International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 November 2023* ***Revised:*** *12 December 2023* ***Accepted:*** *30 January 2024*

An Improved YOLOv8-Based Method for Small Object Detection in UAV perspective

Xiaodong Zhang
Faculty of Engineering, Built Environment, and Information Technology
SEGi University
Kuala Lumpur, Malaysia
zhangxd2641@gmail.com

Yong Chai Tan
Faculty of Engineering, Built Environment, and Information Technology
SEGi University
Kuala Lumpur, Malaysia
tanyongchai@segi.edu.my

Abstract—The target objects in UAV aerial images are usually smaller targets, and the detection of such small targets is an important research area. Although the progress of target detection is constantly improving thanks to the advancement of deep learning technology, the accuracy of small target recognition in images still poses a great challenge. This study proposes an improved YOLOv8 algorithm to improve the performance of the small target object detection algorithm. This method proposes a C2F-DCNv2 module that integrates deformable convolutional network v2 (DCNv2) to replace the C2F module of the original backbone part; in addition, a Dynamic Head (DyHead) with a self-attention mechanism is used on the head) replaced the original detection head. Through training and testing in the VisDrones2019 data set, it is shown that the method proposed in this article reached 37.9% in the mAp50 indicator in the verification data set, and the average detection speed was 33.7 FPS. Compared with the results of the baseline model, the results increased by 3.6%. Experimental results show that the target detection algorithm proposed in this article significantly improves the recognition effect of small targets in UAV aerial images.

Keywords-small-object detection; Deformable Convolutional; Dynamic Head; YOLOv8

I. INTRODUCTION *(HEADING 1)*

In the field of target recognition, the recognition of target objects based on UAV aerial images is a dedicated field, which refers to the use of cameras equipped with UAVs to track and identify ground targets. UAVs have huge application potential in various fields due to their high maneuverability, excellent traffic capacity, and low cost of use. With the development of artificial intelligence technology, the accuracy of target recognition technology based on drone images is constantly improving. Today, drone aerial images have been widely used in traffic management, environmental protection, agricultural monitoring, security, emergency response and urban planning.

However, due to the flight height of UAVs, targets in UAV aerial images often appear as relatively small and tiny features in the images. Recognizing such small objects is still very challenging for target recognition algorithms. At present, some researchers' research on small target recognition algorithms mainly focuses on methods such as data enhancement, scale awareness, feature fusion and super-resolution. However, these algorithms generally do not take into account factors such as the changing shooting conditions and complex scenes of UAV aerial images. General small target recognition algorithms cannot be applied to UAV aerial images. Therefore, this paper considers the special application field of UAV aerial images in order to improve the accuracy and calculation efficiency of recognition. Based on the YOLOv8n model [1], this study uses the C2F-DCNv2 module with deformable convolutional network version 2 (DCNv2) [2] to replace the original C2F, which improves the recognition accuracy of small targets without affecting the calculation efficiency. On the other hand, the DyHead (Dynamic Head) [3] module with a self-attention mechanism is used in the head of YOLOv8n to replace the original detection head to improve the extraction of features of small targets in complex scenes.

II. RELATED WORK

A. Small Object Detection

Small target object detection refers to the identification and detection of targets that are smaller than a certain proportion of the size of the target to be detected in the image. For example, the Society for International Optical Engineering (SPIE) defines targets that are smaller than 80 pixels in a 256×256 image. Become a small target. The COCO data set specifies that small targets refer to targets with pixels smaller than 32×32 [4]. Many scholars in this field are constantly trying to improve existing algorithms to improve the detection capabilities of small targets. The SF-YOLOv5 [5] algorithm proposed by Liu et al. uses the feature fusion technology of PANet and BiFPN to improve the accuracy of detection. Wang et al. [6] integrated the attention mechanism BiFormer module into YOLOv8, improving the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 November 2023* ***Revised:*** *12 December 2023* ***Accepted:*** *30 January 2024*

model's ability to perceive detailed features. In recent years, most of the optimization of algorithm models uses convolution modification, attention mechanism and loss function improvement to improve the original structure of the model, thereby improving the accuracy of the model.

Wherever Times is specified, Times Roman or Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts if possible. True-Type 1 or Open Type fonts are preferred. Please embed symbol fonts, as well, for math, etc.

B. YOLOv8 algorithm introduction

Algorithms using deep learning technology in the field of target detection currently mainly include two-step methods and single-step methods. The two-step method is to first perform regional positioning and then perform target classification. Representative algorithms include RCNN, Fast R-CNN, and Faster R-CNN. On the other hand, the one-step method represented by the YOLO (You Only Look Once) [7] series chooses to directly regress the distribution probability and coordinates of the target. Although the accuracy of this method is slightly lower, its detection speed has been greatly improved. promote.

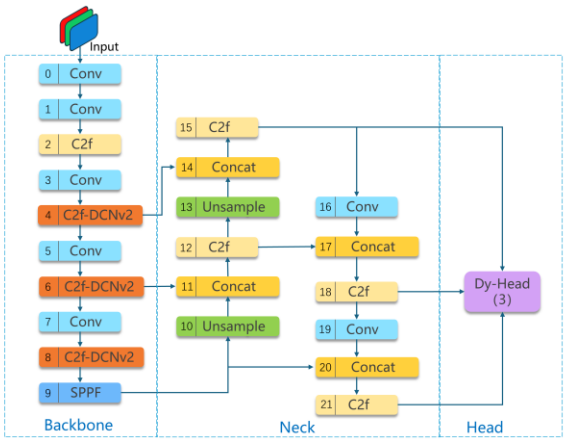
The YOLO series has played an important role in advancing real-time target detection in the field of computer vision. Compared with the previous version, the latest version YOLOv8 has made significant technical improvements, especially the use of anchor-free object center prediction methods, which improves the computational efficiency of non-maximum suppression (NMS). At the same time, YOLOv8 has also improved the original convolution module, replacing the C3 module with the C2F module and merging the ResNet module. These improvements have greatly improved the accuracy and computing efficiency of YOLOv8 compared with the previous version. To meet the needs of different users. YOLOv8 offers a range of model sizes to suit different operational needs and computational constraints, from the fast YOLOv8n to the highly accurate YOLOv8x. Users can choose models of different sizes according to their own accuracy requirements and computing power limitations.

III. METHODS

In order to achieve the goal of improving the accuracy of small target detection in UAV aerial images while taking into account the limitations of computing power, this paper proposes an improved model based on the YOLOv8n model. The improved solution is mainly divided into two parts, by merging the C2F-DCNv2 module into the appropriate layer of the YOLOv8 backbone network, while adopting the DyHead detection head. The C2F-DCNv2 module refers to the deformable convolution v2 (DCNv2) replacing the original convolution in the C2F module; and the DyHead module introduces a self-attention mechanism, which can simultaneously improve the detection capabilities of scale, space and task awareness. As shown in Figure 1, we replaced the C2F module with the C2F-DCNv2 module on the 4th, 6th and 8th layers of the backbone part, and replaced the DyHead module on the head to improve the detection efficiency of small target objects in the public data set. promote.

A. Deformable Convolution Networks

In traditional convolution, since the same convolution kernel is applied at each position, it has good performance in feature recognition of simple objects with relatively regular shapes. However, there are still certain limitations when it comes to some deformed features or more complex features. limitations. The deformable convolutional network (DCN) [8] adds a learnable offset to the convolution kernel in the traditional convolutional network, so that the sampling position of the convolution kernel will also deform accordingly, which can adapt to the extraction of features in more complex situations. In addition, DCN introduces necessary nonlinear processes into the network by integrating batch normalization and activation functions. In the subsequent version of DCNv2, a learnable weight term was added, so that the weight of each sampling point can be obtained on the basis of being able to learn the offset, thus



once again increasing the flexibility of deformable convolution.

Figure 1. Improved network structure diagram

As shown in Figure 2, the basic principle of deformable convolution is to use a learnable offset field so that the convolution kernel of standard convolution can achieve an offset closer to the feature of interest, which can better adapt to different Characteristics. As in formula (1), it represents the operation of traditional convolution, p represents a certain position in the feature map; K represents the total number of convolution samples, pk represents the offset constant with each sampling point in the convolution kernel, $p+pk$ determines the position covered by the convolution kernel. Therefore, since pk is a constant in the standard convolution kernel, it determines that the shape of the convolution kernel may not better match the deformation structure existing in the feature image. In order to solve this problem, deformable convolution introduces a new scientifically learned parameter Δp to expand the fixed pk into an adaptive parameter, as shown in formula (2). This value of Δp corresponds to the offsets in different directions represented in offsets in the figure. In the deformable convolutional network version 2 (DCNv2), a weight term Δmk, called a modulation scalar [2] , is added to each sampling point to further expand its adaptability.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 November 2023* ***Revised:*** *12 December 2023* ***Accepted:*** *30 January 2024*

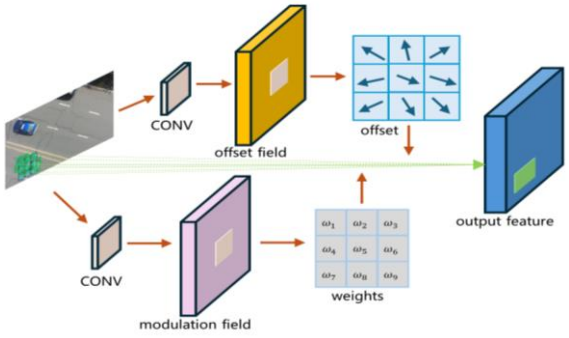


Figure 2. *Deformable convolutional network structure*

$$y(p) = \sum_{k=1}^{K} w_k \cdot x(p + p_k) \quad (1)$$

$$y(p) = \sum_{k=1}^{K} w_k \cdot x(p + p_k + \Delta p_k) \cdot \Delta m_k \quad (2)$$

In the solution proposed in this article, we integrated YOLOv8's original C2F module with DCNv2, replacing the traditional convolution to form the C2F-DCNv2 module, as shown in Figure 3. Through this improvement, when the module performs convolution operations, the sampling position is more consistent with the shape and scale of the target itself, which is more conducive to feature extraction of small targets.

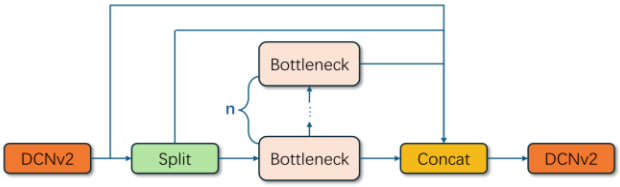


Figure 3. *Structure of C2F-DCNv2*

B. *Dynamic Head*

The Dynamic Head [3] framework represents an innovative approach to object detection that integrates detection heads with attention mechanisms in a cohesive structure. Illustrated in Figure 4, the framework's operational principle commences with a feature pyramid network that supplies multi-scale feature representations. Subsequently, these features are processed through the Dynamic Head, undergoing successive refinements by distinct attention modules. Initially, the scale-aware attention module dynamically adjusts feature representations to optimize the detection of objects across varying sizes. Finally, the spatial-aware attention module selectively concentrates on particular regions within the image, enhancing the focus on areas of interest. The process concludes with the task-aware attention module, which differentially prioritizes output channels based on the specific requirements of various detection tasks, such as classification, center-point regression, and boundary regression. This systematic approach allows the Dynamic Head framework to adeptly manage the complexities inherent in object detection, ensuring tailored attention across scales, spatial domains, and task-specific demands.

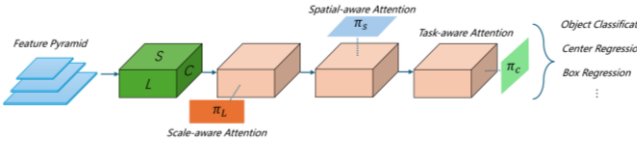


Figure 4. *Dynamic head module diagram*

The DyHead module incorporates sophisticated architecture to embody scale, spatial, and task awareness within object detection frameworks, as delineated in Figure 5. Central to this module is a multi-level self-attention mechanism, systematically formulated to enhance feature processing. This mechanism operates on the input feature vector F, which conforms to the three-dimensional tensor structure defined by equation (4), with dimensions representing the feature level (L), spatial dimensions (S), and channel count (C). Within this framework, self-attention functions are applied across different dimensions of the feature vector, namely channel-wise, spatial-wise, and level-wise attention, detailed by the respective computational formulations in equations (4), (5), and (6). These attention mechanisms enable the DyHead module to dynamically adjust feature emphasis across channels, spatial regions, and hierarchical levels, thereby optimizing the detection model for varying object scales, spatial distributions, and task-specific demands.

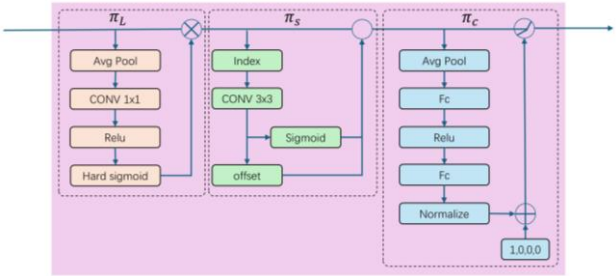


Figure 5. Example of a figure caption. *(figure caption)*

$$W(F) = \pi_C(\pi_S(\pi_L(F) \cdot F) \cdot F) \quad (3)$$

$$\pi_L(F) \cdot F = \sigma\left(f\left(\frac{1}{SC}\sum_{S,C} F\right)\right) \cdot F \quad (4)$$

$$\pi_S(F) \cdot F = \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K} w_{l,k} \cdot F(l; p_k + \Delta p_k; c) \cdot \Delta m_k \quad (5)$$

$$\pi_S(F) \cdot F = \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K} w_{l,k} \cdot F(l; p_k + \Delta p_k; c) \cdot \Delta m_k \quad (6)$$

The YOLOv8 model's original detection head encounters limitations in processing multi-scale target detection, primarily due to its reliance on a single-scale prediction framework and lack of contextual information integration. This approach leads to each prediction being made in isolation, without a comprehensive evaluation across various scales or contextual insights, compromising the detection of small and complexly situated targets.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
***Article Received:** 25 November 2023 **Revised:** 12 December 2023 **Accepted:** 30 January 2024*

To address these deficiencies, integrating the DyHead in place of YOLOv8's original detection head proposes a substantial enhancement. This integration leverages features from the 15th, 18th, and 21st layers of the backbone network, encompassing diverse scales and connecting them to DyHead, as illustrated in Figure 1.

IV. Experimental Results and Analysis

A. Datasets

In this study, to validate the efficacy of our proposed model, a comprehensive analysis and comparison of prevalent datasets from a UAV perspective were undertaken. Among them, notable datasets such as the Stanford Drone Dataset, AU-AIR Dataset, The Highway Drone Dataset, SeaDronesSee Dataset, and VisDrone2019-DET Dataset [9] were meticulously evaluated. Given the unique scenarios encapsulated within certain datasets, a detailed comparison—encompassing scenario types, target classifications, and dataset volumes — was conducted to ascertain the most fitting dataset for assessing our model's versatility.

The decision to utilize the VisDrone2019-DET Dataset was informed by its extensive coverage of UAV aerial imagery, sourced from varied urban settings, environmental conditions, and operational scenarios. This dataset encompasses twelve distinct target categories, including pedestrians, vehicles (bicycles, cars, vans, trucks, buses, and motorcycles), and specialized categories like tricycles and awning-tricycles, among others. With a comprehensive composition of 6,471 training images, 548 validation images, and 1,580 test images [9], the VisDrone2019-DET Dataset offers a robust platform for model validation.

B. Experimental environment parameter settings and evaluation indicators

1) Experimental environment parameter

To ensure a rigorous and unbiased evaluation of experimental outcomes, this study meticulously standardized the training and testing conditions across all comparative analyses. Uniformity was maintained in both hardware specifications and hyperparameter settings to preclude any discrepancies that might influence the results. Table 1 shows the specific configuration of the experimental environment. In addition, the pre-trained model Yolov8n was loaded in the experiment. The input image size was 640x640. The optimizer used was the stochastic gradient descent (SGD) optimizer. Each experimental iteration was 200 cycles.

TABLE I. Experimental Environment Parameters

Parameters	Configuration
Operating System	Ubuntu18.04
CPU	Intel(R) Xeon(R) Silver 4214R CPU
GPU	RTX 3080 Ti(12GB)
Deep Learning Environment	Python3.8+Pytorch11.3+CUDA11.3

2) Evaluation index

When selecting evaluation parameters, we selected mAP50 (average accuracy at IoU threshold of 0.5) to evaluate the performance of the object detection model for different categories in the dataset. It calculates the average of the average precision (AP) values across all categories, thus reflecting the overall performance of the model on all targets. The mathematical representation of mAP50 is given by Equation (7), where N represents the total number of categories and AP_i represents the AP of the category.

In addition, considering that the application object involved in this article may be the airborne computing platform of a drone. In order to meet the requirements of real-time computing, we mainly refer to the inference speed of the model when calculating frames per second (FPS) and considering the computing power and storage capacity of the airborne platform, we use parameter size as a model size evaluation indicator.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \qquad (7)$$

C. Experimental design

This study introduces a series of enhancements targeting both the backbone and detection head of the YOLOv8 model. Our experimental framework is structured around three primary investigations: Firstly, we evaluate various backbone enhancement strategies to ascertain their impact on the model's feature extraction capabilities. Secondly, we explore improvements to the detection head, aiming to refine the model's accuracy and efficiency in object identification. Lastly, we conduct comprehensive ablation studies on both the backbone and detection head modifications. These studies are designed to isolate and evaluate the contributions of individual enhancements to the overall performance improvements observed in the enhanced YOLOv8 model.

1) Comparison of different improvement solutions for C2F-DCNv2

In refining the backbone structure of the YOLOv8 model, we employed the C2F-DCNv2 module, which incorporates Deformable Convolutional Networks version 2 (DCNv2), to replace the existing C2F module. Notably, the potential locations for integrating the C2F-DCNv2 module within the backbone are identified at the 4th, 6th, and 8th layers, as depicted in Figure 1. To determine the most effective layer(s) for this modification, we devised a performance comparison experiment. This experiment aimed to assess the impact of substituting the C2F module with C2F-DCNv2 at different layers on the overall model performance. The outcomes of these comparisons are systematically presented in Table 2.

Analysis shows that replacing the C2F modules in layers 4, 6, and 8 with C2F-DCNv2 resulted in a performance improvement of 0.1% to 0.6% in the model. At the same time, the number of parameters did not increase significantly, and the FPS decrease was also within the acceptable range. It is worth noting that these improvements improve the accuracy without significantly changing the calculation speed and parameter count of the model, indicating that the performance has been effectively improved. Since the improvement accuracy of different replacement schemes has a small difference, we used the top three improvement accuracy schemes for comparative experiments in subsequent experiments. In subsequent experiments we will choose to replace the 4th or 8th layers in the backbone respectively, and replace the 4th, 6th, and 8th layers at the same time. We will choose three options to replace the (4) layer, (8) layer, and (4, 6, 8) layer as a control experiment. and select the best overall improvement solution.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
Article Received: *25 November 2023* ***Revised:*** *12 December 2023* ***Accepted:*** *30 January 2024*

TABLE II. COMPARISON OF C2F-DCNV2 REPLACING DIFFERENT LAYERS

Layers			mAp50(%)	FPS	Parameter(M)
4	6	8			
-	-	-	34.8	344.8	3.01
√	-	-	**35.2**	322.6	3.02
-	√	-	35.0	322.6	3.04
-	-	√	**35.4**	333.3	3.04
√	√	-	35.1	324.6	3.05
√	-	√	34.9	321.5	3.05
-	√	√	35.0	320.4	3.07
√	√	√	**35.3**	318.7	3.09

2) Performance comparison of different numbers of DyHead modules in series

The improvement of the head mainly involves replacing the original detection head with DyHead, but the main consideration here is that DyHead can be used in series. Therefore, a set of experiments was designed to compare the performance of different numbers of DyHead modules in the series. As shown in Table 3, it can be seen from the experimental results that when the number of series connections is 3, the accuracy is the highest and the FPS exceeds 30. But when the number of series connections is increased on this basis, the accuracy decreases and the FPS drops significantly. Therefore, based on the comprehensive measurement of improved detection accuracy and increased computational complexity, we chose to connect 3 DyHead modules in series in subsequent experiments.

TABLE III. COMPARISON OF DIFFERENT NUMBERS OF DYHEADS

DyHead Numbers	mAp50(%)	FPS	Parameter(M)
Baseline(Yolov8n)	34.8	344.8	3.01
1	36.4	72.4	3.49
2	36.9	66.3	3.99
3	**37.4**	36.9	4.48
4	37.2	26.0	4.98

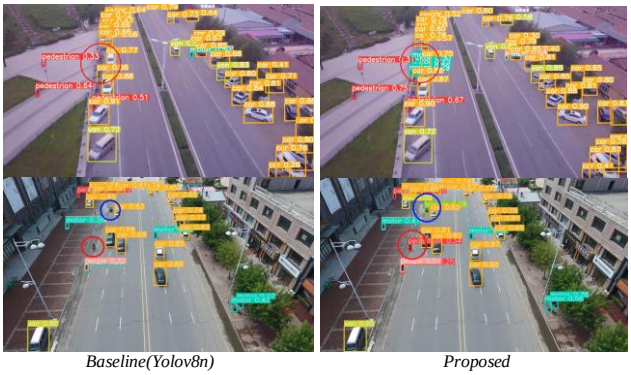
3) Ablation experiment

In order to verify the contribution of the changes in each part of the improvement plan proposed in this article to the final performance, we designed an ablation experiment to gradually remove or "ablate" the improvement of the backbone and head of the model to prove the improvement proposed in this article. effectiveness of the method. As shown in Table 4, 4 control experiments were used. “ √ ” and “-” in the table indicate whether the improvements of this module are used.

The results of Experiment 1 and Experiment 2 show that in the ablation experiment of this article, three C2F-DCNv2 schemes were used to replace the C2F module in the 8th layer of the backbone part as a comparison, replacing the 4th layer, the 8th layer and the 468th layer respectively. In the improvement of the head network, the reasonable number of DyHead modules is 3. Therefore, the data in the table shows that when the backbone part improvement is used alone, the model detection accuracy is improved by 0.6% compared to Baseline, and when the head improvement is used alone, the model detection accuracy is 3.2% improved compared to Baseline, and and the FPS is within the acceptable range. When the backbone and head are improved at the same time, the backbone is selected to replace the C2F modules of layers 4, 6, and 8. The improvement is most obvious. The accuracy rate increased by 3.6% compared with Baseline. Compared with improving the backbone alone, it has increased by 2.1%, and compared with improving the head alone, it has increased by 0.5%. Judging from the results of the ablation experiment, both DCNv2 and DyHead can improve the accuracy, and the rational use of DCNv2 and DyHead at the same time can achieve greater improvements. As shown in Figure 6, the improved model has a higher accuracy in target detection in drone aerial images than the original model.

TABLE IV. COMPARISON OF ABLATION EXPERIMENTS

Improvement Module				mAp50(%)	FPS	Parameter(M)
C2f-DCNv2			DyHead (3)			
4	8	468				
-	-	-	-	34.3	344.8	3.01
-	-	-	√	37.4	36.9	4.48
√	-	-	-	35.2	322.6	3.02
-	√	-	-	35.4	333.3	3.04
-	-	√	-	35.3	318.7	3.09
√	-	-	√	37.4	29.4	4.51
-	√	-	√	37.2	25.7	4.51
-	-	√	√	**37.9**	33.7	4.56



Baseline(Yolov8n) *Proposed*

Figure 6. Improved performance comparison.

V. CONCLUSIONS

Given the emphasis on small target recognition in drone aerial photography, this study introduces a YOLOv8n-based algorithm tailored for drone perspectives to enhance the precision and efficiency of detecting small targets. Key modifications include replacing the original C2F module's convolution with the deformable convolution DCNv2, yielding a novel C2F-DCNv2 module. This adjustment aims to bolster the network's capacity for feature extraction across various target sizes. In addition, the network's detection head is upgraded through the DyHead module, using its scale, space and task-aware attention mechanism to improve the model's performance in identifying targets at different scales and complex backgrounds. Experiments have verified the effectiveness of the

proposed modifications, and the accuracy mAp50 in the VisDrone2019-DET data set is 3.6% higher than the baseline model. Especially in the case of complex scenes and smaller targets, the improvement in recognition accuracy is most obvious. However, given the computational limitations of UAV platforms, future efforts will focus on optimizing model compactness and computational efficiency without compromising accuracy improvements.

REFERENCES

[1] M. Hussain, "YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection," *Machines*, vol. 11, no. 7, p. 677, Jun. 2023, doi: 10.3390/machines11070677.

[2] X. Zhu, H. Hu, S. Lin, and J. Dai, "Deformable ConvNets V2: More Deformable, Better Results," presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 9308–9316. Accessed: Apr. 03, 2024. [Online]. Available: https://openaccess.thecvf.com/content_CVPR_2019/html/Zhu_Deformable_ConvNets_V2_More_Deformable_Better_Results_CVPR_2019_paper.html

[3] X. Dai *et al.*, "Dynamic Head: Unifying Object Detection Heads with Attentions," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA: IEEE, Jun. 2021, pp. 7369–7378. doi: 10.1109/CVPR46437.2021.00729.

[4] B. Mirzaei, H. Nezamabadi-pour, A. Raoof, and R. Derakhshani, "Small Object Detection and Tracking: A Comprehensive Review," *Sensors*, vol. 23, no. 15, p. 6887, Aug. 2023, doi: 10.3390/s23156887.

[5] H. Liu, F. Sun, J. Gu, and L. Deng, "SF-YOLOv5: A Lightweight Small Object Detection Algorithm Based on Improved Feature Fusion Mode," *Sensors*, vol. 22, no. 15, Art. no. 15, Jan. 2022, doi: 10.3390/s22155817.

[6] G. Wang, Y. Chen, P. An, H. Hong, J. Hu, and T. Huang, "UAV-YOLOv8: A Small-Object-Detection Model Based on Improved YOLOv8 for UAV Aerial Photography Scenarios," *Sensors*, vol. 23, no. 16, Art. no. 16, Jan. 2023, doi: 10.3390/s23167190.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.

[8] J. Dai *et al.*, "Deformable Convolutional Networks," presented at the Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 764–773. Accessed: Apr. 04, 2024. [Online]. Available: https://openaccess.thecvf.com/content_iccv_2017/html/Dai_Deformable_Convolutional_Networks_ICCV_2017_paper.html

[9] D. Du *et al.*, "VisDrone-DET2019: The Vision Meets Drone Object Detection in Image Challenge Results," presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019, pp. 0–0. Accessed: Apr. 06, 2024. [Online]. Available: https://openaccess.thecvf.com/content_ICCVW_2019/html/VISDrone/Du_VisDrone-DET2019_The_Vision_Meets_Drone_Object_Detection_in_Image_Challenge_ICCVW_2019_paper.html