



OPEN ACCESS

EDITED BY

Xiaoqiang Zhang,
Beihang University, China

REVIEWED BY

Hongxi Zhou,
University of Electronic Science and
Technology of China, China
Jianxin Hao,
Civil Aviation University of China, China

*CORRESPONDENCE

Gui Fu,
✉ abyfugui@163.com
Liwen Liu,
✉ llw@cafuc.edu.cn

SPECIALTY SECTION

This article was submitted to
Interdisciplinary Physics,
a section of the journal
Frontiers in Physics

RECEIVED 08 March 2023

ACCEPTED 30 March 2023

PUBLISHED 14 April 2023

CITATION

Fu G, Han L, Huang S, Liu L, Wang Y and
Zhu X (2023), Research on infrared target
detection based on neural
network fusion.
Front. Phys. 11:1181928.
doi: 10.3389/fphy.2023.1181928

COPYRIGHT

© 2023 Fu, Han, Huang, Liu, Wang and
Zhu. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Research on infrared target detection based on neural network fusion

Gui Fu^{1*}, Lixiang Han¹, Shaohan Huang¹, Liwen Liu^{2*}, You Wang¹
and Xinyu Zhu¹

¹Institute of Electronic and Electrical Engineering, Civil Aviation Flight University of China, Guanghan, China, ²College of Flight Technology, Civil Aviation Flight University of China, Guanghan, China

Infrared technology can detect targets under special weather conditions, such as night, rain and fog. To improve the detection accuracy of vehicles, pedestrians and other targets in infrared images, an infrared target detection algorithm with fusion neural network is proposed. Firstly, we use Ghost convolution to replace the resunit unit of the convolution layer of the deep residual network layer in YOLOv5s, which can reduce the amount of parameters without losing accuracy. Then, the global channel attention (GCA) is added to the upper sampling layer, the detection accuracy of network is further improved by enhancing the characteristics of the overall goal. Also, the Channel Space Attention (CPA) space attention mechanism is added to the output end to obtain more accurate target location information. The infrared data set taken by the UAV is trained and tested. The accuracy rate of detection based on YOLOv5s and fusion neural network is 96.47%, the recall rate is 91.51%, and the F1 score is 94%, which is 7% higher than YOLOv5s. The results show that the target detection rate of infrared images is improved by proposed method, which has strong research value and broad application prospects.

KEYWORDS

infrared images, target detection, YOLOv5S, UAV, neural network fusion

1 Introduction

Target detection technology has already used in vast quantities in safe, transportation, medical and military fields [1]. However, it can be seen that the light imaging system cannot work day and night, and is easily affected by the severe weather conditions such as smoke and dust. The infrared system uses infrared radiation to obtain information with infrared detectors, therefore, it has the characteristics of all-weather operation, good concealment, easy penetration of smoke and dust and strong anti-noise [2–5]. At present, the target detection in infrared scenes has important applications in the fields of autonomous driving [6, 7], video surveillance [8, 9], military and other fields [10].

Due to infrared images were lack the features of color and texture, low signal-to-noise ratio and contrast, severe background noise and low resolution, etc., the traditional algorithm for infrared target recognition has high false detection rate and poor robustness. Many researchers have studied the detection of infrared targets.

For the problem of low detection accuracy inflicted by inaccurate registration of aerial infrared image, a feature extraction method is designed to feed back the detection information of the previous frame to the current frame. Then, a registration accuracy prediction method based on the distribution concentration of interior points is proposed to

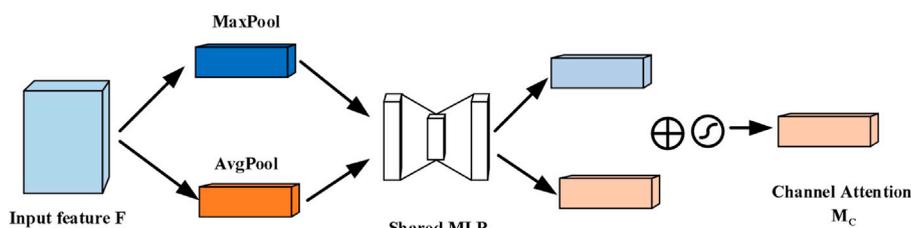


FIGURE 1
The CAM block.

improve the detection accuracy [11]. To improve the detection rate and false-alarm of small target, the spatiotemporal background information in the image is used for low-rank sparse tensor decomposition to detect the targets in complex background [12]. Li, J et al. [13] use motion continuity to solve the problem of eliminating false targets. When the infrared camera has a high gray intensity structural background, Moradi, S [14]. Proposes a directional method to suppress the structural background based on the principle of average absolute gray difference similarity to detect small infrared targets. Bai, X et al. [15] propose a novel algorithm based on a complicated context (DECM) of the guide digital entropy (DECM), it can effectively suppress the noise and enhance the small target in infrared images.

Deep learning has become the focus of image processing, it has a strong automatic feature extraction ability [16, 17]. The deep learning methods for infrared detection can greatly elevate the detection rate of algorithm [18, 19]. In order to further improve the detection rate and enhance the applicability of the detection algorithm, some scholars have proposed fusion network methods for infrared images detection [20–22]. Zhou et al. [23] proposed an approach for improving object detection by designing an adaptive feature extraction module within the backbone network. This module includes a residual structure with stepwise selective kernel, which allows model to extract feature effectively at varying receptive field sizes. Additionally, a coordinate attention feature pyramid network was developed to incorporate location information into the deep feature map using a coordinate attention module, resulting in improved object localization. To further enhance the detection accuracy of weak and dim objects, the feature fusion network was augmented with shallow information. Fan, Y et al. [24] addressed the challenges of vehicle detection in aerial infrared photography, including false detection, missed detection and limited detection capabilities. They achieved this by incorporating Dense Block, Ghost convolution, and SE layer modules into the YOLOv5s. This fusion of neural networks yielded promising results, prompting the authors to propose an improved version of the network that further enhances detection accuracy.

To improve the detection accuracy of vehicles, pedestrians and other targets in infrared images taken by drones, an infrared target detection algorithm with fusion neural network is proposed. The method in this article is to use YOLOv5s as baseline, three improved networks are integrated to the whole model for improve the detection accuracy.

The paper is organized as follows: the relative basic theory is introduced in Section 2, Section 3 details the proposed method, while the experiment and analysis is carried out in Section 4; At last, the conclusions is made in section 5.

2 Relate work

2.1 CAM

The channel attention module, abbreviated as CAM, is a technique that leverages the inter-channel relationship of features to produce a channel attention map. Each channel of feature is regarded as a feature detector [25], CAM focuses on identifying ‘what’ is significant in image. Figure 1 displays the structure of the CAM block.

The channel attention module (CAM) starts by taking the feature map F and applies global maximum pooling and global average pooling operations to it along the H and W dimensions. The resulting feature maps are then fed into a two-level neural network (MLP) with shared weights, where they learn the inter-channel relationships. The dimension between the two MLP layers is reduced by a compression ratio r . Finally, the output of MLP are activated by the sigmoid function to produce the channel weighting coefficients. This can be expressed as the following formula:

$$\begin{aligned} M_c(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))). \end{aligned} \quad (1)$$

2.2 Squeeze-and-Excitation

The Squeeze-and-Excitation block can calculate any transformation relationship $F_{tr}: X \rightarrow U, X \in \mathbb{R}^{H' \times W' \times C'}$, $U \in \mathbb{R}^{H \times W \times C}$. Where F_{tr} is a convolution operator. The outputs of F_{tr} as $U = [u_1, u_2, \dots, u_C]$ can be write:

$$F_{tr}: X \rightarrow u_c = v_c * X = \sum_{s=1}^{C'} v_c^s * x^s. \quad (2)$$

Where $v_c = [v_c^1, v_c^2, \dots, v_c^{C'}]$, $X = [x^1, x^2, \dots, x^{C'}]$, v_c^s is a 2D spatial kernel that operates on v_c , and it is associated with the channel of X . The dependencies are mixed with the spatial correlations and they

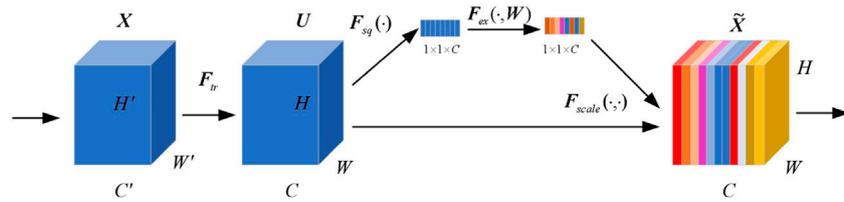


FIGURE 2 The SE building block.

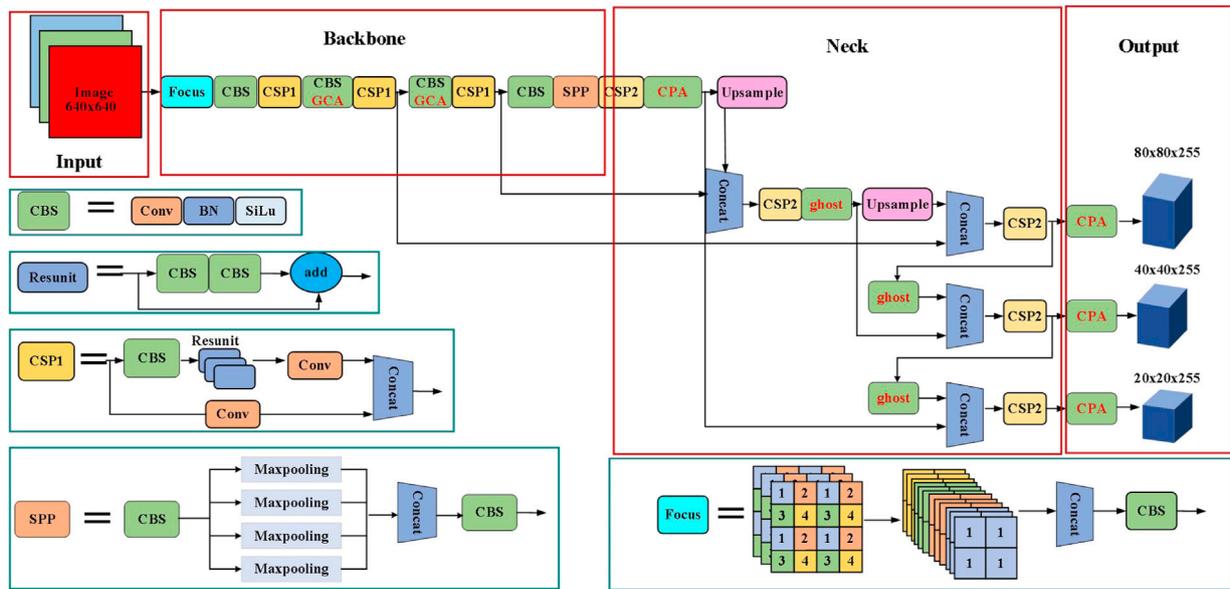


FIGURE 3 The network architecture of proposed method.

are transmitted to the next layer. The SE building block is illustrated in Figure 2.

Global average pooling is used to squeeze global spatial information into a channel descriptor, which generates channel-wise statistics. Specifically, the c -th element of the descriptor is calculated by averaging the activations of all spatial locations in the c -th channel of the feature map.

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j). \quad (3)$$

Where $z_c \in \mathbb{R}^C$ is a statistic.

The exclusion module is composed of two fully connected layers. The first layer reduces the C channels to C/r channels to reduce computational cost (followed by RELU activation), while the second layer restores the C channels and sent to sigmoid activation function. r is the compression ratio. The mathematical expression is given by Eq. 3:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)). \quad (4)$$

The final output of model can be expressed as:

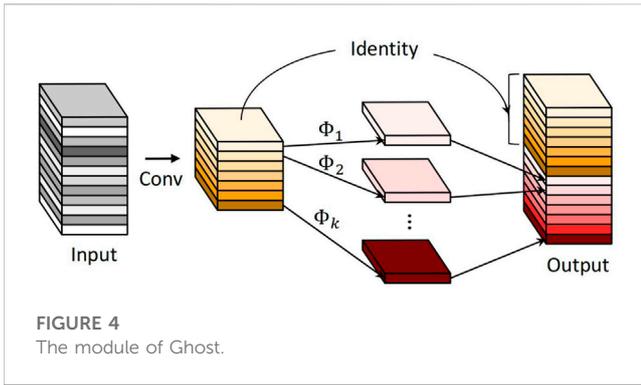
$$\tilde{x}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c. \quad (5)$$

Where $\tilde{X}_c = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_C]$ and $F_{scale}(u_c, s_c)$ is the channel-wise multiplication between $u_c \in \mathbb{R}^{H \times W}$ and s_c .

3 Proposed method

3.1 Network architecture

The network's architecture for infrared target detection is illustrated in Figure 3. The network consists of four main modules. The input module applies Mosaic data augmentation and Focus slicing to the 640×640 input image. For multi-scale feature extraction and fusion, the CSP-Darknet53 network is employed as the backbone network in this study. In the backbone module, Global Channel Attention (GCA) is added to obtain more feature information of target channels and alleviate the network's capability of small targets detection. In the neck module, a channel space attention mechanism (CPA) is built to enhance the



awareness of target location and avoid anchor frame bias and invalid feature information. In this paper, Ghost convolution is used in the neck module to replace the residual unit of CBS modules. CBS is composed of three parts: convolution layer, batch processing layer and SiLu loss function. This module can effectively compensate for the amount of parameters brought by the increased attention mechanism and reduce the redundancy of parameters of neck network.

The main improvement of this article is that we use Ghost convolution to replace the resunit unit of the convolution layer of the deep residual network layer in YOLOv5s, it can effectively reduce the amount of parameters of network without losing accuracy. In order to better obtain the global channel characteristics of the target and provide richer target feature information, the global channel attention (GCA) is added to the upper sampling layer. the detection rate of network is elevated by enhancing the characteristics of the overall goal. The PCA space attention mechanism is added to the output end to obtain more accurate target location information.

3.2 GhostNet

Kai Han et al. [26] introduced the Ghost module as a solution for generating more feature maps from inexpensive operations, which is an essential characteristic of CNN. The general 1×1 Ghost convolution is used to compress the number of channels of the input image, and then the deep separable convolution is used to obtain more feature images, and various feature images are contacted to form a new output. The Ghost bottleneck stacks multiple Ghost modules, which leads to the creation of the lightweight GhostNet.

The Ghost module is depicted in Figure 4 and the cheap operation is denoted by Φ .

3.3 Channel space attention (CPA)

The channel spatial attention is aimed at emphasizing the target's location information by assigning higher weights to the spatial information. To increase the accurate of position information, the input image is pooled in the width and height directions for obtain the image's width and height features.

$$D_s^h(w) = \frac{1}{W} \sum_{i=0}^W X_s(h, i), \quad (6)$$

$$D_s^w(w) = \frac{1}{H} \sum_{i=0}^H X_s(j, w). \quad (7)$$

The paper proposes splicing the feature images of the width and height of FOV, followed by a convolution module to reduce their dimension to the original C/r. The resulting feature maps, denoted as F_1 , passed through batch normalization and sent to sigmoid function to obtain a channel attention map f with the structure of $C/r \times 1 \times (W \times H)$. The formula can be expressed as follows.

$$f = \beta(F_1[D^h, D^w]). \quad (8)$$

Then convolution the image features f obtain the feature graph f^h and f^w which are the same as former channel number. The attention weights of feature g^h and g^w are obtained according to sigmoid function. The formula is as follows.

$$g^h = \partial(F_h(f^h)), \quad (9)$$

$$g^w = \partial(F_w(f^w)). \quad (10)$$

Finally, through multiplication and weighting calculation on the original feature, the feature of attention in the width and height direction is calculated. The improved channel space attention is displayed in Figure 5.

3.4 Global channel attention (GCA)

Because CAM uses maximum pooling to extract features, which is not good for shallow small target features, and is not easy to extract small target features, resulting in small target easy to miss detection. Therefore, an improved CAM method, global channel attention (GCA), is proposed. This method uses global pooling instead of maximum pooling operation to alleviate the network's filtering of small target information, and retain small target information as much as possible. Compared with traditional CAM, the proposed GCA uses a four-layer feature sharing perceptron, which is one more layer than CAM, to fully obtain target feature information and further fuse features. The specific flow chart is shown in Figure 6.

3.5 Loss functions

The loss function is crucial for gradient calculations in deep learning. The proposed method utilizes a loss function consisting of L_{pos} , L_{con} , and L_{cla} . For a detailed introduction of the loss function, please refer [17].

$$L_{pos} = \lambda_{cord} \sum_{i=1}^{S^2} \sum_{j=1}^B \mathbf{1}_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]. \quad (11)$$

Where $\mathbf{1}_{ij}^{obj}$ means there is a target in the image; The presence or absence of targets in the image is represented by 1 and 0; $x_i, y_i, \hat{x}_i, \hat{y}_i$ represent the true coordinates of centre point and predicted one; λ_{cord} defines the weight parameter.

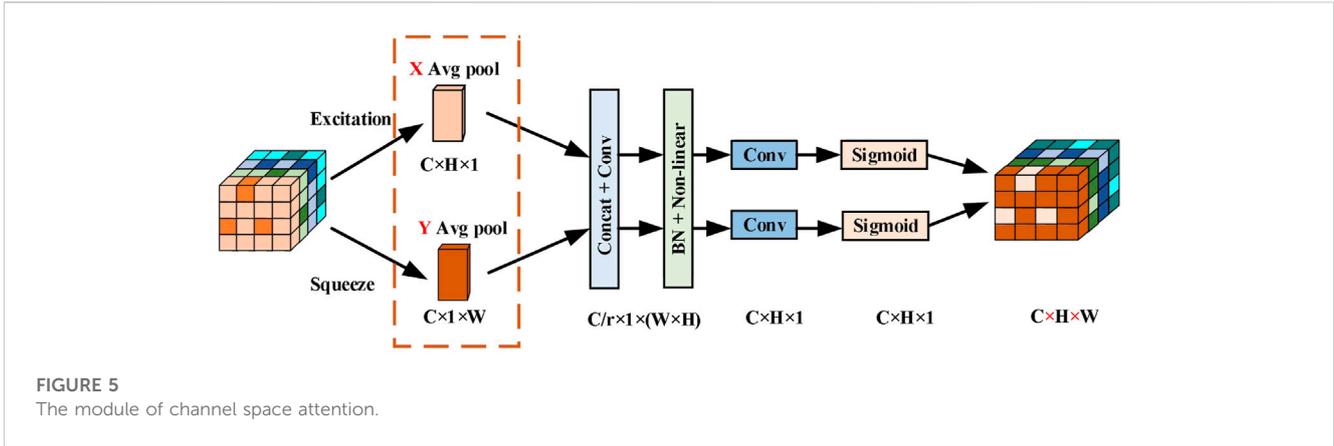


FIGURE 5 The module of channel space attention.

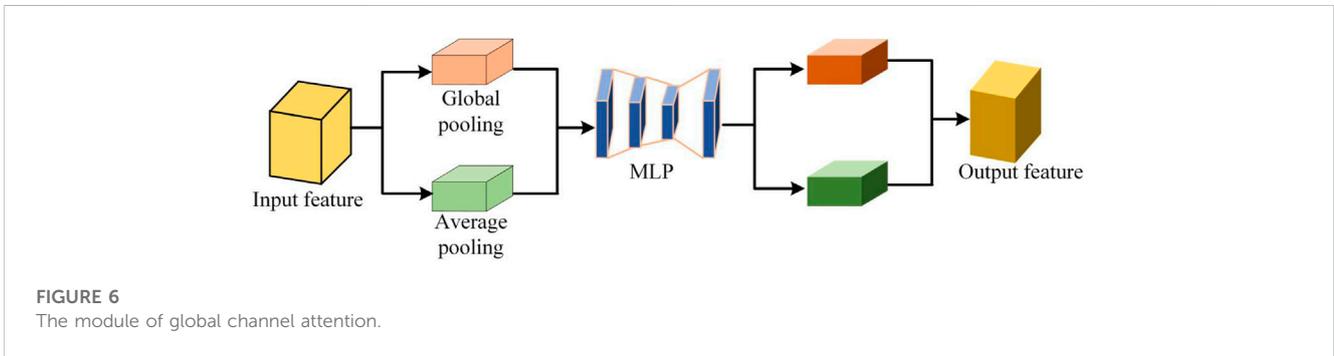


FIGURE 6 The module of global channel attention.

$$L_{con} = \sum_{i=1}^{S^2} \sum_{j=1}^B \mathbf{1}_{i,j}^{obj} [-C_i \log(\hat{C}_i)] + \lambda_{nobj} \sum_{i=1}^{S^2} \sum_{j=1}^B \mathbf{1}_{i,j}^{nobj} [-(1 - C_i) \log(1 - \hat{C}_i)]. \quad (12)$$

Where C_i is the true target's confidence, \hat{C}_i is the predicted target's confidence, and $\mathbf{1}_{i,j}^{nobj}$ indicates that there are objects in the image.

$$L_{cla} = \sum_{i=1}^{S^2} \sum_{j=1}^B \mathbf{1}_{i,j}^{obj} \times \sum_{c \in ccla} [-p_i(c) \log(\hat{p}_i(c)) - (1 - p_i(c)) \log(1 - \hat{p}_i(c))]. \quad (13)$$

Where $\hat{p}_i(c)$ and $p_i(c)$ represent the predicted and true probabilities.

The final loss function can be written as:

$$L_{total} = L_{pos} + L_{con} + L_{cla}. \quad (14)$$

4 Experiment and result analysis

4.1 Experimental environment and parameters

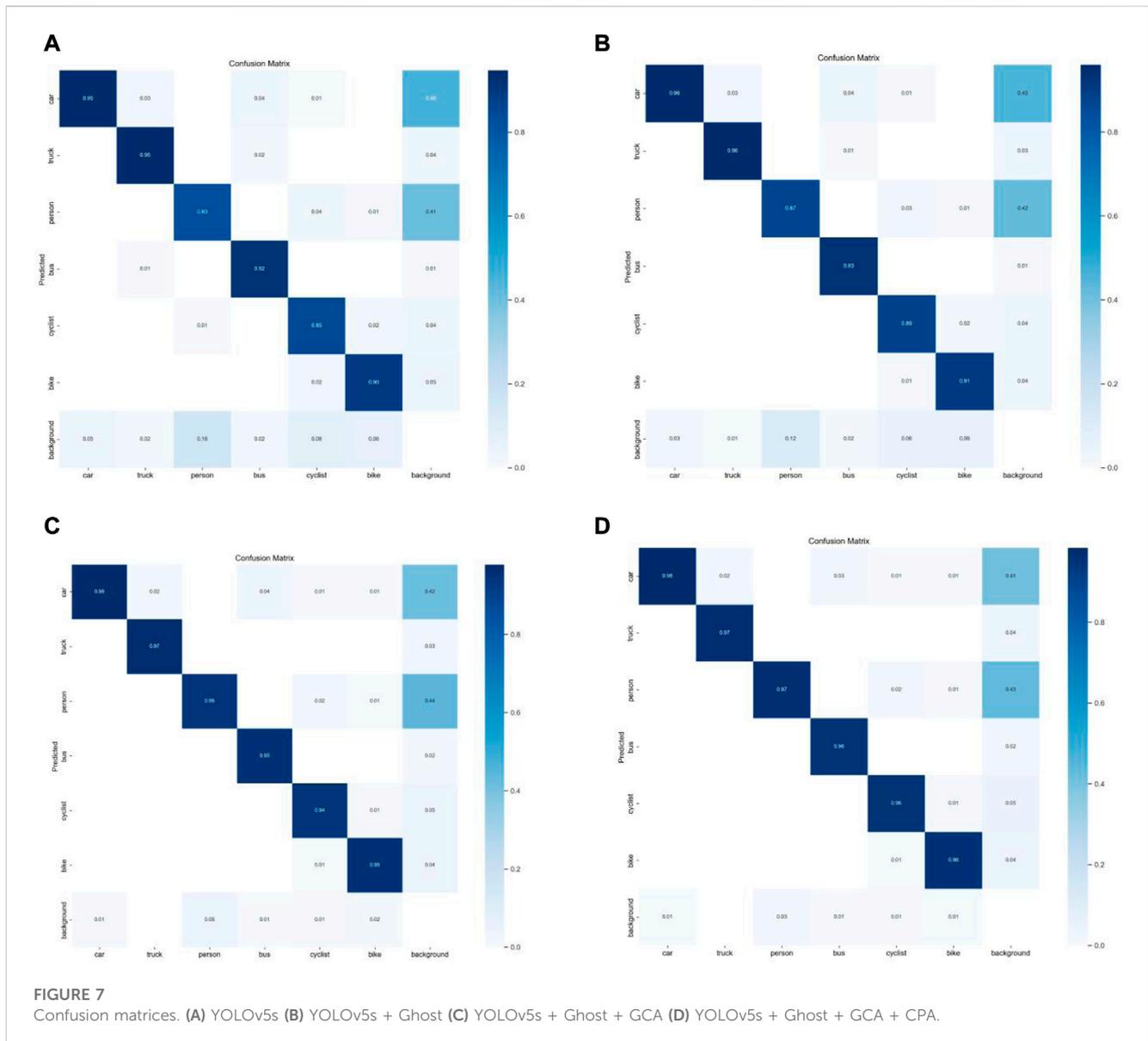
The specific experimental parameters are configured as Table 1. We conduct network training under Ubuntu environment and GPU was invoked.

TABLE 1 The parameters of experiment.

Parameter	Disposition
Environment	Ubuntu
Redaction language	Python 3.8
CPU	Intel i9-10900K
GPU	RTX 3090
Memory	64G
Epoch	200

4.2 Datasets

We used the Yantai IRay Technology Co., Ltd. dataset. This database is created by a drone equipped with a professional infrared camera. It takes pictures of human-vehicle targets in a variety of different scenarios and obtains a great quantity infrared images. The flying height of the drone is about 10–15 m higher than the ground when shooting. Most of the collected data are small infrared targets, which have a wider monitoring perspective in the monitoring range, which is better than the deployment of instruments in ordinary security scenarios. Yantai IRay Technology Co., Ltd. used the inherent 640*512 resolution to collect more than 10000 infrared images in different viewing angles, and then marked the pedestrians, cars, buses, bicycles, cyclists, and trucks that appeared in the images.



4.3 Experiment and analysis

We conducted experiments in this section to directly demonstrate the effect of each module. YOLOv5s was used as the baseline, and Ghost, GCA, and CPA were separately added to the neural network models. According to different experimental results, the effects of different improved method were obtained.

According to the experimental results in Figure 7, various fusion models improve the detection efficiency of different categories, especially for people and cyclists.

To intuitively measure the effect of various improved methods relative to the traditional YOLOv5s algorithm, the precision-recall curves of different methods were shown in Figure 8. With the addition of network, the precision and recall of model have been significantly raised. The precision-recall curve of the final scheme has the largest area surrounded by the

axis, which verifies that the fusion neural network has the best performance. Compared with YOLOv5, the final fusion model's mAP increased by 7.36%, Precision increased by 7.89%, Recall increased by 6.06% and F1 increased by 7%, the results of each improved module are summarized in Table 2.

The infrared detection results of various methods are shown in Figure 9. Compared different improved models with original YOLOv5s version, the detection accuracy of people and vehicles in scenario 1 does not significantly improved by YOLOv5s + Ghost model. The YOLOv5s + Ghost + GCA model has increased by 6% and 7% respectively, while YOLOv5s + Ghost + GCA + CPA model has increased by 8% and 9% respectively. In Scenario 2 and Scenario 3, The YOLOv5s + Ghost and YOLOv5s + Ghost + GCA have greatly improved the detection accuracy compared with YOLOv5s, while the YOLOv5s + Ghost + GCA + CPA has

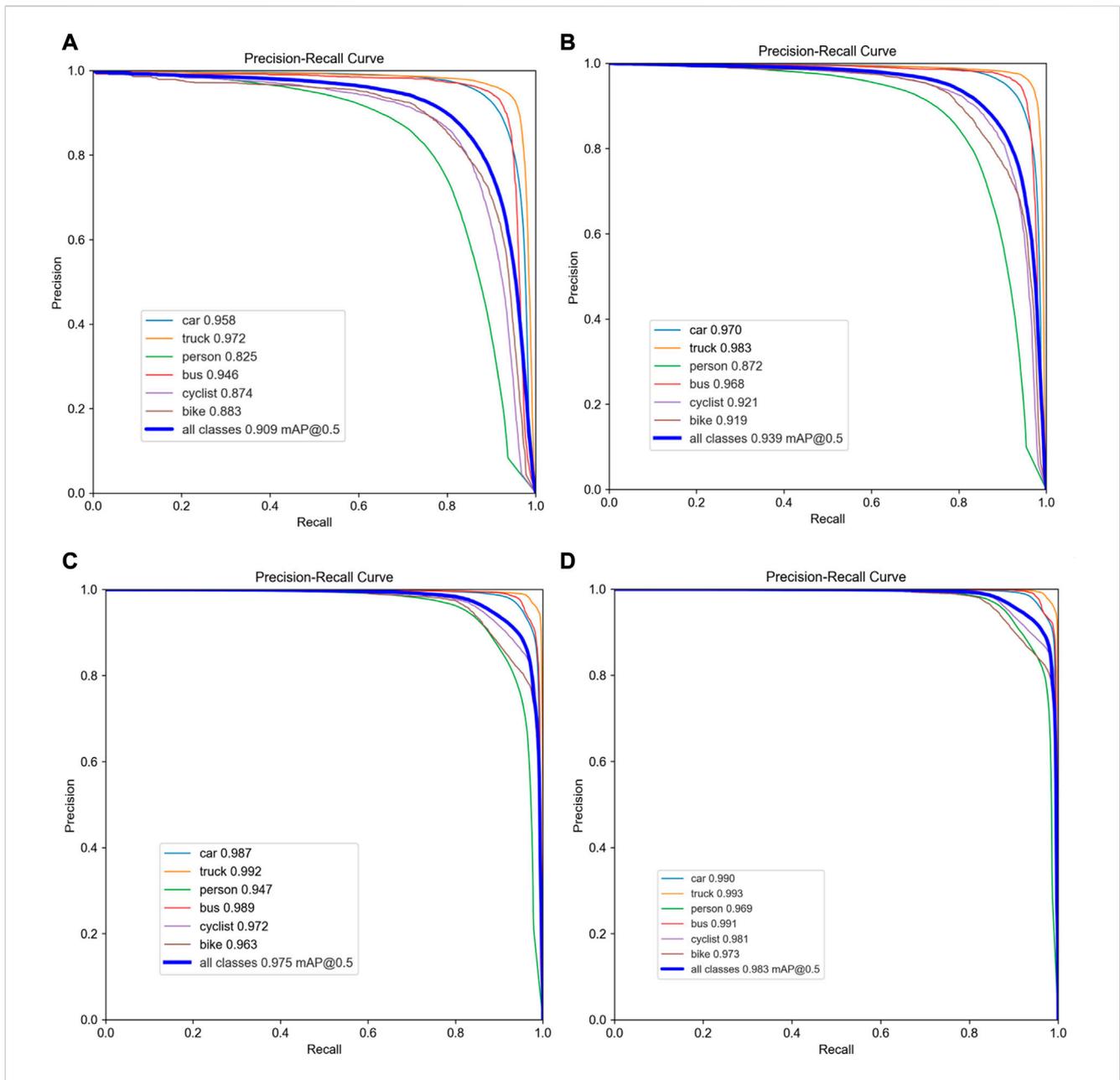


FIGURE 8 P-R curves of various model: (A) YOLOv5s (B) YOLOv5s + Ghost (C) YOLOv5s+Ghost +GCA (D) YOLOv5s+Ghost+GCA+CPA.

TABLE 2 Results of each improved module.

Mode	YOLOv5s	YOLOv5s + Ghost	YOLOv5s + Ghost + GCA	YOLOv5s + Ghost + GCA + CPA
mAP	0.9094	0.9388	0.9751	0.9830
F1	0.87	0.89	0.93	0.94
Recall	0.8545	0.8792	0.9099	0.9151
Precision	0.8858	0.9115	0.9466	0.9647

not further considerably improved. In the cluttered background of Scenario 4, the YOLOv5s + Ghost + GCA + CPA model can detect the distance of cars and trucks, and the detection accuracy

of small targets in the distance has increased 20%. While the nearby targets has not improved much, because the accuracy is already high.

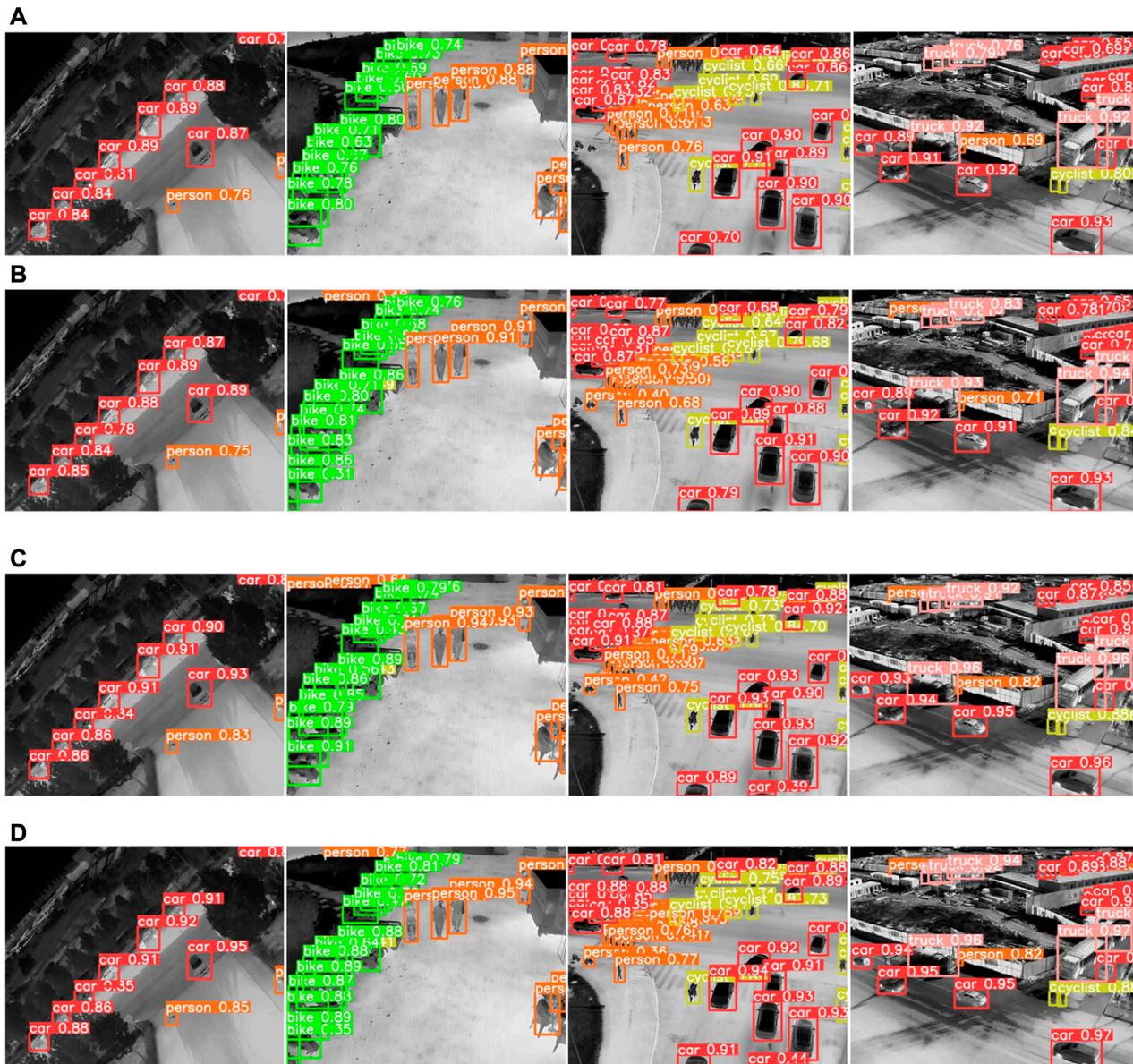


FIGURE 9

The results of various methods. (A) YOLOv5s (B) YOLOv5s + Ghost (C) YOLOv5s+Ghost +GCA (D) YOLOv5s+Ghost+GCA+CPA.

5 Conclusion

This study established a target detection method that integrates fusion neural network to YOLOv5s for infrared images. At first, the Ghost is used to the convolution layer of the deep residual network layer in YOLOv5s, it can decrease the complexity of the network and reduce the calculation. Then, the global channel attention (GCA) is added to the upper sampling layer, the detection rate of method is elevated by enhancing the characteristics of the overall goal. Also, the channel spatial attention (CPA) is incorporated into module for obtain more accurate target location information. By training and testing of

infrared data set taken by UAV, the accuracy rate of detection based on YOLOv5s and fusion neural network is 96.47%, the recall rate is 91.51%, and the F1 score is 94%, which is 7% higher than YOLOv5s. The results prove that our method raised the detection rate of infrared images.

Data availability statement

The datasets for this study can be found at http://iray.iraytek.com:7813/apply/Aerial_mancar.html/. further inquiries can be directed to the corresponding authors.

Author contributions

GF was involved in methodology conceptualization, writing the original draft, reviewing and editing the manuscript. LH is responsible for investigation and data analysis of the work, SH is responsible for validation and editing the manuscript. LL is responsible for validation, data visualization, and reviewing and editing the manuscript. YW and XZ is responsible for reviewing. All authors contributed to the article and approved the submitted version.

Funding

The authors thank the independent research project of Fund project for basic scientific research expenses of central Universities (J2022-024) and the independent research project of Fund project for basic scientific research expenses of central Universities (J2022-07). Open research project of Sichuan province engineering technology research center of general aircraft maintenance (KFXM2021001).

References

- Kang K, Li H, Yan J, Yang B, Xiao T, et al. T-CNN: Tubelets with convolutional neural networks for object detection from videos. *IEEE Trans Circuits Syst Video Tech* (2018) 28(10):2896–907. doi:10.1109/tcsvt.2017.2736553
- Wei ZW, Qiao YL, Lu X. Heat diffusion embedded level set evolution for infrared image segmentation. *IET Image Proc* (2019) 14(2):267–78. doi:10.1049/iet-ipr.2018.6629
- Tang T, Zhou S, Deng Z, Zou H, Lei L. Vehicle detection in aerial images based on region convolutional neural networks and hard negative example mining. *Sensors* (2017) 17:336. doi:10.3390/s17020336
- Mahmood MT, Ahmed SRA, Ahmed MRA. Detection of vehicle with Infrared images in Road Traffic using YOLO computational mechanism. *IOP Conf Ser Mater Sci Eng* (2020) 928:022027. doi:10.1088/1757-899x/928/2/022027
- Li B, Xiao C, Wang L, Wang Y, Lin Z, Li M, et al. Dense nested attention network for infrared small target detection. arXiv 2021, arXiv:2106.00487.
- Dai X, Yuan X, Wei X. TIRNet: Object detection in thermal infrared images for autonomous driving. *Appl Intell* (2021) 51:1244–61. doi:10.1007/s10489-020-01882-2
- Liu Q, Zhuang J, Ma J. Robust and fast pedestrian detection method for far-infrared automotive driving assistance systems. *Infrared Phys Technol* (2013) 60:288–99. doi:10.1016/j.infrared.2013.06.003
- El Maadi A, Maldague X. Outdoor infrared video surveillance: A novel dynamic technique for the subtraction of a changing background of ir images. *Infrared Phys Technol* (2007) 49:261–5. doi:10.1016/j.infrared.2006.06.015
- Zhang H, Luo C, Wang Q, Kitchin M, Parmley A, Monge-Alvarez J, et al. A novel infrared video surveillance system using deep learning based techniques. *Multimed Tools Appl* (2018) 77:26657–76. doi:10.1007/s11042-018-5883-y
- Kim J, Hong S, Baek J, Lee H. Autonomous vehicle detection system using visible and infrared camera. In: Proceedings of the 2012 12th International Conference on Control, Automation and Systems; October 2012; Jeju, Korea. p. 630–4.
- Xu W, Zhong S, Yan L, Wu F, Zhang W. Moving object detection in aerial infrared images with registration accuracy prediction and feature points selection. *Infrared Phys Technol* (2018) 92:318–26. doi:10.1016/j.infrared.2018.06.023
- Liu HK, Zhang L, Huang H. Small target detection in infrared videos based on spatio-temporal tensor model. *IEEE Trans Geosci Remote Sens* (2020) 58(12):8689–700. doi:10.1109/tgrs.2020.2989825
- Li J, Yang P, Cui W, Zhang T. A cascade method for infrared dim target detection. *Infrared Phys Tech* (2021) 117:103768. doi:10.1016/j.infrared.2021.103768

Acknowledgments

The authors would like to thank Yantai IRay Technology Co., Ltd. for helpful feedback and datasets.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Moradi S, Moallem P, Sabahi MF. Fast and robust small infrared target detection using absolute directional mean difference algorithm. *Signal Process* (2020) 177:107727. doi:10.1016/j.sigpro.2020.107727
- Bai X, Bi Y. Derivative entropy-based contrast measure for infrared small-target detection. *IEEE Trans Geosci Remote Sens* (2018) 56(4):2452–66. doi:10.1109/tgrs.2017.2781143
- Zhang Z, Liu Y, Liu J, Wen F, Zhu C. AMP-Net: Denoising-Based deep unfolding for compressive image sensing. *IEEE Trans Image Process* (2021) 30:1487–500. doi:10.1109/tip.2020.3044472
- Sun M, Zhang H, Huang Z, Luo Y, Li Y. Road infrared target detection with I-YOLO. *IET Image Process* (2021) 16(1):92–101. doi:10.1049/ipr2.12331
- Kim J, Huh J, Park I, Bak J, Kim D, Lee S. Small object detection in infrared images: Learning from imbalanced cross-domain data via domain adaptation. *Appl Sci* (2022) 12(21):11201. doi:10.3390/app122111201
- Ye J, Yuan Z, Cheng Q, LiCAA-Yolo X. CAA-YOLO: Combined-Attention-Augmented YOLO for infrared ocean ships detection. *Sensors* (2022) 22(10):3782. doi:10.3390/s22103782
- Ma J, Tang L, Xu M, Zhang H, Xiao G. STDFusionNet: An infrared and visible image fusion network based on salient target detection. *IEEE TRANSACTIONS INSTRUMENTATION MEASUREMENT* (2021) 70. doi:10.1109/TIM.2021.3075747
- Wu D, Cao L, Zhou P, Ning L, Yi L, Wang D. Infrared small-target detection based on radiation characteristics with a multimodal feature fusion network. *Remote Sensing* (2022) 14(15):3570. doi:10.3390/rs14153570
- Li C, Zhang Y, Gao G, Liu Z, Liang L. Context-aware cross-level attention fusion network for infrared small target detection. Zhengzhou, China: Zhongyuan University of Technology, School of Electronic and Information Engineering (2022).
- Zhou L, Gao S, Wang S, Zhang H, Liu R, Liu J. IPD-Net: Infrared pedestrian detection network via adaptive feature extraction and coordinate information fusion. *Sensors* (2022) 22(22):8966. doi:10.3390/s22228966
- Fan YC, Qiu QL, Hou SH, Li YH, Xie JX, Qin MY, et al. Application of improved YOLOv5 in aerial photographing infrared vehicle detection. *ELECTRONICS* (2022) 11(15):2344. doi:10.3390/electronics11152344
- Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: Proceedings of the 13th European Conference CoRR; September 2014; Zurich, Switzerland.
- Han K, Wang Y, Tian Q, Guo J, Xu C, Xu C. GhostNet: More features from cheap operations. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; June 2020; Seattle, WA, USA.