLOv3 target detection network. In order to adapt to complex combat scenarios, this paper improved the feature extraction Network structure of YOLOv3, combined with the construction idea of Residual Network and Dense Network, proposed RDBN-YOLOv3 Network structure, which enhanced the representation capability of the Network. Make full use of all the characteristics of the network layer information, to ensure the continued effectiveness of shallow characteristics, strengthen the network characteristics of the network level fusion, reuse, and greatly reduce the number of parameters, and thus improve the testing efficiency and accuracy of military targets, generated and analysis provide effective for battlefield assistive technology support.

II. Method

A. Production of image data set

In the real combat environment, military action units usually use military camouflage or camouflage similar to complex natural environment to effectively reduce the probability of detection by various detection methods, thus greatly improving the battlefield survivability of military action units. The military target detection training data set is constructed according to common military action units in the combat environment of the land battlefield, which can be roughly divided into ground mobile units, tanks, infantry vehicles, self-propelled artillery, etc., with high mobility and high damage, they are important detection targets in the ground battlefield. Ground combat personnel, ground combat troops is a non-negligible part of the land battlefield, its target is small, the probability of detection is low. Low-altitude flying unit, uav, helicopter and other low-altitude flying units, which mainly play low-altitude reconnaissance, combat guidance, direct attack and other functions in the battlefield, is a direct threat to the ground troops in the land battlefield. Missile force combat unit, with greater damage, but poor mobility, mainly used to attack important military targets.

In order to test the usability of the model in real environment, a large number of military target images with complex backgrounds were downloaded from the Internet to construct training data sets. The data set was disorganized and divided into training set, test set and verification set in a ratio of 8:1:1. Then, LabelImg, an image labeling software, was used to mark the sample data in the dataset, and the label format of the dataset was consistent with Pascal VOC data label format [20]. The classification details of the dataset are shown in Table 1, and some image examples of the dataset created in this paper are shown in Fig 1. At present, the sample data of this dataset is not perfect enough, and it will be extended in the future to enrich the sample data.

Tab. 1 Details of training data sets military target

Military target	Number of military targets	Number of images
tank	9056	6887
Helicopter	2677	1135
Soldier	4854	2735
UAV	2389	1145
Missile	1563	1445
BFV	2066	845
automobile	3435	2675
Sum	26040	16867

Fig. 1 Part of the data set image

B. YOLOv3

Relative to the multi-stage target identification method (R - CNN series), YOLOv3 converts target recognition task goal bounding box and categories to predict regression problems, will be the target area and target category prediction is integrated into a single neural network model, a direct prediction probability of the boundary of the target and categories, to achieve the goal of recognition of the end-to-end testing, The speed and precision of target recognition are guaranteed, which makes it more suitable for application in practical application environment.

Fig. 2 shows an input-output diagram of YOLOv3. According to the resolution of the input image, the image is segmented into $S \times S$ ($S \in$ [13,26,52]) grid. If the center of the detected target falls in a grid, the grid unit detects the target. Bounding boxes are output by each grid and the probability of target category C, where each Bounding box information contains: the center coordinate of Bounding box (x, y), the length and width of Bounding box (W, H), and confidence (indicating the probability of predicting target category and the accuracy of target at that position). Where, the target category C is determined by $P_r(Class \mid Object)$, as shown in Equation (1) :

$$P_r(Class) = P_r(Class \mid Object) \times P_r(Object) \times IOU_{pred}^{truth} = P_r(Class) \times IOU_{pred}^{truth} \quad (1)$$

Finally, boundary Windows with low confidence probability were removed according to the preset threshold, and redundant boundary boxes were removed by non-maximum Suppression method. Finally, target recognition results were output.

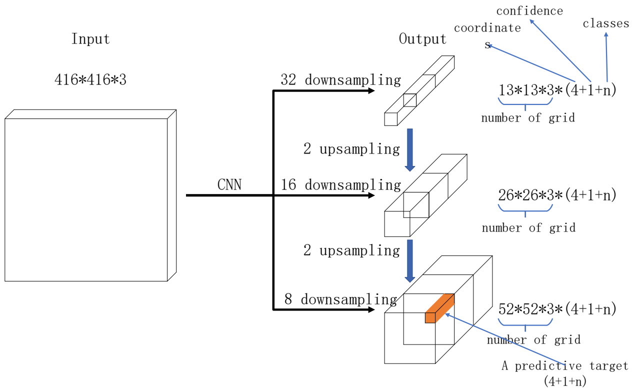

Fig. 2 YOLOv3 Input-Output (regardless of network structure)

C. Improved YOLOv3 network

On VOC2007 and other data sets, YOLOv3 has better performance compared with other target detection algorithms (SSD, R-CNN series, YOLO). However, when the target has small or multi-target local occlusion, there is a possibility of missing detection in the target detection process. In essence, YOLOv3 extracts target image features through convolution operation, and DarkNet53 is used as the basic feature extraction network, which has excellent performance but requires a large amount of computing resources. In view of the above, based on the construction idea of Residual Network (ResNet) and Dense Network (DenseNet), this paper improved the basic feature extraction Network of YOLOv3 and proposed RDBN-YOLOv3 target detection algorithm. Further improve the efficiency and accuracy of military target detection, provide auxiliary support for other battlefield situation awareness systems, and improve the timeliness and accuracy of battlefield situation generation.

1) Residual network

As neural network hidden layer increases, the error of the training set to reduce gradually, the precision of the network gradually become saturated state, when the network layer to increase the number of training error may hold increase, then a sharp degradation, residual network by joining identity mapping, guarantees the accumulation on the basis of the basic characteristics can learn new features, The emergence of residual network effectively solves the degradation problem of neural network with the increase of layers [18]. Fig 3 shows the residual block of the basic module of residual network.

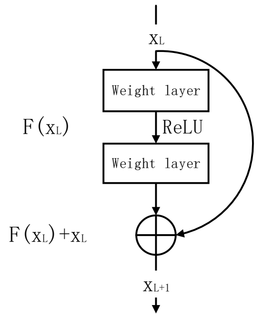

Fig. 3 Residual block

The residual block in Fig. 3 can be expressed as:

$$x_{L+1} = f\left(x_L + F\left(x_L, W_L\right)\right) \quad (2)$$

x_L , x_{L+1} is the input and output of the first residual block, $F(x_L, W_L)$ represent the residual, and f is the activation function of ReLU[21]. When the network parameters reach the optimal performance, the L th residual block is returned to 0, and x_L can continue to pass down through identity mapping, so as to ensure that the network is in the optimal performance state, and is not affected by the degradation caused by the increase in the number of hidden layers. However, residual network is accompanied by the problem of reduced feature reuse in practical application [17].

2) Densely connected network

Traditional convolutional networks have L layers, with one connection between each layer and its subsequent layers, while DenseNet has L(L+1)/2 direct connections. For each layer, the image features of all the previous layers are used as the input, while its own image features are used as the input of all the subsequent layers, so that the high-level feature information contains the low-level feature information, which further strengthens the fusion and reuse of feature information. DenseNet effectively alleviates the problem of gradient disappearance, strengthens the propagation and reuse of features, and greatly reduces the number of parameters [19]. As shown in FIG. 4, it is an example of a 5-layer dense block, whose basic structure includes: dense connection block and conversion layer. The L layer receives the feature information of all the previous layers, and the output of the L layer is shown in Equation (3) :

$$x_l = H_l\left(\left[x_0, x_1, \ldots, x_{l-1}\right]\right) \quad (3)$$

In equation (3), $\left[x_0, x_1, \ldots, x_{l-1}\right]$ represents the characteristic information of $0, \ldots, l-1$ layer. Where H_l is a composite function, including BN[22], ReLU and 3×3 Conv.

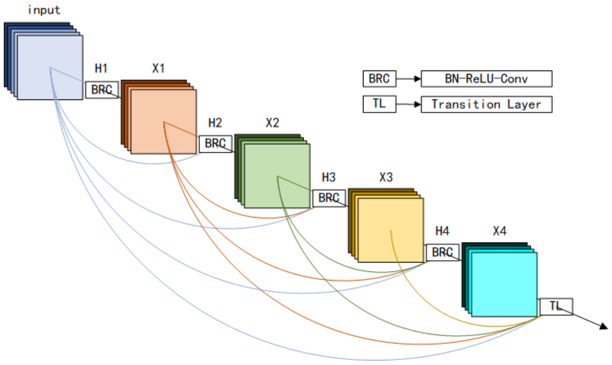

Fig. 4 5 layers Dense block

3) RDBN-YOLOv3 network structure

The analysis of situation elements (discovery, category and location of military targets) is the basis and key of battlefield situation generation. By improving the timeliness and accuracy of target identification and location in complex scenes, the situation generation and analysis ability of battlefield situation awareness system can be effectively improved. Considering the high requirement of timeness and accuracy of battlefield target detection, this paper improved the feature extraction network of YOLOv3 based on YOLOv3 network structure and combined with the construction idea of ResNet network and DenseNet network. After several experiments, the RDBN-YOLOv3 target detection network was proposed. The network parameters of feature extraction are shown in Tab. 2.

Tab. 2 Parameters of RDBN-YOLOv3 Feature Extraction Network

Network layer	Kernel size	output
Conv	[3*3 Conv,Stride 1]*3	208*208
RDB (1)	[1*1 Conv,3*3 Conv] *6*2	208*208
Transition layer	1*1 Conv 3*3 Conv,Stride 2	208*208 104*104
RDB (2)	[1*1 Conv,3*3 Conv] *6*4	104*104
Transition layer	1*1 Conv 3*3 Conv,Stride 2	104*104 52*52
RDB (3)	[1*1 Conv,3*3 Conv] *6*8	52*52
Transition layer	1*1 Conv 3*3 Conv,Stride 2	**52*52** 26*26
Res (1)	[1*1 Conv,3*3 Conv] *8	**26*26**
	3*3 Conv,Stride 2	13*13
Res (2)	[1*1 Conv,3*3 Conv] *4	**13*13**

The RDBN-YOLOv3 network structure proposed in this paper is shown in Fig 5. Features are extracted through an RDBN network composed of a series of convolution, three groups of RDB*N modules and two groups of residual blocks. The basic module structure is shown in Fig 6. The improved network structure can effectively transfer the shallow image features to the deeper network layer through dense connection structure and residual structure, which improves the feature reuse and fusion ability of the network. At the same time, it is conducive to the combination of up-sampling and shallow features in the regression prediction of target detection, and improves the overall performance of the network.

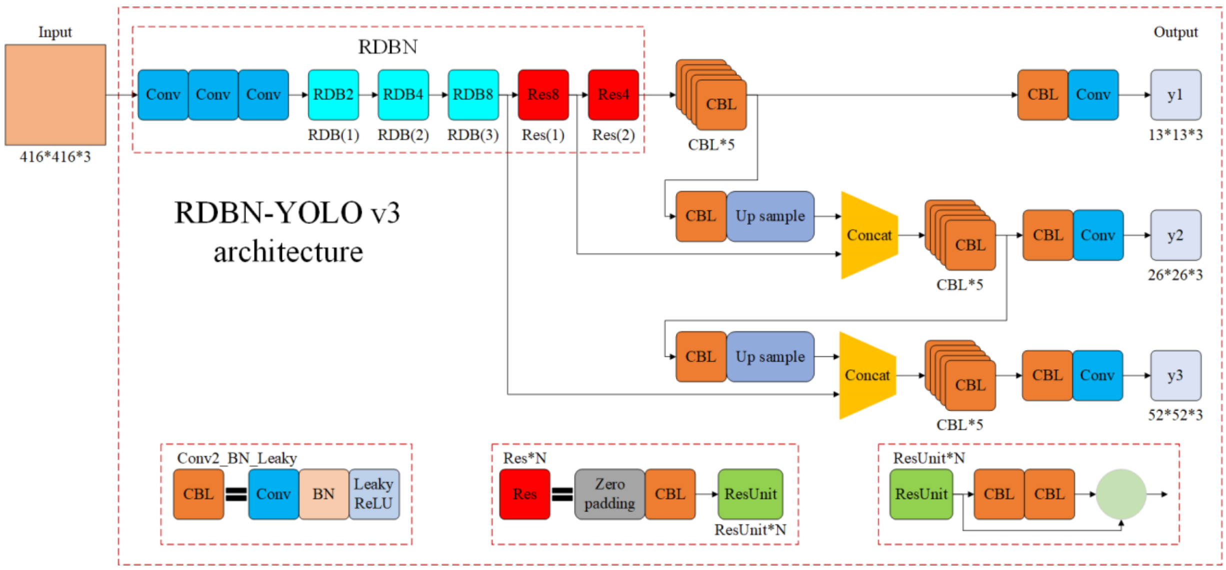

Fig. 5 RDBN-YOLOv3 network structure diagram

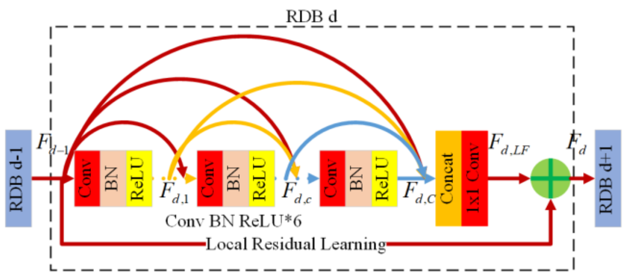

Fig. 6 Residual Dense block (RDB)

In Fig 5, the RDBN network mainly includes three parts: shallow feature extraction (SFE), residual dense block (RDB) and dense feature fusion (DFF). The main process of RDBN is shown in Fig 7.

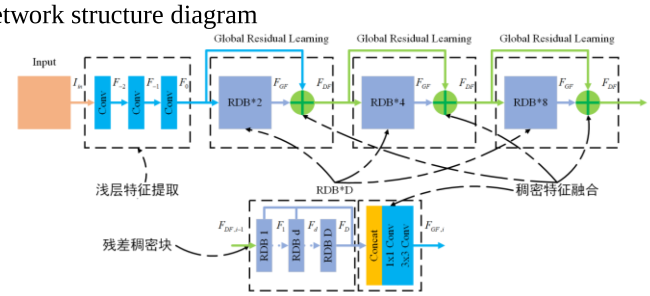

Fig. 7 The process of feature extraction module RDB*D

I_{in} as the input of RDBN, according to the conclusion of literature [23], using a large convolution kernel in the detection task may lose a large amount of image information, resulting in reduced detection effect, especially for small targets. Therefore, three 3*3 convolution layers connected in series are used to extract shallow features:

$$F_{-2} = H_{SFE1}\left(I_{in}\right) \tag{4}$$

$$F_{-1} = H_{SFE2}\left(F_{-2}\right) \tag{5}$$

$$F_{0} = H_{SFE3}\left(F_{-1}\right) \tag{6}$$

In Equation (4)-(6), H_{SFE1}, H_{SFE2}, H_{SFE3} represent convolution operation, and F_0 is the input of the dense residual block. It is assumed that group I RDB*D module has D dense residual blocks, among which the output of the dense residual block D is FFF, as shown in Equation (7) :

$$\begin{aligned} F_{d,i} &= H_{RDB,d,i}\left(F_{d-1,i}\right) \\ &= H_{RDB,d,i}\left(H_{RDB,d-1,i}\left(\cdots\left(H_{RDB,1,i}\left(F_{DF,i-1}\right)\right)\cdots\right)\right) \end{aligned} \tag{7}$$

In equation (7), $H_{RDB,d,i}$ represents the compound function operation in the first RDB, including Conv, BN and ReLU operations. $F_{d-1,i}$ is generated by D-1th RDB using each convolution layer in the module. $F_{d-1,i}$ is regarded as local feature (LF), and the feature graph of the D-1th RDB is splicing with that of the D-1th RDB through cascade, through which the number of features can be effectively reduced. According to the conclusion of literature [24], convolution is introduced to control the information output dimension adaptively. This operation is named as local feature fusion (LFF), as shown in Equation (8), where H_{LFF}^{d} represents the 1*1 convolution operation in the dth RDB.

$$F_{d,LF,i} = H_{LFF}^{d}\left(\left[F_{d-1,i}, F_{d,1,i}, \cdots, F_{d,c,i}, \cdots, F_{d,C,i}\right]\right) \tag{8}$$

Since there are multiple convolutional layers in one RDB, local residual learning (LRL) is introduced to further improve the information flow. The final output of the dth RDB is shown in Equation (9), and the representation ability of the network can be effectively improved through local residual learning.

$$F_{d,i} = F_{d-1,i} + F_{d,LF,i} \tag{9}$$

After extracting locally dense features using a set of RDB, we further perform dense feature fusion (DFF), which utilizes hierarchical features in a global manner, including global feature fusion (GFF) and global residual learning (GRL). Global feature fusion: generate global features $G_{GF,i}$ by fusing all features extracted by RDB in group I, as shown in Equation (10) :

$$F_{GF,i} = H_{GFF}\left(\left[F_{1,i}, \ldots, F_{D,i}\right]\right) \tag{10}$$

In Equation (10), $\left[F_{1,i}, \ldots, F_{D,i}\right]$ is the feature graph generated by cascade of all Dth RDB. According to the conclusion in reference [25], H_{GFF} function is set as 1*1 convolution operation and 3*3 convolution operation. Where, 1*1 convolution operation is used to fuse feature information of different levels, and 3*3 convolution operation is used to further extract feature information for global residual learning.

Finally, through global residual learning, the output of group I RDB*D module is:

$$F_{DF,i} = F_{DF,i-1} + F_{GF,i} \tag{11}$$

In Equation (11), $F_{DF,i-1}$ is the feature graph generated by the previous set of RDB*D modules, and the dense residual blocks can effectively make use of the features extracted from all the convolution layers in the modules to form $F_{GF,i}$ through further fusion operations. After global residual learning, the dense feature $F_{DF,i}$ is obtained.

In order to further enhance the feature extraction capability of RDBN, two residual blocks are used to deepen the depth of feature extraction network. The follow-up feature regression prediction part of detection network is consistent with the original YOLOv3 network structure. The pyramid structure similar to FPN network is used for two up-sampling, feature graph splicing operation, and target boundary box and category regression prediction for three times to achieve multi-scale target detection. RDBN-YOLOv3 can effectively improve the representation capability and learning efficiency of the network by introducing dense residual blocks.

1) Loss function

Loss function L_{S_loss} consists of bounding box loss function L_{S_iou}, confidence loss function L_{S_conf} and category loss function L_{S_cls}, as shown in Equation (12). Images are divided into S*S grids according to feature information:

$$L_{S\_loss} = L_{S\_iou} + L_{S\_conf} + L_{S\_cls} \tag{12}$$

Intersection over Union (IOU) in the regression prediction of boundary box, if there is no overlap between prediction box and marking box, the Loss function gradient will be 0. To solve this problem, this paper introduces the Distance IOU Loss[26] to perform the regression prediction of boundary box. It can effectively solve the problems of inaccurate regression and slow convergence, and its calculation equation is shown in Equation (13) :

$$L_{S\_iou} = 1 - \frac{\left|B_p \cap B_{gt}\right|}{\left|B_p \cup B_{gt}\right|} + \frac{\rho^2\left(b_p, b_{gt}\right)}{c^2} \tag{13}$$

In Equation (13), b_p and b_{gt} represent the center point of ground truth of prediction box $B_p = (x, y, w, h)$ and target box $B_{gt} = \left(x_{gt}, y_{gt}, w_{gt}, h_{gt}\right)$, $\rho(\square)$ represents Euclidean distance, and c represents the diagonal length that can cover the minimum rectangle of target box and prediction box.

In a part of the training sample is relatively easy to classify the sample data, thus makes the model of the optimization direction and we expected in the opposite direction, in order to reduce the effects of negative samples of model optimization, this paper introduced the Focal Loss [27] sample to reduce the negative impact on the model of optimizing the weights, to make it more focus on the harder to more complex classification samples, Its calculation equation is shown in Equation (14) :

$$\begin{aligned} L_{S\_conf} = &-\left(y_{gt} - y_p\right)^2 \times y_{gt} \log y_p \\ &-\left(y_{gt} - y_p\right)^2 \left(1 - y_{gt}\right) \log\left(1 - y_p\right) \end{aligned} \tag{14}$$

In Equation (14), y_p and y_{gt} respectively represent the predicted value and ground truth of the confidence degree of the detected target.

Classification loss is calculated by cross entropy loss, which is only valid for the grid with recognized objects. Its calculation equation is shown in Equation (16) :

$$BC\left(a,\hat{a}\right)=-\left[a\log a+(1-a)\log\left(1-\hat{a}\right)\right] \tag{15}$$

$$L_{S\_cls}=\sum_{i=0}^{S\times S} I_{ij}^{obj} \sum_{c\in classes} BC(p_i(c),\hat{p}_i(c)) \tag{16}$$

In order to ensure the identification accuracy of military targets in a complex environment, military targets are detected at multiple scales. Therefore, the final loss function is shown in Equation (17) :

$$L=L_{13\_loss}+L_{26\_loss}+L_{52\_loss} \tag{17}$$

III. ANALYSIS OF EXPERIMENTAL RESULTS

A. Experimental parameter configuration and detection results

In this paper, TensorFlow is used as the basic framework to build neural network, and training is carried out on the Ubuntu server. The specific configuration of the server is shown in Table 3:

Tab. 3 Details of the server configuration

Name	Details
System	Ubuntu18.04LTS
CPU	Intel 4114@10C*2.2GHz
GPU	GeForce RTX 2080ti
RAM	128G
Python	3.6.8
TensorFlow	2.0.0
OpenCV-Python	3.4
CUDA	10.2
CUDNN	7.6.5

According to the conclusion of literature [28], in order to improve the convergence of network and reduce the error of network training results, optimization techniques combining Adaptive moment Estimation (Adam) and Stochastic gradient Descent (SGD) are adopted in the training process. The shortcomings of slow SGD convergence and large Adam training error were effectively avoided. The training stage was divided into two stages. In the first stage (the first 60K iteration), Adam was used for optimization and in the second stage (the last 40K iteration), SGD was used for optimization.

Tab. 4 Details of the training parameter configuration

	Name	numerical value
	The number of iterations	100k
	Weight attenuation coefficient	0.0005
First stage	All parameters	The initial parameters in literature [28] are applied
Second stage	Initial learning rate	0.0001
	Momentum	0.9
	NMS	DIOU-NMS

In the process of training, more military target samples for network training can be generated through image rotation, scaling, contrast adjustment and other methods, so as to enhance the generalization of network and the accuracy and stability of military target recognition. According to the image samples in the training data set, k-means algorithm was used to cluster the prior frames to obtain 9 template sizes, as shown in Table 5.

Tab. 5 Size of anchor box

feature map	receptive field	Size of anchor box
13*13	Big	219*185；325*380；713*687
26*26	Mid	70*129；139*85；113*41
52*52	Little	23*26；43*55；77*56

The RDBN-YOLOv3 target detection network is applied to conduct target detection on the military target verification set collected in the complex environment in this paper, and the detection effect is shown in Fig 8.

Fig. 8 Detection results of RDBN-YOLOv3

B. Analysis of military target detection results

1) Target detection performance analysis

RDBN-YOLOv3, Faster RCNN, and YOLOv3 are used for target detection of the validation set in the dataset, and the detection results are shown in Table 6.

Tab. 6 Experimental results of the RDBN-YOLOv3 and Faster RCNN、YOLOv3

Military target	AP(%)		
	Faster RCNN	YOLOv3	RDBN-YOLOv3
tank	78.56	73.62	77.32
Helicopter	79.35	72.92	79.60
Soldier	83.11	78.14	82.54
UAV	70.12	62.83	70.66
Missile	77.33	74.60	77.86
BFV	80.15	75.28	78.91
automobile	83.32	77.92	83.12
Backbone	VGG	Darkness	RDBN
FPS	1.46	11.7	10.6
mAP (%)	78.03	73.62	78.44
Model size (MB)	148.7	248.6	218.4

As can be seen from Table 6, RDBN-YOLOV3 proposed in this paper is 4.82% higher than the average accuracy of YOLOv3, and the FPS is decreased by 1.1 frames per second. The average accuracy of RDBN-YOLOV3 is 0.41% higher than that of Faster RCNN, but the speed of RDBN-YOLOV3 is 7.6 times that of Faster RCNN, which can meet the application requirements of real battlefield. The above experimental results can prove that RDBN-YOLOv3 has good detection performance.

2) Model decomposition experiment

The RDBN-YOLOv3 network structure proposed in this paper uses strategies such as local feature fusion, local residual learning and global feature fusion, and adopts multi-scale training in the training process. In order to verify the contribution of each learning strategy and training method, decomposition experiments are carried out in the training data set constructed in this paper. During the experiment, it was found that the network model could not be properly trained if the local feature fusion strategy was missing. Therefore, the local feature fusion strategy was retained by default in the decomposition experiment. The usage and detection results of different strategies are shown in Table 7.

Tab. 7 Experimental results of RDBN-YOLOv3 model decomposition

various combination

Local feature fusion	√	√	√	√	√
Local residual learning	×	√	×	√	√
Global feature fusion	×	×	√	√	√
Multiscale training	×	×	×	×	√
mAP	66.45	71.45	72.67	76.62	78.44

As shown in Table 7, in the absence of local residual learning and global feature fusion strategy, mAP values are low and detection results are poor. On the one hand, the network model is difficult to train [18], and on the other hand, stacking dense blocks in a deep network structure [19] cannot effectively improve the overall network performance. When the local residual learning or global feature fusion strategy is added separately, it is verified that both strategies can effectively improve the detection performance of the network, mainly because they can effectively improve the overall information propagation of the network and improve the process of solving the gradient. Compared with the former case, the target detection accuracy is increased by 5% and 6.22% respectively. Furthermore, the local residual learning and global feature fusion strategy are applied simultaneously, and it can be seen that the algorithm achieves relatively good performance. This strategy further enhances the feature extraction ability, improves the representation ability of the network, and improves the ability of reusing and merging shallow and deep features. Compared with the previous network model using a separate strategy, its detection accuracy is improved by 5.17% and 3.95%, respectively. Finally, a scale is randomly selected for network training every 500 iterations. Combined with multi-scale training, the mAP of RDBN-YOLOv3 detection network is improved by 1.82%, which makes the detection network more robust and enables the detection network to process targets with different resolutions.

Since the Loss function in YOLOv3 has problems of slow convergence, large error and possible gradient of 0, DIOU Loss and Focal Loss are introduced in this paper to solve these problems. In order to verify the influence of the improved loss function on target detection, a comparative experiment is conducted, and the comparison results are shown in Table 8:

Tab. 8 Experimental results of RDBN-YOLOv3 applies different loss function

	Different combination results					
YOLOv3 loss baseline	√	√	√	×	√	×
DIOU loss	×	√	×	√	×	√
Focal loss	×	×	√	√	×	√
RDBN-YOLOv3	×	×	×	×	√	√
mAP	73.62	74.46	74.66	75.12	76.31	77.44

According to the experimental results in Table 8, the improved Loss function of YOLOv3 using DIOU Loss and Focal Loss is 1.5% higher than the original Loss function mAP. However, the improved loss function of RDBN-YOLOv3 proposed in this paper is 2.13% higher than the original loss function mAP. It can be seen that IOU Loss and confidence Loss in the original YOLOv3 Loss function are improved by DIOU Loss and Focal Loss, and the accuracy of military target detection by the detection network is effectively improved.

From what has been discussed above, we can see that RDBN - YOLOv3 network overall performance is better than that of YOLOv3 network, can effectively reduce the target is too small, dense, shade between the impact on the military target detection, by introducing a thick piece of residual, enhance the capacity of the network said, improved the network information flow, all levels of characteristics has made full use of, The feature information fusion and reuse ability of the network are strengthened, while the network parameters are reduced, thus improving the detection efficiency and accuracy of military targets.

IV. CONCLUSION

In this paper, the recognition and location of military targets, which are the key issues affecting the generation and analysis of battlefield situational awareness, are analyzed. The training data set is constructed for complex battlefield environment, and a military target detection method based on convolutional neural network RDBN-YOLOv3 is proposed. First, draw lessons from the characteristics of the residual and dense networks built residual dense connection block, by combining the local residual residuals, global learning strategy to optimize the YOLOv3 detection network, global features fusion of features at the level of information fusion, the reuse capacity, the influence of the external factors on the test results can effectively restrain, improve the detection performance of the detection network. Then, the Loss function of YOLOv3 network was improved based on DIOU Loss and Focal Loss. Combined with DIOU-NMS and multi-scale training, the detection accuracy and robustness of military targets in complex environment were further improved. Finally, the experimental results show that this method has high detection efficiency and accuracy, which can meet the actual combat requirements and provide effective auxiliary technical support for battlefield situation generation and analysis.

REFERENCES

[1] Miller A. Situational awareness – from the battlefield to the corporation[J]. Computer Fraud & Security, 2006, 2006(9): 13-16.

[2] Peng H, Zhang Y, Yang S, et al. Battlefield Image Situational Awareness Application Based on Deep Learning[J]. IEEE Intelligent Systems, 2020, 35(1): 36-43.

[3] Hanratty T P, R. J. Hammell I, Bodt B A, et al. Enhancing Battlefield Situational Awareness through Fuzzy-Based Value of Information[C]. 2013 46th Hawaii International Conference on System Sciences, 2013: 1402-1411.

[4] Liu S, Liu Z. Multi-Channel CNN-based Object Detection for Enhanced Situation Awareness. arXiv e-prints, 2017: arXiv:1712.00075.

[5] Dai W J, Chang T Q, Zhu K X, et al. Video Object Detection Method for Tank Fire Control System Based on Spatial-temporal Convolution Feature Memory Model [J/OL]. Acta Armamentarii:1-10[2020-07-05]. http://kns. cnki.net/kcms/detail/11.2176.TJ.20200618.1320.004.html.

[6] BAO Z Z, ZHAO X J, WANG M F, et al. Multi-Scale Target Detection Network Model Trained from Scratch[J].Computer Engineering,2020,46(06):248-255.(in Chinese)

[7] Dai W J, Chang T Q, Zhang L, et al. Multi-Scale Deformable Target Detection Method for Tank Fire Control System[J]. Journal of Computer-Aided Design & Computer Graphics,2019,31(12):2082-2090. (in Chinese)

[8] Du Z X, Yin J Y, Yang J. Remote Sensing Image Detection Based on Dense Connected Networks[J]. Laster & Optoelectronics Progress,2019,56(22):210-216. (in Chinese)

[9] Yang C D, Liu Z, Shi S B. Research on Object Detection for Missile-borne Image Based on CNN[J]. Tactical Missile Technology,2019(04):85-92. (in Chinese)

[10] Wang Z, Chen P, Pan J X. Object Detection Based on Deep Learning in Complex Background[J]. Journal of Chongqing University of Technology (Natural Science) ,2018,32(04):171-176. (in Chinese)

[11] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. IEEE

Conference on Computer Vision and Pattern Recognition (CVPR),2014:580-587.
[12] Girshick R. Fast R-CNN[C]. The IEEE International Conference on Computer Vision (ICCV).2015:1440-1448.
[13] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[C]. Advances in Neural Information Processing Systems 28,2015:91-99.
[14] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection. arXiv e-prints, 2015: arXiv:1506.02640.
[15] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger. arXiv e-prints, 2016: arXiv:1612.08242.
[16] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement. arXiv e-prints, 2018: arXiv:1804.02767.
[17] DAI J M, TONG J Z. Galaxy Morphology Classification Using Deep Residual Networks[J]. Progress in Astronomy,2018,36(04):384-397.(in Chinese)
[18] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) ,2016:770-778.
[19] Huang G, Liu Z, Van Der Maaten L, et al. Densely Connected Convolutional Networks[C]. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 4700-470.
[20] EVERINGHAM M, GOOL L V, WILLIAMS C K I, et al. The Pascal Visual Object Classes (VOC) Challenge [J]．International Journal of Computer Vision，2010，88(2):303-338.
[21] Glorot X, Bordes A, Bengio Y. Deep Sparse Rectifier Neural Networks[C]. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, 2011: 315--323.
[22] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[C]. Proceedings of the 32nd International Conference on Machine Learning, 2015: 448--456.
[23] Cui H , Wei Z. Multi-Scale Receptive Field Detection Network[J]. IEEE Access,2019,7(7):2169-3536.
[24] Tai Y, Yang J, Liu X, et al. MemNet: A Persistent Memory Network for Image Restoration[C]. The IEEE International Conference on Computer Vision (ICCV), 2017:4539-4547.
[25] Ledig C, Theis L, Huszár F, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 105-114.
[26] Zheng Z, Wang P, Liu W, et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. arXiv e-prints, 2019: arXiv:1911.08287.
[27] Lin T, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
[28] KESKAR N S, SOCHER R. Improving generalization performance by switching from Adam to SGD [J]. arXiv e-prints, 2017: arXiv: 1712.07628.

First Author Name: YU Bo-wen
Publications :
Communication Module of FC-AE-1553 Interface
Improved Apriori Algorithm Based on Matrix Reduction