Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 4, No. 3, 2025

Research on Camouflage Target Detection in Two-Branch Multi-Source Remote Sensing Images

Heng Zhang

School of Resources and Environment, Henan Polytechnic University, Jiaozuo 454000, Henan, China

Abstract: In military operations, the accurate detection of enemy camouflaging targets is of great significance to improve our combat capability. Most of the existing camouflage target detection methods are aimed at a single kind of natural image, and their ability to detect military camouflage image in complex combat environment is limited. In order to solve this problem, this paper proposes a two-branch multi-source remote sensing image camouflage detection method. By combining the fine-grained features of the target in infrared and visible images, the method fuses the camouflage target information to improve the detection capability. Experiments show that the accuracy, recall rate and mAP@.5 of the proposed method are 0.88, 0.859 and 0.923, respectively, which are 0.054, 0.099 and 0.096 higher than the values of YOLOv5s, YOLOv7 and mAP@.5 based on IFCNN method, respectively. Therefore, the method proposed in this paper has outstanding performance in the detection of pedestrian and vehicle camouflage targets, and provides a new method in the field of multi-source information fusion camouflage targets detection.

Keywords: Camouflaged Object Detection; Image Fusion; Remote Sensing Image; Two-Branch Network

1. Introduction

Throughout military history, camouflage techniques inspired by biological concealment have been employed to enhance the survivability of personnel and facilities in dynamic battlefield environments. Compared to conventional targets, camouflaged targets exhibit similar colors and textures to their surroundings, with blurred contours, making them difficult to detect using standard target detection algorithms. Consequently, specialized techniques for detecting military camouflage targets have emerged.

Early camouflage target detection methods primarily focused on low-level features such as color, shape, position, and gradient information in single-source images. These methods relied on handcrafted features to distinguish foreground targets from the background, thereby identifying camouflaged regions [1]. While effective in specific scenarios, these approaches suffered from high computational demands, cumbersome operations, and heavy reliance on manually designed features. With advancements in deep learning, camouflage target detection methods have gradually eliminated manual intervention, employing neural networks to implicitly segment foreground and background regions. Techniques such as SINet [2], PFNet [3], and SINet-V2 [4] leverage deep learning to extract intrinsic relationships between samples and targets from large-scale annotated data, achieving improved detection speed and performance. However, these methods predominantly focus on single-source images, resulting in limited feature diversity and an inability to fully utilize multi-modal target information.

The rise of remote sensing technology has provided rich and diverse features for camouflage target detection. For instance, panchromatic and multispectral images offer spatial and spectral resolution, supplying detectors with physical morphology and material composition information of camouflaged targets. As a result, researchers have explored remote sensing-based methods such as YOLO-FIRI [5], I-YOLO [6], and IARet [7] for camouflage target detection. While these methods leverage remote sensing image characteristics to improve target identification and localization, they are constrained by imaging mechanisms, as single sensors cannot simultaneously capture high spatial and spectral resolution [8]. Therefore, multi-source image fusion techniques that integrate information from different types of remote sensing images to generate fused images with enhanced spatial and spectral resolution are crucial for improving detection performance.

To address the limitations of single-feature extraction pipelines in handling multi-modal images, this paper proposes a two-branch multi-source remote sensing image camouflage target detection method. By exploiting the complementary spatial and spectral resolutions of panchromatic and multispectral images, the method employs dual-branch feature extraction to separately extract contrast features from infrared images and texture details from visible images. The fused image, enriched with fine-grained features, enhances the detection rate of camouflaged targets while reducing false negatives, achieving precise identification and localization.

2. Methodology

The complete network framework mainly completes two tasks: one is feature fusion. Through the dual-branch feature extraction route, the ordinary convolutional layer and the densely connected resblock[9] variant convolutional layer are used to extract infrared contrast features and texture detail features from infrared and visible images respectively. Then, these fine-grained features are input into the feature fusion module, and the feature fusion module uses the stitching strategy to fuse the two fine-grained features to retain as many texture details as possible and highlight the pixel intensity characteristics of the infrared image. The fused features will be sent to the detection network for camouflage target detection. In order to observe the detection results more intuitively, the multi-layer convolution operation is used to reconstruct the fused image from the fused features after feature fusion. The second is target detection. Here, YOLOv7[10] ordinary target detection network framework is used to detect the fused image, and the final detection result is presented on the fused image.

In order to obtain better detection results and guide the training direction of the network, the image fusion loss function and the YOLO loss function are used to jointly guide the detection of camouflaging targets. The fusion results of infrared and visible images are constrained from three aspects: pixel intensity, texture detail and structural similarity. The detection performance of the network is enhanced from three aspects: category confidence, coordinate regression and target confidence. Here the intensity loss, texture loss and structural similarity loss are shown in Equations (1), (2) and (3).

$$L_{int} = \frac{1}{HW} \left\| I_f - \max\left(I_{vi}, I_{ir}\right) \right\| \tag{1}$$

$$L_{text} = \frac{1}{HW} \left\| \left| \nabla I_f \right| - \max\left(\left| \nabla I_{vi} \right|, \left| \nabla I_{ir} \right|\right) \right\| \tag{2}$$

$$L_{SSIM} = 2 - \left[\mathrm{SSIM}\left(I_f, I_{vi}\right) + \mathrm{SSIM}\left(I_f, I_{ir}\right) \right] \tag{3}$$

Where: is the intensity loss, is the texture loss, and is the intensity similarity loss. H and W are the height and width of the image, is the fused image, is the visible image, is the infrared image, is the maximum value operation, is the absolute value operation, is the gradient operation, and is the intensity similarity calculation formula.

The image fusion loss function is the weighted sum of three losses, and the formula is shown in (4).

$$L_{Fusion} = \beta_1 L_{int} + \beta_2 L_{text} + \beta_3 L_{SSIM} \tag{4}$$

Where, is the weight of each sub-loss, and the sum of the three equals 1.

Finally, a joint loss function is formed by the fusion loss and the detection loss to jointly guide the camouflage target detection performance, and the formula is shown in (5).

$$L = \alpha L_{Fusion} + (1 - \alpha) L_{YOLO} \tag{5}$$

Where: is the total loss and is the joint loss weight.

Where VI is the visible light image, IR is the infrared image, C is the ordinary convolution layer with convolution kernel 3, B is the resblock variant convolution layer with dense connection, F is the infrared and visible light fusion image, C 'is the convolution layer for image reconstruction, and the number is the number of channels.

3. Experiments and Analysis

In order to verify the performance index of the above algorithm, this paper designs the following experiments for testing. In this paper, 2096 pairs of registered infrared and visible images are collected from the public datasets TNO[11], MSRS[12] and DroneVehicle[13] for the training, validation and testing of the model. In order to better simulate the camouflage attributes of camouflage targets, the night images with dim lighting are selected, and the target categories are set to pedestrians and vehicles. The PyTorch framework was used to implement the code, the NVIDIA GTX4060 GPU was used for experiments, the Adam optimizer was used to optimize the network, the initial learning rate was 0.001, the learning rate decay was 0.75, and the batch size was 6.

In order to better illustrate the prediction effect of the method in this paper, the current commonly used object detection algorithms YOLOv5[14], YOLOv7 and IFCNN[15] are used to process the same data set, and the predicted value is recorded and compared with the Ground Truth (GT). IFCNN is a general image fusion algorithm. Modify its source code, replace the image fusion module of the algorithm in this paper with it, and explore the improvement effect of the dual-branch fusion module in this paper on the detection ability. Precision (P), Recall (R) and mean Average Precisionm (mAP) are used to measure the detection performance of each algorithm for pedestrian and vehicle targets. The formulas for precision and recall are given in (6) and (7).

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{6}$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{7}$$

Where: TP(True Positive) is the number of positive samples predicted and actually positive samples. FP(False Positive) is the number of samples that were predicted to be positive but were actually negative. FN(False Negative) predicts the number of negative samples, but is actually the number of positive samples.

The proportion of the predicted positive samples that are correctly predicted is defined as precision P, and the proportion of the predicted positive samples to the total number of actual positive samples is defined as recall R. The curve plotted with Precision P as the vertical axis and recall R as the horizontal axis is called the P-R curve, and the area under the P-R curve is defined as the Average Precision, which reflects the training effect of each class in the model. The formula is given in (8).

$$AP = \int P(R)\,\mathrm{d}R \tag{8}$$

The average AP value of each category is defined as mAP, which can reflect the overall training degree of the model. The Intersection over Union (IoU) is defined as the intersection over union (IOU) between the predicted and true pixels. mAP@.5 represents the mAP value at IoU of 0.5, and mAP@.5:.95 represents the mean value of mAP at different IoU thresholds, from 0.5 to 0.95, in steps of 0.05. The pedestrian and vehicle targets are detected using four algorithms, and the detection results are shown in Table 1.

Tab.1 Quantitative comparison of human and vehicle detection accuracy of different algorithms

	CLASS	P			R			mAP@.5	mAP@.5:.95
		preson	car	total	preson	car	total		
YOLOv5s	VI	0.682	0.663	0.673	0.657	0.794	0.726	0.765	0.510
	IR	0.859	0.804	0.832	0.910	0.722	0.816	0.869	0.607
YOLOv7	VI	0.843	0.803	0.823	0.448	0.868	0.658	0.709	0.415
	IR	0.841	0.648	0.745	0.935	0.664	0.800	0.824	0.542
IFCNN	VI+IR(Y)	0.874	0.866	0.870	0.773	0.765	0.769	0.827	0.528
ours	VI+IR(RGB)	0.901	0.858	0.880	0.775	0.943	0.859	0.923	0.629
	VI+IR(Y)	0.908	0.858	0.883	0.848	0.833	0.841	0.922	0.614

From Table 1, it is easy to see that in the image detection models of single route YOLOv5s and YOLOv7. The two detectors have high accuracy for pedestrian target detection in infrared images, but not prominent for vehicle target detection, and the opposite is true for visible images. This means that both detectors can learn more pedestrian contrast features from infrared images and more vehicle texture details from visible images. However, both of them cannot simultaneously learn good pedestrian and vehicle features in a single type of image.

In the IFCNN using dual-branch routes and the proposed algorithm, it has excellent detection results for pedestrian and vehicle targets. It indicates that these two algorithms can integrate pedestrian features in infrared images and vehicle features in visible images. However, in the generated fusion image, the pedestrian target is affected by the texture detail features extracted from the visible light image, and the detection recall rate is reduced. Compared with IFCNN, the proposed algorithm has advantages in the detection accuracy and recall rate for pedestrian targets. For vehicle targets, the accuracy is not as high as IFCNN algorithm, but the difference is very small, and the recall rate has obvious advantages. It shows that when using two kinds of images at the same time, this paper has a stronger ability to integrate the target information in the two kinds of images, can retain more target features, and the detector has better detection performance for the target.

Finally, the influence of using three-channel RGB image and single-channel grayscale image on object detection performance is compared in the proposed algorithm. Although the detection using grayscale image will increase the accuracy of the detector for pedestrian target detection and make the overall detection performance more balanced, from the perspective of mAP@.5 and mAP@.5:.95, the overall detection performance of the detector is better when using three-channel RGB image detection, reaching the leading level in comparison methods.

The first column is the detection results of YOLOv7 on visible images, the second column is the detection results of YOLOv7 on infrared images, the third column is the detection results based on IFCNN method, and the fourth and fifth columns are the test results of the proposed algorithm. It can be seen from (a), (f), and (k) that the detector can detect the vehicle target in the visible light image, but there will be missed detection of the pedestrian target. It can be seen from (b), (g), and (l) that the detector has higher confidence in the pedestrian target in the infrared image than in the visible image, and can detect the missed pedestrian target in the visible image, but the detection level of the vehicle target is not as good as that of the visible image. (c), (h) and (m) are detection methods based on IFCNN, which can detect pedestrian and vehicle targets, but there will be false detection. In contrast, the results of the proposed algorithm can not only retain the detection advantages of infrared and visible images, but also avoid false detection in (m), and the detection level is the best.

4. Conclusion

This paper presents a two-branch multi-source remote sensing image camouflage target detection method. By leveraging dual-branch feature extraction to integrate texture details from visible images and intensity features from infrared images, the method generates fused images with enhanced spatial and spectral resolution, significantly improving detection performance. Experimental results demonstrated superior precision and recall compared to existing methods. Future work will address challenges such as small target pixel ratios and explore further advancements in multi-source image fusion for camouflage target detection.

References

[1] Zhang Dongdong, Wang Chunping, Fu Qiang. Research Progress on Camouflaged Target Detection. Laser Journal, 2024, 45(03): 1-13.

[2] Fan D P, Ji G P, Sun G L, et al. Camouflaged object detection. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020:2777-2787.

[3] Mei H Y, Ji G P, Wei Z Q, et al. Camouflaged object segmentation with distraction mining. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021:8772-8781.

[4] Fan D P, Ji G P, Cheng M M, et al. Concealed object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2021, 44(10): 6024-6042.

[5] Li, S., Li, Y., Li, Y., et al. YOLO-FIRI: Improved YOLOv5 for Infrared Image Object Detection. IEEE Access 2021, 9, 141861–141875.

[6] Sun M, Zhang H, Huang Z, et al. Road infrared target detection with I-YOLO. IET Image Processing, 2022 Jan;16(1):92-101.

[7] Jiang, X.; Cai, W.; Yang, Z.; Xu, P.; Jiang, B. IARet: A Lightweight Multiscale Infrared Aerocraft Recognition Algorithm. Arab. J. Sci. Eng. 2022, 47, 2289–2303.

[8] Tang Linfeng, Zhang Hao, Xu Han. A Review of Image Fusion Methods Based on Deep Learning. Journal of Image and Graphics, 2023,28(01):3-36.

[9] J. Ma, L. Tang, M. Xu, H. Zhang, G. Xiao, STDFusionNet: An infrared and visible image fusion network based on salient target detection, IEEE Trans. Instrum. Meas. 70 (2021) 5009513.

[10] Wang C Y, Bochkovskiy, LIAO H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, 7464-7475.

[11] Alexander Toet. The TNO Multiband Image Data Collection. 2017.

[12] L. Tang, J. Yuan, H. Zhang, et al. Piafusion: A progressive infrared and visible image fusion network based on illumination aware. Inf. Fusion, 2022, 83, 79–92.

[13] Sun Y, Cao B, Zhu P, et al. Drone-Based RGB-Infrared Cross-Modality Vehicle Detection Via Uncertainty-Aware Learning. IEEE Trans. Circuits Syst. Video Technol, 2022, 32, 6700–6713.

[14] Zhan W, Sun C, Wang M, et al. An improved Yolov5 real-time detection method for small objects captured by UAV. Soft Comput, 2022, 26, 361–373.

[15] Y. Zhang, Y. Liu, P. Sun, et al. IFCNN: A general image fusion framework based on convolutional neural network. Inf. Fusion, 2020, 54, 99–118.