

Retraction

Retracted: Research on Dual Mode Target Detection Algorithm for Embedded Platform

Complexity

Received 30 January 2024; Accepted 30 January 2024; Published 31 January 2024

Copyright © 2024 Complexity. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] L. Zhang, S. Wang, H. Sun, and Y. Wang, "Research on Dual Mode Target Detection Algorithm for Embedded Platform," *Complexity*, vol. 2021, Article ID 9935621, 8 pages, 2021.

Research Article

Research on Dual Mode Target Detection Algorithm for Embedded Platform

Li Zhang ^{1,2}, Shaoqiang Wang ³, Hongwei Sun ^{1,2} and Yifan Wang ⁴

¹College of Computer Science and Engineering, Weifang University of Science and Technology, Weifang 262700, China

²Blockchain Laboratory of Agricultural Vegetables, Weifang University of Science and Technology, Weifang 262700, China

³School of Computer and Communication Engineering, China University of Petroleum(East of China), Qingdao 266000, China

⁴Shandong University of Traditional Chinese Medicine, Jinan 250355, China

Correspondence should be addressed to Hongwei Sun; sunhongwei_doctor@wfust.edu.cn

Received 31 March 2021; Revised 24 April 2021; Accepted 2 May 2021; Published 12 May 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Li Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at the problem that the embedded platform cannot meet the real-time detection of multisource images, this paper proposes a lightweight target detection network MNYOLO (MobileNet-YOLOv4-tiny) suitable for embedded platforms using deep separable convolution instead of standard convolution to reduce the number of model parameters and calculations; at the same time, the visible light target detection model is used as the pretraining model of the infrared target detection model and the infrared target data set collected on the spot is fine-tuned to obtain the infrared target detection model. On this basis, a decision-level fusion detection model is obtained to realize the complementary information of infrared and visible light multiband information. The experimental results show that it has a more obvious advantage in detection accuracy than the single-band target detection model while the decision-level fusion target detection model meets the real-time requirements and also verifies the effectiveness of the above algorithm.

1. Introduction

Target detection [1] is an important research content in the field of computer vision. With the rapid development of deep learning, new detection algorithms continue to emerge in the visible light environment. Among them, it is mainly divided into two-stage [2, 3] detection model and one-stage [4] detection model. The two-stage detection model mainly includes the R-CNN series of algorithms, which greatly improves the detection accuracy by generating suggested regions; one-stage detection models mainly include SSD [5, 6] (single shot multibox detector) series and YOLO [7] (you only look once) series, adopting a one-step framework of global regression and classification, while sacrificing certain accuracy, the detection speed gets a big improvement. The above two detection models are based on preset anchor points. Although the detection accuracy and speed have been greatly improved, the limitations of their preset anchor points also hinder the optimization and innovation

of the detection model. To solve these problems, anchor-free detection models have recently been proposed, such as CornerNet [8, 9] and Internet [10], which can complete target detection without presetting anchor points, bringing new ideas for target detection model innovation.

The abovementioned target detection algorithm in the visible light environment is very dependent on the condition of sufficient light and cannot meet the target detection requirements in the insufficient light scene. The infrared camera is based on the target's reflection of infrared light and the target's thermal radiation. It is not affected by the light intensity conditions and can cover most scenes with insufficient light. However, compared with visible light images, infrared images often have problems such as lower resolution and blurred edges of objects, which also greatly reduces the features that can be extracted from infrared images. Therefore, by combining the advantages of the high resolution of visible light images and rich target texture information, multiband information complementation is

realized, thereby improving the performance of the target detection model.

Although some visible light target detection models can achieve high accuracy due to the limitations of the computing power and memory resources of the embedded platform, it is difficult to migrate algorithms to the embedded platform and cannot adapt to the industry's requirements for real-time and portability of target detection algorithms. In response to this problem, Qi and Lu proposed the mobile net model, which uses a deeply separable convolution method to reduce network weight parameters, and in 2019, they proposed a lightweight attention model by integrating the MobileNetV1 [11] deeply separable volume, MobileNetV2 [12, 13] linear bottleneck inverse residual structure, and MobileNetV3 [14, 15] Magnet, while greatly reducing model parameters and further optimizing the target feature extraction and classification network.

To ensure that the decision-level fusion target detection model achieves a good balance of accuracy and speed on the embedded platform, this paper improves the YOLOv4-tiny target detection model and uses the deep separable convolution method to replace the standard convolution to ensure the computing power of the embedded platform and at the same time to reduce the size of the network model; at the same time, using the advantages of multiband information complementation, a decision-level fusion detection model is proposed, which greatly reduces the requirements of the target detection algorithm for actual industrial scenes and significantly improves the robustness of the detection algorithm.

2. MNYOLO Target Detection Network

The MNYOLO network continues to use the idea of YOLO target detection, using the entire image as the input of the entire network, without the need to generate suggested regions, and in the output layer, the position and category of the bounding box are obtained by using the regression idea, and then the redundant bounding box is removed by the nonmaximum suppression algorithm, and the final prediction result is obtained. The whole process is that the detection network directly performs end-to-end prediction, and the detection speed is relatively high.

The YOLOv4 algorithm optimizes the YOLOv3 model from the perspective of data preprocessing, backbone network, training mode, and activation function so that the detection model achieves a good balance between detection speed and detection accuracy. The backbone network CSPDarkNet of YOLOv4 [16] absorbs the advantages of CSPNet [17–19] (cross stage partial network), and the CSP module is added to the backbone network DarkNet53 of YOLOv3, which divides the shallow feature mapping into two parts and then merges them through a cross-level structure, to maintain detection accuracy, reduce computing bottlenecks, and reduce memory costs while lightweight the network. Besides, YOLOv4 absorbs the advantages of PANet [20], spreads the semantic information of high-level features to the low-level network, and merges with the high-resolution information of shallow features to improve

the detection effect of small target objects; then, the low-level information is propagated to the high-level network so that the feature map can obtain richer semantic information, and finally, the feature maps of different layers are merged for prediction.

The YOLO algorithm predicts four coordinate parameters for each detection frame (t_x, t_y, t_w, t_h) , and the predicted rectangular frame is calculated by calculating the offset of the network. (c_x, c_y) is the distance between the grid where the center coordinates of the prediction box are located. (t_x, t_y) is the coordinate of the center point of the predicted bounding box. The function is the activation function for boundary prediction, and the coordinates are uniformly normalized between 0 and 1. (t_w, t_h) is the width and height between the predicted bounding boxes, and (p_w, p_h) is the width and height of the selected anchor point. Figure 1 shows the principle of boundary box prediction.

As a simplified version of YOLOv4, YOLOv4-tiny improves the speed of target detection by sacrificing certain accuracy. The feature extraction stage does not use the Mish activation function, and the feature enhancement stage only uses a feature pyramid, which makes YOLOv4-tiny much faster than YOLOv4 in terms of loading model and detection speed. The MNYOLO network decomposes the standard convolution in the YOLOv4-tiny network into deep convolution and point-by-point convolution in deep separable convolutions, which greatly reduces the number of parameters and calculations of the detection model. Figure 2 shows the operation decomposition diagram.

$$G_s = D_k \times D_k \times M \times N \times D_f \times D_f, \quad (1)$$

$$G_s = D_k \times D_k \times M \times D_f \times D_f + M \times N \times D_f \times D_f.$$

Figure 2(a) shows the standard convolution of $D_k \times D_k \times M$, Figure 2(b) shows the channel-wise convolution of $1 \times 1 \times M$, and Figure 2(c) shows the pointwise convolution of $1 \times 1 \times M$.

The calculation amount of depth separable convolution is as follows:

$$P = \frac{D_k \times D_k \times M \times D_f \times D_f + M \times N \times D_f \times D_f}{D_k \times D_k \times M \times N \times D_f \times D_f} = \frac{1}{N} + \frac{1}{D_k^2}. \quad (2)$$

When $D_k = 3$, the calculation amount of the depth separable convolution is about 1/9 of the standard convolution.

Based on the above content, the YOLOv4-tiny [21] network is improved, and the standard convolution is replaced by the depth separable convolution to reduce the calculation parameters and size of the model; in the forward propagation of the network, the detailed information of the target is seriously lost after multilayer convolution. To improve the ability to express the target, the high-level feature map is convolved, and the larger feature map is up sampled to strengthen the target's feature representation. Figure 3 shows the MNYOLO network structure diagram.

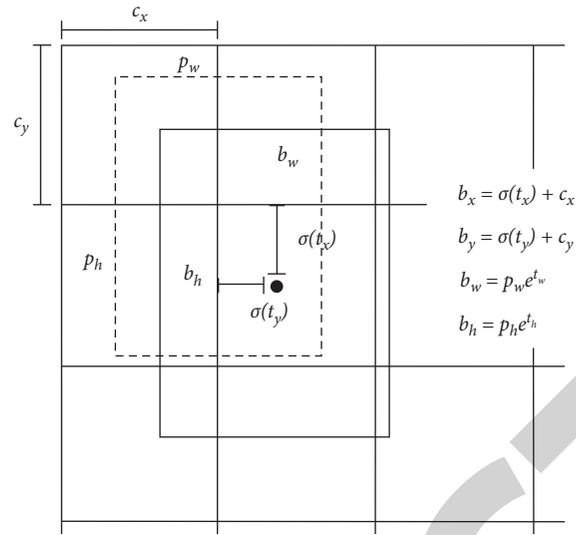


FIGURE 1: Principle of boundary box prediction.

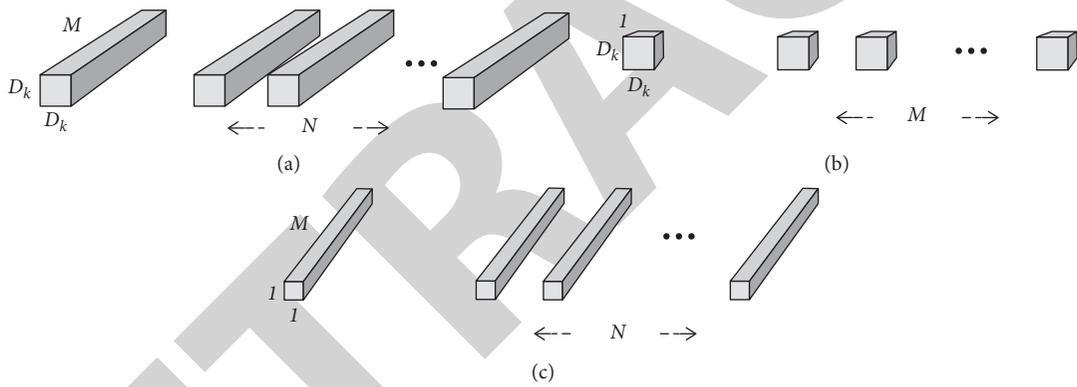


FIGURE 2: Depth separable convolution: (a) standard convolution; (b) channel-by-channel convolution; (c) pointwise convolution.

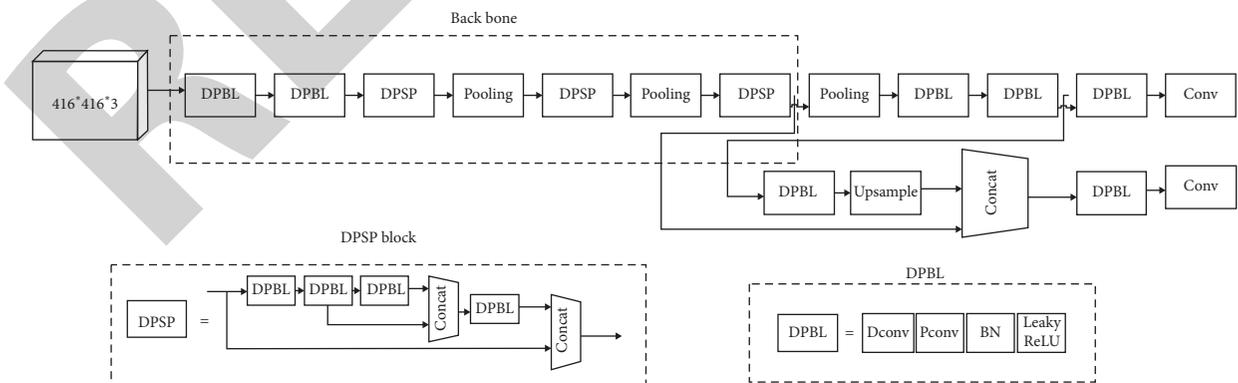


FIGURE 3: MNYOLO network structure diagram.

Compared with other networks, the proposed network has fewer calculations and smaller models.

3. Object Detection Model

Due to the different types of detection targets, the classification layer of the visible light target detection model and the infrared target detection model must be redesigned. In the forward reasoning of the network, the feature maps of the two channels are extracted for prediction, and the size of the feature maps are $13*13*18$ and $26*26*18$, respectively. The pretraining model of the visible light target detection model is generally based on the model trained by ImageNet and MS COCO, and the rapid convergence of the model is achieved through the method of model parameter transfer.

After 3500 iterations of training, the training loss of the visible light target detection model is about 0.5, and its loss transformation curve is shown in Figure 4.

The change curve of the intersection ratio between the predicted bounding box and the actual bounding box is shown in Figure 5. The abscissa represents the number of iterations, and the ordinate represents the intersection ratio between the predicted bounding box and the actual bounding box. With the increase in the number of iterations, the intersection ratio shows an upward trend, eventually approaching 1; at this point, it shows that the predicted bounding box in the visible light scene is very close to the actual bounding box, which meets the training requirements.

In the design of the infrared pretraining model, the initial model is obtained by initializing the parameters according to the model structure. At the same time, to make full use of the target detection capabilities of the visible light model, the parameters of each convolutional layer in the visible light model are transferred to the corresponding convolutional layer of the infrared model to realize parameter sharing. The pretraining model for infrared object detection is extracted from the above visible light detection model and fine-tuned on the infrared data set, thereby obtaining an infrared object detection model based on deep learning. The loss transformation curve is shown in Figure 6.

With the increase in the number of iterations, the intersection of the infrared detection model shows an upward trend compared with the whole and finally approaches 1, which shows that the predicted bounding box is very close to the actual bounding box in the infrared scene, which meets the training requirements. The details are shown in Figure 7.

The target detection model generally uses the mAP [22] (mean Average Precision) index for evaluation, which is the average value of the average detection accuracy (Average Precision, AP) of multiple types of objects. Since the recognition classification of the training model is single class, here $mAP = AP$. The vehicle test sets are tested in different scenarios of visible light and infrared, respectively, and the results are shown in Table 1.

Experimental results show that the proposed model is higher in detection accuracy than YOLOv4-tiny, but compared with YOLOv4, there is a certain gap in detection accuracy. This is due to the simplification of the network

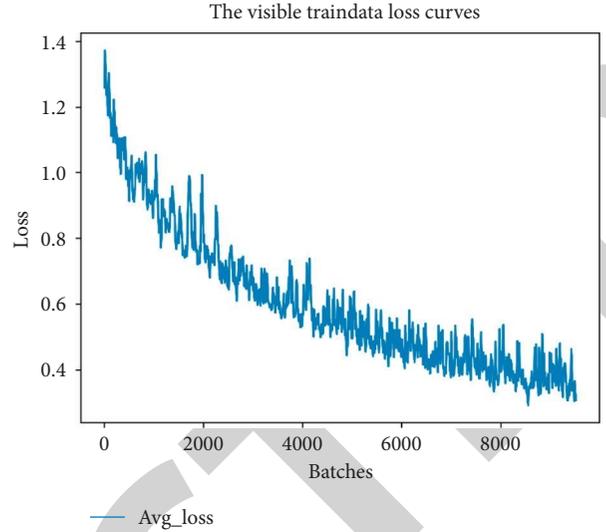


FIGURE 4: Training loss curve of the visible light detection model.

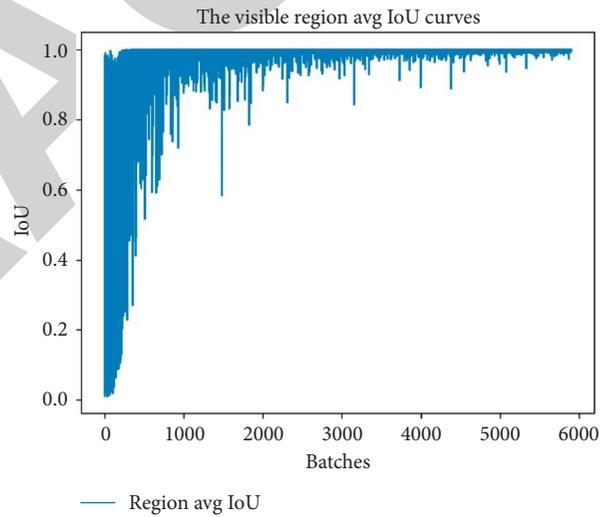


FIGURE 5: Curve of change of intersection ratio of the visible light detection model.

structure and the reduction of network parameters, resulting in a certain loss of accuracy.

4. Decision-Level Image Fusion Detection Model

To improve the target detection performance and reduce the requirements of the detection model on the external environment, the visible light and infrared target detection models are combined to achieve multimodal target detection.

4.1. Model Design. After the infrared image and the visible light, image is collected, they are registered and sent to the visible light scene target detection model and the infrared

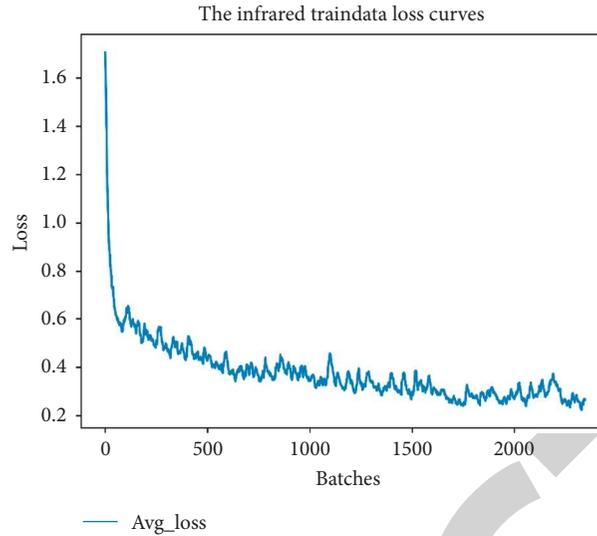


FIGURE 6: Training loss curve of the infrared detection model.

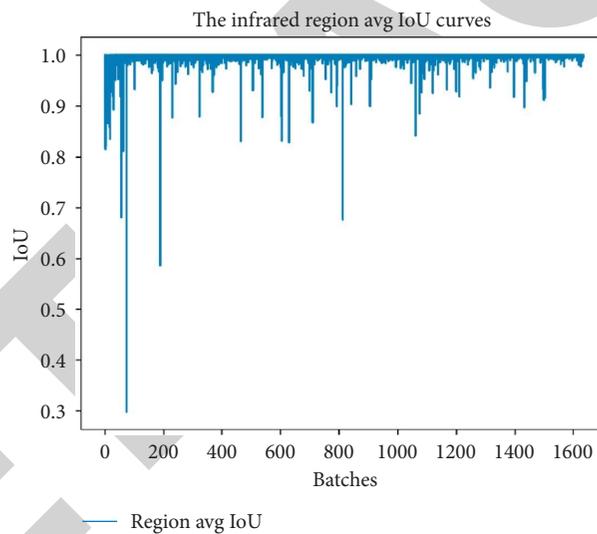


FIGURE 7: Curve of change of intersection ratio of infrared detection mode.

TABLE 1: Visible light.

Model	AP (%)	Size (MB)
YOLOv4	93.28	245
YOLOv4-tiny	75.35	23.1
Ours (visible)	84.82	16

scene target detection model, respectively, and the detection results are subjected to decision-level image fusion. The same target is detected by infrared and visible light detection models, and two detection photos with confidence are obtained, respectively, and the detection regression frame with the highest confidence is displayed. Figure 8 shows the determination of the target bounding box and target confidence.

4.2. Image Registration. Image registration is a process of superimposing two or more images of the same scene obtained from different perspectives or sensors. Assuming that the infrared camera and the visible light camera are placed side by side, make sure that the optical axis is parallel, and the distance between the center of the camera lens is 12 cm. In this case, for a scene 100 meters away from the camera, the parallax between the two cameras can be ignored. The only difference is the difference between the translation, scaling, and rotation of the two cameras before the image is obtained. The difference between the images can be eliminated by an affine transformation.

4.3. Decision-Level Fusion. The decision-level fusion is carried out based on dual-band region merging, and the bounding box of the target with a high overlap rate is

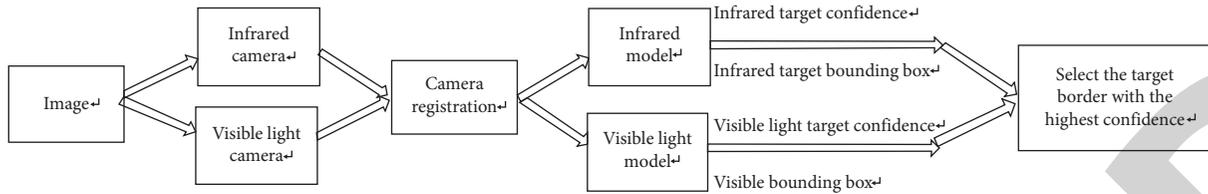


FIGURE 8: Determination of the target bounding box and target confidence.

TABLE 2: Infrared.

Model	AP (%)	Size (MB)
YOLOv4	90.39	245
YOLOv4-tiny	72.16	23.1
Ours (infrared)	82.36	16



FIGURE 9: Decision-level fusion model detection effect diagram.

removed from the detection results of the visible light and infrared models. The confidence of the same target is used in the visible and infrared images (visible light, μ_i , where i is the number of targets in the visible image; infrared, η_i , where i is the number of targets in the infrared image) to calculate the weighting coefficient of the fusion image.

Because different models detect the same target, the position and size of the detected target are the same in different scenarios, and the displayed images are displayed differently because of the different wavebands, so they play a complementary role in different wavebands. The weighted fusion of bimodal images with the confidence of the

Input: the total number of detection targets in the visible light image n , the detection target confidence level in the visible light image μ_i (i : a single target)
 The total number of detected targets in the infrared image m , the confidence level of the detected targets in the infrared image η_i (i : a single target)
 Output: visible light image weighting factor α ; infrared image weighting factor β
 Step 1. Normalize the confidence of the detection targets in different images $\mu'_i = \mu_i / \sum_{i=1}^{i=n} (\mu_i + \eta_i)$, $\eta'_i = \eta_i / \sum_{i=1}^{i=n} (\mu_i + \eta_i)$
 Step 2. Calculate the weighted fusion coefficient of visible light image and infrared image $\alpha = \sum_{i=1}^{i=n} \mu'_i$, $\beta = \sum_{i=1}^{i=n} \eta'_i$

ALGORITHM 1: Decision-level model fusion.

TABLE 3: Comparative experiments of different models.

Model	Frame (FPS)	AP (%)
YOLOv4	6	92.2
YOLOv4-tiny	26	76.4
Ours	32	90.3

detection target can not only alleviate the image thermal noise problem in thermal infrared images to a certain extent but also achieve a clearer and more accurate description of the scene information.

5. Analysis of Results

The experimental hardware configuration is NVIDIA's embedded platform Jetson TX2 [23, 24], and the experimental data use the FLIR vehicle thermal infrared data set. Experiments were conducted on the speed and accuracy of the decision-level fusion detection model and compared with other detection models. Table 2 shows the experimental results. Figure 9 shows the effect diagram of using the model in this paper to detect in the test set experiment. Algorithm 1 shows the decision-level model fusion.

As shown in Table 3, the detection speed of the MNYOLO decision-level fusion detection model meets the real-time requirements of detection. Although the detection accuracy is slightly lower than that of the single-mode YOLOv4 model, it is much higher than that of the single-mode YOLOv4-tiny model. As shown in Figures 9(h)–9(j), the visible light detection model fails to detect the vehicle ahead, but the infrared detection model can detect the vehicle ahead, and the confidence of the same target is much higher than that of the visible light detection model. As shown in Figures 9(e)–9(g), although both the visible light detection model and the infrared detection model can detect all vehicles in the scene, the target confidence of the infrared detection model is significantly lower than that of the visible light detection model. Because the ambient temperature during the day is relatively high, it is easy to interfere with infrared imaging. The MNYOLO decision-level fusion [25] detection model can adapt to environmental changes well, make full use of the advantages of the two detection models,

realize the complementary information of different bands, and improve the robustness of the detection model.

6. Conclusion

This paper studies the current mainstream target detection algorithms in the industry. Based on YOLOv4-tiny, the embedded platform has weak computing power and cannot meet the problem of real-time detection of multisource images. Taking advantage of the small amount of calculation of the depth separable convolution, the ordinary convolution in YOLOv4-tiny is replaced with the depth separable convolution; at the same time, it makes full use of the complementary advantages of different waveband detection models to propose a decision-level fusion detection model, which greatly improves the robustness of the detection algorithm and can meet the requirements of the industry for real-time and high accuracy of target detection. However, because the model design is relatively simple, the next step is to study the use of the attention mechanism to further feature a fusion of the backbone network, enhance the information channel, and improve the network recognition rate.

Data Availability

The data used to support the findings of this study are included in the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Nature Science Foundation of China under no. 61973180.

References

- [1] Z. Y. Zhao, "Data link system and technology," *Publishing House of Electronics Industry of China*, vol. 71, no. 8, pp. 20–32, 2015.
- [2] G. X. Li, *Research on Recognition Algorithm of Link16 Data Link Signal*, Xidian University, Xi'an, China, 2014.

- [3] Y. Chen, F. Zhong, and S. X. Guo, "Blind compressed sensing estimation of non-cooperative frequency hopping signal parameters," *Journal of Radar*, vol. 5, no. 5, pp. 531–537, 2016.
- [4] C. J. Lv, *Frequency Hopping Signal Detection and Parameter Estimation Technique Based on Time-Frequency Analysis*, People's Liberation Army Information Engineering University, Zhengzhou, China, 2015.
- [5] Y. Zeng, "Application and performance analysis of RS coding in mobile communication," *Shandong Industrial Technology*, vol. 55, no. 16, pp. 200–201, 2015.
- [6] Q. P. Song, "Research on anti-interference soft spread spectrum algorithm for measurement and control communication," *Aerospace Control*, vol. 31, no. 6, pp. 72–78, 2016.
- [7] Q. Zhan, "Study on the rapid detection and the symbol timing technology of MSK communication," *Applied Mechanics & Materials*, vol. 530–531, pp. 764–767, 2014.
- [8] X. H. Chen, "Research and design of MSK modem," *Journal of Chinese New Communications*, vol. 20, no. 9, pp. 74–75, 2018.
- [9] W. P. Shao, X. Wang, Z. R. Cao, and F. Bai, "Design of lightweight convolutional neural network based on Mobile Net and YOLOv3," *Computer Applications*, vol. 50–53, no. 9, pp. 774–777, 2012.
- [10] W. H. Fu, L. Wang, K. Jia, and G. Z. Lu, "Blind estimation algorithm for frequency hopping parameters based on STFT and SPWVD," *Journal of Huazhong University of Science and Technology*, vol. 32, no. 9, pp. 59–63, 2014.
- [11] Y. Qi and L. X. Lu, "Frequency hopping signal detection based on STFT," *Research on Telecommunication Technology*, vol. 71, no. 6, pp. 35–42, 2013.
- [12] J. He, "Design of frequency hopping signal detection system based on STFT," *Vehicle and Power Technology*, vol. 5, no. 4, pp. 32–35, 2015.
- [13] W. C. Li, Z. W. Tian, B. Wang, and Q. Tang, "Edge detection method based on short time Fourier transform data link signal," 2016.
- [14] X. C. Wang, *Analysis and Research on Multi-Signal Modulation Recognition Technology*, Qingdao University, Qingdao, China, 2015.
- [15] A. E. Sherme, "Anovel method for automatic modulation recognition," *Applied Soft Computing Journal*, vol. 71, no. 121, pp. 453–461, 2011.
- [16] L. M. Zhang, W. J. Yan, Z. G. Zhong, and Q. Ling, "A method for blind recognition of orthogonal space-time block coding(STBC) using higher-order cumulants," *Acta Electronica Sinica*, vol. 44, no. 8, pp. 1258–1264, 2016.
- [17] R. F. Sun, *Digital Modulation Signal Recognition Based on Higher Order Cumulant and Spectral Analysis*, Hangzhou University of Electronic Science and Technology, Hangzhou, China, 2016.
- [18] K. Shi, Y. Tang, X. Liu, and S. Zhong, "Secondary delay-partition approach on robust performance analysis for uncertain time-varying Lurie nonlinear control system," *Optimal Control Applications and Methods*, vol. 38, no. 6, pp. 1208–1226, 2017.
- [19] Q. Jiang, F. Shao, W. Gao, Z. Chen, G. Jiang, and Y.-S. Ho, "Unified no-reference quality assessment of singly and multiply distorted stereoscopic images," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1866–1881, 2019.
- [20] C. Zuo, Q. Chen, L. Tian, L. Waller, and A. Asundi, "Transport of intensity phase retrieval and computational imaging for partially coherent fields: the phase space perspective," *Optics and Lasers in Engineering*, vol. 71, no. 8, pp. 20–32, 2015.
- [21] C. Zuo, J. Sun, J. Li, J. Zhang, A. Asundi, and Q. Chen, "High-resolution transport-of-intensity quantitative phase microscopy with annular illumination," *Scientific Reports*, vol. 7, no. 1, pp. 7622–7654, 2017.
- [22] B.-H. Li, Y. Liu, A.-M. Zhang, W.-H. Wang, and S. Wan, "A survey on blocking technology of entity resolution," *Journal of Computer Science and Technology*, vol. 35, no. 4, pp. 769–793, 2020.
- [23] J. Zhang, Q. Chen, J. Sun, L. Tian, and C. Zuo, "On a universal solution to the transport-of-intensity equation," *Optics Letters*, vol. 45, no. 13, pp. 3649–3652, 2020.
- [24] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 6062–6071, 2015.
- [25] Z. Niu, B. Zhang, J. Wang et al., "The research on 220GHz multicarrier high-speed communication system," *China Communications*, vol. 17, no. 3, pp. 131–139, 2020.