Geomatica 77 (2025) 100049

Contents lists available at ScienceDirect

Geomatica

journal homepage: www.sciencedirect.com/journal/geomatica

LW-UAV–YOLOv10: A lightweight model for small UAV detection on infrared data based on YOLOv10

Phat T. Nguyen [a,*], Giang L. Nguyen [b,1], Duy D. Bui [a,1]

[a] *Le Quy Don Technical University, 236 Hoang Quoc Viet, Ha Noi, 100000, Viet Nam*
[b] *Institute of Information Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Ha Noi, 100000, Viet Nam*

ARTICLE INFO

Keywords:
UAV
Small target detection
Deep learning
YOLOv10
Lightweight model

ABSTRACT

Advancements in unmanned aerial vehicle (UAV) technology have driven their widespread use in both civil and military sectors. Among various UAV types, small UAVs pose significant threats to global security, necessitating effective detection solutions. Real-time detection of small UAVs, especially under challenging conditions, remains a critical issue in computer vision. This study introduces LW-UAV–YOLOv10, an enhanced YOLOv10-based detection model optimized for small UAV detection using infrared data in mountainous terrain. Architectural improvements in the Backbone and Head modules enhance detection accuracy while maintaining a lightweight structure. Experimental results show that LW-UAV–YOLOv10 surpasses existing YOLO models in accuracy, speed, and suitability for real-time applications, offering a promising solution for UAV detection in complex environments.

1. Introduction

Recent years, it have marked the rapid development of the UAV manufacturing and trading industry. UAVs are increasingly asserting their important role in supporting the development of agriculture, forestry, fisheries; television, rescue, transportation, and national defense and security on a global scale. However, the use of UAVs is also causing many serious threats to security and safety such as in aviation safety, drug trafficking crimes, infringement of people's privacy, and violations of border security [1]. Many countries around the world have issued legal documents to manage the operation of UAVs, but security threats still occur frequently, especially with small, low-flying UAVs. One of the measures to minimize the negative impacts and harms from UAV operations is to accurately detect them. Various approaches have been used to detect UAVs, including radar [2], radio frequency analysis [3], acoustic methods [4], and visual techniques [5]. Although the above methods have achieved many good results, there are still many limitations such as high cost, difficult implementation, and poor accuracy when detecting small, low-flying UAV targets. Thanks to advances in the field of computer science, the field of computer vision has made great strides in recent times. Machine learning methods such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF), as well as deep learning models, have been effectively applied to detect UAVs, outperforming traditional methods [6,7]. YOLO network introduced in 2015 [8] is a deep learning model widely applied in the field of real-time target detection on video and image data.

YOLO is a one-stage target detection algorithm, which is outstanding in its ability to learn complex features and detect targets in real time, outperforming two-stage algorithms such as R-CNN, Faster R-CNN [9], and SSD [10]. The YOLOv10 version [11,12], released in May 2024, significantly improves its structure and performance. The main advantage of the YOLO network is its ability to divide the image into a grid matrix, and predict the bounding box and probability at the same time, which optimizes real-time target detection. This mechanism has become an important theoretical foundation for the development of subsequent versions of YOLO. In less than 10 years, the research community has upgraded from YOLOv1 to YOLOv11, demonstrating great progress in architectural design, training strategies, and optimization techniques. Later versions of YOLO have focused on addressing challenges such as small target detection, occlusion, and improving performance on different datasets [13]. For the problem of UAV target detection, the use of YOLO versions have received a lot of research attention recently.

Dewangan et al. [14] used RGB image datasets to detect small UAVs using YOLOv5 and YOLOv7 models. Instead of focusing on improving the YOLO model, the authors proposed data preprocessing techniques, including grayscale adjustment, color enhancement, and edge enhancement, before feeding them into the YOLO models. The results showed

[*] Corresponding author.
E-mail address: nguyenphat@lqdtu.edu.vn (P.T. Nguyen).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.geomat.2025.100049
Received 24 October 2024; Received in revised form 23 January 2025; Accepted 27 January 2025
Available online 7 February 2025

that this preprocessing method significantly improved the performance of YOLOv5 and YOLOv7 in UAV detection on RGB data.

Yuan et al. [15] focused on small UAV detection on infrared data under different conditions such as mountains, sea, sky, and urban areas. The authors proposed the Infrared Small Target Detection Module (IRSTDM), which combines semantic information from deep layers and spatial information from shallow layers. This method effectively extracts the features of small targets and minimizes the loss of information during multi-layer downsampling. In addition, the team uses the Wasserstein distance and Gaussian distribution to optimize the loss function. The contributions from this research include an effective target detection method and a new infrared dataset, which is valuable for further research.

The latest paper on the problem of detecting small UAV targets using the YOLO network is introduced in [16]. The authors optimized YOLOv8 for detecting small UAVs by improving the network architecture. In which, the Backbone part uses the SPD-conv module instead of Conv to reduce the number of calculations. The Neck part integrates the GAM attention mechanism to increase the ability of feature fusion, and the Head part adds the function of detecting very small targets instead of large targets. Experimental results on RGB data show that the proposed method is highly effective in detecting small UAVs under different lighting and airspace conditions. However, the integration of the GAM module increases the inference time, affecting the detection speed of the model. Recent research results have potential for practical application and deployment. However, there are still some unresolved issues, such as real-time detection of small UAV targets and efficient scalability on devices with limited resources. The problem is that the development of more optimized models should focus on reducing computational complexity without significantly reducing accuracy. This is especially important for applications where hardware resources are limited and real-time processing requirements are high. To address this issue, we proposed the YOLOv10 network optimization model for the problem of detecting small UAV targets using infrared data in mountainous conditions. Our research contributions will be presented in detail later and can be summarized as follows:

- Improve the Backbone structure of the YOLOv10 network to lighten the proposed model. The model will be more lightweight, leading to faster inference speed. This is especially important in real-time applications, where low latency is crucial.

- Improve the Head network structure of the YOLOv10 network by adding very small target parts and removing large targets. To enhance the ability to detect very small targets and increase the inference speed.

The subsequent sections of this paper are structured as follows: Section 2 describes the methodology, Section 3 displays the experimental results, and Section 4 summarizes the conclusions.

2. Methodology

2.1. Original YOLOv10

The overall architecture of YOLOv10 is illustrated in Fig. 1 [17]. Compared to prior versions, YOLOv10 introduces key improvements, most notably the removal of the Non-Maximum Suppression (NMS) post-processing step. This is achieved through an NMS-independent training method, employing a consistent double assignment technique to reduce inference latency. This approach generates multiple predictions per target, selecting the bounding box with the highest IOU or confidence during inference, thus minimizing processing time while preserving accuracy. Additionally, YOLOv10 integrates architectural and optimization enhancements, including adjustments to convolutional layers and the incorporation of partial self-attention (PSA) modules, which improve performance without increasing computational costs. A significant improvement is the introduction of a one-to-one head, shown in Fig. 2. This head maintains the one-to-many head structure but is optimized with the other head during training, allowing the model to process multiple potential bounding boxes per target, providing more information for the Backbone and Neck components. A consistency metric is introduced to optimize the one-to-one head towards the one-to-many head. This metric evaluates the IOU consensus between the two heads, refining predictions accordingly. During inference, only the one-to-one head is used, enhancing prediction efficiency. Finally, YOLOv10 introduces an intrinsic ranking metric to analyze overlap across model stages. The results indicate that deep stages and previous YOLO versions tend to overlap, which hampers performance. To address this, a compact inverted block (CIB) design is implemented.

2.2. Improved YOLOv10 model

The optimized consistent matching metric considers the one-to-one head in the direction of the one-to-many head. A metric that measures the IOU agreement between both heads and adjusts their predictions. During inference, the one-to-many head is discarded and we use the one-to-one head to make predictions. In addition, the YOLOv10 model introduces an intrinsic ranking metric to analyze the redundancy of model stages. The analysis shows that the deep stages and the previous YOLO models tend to be more redundant. This causes a negative impact on the efficiency and detection performance of the model. To address this issue, they introduce the compact inverted block (CIB) design [Tan et al. 2024].

2.2.1. Improvement of the backbone network

Specifically, in the Backbone architecture of YOLOv10, the feature extraction network structures SCDown (Spatialchannel decoupled downsampling) and C2fCIB are used. The principle and role of SCDown is to separate the spatial and channel dimensions. First, increase the number of channels by using 1 × 1 point-by-point convolution, followed by spatial downsampling by using 3 × 3 convolution for each depth. This method reduces the computational cost while still preserving the target feature information. On the other hand, CIB (Compact Inverse Block) replaces the original Bottleneck structure in C2f to reduce redundant information. Although the Backbone structure of YOLOv10 has many improvements compared to previous versions, balancing the weight and lightness of the model with the accuracy of the model for specific problems still requires appropriate designs. In the structure of YOLO network, the Backbone part of YOLO network is a place where the architecture is heavily modified to suit the specific problem characteristics. The backbone of the YOLO network is responsible for extracting features from the input image through convolutional layers. A popular method is to improve the Backbone structure by replacing it with a popular CNN architecture. Examples of CNN architectures that are commonly used as replacements in Backbone networks include GoogleNet, Vgg-19, AlexNet, SqueezeNet, ResNet-18, Inception-v3, ResNet-50, Vgg-16, ResNet-101, DenseNet-201 and Inception-ResNet-v2 [18].

In this paper, we replace the YOLOv10 Backbone with the MobileNet V3 architecture. There are reasons to choose the MobileNet V3 architecture, that is, MobileNet V3 is capable of producing high accuracy as well as being lightweight and can run on low-performance computing devices (personal computers or laptops). MobileNetV3 [19] was proposed by a group of scientists from Google Corporation. In applications, MobileNetV3 has been demonstrated to be highly efficient in image classification, target detection, and other tasks. The MobileNetV3s architecture offers several advantages over the default Backbone used in YOLOv10. First, MobileNetV3 is specifically designed to be lightweight, suitable for resource-constrained environments. The parameters of the MobileNetV3 model are inherited from MobileNetV1 [20] and MobileNetV2 [21], in particular, these parameters are derived using the Network Search Architecture (NAS). The MobileNetV3 network also uses a Squeeze-and-Excitation (SE) channel attention mechanism [Howard et al. 2022]. This feature is especially important in

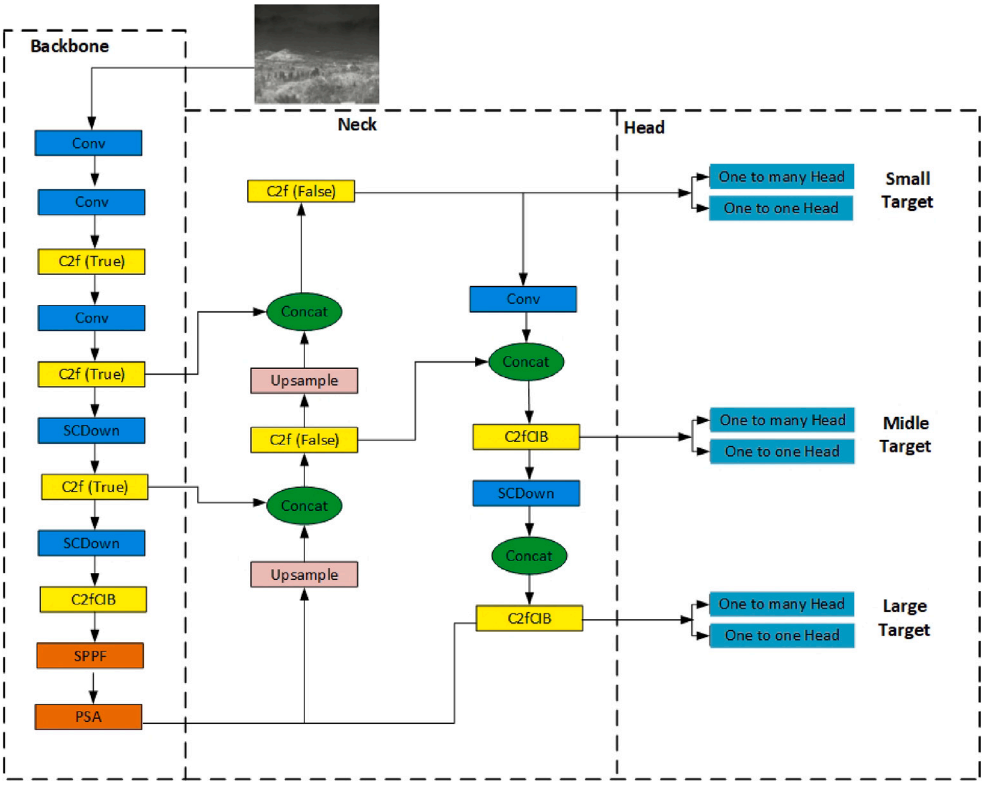

Fig. 1. YOLOv10 network structure diagram.

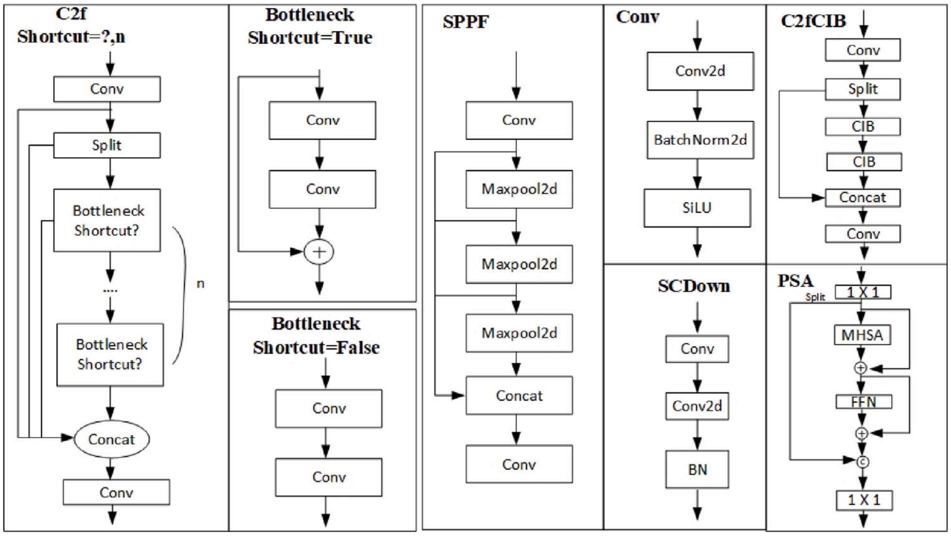

Fig. 2. YOLOv10 network detail structure diagram.

applications such as attribute detection, which may need to be deployed on edge devices or devices operating on low-power platforms.

The MobileNetV3 architecture achieves a balance between computational efficiency and model accuracy. By improving the Backbone of YOLOv10 by integrating MobileNetV3 into YOLOv10, we can leverage these advances to improve the accuracy of small UAV target detection on infrared images while maintaining real-time performance. The architecture of the MobilenetV3 model is depicted in Fig. 4. First, the input image is subjected to a 1 × 1 convolution to increase the number of channels. Next, it undergoes a deep convolution in the high-dimensional space and is optimized using the SE attention mechanism. Then, the number of channels is reduced by a 1 × 1 convolution (linear activation function). Residual connections are used when the step size is 1 and the input and output sizes are equal. The feature map is downsampled when the step size is 2 (downsampling stage).

2.2.2. Improvement of the head network

In the Head section, we added the small target detection module x-small target and removed the large-head module. In the head section of the original YOLOv10 model, there are 3 target detection modules: small target, medium target, and large target [17].

These three detection modules correspond to the connections from P3, P4, and P5. In the detection pipeline, the small target detection module connected to P3 undergoes downsampling by a factor of 8,

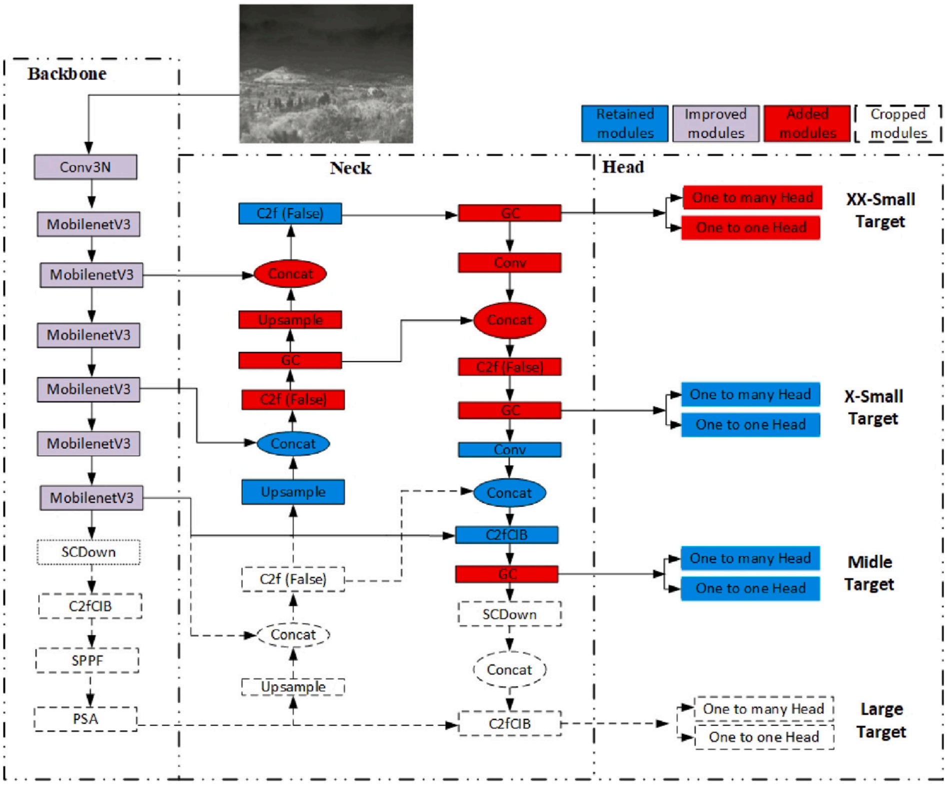

Fig. 3. Improved YOLOv10 (LW-UAV-YOLOv10) network structure diagram.

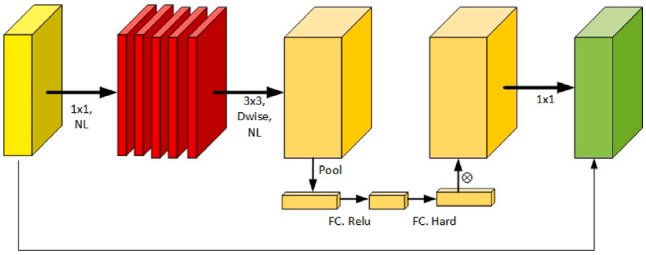

Fig. 4. MobilenetV3 network structure [22].

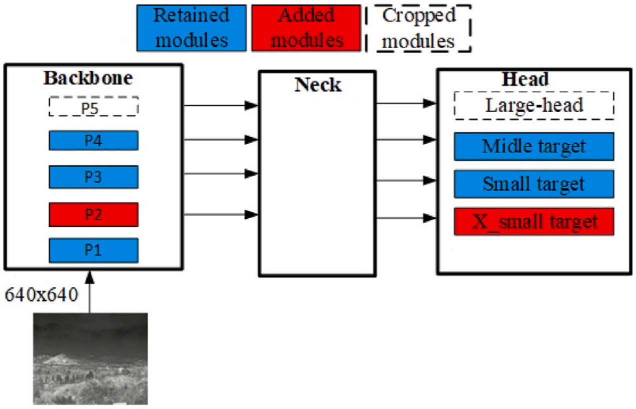

Fig. 5. Improvement scheme at the head.

reducing the original image size from 640 × 640 to 80 × 80. The medium target connected to P4 is downsampled by a factor of 16, resulting in a size of 40 × 40. For larger targets connected to P5, downsampling is performed by a factor of 32, resulting in an image size of 20 × 20. It is evident that when the target size is small (less than 32 pixels), the feature maps for the Large-head component may only capture a single point or no points at all. To address this, we propose removing the Large-head component along with the associated feature extraction and feature fusion layers from the YOLO network structure, as illustrated in Fig. 3. In the proposed model, only the head modules for medium and small targets are retained. To improve the detection of even smaller targets, we propose adding an x-small target module to the head along with the associated feature extraction layers. This x-small target module will perform downsampling with a factor of 4, resulting in an image size of 160 × 160 for targets smaller than 32 pixels. The additional head section is designed to sample 5 points, thereby enhancing the network's ability to effectively detect small targets. The updated head section is illustrated in Fig. 5.

2.2.3. Infrared small UAV target detection using LW-UAV-YOLOv10 model

The infrared small UAV target detection process based on the LW-UAV-YOLOv10 model, including dataset construction, model training, and UAV detection, is depicted in Fig. 6. First, the data is collected, preprocesses, labeled, and then divided into training and validation sets. Next, the training parameters are configured, and the convolutional neural network is initialized. The loss function is determined by the sum of squared errors in regression and the binary cross-entropy error in classification. The model weights are updated iteratively to ensure the loss function converges, ultimately resulting in the LW-UAV-YOLOv10 model for UAV target detection.

3. Experimental preparation and results

3.1. Dataset introduction

Currently, in popular target detection datasets such as MSCOCO, PASCAL VOC, and ImageNet, the number of small target samples is usually quite limited. The size difference between large and small

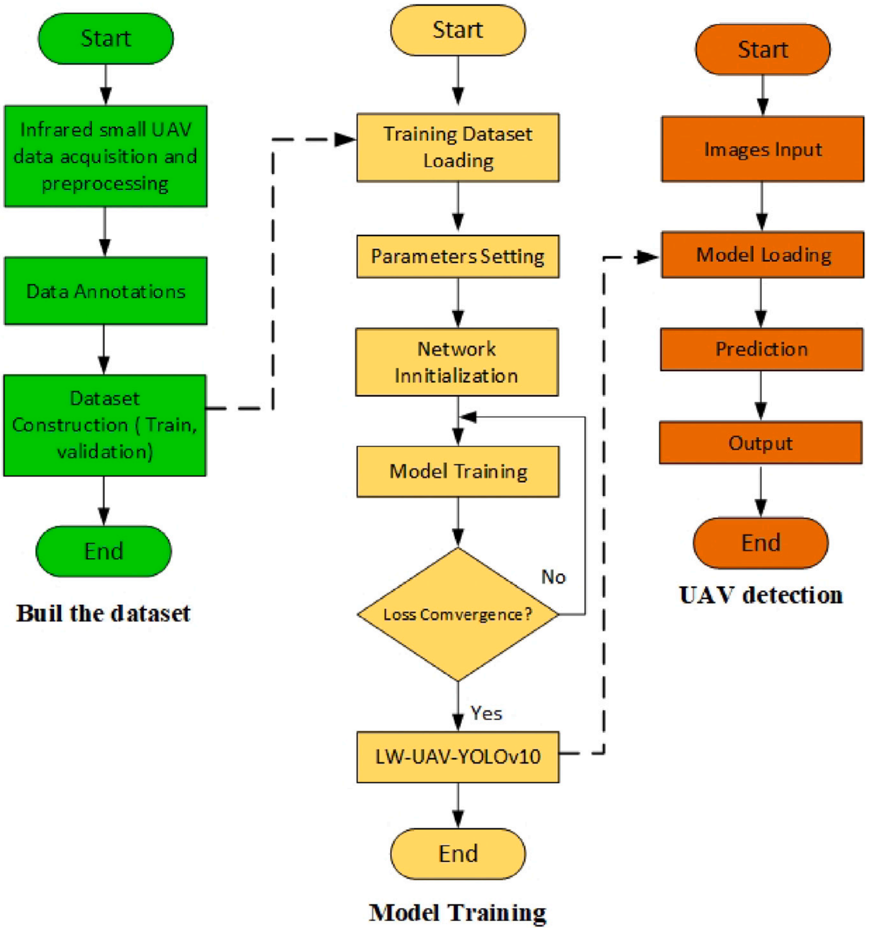

Fig. 6. The UAV detection process based on LW-UAV-YOLOv10.

targets is significant, and the absence of targets like UAVs makes it difficult to build deep learning models that are adapted to detecting small UAV targets. The data used by the proposed model is shared by the authors in the publication (https://github.com/Dang-zy/SIDD.git) [15]. The SIDD dataset contains 1,093 infrared images taken in mountainous conditions. This dataset is dedicated to detecting small UAV targets in infrared image data and is aimed at applied research related to the detection of weak and low-flying UAVs.

3.2. Implementation details

The hardware configuration for the experiments includes a GPU (RTX4070 Ti Super 16 GB), a CPU (i5-12600KF), 32 GB of RAM, PyTorch 1.13.1 as the deep learning framework, and Python version 3.10.13. The dataset was trained over 150 epochs using the Adam optimization method with an initial learning rate of 0.01. Given the presence of small targets in the sample images and the need to balance real-time performance with accuracy during detection, the input image size was normalized to 640 × 640. This size helps the model perform effectively on edge devices while preserving essential features from the input images. To ensure fair and objective comparison of the model performance, no pre-trained weights were used, and all training processes employed consistent hyperparameter settings. Stochastic Gradient Descent (SGD) Momentum is a parameter in SGD that accelerates convergence by retaining more past gradient information. The value of 0.937 is the momentum coefficient, ranging from 0 to 1, which influences the current update based on previous gradients. A higher momentum value (closer to 1) helps maintain a consistent direction and speeds up the learning process. Key parameter settings for the training process are detailed in Table 1.

Table 1
Important parameter setting table.

STT	Parameters	Setup
1	Epochs	150
2	Batch Size	16
3	Imgsize	640 × 640
4	Learning rate	0.01
5	Optimizer	Adam
6	Momentum	0.937 (SGD momentum)

3.3. Evaluation metric

To evaluate the performance of the proposed model in comparison with the newly introduced YOLO versions (YOLOv10), this paper employs statistical indicators for model comparison. We utilize Intersection over Union (IoU) to assess the accuracy of positive or negative sample detection. Additionally, Average Precision (AP) is computed across various IoU thresholds, which can be expressed mathematically as follows:

$$AP = \int_0^1 Pr_m(Re_m)d\,Re_m \quad (1)$$

Where Re_m and Pr_m denote the recall and precision for target class m and N denotes the number of targets. The average precision mean is calculated as follows:

$$mAP = \frac{1}{N}\sum_{k=1}^{N} AP_i \quad (2)$$

where N is the number of categories and AP is the average accuracy of each category. In our UAV detection task, $N = 1$. A high precision value

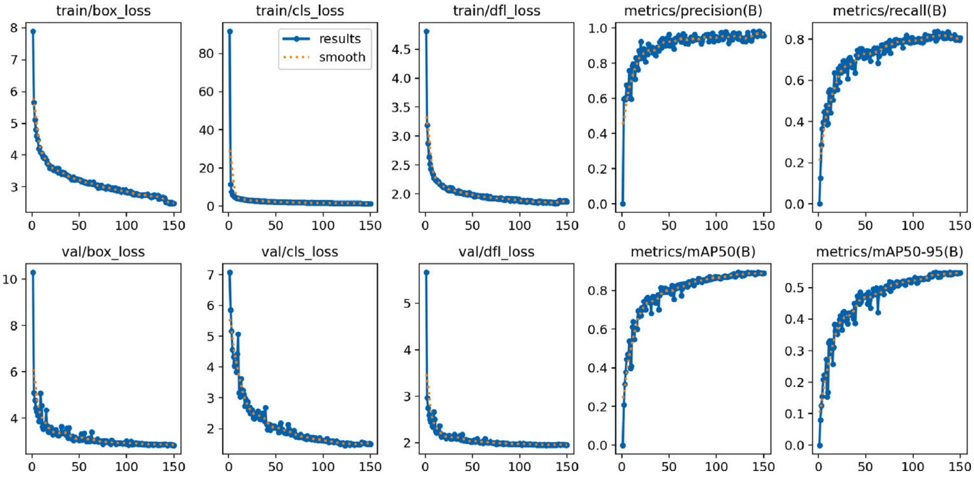

Fig. 7. Graph of result values for LW-UAV-YOLOv10 showing changes in key indicators across the training epochs.

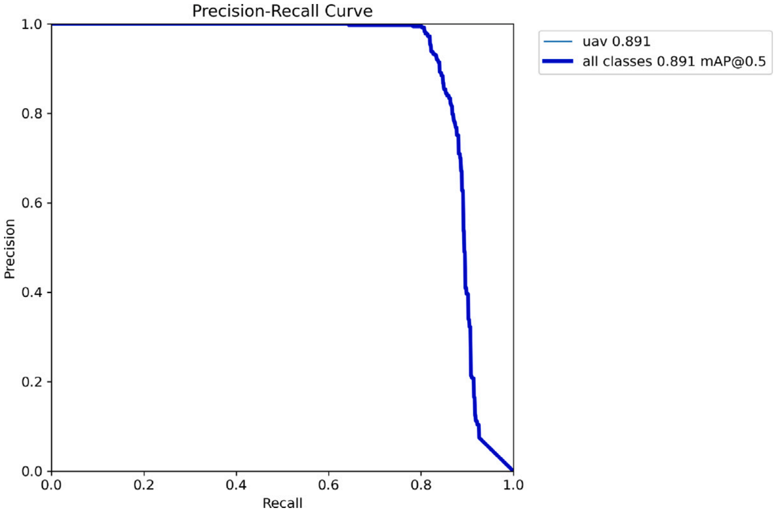

Fig. 8. Precision–Recall curve of the LW-UAV-YOLOv10 training process.

signifies that the model is effective at minimizing the misclassification of false positives as true positives. Recall measures the model's capability in identifying positive examples, with a high recall value indicating a lower incidence of false negatives. The precision (Pr) and recall (Re) metrics can be mathematically defined by the following formulas [23]:

$$Pr = \frac{TP}{TP + FP} \tag{3}$$

$$Re = \frac{TP}{TP + FN} \tag{4}$$

where TP, FN, and FP denote the number of true positives, false negatives, and false positives, respectively.

FPS: A higher FPS value indicates better real-time processing capabilities of the model and that the model is lighter when running the detection algorithm on the same hardware. FPS is calculated as 1/latency, where latency refers to the time the network takes to predict an image.

3.4. Main results

The results after running the model on the training and validation sets have demonstrated that the proposed model both meets the requirements of being lightweight and ensures accuracy when detecting small UAV targets in mountainous conditions, with superior performance compared to the versions of the YOLOv10 model. Fig. 7 and Fig. 8 depict the results of the training process of the proposed model. Describes the convergence process and the process of achieving UAV target detection accuracy of the proposed model. After evaluation, our LW-UAV –YOLOv10 model achieved an accuracy of up to 98.3%, a recall of 81%, and a mAP0.5 score of 89.1%. Fig. 7 shows that, with 150 epochs, the model performance evaluation indexes such as Precision, recall, mAP0.5 as well as the values of the loss functions also reach stable values.

3.4.1. Ablation experiment

Table 2 shows the results to demonstrate the impact of the improvement method on the YOLOv10 model and prove the effectiveness and necessity of the proposed model. The ablation experiments were conducted on the same dataset using the above-mentioned improvement modules such as removing the P5 head, adding the P2 detection head, or replacing the backbone with MobinetV3. "x" indicates the module that has been added to the YOLOv10 model and "–" indicates the module that has not been added. When each module was applied separately to the YOLOv8 model, the ablation experiments were conducted to determine the detection accuracy and detection speed of that module, as shown in Table 2. The results show that each improvement affects the detection performance and detection speed of that module to a certain extent. When using the YOLOv10s model denoted as M1, the mAP@0.5 score of this module is 87.4% and the FPS value is 152. After removing the large detector head, which is then denoted as M2,

Table 2
Results of the various ablation experiments.

Components	M1	M2	M3	LW-UAV-YOLOv10 model
- P5	–	x	x	x
+ P2	–	–	x	x
+ MobinetV3	–	–	–	x
Precision	0.93	0.949	0.965	0.983
Recall	0.8	0.773	0.819	0.81
mAP	0.874	0.868	0.9	0.891
FPS (1/s)	152	183	132	164
Parameters	8,035,734	7,186,196	7,457,174	3,840,010
Model size (Mb)	16.5	14.8	15.5	8

Table 3
Results of the proposed models LW-UAV-YOLOv10 and some other version YOLO models.

Model	mAP@0.5	mAP 0.5 – 0.95	Precision	Recall	Parameters	Layers	FPS (f/s)	Model Size (MB)	Model training time (h)
YOLOv8n	0.857	0.51	0.94	0.791	3,005,843	168	157	6.2	0.731
YOLOv9t	0.812	0.471	0.868	0.789	1,970,979	486	97	4.6	1.566
YOLOv11n	0.836	0.494	0.93	0.768	2,582,347	238	152	5.5	0.737
YOLOv10s	0.874	0.535	0.93	0.8	8,035,734	293	152	16.5	1.205
YOLOv10m	0.88	0.544	0.95	0.797	16,451,542	369	135	33.5	1.329
YOLOv10n	0.83	0.5	0.89	0.754	2,694,806	285	152	5.8	0.723
YOLOv10b	0.864	0.54	0.946	0.766	20,412,694	383	139	41.5	1.583
YOLOv10l	0.864	0.545	0.945	0.797	25,717,910	461	106	52.2	1.866
YOLOv10x	0.876	0.549	0.975	0.777	31,586,006	503	88	64.1	21.726
LW-UAV-YOLOv10	0.891	0.549	0.983	0.81	3,840,010	307	164	8	1.407

the detection accuracy is not significantly affected, but the detection frame rate per second is improved. Combining the design of removing the large detector head and adding a very small target detector head (the model denoted as M3), the detection accuracy increases, while the detection speed is slightly reduced. Replacing the backbone of M3 with the MobinetV3 module, we get the proposed model LW-UAV-YOLOv10. From Table 2, it can be seen that the proposed model LW-UAV-YOLOv10 exhibits the characteristics of a lightweight model when the number of parameters and model size of the model are only about 50%compared to the models M1, M2, M3. Although the proposed model has a slightly lower accuracy than the M3 model through the mAP@0.5 index of 0.9%, and the detection speed of the proposed model is worse than M2 through the FPS index. Finally, the overall performance of the proposed model has been improved in terms of infrared small UAV target detection efficiency, which is suitable for the hardware requirements and the ability to detect targets in real time. The improvements of the proposed model balance the M2 and M3 models in terms of detection speed and detection accuracy, meeting the requirements of light weight and high accuracy.

3.4.2. Comparative experiments on detection of different models

From Table 3, we can see that our proposed model has the highest detection efficiency compared to all versions of YOLOv10. For the evaluation of the lightweight characteristics and Real-time detection capabilities of the model, we base on the Parameters, Layers; FPS; Model Size. The lightweight the model and the better the real-time detection capabilities, the smaller the values of Parameters, Layers; Model Size, and the larger the FPS. From Table 3, it is clear that the proposed model LW-UAV-YOLOv10 has a lighter model size than the versions YOLOv10s, m, l, b, x. More specifically, the proposed model has a lighter model size than the YOLOv10l model by more than 6.5 times, YOLOv10b by 5 times, YOLOv10m by 4 times, YOLOv10s by 2 times, and YOLOv10x by nearly 8 times. However, the proposed model has a heavier model size than the YOLOv10n version, but in return, the accuracy (mAP0.5) of the proposed model LW-UAV-YOLOv10 is much better (6.1%) than YOLOv10n. In addition, the FPS inference speed index of the proposed model LW-UAV-YOLOv10 reaches the highest value (FPS = 164) and the number of Parameters is much less than the YOLOv10s, m, l, b, x versions and only more than the YOLOv10n version. From Table 3, the proposed model is compared with YOLOv8n, YOLOv9t and YOLOv11n [24], which are the lightest versions of YOLOv8, YOLOv9 and YOLOv11, respectively. The proposed model has superior accuracy (mAP0.5) and detection rate (FPS) compared with YOLOv8n, YOLOv9t and YOLOv11n models. However, the model size and number of parameters of the proposed model are larger than YOLOv8n, YOLOv9t and YOLOv11n. These results demonstrate the practicality of the proposed model and its ability to be deployed on edge devices for the problem of detecting small UAV targets on infrared data. The results in Table 3 show that the proposed model balances the lightweight feature well; the ability to detect targets in real time and the accuracy compared to YOLOv10 versions.

Fig. 9 illustrates the UAV target detection results on the research dataset, comparing the proposed model with the YOLOv11n, YOLOv10n, and YOLOv9t models. Specifically, in Fig. 9a, the YOLOv9t model misses 4 targets (marked by red crosses) and mistakenly detects 1 target (marked by a red circle). Meanwhile, the YOLOv10n model (Fig. 9b) misses 1 target and mistakenly detects 2 targets. Both the YOLOv11n (Fig. 9c) and the proposed LW-UAV-YOLOv10 models (Fig. 9d) miss only 2 targets and do not mistakenly detect any targets.

4. Conclusions

Our research advances AI solutions for real-time small target detection, especially for small UAVs using infrared data in mountainous environments. The proposed LW-UAV-YOLOv10 model addresses two key challenges in deep learning: achieving lightweight models without compromising accuracy. Experimental results demonstrate that LW-UAV-YOLOv10 achieves higher accuracy, with mAP values increasing from 1.1% to 6.1%, and delivers the highest FPS detection rate among all YOLOv10 versions on the same dataset. In addition, its model size is 2 to 8 times smaller than YOLOv10, m, l, b, and x versions. The proposed model LW-UAV-YOLOv10 has a heavier model size than the YOLOv10n version, but its accuracy outperforms YOLOv10n by 6.1%.

Furthermore, the proposed model outperforms the lightest versions of YOLOv8, YOLOv9, and YOLOv11 in terms of mAP accuracy and FPS detection rate, despite having a slightly larger model size. Moving forward, research should focus on further improving the model accuracy for small target detection, especially in terms of inference speed and

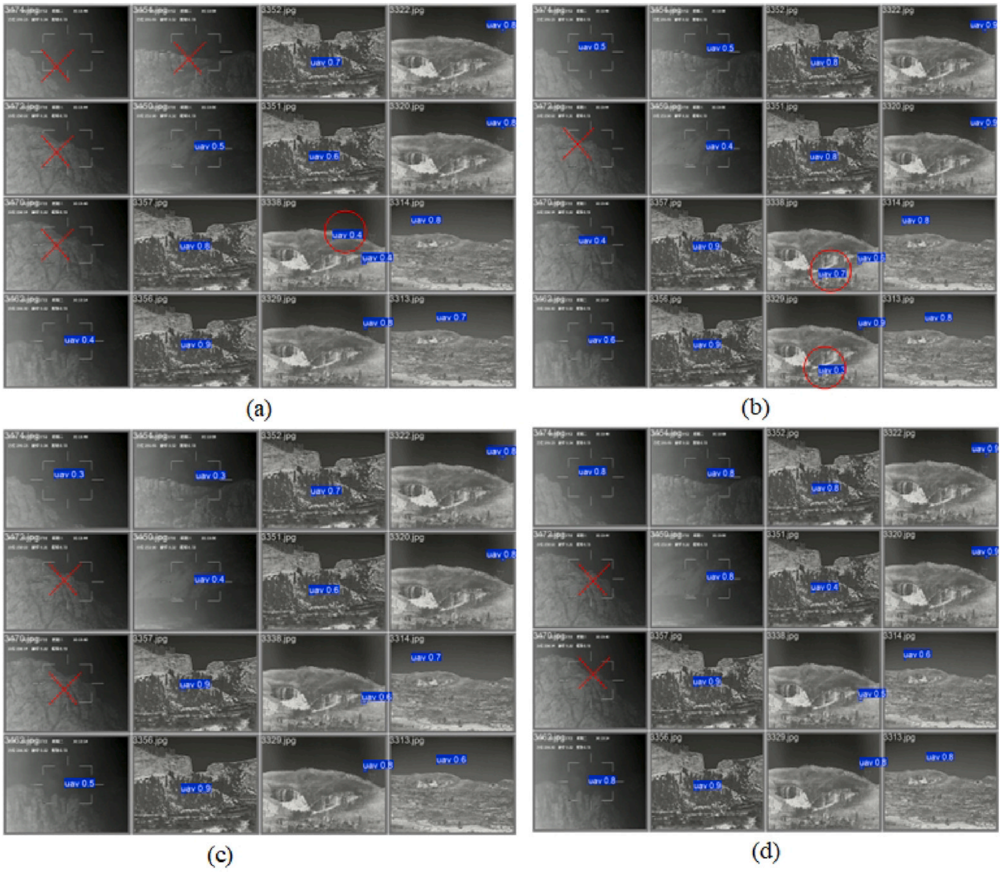

Fig. 9. Example of YOLOv9t model (a);YOLOv10n model (b);YOLOv11n model (c) and LW-UAV-YOLOv10 model (d).

model efficiency for edge devices. Future work could explore incorporating attention modules into Backbone, Neck, or data processing pipelines.

CRediT authorship contribution statement

Phat T. Nguyen: Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Formal analysis, Conceptualization. **Giang L. Nguyen:** Writing – review & editing, Methodology, Investigation. **Duy D. Bui:** Software, Formal analysis, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] J. Chenchen, R. Huazhon, Y. Xin, Z. Jinshun, Z. Hui, N. Yang, S. Min, R. Xiang, H. Hongtao, Object detection from UAV thermal infrared images and videos using YOLO models, Int. J. Appl. Earth Obs. Geoinf. 112 (2022) http://dx.doi.org/10.1016/j.jag.2022.102912.

[2] D. Raval, E. Hunter, S. Hudson, A. Damini, B. Balaji, Convolutional neural networks for classification of drones using radars, Drones 5 (4) (2021) 149, http://dx.doi.org/10.3390/drones5040149.

[3] I. Nemer, T. Sheltami, I. Ahmad, A.U.-H. Yasar, M.A.R. Abdeen, RF-based UAV detection and identification using hierarchical learning approach, Sensors 21 (6) (2021) 1947, http://dx.doi.org/10.3390/s21061947.

[4] M.Z. Anwar, Z. Kaleem, A. Jamalipour, Machine learning inspired sound-based amateur drone detection for public safety applications, IEEE Trans. Veh. Technol. 68 (2019) 2526–2534, http://dx.doi.org/10.1109/TVT.2019.2893615.

[5] U. Seidaliyeva, D. Akhmetov, L. Ilipbayeva, E.T. Matson, Real-time and accurate drone detection in a video with a static background, Sensors 20 (14) (2020) 3856, http://dx.doi.org/10.3390/s20143856.

[6] B. Taha, A. Shoufan, Machine learning-based drone detection and classification: State-of-the-art in research, IEEE Access 7 (2019) 138669–138682, http://dx.doi.org/10.1109/ACCESS.2019.2942944.

[7] J. Yousaf, H. Zia, M. Alhalabi, M. Yaghi, T. Basmaji, E.A. Shehhi, A. Gad, M. Alkhedher, M. Ghazal, Drone and controller detection and localization: Trends and challenges, Appl. Sci. 12 (24) (2022) 12612, http://dx.doi.org/10.3390/app122412612.

[8] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: IEEE Conf. on Comput. Vis. and Pattern Recognit., CVPR, Las Vegas, NV, USA, 2016, pp. 779–788, http://dx.doi.org/10.1109/CVPR.2016.91.

[9] X. Ren, M. Sun, X. Zhang, L. Liu, H. Zhou, X. Ren, An improved mask-RCNN algorithm for UAV tir video stream target detection, Int. J. Appl. Earth Obs. Geoinf. 106 (2022) 102660, http://dx.doi.org/10.1016/j.jag.2021.102660.

[10] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A. Berg, SSD: Single shot MultiBox detector, in: Proc. of the Eur. Conf. on Comput. Vis., Vol. 9905, ECCV, 2016, pp. 21–37, http://dx.doi.org/10.1007/978-3-319-46448-0_2.

[11] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, G. Ding, YOLOv10: Real-time end-to-end object detection, 2024, http://dx.doi.org/10.48550/arXiv.2405.14458.

[12] M. Hussain, R. Khanam, In-depth review of YOLOv1 to YOLOv10 variants for enhanced photovoltaic defect detection, Solar 4 (2024) 351–386, http://dx.doi.org/10.3390/solar4030016.

[13] L. Jiao, F. Zhang, F. Liu, S. Yang, L. Li, Z. Feng, R. Qu, A survey of deep learning-based object detection, IEEE Access PP (1) (2019) http://dx.doi.org/10.1109/ACCESS.2019.2939201.

[14] V. Dewangan, A. Saxena, R. Thakur, S. Tripathi, Application of image processing techniques for UAV detection using deep learning and distance-wise analysis, Drones 7 (3) (2023) 174, http://dx.doi.org/10.3390/drones7030174.

[15] S. Yuan, B. Sun, Z. Zuo, H. Huang, P. Wu, C. Li, Z. Dang, Z. Zhao, IRSDD-YOLOv5: Focusing on the infrared detection of small drones, Drones 7 (393) (2023) http://dx.doi.org/10.3390/drones7060393.

[16] X. Zhai, Z. Huang, T. Li, H. Liu, S. Wang, YOLO-drone: An optimized YOLOv8 network for tiny UAV object detection, Electronics 12 (3664) (2023) http://dx.doi.org/10.3390/electronics12173664.

[17] A. Ahmed, A. Manaf, Pediatric wrist fracture detection in X-rays via YOLOv10 algorithm and dual label assignment system, 2024, http://dx.doi.org/10.48550/arXiv.2407.15689.

[18] O. Elharrouss, Y. Akbari, N. Almaadeed, S. Al-ma'adeed, Backbones-review: Feature extraction networks for deep learning and deep reinforcement learning approaches, 2022, http://dx.doi.org/10.48550/arXiv.2206.08016.

[19] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q.V. Le, H. Adam, Searching for MobileNetV3, in: Proc. IEEE/CVF Int. Conf. on Comput. Vis., ICCV, October, 2019, http://dx.doi.org/10.1109/ICCV.2019.00140.

[20] A.G. Howard, Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017, arXiv preprint arXiv:1704.04861.

[21] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, MobileNetV2: Inverted residuals and linear bottlenecks, in: IEEE/CVF Conf. on Comput. Vis. and Pattern Recognit., Salt Lake City, UT, USA, 2018, pp. 4510–4520, http://dx.doi.org/10.1109/CVPR.2018.00474.

[22] Y. Wu, Y. Sun, S. Zhang, X. Liu, K. Zhou, J. Hou, A size-grading method of antler mushrooms using YOLOv5 and pspnet, Agronomy 12 (2601) (2022) http://dx.doi.org/10.3390/agronomy12112601.

[23] J. Chen, Y. Fu, Y. Guo, X. Yue, X. Zhang, F. Hao, An improved deep learning approach for detection of maize tassels using UAV-based RGB images, Int. J. Appl. Earth Obs. Geoinf. 130 (2024) 103922, http://dx.doi.org/10.1016/j.jag.2024.103922.

[24] L. He, Y. Zhou, L. Liu, J. Ma, Research and application of YOLOv11-based object segmentation in intelligent recognition at construction sites, Buildings 14 (12) (2024) 3777, http://dx.doi.org/10.3390/buildings14123777.