Sensors and Materials, Vol. 36, No. 8 (2024) 3275–3292
MYU Tokyo

S & M 3732

Multiscale Object Detection Using Adaptive Context Redetecting in Remote Sensing Systems

Yanjun Feng,[1] Jun Liu,[2*] and Yonggang Gai[2]

[1]School of Information Science and Engineering, Shenyang Ligong University, Shenyang 110159, China
[2]School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang 110159, China

(Received January 19, 2024; accepted March 14, 2024)

Keywords: remote sensing, object detection, unmanned aerial vehicle (UAV) system

Object detection is one of the critical technologies in intelligent remote sensing, integrating sensor design, image acquisition, and analysis to locate targets in images, especially crucial for the perception and communication of unmanned aerial vehicle (UAV) systems. In recent years, significant progress has been made in object detection based on deep learning, such as convolutional neural networks (CNNs). However, existing methods, particularly in adapting to remote sensing sensor systems, still face two main challenges: deep learning framework overfitting and inefficient multiscale object detection. To bridge the gap between remote sensing sensor design and image analysis algorithms, we propose a lightweight detector based on context feature adaptive redetecting, ReAC-DETR, an enhanced version of the baseline detector Faster-RCNN. ReAC-DETR further improves with Fast-RCNN by optimizing the feature extraction and fusion branches, making it more suitable for small object detection. It also alleviates the problem of vague object features in remote sensing perception through an object saliency enhancement algorithm. Finally, we propose a spatial context adaptive analysis algorithm to enhance the detector's capability to adjust to multiscale object detection and improve detection precision. ReAC-DETR can provide a more robust detection technology for remote sensing, enabling the system to adapt to various shooting devices, methods, targets, and environments. We tested ReAC-DETR on two challenging datasets and achieved excellent performance.

1. Introduction

Drones are widely used to acquire remote sensing images owing to their high maneuverability, rapid deployment capability, and broad surveillance range.(1) Drone technology is a cross-cutting field, including drone platforms, sensors, and computer vision. With the development of related research, drone-based object detection has been widely deployed in scenes such as road traffic, crop production, and search and rescue, and has far-reaching significance.(2,3) Such devices can collect remote sensing images in real time and extract important object information.

In the modern field of computer vision, deep learning methods specifically designed for UAV sensors are crucial for enhancing object detection technologies. These methods emphasize

*Corresponding author: e-mail: lj_mail_sut@163.com

https://doi.org/10.18494/SAM4996

ISSN 0914-4935 © MYU K.K.
https://myukk.org/

adapting to the needs of UAV sensors to more effectively perform tasks such as intelligent surveillance, autonomous driving, and motion capture. However, the UAV sensors also present challenges.(4) As shown in Fig. 1, the vibration of the drone while hovering changes in the zoom of the visual sensors, and the movements of the subject can all lead to object blurring; the size of the target and the shooting distance can result in different sizes of objects in the image; targets located at the edge, or being obscured, lead to targets being only partially visible or distorted.

Deep learning, particularly CNN technology, has become vital in addressing these challenges. By customizing deep learning models for UAV sensors, small object detection in complex environments can be effectively addressed.(5–7) These models adapt to the image characteristics captured by the sensors, such as changing backgrounds and lighting conditions, thereby improving the accuracy and efficiency of object detection. Therefore, enhanced R-CNN and its derivatives, such as Fast R-CNN and Faster R-CNN, can be employed to optimize these systems further.(8,9) These models handle high-resolution, multi-angle images captured by UAV sensors better and achieve rapid and accurate object detection based on them.

Considering the challenges of remote sensing systems, we have made three improvements to Faster R-CNN. The most significant challenge in remote sensing is the small target problem, for which we have focused on combining detailed and semantic information in visual information analysis.(10) To address the issue of non-prominent object features caused by blurring, we

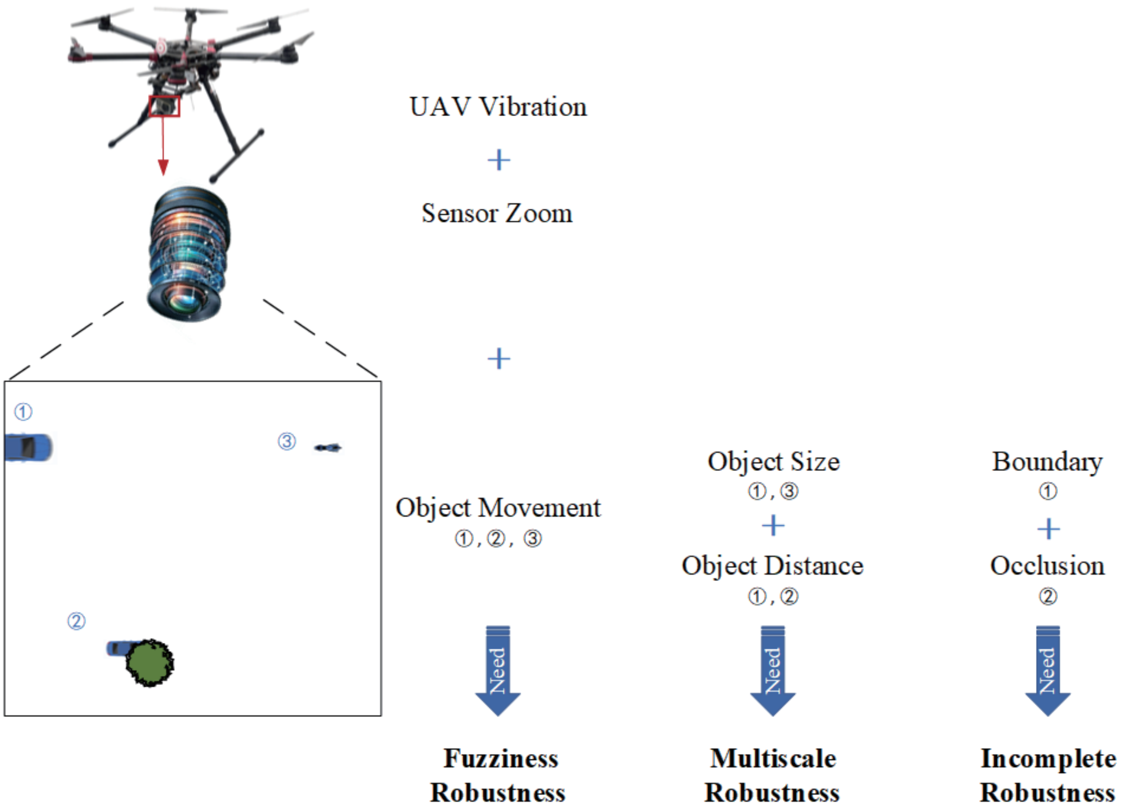


Fig. 1. (Color online) Challenges faced by the intelligent remote sensing system.

perform saliency enhancement based on detail-enhanced feature maps. Finally, redetection is conducted for less reliable targets on the basis of contextual information. Ultimately, this deep learning method ReAC-DETR, specially designed for UAV sensors, not only improves the performance of object detection technology but also brings revolutionary improvements to application fields such as intelligent surveillance and autonomous driving. By closely integrating deep learning algorithms and UAV sensor technology, we can capture, process, and analyze remote sensing images more effectively, achieving efficient object detection in various environments and conditions.

The main contributions of this paper can be summarized as follows:

(1) We propose an intelligent remote sensing system. The communication methods and connections between UAVs, visual sensors, and processors are designed to achieve the real-time object detection and transmission of remote sensing images.

(2) To prevent missed detection, we enhance details and object saliency. On the basis of Faster R-CNN, we designed multilevel feature fusion and proposals, rationally using the relationship between features and collected data. Subsequently, after feature fusion, the estimated objects undergo feature enhancement, generating feature map proposals.

(3) We adjust detection results by incorporating spatial context information. The spatial relationships between similar or dissimilar objects serve as spatial context information, which changes the confidence of less reliable objects to optimize detection accuracy.

(4) Experiments are conducted on two UAV image datasets, the CARPK and Visdrone datasets. Compared with other advanced detectors, ReAC-DETR is more effective in addressing the challenges faced by remote sensing.

2. Related Works

2.1 Object detection for remote sensing

We present a synopsis of several significant research contributions made in recent years to encapsulate the rapid advancements and diverse approaches in object detection for remote sensing. Below is a detailed overview of these pivotal research endeavors:

Chen *et al.*(11) introduced the Domain Adaptation Faster R-CNN to address data scarcity in remote sensing images. This work utilized adversarial training to combat overfitting, proving effective under low-light conditions and enhancing accuracy for low-quality images. Dong *et al.*(12) integrated deformable convolution lateral connection modules and attention-based multilevel feature fusion modules for managing diverse shapes and enhancing feature fusion. Liu *et al.*(2) detected salient objects in remote sensing images by improving multilevel information flow and fusing global and local attention, showing superior performance on public datasets. Aside from CNNs, Transformer has been explored to enhance object detection. Li *et al.*(5) customized Transformer for global spatial relationship capture with an attention-based transfer CNN. Zhang *et al.*(6) proposed EIA-pyramid Vision Transformer, which included an adaptive multigrained routing mechanism, a compact dual-path encoding architecture, and an angle tokenization technique.

In the ever-evolving domain of remote sensing, detecting small objects presents a unique set of challenges, primarily due to their minute size and often complex backgrounds. Recent research has made significant strides in this area, leveraging advanced deep-learning models and innovative techniques to enhance detection accuracy and efficiency.(13)

Xiaolin *et al.*(14) used a super-resolution enhancement module to improve the SA-NET network for small objects. Zhang *et al.*(15) utilized an improved data augmentation method, multigranular deformable convolution, and a High-Resolution Block-based Feature Pyramid for efficient feature extraction. Unlike the above method, some designed one-stage object detectors. Wang *et al.*(16) optimized the YOLOv8 model by creating Wise Intersection over Union (IoU) v3 for bounding box regression loss, a BiFormer attention mechanism for enhanced backbone networks, and a Focal FasterNet block for multiscale feature fusion. This task has also used generative adversarial networks (GANs). Bosquet *et al.*(17) introduced a comprehensive data augmentation pipeline based on small object detection. The DS-GAN architecture generates realistic small objects, alleviating the scarcity of small object instances in datasets.

Each of these studies contributes uniquely to remote sensing object detection, showcasing innovations in model architecture, data handling, and algorithmic efficiency. These advancements collectively push the boundaries of accuracy and performance in detecting various objects in remote sensing imagery.

2.2 Robust object detection in IoT

IoT is rapidly expanding, with robust object detection becoming an increasingly critical component in various IoT applications.(18) This requirement stems from the need to ensure the accurate and reliable identification of objects in diverse and often challenging environments.

For precise object detection in remote surveillance, Gautam and Singh(19) proposed a framework that integrates the YOLO-Lite + SPP model and IoT. Zhang *et al.*(20) incorporated spatial attention into the YOLOv3 framework. Dong *et al.*(21) modified YOLOv5 by containing low-computational ghost modules and a streamlined backbone network. Different from the above methods, Hu and Ni(22) focused on intelligent vehicle license plate recognition and vehicle detection; their unified method efficiently identified high-energy frequency areas in images, facilitating rapid and accurate object detection. Gao *et al.*(23) introduced the Spatial Attention-powered Multi-domain Network by incorporating spatial attention and a multidomain network. Fu *et al.*(24) presented a framework that offloads computation to edge nodes for improved reliability, real-time capability, and flexibility compared with cloud computing. Moreover, Petros *et al.*(25) investigated a cooperative task execution system in edge computing for IoT applications. The proposed system offloads workloads to end devices, collaboratively executing object detection on transmitted sets of images. For multiple object detection, Imran *et al.*(26) employed a collaborative approach involving drones, deep learning, and IoT to enhance surveillance applications in smart cities. Deep learning is coupled with data augmentation and transfer learning.

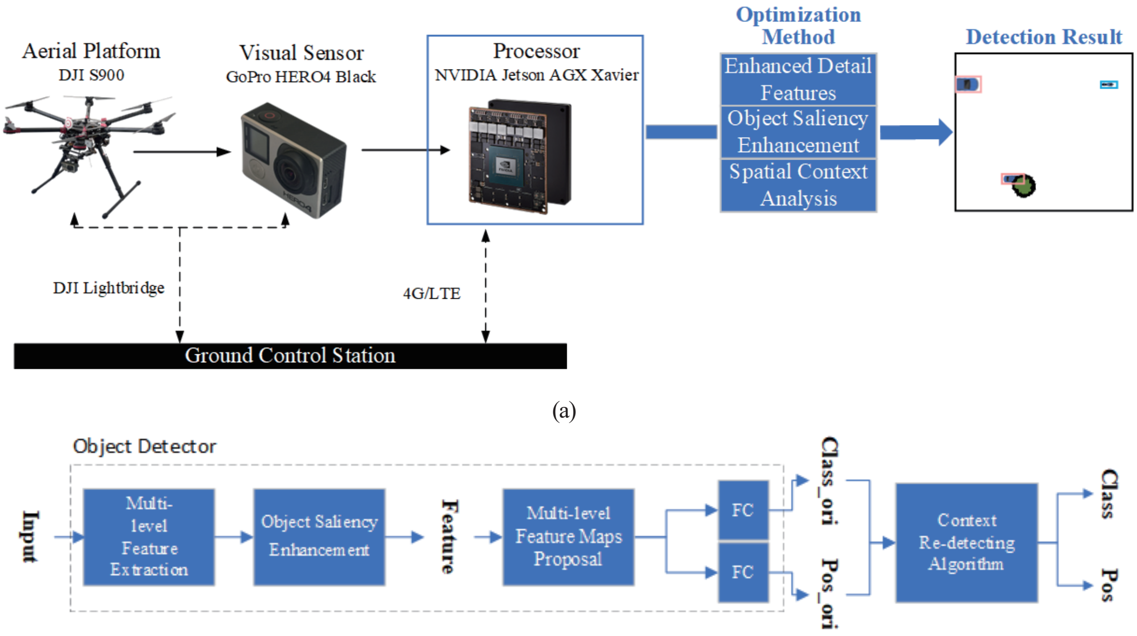
Collectively, these research efforts represent significant advancements in robust object detection for IoT applications. Our work builds on previous work experience to further enhance the robustness of remote sensing.

3. Proposed Method

3.1 Remote sensing systems

As shown in Fig. 2(a), in an intelligent remote sensing system, the drone serves as the carrying platform for the visual sensor. The processor performs object detection on the collected images and enhances the entire system's performance through a built-in optimization algorithm. The drone model used in this system is DJI S900. It is a professional-grade drone with various advanced flight control systems and a three-axis gimbal, ensuring stability and flexibility during filming. It is a hexacopter with a lightweight and high-strength frame design, allowing it to carry relatively heavy equipment. It achieves long-distance communication via Lightbridge. The visual sensor model is GoPro HERO4 Black, known for its compact size and robust performance. It is a portable camera capable of capturing high-quality videos and stills under extreme conditions. It has a 12-megapixel camera for still images with a burst mode of 30 photos per second. It is mounted on the drone and controlled through the drone's remote controller or flight control system.

Edge computing can reduce communication delays, increasing response speed. Additionally, it does not rely on ground infrastructure, which improves operational capabilities in remote or poor network areas. The AI computing platform model is NVIDIA Jetson AGX Xavier, equipped with a high-performance processor and supporting various popular AI frameworks and libraries, making it suitable for complex deep learning. Despite its powerful performance, it is designed for energy efficiency with relatively low power consumption. It provides various I/O and connectivity options for easy integration with visual sensor and ground control terminal devices.



(a)

(b)

Fig. 2. (Color online) (a) System flowchart and (b) ReAC-DETR algorithm framework.

We integrate the AI computing platform with 4G/LTE modules, providing stable remote data transmission.

Traditional drone visual data acquisition systems often face challenges such as multiscale objects or blurred images. This system improves the detection capability of the visual sensing system through methods such as enhancing detail features, enhancing object saliency, and spatial context analysis. Specifically, we have improved the traditional object detector Faster RCNN. The proposed algorithm is primarily deployed on the processor, enhancing the robustness of the target detector, thereby improving the adaptability and accuracy of the entire system to the shooting environment. The proposed lightweight detector ReAC-DETR is shown in Fig. 2(b). This algorithm enhances details through multilevel feature extraction and multilevel feature map proposal, improving the system's ability to detect small objects. It performs object saliency enhancement on the expanded feature maps to alleviate the issue of missed detection due to unclear or small-scale object features. On the basis of the distance information of similar and dissimilar reliable objects, the confidence of less reliable objects is adjusted to enhance the system's detection ability for complex objects. ReAC-DETR mainly consists of an object detector and a context redetection algorithm, which will be introduced in the following two subsections.

3.2 Object detector suitable for remote sensing

As shown in Fig. 2(b), the object detector first performs feature extraction on the data collected by the airborne visual sensor. Considering the application of drone platforms, a lightweight network, MobileNet-v2, is required for feature extraction.(27) It has a higher universality than the recently published MobileNet-v3, as the latter's higher accuracy is dependent on network architecture search to determine parameters. As shown in Fig. 3, to enhance the multiscale robustness of the collection system, we designed the feature extractor with a multilevel feature fusion structure. MobileNet-v2 is divided into four parts (C1, C2, C3, and C4), and the numbers in parentheses in Fig. 3 indicate the number of layers in each part. Then, the features extracted from each part are fused from top to bottom. We use 1 × 1 Conv to fix the number of channels to 256, which can be directly used for subsequent candidate box prediction, then use Content-Aware Feature Up-Sampling (CAFUS) to up-sample the previous feature layer two times.(28) This method introduces learnable parameters to fine-tune the interpolated image, thereby preserving the internal details of the image and enhancing the representation ability of the feature map; after the adjusted two-level features are added, Efficient Channel Attention (ECA) is used for the interaction of information between channels.(29) ECA significantly reduces the model complexity while maintaining performance by adaptively selecting the convolution kernel size.

Inspired by a study,(30) we enhance the detail features after fusion with object saliency enhancement, to enhance the ability of the drone visual information collection system to overcome object blurriness. This method measures the degree of membership of each feature point f_p in the feature map F belonging to the foreground based on the estimated background mean:

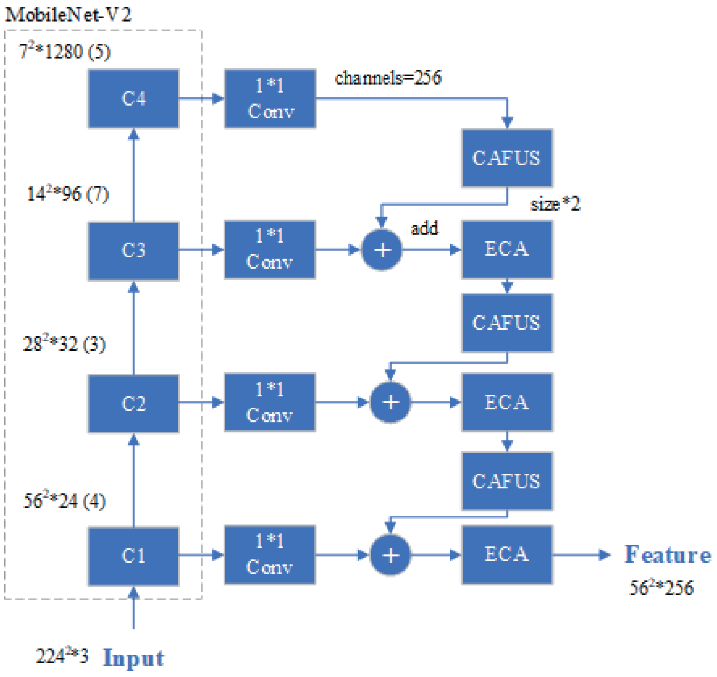


Fig. 3. (Color online) Details of multilevel feature fusion.

$$V = F - mean\left(T \odot F\right). \tag{1}$$

In this equation, T is a learnable tensor used to estimate the background area. The shape of this tensor is the same as that of feature F, with edge elements initialized to 1 and other elements to 0. The difference between each feature point and the mean of the background features represents the saliency of each feature point. Then, on the basis of the saliency, the feature map F is enhanced with object saliency enhancement:

$$\hat{F} = \begin{cases} f_p + f_p \times \left(1 - e^{-\alpha \times V_p}\right), & V_p > 0 \\ f_p. & \text{otherwise} \end{cases} \tag{2}$$

In this equation, features with a degree of membership greater than 0 are enhanced. On the basis of impact of hyperparameter α on model performance (measured by mAP), α is set to 0.3.

Then, the object detector uses the enhanced features to predict candidate boxes, mapping them onto the feature map for the final detection box regression. As shown in Fig. 4, for different candidate box sizes, we use feature points containing corresponding pixel information to generate the scores and coordinates of the candidate boxes, and regress them on the previous level feature map. Compared with the original method, our approach prevents the loss of information for small-scale objects. It ensures the completeness of information for large-scale objects, thereby preventing the missed detection caused by target blurriness or size. Specifically, we combine Region Proposal Network (RPN) and Region of Interest (RoI) Align into one

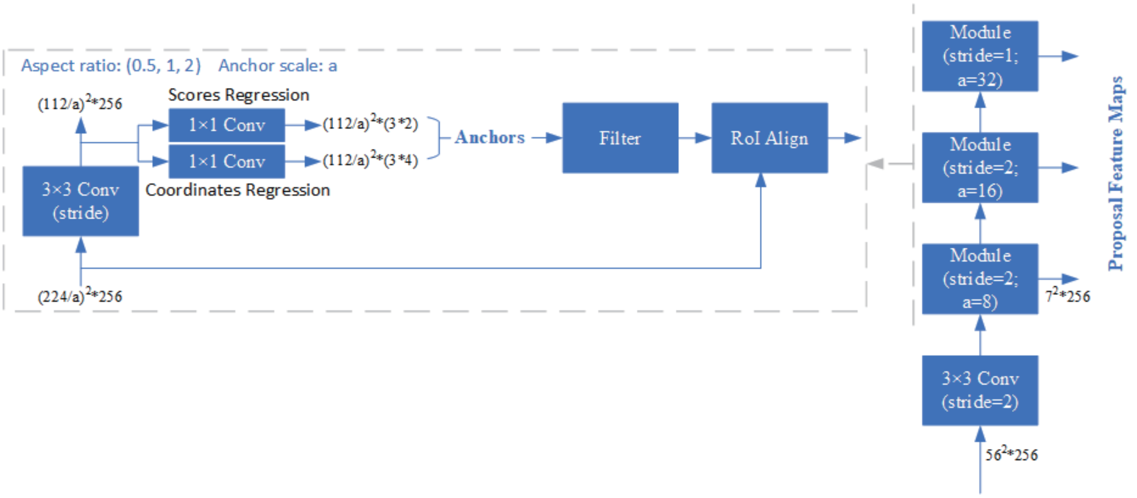


Fig. 4. (Color online) Details of multilevel feature map proposal.

module, and the method used to generate proposal feature maps at each level is similar to the original method. The difference lies in using RoI Align to reduce the dual quantization error of RoI pooling.(31) Each module uses 3 × 3 Conv for feature extraction, followed by 1 × 1 Conv to generate candidate boxes of a specified size. RoIs filtered through a threshold are fed into RoI Align to generate proposal feature maps. We use three consecutive modules of the above type to generate proposal feature maps containing information on the corresponding scale images. Among these, the stride of 3 × 3 Conv in the first two modules is 2 to achieve down-sampling; in the last module, the stride is 1 to prevent missed detection.

Finally, on the basis of the proposal feature maps, two fully connected layers are used to regress the category confidence and coordinate offsets of the detection boxes, thereby completing the entire object detection.

3.3 Context redetecting algorithm

Inspired by a previous work,(32) in drone-captured scenes, objects have a specific distribution pattern due to human conventions, i.e., objects belonging to the same or similar categories are often closer than those belonging to different categories. Therefore, we incorporate the spatial relationship of targets into the context redetecting algorithm. Objects with a confidence level between 0.4 and 0.6 are considered less reliable, whereas those with a confidence level above this range are considered reliable. By calculating the inter- and intraclass distances between different object instances as spatial context, we achieve the redetection of less reliable objects using detection objects with high category confidence. This effectively alleviates the problem of missed detection caused by multiscale targets or blurriness. To the target object *a*, let the same category set of reliable objects be *B* and the different category set be *C*. The process of the context redetecting algorithm is shown in Fig. 5.

First, measure the distance between objects in *B* and *C* and target *a*. Use the Euclidean distance of the center points of the bounding boxes to measure the distance between non-

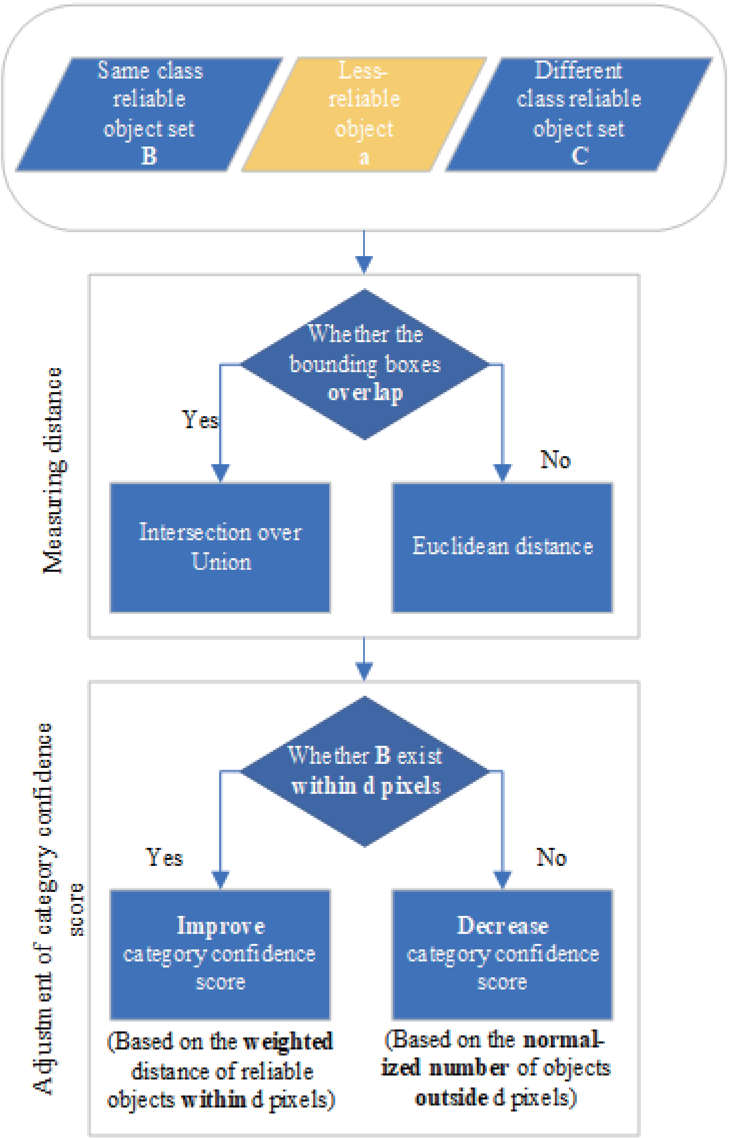

Same class reliable object set B
Less-reliable object a
Different class reliable object set C
Measuring distance
Whether the bounding boxes overlap
Yes
No
Intersection over Union
Euclidean distance
Adjustment of category confidence score
Whether B exist within d pixels
Yes
No
Improve category confidence score
Decrease category confidence score
(Based on the weighted distance of reliable objects within d pixels)
(Based on the normalized number of objects outside d pixels)

Fig. 5. (Color online) Process of the context redetecting algorithm.

overlapping objects. For overlapping objects, their distance is affected by the size and ratio of the bounding boxes, so using the above method will produce errors. Therefore, we introduce IoU to measure the distance between overlapping objects, calculated as

$$D_{IoU} = 1 - \frac{R(a) \cap R(b)}{R(a) \cup R(b)}. \tag{3}$$

In this equation, R() is the area of the object-bounding box and the ratio of the overlapping area to the combined area of two object-bounding boxes is the IoU.

Then, on the basis of the category and spatial relationship of objects, adjust the score *S*() of the less reliable object *a*. The main principle is as follows: If a less reliable object within a certain distance has reliable objects of the same category nearby, then the confidence of the object's existence will be higher. The distance threshold d is set to 500 pixels on the basis of the average distance between different object instances. In this case, the formula for adjusting the confidence is

$$\begin{aligned}\hat{S}(a) = S(a) + \lambda \times mean\left(\left\{\hat{D}(a,b) \mid D(a,b) < d, b \in B\right\}\right) \\ +(0.4-\lambda) \times mean\left(\left\{\hat{D}(a,c) \mid D(a,c) < d, c \in C\right\}\right).\end{aligned} \tag{4}$$

In this equation, $\hat{D}()$ is the normalized distance between objects to [0, 1]. To emphasize the effect of the same-category object, we set the influence weight $\lambda = 0.25$. In the formula, we use the weighted average distance of reliable objects of the same and different categories within the radius d of object a, which increases the confidence score of object a.

Conversely, if there are no reliable objects within the radius d of object a, but there are reliable objects of the same category outside the radius d, then the confidence of the object's existence will be lower. In this case, the formula for adjusting the confidence is

$$\hat{S}(a) = S(a) - \lambda \times \left(\frac{e^{N_b}}{e^{N_b} + e^{N_c}} \right). \tag{5}$$

In this equation, N is the number of same- or different-category reliable objects outside the radius d of object a.

4. Experiments and Discussion

4.1 Datasets and evaluation indexes

4.1.1 Dataset introduction

To validate the performance of the model, we conducted experiments on two representative datasets of UAV images.

(1) CARPK dataset for car detection.[33] This dataset is the largest drone-view parking lot dataset, comprising 1,448 aerial images collected by drones from four parking lots. The dataset only contains a single car target and records the top-left and bottom-right points of each car's bounding box. These images were collected at about 40 m, with a resolution of 1280 pixels × 720 pixels, totaling 89777 cars. We use 989 images for training and validation and the remaining for testing.

(2) Visdrone dataset for multiclass object detection.[34] This challenging dataset was collected using different drone platforms and cameras under various scenes and weather and lighting conditions. The objects in the dataset are divided into ten categories, including some less numerous vehicles, such as awning-tricycles, tricycles, and buses. The dataset contains 10209 static images, of which 6471 are used for training, 548 for validation, and 3190 for testing.

The dataset presents several difficulties: dense detection objects, small-scale objects, uneven data distribution, and severe target occlusion.

4.1.2 Evaluation indices

We set different evaluation metrics for other datasets to measure the model's performance: For the binary CARPK dataset, we use Precision and Recall to measure detection accuracy. Precision indicates the proportion of actual positive samples among the predicted positive samples; Recall suggests the proportion of actual positive samples in the expected positive results to all positive samples. Additionally, we use mAP@0.5:0.95 and mAP@0.5 to measure this dataset. At a fixed IoU threshold, a PR curve can be obtained with Precision values on the y-axis and Recall values on the x-axis. Average precision (AP) is calculated by integrating the Precision values on the PR curve, while mAP is the mean of the AP for all categories. mAP@0.5 sets the IoU to 0.5; mAP@0.5:0.95 represents the average mAP at different IoU thresholds (from 0.5 to 0.95, in steps of 0.05). The general formula for mAP is as follows:

$$mAP = \frac{\sum_{c=1}^{C} AP_c}{C} = \frac{\sum_{c=1}^{C} \int P_c(r)\, dr}{C}, \tag{6}$$

where C represents the total number of categories and $P_c(r)$ stands for the PR curve for category c.

For the multiclass Visdrone dataset, we use the COCO evaluation metrics: AP, AP@0.5, AP@0.75. AP, as previously described, is the mean of mAP values at different IoU thresholds (from 0.5 to 0.95, in steps of 0.05) and is the primary challenge metric; AP@0.5 is the AP value at IoU = 0.5, a Pascal VOC metric; AP@0.75 is the AP value at IoU = 0.75, a strict metric.

4.2 Main comparison results

In this experiment, we compared the proposed object detection model with other advanced detectors on two UAV image datasets to measure the object detection capability of the intelligent remote sensing system.

On the CARPK dataset, we compared the model on the basis of two types of evaluation. First, we compared on the basis of Precision and Recall, as shown in Table 1. The method proposed in

Table 1
Comparison of results on the CARPK dataset, evaluated on the basis of Precision and Recall.

Method	Precision (%)	Recall (%)
Faster-RCNN[8]	84.85	81.27
SDD[35]	89.09	87.52
FCN8s[36]	89.02	86.03
SegNet[37]	87.77	86.89
U-Net[38]	91.43	89.61
FICLAR-Net[39]	94.19	90.14
ReAC-DETR	**95.30**	**92.40**

this paper is the best among the seven models in both metrics, with Precision 1.11% higher than the second-best FICLAR-Net and Recall 2.26% higher.(39) Compared with the original model Faster-RCNN,(8) it is 10.45 and 11.13% higher in these two metrics, respectively, indicating the effectiveness of the improvement module. Then, the models are compared on the basis of mAP at different IoU thresholds, as shown in Table 2. Our model also ranks the best in both metrics.

As shown in Table 3, on the Visdrone dataset, we compared multicategory predictions using AP at three IoU thresholds. YOLOv8 is the latest generation of the popular object detection network YOLO series.(47) The YOLO series treats object detection as a single regression problem, directly predicting the image's bounding boxes and class probabilities. Although the detection speed has been improved, the accuracy is relatively lower. The YOLOv8 algorithm has also been enhanced for small object detection. However, our method still shows an improvement of more than 7% in all three metrics compared with YOLOv8. In particular, AP@0.5 is improved by 12.2%. Compared with the most outstanding comparison model, AMRNet,(49) our method also shows improvements in all three metrics.

4.3 Visualization results

To validate that the model can successfully address the many challenges in drone object detection, we selected representative challenging images for visualization and analyzed the contribution of each optimization algorithm.

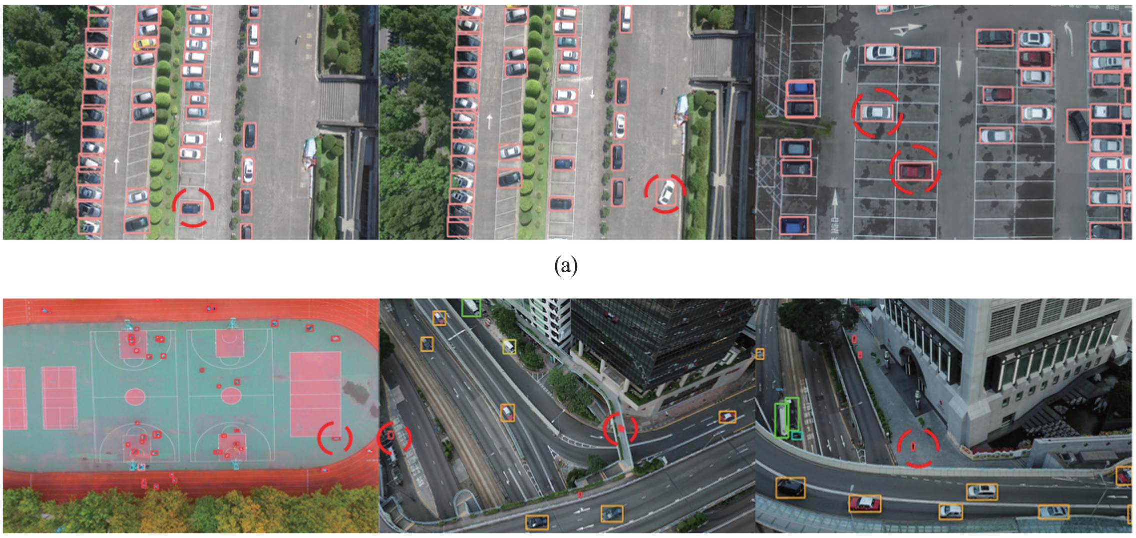
Figure 6 shows small, isolated samples from the two datasets. The model's ability to successfully detect these objects indicates its strong capability in small sample detection. Moreover, in Fig. 6(b), the isolated objects near the overpass are far from similar objects, exceeding 500 pixels. Thus, the context redetecting algorithm has a lesser effect. This illustrates that the multilevel feature extraction and feature map proposal positively enhance the system's

Table 2
Comparison of results on the CARPK dataset, evaluated on the basis of mAP.

Method	mAP@0.5 (%)	mAP@0.5:0.95 (%)
SF-SSD(40)	90.10	—
CD-YOLOv5(41)	95.80	63.10
Modified YOLOv5(42)	94.90	61.10
ReAC-DETR	**96.60**	**64.40**

Table 3
Comparison of results on the Visdrone dataset, evaluated on the basis of AP(COCO).

Method	AP (%)	AP@0.5 (%)	AP@0.75 (%)
FSAF(43)	20.8	36.4	20.5
VFNet(44)	23.1	37.3	24.1
TOOD(45)	24.4	39.8	25.3
DDOD(46)	23.3	38.2	24.2
YOLOv8(47)	25.9	42.9	26.4
DroneNet(48)	29.6	50.4	29.6
AMRNet(49)	31.7	52.7	33.1
ReAC-DETR	**33.5**	**55.1**	**34.7**



(a)

(b)

Fig. 6. (Color online) Small isolated samples were not missed (emphasized with red dashed lines): from the (a) CARPK and (b) Visdrone datasets.

detection robustness across multiple scales. Object saliency enhancement is precisely the enhancement of small target object features through the detail-enhanced feature map output after multilevel fusion.

Figure 7 lists three examples from the CARPK dataset containing severely occluded vehicles. Our method effectively prevents the missed detection of occluded objects. In the first example, although the color of the vehicle is similar to the occluding object, it can still be effectively detected, and the detection box only includes the unobstructed part of the vehicle. This indicates that object saliency enhancement effectively distinguishes foreground and background, and enhances vehicle features. In the third example, the vehicle is severely impeded, retaining only a small part of the visual information. Owing to the model's multilevel feature structure, this part of the feature is preserved, thus preventing missed detection. Furthermore, Fig. 8 lists vehicles at the edge of the image. These vehicles can still be correctly detected, even if only partially present in the frame. This is also a good demonstration of the model's robustness to objects with weakened visual information. Additionally, this shows that object saliency enhancement, initialized with edge features as background features, has learned during training to distinguish between background and foreground correctly.

Figure 9 lists images from the two datasets containing densely packed objects. Dense distribution can lead to occlusion, feature confusion, and thus a higher false-positive rate. However, the context redetecting algorithm can adjust the confidence of less reliable objects using surrounding similar objects, thereby improving the object recall rate. For example, in Fig 9(b)'s first image, some pedestrian objects overlap, but the object distinction is still well accomplished. In addition to dense distribution, the impact of the context redetecting algorithm is more significant, playing an essential role in detecting other challenging objects and effectively mitigating the problems of missed and false detections.

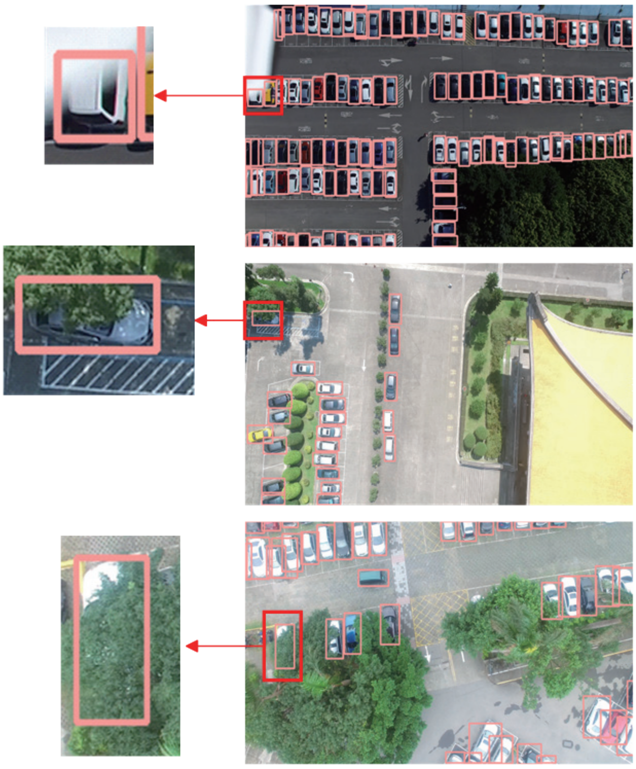
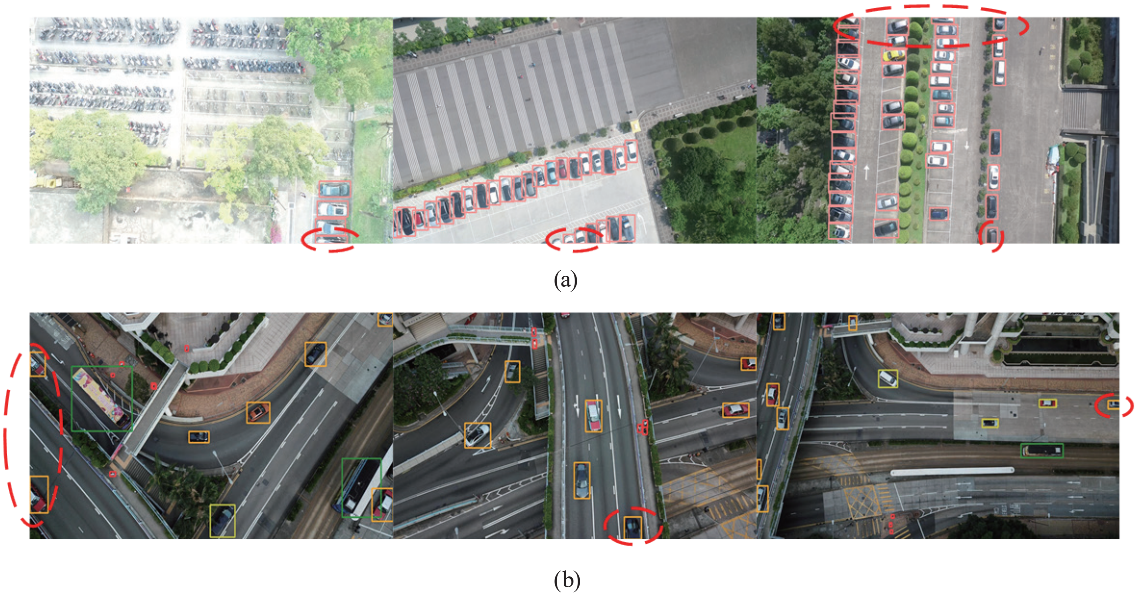


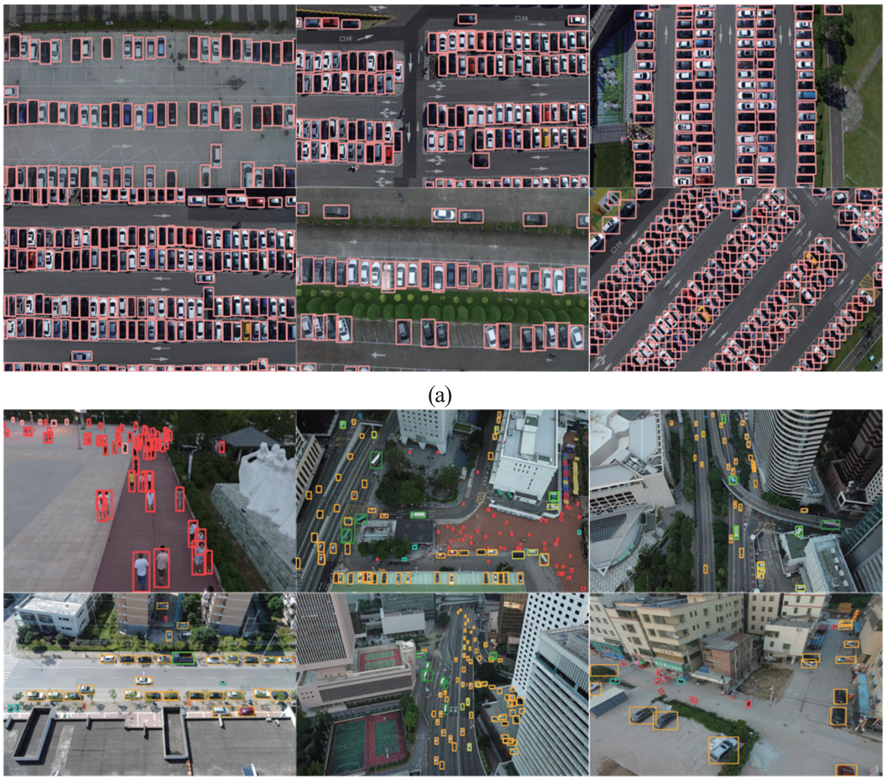
Fig. 7. (Color online) Objects were disturbed but correctly detected.



(a)

(b)

Fig. 8. (Color online) Incomplete objects at the edges were correctly detected (emphasized with red dashed lines), from the (a) CARPK and (b) Visdrone datasets.


(a)

(b)

Fig. 9. (Color online) Detection of densely packed objects: from the (a) CARPK and (b) Visdrone datasets.

4.4 Discussion

On the basis of the designed method, we improve the classic object detector Faster-RCNN and build ReAC-DETR. In comparison using the CARPK dataset, this method has improved Faster-RCNN's Precision and Recall by more than 10%. To implement the detail enhancement strategy, we specifically utilize the characteristics of different levels of features, designing feature extraction and feature map proposals to use shallow detail information and deep semantic information fully. The detection results of isolated small samples in Fig. 6 verify that this strategy and implementation can effectively improve the system's ability to detect small samples. To enhance object saliency, we assess the probability of foreground feature points on the detail-enhanced feature maps and assign values to regions accordingly. This strategy enables the system to detect objects with incomplete or unobvious visual information, as shown in the occluded samples in Fig. 7 and the insufficient edge samples in Fig. 8. After regressing detection boxes on the basis of proposal feature maps, we adjust the confidence of the regressed less reliable detection boxes on the basis of the distance information of similar and dissimilar reliable objects. In the dense object examples of Fig. 9, which contain many challenging samples such as occlusions, spatial context analysis mitigates the system's problems of false negatives and false positives for such samples.

With these strategies, this method enhances the system's robustness against numerous challenges. To measure this method's detection level, it was compared with several advanced models on the CARPK and Visdrone datasets. Our model achieved the best performance under multiple evaluation metrics, further proving the effectiveness of the three optimization strategies in this system and the overall advanced design of the system.

5. Conclusion

Remote sensing has been widely applied. We have designed an intelligent remote sensing system whose main components are drones, visual sensors, and an AI algorithm platform. In the system above, we elucidate these components' interconnections and communication methods. Moreover, to address issues affecting the detection accuracy of the intelligence remote sensing system, such as blurred or multiscale collection objects, we proposed ReAC-DETR based on Faster-RCNN. The improvement strategies mainly include detail enhancement, object saliency enhancement, and spatial context analysis. These three strategies complement each other, enabling the system to effectively cope with the issues above.

To validate the proposed method, we conducted extensive experiments on the binary-class CARPK dataset, which contains many dense objects, and the multiclass Visdrone dataset, captured with various devices and scenarios. Our model achieved the best performance under multiple evaluation metrics compared with the state-of-the-art methods. Additionally, we visualized the detection results of different challenging samples, proving that each optimization method provides a positive impact on the system.

In our future work, we will consider an object detection architecture that better balances computational demands and accuracy to enhance the real-time performance of the intelligent remote sensing system. Furthermore, we plan to explore video-based object detection, utilizing temporal video information to improve detection accuracy.

Acknowledgments

This work was supported by the General Science Research Project from the Liaoning Provincial Department of Education under Grant no. LJKMZ20220610.

References

1 C. Zhen, L. Kooistra, W. Wang, L. Guo, and J. Valente: Drones **7** (2023) 620. https://doi.org/10.3390/drones7100620

2 Y. Liu, S. Zhang, Z. Wang, B. Zhao, and L. Zou: IEEE Trans. Geosci. Remote Sens. **60** (2022) 1. https://doi.org/10.1109/TGRS.2022.3141953.

3 J. Yu, H. Gao, D. Zhou, J. Liu, Q. Gao, and Z. Ju: IEEE Trans. Cybern. **52** (2022) 13738. https://doi.org/10.1109/tcyb.2021.3114031.

4 X. Wang, A. Wang, J. Yi, Y. Song, and A. Chehri: Remote Sens. **15** (2023) 3265. https://doi.org/10.3390/rs15133265

5 Q. Li, Y. Chen, and Y. Zeng: Remote Sens. **14** (2022) 984. https://doi.org/10.3390/rs14040984

6 C. Zhang, J. Su, Y. Ju, K. -M. Lam, and Q. Wang: IEEE Trans. Geosci. Remote Sens. **61** (2023) 1. https://doi.org/10.1109/TGRS.2023.3292418.

7 J. Yu, Y. Xu, H. Chen, and Z. Ju: IEEE Trans. Neural Networks Learn. Syst. **35** (2022) 8869. https://doi.org/10.1109/tnnls.2022.3216084

8 Z. Kaihua and H. Shen: Remote Sens. **14** (2022) 579. https://doi.org/10.3390/rs14030579

9 J. Yu, W. Zheng, Y. Chen, Y. Zhang, and R. Huang: Front. Neurosci. **17** (2023) 1219363. https://doi.org/10.3389/fnins.2023.1219363.

10 F. Changhong, L. Kunhan, Z. Guangze, Y. Junjie, C. Ziang, L. Bowen, and L. Geng: Artif. Intell. Rev. **56** (2023) 1417. https://doi.org/10.1007/s10462-023-10558-5

11 J. Chen, J. Sun, Y. Li, and C. Hou: Multimedia Tools Appl. **81** (2022) 12093. https://doi.org/10.1007/s11042-021-10833-z

12 X. Dong, Y. Qin, Y. Gao, R. Fu, S. Liu, and Y. Ye: Remote Sens. **14** (2022) 3735. https://doi.org/10.3390/rs14153735

13 J. Yu, H. Gao, Y. Chen, D. Zhou, J. Liu, and Z. Ju: IEEE Trans. Hum.-Mach. Syst. **52** (2022) 784. https://doi.org/10.1109/thms.2022.3144951

14 F. Xiaolin, H. Fan, Y. Ming, Z. Tongxin, B. Ran, Z. Zenghui, and G. Zhiyuan: Pattern Recognit. Lett. **153** (2022) 107. https://doi.org/10.1016/j.patrec.2021.11.027

15 H. Zhang, M. Li, D. Miao, W. Pedrycz, Z. Wang, and M. Jiang: Pattern Recognit. **143** (2023) 109801. https://doi.org/10.1016/j.patcog.2023.109801

16 G. Wang, Y. Chen, P. An, H. Hong, J. Hu, and T. Huang: Sensors **23** (2023) 7190. https://doi.org/10.3390/s23167190

17 B. Bosquet, D. Cores, L. Seidenari, V. M. Brea, M. Mucientes, and A. D. Bimbo: Pattern Recognit. **133** (2023) 108998. https://doi.org/10.1016/j.patcog.2022.108998

18 J. Yu, H. Gao, Y. Chen, D. Zhou, J. Liu, and Z. Ju: IEEE Trans. Cognit. Dev. Syst. **14** (2022) 1654. https://doi.org/10.1109/tcds.2021.3131253.

19 A. Gautam and S. Singh: Wireless Pers. Commun. **118** (2021) 2121. https://doi.org/10.1007/s11277-021-08071-5

20 Y. Zhang, P. Zhao, D. Li, and K. Konstantin: IEEE Access **8** (2020) 165863. https://doi.org/10.1109/ACCESS.2020.3022645.

21 C. Dong, C. Pang, Z. Li, X. Zeng, and X. Hu: IEEE Access **10** (2022) 123736. https://doi.org/10.1109/ACCESS.2022.3223997.

22 L. Hu and Q. Ni: IEEE Internet Things J. **5** (2018) 747. https://doi.org/10.1109/JIOT.2017.2705560.

23 H. Gao, L. Yu, I. A. Khan, Y. Wang, Y. Yang, and H. Shen: IEEE Internet Things J. **10** (2023) 2811. https://doi.org/10.1109/JIOT.2021.3099855.

24 M. Fu, S. Sun, K. Ni, and X. Hou: Proc. 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conf. (2019, APSIPA ASC) 1838–1842. https://doi.org/10.1109/APSIPAASC47483.2019.9023187.

25 A. Petros, D. Karampatzakis, G. Iosifidis, T. Lagkas, and A. Nikitas: Appl. Sci. **13** (2023) 4982. https://doi.org/10.3390/app13084982

26 A. Imran, M. Ahmad, A. Chehri, M. M. Hassan, and G. Jeon: Remote Sens. **14** (2022) 4107. https://doi.org/10.3390/rs14164107

27 M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen: Proc. 2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition (2018, IEEE) 4510–4520. https://doi.org/10.1109/CVPR.2018.00474.

28 Z. Kaihua and H. Shen: Remote Sens. **14** (2022) 579. https://doi.org/10.3390/rs14030579

29 Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu: Proc. 2020 IEEE/CVF Conf. Computer Vision and Pattern Recognition (2020, CVPR) 11531–11539. https://doi.org/10.1109/CVPR42600.2020.01155.

30 W. Ma, N. Li, H. Zhu, L. Jiao, X. Tang, Y. Guo, and B. Hou: IEEE Trans. Geosci. Remote Sens. **60** (2022) 1. https://doi.org/10.1109/TGRS.2022.3140856.

31 K. He, G. Gkioxari, P. Dollár, and R. Girshick: Proc. 2017 IEEE Int. Conf. Computer Vision (2017, ICCV) 2980–2988. https://doi.org/10.1109/ICCV.2017.322.

32 X. Liang, J. Zhang, L. Zhuo, Y. Li, and Q. Tian: IEEE Trans. Circuits Syst. Video Technol. **30** (2020) 1758. https://doi.org/10.1109/TCSVT.2019.2905881.

33 M.-R. Hsieh, Y.-L. Lin, and W.-H. Hsu: Proc. 2017 IEEE Int. Conf. Computer Vision (2017, ICCV) 4145–4153. https://arxiv.org/pdf/1707.05972v1

34 P. Zhu, L. Wen, X. B., H. Ling, and Q. Hu: arXiv. (2018) 1804.07437. https://arxiv.org/abs/1804.07437

35 S. Ren, K. He, R. Girshick, and J. Sun: IEEE Trans. Pattern Anal. Mach. Intell. **39** (2017) 1137. https://doi.org/10.1109/TPAMI.2016.2577031.

36 T. Tianyu, S. Zhou, Z. Deng, L. Lei, and H. Zou: Remote Sens. **9** (2017) 1170. https://doi.org/10.3390/rs9111170

37 E. Shelhamer, J. Long, and T. Darrel: IEEE Trans. Pattern Anal. Mach. Intell. **39** (2017) 640. https://doi.org/10.1109/TPAMI.2016.2572683.

38 V. Badrinarayanan, A. Kendall, and R. Cipolla: IEEE Trans. Pattern Anal. Mach. Intell. **39** (2017) 2481. https://doi.org/10.1109/TPAMI.2016.2644615.

39 O. Ronneberger, P. Fischer, and T. Brox: Proc. Medical Image Computing and Computer-Assisted Intervention (2015, MICCAI) 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

40 Q. Song, X. Chen: IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **15** (2022) 7988. https://doi.org/10.1109/JSTARS.2022.3206036.

41 J. Yu, H. Gao, J. Sun, D. Zhou, and Z. Ju: IEEE Trans. Cognit. Dev. Syst. **14** (2022) 1574. https://doi.org/10.1109/TCDS.2021.3124764.

42 D. L. Nguyen, X. T. Vo, A. Priadana, and K. H. Jo: Proc. The 12th Conf. Information Technology and Its Applications (2023, CITA). https://doi.org/10.1007/978-3-031-36886-8_9

43 M. H. Hamzenejadi and H. Mohseni: Proc. 2022 12th Int. Conf. Computer and Knowledge Engineering (2022, ICCKE) 231–236. https://doi.org/10.1109/ICCKE57176.2022.9960099.

44 C. Zhu, Y. He, and M. Savvides: Proc. 2019 IEEE/CVF Conf. Computer Vision and Pattern Recognition (2019, CVPR) 840–849. https://doi.org/10.1109/CVPR.2019.00093.

45 H. Zhang, Y. Wang, F. Dayoub, and N. Sünderhauf: Proc. 2021 IEEE/CVF Conf. Computer Vision and Pattern Recognition (2021, CVPR) 8510–8519. https://doi.org/10.1109/CVPR46437.2021.00841.

46 C. Feng, Y. Zhong, Y. Gao, M. R. Scott, and W. Huang: Proc. 2021 IEEE/CVF Int. Conf. Computer Vision (2021, ICCV) 3490–3499. https://doi.org/10.1109/ICCV48922.2021.00349.

47 Z. Chen, C. Yang, Q. Li, F. Zhao, Z.-J. Zha, and F. Wu: Proc. 29th ACM Int. Conf. Multimedia (2021, MM) 4939–4948. https://doi.org/10.1145/3474085.3475351

48 W. Xiandong, F. Yao, A. Li, Z. Xu, L. Ding, X. Yang, G. Zhong, and S. Wang: Drones **7** (2023) 441. https://doi.org/10.3390/drones7070441

49 Z. Wei, C. Duan, X. Song, Y. Tian, and H. Wang: arXiv. (2020) 2009.07168. https://arxiv.org/abs/2009.07168

About the Authors

Yanjun Feng received her B.E. degree from Liaoning University of Technology in 1997 and her M.E. degree from Shenyang University of Technology in 2000. She is currently an associate professor at Shenyang Ligong University. Her main research interests include IoT technology and intelligent information processing. (braverfyj@126.com)

Jun Liu received his B.E. and M.E. degrees from Shenyang University of Technology in 1995 and 2000, respectively, and his Ph.D. degree from the Graduate University of Chinese Academy of Sciences and the Shenyang Institute of Automation, Chinese Academy of Sciences, in 2010. He is currently a professor at Shenyang Ligong University. His main research interests include intelligent sensors and detection technology, image and signal processing, and intelligent robots. (lj_mail_sut@163.com)

Yonggang Gai received his B.E. degree from Shenyang University of Technology in 1993. He is currently a senior experimentalist at Shenyang Ligong University. His main research interests include control and image processing. (12576962@qq.com)