Research On Multi-Source Image Fusion Target Detection Technology Based on Neural Network

Hao Lv*, Bao Deng, Chenguang Shi
Xi'an Aeronautics Computing Technique Research Institute, Aviation Industry Corporation of China, Xi'an, China

*Corresponding author: lvhao20032003@mail.nwpu.edu.cn

Abstract. Target detection and recognition technology based on multi-sensor fusion such as radar, visible light, and infrared has become a development trend in industries and military fields. Deep learning networks have strong characterization learning capabilities and are good at extracting various complex features from multi-source heterogeneous data. This article reviews the research and development status of multi-source image fusion and target detection at home and abroad, and discusses the multi-source information fusion technology and multi-sensor system construction plan. Aiming at the fusion target detection problem of visible light image and infrared image, a target detection algorithm based on decision-level fusion is introduced. The YOLOv3 network is used to detect the visible light image and the infrared image separately, and then achieve rapid target detection based on decision-level fusion.

Keywords: Deep Learning, Image Fusion, Target Detection, CNN.

1. Introduction
As an important research content in the field of computer vision, target detection is widely used in target tracking, pedestrian recognition, medical images, video surveillance, intelligent machines, and military welders. The main task of target detection is to efficiently and accurately identify and locate instance targets of predefined categories in images [1].

In recent years, with the rapid development of deep learning technology, more and more target detection algorithms have emerged. The solution of the detection system generally uses radar as the core, supplemented by other sensors such as photoelectric (visible light and infrared), acoustics, and radio detection. People's demand for fusion of multi-source information in various applications is increasing day by day, and data fusion technology has received widespread attention. The goal of data fusion is to make up for the weakness of a single sensor, thereby obtaining more accurate detection results. On the other hand, artificial intelligence and deep neural networks (DNNs) have become a very attractive data processing method, which can find high-level abstract features that are difficult to find by typical feature extraction methods. Therefore, they are used in massive multi-source data processing. It is widely used, and its application in target detection and classification has begun to emerge [2].
2. Multi-Source Image

2.1. Radar Image
As the main means of monitoring and early warning of air and sea targets, radar is widely used in the fields of national defense and public safety. Compared with other technologies, radar is actually the only long-distance capable of several kilometers to tens of kilometers. A technology that detects and is almost unaffected under adverse light and weather conditions.

Traditional surveillance radars usually use mechanical scanning antennas to search the entire airspace. They stay on a single target for a short time. They can only obtain limited data such as the target's azimuth, angle, radial velocity, RCS, etc., making it difficult to detect "low, small, slow" targets. However, with the continuous development of radar electronic technology, its ability to detect weak targets has steadily improved, providing a new way for the detection and identification of "low, small, and slow" targets.

A series of UAV detection and identification studies have been carried out using advanced radar technologies such as active phased array, multiple input multiple output (MIMO), frequency modulated continuous wave (FMCW), holographic radar, and external source radar. Some systems have been preliminary Possess engineering application capabilities.

2.2. Visible Light Image
The target image information acquired by visible light detection equipment is more abundant than that of radar, and it has advantages in target recognition. However, a single visible light camera usually does not have the target ranging function. It needs to obtain the target distance through multi-station cross positioning or adding laser ranging equipment. It is easily affected by environmental factors and the detection range is much lower than that of radar. Under normal circumstances, it is difficult to use it alone. Therefore, the visible light camera is usually only used as a supplementary confirmation device for the radar. A typical photoelectric detection system can obtain both visible light and infrared images of the target at the same time.

2.3. Infrared Image
Unlike visible light sensors, infrared sensors are more sensitive to the infrared spectrum with a wavelength of 9 μm to 14 μm in the electromagnetic spectrum. Its main advantage is that it is not affected by light or weather conditions, and can still operate normally even in complete darkness [3]. In general, the resolution of infrared cameras is lower, but the cost is higher. Therefore, they were
initially only used in the military, but with the advancement of technology, their costs have been reduced, allowing them to be used in industrial and research sectors.

Like a visible light camera, an infrared camera cannot give the distance of the target, so it cannot locate the target. In most multi-sensor systems, infrared cameras are usually combined with visible light as an auxiliary device for radar.

![UAV infrared image](image)

**Figure 2.** UAV infrared image

3. Image Fusion Method Based on Deep Learning

The rapid development and good application of convolutional neural networks (convolutional neural networks, CNN) have led researchers to consider the use of CNN to solve multi-source image fusion. For example, literature [4] uses deep convolutional neural networks to solve the problem of multi-focus image fusion. Literature [1] solves the problem of infrared and visible light image fusion with the help of deep learning theory; literature [5] proposes a battlefield situation display technology based on information fusion to realize the fusion settlement and effect display of missile-borne sensor information. The information fusion part is divided into two aspects: data fusion and image fusion.

Literature [1] proposed a YOLO3 improved missile-borne image target detection algorithm. This algorithm improved the YOLO v3 method and improved the multi-scale prediction for the characteristics of rapid change of the missile-borne image scale, high positioning accuracy, and strong real-time requirements. K-means dimensional clustering is performed on the a priori box size on the branch feature map, which enhances the scale adaptability; the position loss function is improved to improve the position positioning ability; the fast NMS algorithm is used to accelerate the prediction process and improve the real-time performance of the network.

Image fusion is a process in which the fusion image is obtained by adding the information of multiple original images, and then the fusion image is analyzed and processed. According to the characteristics of fusion processing and the degree of abstraction, image fusion methods are divided into three categories: pixel-level image fusion, feature-level image fusion, and decision-level image fusion. Pixel-level image fusion refers to the process of selecting a fusion strategy to process the pixels of a strictly registered source image to obtain a fused image; feature-level image fusion refers to extracting the target features of the region of interest in the source image, and comparing these features Information is fused, and finally the fused image is obtained; decision-level image fusion is a fusion method based on a cognitive model, based on feature extraction, to identify and judge the feature information of the image, and select the appropriate one according to actual needs Decision-making The process of global optimal processing of images.

The target detection algorithm based on decision-level fusion mainly includes three parts: establishing an experimental data set; using YOLOv3 as the basic network to detect visible light images and infrared images separately; designing a decision-level fusion method to fuse the detection results. Determine whether the target A is accurately detected in the infrared image Determine whether
the target A is accurately detected in the visible light image. The detection result of the weighted fusion target A, where the calculation of the weighted fusion is expressed as:

$$R_f = \theta_v R_v + \theta_r R_f$$

Where: $R_v$ and $R_f$ are the accurate detection frame position of target A in visible light image and infrared image respectively; $\theta_v$ and $\theta_r$ are the weights corresponding to $R_v$ and $R_f$ respectively; $R_f$ is the detection frame position of target A corresponding to the target in the fusion image.

![Diagram](image)

**Figure 3. Infrared and visible light fusion algorithm framework based on YOLOv3**

4. Conclusion

This paper reviews the research work of UAV detection and recognition methods based on radar, visible light, infrared and other sensors and multi-sensor information fusion algorithms. The focus is on a target detection algorithm based on decision-level fusion. The algorithm uses the YOLOv3 network to train the visible light image and the infrared image separately, and processes the target detection results based on the decision-level fusion method, and then obtains the accurate detection results of the fused image.

References

[1] BAI Yu, et al. An Object Detection Algorithm Based on Decision-Level Fusion of Visible Light. Image and infrared Image. Journal of air force engineering university (Natural Science Edition), China, 2020, 21(6):53-59.

[2] CHEN Weishi, HUANG Yifeng, LU Xianfeng. Survey on application of multi-sensor fusion in UAV detection technology. Modern Radar, China, 2020, 42(6): 15-29.

[3] XUE Song, et al. Review of Key Technologies for Target Detection of Multi-source Terminal Missile-based Fusion Image. Journal of Projectiles, Rockets, Missiles and Guidance. China. 2021, 41(3):176-181.

[4] Venieris S I, Bouganis C S. Latency-driven design for FPGA-based convolutional neural networks. 2017 27th International Conference on Field Programmable Logic and Applications (FPL). Piscataway: IEEE, China, 2017: 1-8.

[5] MAX, OH BS, SUN L, et al. EMD-based entropy features for micro-Doppler mini-classification. / 2018 24th International Conference on Pattern Recognition (ICPR). Beijing: IEEE Press, China, 2018: 1295-1300.