High-Resolution Remote Sensing Image Object Detection System for Small Unmanned Aerial Vehicles Based on MPSOC

Hui Xia
College of Marine Engineering, Electrization and Intelligence, Jiangsu Maritime Institute Nanjing, 211199, China

***Abstract*—With the maturation of remote sensing, the applications of small unmanned aerial vehicles are rapidly expanding. Efficient image object detection algorithms have become crucial for information extraction in unmanned aerial vehicles. To meet this demand, an improved YOLOv5s algorithm was developed and deployed within a multi-processor system to optimize the performance of object detection in high-resolution remote sensing images captured by small unmanned aerial vehicles. Through adjustments to the structure and parameters of YOLOv5s, the algorithm was enhanced to improve object recognition capabilities in high-resolution remote sensing imagery. Experimental results demonstrated that the improved YOLOv5s (I-YOLOv5s) algorithm effectively mitigates interference from shadows and other external factors, enabling precise identification of objects. During training, I-YOLOv5s exhibited faster convergence, reaching optimal status after approximately 176 iterations. In performance evaluation, the algorithm achieved F1 and Recall values of 0.92 and 0.94, respectively, significantly outperforming single-shot multibox detectors. I-YOLOv5s attained a maximum average precision of 0.96, markedly higher than comparative algorithms, with its Loss value reduced to a mere 0.06. The introduction of this enhanced algorithm not only enhances the accuracy and efficiency of object detection but also profoundly advances the further application of unmanned aerial vehicles in fields such as environmental monitoring, traffic management, and disaster assessment.**

Keywords*—UAVs; remote sensing images; object recognition; deep learning

I. INTRODUCTION

Small Unmanned Aerial Vehicles (UAVs) play pivotal roles in various domains, such as agriculture, environmental monitoring, and military reconnaissance [1]. Particularly in high-resolution remote sensing image capture and object detection, small UAVs are indispensable. However, achieving high-resolution remote sensing image object detection on small UAVs remains a challenging task due to limitations in size and payload capacity [2]. In high-resolution remote sensing image detection tasks, targets often exhibit characteristics such as small, complex, and similar, posing great challenges to detection algorithms. Firstly, small targets are easily affected by image noise and scale changes, leading to a decrease in detection accuracy. Secondly, the diversity of complex backgrounds and target shapes requires detection algorithms to have strong adaptability. In addition, high-resolution remote sensing images have high similarity between targets, making it difficult to achieve accurate differentiation solely based on traditional feature extraction and classification methods. These issues all pose higher requirements for high-resolution remote sensing image detection algorithms. The existing high-resolution remote sensing image detection algorithms are mainly divided into two categories: based on traditional computer vision methods and based on deep learning methods. Traditional computer vision methods, such as edge detection and region growing, have achieved certain results in object detection, but they have limitations such as high computational complexity, poor robustness, and low detection accuracy. With the rapid development of deep learning technology, especially the application of Convolutional Neural Networks (CNN), high-resolution remote sensing image detection has achieved significant improvement. However, the detection results of existing deep learning methods may be affected by noise and interference, leading to a decrease in accuracy. The Multi-Processor System-on-Chip (MPSOC) offers a solution to this problem, possessing exceptional performance and a highly integrated design to meet the demands of high-precision object detection in complex environments with real-time requirements [3]. The proposed method in this study is based on MPSOC technology, which integrates processors, memory, interfaces, and other components into a system-level chip, offering greater computational power and higher energy efficiency [4]. By embedding the remote sensing image object detection algorithm into MPSOC, efficient and precise object detection can be achieved while meeting the payload and energy consumption constraints of UAVs [5]. This design exhibits innovation in several aspects: firstly, adopting MPSOC technology overcomes the speed and energy efficiency issues of traditional single-processor systems when processing high-resolution remote sensing images. Secondly, employing the YOLOv5s-based object detection algorithm not only ensures efficient object detection but also meets the payload and energy consumption limitations of UAVs. Finally, through optimizing the YOLOv5s algorithm, high-precision object detection in complex environments with real-time requirements is achieved. This research can drive the development of UAVS remote sensing image object detection technology and elevate the application levels of UAVs in agriculture, environmental monitoring, military reconnaissance, and other fields. Moreover, it holds significant theoretical and practical significance for understanding and optimizing the application of MPSOC. The study is divided into five sections. Section II provides a summary of MPSOC and object detection domains. Section III is the implementation of the method proposed by the research.

Section IV is the verification of the UAV target detection system and algorithm proposed by the research. Section V is a summary and outlook of the research content.

II. Related Work

MPSOC technology is a system-level integrated circuit technology that integrates multiple processor cores and other peripherals. Its design goal is to integrate multiple processor cores on a single chip to provide higher computational capabilities and enhanced system performance. The core idea of MPSOC technology is to achieve parallel computation and distributed processing by integrating multiple processor cores on the same chip. Each processor core can independently execute different tasks and communicate and share resources through internal communication channels. This parallel computing architecture enables MPSOC systems to simultaneously handle multiple tasks, thereby improving overall system performance and efficiency. Gomez F et al. proposed a novel platform supported by the MPSOC platform. The research results indicated that it meets the security and criticality requirements for space missions, supports performance verification and diagnostics, and is expected to reach commercial maturity by 2022. The platform will be evaluated for space use cases [6]. Gkeka M R and colleagues, leveraging the efficient performance of MPSOC, introduced a posture optimization module based on RGB features. The research results demonstrated that this module can recover the posture of a robot with an unstable gait when tracking fails, achieving real-time tracking exceeding 30 fps without sacrificing the accuracy and efficiency of tracking and map building [7]. Bruno Sá et al. presented the first public implementation and evaluation of the RISC-V super extension (H-extension v0.6.1) on the Rocket chip core in the MPSOC platform. The results showed that, by enhancing the timer infrastructure, direct interrupt injection and low latency are achieved, supporting the systematic requirements [8]. Nehnouh C proposed an online fault detection and isolation mechanism to enhance the reliability of MPSOC. The research results indicated that this mechanism improves protective performance by 22 times with a 27% area overhead, ensuring high reliability of the network chip. The throughput was only reduced by 5.19%, and the average latency slightly increased by 2.40% [9]. Spieck J et al. introduced a hybrid application mapping method based on MPSOC for data-aware scenarios. The research results showed that machine learning-optimized mapping, significantly reduced the miss rate and energy consumption of soft real-time streaming applications, outperforming existing technologies in a multi-application environment [10].

Object detection technology is a crucial technology in the field of computer vision, used to accurately locate and identify objects of interest in images or videos. This technology finds wide applications in various fields such as intelligent surveillance, autonomous driving, drones, and facial recognition. Zhu Y et al. proposed a weighted truncated Schatten-p norm minimization model to enhance denoising effects in object detection. The research results demonstrated that the model, optimized through adaptive thresholds and alternating direction multiplier methods, effectively improves the accuracy of infrared imaging object detection [11]. Ji Y et al. introduced a local-to-global context-aware feature enhancement network. The research results showed that, through a dual-branch attention mechanism combined with pixel-level self-attention, the method outperforms 18 advanced methods on six benchmark datasets, demonstrating superior object detection performance [12]. Wan Y and colleagues proposed a fine-grained small target detection method with density-aware scale adaptation to overcome occlusion and scale issues in weak small target detection. Research results indicated that this method outperforms existing technologies with high precision on AI-TOD, VisDrone, and UAVDT datasets [13]. Zheng Q and others introduced a cascaded fully convolutional network combined with motion attention to enhance the accuracy of video target detection. The results showed that this method achieves higher accuracy on DAVIS, ViSal, and FBMS datasets compared to existing technologies and simultaneously achieves real-time performance at 27 frames per second [14]. Liang Y and the team addressed the challenge of handling fuzzy contours in RGB-based object detection algorithms by proposing a unified framework applicable to RGB-D and RGB-T saliency detection tasks. Results demonstrated that this framework performs exceptionally well in handling fuzzy contours and low-contrast scenes, exhibiting good generalization and surpassing existing advanced methods across multiple datasets [15].

Although MPSOC technology has significant advantages in improving system performance and efficiency, existing research results still have some shortcomings. Firstly, most of the research focuses on theoretical analysis and simulation verification, and the practical application and commercialization level still need to be improved. Secondly, although MPSOC technology continues to make breakthroughs in the number and performance of processor cores, the related system design and optimization techniques still need to be further improved. In addition, there are few customized MPSOC systems for specific application scenarios, which limits their widespread promotion in practical applications. In response to these shortcomings, a small UAV high-resolution remote sensing image target detection system design based on MPSOC has been proposed. This study aims to address the following issues: 1) How to use MPSOC technology to achieve efficient object detection algorithms to improve system performance and real-time performance; 2) How to design a hardware platform with high integration and low power consumption based on the characteristics and requirements of small UAVs; 3) How to optimize algorithms and system design to adapt to complex scenes in high-resolution remote sensing images.

III. Design of High-Resolution Remote Sensing Image Target Detection System for Small UAVs

The research focuses on constructing and designing a target detection system for high-resolution remote sensing images using small UAVs. Initially, the construction of the small UAVS target detection system is elucidated, emphasizing the design and construction process. Subsequently, the study delves into the construction and optimization of UAVS target detection algorithms, encompassing the selection of initial algorithms, optimization strategies and methods, and their application within the system. The overarching goal of the entire process is to achieve efficient and accurate target detection, aiming to open new research avenues in the field of UAVS remote sensing.

A. Construction of Small UAVs Target Detection System

With the rapid advancement of technology, especially in UAVS and remote sensing technologies, the construction of small UAVS high-resolution remote sensing image target detection systems have become increasingly crucial. The application of this system spans various fields, including environmental protection, disaster management, urban planning, and agricultural monitoring [16]. The small UAVS target detection system enables rapid, efficient, and accurate detection of ground targets, significantly enhancing the efficiency of information acquisition and processing. Moreover, it can automate monitoring in environments where there are high safety requirements or are difficult for humans to access, providing real-time and accurate data support for decision-making [17]. The system must possess the capability to rapidly process and analyze images and data in complex environments while incorporating image target detection algorithms for target detection. MPSOC is a hardware platform that integrates multiple processor cores, enabling efficient and flexible system level performance optimization. The study applies MPSOC to the processing and analysis of UAV remote sensing images, and achieves fast and accurate target detection of high-resolution remote sensing images by assigning different tasks to each processor core. As depicted in Fig. 1, the flow chart illustrates the construction of a small UAVS high-resolution remote sensing image target detection system based on MPSOC.

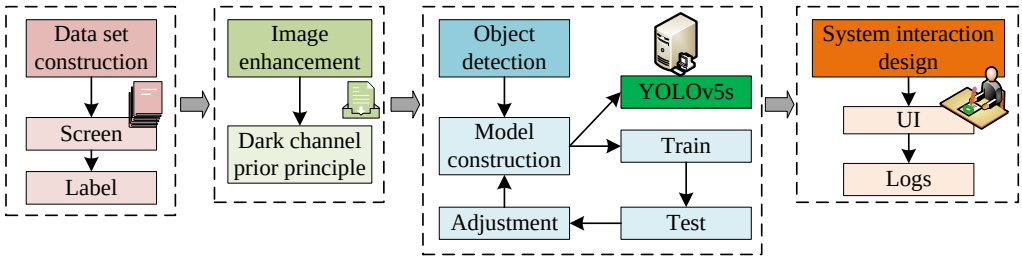


Fig. 1. Flow chart of the construction of a small UAVs high-resolution remote sensing image target detection system.

Considering practical considerations, the process comprises four main parts. Firstly, it involves selecting a dataset suitable for the research content based on actual application scenarios, thereby reducing the training cost of subsequent models and improving training efficiency. The next step involves image enhancement to minimize the impact of complex weather conditions on images, thereby improving image recognition efficiency. Subsequently, the construction of target detection algorithms is addressed. The study primarily utilizes the YOLOv5 algorithm as the foundational target detection algorithm, with improvements tailored to practical situations to enhance image detection efficiency. The final part involves the interactive design of the small UAVS high-resolution remote sensing image target detection system. The initial step in this part is dataset construction. Table I illustrates the integrated dataset constructed for the study. Due to the often limited and singular nature of existing datasets, which may not fulfil new research and application requirements regarding the identification of object categories, quantities, distributions, etc., the research opts to rebuild an integrated dataset. As UAVS camera resolutions improve, existing dataset resolutions may prove insufficient to leverage these high-resolution images. Advances in computational capabilities make it feasible to process and analyze high-resolution images, necessitating the reconstruction of datasets to capitalize on this advantage. Therefore, the study chooses to reconstruct an integrated dataset.

TABLE I. UAVS HIGH RESOLUTION REMOTE SENSING TARGET DETECTION INTEGRATED DATA SET

Data set name	Year of presentation	Target	Number of videos	Total frame count	Detail
UAVDT	2019	Pedestrians, vehicles	100	80K	Covers all weather and light conditions
VisDrone-DET	2019	Pedestrians, vehicles	263	179K	With a variety of scenarios and weather conditions
DTB70	2017	Pedestrians	70	36.5K	A variety of weather, lighting and scenes are included
UCF-ARG	2012	Pedestrians	50	10K	Provides a variety of activity scenarios in the real world
UCSD Birds 200	2010	Birds	-	12K	Images of 200 different bird species are available
Stanford Drone Dataset	2016	Pedestrians, vehicles	60	70K	Suitable for UAVS target detection and tracking
UAV123	2016	Pedestrians, vehicles	123	110K	Covers a variety of weather, lighting, target sizes and speeds
DOTA	2018	Planes, boats, vehicles	-	280K	Designed for detection of large scale ground targets
Okutama-Action	2017	Pedestrians	43	66K	Offers a variety of complex outdoor environments and weather conditions
Aerial Maritime Drone Dataset	2020	Boats	7	25K	Various boat types and weather conditions are included

As shown in Table I, a high-resolution UAV remote sensing image target detection dataset was reconstructed to meet practical requirements. The reconstructed dataset allows for the expansion of detected target types, enabling more detailed analysis. The optimized integrated dataset also promotes research in the remote sensing field, fostering algorithm innovation and optimization. By constructing datasets that include different geographical, climatic, and environmental conditions, the model's adaptability and robustness to different scenarios can be enhanced.

B. Construction of UAVs Target Detection Image Enhancement Methods

For high-resolution images acquired by UAVs, it is crucial to develop adaptive detection algorithms to guide the refinement and expansion of datasets. This process involves not only improving image quality but also considering complex environmental factors to ensure the effectiveness and reliability of the system in practical applications. The first step involves image enhancement processing. An image enhancement method was proposed under the dark channel prior principle [18]. The principle is a dehazing algorithm in computer vision. This principle suggests that in natural scenes, distant objects in an image appear blurry and color-distorted due to factors like light scattering and occlusion. The dark channel prior principle estimates atmospheric light and transmittance by analyzing the dark channel in the image, achieving image dehazing. Fig. 2 illustrates the schematic diagram of atmospheric light scattering.

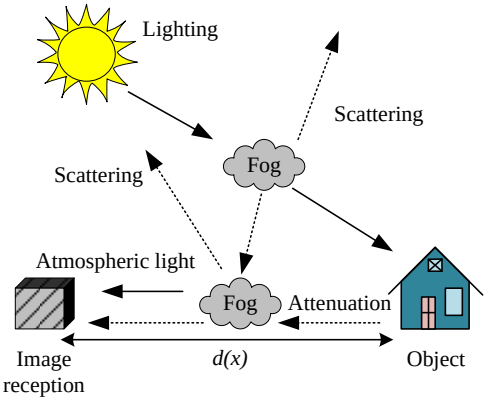


Fig. 2. Schematic diagram of atmospheric light scattering model.

As shown in Fig. 2, the atmospheric light scattering model is depicted. Based on this model and computer vision techniques, the atmospheric scattering model for describing haze can be represented as Formula (1).

$$I(x) = J(x)t(x) + A\left(1 - t(x)\right) \tag{1}$$

In Formula (1), I represents the haze image in the picture, J represents the reflected light of the scene in the image, t represents the transmittance of light in the air, A represents the global atmospheric light intensity, and $t(x)$ represents the transmittance. The representation of transmittance is given by Formula (2).

$$t(x) = e^{-\beta d(x)} \tag{2}$$

In the study of RGB channels for outdoor images, it is found that at least one channel among the three color channels has a very low brightness value close to zero, appearing as dark pixels. This is specifically represented as Formula (3).

$$J_{dark}(x) = \min_{y \in \Omega(x)} \left(\min_{C \in \{r,g,b\}} \left(J^C(Y) \right) \right) \tag{3}$$

In Formula (3), J_{dark} represents the dark original color of the image, J^C represents the color channel of the image, and $\Omega(x)$ represents a square region calculated with x as the center point. In image enhancement calculation, the maximum transmittance value in the image is selected as the initial transmittance to achieve image enhancement. At this point, J_{dark} can be represented as Formula (4).

$$J_{dark} \to 0 \tag{4}$$

Assuming A is a fixed value and constant, taking a local image and dividing both sides of Formula (1) by A yields Formula (5).

$$\min_{y \in \Omega(x)} \left(\min_{c} \frac{I^c(x)}{A^c} \right) = \tilde{t}(x) \min_{y \in \Omega(x)} \left(\min_{c} \frac{J^c(x)}{A^c} \right) + 1 - \tilde{t}(x) \tag{5}$$

Substituting the minimum grayscale value into Formula (5), Formula (6) is obtained.

$$\min_{y \in \Omega(x)} \left(\min_{c} \frac{J^c(x)}{A^c} \right) \to 0 \tag{6}$$

Combining Formula (5) and (6), the real scene transmittance can be calculated, as shown in Formula (7).

$$\tilde{t}(x) = 1 - \min_{y \in \Omega(x)} \left(\min_{c} \frac{I^c(x)}{A^c} \right) \tag{7}$$

To avoid image distortion, it is necessary to maintain the depth of field in the image. Assuming the depth of field adjustment factor is denoted by $\omega \in (0,1)$, introducing it into Formula (7) yields Formula (8).

$$\tilde{t}(x) = 1 - \omega \min_{y \in \Omega(x)} \left(\min_{c} \frac{I^c(x)}{A^c} \right) \tag{8}$$

To further improve the accuracy of image enhancement and dehazing, research explores the use of a soft matting algorithm for optimization [19]. Assuming the optimized transmission image is denoted by t, the optimal value based on the soft matting algorithm's principles can be calculated using Formula (9).

$$(L + \lambda U)\tilde{t} = \lambda t \tag{9}$$

In Formula (9), L represents the Laplacian matrix, λ is a regularization parameter, and U represents a unit matrix of the same size as L. To ensure no distortion in the image, the values

of $J(x)$ are restricted, and the final value can be represented as Formula (10).

$$J\left(X=\frac{I(X)-A}{\max\left(t(x),t_0\right)}\right)+A \tag{10}$$

In practical situations, t_0 is typically set to 0.1. Combining the above calculations, the final process of image enhancement and dehazing can be represented as shown in Fig. 3.

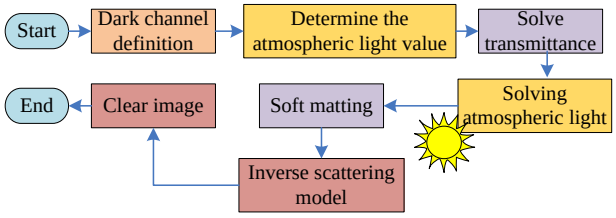


Fig. 3. Schematic image enhancement of the dark channel prior principle.

C. Construction and Optimization of UAVs Target Detection Algorithm

Image enhancement provides a clearer and higher-contrast image input for the algorithm. Next, through algorithm design, the effective identification and localization of objects in the image are achieved. Considering practical needs, YOLOv5s algorithm is chosen as the target detection algorithm. YOLOv5s, as a lightweight version in the YOLO algorithm family, effectively reduces computational burden through network structure and parameter simplification, making it an ideal choice for high-speed detection under limited resources. The end-to-end design of this algorithm achieves one-time detection, avoiding time delays in traditional multi-stage detection methods. The study optimizes the YOLOv5s algorithm by introducing a compression excitation module and a conical feature fusion structure to improve the algorithm's functional utilization and detection accuracy. At the same time, selecting CIOU-Loss as the loss function accelerates the convergence speed of the model and improves the accuracy of regression localization. These improvements enable YOLOv5s to achieve high-speed and accurate object recognition and localization within limited resources. To meet specific requirements, the study optimizes and adjusts YOLOv5s by introducing a compression excitation module, enhancing its feature utilization, as shown in Fig. 4.

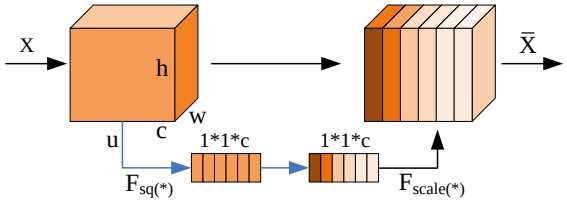


Fig. 4. Structure diagram of compression excitation module.

As shown in Fig. 4, it first compresses features, then maps and weights them. This is mainly achieved through the use of global average pooling [20]. The result can be expressed as Formula (11).

$$z_c=F_{sq(u_c)}=\frac{1}{h\times w}\sum_{i=1}^{h}\sum_{j=1}^{w}u_c(i,j) \tag{11}$$

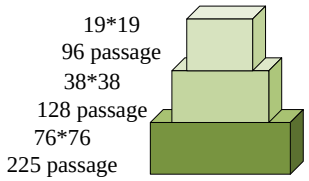
In Formula (11), u_c represents the output of the c-th feature, z_c represents the one-dimensional vector value of the c-th feature, and h and w represent the two dimensions of the feature map. To meet practical requirements, the sigmoid function is chosen as the activation function, obtaining normalized weights, as specified in Formula (12).

$$s=F_{ex}(z,w)=\sigma\left(g(z,w)\right)=\sigma\left(w_2\delta(w_1z)\right) \tag{12}$$

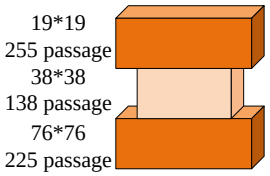
In Formula (12), σ represents the Sigmoid activation function, δ represents the ReLU function, $g(z,w)$ represents a structure composed of two fully connected layers, where the dimension of w_1 is $\frac{c}{r}\times c$, the dimension of w_2 is $c\times\frac{c}{r}$, and r represents a parameter whose main function is scaling. The final output of this module is obtained through rescaling the output, as specified in Formula (13).

$$\bar{X}_c=F_{scale}(u_c,s_c)=u_c\cdot s_c \tag{13}$$

In Formula (13), $\bar{X}_c$ represents the final output, and s_c represents the normalized weight processing result for the c-th feature. Through these steps, the compressed excitation module can be added to the YOLOv5s object detection algorithm, thereby improving its feature utilization. To further enhance the algorithm's ability to detect targets, a conical feature fusion structure is introduced, and its comparison with traditional pyramid feature fusion structures is illustrated in Fig. 5.



(a) Pyramid feature fusion structure



(a) Double cone platform feature fusion structure

Fig. 5. Feature fusion structure comparison diagram.

From Fig. 5, it can be observed that the study adopts a feature fusion structure with a special design, allowing the transmission of features from three layers of different sizes. This enhances the efficiency of feature fusion and improves the accuracy of the network. To further accelerate model convergence, the study opts for CIOU_Loss as its loss function, as defined in Formula (14).

$$\begin{cases} CIoU = IoU - \dfrac{\rho^2\left(b, b^{gt}\right)}{c^2} - \alpha v \\ L_{CIoU} = 1 - CIoU \end{cases} \tag{14}$$

In Formula (14), IoU represents the intersection over union of the bounding box and the true bounding box, b represents the center point of the bounding box, b_{gt} represents the center point of the true bounding box, $\rho^2(\cdot)$ represents the Euclidean distance, c represents the shortest minimum enclosing rectangle diagonal length of the two, α represents a positive balancing parameter, and v represents the aspect ratio consistency between the two. Furthermore, these can be expressed as shown in Formula (15).

$$\begin{cases} \alpha = \dfrac{v}{(1 - IoU) + v} \\ v = \dfrac{4}{\pi^2}\left(\arctan\dfrac{w^{gt}}{h^{gt}} - \arctan\dfrac{w}{h}\right)^2 \end{cases} \tag{15}$$

In Formula (15), w^{gt} represents the width of the true bounding box, h^{gt} represents the height of the true bounding box, w represents the width of the bounding box, and h represents the height of the bounding box. CIOU_Loss incorporates penalties for distance from the center and aspect ratio into its loss term for both the bounding box and the true bounding box. This effectively improves the convergence speed of the predicted box during training, thereby enhancing the model's regression localization accuracy.

IV. Performance Testing of UAVs High-Resolution Remote Sensing Image Object Detection System

To evaluate the usability of the proposed UAVs high-resolution remote sensing image object detection system based on the YOLOv5s algorithm and assess the excellence of the optimization conducted in the study, a cost-effective approach was chosen, utilizing the cloud server platform provided by Amazon for testing. The dataset used for testing is an integrated dataset constructed in the study, with 80% randomly selected for training and the remaining 20% for testing. For a comprehensive comparison of research methods, Faster Region-based Convolutional Neural Networks (Fast R-CNN) and Single Shot MultiBox Detector (SSD), which are faster alternatives, were chosen for comparison with the proposed improved YOLOv5s (I-YOLOv5s). Table II shows the software and hardware details, as well as parameter settings used in the testing.

TABLE II. Software and Hardware Details and Parameter Settings

Hardware			**Software**		
Name	***Type***	***Argument***	***Name***	***Type***	***Argument***
Cloud service	Amazon Web Services		OS	Ubuntu	20.04 LTS
Instance type	g4dn.xlarge		Deep learning framework	PyTorch	1.8.0
CPU	Intel Xeon Platinum	8259CL	Algorithm	YOLOv5s	V6.1
GPU	NVIDIA T4 Tensor Core	242 teraFLOPS*	Python	3.8	
RAM	16GB		CUDA	11.0	
MEM	EBS	125GB	cuDNN	Compatible version	
Network performance	-	25Gbps	Other	Numpy, OpenCV, Matplotlib...	
Parameter setting					
Type	***Argument***	***Type***	***Argument***	***Type***	***Argument***
batch_size	16	lr_scheduler	Cosine	label_smoothing	0.0
img_size	640	warmup_lr	0.0	anchor_t	4.0
subdivisions	1	min_lr	0.00001	iou_t	0.2
epochs	300	mosaic	True	cls_pw	1.0
optimizer	SGD	mixup	True	obj_pw	1.0

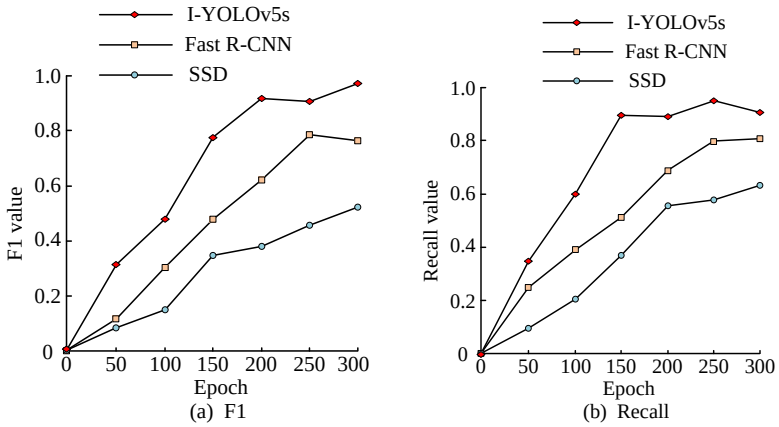


Fig. 6. F1 value and recall value test results of the three algorithms.

Firstly, the convergence performance of the three algorithms was tested, with F1 and Recall values as the metrics. The test results are shown in Fig. 6. From Fig. 6, it can be observed that the I-YOLOv5s algorithm developed by the research achieved optimal state at a faster convergence rate. It reached its best state around the 176th training iteration. Moreover, compared to the Fast R-CNN and SSD algorithms, I-YOLOv5s exhibited superior F1 and Recall values, with F1 value reaching 0.92 and Recall value reaching 0.94.

Next, the average precision and loss variation of the three algorithms were tested, and the results are presented in Fig. 7. It is evident from Fig. 7 that the proposed I-YOLOv5s algorithm attains a maximum average precision of 0.96, surpassing Fast R-CNN and SSD by 0.24 and 0.37, respectively. The lowest loss value for I-YOLOv5s was 0.06, which is 0.09 and 0.13 lower than Fast R-CNN and SSD, respectively.

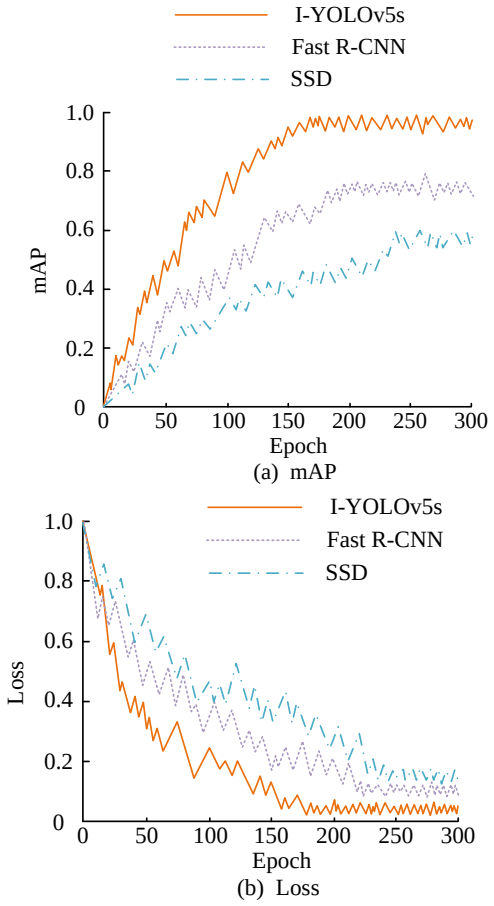


Fig. 7. MAP value and loss value test results of three algorithms.

The ROC curves and P-R curves of the three algorithms were tested, and the results are illustrated in Fig. 8. From Fig. 8, it can be concluded that the curves of the I-YOLOv5s algorithm performed well, encompassing the curves of the other two algorithms in both ROC and P-R curves. This indicated that I-YOLOv5s overall performance is superior to Fast R-CNN and SSD.

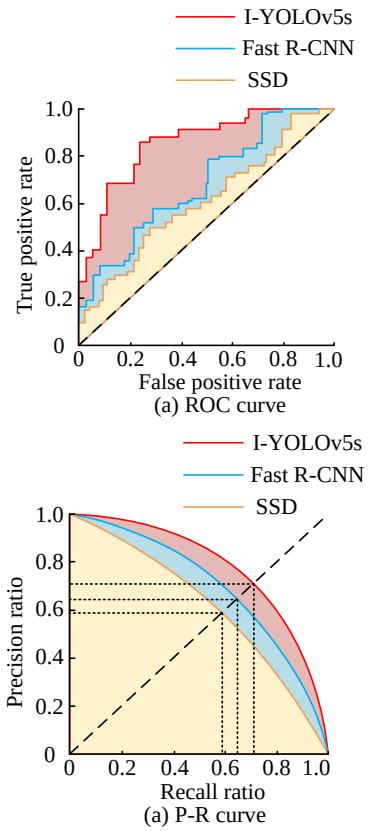
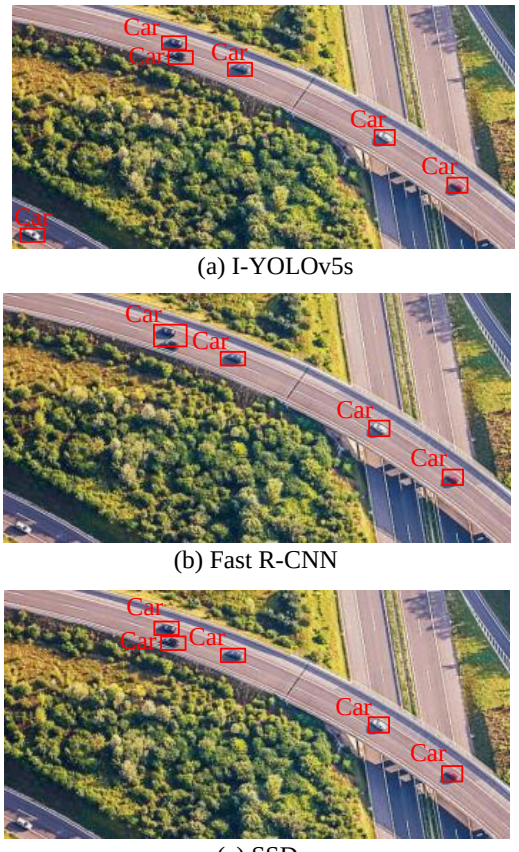


Fig. 8. Comparison results of AOC curves and P-R curves of the three algorithms.

Testing of the image processing results for the three algorithms was conducted to evaluate their image enhancement capabilities. To ensure testing accuracy, two scenes were randomly selected from the dataset for testing, minimizing the impact of experimental errors. The parameters tested included Mean Gradient (MG), Gray Value (GV), Structural Similarity (SS), and Articulation (AR). The results are shown in Table III. It is evident that I-YOLOv5s effectively increases image clarity, with a significant improvement in both mean gradient (97% increase) and articulation (228% increase), while maintaining relatively stable structural similarity, indicating stronger image fidelity.

TABLE III. EVALUATION OF ACTUAL IMAGE PROCESSING CAPABILITY OF THREE ALGORITHMS

Scenario	Algorithms	MG	GV	SS	AR
Scenario 1	Original	2.6514	86.5182	1	2.1547
	I-YOLOv5s	4.6298	91.2647	0.9751	6.9328
	Fast R-CNN	4.1852	85.2648	0.9432	4.2518
	SSD	3.9541	87.6249	0.9215	3.5184
Scenario 2	Original	2.1659	83.2614	1	2.3591
	I-YOLOv5s	4.9251	92.0518	0.9820	6.8521
	Fast R-CNN	3.2691	86.9248	0.9532	5.1244
	SSD	2.9518	85.2694	0.9152	4.2697



(a) I-YOLOv5s

(b) Fast R-CNN

(c) SSD

Fig. 9. The actual target detection test of three algorithms.

Finally, the practical object detection performance of the three algorithms was tested, as in Fig. 9. Only the I-YOLOv5s algorithm proposed by the research achieved complete recognition of vehicle targets in the images. Other algorithms faced challenges in recognizing shadow targets and exhibited difficulties in identifying adjacent targets. In conclusion, the proposed UAV object detection system based on the YOLOv5s algorithm exhibits excellent performance and demonstrates precise target recognition, showcasing strong practicality.

According to the above test results, it can be seen that the proposed UAV high-resolution remote sensing image object detection system based on the YOLOv5s algorithm exhibits excellent performance and strong practicality. The I-YOLOv5s algorithm outperforms traditional Fast R-CNN and SSD algorithms in terms of convergence performance, average accuracy, loss function, ROC curve, and P-R curve. In addition, I-YOLOv5s significantly improves the clarity and fidelity of images. In practical object detection applications, the proposed I-YOLOv5s algorithm can achieve complete recognition of vehicle targets in images, while other algorithms face certain challenges in identifying shadow targets and adjacent targets. These results fully demonstrate the superiority and feasibility of the research method. By combining cloud computing platforms, this efficient and accurate object detection system can be applied to UAV remote sensing image analysis, providing accurate data support for fields such as agriculture, forestry, and urban planning. In summary, the study provides a cost-effective solution for target detection in high-resolution remote sensing images of drones. The proposed detection system based on YOLOv5s algorithm has shown superiority in both performance and practicality, and is expected to contribute to the development of UAV remote sensing applications in China.

V. Conclusion

The application of small UAVs is becoming increasingly widespread. Achieving precise target detection with high-resolution remote sensing imagery on small UAVS platforms is a highly meaningful topic, holding significant importance in fields such as resource monitoring, disaster assessment, and environmental monitoring. Therefore, research was conducted to enhance the YOLOv5s algorithm based on the MPSOC system, thereby improving the accuracy and efficiency of target detection on UAVs. The study involved the modification of the original YOLOv5s algorithm to optimize the performance of target detection in high-resolution images acquired by small UAVs. The I-YOLOv5s algorithm, through structural and parameter adjustments, significantly enhanced the efficiency and accuracy of target detection. Data analysis demonstrated that the algorithm effectively identified objects in images, overcoming interference from shadows and other external factors. After training optimization, I-YOLOv5s achieved its optimal performance state after approximately 176 training iterations, with an F1 value of 0.92 and a Recall value of 0.94, significantly outperforming Fast R-CNN and SSD algorithms. Additionally, the mAP reached 0.96, and the Loss value decreased to 0.06, indicating superior performance compared to similar algorithms. This research provides an efficient solution for high-resolution remote sensing target detection on small UAVs. The optimized algorithm accelerates convergence, offering feasibility for real-time remote sensing data processing. The main limitation of this study is the lack of robustness testing for algorithms and systems in complex environments. Future research can address this issue by testing the performance of algorithms and systems under challenging conditions to validate the practicality and reliability of UAV target detection systems in various complex environments. In addition, algorithms and system design can continue to be optimized to adapt to more complex high-resolution remote sensing image scenes, further improving the accuracy and efficiency of object detection.

Acknowledgment

The research is supported by the study of modular teaching mode and methods of the teamwork of electrical automation major (ZI2021030104).

References

[1] J. Chebbi and Y. Briere, “Robust active disturbance rejection control for systems with internal uncertainties: Multirotor UAV application,” J. Field Robot., vol. 39, no. 4, pp. 426-456, January 2022.

[2] J. Qi, H. Chen, and F. Chen, “Extraction of landslide features in UAV remote sensing images based on machine vision and image enhancement technology,” Neural Comput. Appl., vol. 34, no. 15, pp. 12283-12297, September 2022.

[3] J. Spieck, S. Wildermann, and J. Teich, “A learning-based methodology for scenario-aware mapping of soft real-time applications onto heterogeneous MPSoCs,” ACM Trans. Design Autom. Electr. Syst., vol. 28, no. 1, pp. 4-40, December 2023.

[4] P. S. Titare and D. G. Khairnar, “MPSoC design and implementation using microblaze soft core processor architecture for faster execution of

arithmetic application," Int. J. High Perform. Syst. Architect., vol. 11, no. 3, pp. 156-168, April 2023.

[5] J. A. Belloch, G. Leon, J. M. Badia, A. Lindoso, and E. S. Millan, "Evaluating the computational performance of the Xilinx Ultrascale+ EG Heterogeneous MPSoC," J. Supercomput., vol. 77, no. 2, pp. 2124-2137, June 2021.

[6] F. Gomez, M. Masmono, V. Nicolau, J. Andersson, J. L. Rhun, and D. Trilla, "De-RISC-dependable real-time infrastructure for safety-critical computer systems," Ada User J., vol. 41, no. 2, pp. 107-112, June 2020.

[7] M. R. Gkeka, A. Patras, N. Tavoularis, S. Piperakis, E. Hourdakis, and P. Trahanias, "Reconfigurable system-on-chip architectures for robust visual SLAM on humanoid robots," ACM Trans. Embed. Comput. Syst., vol. 22, no. 2, pp. 24-29, February 2023.

[8] B. Sá, J. Martins, and S. Pinto, "A first look at RISC-V virtualization from an embedded systems perspective," IEEE Trans. Comput., vol. 71, no. 9, pp. 2177-2190, November 2022.

[9] C. Nehnouh, "A new architecture for online error detection and isolation in network on chip," J. High Speed Netw., vol. 26, no. 4, pp. 307-323, December 2020.

[10] J. Spieck, S. Wildermann, and J. Teich, "A learning-based methodology for scenario-aware mapping of soft real-time applications onto heterogeneous MPSoCs," ACM Trans. Design Autom. Electr. Syst., vol. 28, no. 1, pp. 4-40, December 2023.

[11] Y. Zhu, C. Gong, S. Liu, Z. Yu, H. Shao, and G. Yu, "Infrared object detection via patch-tensor model and image denoising based on weighted truncated Schatten-p norm minimization," IET Image Process., vol. 17, no. 6, pp. 1762-1774, February 2023.

[12] Y. Ji, H. Zhang, F. Gao, H. Sun, H. Wei, N. Wang, and B. Yang, "LGCNet: A local-to-global context-aware feature augmentation network for salient object detection," Inform. Sci., vol. 584, no. 1, pp. 399-416, January 2022.

[13] Y. Wan, Z. Liao, J. Liu, W. Song, H. Ji, and Z. Gao, "Small object detection leveraging density-aware scale adaptation," Photogramm. Rec., vol. 38, no. 182, pp. 160-175, May 2023.

[14] Q. Zheng, Y. Li, L. Zheng, and Q. Shen, "Progressively real-time video salient object detection via cascaded fully convolutional networks with motion attention," Neurocomputing, vol. 467, no. 7, pp. 465-475, January 2022.

[15] Y. Liang, G. Qin, M. Sun, J. Qin, J. Yan, and Z. Zhang, "Multi-modal interactive attention and dual progressive decoding network for RGB-D/T salient object detection," Neurocomputing, vol. 490, no. 14, pp. 132-145, June 2022.

[16] S. Hartling, V. Sagan, and M. Maimaitijiang, "Urban tree species classification using UAV-based multi-sensor data fusion and machine learning," GISci. Remote Sens., vol. 58, no. 7/8, pp. 1250-1275, September 2021.

[17] R. R. Bisset, P. W. Nienow, D. N. Goldberg, W. Oliver, A. Loayza-Muroraúl, J. L. Wadham, M. L. Macdonald, and R. G. Bingham, "Using thermal UAV imagery to model distributed debris thicknesses and sub-debris melt rates on debris-covered glaciers," J. Glaciol., vol. 69, no. 276, pp. 981-996, December 2023.

[18] M. Hasanvand, M. Nooshyar, E. Moharamkhani, and A. Selyari, "Machine learning methodology for identifying vehicles using image processing", Artifi. Intell. Appl., vol. 1, no. 3, pp. 170-178, April 2023.

[19] K. Borkar and S. Mukherjee, "Single image dehazing by approximating and eliminating the additional airlight component," Neurocomputing, vol. 400, no. 4, pp. 294-308, March 2020.

[20] J. Yao, Y. Li, B. Yang, and C. Wang, "Learning global image representation with generalized-mean pooling and smoothed average precision for large-scale CBIR," IET Image Process., vol. 17, no. 9, pp. 2748-2763, May 2023.