Application and Implementation of CNN in Artillery Countermeasure Training System

Dong Chen¹, Chuandong Yang¹,², Hang Ji², Bin’an Jiang¹, Zhen Liu¹,*

¹Ammunition Technology Office, PLA Army Academy of Artillery and Air Defense, Hefei Anhui, China
²Postgraduate Brigade, PLA Army Academy of Artillery and Air Defense, Hefei Anhui, China

*Corresponding author e-mail: 1092850034@qq.com

Abstract. Blasting point recognition is the basic task in artillery countermeasure training system. In order to solve the inefficiency and inaccuracy caused by manual recognition, this paper proposes a blast point area recognition algorithm based on convolution neural network. Aiming at the relatively small area of blasting point, the design idea of transfer learning is adopted. Firstly, based on the Mask-RCNN model, the structure of feature pyramid network is modified. By comparing the recognition effects of different feature fusion methods, the optimal one is selected. The anchor parameters are adjusted according to the target characteristics as well. Finally combining the mask branch network and the target detection network, the contour of the blasting point area and the center of it are calculated. The experimental results show that the method proposed in this paper achieve 97.6% mAP on the blasting points test dataset we built at 4.7 FPS, which solved a basic problem in artillery countermeasure training system well.

1. Introduction

At present, virtual confrontation training still stays on the software interface level, lacking the real sense of confrontation [1]. In real combat training, ammunition can't really hit equipment on both sides of the drill, so it’s hard to embody the tactical effect of ammunition authentically, which leads to the inaccuracy of the evaluation of operational results and affects the effect of military training.

Aiming at the existing problems of countermeasure training, an artillery countermeasure training system based on augmented reality is constructed. The coordinates of actual targets are mapped to the real blasting point area by virtual reality technology. The guidance and coordination department can observe battlefield and evaluate attacking effect by using display and evaluation system. Among them, the recognition of blasting point as the primary task plays a fundamental role in the damage assessment in the follow-up confrontation. So it’s important to quickly and accurately capture the blasting point to the system. In this paper, a method of blasting point recognition based on CNN is proposed, which can obtain the detailed information of blasting points.

2. Countermeasure training system

The artillery countermeasure training system based on VR includes battlefield monitoring system, guidance and evaluation system and coordinate mapping system. The battlefield monitoring system...
captures the blasting point by using TV theodolite and unmanned rotorcraft, and calculates the blasting point positions. The guidance evaluation system is used to connect confrontation units in series, acquire confrontation data, formulate confrontation rules and evaluate confrontation effectiveness. The coordinate mapping system is used to construct digital maps, projecting coordinate information through augmented reality technology and obtain real-time images. Figure 1 shows the firing range division which contains 4 regions.

Figure 1. Recognition framework of Mask-RCNN.

3. Blasting point target recognition algorithm based on CNN
The explosive effects of ammunition usually appears in the form of fire, smoke, craters and other near-circular appearance, and the explosion process is very short. If blasting points are captured by military personnel, there would be many shortcomings, such as substantial energy consumption of trainers and different results due to different judgment criteria for different personnel. Therefore, automatic target detection is urgently necessary.

In recent years, target detection algorithms based on convolution neural network (CNN) are superior to traditional algorithms which quickly obtained the rapid and the extensive application in military field. In order to obtain the location of explosion point quickly and accurately, an instance segmentation method based on CNN is adopted. In view of the characteristics of the small-scale blasting point targets, the feature pyramid structure is used to fuse the features of different layers and the parameters of anchors are targeted designed. Finally, based on the segmentation result, the blasting point size and its mass center can be obtained.

3.1. Target recognition model
Mask-RCNN [2] is a target recognition model based on CNN which can output the target class, location and segmentation. It adopts Faster-RCNN’s [3] structure which use the classification branch and the bounding box regression branch to output class labels and bounding boxes’ offset. Besides, a third mask branch network is added to output binary mask for each Region of Interest(RoI), which realize the basic target recognition as well as instance segmentation. Weighing the speed and precision of target recognition, Resnet101+FPN [4] structure is selected as the basic feature extraction network.

3.1.1. Multi-scale feature fusion for small targets. Because the acquisition equipment is usually far from the explosion area, the target size of the explosion point is relatively small. In a multi-layer convolution neural network, low-level features can better represent image details such as textures, contours and edges.
After multiple convolution operation, the receptive field of neurons expands and the semantic information can be better represented by high-level features, but this will result in some loss of image details accordingly. For blasting point targets, high-level semantic information is relatively simple so there is no need for neurons to have such a large receptive field. Instead, we rely more on low-level features with image details to identify blasting point.

FPN structure builds a feature pyramid which make full use of low-level image information and integrate it with high-level information to achieve better detection results. The basic structure of FPN is shown in Figure 3. Ci is the feature activations output by stage i’s last residual block of ResNets [5] used as basic feature extraction network in our model. {C2, C3, C4, C5} are the feedforward outputs for conv2, conv3, conv4, and conv5 outputs and they have strides of {4, 8, 16, 32} pixels with respect to the input image. The upsampled feature map is then merged by element-wise addition with the corresponding bottom-up map which undergoes a 1x1 convolutional layer to reduce channel dimensions. The final set of feature map is called {P2, P3, P4, P5, P6}. Because of the small scale of the blasting point, through comparative experiments, the feature of P6 layer is removed. In RPN, we assign anchors with areas of {32², 64², 128², 256²} of a single scale to each level {P2, P3, P4, P5} respectively. And we use anchors of multiple aspect ratios {1:1.3, 1:1, 1.3:1} at each level. Compared with 15 anchor points of the original structure, only 12 anchor points are used in this structure, which can improve the speed and ensure the accuracy. In order to assign RoIs of different scales to the pyramid levels, we use formula (1) to calculate the level Pk of the feature pyramid corresponding to a RoI of width w and height h:

$$k = \left\lfloor k_0 + \log_2\left(\frac{\sqrt{wh}}{224}\right) \right\rfloor$$

Where k₀ is set to 4.

3.1.2. Mask branch network. The mask branch network is a fully convolutional network with few parameters and more accurate spatial information, whose structure is shown in Fig 2. The network generates a 28*28 mask for the blasting point target from each input RoI feature. First, four consecutive convolution operation is performed on the fixed-size feature map (14*14*256), then the deconvolution operation is used to increase the dimension of the feature map(28*28*256). And finally the 1*1 convolution is used and activated by sigmoid activation function. The function obtains a mask with a resolution of 28*28*C, where C is 1 due to there is only one type of target. And the value of each position is a float number ranged from zero to one, which will be binarized at a threshold of 0.5.
3.2. Model Training

The target recognition algorithm consists of two parts: model training and model testing. The model is pre-trained on the canonical COCO dataset, and then fine-tuned on the blasting point target dataset we built. After the model training is completed, the fine-tuned model is used for target recognition on test images, which output the predicted target type, position and masks of the targets in the image.

Model training method are described as follows. The input training samples are paired data and we use the pre-trained weight to initialize the parameters. We trained the RPN network and the Fast-RCNN module separately. Before the stop criterion is reached, we perform the following four steps:

Step1: Obtain anchors by the RPN.
Step2: Use non-maximum suppression algorithm (NMS) to select the RoI and use the RoIAlign layer to output accurate feature maps.
Step3: Calculate the model loss via Eq. (2).
Step4: Use the back propagation algorithm to update the model parameters.

4. Experiment and analysis

4.1. Dataset and implementation details

Various blasting points were photographed by TV theodolite and unmanned rotorcraft at different altitudes and distances in different scenarios such as grassland, desert and mud land, which form the blasting point dataset. The data augmentation method [7] is used which including transformation of brightness, saturation, hue and distortion. Finally about 3000 images are obtained, of which 2500 images are used as training set and 500 images are used as test set. All images are annotated by labelme toolkit to meet the COCO data format for training and validation.

During the training process, some parameters are set as follows: for the first 30k mini-batches the learning rate is set to 0.02 and 0.002 for the following 10k. A momentum of 0.9 and a weight decay of 0.0001 are used. A mini-batch involves 2 images. To reduce redundancy in RPN proposals, the IoU threshold for NMS is fixed at 0.7 and after NMS the top-N ranked proposals regions for recognition is
used. The network is implemented using pytorch and trained on a Dell precision tower 7810 workstation with an Intel Xeon E5-2683 V3 and an Nvidia GeForce 1080 GPU.

4.2. Comparison and result
In order to select the optimal pyramid feature fusion method for blasting point targets, the comparative experiments of different models were carried out. The results are shown in Table 1. mAP indicates the mean average accuracy of blasting points. Experiment 6 is equivalent to the Mask-RCNN structure using a single scale feature. The experimental results show that the recognition results with multi-scale feature fusion method are better than those of single-scale feature, and the best recognition results can be achieved by combining the feature information of P2-P5 features. Experiments 3-6 show that increasing low-level feature information can effectively improve the accuracy of prediction; Experiments 1-4 and 6 show that the accuracy of recognition results is higher with the increase of the number of fused features with different scales, but when the fused information is comprehensive enough, adding features of other scales will not have a significant impact on the detection results of blasting points. At the same time, the contrast experiment 7 shows that the detection model using only the low-level feature cannot achieve good results.

| Index | P2 | P3 | P4 | P5 | P6 | mAP |
|-------|----|----|----|----|----|-----|
| 1     | ✓  | ✓  | ✓  | ✓  | ✓  | 0.977|
| 2     | ✓  | ✓  | ✓  | ✓  | ✓  | 0.976|
| 3     | ✓  | ✓  | ✓  | ✓  | ✓  | 0.970|
| 4     | ✓  | ✓  | ✓  | ✓  | ✓  | 0.962|
| 5     | ✓  | ✓  | ✓  | ✓  | ✓  | 0.955|
| 6     | ✓  | ✓  | ✓  | ✓  | ✓  | 0.893|
| 7     | ✓  | ✓  | ✓  | ✓  | ✓  | 0.824|

Figure 3 shows the recognition results of blasting point using the model with P2-P5 level features. The white point in the blasting point area is the center of burst point which is calculated by a centroid calculation. It can be seen that for small-scale blasting points, the model can obtain the blasting point area by drawing the outlines and mark the location of its center. On the blasting point test dataset, the mAP reaches 97.6 and the recognition speed is 4.7 FPS in the experimental environment. It can get the exact position of blasting point accurately and relatively quickly, and meet the needs of the artillery countermeasure training system.

Figure 3. Recognition examples on blasting points test dataset.
5. Conclusion
Aiming at the difficulty of blasting point recognition in artillery countermeasure training system, a blasting point area segmentation method based on CNN is proposed. Based on the Mask-RCNN model, the structure and parameters of the network are modified and fine-tuned. Combining the mask branch network and the target detection network, the contour of the blasting point area and the center of it are calculated. At the same time, aiming at the problem of small scale of blasting point, the best recognition results on the blasting points test dataset we built can be achieved by fusing features from different layers of feature pyramid network, which can solve a basic problem in artillery countermeasure training system well.

Acknowledgments
This work was financially supported by military science technology fund of 13th five-year plan.

References
[1] W. Han, ZH. Hongxu, ZH. Shuangxi, Antagonistic command training based on the virtual reality technology, Defense Technology Review, 2018, 39 (01): 117 - 119.
[2] H. Kaiming, G. Gkioxari, P. Dollár: Mask R-CNN. In:Proceedings of the IEEE international conference on computer vision.2017, pp: 2961 - 2969.
[3] R. Shaoqing, H. Kaiming, R. Girshick, Faster R-CNN: towards real-time object detection with region proposal networks, IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39 (6), pp: 1137 - 1149.
[4] T. Y. Lin, P. Dollár, R. Girshick, Feature pyramid networks for object detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017, pp: 2117 - 2125.
[5] H. Kaiming, Zh. Xiangyu, R. Shaoqing, Deep residual learning for image recognition, Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp: 770 - 778.
[6] L. Wei, D. ANGUELOV, D. ERHAN, SSD: single shot multibox detector, European Conference on Computer Vision, 2016, pp: 21 - 37.