International Journal of Cloud Computing and Database Management

E-ISSN: 2707-5915
P-ISSN: 2707-5907
IJCCDM 2025; 6(2): 01-06
Journal's Website
Received: 15-05-2025
Accepted: 16-06-2025

Muhammad D Hassan
Quality Assurance Division, College of Health and Medical Technology, Northern Technical University, Iraq

Corresponding Author:
Muhammad D Hassan
Quality Assurance Division, College of Health and Medical Technology, Northern Technical University, Iraq

UAV image recognition YOLOv8 detection network improvement

Muhammad D Hassan

DOI: https://www.doi.org/10.33545/27075907.2025.v6.i2a.91

Abstract
An important skill in both civilian and military realms is the ability for UAVs to identify numerous targets. Even if deep learning approaches provide a better answer to the problem, there are still major obstacles in this area, such as variations in target size, form changes, occlusion, and illumination conditions as seen from the drone's viewpoint. The research presents a model for aerial picture identification that is very resilient and performs exceptionally well. These problems are the foundation of the model. The goal of introducing the Bi-PAN-FPN concept is to enhance the neck area. This is done because there is a common issue with aerial pictures where tiny targets might be missed or misdetected. A more sophisticated and comprehensive feature fusion technique may be developed with the goal of fully lowering parameter costs by taking into account and reapplying multiscale features in their entirety. One part of the C2f module is replaced with the GhostblockV2 structure in the benchmark model's backbone. In addition to drastically cutting down on the total number of model parameters, this also reduces the likelihood of data loss as features are sent over long distances. A globally famous dataset is used to analyze and examine the method's performance. To further confirm the efficacy and practicability of the suggested model, extensive ablation studies, comparative analyses, interpretability evaluations, and experiments with bespoke datasets are carried out.. This introduces a fresh approach to multitarget identification by unmanned aerial vehicles (UAVs) utilizing deep learning.

Keywords: Deep learning, multitarget detection, UAV

Introduction

Currently, a prevalent research avenue is integrating deep learning detection algorithms with aerial imagery obtained from unmanned aerial vehicles (UAVs). The use of UAV surveillance offers several benefits, including an extensive monitoring area, high efficiency, and reduced costs. These advantages stem from the attributes of mobility and flexibility, together with the capacity to overcome constraints imposed by natural factors such as the environment and location. Conversely, deep learning methodologies have often posed challenges for implementation in actual UAV object identification applications. This arises from the intricacy of the issue. This mostly stems from two distinct reasons: Aerial photography captured by unmanned aerial vehicles (UAVs) has distinct qualities that differentiate it from terrestrial photography. These attributes include expansive vistas, diminutive targets, multiscale dimensions, complex backdrops, and overlapping occlusions [1, 2, 3]. Conversely, ground-level photography lacks these attributes. Conversely, such detection tasks often need the integration of inference methodologies into embedded systems, imposing rigorous demands for both precision and real-time speed. Identifying certain items reliably is tough. The identification of specific products, however, poses significant challenges. In the near future, companies enhancing their intelligence will likely confront a new phase. During this phase, we will shift from the quick development of the first stage to efficient functioning. Throughout this process, the underutilization of deep models will inevitably evolve into a significant tool for these firms. Consequently, to mitigate the limitations posed by deep learning in the use of UAV aerial photos, it is very beneficial to create a model that considers both detection accuracy and model weight. This is attributable to deep learning being a bottleneck, which is the cause of the limitation. This work primarily addresses the previously described issues and tries to improve the flexibility and efficacy of a multitarget detection model for aerial photographs captured by unmanned aerial vehicles (UAVs) [4, 5].

Relevant Job

Despite its widespread use, target identification in the context of unmanned aerial vehicles (UAVs) has several challenges, significantly impacting both research and practical implementation. Efficient methodologies for identifying photographs recorded by unmanned aerial vehicles (UAVs) have emerged due to the continuous advancement of target recognition technology. The reference proposed a method for object detection in drone imagery known as UFPMP-Net. This approach involves the construction of a unified foreground packing (UFP) module to aggregate the subregions provided by the coarse detector, therefore mitigating the background. This was executed with regard to the unique characteristics. The SSD detector was the basis for the methodology, which included the enhancement of the feature fusion module via the incorporation of an extra branch of the deconvolution module and the use of average pooling to generate a distinctive feature pyramid. Moreover, the methodology integrates the spatial relationships of objects into the detection process, hence enhancing detection accuracy [6, 7].

Direct implementation on low-power image processors, such as edge devices, is challenging due to the substantial memory overhead and processing requirements of most systems. Although advanced target recognition methods have significantly enhanced UAV multitarget detection capabilities, these technologies are not readily accessible. The advancement of YOLO series detection networks has addressed this problem. This series of models has undergone eight official modifications, along with many branch variations. The YOLO model structure has three unique components: the backbone, the neck, and the head. The neck amalgamates features derived from the backbone, enhancing feature diversity and augmenting the detection network's performance; the head enables accurate predictions by utilizing the previously refined high-quality features; and the backbone functions as a network for image feature extraction. Backbone is a feature extraction network used for obtaining feature information from photos. Nearly every version of YOLO models has included similar modifications and enhancements in these three frameworks. The updates and upgrades have been included. The YOLO series models have gained substantial use across several sectors, including medicine, remote sensing, transportation, and manufacturing, due to their exceptional detection accuracy and speed. Researchers in the academic community are presently exploring the use of YOLO and other lightweight models for aerial picture recognition using unmanned aerial vehicles (UAVs). The model's parameters and calculations were reduced by the use of a proved adaptive model compression approach. This activity was undertaken to address the conflicting issue of limited resources in the UAV deployment platform amongst the substantial demands for real-time reasoning. The integration of a "transfer factor" into the model pruning process enables the approach to ascertain the need of pruning a specific channel based on the scale factor [8, 9, 10]. Moreover, using the transfer factor may effectively mitigate the impact of pruning on the subsequent structure.

The channels comprising the convolutional layer. The YOLOv3-SPP3 model is used to assess the effectiveness of the approach. This reference emphasizes. The solid-state drive (SSD) technology was used in the creation of the UAV aerial image recognition model referred to as UAV-Net. The alterations to the backbone and neck, together with the adoption of the automatic pruning technique, have decreased the model's size to a mere 0.4 megabytes, hence enhancing its potential for general applicability. The reference introduced an innovative method for the large-scale detection of marine targets for unmanned aerial vehicles (UAVs) with YOLOv5. This approach was created to achieve a compromise, hence improving the detection accuracy of tiny and obstructed objects. A segment of the convolutional framework is substituted with linear transformation, characterized by its simplicity and rapid computation. Consequently, the quantity of parameters included in the model is decreased. The experimental findings indicate that this technique has some significant advantages over other advanced models regarding detection accuracy, recall rate, average precision, and parameter count. A method for assessing the integrity of an insulator was proposed by Reference. This method offers the technique of integrating deep learning with mobile edge computing. The YOLOv4 detector is the core of this system, while the lightweight MobileNetv3 network replaces the original backbone. This leads to a substantial decrease in the network's parameters. The refinement of the activation function in MobileNetv3 and the optimization of the model's loss function are two elements that have contributed to the overall increase in model verification quality. The use of particle swarm optimization enables the approach to efficiently segment the deep neural network while minimizing time and processing resources [11, 12, 13].

Aerial Image Detection with Enhanced Results:

Three components are used to develop a model for the accurate and fast identification of aerial photos by unmanned aerial vehicles (UAVs): To address the widespread issue of misdetection and missed detection of tiny objects in aerial photos. An upsampling method has been developed to emphasize minute target characteristics. By meticulously evaluating and reapplying

1. Enhancement of the Cervical Region: Consequently, the following modifications were implemented to the neck structure for the UAV aerial photography dataset:

Initially, we redirected our focus to large-scale feature maps. To improve the detection performance of small targets, an upsampling process was included into the FPN and coupled with the attributes of the B2 layer in the backbone matrix. Analogous to the upsampling technique previously used in FPN, the C2f module was implemented to augment the quality of feature extraction subsequent to feature fusion. The C2f module enhances the previous C3 module, particularly by using the advantages of the ELAN structure in YOLOv7, resulting in improved gradient information. The upgrade centered on the C2f module. This module diminishes a standard convolutional layer, maximizing the bottleneck module to enhance the gradient branch. This facilitates the gathering of more nuanced gradient flow information while preserving a lightweight framework.

The second thing that we did was present the concept of Bi-PAN-FPN. In order to achieve a greater level of detection accuracy, the fundamental concept behind this structure is to enhance the possibilities and durations of multiscale feature fusion. The following are the stages involved in its implementation: There is no further processing that is carried out for feature maps that have just one input route. Generally speaking, these kinds of features make a rather little contribution to feature engineering. An extra route is created from the features in the backbone, and the features in the PAN are fused, in the case of feature maps that have two input pathways. This occurs only if the size of the feature maps is the same. Such a processing procedure does not result in any extra costs associated with the parameters. Finally, in order to optimize blending, consider each bidirectional route (top-bottom and bottom-top) as a unit and reuse this unit numerous times if you want to get the desired results. The only additional routes of B3-N3 and B4-N4 were added here, and just one unit was employed. This was done in consideration of the fact that the model was very lightweight.

2. Enhancement of the spine: YOLOv8 used both the conventional convolution module and the C2f module to effectively extract high-quality features and downsample images. Conversely, the integration of an upsampling technique in the neck section and the implementation of Bi-PAN-FPN resulted in an increase in both the number of parameters and the model's complexity. This post will introduce the notion of Ghostblock inside the Backbone framework and using this structure to substitute certain C2f components. Ghostblock is an optimization strategy for the lightweight convolutional neural network referred to as GhostNet. The predominant advantages may be classified into two groups. Ghostblock, on one hand, is an extension of the fundamental ideas of GhostNet. The first feature map is constructed by a preliminary classic convolution technique. Subsequently, it integrates many linear transformation techniques to enhance the information inside the feature map. This ensures excellent feature extraction while preserving a varied array of characteristics. A decoupled fully connected (DFC) attention mechanism is proposed as an option for consideration. This method, due to its specificity, bypasses the limitations imposed by traditional attention algorithms regarding computational expense, doing this by gathering feature information across vast distances. The advantages of the structure lead to an enhanced grade of feature engineering across the whole framework. In the context of GhostNet, the convolution method used is termed the inexpensive operation.

4. Outcomes

4.1 Preprocessing the dataset: The experimental verification object employed in this research is VisDrone, a reputable dataset in the area of worldwide drone vision. Drone use is now common in a number of industries, including agricultural, aerial photography, and personal surveillance. The complete effect of shooting angle, light, backdrop, and other elements makes it more challenging to intelligently comprehend UAV visual data than traditional computer vision jobs. The dataset includes 2.6 million hand annotations of bounding boxes, and each picture may include hundreds of items to be spotted, in contrast to traditional detection datasets. In order to enhance the use of data in a variety of tasks, VisDrone also offers several crucial features including scene visibility, object class, and occlusion.

Each detection sample comes from the test set. The constructed model exhibited exceptional detection skills across many environments, and its robustness aligned with strict technical specifications. Nevertheless, the model consistently encountered both missed detections and erroneous detections during identification tasks involving objects that were diminutive and closely clustered. Numerous targets were overlooked during the detection phase, mostly due to the considerable resemblance between the truck and bus categories when seen from an elevated perspective. If the categories for automobiles and pedestrians are insufficiently sized, the model may categorize them as background, leading to missed detections.

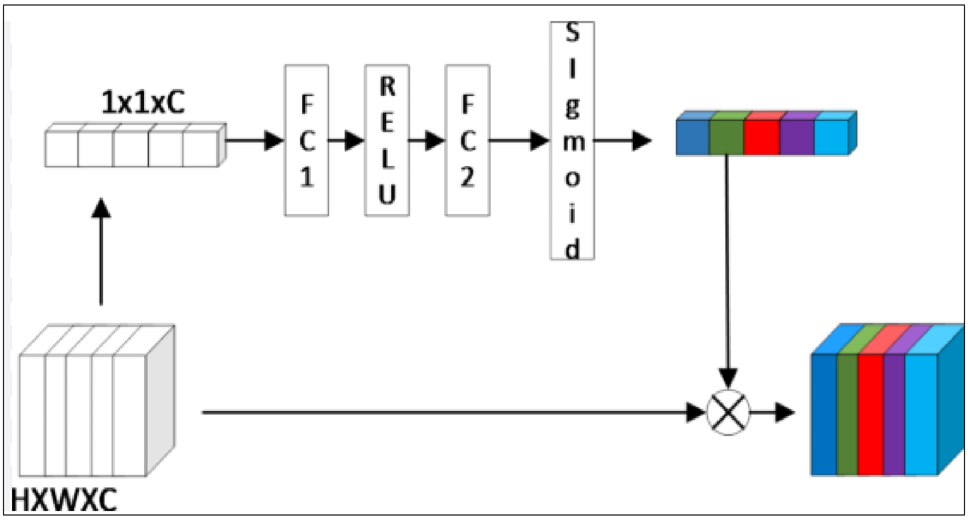

Fig 1: target of UaVs

4.2 Comparison of Deep Learning Model Performance

In the domain of target identification, deep learning methodologies are principally categorized into one-stage and two-stage methods. This categorization is predicated on the diverse anchor creation procedures used. The real-time analysis of aerial photos aligns more closely with the conditions faced in actual engineering. Consequently, using the one-stage target detection method is more pragmatic, since it exhibits less reliance on hardware while accounting for accuracy. The YOLO series and the solid-state drive (SSD) were selected for the comparative test in the experiment. Both goods exhibit more sophistication and prevalence within the market. Specifically, it included YOLOv5-s, YOLOv4-s, YOLOv7-tiny, MobilNetv2-SSD, and YOLOX-sall of which have been widely used in various embedded applications and documented in multiple publications. The YOLOv5-m model was used as the benchmark for the investigation. This choice was taken to illustrate the superiority of the paradigm proposed in this research. Table 1 displays the findings of the trials that were conducted to compare the two.

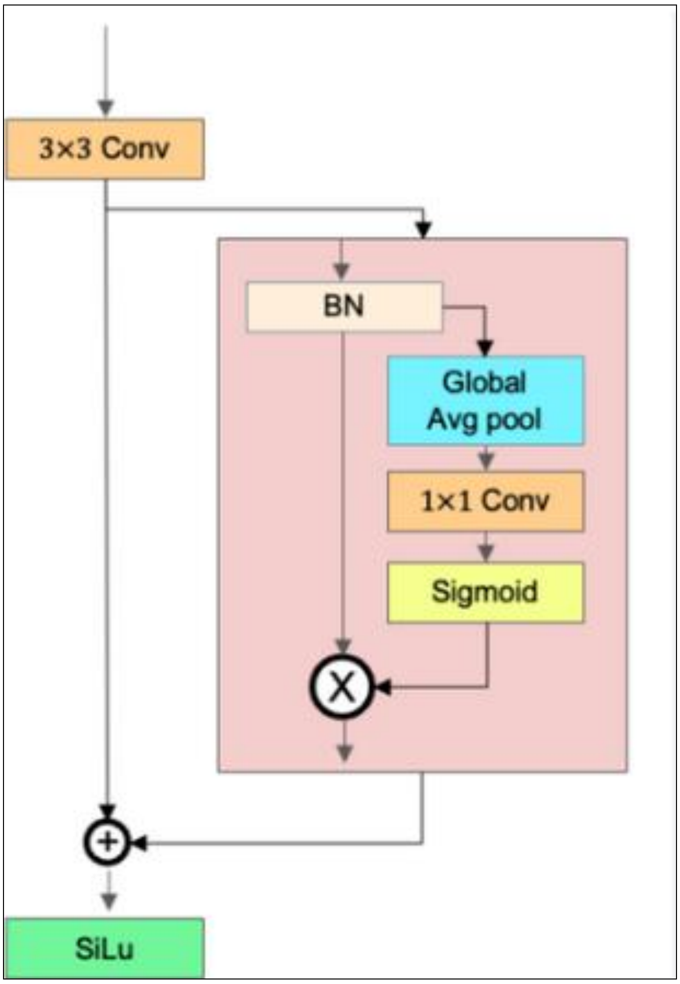

Fig 2: Target detection by using dataset

Table 1: The comparison of the findings of the experiment.

Data set	**Indicators**	1	2	3	4	5	6	7
Val	P	0.502	0.437	0.502	0.611	0.790	0.407	0.615
	R	0.17	0.418	0.425	0.409	0.264	0.429	0.507
	MAP	0.124	0.410	0.455	0.506	0.417	0.413	0.500
check	MAP	0.410	0.434	0.511	0.513	0.756	0.422	0.543
	p	0.100	0.410	0.423	0.430	0.223	0.409	0.428
	R	0.103	0.343	0.421	0.426	0.352	0.257	0.426
FPS/f.s		202	204	274	232	63	243	256
Parameters/million model		4.839	10.009	10.004	30.940	9.831	7.104	10.558
sizze/MB		20.219	19.152	19.152	50.184	40.072	15.100	20.046

The following is a summary of the performance and comparative findings of each model, which can be obtained by doing an exhaustive analysis of the pertinent data in Table 1:

Overall, MobilNetv2-SSD had the weakest performance of all the networks. When it came to the job of target detection, the R index of the model was the lowest value in both the test and validation sets. This indicates that there were a large number of false positives in the model. On the other hand, the p-value showed that it performed well, and it's clear that the model's recognized parts were easier to see than the ignored ones. The reason for the current state of affairs is mainly because the target identification model is subject to strict limitations in the VisDrone dataset regarding firing angle, target size, and environmental complexity. The MobileNetv2-SSD is likely to have a high degree of application in activities that are less complicated, but it is not appropriate for this kind of sophisticated activity.

The aforementioned issues were also experienced by YOLOX-s

Both of the lightweight varieties of the YOLOv5-s and the YOLOv7-tiny, which were used in the test set, produced results that were very impressive. After the official release of a number of different versions, there was a significant improvement in the general performance of YOLOv5-s. Both the R and P indicators of the two models were found to be in a condition that was somewhat balanced, with the detection rate and the accuracy rate being aligned in a manner that was comparable to one another. The YOLOv7-tiny model was the one that had the fewest number of dimensions. This study's lightweight model provided improved performance on the test set, whereas YOLOv5-s and YOLOv7-tiny also demonstrated adequate effectiveness for target detection tasks in UAV aerial photography. The lightweight model was suggested in this work.

Applying the MAP index of the model delineated in this work to both the validation and test sets yielded optimal results. The MAP measurements exhibited superior performance, however the P, R, and MAP measures were comparable to or even exceeded those of the non-lightweight model YOLOv5-m. The model's performance was exemplary, indicating it achieved the greatest level of efficacy. To evaluate the model's performance, three indicators were utilized: frame rate per second, parameters, and model size. The proposed model demonstrated sufficient detection accuracy and effectively addressed deployment challenges, meeting the requirements of real-world production environments. Moreover, it exhibited a significant degree of robustness and practicality. This work successfully fulfilled the UAV aerial picture target identification job.

4.2 The Experiments Concerning Interpretability

"Black box" is a term that is often used to allude to deep learning. DL has not made significant progress in many high-tech industries, Despite the widespread use of deep learning models across several engineering fields.

4.3 Experiment with a Self-Constructed Dataset

A drone multitarget identification dataset is constructed in this article based on a variety of situations that took place in Guiyang, which is located in Guizhou Province, China. The objective of this study is to demonstrate the technique's universality. The performance impacts and comparative outcomes of each model, derived from the thorough examination of the pertinent data in Table 1, may be described as follows:

1. Both YOLOv4 s and YOLOv7-tiny got comparably low mean average precision scores on the test set, indicating analogous performance on the dataset generated by the researchers. Despite YOLOv5-tiny's minimal model size and reduced parameter count, its overall performance was inadequate. Nonetheless, these two kinds may still be used in situations when accuracy is not essential. Each achieved a frame rate above 150 frames per second and is suitable for deployment on edge devices.
2. The performance of YOLOv8-s, YOLOX-s, and YOLOv5-s was commendable, with each demonstrating equivalent detection accuracy. Nonetheless, YOLOv7-s attained the highest accuracy. In terms of frames per second (FPS), both YOLOv5-s and YOLOv8-s achieved 275, however YOLOX-s failed to reach 125. This indicates that the first two possess significant advantages in terms of detection accuracy and speed.
3. In the test set, the model presented in this article outperformed the original YOLO7 by around two percentage points and surpassed the YOLOv7-tiny, which exhibited the lowest performance, by twenty-one point four percent. During this period, the frame rate reached 302 frames per second, representing a reasonable balance between speed and detection accuracy. This indicates that the model developed in this study has attained superior detection performance across many datasets and conditions, demonstrating a significant level of universality. Conversely, there are situations when duplicate detection boxes may arise, as well as occasions where similar backgrounds might be erroneously identified as targets.

Conclusions

This study introduces a YOLOv8-s based approach to aerial picture recognition. While allowing the deployment of edge devices, this technique can accurately recognize targets in aerial images in real-time. By reducing the influence of factors like shot angle, lighting, and background, this method enhances the identification process. For the sake of further explanation: To address the common problem of misdetection and the inability to recognize small objects in aerial photography, the FPN - PAN -Bi idea is put up to improve the YOLOv8-s neck component. A more intricate and thorough feature fusion procedure may be achieved by thoroughly assessing and reimplementing multiscale features, all while minimizing parameter expenditures to the maximum degree possible. Incorporating a dynamic nonmonotonic focusing mechanism, the suggested approach is evaluated for its practicality and efficacy from several angles via the use of ablation, comparison, and interpretability tests. The data shows that the proposed improved method greatly helps in aerial picture recognition. Our method reduces the number of parameters by 12.87 percent and improves MAP performance by 11.15% when applied to the test set, in comparison to the baseline model. When compared to the six other algorithms that were used for comparison, the technique that was given in this study achieved higher accuracy. This approach exhibits a significant degree of interpretability about its performance. Moreover, the model achieved the highest detection accuracy 93.5% In the dataset it produced, functioning at a frame rate of up to 305 frames per second. The suggested approach is broadly applicable in intricate working situations, demonstrating significant versatility and resilience.

However, a problem arose throughout the experiment: the model presented in this paper did not outperform rival structures in any of the minor categories evaluated during the ablation sessions. Both the tricycle and the bus exhibit performance that is inferior than the combination of A, B, and C. In contrast, the van's performance exceeds that of the combination of A and B. The highlighted concerns will be the primary focus of future study, which will also include the incorporation of tailored detection tasks and the investigation of adaptive alterations to the model architecture regarding hyperparameter optimization and network design. Furthermore, the training of DL networks sometimes requires a considerable amount of labeled data, which might be prohibitive for applications related to aerial picture recognition. Future study will focus on using unsupervised theory to influence public and self-assembled datasets. This will be accomplished by reducing the disparities in data distribution across source and target domains, so ultimately lessening deep learning's need on labeled data.

References

1. Bao Z. The UAV target detection algorithm based on improved YOLO V8. In: Proceedings of the International Conference on Image Processing, Machine Learning and Pattern Recognition; 2024 Sep; p. 264-9.
2. Wang G, Chen Y, An P, Hong H, Hu J, Huang T. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. Sensors (Basel). 2023;23(16):7190.
3. Bakirci M. Enhancing vehicle detection in intelligent transportation systems via autonomous UAV platform and YOLOv8 integration. Appl Soft Comput. 2024;164:112015.
4. Li H, Li Y, Xiao L, Zhang Y, Cao L, Wu D. RLRD-YOLO: An improved YOLOv8 algorithm for small object detection from an unmanned aerial vehicle (UAV) perspective. Drones. 2025;9(4):293.
5. Rahman S, Rony JH, Uddin J, Samad MA. Real-time obstacle detection with YOLOv8 in a WSN using UAV aerial photography. J Imaging. 2023;9(10):216.
6. Feng F, Hu Y, Li W, Yang F. Improved YOLOv8 algorithms for small object detection in aerial imagery. J King Saud Univ Comput Inf Sci. 2024;36(6):102113.
7. Nie H, Pang H, Ma M, Zheng R. A lightweight remote sensing small target image detection algorithm based on improved YOLOv8. Sensors (Basel). 2024;24(9):2952.
8. Serageldin AF, Elsayed HA, Abdel-Hamid L. Real-time UAV detection using an enhanced YOLO v8 model. In: 2025 42nd National Radio Science Conference (NRSC); 2025 May; Vol. 1, p. 193-201. IEEE.
9. Qu J, Li Q, Pan J, Sun M, Lu X, Zhou Y, Zhu H. SS-YOLOv8: Small-size object detection algorithm based on improved YOLOv8 for UAV imagery. Multimed Syst. 2025;31(1):1-17.
10. Zhu J, Wu Y, Ma T. Multi-object detection for daily road maintenance inspection with UAV based on improved YOLOv8. IEEE Trans Intell Transp Syst. 2024.
11. Bi J, Li K, Zheng X, Zhang G, Lei T. SPDC-YOLO: An efficient small target detection network based on improved YOLOv8 for drone aerial image. Remote Sens. 2025;17(4):685.
12. Enhancement for heat pipe using different configurations: review study. NTU-JET. 2025 Mar;4(1). doi:10.56286/t0kbhe55.
13. Early prediction of stroke risk using machine learning approaches and imbalanced data. NTU-JET. 2025 Mar;4(1). doi:10.56286/1vf19469.
14. Design, modeling, and comprehensive analysis for the wing of an aircraft. NTU-JET. 2025 Mar;4(1). doi:10.56286/txv7rh29.