

Research on fast detection method of infrared small targets under resource-constrained conditions

Zhang Rui¹, Liu Min¹, Li Zheng²

(1. School of Opto-Electronic and Communication Engineering Xiamen University of Technology Xiamen, China;
2. Shanghai Institute of Technical Physical Chinese Academy of Sciences)

Abstract: Infrared small target detection is a common task in infrared image processing, and computational resources are limited in application scenarios that require the use of the infrared small target detection. Traditional IR small target detection methods face the problem of balancing detection rate and accuracy. This paper presents a swift infrared small target detection method designed for resource-constrained conditions, in the YOLOv5s model that adds a tiny target detection head and replaces the original Intersection over Union (IoU) with a Normalized Gaussian Wasserstein Distance (NWD), considering both the detection accuracy and rate of the infrared targets. The experimental results show that the algorithm in this paper realizes the maximum 95 FPS effective detection speed of the infrared small targets on a 15W TPU, and at the same time achieves the maximum 91.9 AP@0.5 effective detection accuracy, which effectively improves the detection efficiency of the infrared small targets under resource-constrained conditions.

Key words: *infrared UAV image, fast small object detection, low impedance, loss function*

PACS:

资源受限条件下的红外小目标快速检测方法研究

张 瑞¹, 刘 敏¹, 李 争²

(1. 厦门理工学院, 光电学院;
2. 中科院上海技术物理研究所)

摘要: 红外小目标检测是红外图像处理中的一项常见任务。计算资源是有限的条件下, 传统的红外小目标检测方法面临着检测率和检测精度的平衡问题。本文在 YOLOv5s 模型中提出了一种针对资源受限条件设计的快速红外小目标检测方法, 该方法增加了一个小目标检测头, 并用归一化高斯瓦瑟斯坦距离(NWD)取代了原来的交并比(IoU), 同时考虑了红外目标的检测精度和检测速率。实验结果表明, 本文算法在 15W TPU 上实现了最大 95 FPS 的红外小目标有效检测速度, 同时达到了最大 91.9 AP@0.5 的有效检测精度, 有效提高了资源受限条件下的红外小目标检测效率。

关 键 词: 红外无人机; 快速小目标检测; 低功耗; 损失函数

1 Introduction

The application range of unmanned aerial vehicles (UAVs) is constantly expanding, encompassing areas such as military reconnaissance, outdoor photography, and power line inspections, among others. However, at the same time, it has also given rise to a series of social issues. These include concerns about privacy infringement due to UAVs being used for illicit filming and the potential threat to national security posed by the military application of UAVs. Therefore, anti-drone technology research has important practical significance. Infrared UAV target detection technology is a technique that uses infrared imaging to continuously monitor UAVs. It en-

ables UAV target detection based on their temperature characteristics, and also has obvious advantages in low-light conditions [1]. In recent years, this technology has become an important research direction and provides an effective complement to radar target detection technology.

However, infrared UAV target detection faces challenges. The distance between the UAV and the sensor makes small infrared targets lack distinctive texture and shape features, hampering detection. Additionally, background clutter and noise, like clouds and buildings, can cause confusion with obstacles [2]. Infrared images have high noise, poor spatial resolution, and are sensitive to environmental temperature changes. These chal-

lenges add complexity to infrared UAV target detection. Infrared small UAVs are shown in Fig. 1.

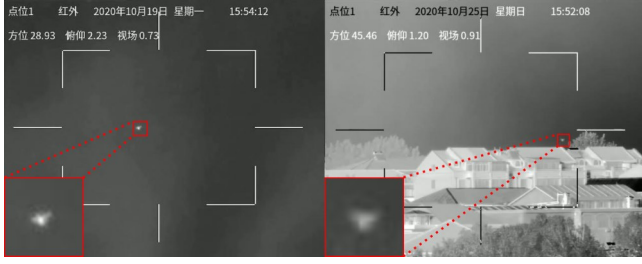


Fig. 1 Infrared small UAVs.
图1 红外无人机目标

In recent years, there has been a continuous emergence of object detection methods based on deep learning, which have achieved impressive detection performance. These methods can be categorized into two types based on how they handle input images. The first type is the two-stage detection methods, such as region-based R-CNN and its variants [3]. The second type consists of one-stage detection methods, including RetinaNet [4], SSD [5], and YOLO series. During the flight of UAVs, real-time transmission of infrared images is required for the infrared cameras. In scenarios that demand high real-time performance, methods from the YOLO series [6], known for their fast speed and high accuracy, have been widely adopted. Among them, YOLOv5 stands out as an advanced detector with strong real-time processing performance and lower hardware computing requirements, allowing for easy deployment on mobile devices. Therefore, using YOLOv5 can significantly enhance the detection accuracy and real-time performance of infrared small UAV [7].

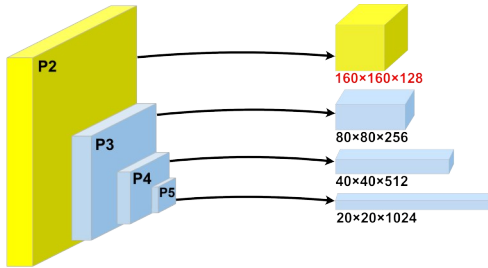


Fig. 3 Small target detection head.
图3 小目标检测头

This paper proposes an UAV detection method based on an improved YOLOv5s model to address the challenges of detecting small targets. The original YOLOv5 structure only includes three feature detection heads, which are not effective in extracting the feature information of small target UAVs captured by infrared cameras at long distances. To address this issue, this paper adds a feature detection head suitable for detecting small targets to YOLOv5. Additionally, the intersection over union (IoU) in the original YOLOv5 model is not a good metric for small target detection tasks. Therefore, this

paper replaces it with a more suitable metric for small targets called the Normalized Gaussian Wasserstein Distance (NWD) [8]. This metric calculates the similarity between bounding boxes using the Gaussian distribution of their corresponding boxes.

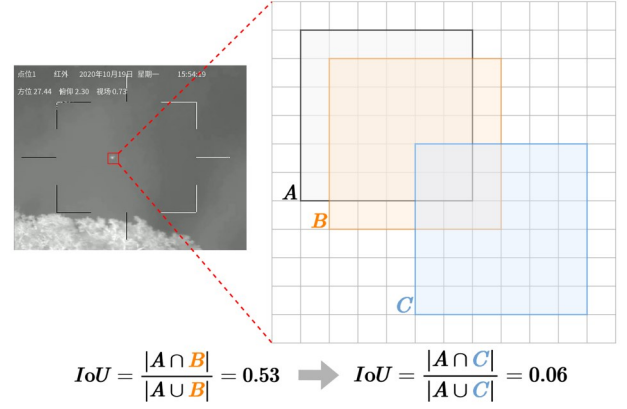


Fig. 4 The sensitivity analysis of IoU on infrared small UAV.
图4 IOU在红外无人机上性能分析

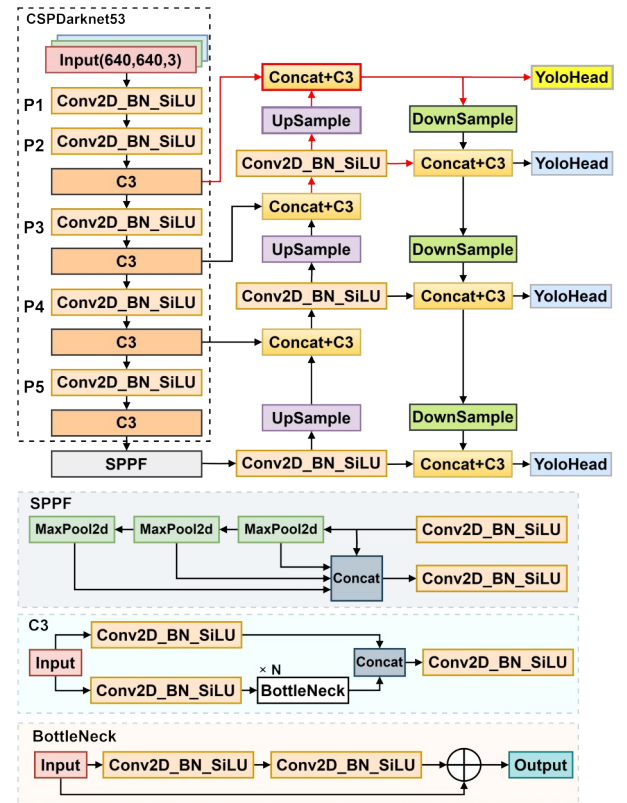


Fig. 2 Improved YOLOv5s network architecture.
图2 改进的yoloV5s 网络架构

II Method to improve YOLOv5

YOLOv5 can be classified into four architectures based on the depth and width of the model: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. Among them, YOLOv5s has the smallest number of parameters and computations, making it more suitable for real-time de-

tection on embedded devices. The network structure of YOLOv5 consists of three parts: CSP-Darknet53 as the feature extraction network to extract features from the input images. CSP-Darknet53 is an improved version of Darknet53 from YOLOv3, incorporating Cross Stage Partial (CSP) network strategy to reduce the number of parameters and computations, thus enhancing the inference speed. In the neck part, YOLOv5 incorporates two modules: SPPF and PANet [9]. backbone, neck, and head. The SPPF module increases the receptive field and diversifies backbone of YOLOv5 utilizes. the feature pyramid, while the PANet module enables bottom-up and top-down feature fusion, thereby improving the detection capability for targets. In the head part, YOLOv5 adopts the head structure from YOLOv4. This structure applies anchor boxes on the feature map and outputs predictions such as class probabilities, confidence scores, and bounding boxes.

A Small Object Detection Head

The YOLOv5 model uses a backbone network that undergoes five downsampling stages, producing five feature maps (P1-P5) with resolutions of 1/2, 1/4, 1/8, 1/16, and 1/32 of the input image size, respectively. The neck network combines multi-scale features in a top-down and bottom-up manner without changing the feature map sizes. The detection head operates on the P3-P5 feature maps for object detection. This design is based on the relationship between the feature layer size and the receptive field size in YOLOv5. The receptive field refers to the size of the input image region corresponding to each output unit in a convolutional neural network. A larger receptive field captures more object features, making it suitable for detecting larger objects. On the contrary, a smaller receptive field can only capture a limited number of object features, making it suitable for detecting smaller objects. A smaller receptive field implies that each pixel in the feature map is influenced by a smaller region in the original image. This enables more precise localization of object positions and boundaries, unaffected by irrelevant regions. Additionally, a smaller receptive field corresponds to a larger feature map size, preserving more spatial information and avoiding the loss of fine details of small objects. Therefore, a smaller receptive field is better suited for detecting smaller objects.

We have added a new detection head for small objects after the P2 feature layer in the YOLOv5 model. The detection head operates at a resolution of 160x160 pixels in the P2 layer, which corresponds to two downsampling operations in the backbone network. Each pixel in the P2 layer has a receptive field of 10x10 pixels, which is the smallest receptive field among the P2-P5 feature extraction layers. Additionally, we have assigned different loss function weights to the P2-P5 feature layers based on the target sizes. Specifically, we have assigned a weight of 4.0 to the feature layers for P2 and P3, a weight of 1.0 to the feature layer for P4, and a weight of 0.4 to the feature layer for P5. The purpose of this weighting scheme is to enhance the focus on small and tiny objects while reducing overfitting to large objects.



Fig. 5 Dataset analysis.

图5 数据集分析

The weighted formula for the object loss function as shown in equation (1). The experimental results for different weighting coefficients are shown in Table I. Although adding the new detection head increases the computational and memory overhead of the model, it significantly improves the detection performance for small objects. The Improved YOLOv5s network architecture is shown in Fig. 2. Small target detection head is shown in Fig. 3.

$$L_{obj} = 4 \cdot L_{P2} + 4 \cdot L_{P3} + 1 \cdot L_{P4} + 0.4 \cdot L_{P5} \quad (1)$$

B Normalized Gaussian Wasserstein Distance

The IoU in YOLOv5 is used as Ap2n indicator to measure the degree of matching between predicted bounding boxes and real bounding boxes. It is obtained by calculating the ratio of the intersection area and union area of the two. However, in the UAV images obtained by infrared devices, the overlapping part of the bounding boxes of small targets is often very small, which will result in lower IoU values. As shown in Fig. 4, IoU is very sensitive to the scale of small targets. For infrared small UAV target with a size of 6 x 6 pixels, only a small position deviation can cause IoU to decrease from 0.53 to 0.06, thereby affecting the accuracy of label allocation and reducing the performance of detection [8]. So, for small target objects, IoU does not measure their matching degree effectively.

We used a metric known as Normalized Gaussian Wasserstein Distance named NWD, which is suitable for small object detection. It is insensitive to the scale of the targets, allowing for better assessment of similarities between small objects. Specifically, this method models the object bounding boxes as two-dimensional Gaussian distributions and calculates the NWD between the predicted and ground truth distributions, as shown in equation (2). Modeling the bounding boxes with Gaussian distributions enables the representation of the object's po-

sition as the mean point and its shape as the standard deviation of the Gaussian distribution. This approach has the advantage of better describing the spatial distribution of the objects, not solely relying on their geometric shapes. NWD takes into account the geometric features and spatial distribution between two bounding boxes, even when they do not overlap, by considering their geometric relationships [10]. This measurement approach avoids the sensitivity issues of IoU with small objects, making it more suitable for quantifying small object matches.

$$\text{NWD}(N_A, N_B) = \exp\left(-\frac{\sqrt{W_2^2(N_A, N_B)}}{C}\right) \quad (2)$$

$$W_2^2(N_A, N_B) = \left\| \left[\begin{matrix} cx_A, cy_A, \frac{w_A}{2}, \frac{h_A}{2} \end{matrix} \right]^T, \left[\begin{matrix} cx_B, cy_B, \frac{w_B}{2}, \frac{h_B}{2} \end{matrix} \right]^T \right\|_2^2 \quad (3)$$

Where N_A and N_B are Gaussian distributions modeled by $A = (cx_A, cy_A, w_A, h_A)$ and $B = (cx_B, cy_B, w_B, h_B)$, $W_2^2(N_A, N_B)$ is a distance measure, C is a constant closely related to the dataset.

III Experiment

A Dataset Introduction

This article is based on experimental analysis of selected subsets from the 3rd Anti-UAV Workshop & Challenge [11]. The dataset consists of a total of 9841 small target drone images captured under thermal infrared, involving complex environmental factors such as dynamic backgrounds, complex motions, scale changes, and small targets. The labels in this dataset only include the

category of drones. It is noteworthy that the average signal-to-noise ratio (SNR) of this dataset is 12.615 dB. The dataset was divided into training and testing sets in a 7:3 ratio. The distribution of the dataset is shown in Fig. 5.

B Evaluating Evaluation Index

TABLE I. The Different Weighting Coefficients Results

To verify the performance of the model, this article selects Average Precision (AP) to evaluate the model performance, where the AP@0.5 and AP@0.5:0.95 represent the detection accuracy of two models under different IoU thresholds. AP@0.5 is the average accuracy of a certain category when IoU is 0.5. The average accuracy is based on different confidence levels. The curve area of Precision (P) and Recall (R). And AP@0.5:0.95 is the average accuracy of a certain category when IoU is taken every 0.05 from 0.5 to 0.95. This indicator requires a higher degree of overlap in the target box. The False Alarm Rate (FAR), as depicted in equation (6), the Miss Rate (MR), represented by equation (7), with the Average Precision (AP) illustrated in equation (8).

$$P = \left(\frac{TP}{TP + FP} \right) \quad (4)$$

$$R = \left(\frac{TP}{TP + FN} \right) \quad (5)$$

$$\text{FAR} = \left(\frac{FP}{TN + FP} \right) \quad (6)$$

$$\text{MR} = 1 - R \quad (7)$$

$$\text{AP} = \int_0^1 P(R) dR \quad (8)$$

表1 不同权重系数的结果对比

Weighting Coefficients	AP@0.5 (%)	AP@0.5:0.95(%)	FAR (%)	MR (%)
$1 \cdot L_{p2} + 1 \cdot L_{p3} + 1 \cdot L_{p4} + 0.4 \cdot L_p$	81.2	44.9	6.4	28.4
$1 \cdot L_{p2} + 4 \cdot L_{p3} + 1 \cdot L_{p4} + 0.4 \cdot L_p$	86.8	47.6	4.7	19.4
$4 \cdot L_{p2} + 1 \cdot L_{p3} + 1 \cdot L_{p4} + 0.4 \cdot L_p$	84.2	46.0	7.8	22.8
$4 \cdot L_{p2} + 4 \cdot L_{p3} + 1 \cdot L_{p4} + 0.4 \cdot L_p$	88.4	48.6	4.0	17.4

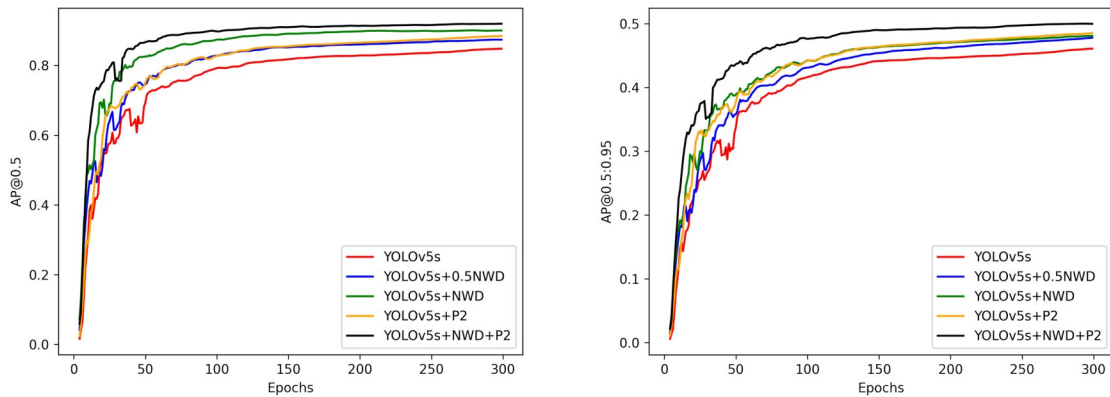


Fig. 6 Performance comparison of the AP. The left curves are AP@0.5. The right curves are AP@0.5:0.95.

图6 模型在AP值上的性能表现,左边的是AP@0.5,右边的曲线是AP@0.5:0.95

Where TP means true positive, TN means true negative, FP means false positive, and FN means false negative.

C Experimental Environment and Parameter Settings

In this work, all experiments were conducted on the Ubuntu 22.04 operating system, with 128GB of RAM and an Intel i9-13900K processor. The system was equipped with an NVIDIA RTX 3090 Ti graphics card with 24GB of video memory; the deep learning framework used was Pytorch 1.12.1; and the programming language was Python 3.10.

The optimization algorithm used for model training was stochastic gradient descent (SGD). The initial learning rate was 0.01, the moment was 0.937, and the weight decay coefficient was 0.0005. In addition, the model was trained for 300 epochs, the batch size of the dataset was set to 64.

D Experimental Results

From the ablation experiment results in Table II, it can be observed that adding the NWD metric with a coefficient of 0.5 resulted in a 3.3% improvement in AP@0.5 compared to the original model. Furthermore, when using the complete NWD metric, AP@0.5 improved by 5.2%. Moreover, introducing a small object detection head with a resolution of 160x160 pixels from the P2 layer also led to a 3.7% increase in AP@0.5 compared to the baseline results. Finally, by combining the complete NWD loss function with the small object detection head, the overall model achieved a 7.2% improvement. The improved model also showed a 3.9% increase in AP@

0.5:0.95 compared to the original model.

TABLE II. The Comparison of YOLOv5s Improvements

表2 模型的消融实验对比

Models	AP@	AP@	FAR (%)	MR (%)
	0.5 (%)	0.5: 0.95 (%)		
YOLOv5s	84.7	46.1	3.1	23.6
YOLOv5s+0.5NWD	87.4	47.9	4.0	17.4
YOLOv5s+NWD	89.9	48.1	4.4	14.6
YOLOv5s+P2	88.4	48.6	4.0	17.4
YOLOv5s+NWD+P2	91.9	50.0	4.2	12.9

The performance comparison chart of AP is shown in Fig. 6. The partial experimental result is shown in Fig. 7. To

validate the effectiveness of the improved YOLOv5s in this paper, a comparison was made with various mainstream detection networks using the official sub-dataset from the 3rd Anti-UAV Workshop & Challenge. The comparison results are shown in Table III.

E Deployment

Our algorithm is deployed on the BM1684X TPU, and the specific flow is shown in the Fig. 8. Generally speaking, deep learning-based algorithms have two main

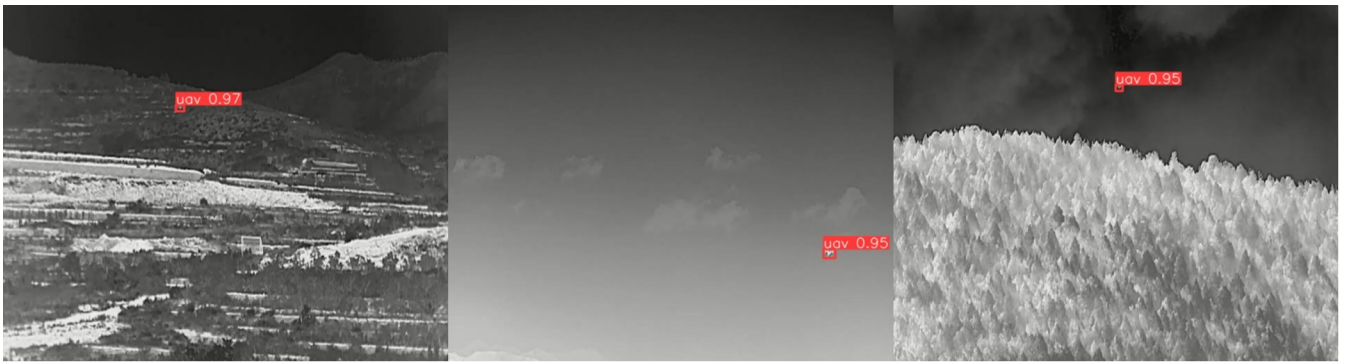


Fig. 7 Some examples of the detection result on the improved YOLOv5s.

图7 改进的模型的案例表现。

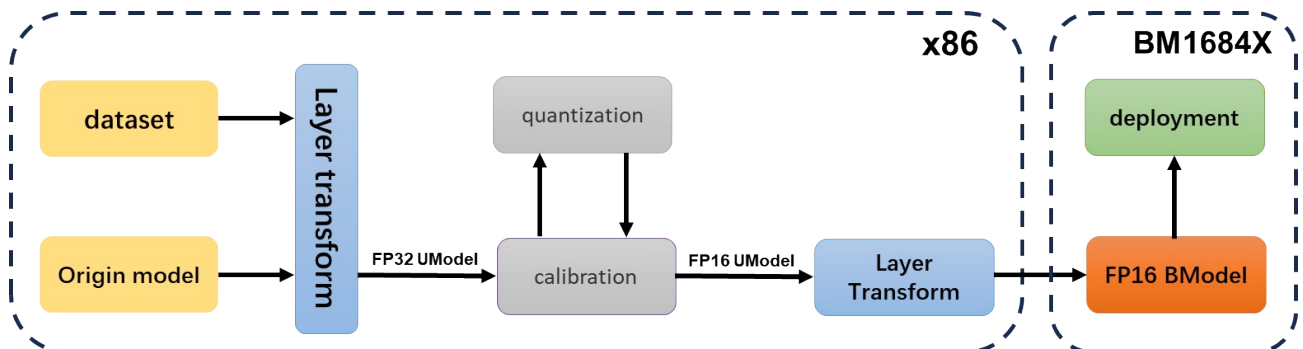


Fig. 8 Framework of deployment.

图8 模型部署的架构流程

TABLE III. Comparison of Improved YOLOv5s with Other Methods
表3 改进的模型与其他同类算法的比较

Models	AP@0.5 (%)	AP@0.5; 0.95 (%)	FAR (%)	MR (%)	Parameter (M)	GFLOPs	Speed (FPS)	Weights (MB)
SSD-ResNet50	60.4	22.8	29.7	46.2	13.1	15.0	200	105.1
Faster-RCNN-ResNet50	78.3	30.9	21.2	40.0	41.1	134.5	50	330.3
RetinaNet-ResNet50	82.1	33.8	18.5	33.2	32.0	127.5	43	257.3
YOLOv3	83.3	45.2	8.7	22.4	9.3	23.1	526	18.9
YOLOv5s	84.7	46.1	3.1	23.6	7.0	15.8	625	14.4
YOLOv5m	86.6	48.8	3.1	20.4	20.8	47.9	303	42.2
YOLOv5l	87.7	49.1	3.8	17.6	46.1	107.6	196	92.8
YOLOv5x	89.5	48.9	6.3	17.3	11.1	28.4	435	22.5
YOLOv5s+NWD+P2	91.9	50.0	4.2	12.9	7.7	26.8	400	16.3

results in the quantization process: int8 and FP16. According to the characteristics of infrared small targets, learning-based target detection algorithms will have a significant accuracy decline in the quantization process. The experimental results show that when the algorithm accuracy is quantized to FP16, the target detection efficiency is the highest, while the quantization to int8 has the weakest effect to the extent that no valid data can be counted.

TABLE IV. Result of Deployment

表4 嵌入式部署的实验结果

BM1684X	FPS	AP50	AP95
FP32	12	91.9	50.0
FP16	95	87.8	49.8
INT8	163	-	-

IV Conclusion

To address the challenge of small drone detection in infrared devices, this paper proposes an light weight detection model. The model introduces a small object feature extraction layer at the P2 level of the backbone network, connecting it to a high-resolution detection head, thereby enhancing the network's capability to perceive small objects. Additionally, the paper adopts the NWD metric to replace the original IoU-based metric, as the NWD metric provides better measurements for small object instances and improves the model's detection accuracy. The paper conducts several comparative experiments on a partial sub-dataset provided by the 3rd Anti-UAV Workshop & Challenge. The results demonstrate that the proposed model outperforms other mainstream detection models in terms of both AP@0.5 and AP@0.5:0.95 evaluation metrics, validating the effectiveness of the

proposed approach. Furthermore, the proposed method maintains high performance levels concerning parameter count, computational complexity, and model weight file size.

References

- [1] S. S. Rawat, S. Alghamdi, G. Kumar, Y. Alotaibi, O. I. Khalaf, and L. P. Verma, "Infrared small target detection based on partial sum minimization and total variation," *Mathematics*, 2022, vol. **10**, pp. 671.
- [2] H. Fang, M. J. Xia, G. Zhou and Y. Chang, "Infrared small UAV target detection based on residual image prediction via global and local dilated residual networks," *IEEE Geoscience and Remote Sensing Letters*, 2021, vol. **19**, pp. 1 – 5.
- [3] S. Ren, K. He, R. Girshick and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, 2015, 28.
- [4] T. Y. Lin, P. Goyal, R. Girshick, G. Zhou and Y. Chang, "Infrared small UAV target detection based on residual image prediction via global and local dilated residual networks," *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980 – 2988.
- [5] W. Lin, D. Anguelov, D. Erhan, C. Szegedy and S. Reed, "SSD: Single shot multibox detector," *Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11 – 14, 2016, Proceedings, Part I 14*. Springer International Publishing, 2016, pp. 21 – 37.
- [6] J. Redmon, S. Divvala, R. Girshick, G. Zhou and Y. Chang, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779 – 788.
- [7] K. Guo, C. He, M. Yang, G. Zhou and Y. Chang, "A pavement distresses identification method optimized for YOLOv5s," *Scientific Reports*, 2022, vol. **12**, pp. 3542.
- [8] J. Wang, C. Xu, W. Yang and L. Xu, "A normalized Gaussian Wasserstein distance for tiny object detection," *arXiv preprint arXiv: 2110.13389*, 2021.
- [9] S. Liu, L. Qi, H. Qin, J. Shi and J. Jia, "Path aggregation network for instance segmentation," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8759–8768.
- [10] J. Wang, W. Yang, H. C. Li, H. Zhang and G. S. Xia, "Learning center probability map for detecting objects in aerial images," *IEEE Transactions on Geoscience and Remote Sensing*, 2020, vol. **59**(5), pp. 4307–4323.
- [11] J. Zhao, J. Li, L. Jin, J. Chu, Z. Zhang and J. Wang, "The 3rd Anti-UAV Workshop & Challenge: Methods and Results," *arXiv preprint arXiv:2305.07290*, 2023.