Feature Enhancement Algorithm for Recognition of Small Targets in Low Resolution Images

Jingjing Wang
Research China, Signify N.V.
Shanghai, China
wjj0219@naver.com

2023 IEEE 6th International Conference on Pattern Recognition and Artificial Intelligence (PRAI) | 979-8-3503-2548-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/PRAI59366.2023.10332092

***Abstract*—Small object detection is a challenging task in computer vision due to limited pixels, low resolution, and sparse feature information. Existing infrared-based methods for small object detection suffer from low accuracy and susceptibility to environmental disturbances. This paper proposes a novel approach using an enhanced lightweight YOLOv5 network for high-precision infrared small object detection. The method involves capturing depth data using a ToF sensor, converting it to a four-channel RGB image through pseudo-color encoding, and applying HSV conversion. By preprocessing the depth values and incorporating them into the lightweight YOLOv5 model, the proposed approach extracts infrared small object detection features and improves the network's efficiency. Experimental results demonstrate that the proposed method achieves a significant improvement in detection performance, with an average precision of 97.4% on an enhanced dataset compared to pseudo-color encoded images.**

***Index Terms*—Object recognition; Image enhancement; YOLOv5; Low resolution; Small object detection**

I. Introduction

Small object detection is one of the hot research topics in the field of computer vision. Current object detection techniques based on deep learning have made rapid progress, but small object detection remains a challenging problem that requires improvement. Compared to large objects, small object detection involves identifying the positions and categories of objects with fewer pixels, lower resolution, and limited feature information in images. Many general object detection algorithms cannot be directly applied to small object detection. Therefore, enhancing the ability of small object detection is a challenging and important research direction in the field of object detection. Currently, there are still several limitations in object detection techniques, such as occlusion, complex backgrounds, object overlap, and low object resolution. Small objects often suffer from low recognition rates due to their small size and limited number of pixels in the original image, resulting in feature maps with single-digit dimensions. This leads to poor detection performance, unclear classification, and even missed or false detections in cases of occlusion or overlap. To address these issues, numerous researchers have conducted extensive studies.

II. Related Work

The field of computer vision encompasses fundamental tasks such as object detection, image segmentation, scene understanding, object tracking, and image captioning. Among these tasks, object detection serves as the cornerstone of image understanding and forms the basis for other advanced computer vision tasks. In recent years, deep learning-based object detection methods have made significant progress, and object detection has been widely applied in various domains, including robot navigation, intelligent video surveillance, aerospace, and more. Deep learning methods have achieved impressive performance in terms of detection speed and accuracy in the field of object detection. Currently, representative object detection networks such as Faster R-CNN [1], You Only Look Once (YOLO) [2], and SSD [3] have demonstrated good performance on various datasets. However, the detection of small and weak infrared targets in complex backgrounds generally yields unsatisfactory results. It is necessary to improve the detection of small targets with limited features and information. In the domain of infrared small target detection, a number of studies have emerged that utilize deep learning methods.

CornerNet [4] treats object detection as a pair of keypoints (top-left and bottom-right corners of bounding boxes) and groups them based on proximity to obtain the final detection results. On the other hand, CenterNet predicts the center points of objects and regresses the size of each object based on the center point of the bounding box. Redmon proposed the YOLOv3 [5] algorithm in 2018, which incorporated deeper networks, resulting in improved detection efficiency and accuracy. He Xiaowei et al. [6] proposed an improved algorithm based on YOLOv3 to address the challenges of low accuracy, slow detection speed, and complex model in detecting small objects in traffic scenarios. They introduced a lightweight yet powerful backbone network named SA-MobileNetXt, which incorporates spatial and channel attention mechanisms to enhance expressive feature extraction while introducing a minimal number of parameters.

The difficulties of small target detection can be summarized in the following five aspects:

- Limited spatial info: Small size hampers precise localiza-

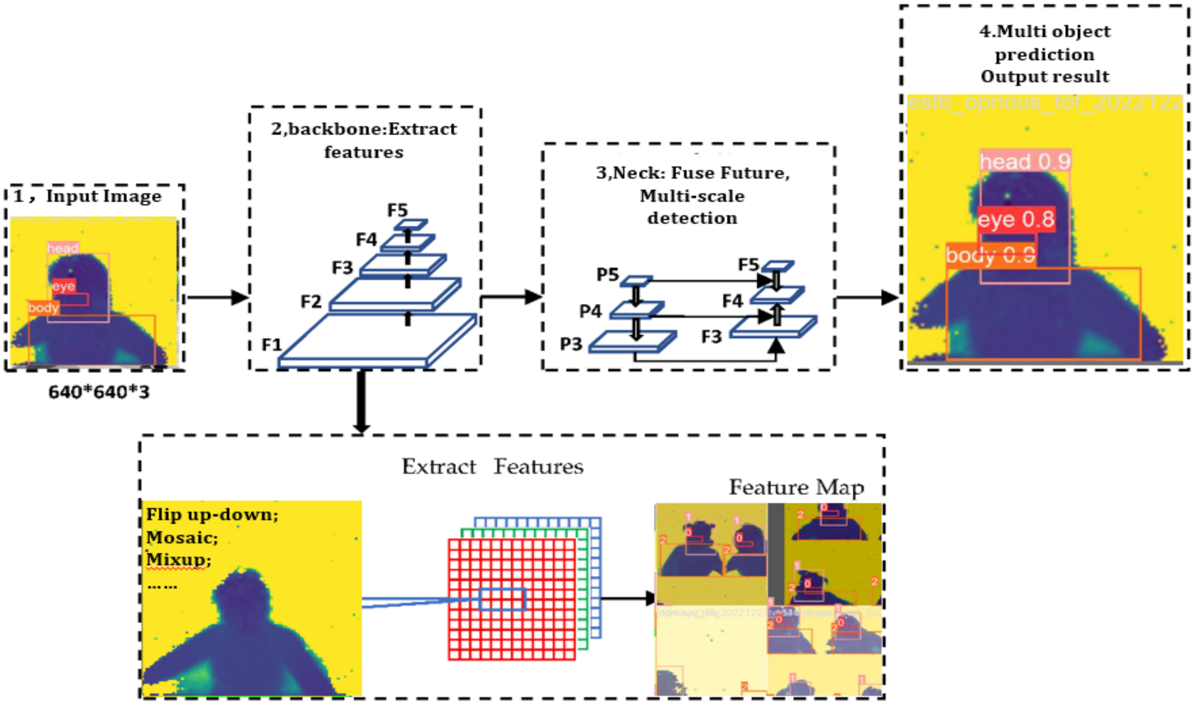


Fig. 1. Algorithm architecture diagram of YOLOv5 model based on pseudo-color coded ToF data.

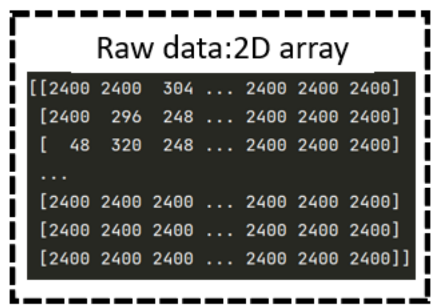
tion due to minimal pixel coverage.
- Low res & sparse features: Sparse traits challenge detection and classification.
- Occlusion & backgrounds: Overlap and complex backgrounds hinder identification.
- Object overlap: Overlapping causes blurred boundaries, leading to misses and false alarms.
- Low object resolution: Limited detail extraction complicates accurate detection and classification.

The existing infrared small target detection methods suffer from low detection accuracy and susceptibility to environmental interference. Improving the capability of infrared small target detection is a challenging problem and an important research direction in the field of object detection. To achieve high-precision infrared small target detection, this paper proposes a novel method based on enhanced data and a lightweight YOLOv5 network.

III. Proposed Algorithm for Feature Enhancement

A. Algorithm architecture diagram of YOLOv5 model

The YOLO algorithm is a fast and efficient deep learning approach for object detection. It operates in a single stage without region proposals, resulting in fast execution speed. Although its accuracy is slightly lower compared to two-stage algorithms, YOLO has been continuously updated and optimized to improve precision while maintaining its speed advantage. YOLOv4, in particular, is a powerful and efficient one-stage detection model suitable for various industrial projects. Its high customization and flexibility make it widely applicable, especially for small object detection. In this paper, the YOLOv5 model is used to detect small targets. as shown in Figure 1.


Raw data:2D array

```
[[2400 2400  304 ... 2400 2400 2400]
 [2400  296  248 ... 2400 2400 2400]
 [  48  320  248 ... 2400 2400 2400]
 ...
 [2400 2400 2400 ... 2400 2400 2400]
 [2400 2400 2400 ... 2400 2400 2400]
 [2400 2400 2400 ... 2400 2400 2400]]
```

Fig. 2. Raw two-dimensional data collected from the ToF sensor

B. RGB to HSV Conversion Principle

Figure 2 shows the original two-dimensional data acquired by the ToF sensor. Figure 3 shows the ToF sensor data processed by pseudo-color coding. To further improve the detection rate of the YOLO model for small targets, we process the pseudo-color encoded data based on the principle of RGB data to HSV conversion.

The RGB color model [7] is an industry-standard color space that is used to create a wide range of colors by varying the intensity of three color channels: red (R), green (G), and blue (B). The colors are achieved by combining and adjusting these channels. RGB represents the three color channels and encompasses almost all colors perceptible to the human eye. It is one of the most widely used color systems today. HSV (Hue, Saturation, Value) [8] is a color space created by A. R. Smith in 1978, based on the intuitive properties of color. It is also known as the hexcone model. The parameters of this model are hue (H), saturation (S), and value (V). The conversion principle is straightforward. For any arbitrary coordinate point in an image, with RGB color space values (R, G, B) and HSV color space values (H, S, V), the first step is to normalize the R, G, and B values to the range of 0 to 1:

$$\begin{aligned} R &= R/255 \\ G &= G/255 \\ B &= B/255 \end{aligned} \tag{1}$$

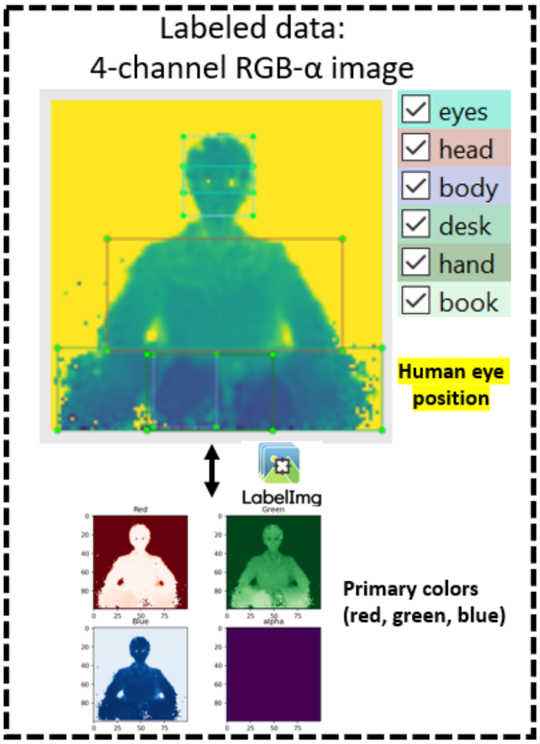


Fig. 3. RGB data processed by pseudo-color encoding

Next, the H, S, and V values are computed:
And, Value (V) is denoted as:

$$V = \max(R, G, B) \tag{2}$$

Saturation (S) is denoted as:

$$S = \begin{cases} \frac{V - \min(R,G,B)}{V}, & \text{if } V \neq 0 \\ 0, & \end{cases} \tag{3}$$

Hue (H) is denoted as:

$$H = \begin{cases} 60^*(G-B)/(V-\min(R,G,B)) & ,\text{if } V = R \\ 120 + 60^*(B-R)/(V-\min(R,G,B)) & ,\text{if } V = G \\ 240 + 60^*(R-G)/(V-\min(R,G,B)) & ,\text{if } V = B \end{cases} \tag{4}$$

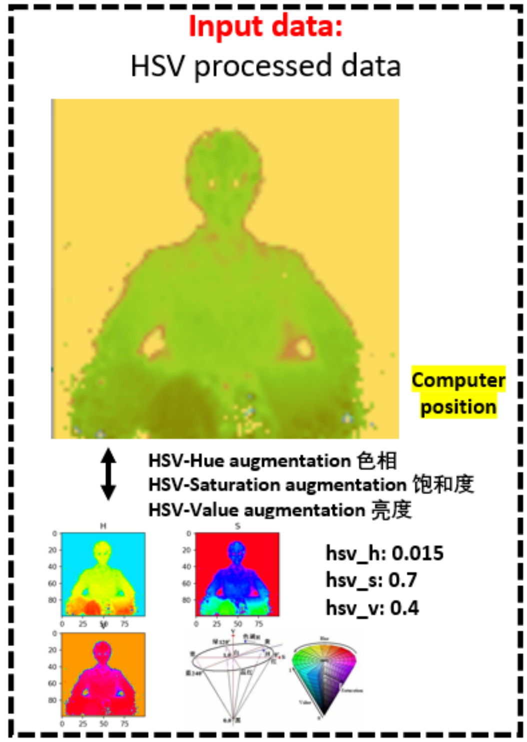


Fig. 4. HSV data after conversion processing.

If the calculated H value is less than 0, it is then incremented by 360 to obtain the final H value.

$$H = H + 360 \tag{5}$$

Finally, since OpenCV requires HSV images for visualization, it is necessary to map the values back to the range of 0 to 255:

$$\begin{aligned} H &= \frac{H}{2} \\ S &= S \times 255 \\ V &= V \times 255 \end{aligned} \tag{6}$$

HSV is a method of representing points in RGB color space in an inverted cone. The pseudo-color encoded image is converted into an HSV image, as shown in Figure 5.

IV. Experimental Environment and Results

A. Experimental Environment and Evaluation Metrics

1) Experimental Platform: All experiments in this paper were conducted using Python 3.7. This paper performed training and testing on an NVIDIA RTX3080 GPU. This paper employed the SGD optimization algorithm with a learning rate of 1e-2, weight decay of 5e-4, momentum set to 0.937, and a batch size of 16. We set the total number of epochs to 1328.

2) Experimental Dataset: This paper extracted a total of 6958 static images from the video frames, which we acquired in various indoor scenes. The dataset was split into 4480 training images, 548 validation images, and 1930 test images. Adjustments in ToF sensor position were made within ±10cm

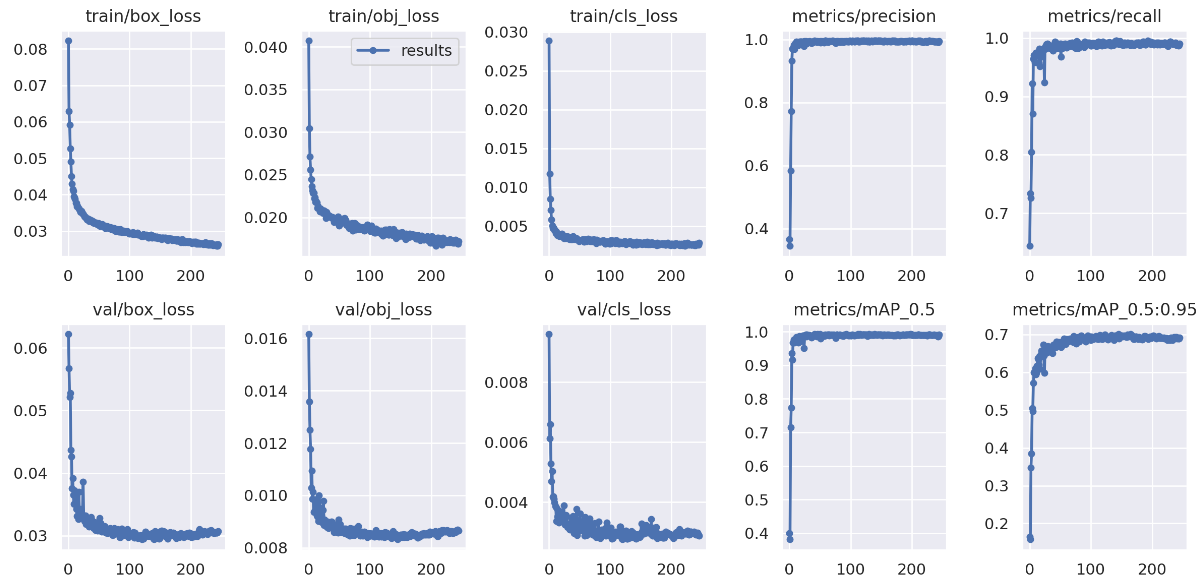


Fig. 5. Visualization training results of object detection model.

of the head during the capture process. The captured images exhibited significant variations in scale and background due to multiple individuals and varying lighting conditions. The dataset comprised 48 classes with an imbalanced distribution. This paper trained and evaluated the proposed model on this dataset.

3) Evaluation Metrics: The performance of the algorithm was assessed using metrics such as mean average precision (mAP), parameter count, and computational complexity. mAP was calculated as the average precision over 10 intersection over union (IoU) thresholds ranging from 0.50 to 0.95 with a uniform step size of 0.05. It served as the primary evaluation metric for algorithm detection performance. mAP at various IoU thresholds provided a comprehensive measure of detection accuracy.

B. Experimental Results

The performance of the model in different aspects is evaluated by mAP, precision, recall, etc. The result is shown in Figure 5. In the object detection domain, precision and recall are used as the axes to construct a P-R (precision-recall) curve. The area under the curve, which is bounded by the curve and the axes, represents the average precision (AP) of the model for a particular class. By averaging the AP values for all classes, the mAP value is obtained, which evaluates the overall detection accuracy of the model on the entire dataset. The comparative results are shown in Table I.

TABLE I
COMPARISON OF THE ACCURACY OF TARGET DETECTION ALGORITHMS ON DIFFERENT TRAINING SETS

Mean Average precision	**mAP (0.5)**	**mAP (0.5:0.95)**
Pseudo-color encoded image	84.8%	57.4%
HSV image (Proposed method)	97.4%	66.0%

V. CONCLUSION

This paper presents a method to enhance the detection capability of YOLO models for small objects by employing data augmentation techniques. The results show an overall improvement in detection accuracy for objects of different sizes, with a significant 12.6 % boost in accuracy specifically for small objects. However, the model's performance could not be further enhanced due to the absence of integration with IR images obtained from the ToF sensor. Future research aims to address this limitation and improve the model accordingly.

REFERENCES

[1] Girshick R. Fast r-cnn. InProceedings of the IEEE international conference on computer vision 2015 (pp. 1440-1448).
[2] Jiang P, Ergu D, Liu F, Cai Y, Ma B. A Review of Yolo algorithm developments. Procedia Computer Science. 2022 Jan 1;199:1066-73.
[3] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC. Ssd: Single shot multibox detector. InComputer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14 2016 (pp. 21-37). Springer International Publishing.
[4] Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q. and Tian, Q., 2019. Centernet: Keypoint triplets for object detection. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 6569-6578).
[5] Redmon J, Farhadi A. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767. 2018 Apr 8.
[6] He, X., Cheng, R., Zheng, Z. and Wang, Z., 2021. Small object detection in traffic scenes based on YOLO-MXANet. Sensors, 21(21), p.7422.
[7] Ibraheem NA, Hasan MM, Khan RZ, Mishra PK. Understanding color models: a review. ARPN Journal of science and technology. 2012 Apr;2(3):265-75.
[8] Smith AR. Color gamut transform pairs. ACM Siggraph Computer Graphics. 1978 Aug 23;12(3):12-9.