Computer Science and Application 计算机科学与应用, 2024, 14(5), 137-146
Published Online May 2024 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2024.145122

Hans汉斯

基于ASF-WIoU-YOLOv8的无人机航拍图像多目标检测算法

殷　波

贵州交通职业技术学院机械电子工程系，贵州　贵阳

收稿日期：2024年4月25日；录用日期：2024年5月23日；发布日期：2024年5月31日

摘　要

针对无人机航拍图像的多目标检测问题，本文提出了一种基于改进YOLOv8的目标检测算法ASF-WIoU-YOLOv8。首先，在YOLOv8的基础架构上，加入一种注意尺度序列融合机制(Attentional Scale Sequence Fusion-ASF)，该机制能够对不同尺度的特征图进行融合，从而获得更好的图像特征，提取出更丰富、更准确的特征信息。然后，对损失函数进行改进，引入Wise-IoU机制，该机制通过自适应地调整权重系数提高目标检测的灵活性和鲁棒性，从而进一步提高算法的检测精度。实验结果表明，在VisDrone数据集上，本文所提算法比YOLOv8算法的平均精度mAP50提升了2.0%，该算法在无人机航拍图像上具有更高的检测精度。

关键词

目标检测，YOLOv8，注意尺度序列融合，Wise-IoU

Multi-Object Detection Algorithm for UAV Aerial Images Based on ASF-WIoU-YOLOv8

Bo Yin

Department of Mechanical and Electronic Engineering, GuiZhou Communications Polytechnic, Guiyang GuiZhou

Received: Apr. 25th, 2024; accepted: May 23rd, 2024; published: May 31st, 2024

Abstract

For the problem of multi-object detection in UAV aerial images, this paper presents an improved YOLOv8 object detection algorithm named ASF-WIoU-YOLOv8. Firstly, on the infrastructure of

文章引用: 殷波. 基于 ASF-WIoU-YOLOv8 的无人机航拍图像多目标检测算法[J]. 计算机科学与应用, 2024, 14(5): 137-146. DOI: 10.12677/csa.2024.145122

YOLOv8, an attention-scale sequence fusion mechanism (Attentional Scale Sequence Fusion-ASF) is added, which can integrate feature maps at different scales, so as to obtain better image features and extract richer and more accurate feature information. Then, the loss function is improved by using the Wise-IoU mechanism, which improves the flexibility and robustness of target detection by adaptively adjusting the weight coefficients. Wise-IoU can further improve the detection accuracy of the algorithm. The experimental results show that the average accuracy of the proposed algorithm is 2.0% higher than the YOLOv8 algorithm, which has higher detection accuracy on aerial images of UAV.

Keywords

Target Detection, YOLOv8, Attention-Scale Sequence Fusion, Wise-IoU

1. 引言

随着无人机技术的蓬勃发展，无人机凭借其机动灵活的特点，能够实现大范围的区域监测，现已广泛应用于各行各业当中[1]。针对无人机航拍图像的目标检测是无人机的一个重要应用，在民用和军事领域中发挥着重要作用，有助于测绘航测、应急救援、危险区域监测和识别易受灾地区等诸多方面[2]。

基于深度学习的目标检测算法一般可分为两类：以 R-CNN 为代表的二阶段(two-stage)检测方法和以 YOLO 系列为代表的一阶段(one-stage)检测方法。两阶段方法在数据特征提取之后先生成区域提取(Region Proposal)网络，再进行样本的分类与定位回归，代表性算法有区域卷积神经网络(Region based Convolutional Neural Network, R-CNN) [3]、快速区域卷积神经网络(Fast Region based Convolutional Neural Network, Fast R-CNN) [4]、更快的区域卷积神经网络(Faster Region based Convolutional Neural Network, Faster RCNN) [5]。一阶段检测方法从最开始的提取特征到最后预测类别和边界框回归信息，是一个整体的过程，代表性算法代表算法有 SSD 系列算法[6]和 YOLO 系列算法[7]。

尽管基于深度学习的目标检测已经具备了很好的效果，但其在无人机目标检测中表现不佳。由于无人机航拍图像相比于自然场景的图像，具有大场景、多尺度、小目标、背景复杂和相互遮挡的特点，使得目标检测的精度不是很高。文献[8]为了充分利用可见光图像和红外图像的优点，提出了一种图像融合的目标检测算法，通过优化和重新设计 YOLOv2 算法，提高了其在嵌入式平台上的性能。文献[9]设计了一种新的自我关注机制，将查询向量和周围环形区域的关键向量分开计算，提高了旋翼无人机数据集上的检测精度。为了更好地完成无人机航拍图像多目标检测，近几年来的主流方法大多都是基于 YOLO 系列算法来实现的。在文献[10]中，研究者对 YOLOv5 算法进行了优化，将原有的 CIoU 替换为 Focal EIoU，此举显著提升了模型的收敛速度和回归精确度。而文献[11]则专注于提升无人机航拍图像中小目标物体的检测效果。通过在 YOLOv7 网络中加入 SPPFS 金字塔池化模块、优化损失函数以及引入 CBAM 注意力机制等手段，该网络对小目标物体的检测精度得到了显著提升。然而，这样的改进也导致了网络结构的复杂化。在文献[12]中，同样为了提高对小目标物体的检测能力，研究者对 YOLOv7 网络进行了改进，增加了专门用于小目标检测的网络层，并引入了注意力机制。这些措施确实提高了网络的检测精度，但相应地也增加了网络的复杂度，提高了所需的参数量，并增加了网络层数。

由于 YOLO 系列算法的快速发展，YOLOv8 算法在目标检测领域已经展现出强大的性能，因此针对无人机航拍图像的多目标检测问题，本文在 YOLOv8 的算法基础上进行改进，将注意尺度序列融合机制(ASF) [13]融入到 YOLOv8 算法中，以增强网络对多个不同种类不同尺度的目标的检测能力，另外引入 Wise-IoU 机制[14]，通过自适应调整权重系数，进一步提高算法的检测精度。

2. 本文总体方案

2.1. 总体网络结构

本文提出了一种基于改进 yolov8 的多目标检测算法 ASF-YOLv8，该算法模型主要包括三大模块：主干特征提取模块(Backbone)、特征加强模块(Neck)、检测模块(Detect)三个部分构成，如图 1 所示。

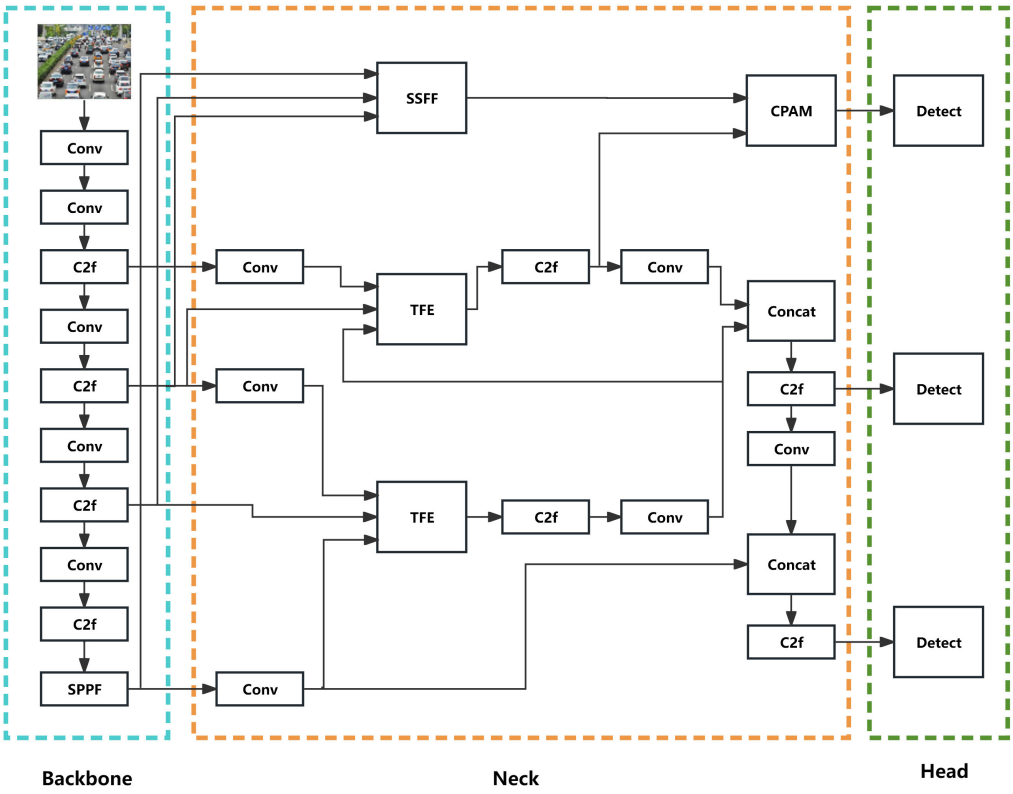


Figure 1. Network structure of the proposed algorithm
图 1. 本文所提算法的网络结构

在图 1 所示的 ASF-YOLOv8 算法网络结构图中，Conv、C2f、SPPF、Concat 和 Detect 模块的操作流程和 YOLOv8 一样，在此本文不做过多说明。SSFF-YOLOv8 算法的主干特征提取模块 Backbone 仍然沿用了 YOLOv8 的 CSPDarkNet 结构；特征加强模块(Neck)则进行了重新设计，主要是在 Neck 部分加入了 TFE、SSFF 和 CPAM 模块，其中 TFE 模块的操作流程如图 2 所示，SSFF 和 CPAM 模块在 2.2 节和 2.3 节详细说明；检测模块(Detect)采用了和 YOLOv8 一样的三个解耦检测头，分别用来检测大目标、中目标和小目标。

本文所做改进重点在 Neck 部分：

(1) 本文在 TFE 模块前增加了 Conv 模块，该模块的一个作用是加强特征信息的融合度，第二个作用是通过调整 Conv 模块的参数使 TFE 模块的三个输入能够达到尺度统一，方便后续处理；

(2) 本文用 C2f + Conv 模块代替 ASF 算法里的 CSP 模块，主要原因是 CSP 模块里面用的是 C3 结构

(YOLOv5 算法用的)，本文用 C2f 结构代替 C3 模块，可以增加网络性能，提高估计精度，另外后续可以继续改进 C2f 模块，增加了网络的灵活性；

(3) 本文的 CPAM 模块的输出直接进入检测层(即 Head 部分的最上面的 Detect)，不再向下面两层传递信息，而在 ASF 算法里面 CPAM 模块的输出信息还会继续向下面两层传输，参与 Head 部门的下面两个 Detect。本文的这种改进降低了网络的复杂度，大大降低了运算量，而且还能保持相对高的精度。这样改动的主要原因是 CPAM 模块的输出原本就是通过下面两层信息不断上采样之后与最上层信息融合的，因此最下面两层检测再通过 CPAM 模块得到有效的信息量就比较少。

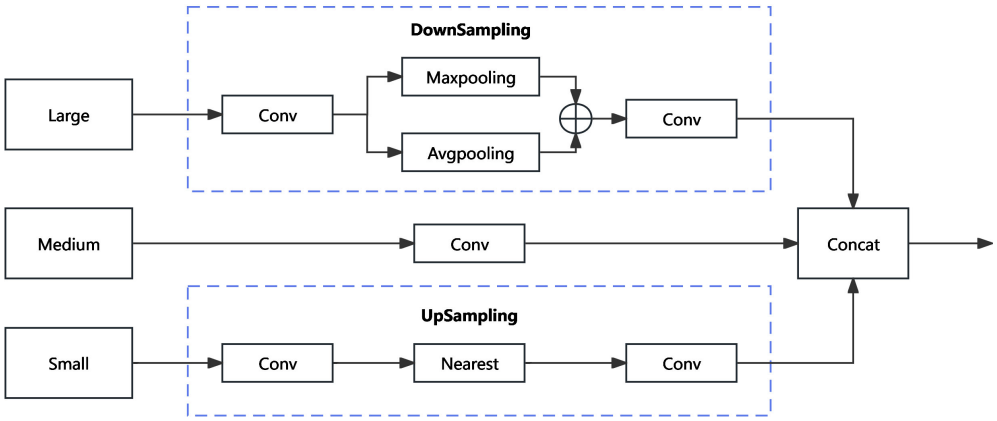


Figure 2. TFE modular structure
图 2. TFE 模块结构

TFE 模块的结构如图 2 所示，它的主要作用是为了将大、中、小三个尺度的特征图信息进行融合。大尺度的特征图 large 依次经过卷积 Conv 模块、下采样 DownSampling 模块和卷积 Conv 模块后在 Concat 模块与另外两个支路融合，其中 DownSampling 模块采用两类下采样方式：最大池化 Maxpooling 和平均池化 Avgpooling。中尺度的特征图则经过一个卷积 Conv 模块直接融合，小尺度的特征图依次经过卷积 Conv 模块、上采样 UpSampling 模块和卷积 Conv 模块后进行融合，其中上采样方式为最近邻插值方法。TFE 模块通过将不同尺度的特征图信息进行融合，能够改善网络对不同尺寸目标的识别检测能力，尤其是提高对小目标信息的提取能力。

2.2. 尺度序列特征融合模块(SSFF)

SSFF 模块的结构如图 3 所示，它的主要作用是将多个尺度图像的全局或高级语义信息进行融合。SSFF 模块有三个输入，分别对于大、中、小尺度的特征图，三个特征图经过前期的处理后达到相同的尺寸，在 Stack 模块进行堆叠，之后进入 3D Conv 模块进行 3d 卷积操作，然后经过 BN 模块归一化、SiLU 模块函数激活后得到输出。SSFF 模块通过对三个尺度的特征图进行融合，增强了网络的多尺度信息提取能力。

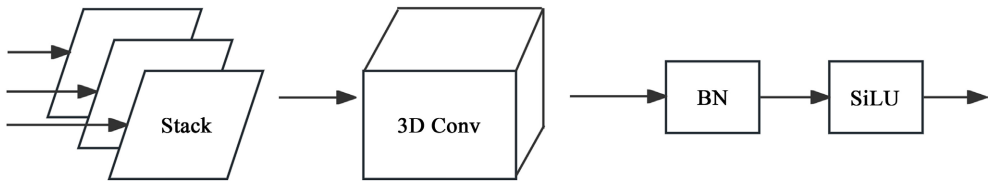


Figure 3. SSFF modular structure
图 3. SSFF 模块结构

2.3. 通道和注意力模块(CPAM)

CPAM 模块的结构如图 4 所示，它包括两部分网络结构信道注意力网络 Channel attention networks 和位置注意力网络 Position attention networks。CPAM 模块有两个输入 Input1 和 Input2，其中 Input1 来自 TFE 模块的输出，Input2 来自 SSFF 模块的输出。

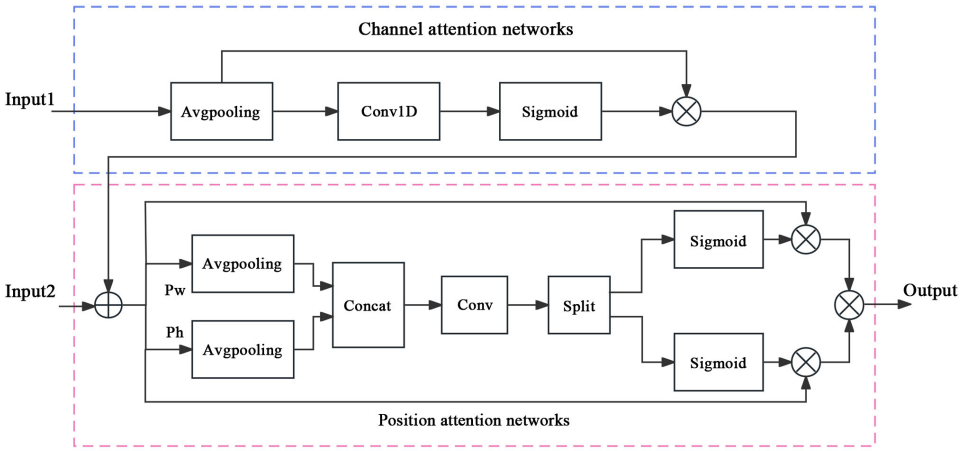


Figure 4. CPAM modular structure
图 4. CPAM 模块结构

信道注意力网络 Channel attention networks 首先对每个信道采用全局平均池化 Avgpooling，然后经过 1d 的卷积 Conv1D 模块和激活函数 Sigmoid 模块，最后与没经过 Conv1D 和 Sigmoid 模块的支路相乘。信道注意力网络可以增强信道之间的相互作用信息。位置注意力网络 Position attention networks 将信道注意力网络的输出和 Input2 进行合并作为输入，之后分成两个支路进行处理(Pw 和 Ph 支路，其中 Pw 支路提取网络的宽度信息，Ph 支路提取网络的高度信息)，先通过平均池化 Avgpooling，然后进入 Concat 模块进行拼接，之后经过卷积 Conv 模块，在 Split 模块进行信道又分离成两条支路，经过激活函数 Sigmoid 模块后和 Position attention networks 的输入相乘，得到每个支路的输出，两个支路输出再相乘得到 Position attention networks 的输出 Output。

2.4. 损失函数 Wise-IoU

Wise-IoU 参数示意图如图 5 所示，Wise-IoU 共有三个版本，即 Wise-IoUv1、Wise-IoUv2、Wise-IoUv3，不同版本适用的场景各有不同，本文使用 Wise-IoUv1 (简写为 WIoUv1)损失函数替代原 YOLOv8 模型所使用的 IoU，以平衡不同质量图像的模型训练结果，获得更准确的检测结果。

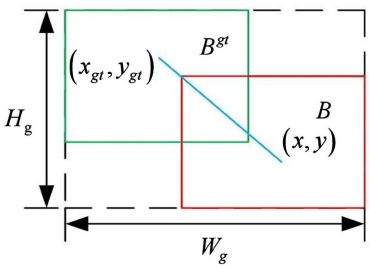


Figure 5. Schematic diagram of Wise-IoU
图 5. Wise-IoU 示意图

WIoUvl 的计算公式如下所示：

$$L_{WIoUv1} = R_{WIoU} L_{IoU} \tag{1}$$

$$R_{WIoU} = \exp\left(\frac{\left(x - x_{gt}\right)^2 + \left(y - y_{gt}\right)^2}{\left(W_g^2 + H_g^2\right)^*}\right) \tag{2}$$

其中， x 和 y 为锚框的中心点坐标， x_{gt} 和 y_{gt} 表示目标框的中心点坐标， W_g 和 H_g 表示最小包围框的宽和高， R_{WIoU} 为惩罚项。为了防止 R_{WIoU} 产生阻碍收敛的梯度，将 W_g 和 H_g 从计算图中分离(公式 2 中上标 * 表示此操作)。

3. 仿真实验与结果分析

3.1. 实验数据集

VisDrone 数据集是一个基于无人机视角拍摄的复杂交通场景数据集，如图 6 所示。该数据集包括 10 个类别：行人、人群、自行车、汽车、货车、卡车、三轮车、遮阳三轮车、公共汽车和摩托车，同时还包括了许多有用的场景信息，例如天气、地形和时间等。该数据集包括 6471 张训练集图片、548 张验证集图片、1610 张测试集图片。

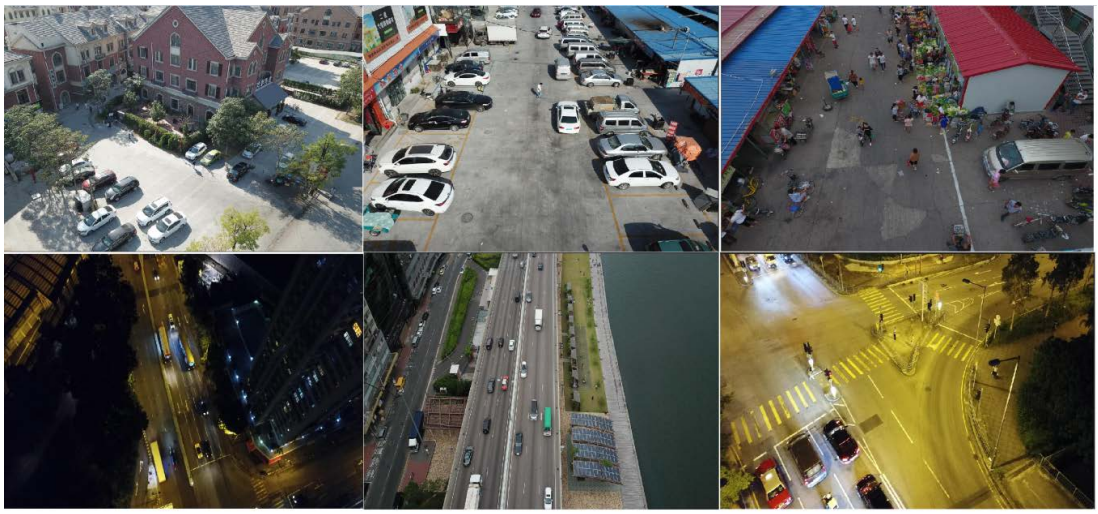


Figure 6. Dataset images
图 6. 数据集图片

3.2. 仿真实验

通过仿真实验来验证本文算法的性能，具体的实验参数如表 1 所示。

Table 1. Experimental parameters
表 1. 实验参数

参数名	参数值
训练集图片	6471
验证集图片	548
测试集图片	1610
epoch	100

续表

学习率	0.01
batchsize	16
优化器	SGD

本文所提算法的训练过程如图 7 所示，在训练过程中可以看到，随着跌到次数 epoch 的增加，分类损失 clc_loss、预测框损失 box_loss 和分布特征损失 dlf_loss 都在不断降低，而预测精度 precision、召回率 recall 和全类平均正确率 mAP 都在不断上升，最后达到收敛。在训练过程中，每训练一个 epoch 后都会获得一个参数模型，然后用该参数模型在验证集上进行验证，从图 7 的第二排前 3 个小图中可以看到：预测框损失 box_los 不断降低并收敛在 0.7 左右，分类损失 clc_loss 不断降低并收敛在 1.15 左右，分布特征损失 dlf_loss 不断降低也收敛在 1.15 左右。算法的精度指标：精度 precision 达到 47%左右，召回率 recall 达到 35%左右，mAP50 和 mAP50-95 分别达到 35%和 20%左右。

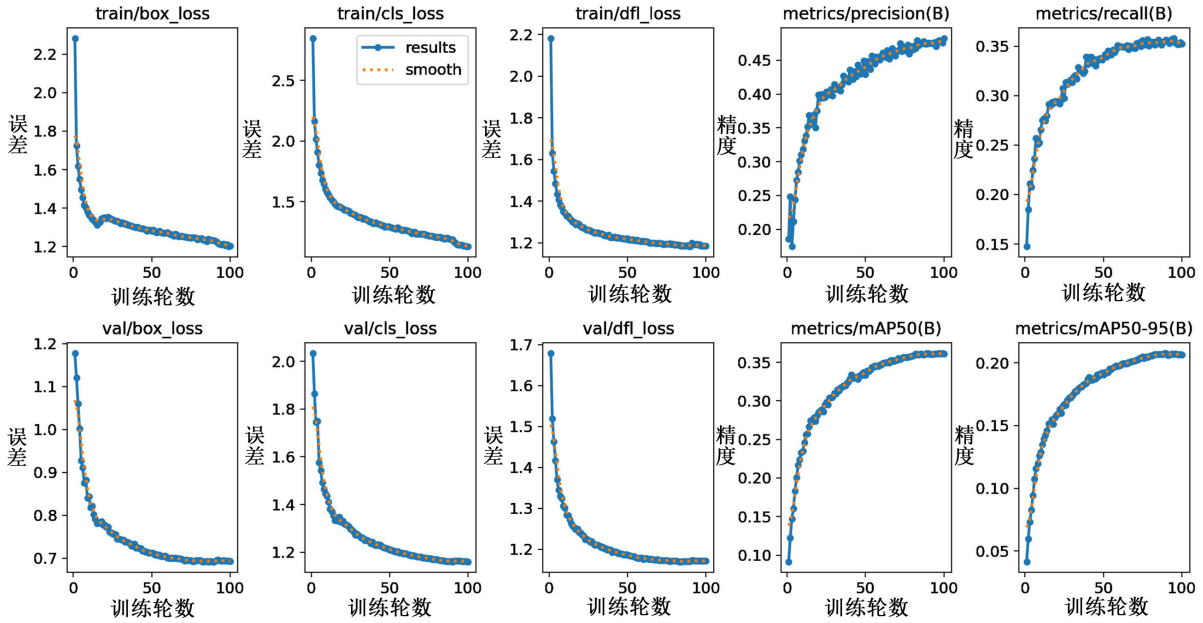
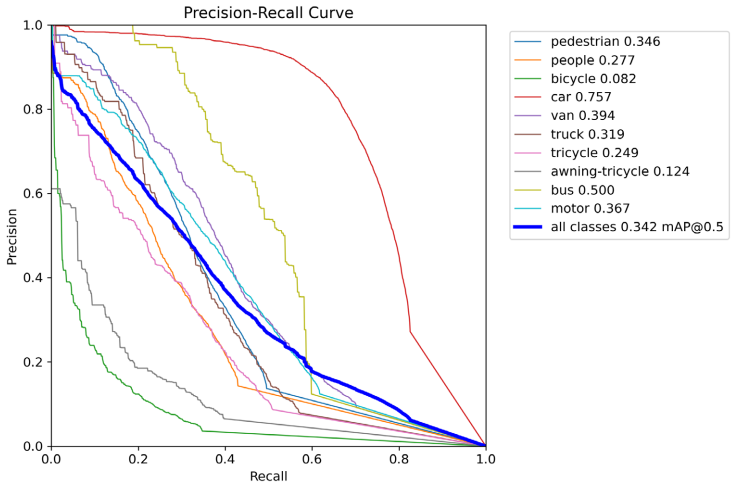


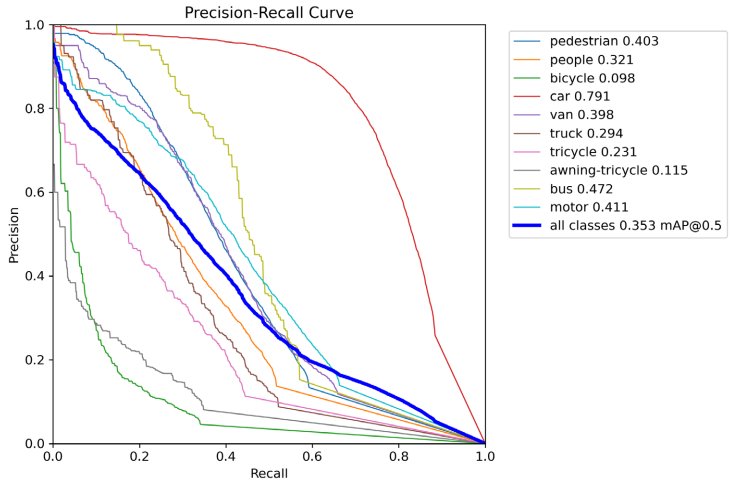
Figure 7. Training procedure of the proposed algorithm
图 7. 本文所提算法的训练过程

为了验证本文所提算法的性能，我们做了消融实验，在相同的实验条件和参数下，分别对比验证了 YOLOv8、ASF + YOLOv8 和 ASF + YOLOv8 + WIoUv1 的性能，实验结果如图 8 所示。从图 8 中可以看到，ASF 机制让 YOLOv8 算法的性能提高了 1.1% (mAP50：34.2%→35.3%)，而 Wise-IoU 机制让 ASF + YOLOv8 算法的性能提高了 0.9%(mAP50：35.3%→36.2%)，因此本文所提算法的性能相比 YOLOv8 算法提升了 2.0%。另外观察这 10 类不同目标的检测精度，本文算法与 YOLOv8 算法进行对比后发现，在 pedestrian：行人、people：人、bicycle：自行车这几类目标的检测精度提升较多，说明本文算法在密集型小目标的检测能力方面更为突出。

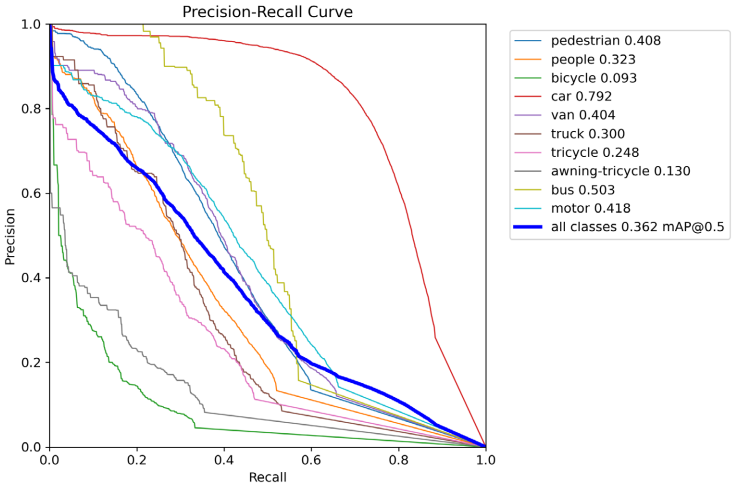
图 9 显示的是本文所提算法在测试集上的一些测试情况，从图 9 中可以看到本文所提算法能够准确地识别行人、人群、自行车、汽车、货车等不同种类的事物，而且在不同场景(白天、夜晚、公路、足球场等)下都能准确识别，对于一些比较密集的小目标群也能有较好的分辨和识别能力。图 9 右上角图片展示了本文算法对密集型小目标的检测能力，图 9 左下角图片展示了本文算法在夜景下的检测能力。



(a) YOLOv8



(b) ASF + YOLOv8



(c) ASF + YOLOv8 + WIoUv1

Figure 8. Algorithm performance comparison (mAP50)
图 8. 算法性能对比(mAP50)

Figure 9. Testing situation of the proposed algorithm
图 9. 本文所提算法的测试情况

4. 结论

本文提出了一种基于改进 YOLOv8 的目标检测算法 ASF-WIoU-YOLOv8，主要用于解决无人机航拍图像中的多目标检测问题。首先，在 YOLOv8 的基础架构上，加入一种注意尺度序列融合机制(Attentional Scale Sequence Fusion-ASF)，该机制能够对不同尺度的特征图进行融合，从而获得更好的图像特征，提取出更丰富、更准确的特征信息。然后，对损失函数进行改进，引入 Wise-IoU 机制，该机制通过自适应地调整权重系数提高目标检测的灵活性和鲁棒性，从而进一步提高算法的检测精度。实验结果表明，在 VisDrone 数据集上，本文所提算法比 YOLOv8 算法的平均精度 mAP50 提升了 2.0%，该算法在无人机航拍图像上具有更高的检测精度。

基金项目

省级科研平台项目资助(黔科合平台人才-CXTD[2021]008)。

参考文献

[1] Chai, X.H., Hu, Y., Lei, Y.L., *et al*. (2019) Review of UAV Intelligent Measurement and Control Technology. *Radio Engineering*, **49**, 855-860. https://doi.org/10.3969/j.issn.1003-3106.2019.10.003

[2] Surmann, H., Worst, R., Buschmann, T., *et al*. (2019) Integration of UAVs in Urban Search and Rescue Missions. 2019 *IEEE International Symposium on Safety, Security, and Rescue Robotics* (*SSRR*), Würzburg, 02-04 September 2019. https://doi.org/10.1109/SSRR.2019.8848940

[3] Girshick, R., Donahue, J., Darrell, T., *et al*. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014. https://doi.org/10.1109/CVPR.2014.81

[4] Girshick, R. (2015) Fast R-CNN. 2015 *IEEE International Conference on Computer Vision* (*ICCV*), Santiago, 07-13

December 2015. https://doi.org/10.1109/ICCV.2015.169

[5] Ren, S., He, K., Girshick, R., *et al*. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **39**, 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

[6] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y. and Berg, A.C. (2016) SSD: Single Shot MultiBox Detector. *Computer Vision—ECCV* 2016, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2

[7] Redmon, J., Divvala, S., Girshick, R., *et al*. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016. https://doi.org/10.1109/CVPR.2016.91

[8] Wang, H., *et al*. (2022) A UAV-Based Energy-Efficient and Real-Time Object Detection System with Multi-Source Image Fusion. *Journal of Circuits, Systems and Computers*, **31**, Article 2250166. https://doi.org/10.1142/S0218126622501663

[9] Fan, Y., Li, O. and Liu, G. (2022) An Object Detection Algorithm for Rotary-Wing UAV Based on AWin Transformer. *IEEE Access*, **10**, 13139-13150. https://doi.org/10.1109/ACCESS.2022.3147264

[10] 郭业才, 孙京东, Amitave, S. 基于 YOLOv5 改进的航拍图像目标检测算法[J/OL]. 系统仿真学报: 1-14. https://doi.org/10.16182/j.issn1004731x.joss.23-1564, 2024-04-09.

[11] Zhao, L.L. and Zhu, M.L. (2023) Ms-Yolov7: Yolov7 Based on Multi-Scale for Object Detection on UAV Aerial Photography. *Drones*, **7**, Article No. 188. https://doi.org/10.3390/drones7030188

[12] Tang, F., Yang, F. and Tian, X. (2023) Long-Distance Person Detection Based on Yolov7. *Electronics*, **12**, Article No. 1502. https://doi.org/10.3390/electronics12061502

[13] Kang, M., Ting, C.M., Ting, F., *et al*. (2023) ASF-YOLO: A Novel YOLO Model with Attentional Scale Sequence Fusion for Cell Instance Segmentation. *Image and Vision Computing*, **147**, Article 105057. https://doi.org/10.1016/j.imavis.2024.105057

[14] Tong, Z., Chen, Y., Xu, Z., *et al*. (2023) Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism. https://doi.org/10.48550/arXiv.2301.10051