Research on Small Target Recognition Technology Based on Deep Learning

Yuewu Li*, Jianwang Hu and Bing Ji

The Army Engineering University of PLA, Shijiazhuang, Hebei, 050003, China
*Corresponding author’s e-mail: 1315207100@qq.com

Abstract. All small target recognition has always been a difficult problem in the field of target recognition. Traditional target detection methods have met a bottleneck. With the development of deep learning, the field of target recognition has witnessed a boom based on deep learning. According to the characteristics of the small target recognition, this paper briefly introduces the traditional target recognition method, and studies the methods of target detection based on the deep learning. It mainly summarized the development of research on target detection algorithm based on deep learning. Finally, it summarized the influence of the depth study on target recognition and the trend of development in the future.

1. Introduction

Target recognition technology refers to the judgment of the attribute category of the target based on computer image technology. It is one of the basic topics in the field of computer vision, and is also a hot issue. However, the identification of small targets has always been a difficult problem in this field. The research on the identification of small targets has attracted more and more attention. Early target recognition technology is usually based on manual retrieval feature and classical classification algorithm. It has the advantages of convenient implementation and fast calculation speed, but it is often not accurate enough, and it is not robust to interference such as direction, scale change, noise and occlusion. Meantime, it is sensitive to data sets, not suitable for detection and identification of weak small targets. In recent years, the increasingly extensive application of deep learning has gradually become an effective tool for experts and scholars to help people better explore and research in the field of target detection. Compared with traditional visual algorithms, deep learning has great advantages in the field of target detection. The characteristics obtained by autonomous learning under big data exceed the performance of human-designed algorithms. The precision is greatly improved and the learning characteristics are richer, even the feature ability is expressed more, and the requirements of small target recognition are met. Target detection based on deep learning has gradually taken the lead, and the performance has been continuously improved. More and more scholars try to solve the difficult problems in the target detection field from the perspective of deep learning.

2. Small target recognition features

Small target recognition is a difficult point in target recognition, and has the following characteristics [1]:

- The signal-to-noise ratio (SNR) is very low: at long distances, the target signal received by the detector is weak and the image signal-to-noise ratio is low;
• The amount of information available is small: the detector is far from the target, and the target image is point-like, so that it is difficult to distinguish the target from the noise. The basis for distinguishing between the two is generally only the motion characteristics of the target (speed, direction, trajectory) and the difference in gray scale between the target point and background. There basically no shape information can be utilized.

• The background information is complicated: the interference of the cloud layer in the atmosphere, the natural topography of the ground and the building are not only strong, but also have a strong spatial structure.

• Large amount of information processing: the target needs to be detected in the entire graphics space. Since the signal-to-noise ratio of the image is low, in order to correctly detect the target and determine its position in the image, it is necessary to utilize multi-frame image information, so that it make the amount of data information to be processed bigger.

• Stable target information: When detecting targets at a long distance, since the distance is relatively large, the obtained target information (grayscale, motion characteristics) is relatively stable, and generally no mutation occurs, which is an important basis for detecting and identifying weak targets.

3. Target detection and recognition methods

3.1 Traditional target detection and identification

At present, the detection method of moving small targets in a complex background is mainly carried out in the case of multi-frame images. At home and abroad, a lot of researches have been done on the detection of small moving targets under low SNR conditions. Experts and scholars have proposed a variety of detection methods, which can be roughly divided into two categories:

• Detection before tracking (DBT): Firstly, each frame of image is single framed, then several single frame detection results are correlated, and finally the target information stability is used to confirm the real target, which can be seen Signal-to-noise ratio has a great relationship, mainly used for weak target detection in images with complex background and low noise;

• Tracking before detection (TBD): This method does not first detect the presence of a target in a single-needle image. Rather, it directly associates multiple images, obtains several possible targets, and then analyzes the information of each possible target, and then detects whether the target exists. It is mainly used for weak target detection in images with simple background and strong noise.

3.2 Target detection and recognition based on deep learning

In order to solve the problems caused by traditional target detection methods, scholars have proposed some new methods, especially due to the rise of deep learning, which has set off a craze based on deep learning target detection methods. It is mainly a convolutional neural network (CNN) and a region proposal algorithm. Since 2014, the target detection has made a huge breakthrough. Deep learning algorithms are usually based on a large number of training samples, relying on the powerful feature extraction capabilities of convolutional neural networks to achieve classification. Although the training speed is slower than the traditional algorithm, and the required resources are more, the accuracy of the implementation is relatively high.

3.2.1 RCNN. In 2014, RBG and his students proposed the R-CNN algorithm, which replaces the sliding window + manual design features used in traditional target detection. The algorithm is basically divided into four steps: a. input image; b. extract region proposals; c. compute CNN features; d. classify regions. The framework of the target detection process is shown in Figure 1[3].
R-CNN algorithm solves some problems existing in traditional target detection, but it still has some disadvantages: training is divided into many stages and steps are cumbersome; training takes time and takes up a lot of disk space; slow speed, it takes 47s to process an image using GPU and VGG16 model[24].

3.2.2 SPP-NET. In 2015, Kaiming He and his team proposed SPP-NET algorithm, which effectively solved the problem of slow processing speed of RCNN. SPP-NET has the following characteristics[4]: combined with the spatial pyramid method, the multi-scale input of CNN is realized; only one convolution feature is extracted from the original image.

SPP-NET saves a large amount of computing time, and has about one hundred times faster than the R-CNN. The SPP-NET network structure is shown in figure 2.

Compared with R-CNN, SPP-NET can greatly accelerate the speed of target detection, but there are still many problems: a. training is divided into many stages, the steps are cumbersome; b. SPP-NET fixes the convolution layer when fine-tuning the network, and only fine-tunes the fully-connected layer. For a new task, it is necessary to fine-tune the convolutional layer.

3.2.3 Fast RCNN. In 2015, RBG mainly adopted the idea of SPP-NET and proposed the Fast R-CNN algorithm alone for the problem of slow R-CNN. He proposed a network layer that can be regarded as a single-layer SPP-NET, called ROI Pooling[25]. Fast R-CNN combines the advantages of R-CNN and SPP-NET and introduces a multi-task loss function, making it easier to train and test the entire network. The frame diagram is shown in Figure 3[5].
However, Fast R-CNN still has some shortcomings: the extraction of region proposal uses selective search, resulting in that the target detection time is mostly consumed above, which can not meet the real-time application, and there is no real end-to-end training test.

3.2.4 Faster RCNN. In 2015, Kaiming He and Ross Girshick collaborated to propose Faster R-CNN, the third version of R-CNN, which can be used to directly generate and classify region proposals using CNN.

The core idea of RPN is to directly generate a region proposal using a convolutional neural network. The method used is essentially a sliding window. However, the design of the RPN is ingenious. It only needs to be swiped over the last convolutional layer. The overall flow chart shown in Figure 4 [6].

![Figure 4. The overall flow chart](image)

Faster RCNN uses end-to-end network for target detection, which has been greatly improved in both speed and accuracy. However, it still cannot achieve real-time target detection, obtain region proposal in advance, and then classify and calculate each proposal in a small amount.

In order to achieve real-time target detection, the proposed YOLO method makes it possible.

3.2.5 YOLO. In 2015, Joseph Redmon proposed a new object detection method YOLO (You Only Look Once). As the name implies, you can know whether it exists and where only once, and further combine target judgment and target recognition into one. So the recognition performance has been greatly improved, reaching 45 frames per second, and in the fast version of YOLO (Fast YOLO, less convolutional layer), can reach 155 frames per second. The YOLO model is shown in Figure 6[7].

![Figure 5. The YOLO model](image)

YOLO converts the target detection task into a regression problem, which greatly speeds up the detection, enabling YOLO to process 45 images per second. And the false positive ratio is greatly reduced because each network predicts the target window using full-image information[26]. However, YOLO does not have a region proposal mechanism. Using only 7*7 mesh regression will make the target not very accurate, which leads to the YOLO detection accuracy is not very high.

3.2.6 SSD. In 2016, Wei Liu and his team proposed the SSD network, and the SSD network improved the YOLO network accordingly. The framework is shown in Figure 6. Figure 6(a) shows an input picture with two Ground Truth borders, and Figures 6(b) and (c) show an 8x8 grid and a 4x4 grid, respectively. Obviously the former is suitable for detecting small targets, such as the cat in the picture. The latter is suitable for detecting large targets, such as dogs in pictures. There is a series of fixed-size Boxes on each grid. These are called Default Boxes in the SSD and are used to frame the position of objec
the target object. During training, Ground Truth will be given to a fixed Box, as shown the blue box in Figure 6(b) and the red box in Figure 6(c)[8].

Figure 6. The SSD framework

SSD also improves the YOLO speed while keeping the YOLO speed. It mainly uses the Anchor mechanism in Faster R-CNN and uses multi-scale. However, since the shape of the Default Box and the size of the grid are fixed in advance, its effect to the small target detection of the specific picture is relatively poor.

3.2.7 Mask RCNN. In 2017, Kaiming He proposed the Mask RCNN, which is a further improvement of Faster R-CNN. It's conceptually simple, just like the previous version of Faster R-CNN, but with an additional output: Faster R-CNN has two outputs per candidate, label and bounding-box offset. To improve accuracy, a third branch of the output object mask (binary mask) is added. But the additional Mask output is different from the class and box output, and you need to extract a more elaborate spatial layout of the object. At the same time, the classification also depends on the mask prediction. The network frame diagram of Mask RCNN is shown in Figure 7[9].

Figure 7. The Mask RCNN network framework diagram

Mask RCNN applies image segmentation to the detection, further enhancing the detection function, and using non-competitive sigmoid instead of soft-max is a big innovation.

The above are some of the more popular deep learning-based target detection methods proposed by scholars and experts in recent years. In general, they can be divided into two groups: R-CNN series based on regional nomination; YOLO and SSD series without regional nomination. Compared with the traditional target detection method, these methods have great advantages, and with the development of technology, the problems in the detection process are continuously solved, so that the target detection speed is faster, the precision is higher, and the real-time performance is stronger.

Recently, there are still many scholars working hard to make some progress in target detection, and proposed some new target detection methods based on deep learning. Tang Cong et al. proposed a multi-window SSD target detection method based on deep learning[10]; Cai Z et al. proposed a multi-level target detection architecture, Cascade R-CNN [11]; H Hu and his team proposed an object relation module using the state-of-the-art object detection system in the learning process [12]; B Singh et al. proposed a large-scale real-time target detector R-FCN-3000, which successfully applied the R-FCN algorithm in scenarios with many detection categories [13] and so on. However, there is still a long way to go for target detection, and more research on small target detection is needed.

4. Conclusion
Target recognition is an important topic in the field of computer vision. Weak target recognition is one
of the most difficult points. This paper first summarizes the characteristics of several weak target recognition, and briefly introduces some traditional target recognition methods. Although the implementation speed is fast and easy to implement through manual retrieval and classical classification algorithm design, the accuracy is not very high, and the robustness to interference is not strong, which is difficult to meet the current needs. And the target recognition based on the traditional method is more difficult to apply to small target recognition.

With the rapid development of deep learning, some science and technology developed by deep learning have had a profound impact on the image field. Compared with the traditional target detection method, target recognition based on deep learning has great advantages, can achieve higher precision, stronger robustness, and is more suitable for small target detection. Although it has the disadvantage of long training time, it can be by continuously improving the algorithm, the training time is reduced, making it faster and more real-time.

With the application of deep learning, the research of target recognition has been greatly developed. Experts and scholars have continuously improved and innovated the deep learning algorithm. However, due to the characteristics of small target recognition, it is still a difficult problem in the field of target recognition. Further research is needed for the identification of small targets. For example, the depth needs to be deepened, the convolution module function needs to be enhanced, and new functional units, loss functions need to be designed, and so on.

References
[1] Tan X. (2008) Research on detection and recognition technology of weak targets in complex environment. Nanjing University of Aeronautics and Astronautics.
[2] Wang T. (2017) Research on high reliability detection method of infrared weak target in complex space environment. University of Chinese Academy of Sciences.
[3] Jan H., et al. (2016) What Makes for Effective Detection Proposals. IEEE Transactions on Pattern Analysis & Machine Intelligence, 38(4):814.
[4] He K., Zhang X., Ren S., et al. (2015) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Trans Pattern Anal Mach Intell, 37(9):1904-1916.
[5] Girshick R. (2015) Fast R-CNN. Computer Science.
[6] Ren S., He K., Girshick R., et al. (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.
[7] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015) You only look once: unified, real-time object detection.
[8] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., & Fu, C. Y., et al. (2015) SSD: single shot multibox detector.:21-37.
[9] Kaiming, H., Georgia, G., Piotr, D., & Ross, G. (2018) Mask r-cnn. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1-1.
[10] Tang C., Ling Y., Zheng K., et al. (2018) Multi-window SSD target detection method based on deep learning. Infrared and Laser Engineering, 003(1): 290-298.
[11] Cai, Z. , & Vasconcelos, N. (2017). Cascade r-cnn: delving into high quality object detection.
[12] Hu H, Gu J, Zhang Z, et al. (2017) Relation Networks for Object Detection.
[13] Singh B, Li H, Sharma A, et al. (2017) R-FCN-3000 at 30fps: Decoupling Detection and Classification.
[14] Luan S. (2018) Research on the development trend of deep learning target recognition algorithm. Telecommunication network technology.(4).
[15] Zhao H. (2017) Research on target detection algorithm based on deep learning. Beijing University of Posts and Telecommunications.
[16] Fu R. (2017) Research on Target Detection Based on Deep Learning. Beijing Jiaotong University.
[17] Wan W. (2015) Research and application of target detection algorithm based on deep learning. University of Electronic Science and Technology of China.
[18] Zhou X., Wang., Li L. (2017) Overview of Target Detection Algorithm Based on Deep Learning.
Electronic Measurement Technology,(11): 89-93.
[19] Xiong L. (2017) Research on target recognition based on deep learning. Nanchang Hangkong University.
[20] Liang D. (2015) Research on target recognition based on deep learning and its multi-robot formation. Harbin Institute of Technology.
[21] Zhao X. (2018) Remote sensing image target detection and recognition based on convolutional neural network. University of Science and Technology of China.
[22] Sun G. (2017) Research on road scene object detection method based on deep convolutional neural network. North China University of Technology.
[23] Zheng J. (2017) Research and design of well descending human detection system based on DCNN. Xi'an University of Science and Technology.
[24] Cao W. (2017) A review of deep convolutional neural networks based on regional information. Electronic World,(06):32+36.
[25] Liu J. (2016) Detection and identification of traceable video targets based on deep learning. Southeast University.
[26] Wang Z. (2017) Research on fast target detection technology based on deep learning. Tianjin University of Technology.