

中图法分类号: TP751.1 文献标识码: A 文章编号: 1006-8961(2025)01-0282-15

论文引用格式: Wang X Y, Liang W H, Li J, Mu J H and Wang X Y. 2025. Fusion of bounding box Gaussian modeling and feature aggregation distribution for fine-grained recognition of remote sensing aircraft images. Journal of Image and Graphics, 30(01):0282-0296(王晓燕, 梁文辉, 李杰, 牟建宏, 王禧钰. 2025. 融合边界框高斯建模与特征聚合分发的遥感飞机细粒度识别. 中国图象图形学报, 30(01):0282-0296)[DOI: 10.11834/jig.230862]

融合边界框高斯建模与特征聚合分发的 遥感飞机细粒度识别

王晓燕¹, 梁文辉², 李杰^{3*}, 牟建宏², 王禧钰²

1. 北京物资学院统计与数据科学学院, 北京 101149; 2. 北京物资学院信息学院, 北京 101149;

3. 北京建筑大学机电与车辆工程学院, 北京 102616

摘要: 目的 遥感飞机影像由于目标尺寸差距大, 采集过程中受光照、遮挡等因素的影响, 导致不同型号飞机特征相似, 小目标检测效果不好, 类内无法实现细粒度区分。为了解决上述问题, 提出了一种融合边界框高斯建模与特征聚合分发的YOLOv5s(you only look once)遥感飞机细粒度识别算法。**方法** 首先, 将归一化高斯瓦瑟斯坦距离(normalized Gaussian Wasserstein distance, NWD)与交并比(intersection over union, IoU)及其衍生指标相结合, 并合理地设置两者的比例参数, 改进YOLOv5s位置损失的计算方式, 从而提高模型对小目标的敏感度。其次, 在YOLOv5s的Neck部分引入特征聚合分发模块(gather-and-distribute, GD), 在原网络“自顶向下、横向连接”的特征融合方式的基础上做到了跨层融合信息, 增强网络细粒度特征、全局特征提取能力, 提高了整体检测精度。为检验本算法在军用飞机上的细粒度和小目标识别优势, 使用遥感飞机细粒度数据集MAR20(military aircraft recognition 20)和遥感飞机小目标数据集CORS-ADD(complex optical remote-sensing aircraft detection dataset)进行实验。**结果** 实验结果显示: 对于数据集MAR20和CORS-ADD, 本文模型精确度分别达到了99.10%和95.36%, 与原YOLOv5s、YOLOv8s、Gold-YOLO和Faster R-CNN相比, 检测精度最佳。实验验证了模型在细粒度和小目标检测方面性能更加优秀, 在检测结果上与真实结果更加接近, 改进算法细粒度和小目标检测精度最佳。**结论** 实验结果表明, 本文算法在检测性能和模型精度上的表现优于上述4种目标检测算法, 模型具有良好的实用价值。

关键词: 目标检测; 改进YOLOv5s; 遥感飞机影像; 细粒度识别; 特征融合

Fusion of bounding box Gaussian modeling and feature aggregation distribution for fine-grained recognition of remote sensing aircraft images

Wang Xiaoyan¹, Liang Wenhui², Li Jie^{3*}, Mu Jianhong², Wang Xiyu²

1. School of Statistics and Data Science, Beijing Wuzi University, Beijing 101149, China;

2. School of Information, Beijing Wuzi University, Beijing 101149, China; 3. School of Mechanical, Electrical and Vehicle Engineering, Beijing University of Civil Engineering and Architecture, Beijing 102616, China

Abstract: Objective As a basic branch of computer vision, object detection plays an important role in subsequent tasks such as image segmentation and object tracking. It aims to find all the objects in the image and determine the location and category of the objects. It is used in industrial testing and has profound and extensive applications in aerospace, auto-

收稿日期: 2023-12-29; 修回日期: 2024-05-13; 预印本日期: 2024-05-20

* 通信作者: 李杰 lijie1@bucea.edu.cn

基金项目: 国家自然科学基金项目(51675494); 北京物资学院校内项目(2023XJKY14); 北京建筑大学金字塔人才培养工程项目(JDJQ20200308)

Supported by: National Natural Science Foundation of China(51675494)

mous driving, and other fields. Aircraft detection in remote sensing images is of great significance to both military and civilian fields such as air traffic control and battlefield dynamic monitoring. As a result of the large differences in object size in remote sensing aircraft images, the acquisition process is affected by factors such as lighting and occlusion, resulting in similar characteristics of different types of aircraft, poor detection of small objects, and the inability to achieve fine-grained distinction within categories. In object detection, the loss function is used to measure the difference between the model prediction and the actual object, which directly affects the performance and convergence speed of the model. Adjusting the model parameters so that the value of the loss function reaches the minimum value can improve the accuracy of the model in the test set. The loss function of YOLOv5 consists of position loss, category loss and confidence loss. YOLOv5 uses the intersection over union (IoU) and the derivative algorithm complete IoU by default, and provides IoU, generalized IoU, and distance IoU for replacement. However, for small object detection, especially with anchor box-based algorithms such as YOLOv5, the IoU series indicators cannot meet application needs well. Different types of remote sensing aircraft have fine-grained characteristics, which are reflected in subtle differences between classes, large differences within classes, and detail accuracy within classes. For fine-grained recognition tasks, extracting local information is crucial. The feature fusion module PANet used by YOLOv5s cannot achieve global feature fusion and is not conducive to extracting fine-grained features. To solve the above problems, this article proposes a model improvement algorithm based on YOLOv5s. **Method** In view of the shortcomings of IoU in small object detection based on YOLOv5, this article introduces Gaussian Wasserstein distance into the calculation of bounding box overlap to improve the detection performance of the network. Different from the IoU series of algorithms that calculate the similarity between different prediction boxes and real boxes based on the set of pixels contained in the bounding box, the Gaussian Wasserstein distance abandons the set, models the bounding box as a two-dimensional Gaussian distribution, and proposes a new metric called normalized Gaussian Wasserstein distance to calculate the similarity between frames, which fundamentally solves the problem of IoU in small object detection based on YOLOv5. In response to PANet's shortcomings in fine-grained detection, this article introduces the gather-and-distribute feature aggregation module in Gold-YOLO into YOLOv5s to enhance the YOLOv5s network's ability to extract fine-grained features through convolution and self-attention mechanisms. 1) The method combining Gaussian Wasserstein distance and traditional IoU is used to improve the loss function of YOLOv5s. 2) The gather-and-distribute feature aggregation module is introduced in the neck part of YOLOv5s to enhance the network's local feature extraction capabilities. Through the above two methods, the overall detection accuracy is improved. To test the advantages of this algorithm in fine-grained and small object recognition on military aircraft, this paper uses the remote sensing aircraft fine-grained classification dataset MAR20 and the remote sensing aircraft small object dataset CORS-ADD to conduct experiments. In the field of remote sensing military aircraft identification, different types of aircraft often have similar characteristics, resulting in different types of aircraft having similar characteristics, making it difficult to achieve intra-class identification. This article uses the open-source object detection remote sensing image dataset military aircraft recognition 20(MAR20) to achieve fine-grained recognition of remote sensing military aircraft. The dataset contains a total of 3 842 images, including 20 military aircraft models (SU-35, C-130, C-17, C-5, F-16, TU-160, E-3, B-52, P-3C, B-1B, E-8, TU-22, F-15, KC-135, F-22, FA-18, TU-95, KC-10, SU-34, SU-24). The CORS-ADD dataset is a complex optical remote sensing aircraft small object dataset that is manually annotated and constructed by the Space Optical Engineering Research Center of Harbin Institute of Technology. It contains a total of 7 337 images, including 32 285 aircraft instances, and the object size ranges from 4×4 pixels to 240×240 pixels. Different from the single data source of previous remote sensing datasets, the CORS-ADD dataset comes from satellite platforms such as Google Maps, WorldView-2, WorldView-3, Pleiades, Jilin-1, and IKONOS, covering airports, aircraft carriers, oceans, land, and other scenarios, as well as aircraft objects such as bombers, fighter jets, and early-warning aircraft at typical airports in China and the United States. **Result** To test the algorithm improvement effect of the two improved modules on remote sensing aircraft recognition based on YOLOv5s, this article compares the model performance of the original YOLOv5s with the introduction of normalized Gaussian Wasserstein distance (NWD) (r is the weight parameter used to adjust the ratio of IoU and NWD) and GD. The experimental result shows that the introduction of NWD and GD can improve the recognition accuracy to varying degrees, and the improvements are effective. When the ratio of IoU to NWD is 1:1, the recognition effect of the MAR20 dataset is the best; when the ratio of IoU to NWD is 1:9, the rec-

ognition effect of the CORS-ADD dataset is the best. Experimental results show the following: For the MAR20 dataset, compared with that of YOLOv5s, YOLOv8s, and Gold-YOLO, the mAP of improved YOLOv5s increased by 1.1%, 0.7% and 1.8% respectively; for the CORS-ADD dataset, mAP increased by 0.6%, 1.7%, and 3.9%, respectively. **Conclusion** An improved YOLOv5s network is proposed to solve the problems of large object size differences and high intra-class similarity in the process of remote sensing aircraft image recognition. On the basis of YOLOv5s, the loss function of YOLOv5s is improved by combining the Gaussian Wasserstein distance with the traditional IoU metric, which improves the detection effect of objects of different sizes, thereby improving the detection accuracy of the model. At the same time, to solve the problem of the characteristics of different types of aircraft being similar and the difficulty of distinguishing between sub-categories, this article uses the gather-and-distribute feature aggregation module in Gold-YOLO to enhance the ability of the YOLOv5s network to extract fine-grained features. A comparison shows that the improved YOLOv5s has a better model detection accuracy than that of YOLOv5s, YOLOv8s, Gold-YOLO, and Faster R-CNN. To improve the image processing speed of the model without reducing the accuracy of the model and to reduce the consumption of computing resources as much as possible to achieve lightweight deployment in the future, this article will consider using the C3_DSConv network to replace the C3 network of the YOLOv5s detection part to improve the model check speed and make it lightweight.

Key words: object detection; improved YOLOv5s; remote sensing aircraft imagery; fine-grained recognition; feature fusion

0 引言

目标检测(object detection)作为计算机视觉的一个基础分支,对后续的图像分割(Minaee 等, 2022)、目标跟踪(Yilmaz 等, 2006)等任务起着重要作用,其任务是找出图像中所有感兴趣的目标,并且确定目标的位置和类别。遥感图像目标检测(刘扬等, 2015)任务是遥感领域内的一个研究热点内容,在情报侦查、灾害救援和工业应用等领域发挥着重要作用。其中可见光遥感图像目标检测关注的对象包括飞机、机场跑道线、舰船、车辆和农作物等。

在可见光遥感影像检测中,传统机器学习检测主要为人工特征提取和分类器分类。即先进行人工特征提取(Sugiarto 等, 2017),再使用 Adaboost 和支持向量机(support vector machine, SVM)(Pang 等, 2014)等算法搭建对应的分类器。

随着计算机计算能力不断提升和大数据时代的来临,深层神经网络的出现使得遥感目标检测的发展进入了新阶段,基于深度学习的遥感影像分类方法(刘钦龙 等, 2009)开始兴起,特别是基于卷积神经网络(convolutional neural network, CNN)(黄洁等, 2017)的识别方法表现出了良好的性能。然而,遥感图像的目标分布往往不固定且目标尺寸差距大,在这种情况下, CNN 一次处理整个图像的特性使其效率低下,不利于实际应用。

近年来,基于深度学习的遥感影像目标检测算法可以大致分为两类:基于区域的分类算法和基于回归的分类算法。其中,基于区域的分类算法又称为两阶段算法,即先在目标图像上形成候选框,再在候选框中进行图像的识别与分类。绝大多数两阶段的目标检测算法都属于 R-CNN(region with CNN features)系列(李鑫 等, 2021), R-CNN 在 CNN 的基础上引入了区域建议网络(region proposal network, RPN)进行候选区域生成,其优点是在精度和检测效果上具有相对优势,但往往涉及复杂的模型,且运行速度较慢。因此研究人员将重点放在了速度和精度兼顾的基于回归的分类算法上,也可以称为单阶段算法或端到端的算法,如 YOLO(you only look once)系列(化嫣然 等, 2020)、SSD(single shot multibox detector)(史文旭 等, 2020)等。其中以 YOLO 系列算法为典型代表,在原 YOLO 网络上做出改进让其更适应遥感目标检测的算法也层出不穷。

Wang 等人(2023)使用 YOLOv2 针对识别速度不高优化了 YOLOv2 的损失函数,并且在主干网络中删除 Dropout 层,改为使用 1×1 卷积。针对识别精度不高,改进了重叠框置信度策略,增加线性衰减函数和高斯衰减函数。Zhao 和 Li(2020)基于 YOLOv3 开发了一种新的聚类方法来估计预测边界框的初始宽度和高度,提升了模型检测精度。Gai 等人(2023)提出了一种改进的 YOLOv4 算法来检测樱桃果实,首先,使用更符合樱桃果实形状的圆形标记

框进行训练。其次,在YOLOv4骨干网络CSPDarknet53网络的基础上结合了DenseNet层间密度。基于改进后的YOLOv4模型,特征提取得到加强,网络结构得到深化,检测速度得到提高。Zhu等人(2021)在YOLOv5的基础上,增加了一个预测头来检测不同尺度的物体,并且用变形预测头(Transformer prediction head,TPH)替换原有的预测头,同时整合了卷积块注意力模型(convolutional block attention module,CBAM),在无人机拍摄的遥感影像中具有良好的性能。

综上,研究人员在YOLO系列算法的改进上主要有以下几方面:改进或替换YOLO主干网络结构、增加特征注意力机制以及优化损失函数等。

细粒度识别(fine-grained recognition)(Šulc和Matas,2017)是计算机视觉领域中的一个任务,旨在区分属于同一大类的物体或图像,但这些物体在细节上具有微小但显著的差异。与传统的目标分类任务不同,细粒度识别需要模型具备更高的分辨能力,能够区分同一类别下的不同子类(阳治民和宋威,2023)。这对于模型的性能和特征表示能力提出了更高的要求。细粒度识别的物体差异可能非常微小,需要模型学习捕捉这些微妙但重要的特征。同时,细粒度类别内部的差异可能比类别间的差异更大,需要模型能够平衡处理。通常采用深度学习模型、特征增强方法和迁移学习等技术来提高模型的识别性能。

本文针对遥感影像识别过程出现的以下问题进行改进:1)由于遥感影像采集过程的特殊性,即相对较大的视野、高海拔拍摄等特性,会出现成像目标尺寸差距大、部分目标过小不易识别(袁翔等,2023)的问题。本文使用归一化高斯瓦瑟斯坦距离(normalized Gaussian Wasserstein distance, NWD)(Wang等,2022)与传统的交并比(intersection over union, IoU)(Zheng等,2020)度量相结合的方式,改进YOLOv5s的损失函数。2)针对不同种类飞机特征相似、子类间难以区分的问题,本文利用Gold-YOLO(Wang等,2023)中的gather-and-distribute(GD)特征聚合模块,增强YOLOv5s网络提取细粒度特征的能力。

1 YOLOv5s模型改进

1.1 YOLOv5s网络结构简介

YOLO系列算法是Joseph Redmon为首的团队提出的一种目标检测算法(Redmon等,2016),主要创新点在于将目标检测看做回归问题,从空间上定位边界框并预测该框的类别概率。YOLO使用单个神经网络,在一次评估中直接从完整图像上预测边界框和类别概率。由于整个检测流程仅用一个网络,所以可以直接对检测性能进行端到端的优化。现阶段YOLO系列已经更新到YOLOv8,演变过程如图1所示。

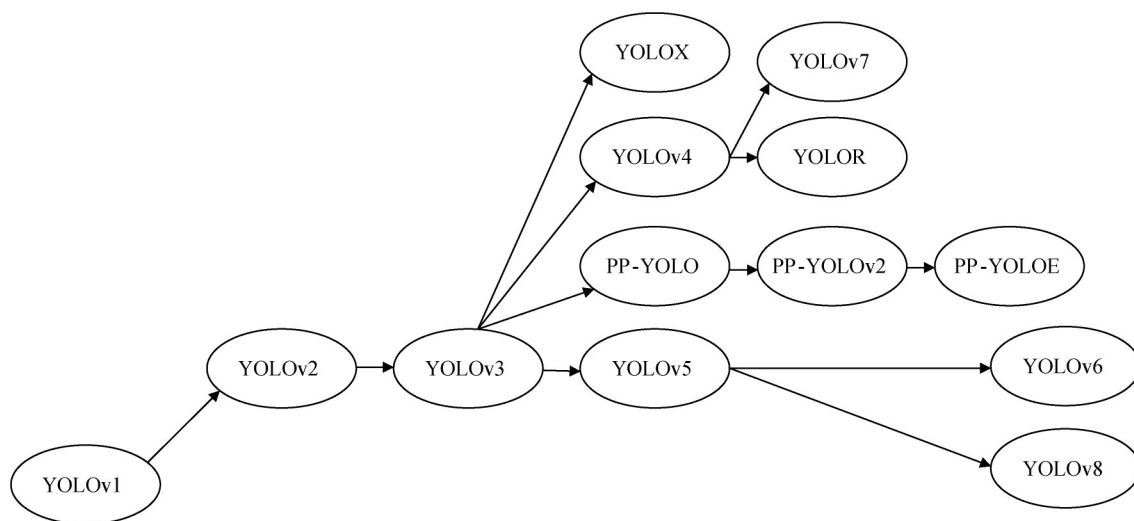


图1 YOLO版本演变

Fig. 1 Evolution of YOLO versions

虽然目前YOLO系列算法已经更新到v8,但是YOLOv5以其更易读懂、调参的源代码、更低的功耗

以便部署到移动设备以及训练快速等优点,在目标检测领域依旧是热门算法,在YOLOv5的基础上进

行改进的算法层出不穷, YOLOv5s 网络结构如图 2 所示。

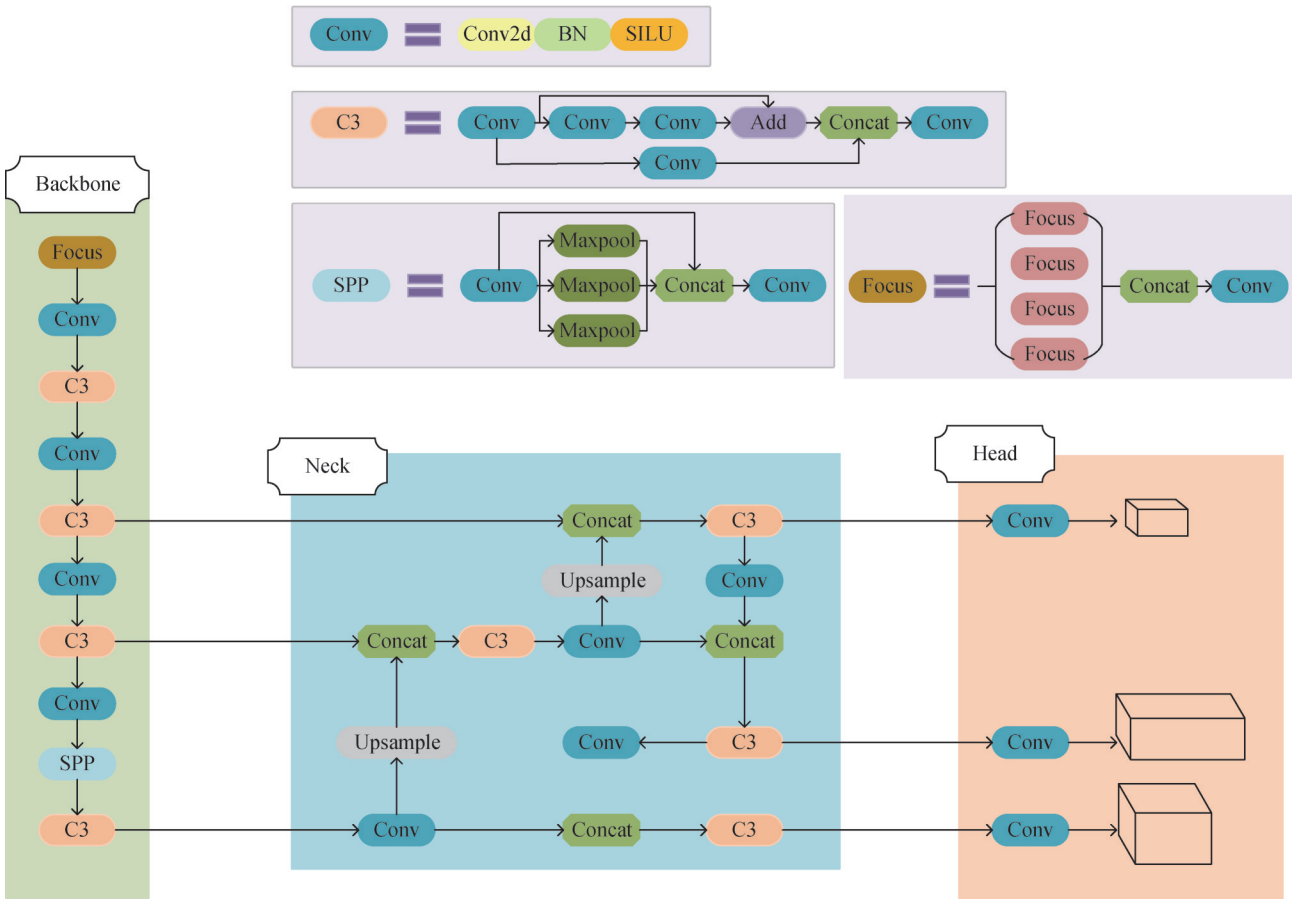


图2 YOLOv5s网络结构
Fig. 2 Architecture of YOLOv5s network

1.2 YOLOv5s 网络结构改进

1.2.1 YOLOv5s 损失函数改进

在目标检测中,损失函数用于衡量模型预测与实际目标之间的差异,直接影响着模型的性能和收敛速度。YOLOv5的损失函数由位置损失(localization loss)、类别损失(class loss)和置信度损失(confidence loss)构成。其中,位置损失的计算使用到IoU, YOLOv5 默认使用IoU的衍生算法CIoU(complete intersection over union), 并且提供IoU、GIoU(generalized intersection over union)和DIoU(distance intersection over union)用于替换。

但是,对于遥感小目标检测,特别是YOLOv5这种基于锚框的算法(潘晓英等,2023),IoU系列指标不能很好地满足应用要求。本文将高斯瓦瑟斯坦距离(Gaussian Wasserstein distance)引入边界框重叠度计算中,提高了网络的检测性能,改进原理如下所述。

1)IoU 系列指标。IoU 是衡量目标检测算法性能的重要指标之一,是通过计算预测框A和真实框B之间的重叠程度来衡量的。具体来讲,就是A与B像素点的交集与并集之比,具体为

$$IoU_{AB} = \frac{Intersection(A, B)}{Union(A, B)} \tag{1}$$

式中, IoU 取值范围为: $0 \leq IoU \leq 1$, 0表示完全没有重叠, 1表示完全重叠。IoU越大,表示预测框A和真实框B重叠度越高,即目标检测的准确性越高。在实际应用中,预测框A和真实框B之间的位置关系有许多种复杂情况,如图3所示。

IoU 只能解决其中最简单的情况(图3(a)), 此时损失函数计算为

$$Loss_{IoU} = 1 - IoU \tag{2}$$

其余4种情况出现时,即预测框与真实框没有交集或完全重合时,计算出来的IoU都为0,损失都为1, IoU失效,无法比较预测框A与a的优劣。

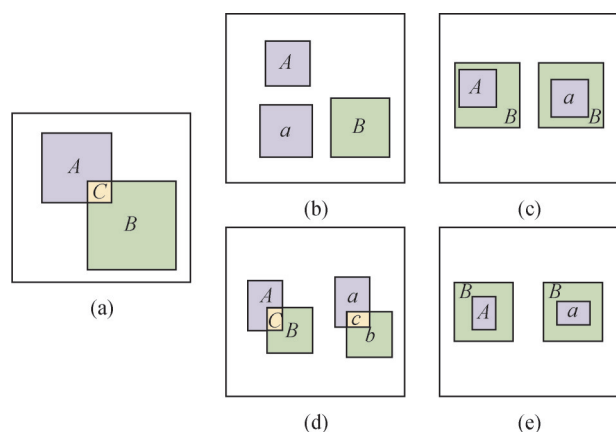


图3 预测框与真实框之间的多种位置关系

Fig. 3 Various positional relationships between predicted boxes and real boxes

针对图3(b),即当多个预测框与真实框都没有交集时,GIoU引入了“最小闭包区”的概念,即能将预测框A、a和真实框B包裹住的最小矩形框,如图4所示。

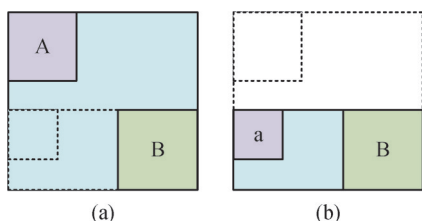


图4 最小闭包区

Fig. 4 Minimum closure area

GIoU 及其损失函数计算为

$$GIoU = IoU - \frac{|A_c - U|}{A_c} \quad (3)$$

$$Loss_{GIoU} = 1 - GIoU \quad (4)$$

式中, A_c 为最小闭包区, U 为预测框和真实框的并集。由式(3)可得,GIoU认为直观上与真实框距离更近的预测框a效果更好。

针对图3(c),即两个预测框都是真实框的子集且大小相同时,IoU和GIoU计算结果都是一样的,但实际上认为预测框与真实框的中心点距离越近效果越好,即预测框a效果优于预测框A。此时引入DIoU可以解决,具体为

$$DIoU = IoU - \frac{\rho^2(b, b^*)}{c^2} \quad (5)$$

式中, b 和 b^* 分别代表预测框和真实框的中心点, ρ 代表的是计算两个中心点间的欧氏距离, c 代表的是

能够同时包含预测框和真实框的最小闭包区域的对角线距离,如图5所示。

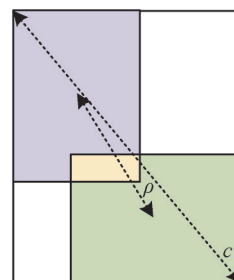


图5 中心点距离

Fig. 5 Center point distance

DIoU 损失函数计算为

$$Loss_{DIoU} = 1 - DIoU \quad (6)$$

针对图3(d)(e),当两个预测框A和a大小相等,与真实框的中心点都重合但是长宽比不同,计算出来的DIoU、GIoU、IoU都一样,此时CIoU在DIoU的基础上增加了对长宽比的考虑,认为预测框的长宽比越接近越好,具体为

$$Loss_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^*)}{c^2 + \alpha v} \quad (7)$$

式中, α 是权重函数, v 用来度量长宽比的相似性。

对小目标检测而言,“预测框与真实框没有重叠或者完全重叠”是非常常见的。虽然CIoU和DIoU可以处理上述情况,但由于它们都基于IoU,对微小物体的位置偏差很敏感,特别是在YOLOv5这种基于锚点的检测器中。

2)高斯瓦瑟斯坦距离。瓦瑟斯坦距离(Wasserstein distance)也称为地球移动距离(earth mover's distance, EMD),是一种用于测量两个概率分布之间差异的指标,在运动规划、图像识别和信号处理等领域有广泛研究。其基本思想是将一个概率分布变换到另一个概率分布所需要的最小代价(两个分布对应观测值之间的距离乘以权重)。高斯瓦瑟斯坦距离是瓦瑟斯坦距离的一种变体,其使用高斯核函数来平滑分布并计算距离。

若将其应用到目标检测中,首先需要将边界框 $R = (c_x, c_y, w, h)$ 转换为二维高斯分布 $N(\mu, \Sigma)$,具体为

$$\mu = \begin{bmatrix} c_x \\ c_y \end{bmatrix} \quad (8)$$

$$\Sigma = \begin{bmatrix} \frac{w^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{bmatrix} \quad (9)$$

式中, c_x 和 c_y 分别表示边界框中心点的横纵坐标, w 和 h 分别表示边界框的宽度和高度, μ 表示二维高斯分布的均值向量, Σ 表示其协方差矩阵。

对于由边界框 A 和边界框 B 建模的高斯分布 N_a 和 N_b , 其二阶瓦瑟斯坦距离 (Wasserstein-2 distance) 为

$$W_2^2(N_a, N_b) = \left\| \left(\begin{bmatrix} cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2} \end{bmatrix}^T, \begin{bmatrix} cx_b, cy_b, \frac{w_b}{2}, \frac{h_b}{2} \end{bmatrix}^T \right) \right\|_2^2 \quad (10)$$

式中, 边界框 A 和 B 分别为

$$A = (cx_a, cy_a, w_a, h_a) \quad (11)$$

$$B = (cx_b, cy_b, w_b, h_b) \quad (12)$$

式(10)是距离度量, 不能直接作为相似性度量 (例如 IoU 是 0~1 的值)。因此, 利用它的指数形式归一化, 得到了新的度量, 称为归一化高斯瓦瑟斯坦距离 (NWD), 具体为

$$NWD(N_a, N_b) = \exp \left(- \frac{\sqrt{W_2^2(N_a, N_b)}}{C} \right) \quad (13)$$

式中, C 是与数据集密切相关的常数。

综上所述, 与 IoU 系列指标基于边界框所含像素点的集合来计算不同预测框和真实框之间的相似性不同, 高斯瓦瑟斯坦距离摒弃了集合的思想, 将边界框建模为二维高斯分布, 并提出了一种新度量 NWD 来计算框间的相似性, 从根本上解决了 IoU 系列指标在基于 YOLOv5 的小目标检测中对微小物体位置偏差敏感度的问题。

IoU 在一定程度上反映了目标定位的准确性, 而 NWD 能够弥补 IoU 在小目标检测下的不足。在遥感飞机数据集中, 目标尺寸差异大, 并非全部是小目标。故本文并未直接用 NWD 代替 IoU 系列指标进行计算, 而是在计算 YOLOv5s 边界框位置损失 (l_{box}) 时, 将二者以不同比例加权计算, 更好地综合利用 IoU 和 NWD, 从而达到对飞机目标更好的识别精度。

具体来讲, 本文在后续的实验中设置一个参数 r ($0 \leq r \leq 1$), 用来调节二者所占的权重, 得到边界框位置损失 l_{box} 的计算方式, 具体为

$$l_{box} = (1 - r) \times (1 - NWD) + r \times (1 - CIoU) \quad (14)$$

理论上, 当数据集中小目标含量越多, r 取值越大。

1.2.2 YOLOv5s 引入特征聚合模块

不同型号的遥感飞机具有细粒度特征, 体现在类间差异细微、类内差异较大、类别精度细致。对于细粒度识别任务, 提取局部信息是至关重要的。本文在 YOLOv5s 中引入 Gold-YOLO 中的 gather-and-distribute (GD) 特征聚合模块, 通过卷积和自注意机制增强 YOLOv5s 网络提取细粒度特征的能力。

YOLOv5s 在 Neck 部分使用 PANet (path aggregation network) (Liu 等, 2018) 来进行特征融合操作, PANet 是特征金字塔网络 (feature pyramid network, FPN) (Lin 等, 2017) 的变体, 使用自顶向下和自底向上的双向融合骨干网络生成高分辨率的特征图, 使得网络能在多个尺度上同时感知目标。

然而, 这种特征融合方法有一个缺点, 即无法跨层整合信息, 使得信息融合的整体效果不好。

GD 模块放弃了原有的“逐层递归”方法, 构建了一种全新的特征收集与分配方式, 通过使用统一的模块对各层信息进行收集和融合, 再将其分发到不同层次。

在具体实施过程中, 以上过程对应着网络的 3 个模块: 特征对齐模块 (feature alignment module, FAM)、信息融合模块 (information fusion module, IFM) 与信息注入模块 (information injection module, Inject)。FAM 收集并对齐不同层次的特征, 由 IFM 将对齐的特征进行融合生成全局特征信息, 再由 Inject 将特征分发到各个层级, 并由注意力机制进行注入, GD 为了增强对不同尺寸目标的识别能力, 提供了低阶段采集一分发 (Low-GD) 和高阶段采集一分发 (High-GD) 两个分支。

1) Low-GD 内部结构。如图 6(a) 所示, B2、B3、B4、B5 代表 YOLOv5s 在 backbone 层输出的特征图。低阶特征对齐 (Low-FAM) 将特征图大小统一为 B4 的尺寸, 分别对 B5 层做双线性上采样插值 (bilinear upsampling) (Shelhamer 等, 2017), 对 B3、B2 层进行平均池化 (average pooling), 最后通过 Concat 进行特征图合并。低阶信息融合 (Low-IFM) 将上层 channel 的输出经过 1 个卷积 (Conv) 层、多个 RepConv 层和 1 个卷积层提取信息, 最后在 Split 层切片, 得到 Inject_P4 和 Inject_P3, 即后续 Inject 层的多层融

合特征输入。特征分发(Inject)将当前尺度下的特征与 Low-IFM 层的全局特征进行进一步提取和融合。

2) High-GD 内部结构。如图 6(b) 所示, High-GD 与 Low-GD 的网络结构大体上类似。高阶特征对齐(High-FAM)通过平均池化将 Low-GD 的 Inject 层输出 P3 和 P4 降维到 P5(特征图 B5)大小,再通过 Concat 进行特征图合并。高阶信息融合(High-IFM)将 High-FAM 得到的特征在由 1 个多头注意力(multi-head attention)模块(Zhang 和 Gao, 2020)、1 个前馈网络(feed-forward network)模块(Oh 和 Park, 2011)和残差连接(He 等, 2016)组成的 Trans-

former 模块进行融合后,在 Split 层切片,得到 Inject_N4 和 Inject_N5。High-GD 和 Low-GD 的 Inject(特征分发)模块是完全相同的。

综上所述,本文使用的 GD 特征聚合模块相比于原 YOLOv5s 的 Neck 部分,最大的改进就是使用 Low-GD 代替 PANet 的上采样融合阶段,用 High-GD 代替 PANet 的下采样融合阶段。通过统一的模块对各个尺度的信息进行采集和融合,并将融合的特征分发到不同的层,既避免了传统 FPN 结构固有的信息丢失,也在不显著增加延迟的情况下增强了 Neck 部分的信息融合能力,更有效地利用了对 backbone 提取到的特征。

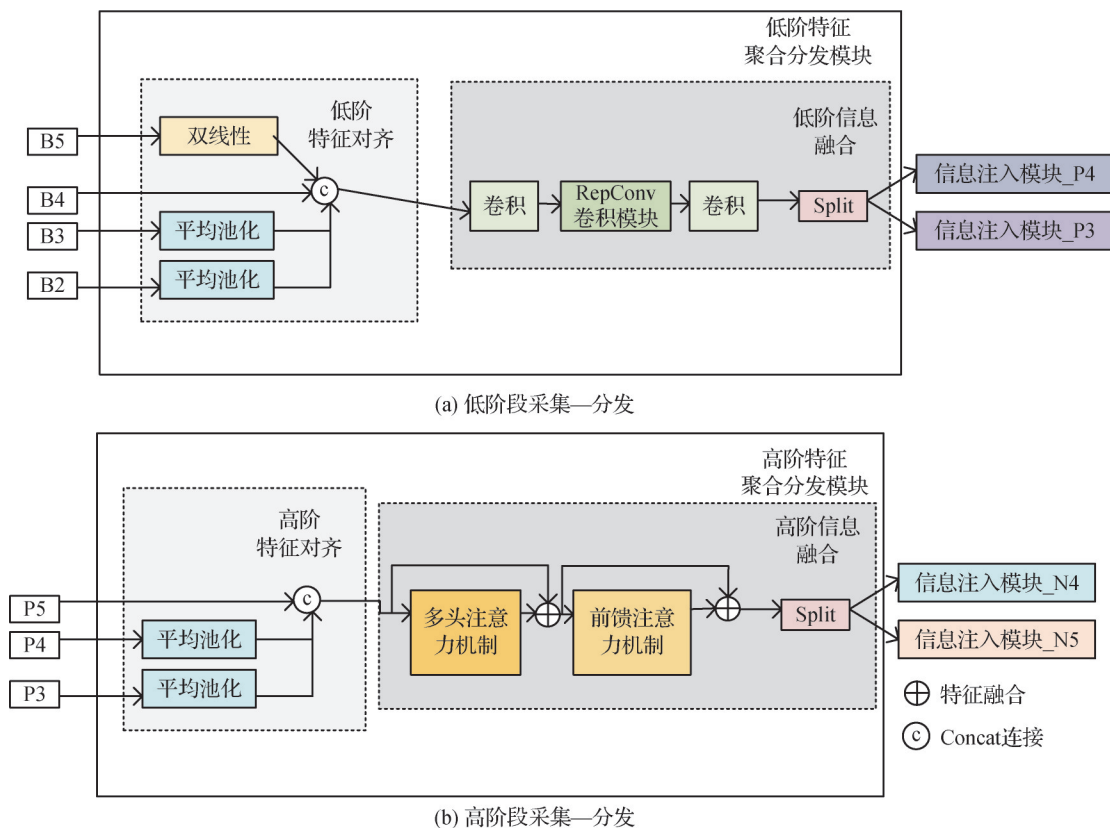


Fig. 6 Two branches of the GD module ((a) Low-GD; (b) High-GD)

2 实验与分析

2.1 研究对象与数据集

2.1.1 MAR20

在遥感军用飞机识别领域,不同型号的飞机往往具有相似的特点,导致不同型号的飞机具有相似的特征,不易实现类内识别。本文使用开源的目标

检测遥感影像数据集 MAR20(military aircraft recognition 20),实现遥感军用飞机细粒度识别。该数据集共计 3 842 幅图像,包括 20 种军用飞机型号(SU-35、C-130、C-17、C-5、F-16、TU-160、E-3、B-52、P-3C、B-1B、E-8、TU-22、F-15、KC-135、F-22、FA-18、TU-95、KC-10、SU-34、SU-24),简称分别为 A1—A20,样例如图 7 所示。

MAR20 共包含 20 种飞机型号的 22 341 个实例,

实例在数据集中的数目分布情况如表 1 所示。



图7 MAR20数据集中20种飞机型号样例
Fig. 7 Examples of 20 aircraft models in the MAR20 dataset

表 1 MAR20 各型号飞机实例数目分布情况
Table 1 The distribution of the number of MAR20 aircraft instances of each type

飞机型号	实例个数	飞机型号	实例个数
A1	1 646	A11	507
A2	1 729	A12	702
A3	1 176	A13	1 652
A4	642	A14	1 778
A5	1 262	A15	618
A6	441	A16	2 632
A7	680	A17	1 397
A8	944	A18	308
A9	1 086	A19	1 236
A10	924	A20	981

2. 1. 2 CORS-ADD

CORS-ADD (complex optical remote-sensing aircraft detection dataset)数据集(Shi 等, 2023)是哈尔滨工业大学空间光学工程研究中心手动标注并构建的复杂光学遥感飞机小目标数据集,共包含 7 337 幅图像,包含飞机实例 32 285 个,目标的大小范围从 4 × 4 像素到 240 × 240 像素不等。不同于以往遥感数据集单一的数据来源,CORS-ADD 数据集的数据来源于 Google Map、WorldView-2、WorldView-3、Ple-

iades、Jilin-1、IKONOS 等卫星平台,涵盖机场、航空母舰、海洋和陆地等多种场景,覆盖中国和美国等典型机场中的轰炸机、战斗机和预警机等飞机目标,样例如图 8 所示。



图8 CORS-ADD数据集飞机样例
Fig. 8 Examples of aircraft models in the CORS-ADD dataset

本次实验将数据集按照训练集、验证集和测试集随机划分,相互之间无重复图像,保证实验结果的客观性。训练集、验证集和测试集所用图像分别为 3 457 幅、192 幅和 193 幅,在数据集中分别占 90%、5%和 5%。

2. 2 环境配置

实验采用 Linux 操作系统,深度学习 PyTorch 1. 10. 0 框架, cuda 版本 11. 3。以 NVIDIA GeForce RTX 3090 GPU 作为硬件支持,显存 12 GB,实验使用的开发环境为 Pycharm,语言为 Python 3. 8。

2. 3 目标检测算法评估指标

为了评估改进算法的性能,本文选用的算法精度评价指标为平均精度均值(mean average precision, mAP),精度 AP 表示准确率(precision, P)和召回率(recall, R)曲线下的面积,mAP 由各类别 AP 的平均值得出,mAP 越大表示模型精度越高(石争浩等,2023)。

2. 4 消融实验与对比实验

2. 4. 1 模型精度

为了测试基于 YOLOv5s 的遥感飞机识别,两个

改进模块独立改进效果及最终改进 YOLOv5s 效果, 本文分别在 MAR20 和 CORS-ADD 两个数据集上进行实验验证。实验在相同硬件资源下进行, 均设置学习率为 0.001, 批量大小为 32, 经过 100 次迭代训练, 网络最终收敛。

本文比较了原 YOLOv5s 与分别引入 NWD 和 GD 后的模型性能, 实验结果见表 2。表 2 中, r 为上文提到的用于调节 IoU 与 NWD 比例的权重参数, 为了更全面地评估所用参数 r 的影响, 在此进行敏感性分析。对遥感飞机小目标数据集 CORS-ADD 而言, 当 IoU 所占比例下降、NWD 所占比例上升时, 识别精度随之提升, 当 IoU 与 NWD 比例为 1:9 时, 识别效果最好, 证明 NWD 的引入对小目标识别有效。实验结果表明, NWD 和 GD 的引入均能够使识别精度有不同程度的提升, 单模块改进均有效。

表 2 消融实验 Table 2 Ablation experiment		
方法	mAP@0.5/%	
	MAR20	CORS-ADD
YOLOv5s	96.90	94.70
引入 NWD($r = 0.1$)	98.20	95.15
引入 NWD($r = 0.5$)	99.00	95.06
引入 NWD($r = 0.9$)	98.48	94.75
引入 GD	98.20	95.29
改进 YOLOv5s	99.10	95.36

注: 加粗字体表示各列最优结果。

基于消融实验的最好结果及其参数设置, 本文在 YOLOv5s 中同时引入 NWD 和 GD 模块, 分别在 MAR20 和 CORS-ADD 两个数据集上将改进 YOLOv5s 算法与原 YOLOv5s、YOLOv8s、Gold-YOLO 和 Faster R-CNN 做对比实验, 实验结果如表 3 所示。实验表明本文模型对于数据集 MAR20 和 CORS-ADD 具有良好的识别精度。

对于数据集 MAR20, 改进 YOLOv5s 的 mAP 达到了 99.10%, YOLOv5s、YOLOv8s、Gold-YOLO 和 Faster R-CNN 分别为 96.90%、98.30%、97.20% 和 98.70%, mAP 分别提高了 2.20%、0.80%、1.90% 和 0.40%。

对于数据集 CORS-ADD, 改进 YOLOv5s 的 mAP 达到了 95.36%, YOLOv5s、YOLOv8s、Gold-YOLO 和

Faster R-CNN 分别为 94.70%、93.61%、91.40% 和 89.10%, mAP 分别提高了 0.66%、1.75%、3.96% 和 6.26%。

表 3 对比实验 Table 3 Comparative experiment		
方法	mAP@0.5/%	
	MAR20	CORS-ADD
YOLOv5s	96.90	94.70
YOLOv8s	98.30	93.61
Gold-YOLO	97.20	91.40
Faster R-CNN	98.70	89.10
改进 YOLOv5s	99.10	95.36

注: 加粗字体表示各列最优结果。

2.4.2 模型细粒度识别效果

图 9—图 11 分别为 YOLOv5s、YOLOv8s 和改进 YOLOv5s 生成的混淆矩阵的可视化图像, 其对角线上的数值表示模型正确分类的样本比例, 而非对角线上的数值则表示模型错误分类的样本比例, 理想的混淆矩阵应为一个对角矩阵。

本文实验的混淆矩阵是 21×21 的矩阵, 每一列代表一种真实类别, 每一行代表一种预测类别, 每一个元素的真实类别为行对应的类别, 预测类别为列对应的类别。例如, 图 9 中第 1 列第 19 行的值“0.26”表示有 26% 的 A1 被误分类为 A19, 而第 21 列第 20 行的值“0.04”表示有 4% 的背景被误分类为 A20。

如图 9 所示, YOLOv5s 有 6 种类别的分类正确率超过了 90%, 部分类别如 A11、A15、A18 的分类成功率分别为 45%、49%、43%, 均在 50% 以下, 混淆矩阵非对角线元素多, 分类效果差。如图 10 所示, YOLOv8s 的效果较 YOLOv5s 有较大提升, 14 种类别的分类正确率超过了 90%, 但混淆矩阵非对角线元素较多且分散, 可见 YOLOv8s 对部分类别的飞机不能很好分辨。如图 11 所示, 本文改进的 YOLOv5s 分类正确率超过 90% 的有 18 种, 5 种类别的分类正确率达到了 100%, 混淆矩阵非对角线元素较 YOLOv8s 数量下降明显, 分类效果更好。

通过图 9—图 11 的对比, 可以看出改进 YOLOv5s 算法生成的混淆矩阵最接近对角矩阵, 非对角线元素较少且元素值较小。实验结果表明, 对

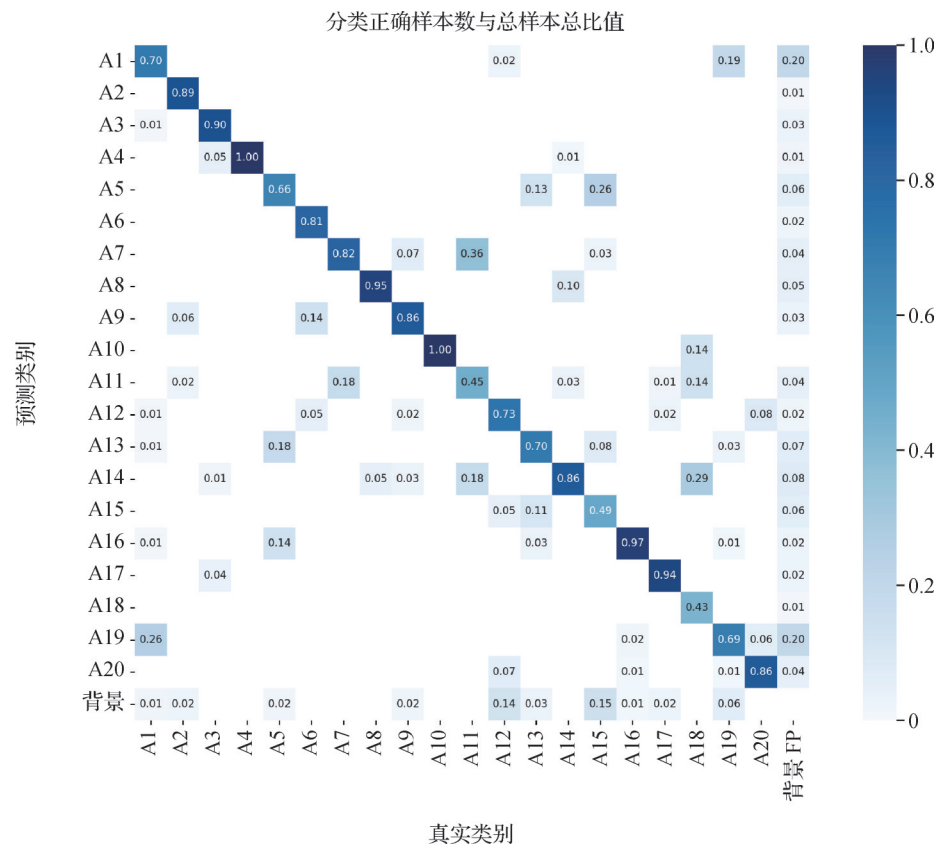


图9 YOLOv5s的混淆矩阵可视化
Fig. 9 Visualization of confusion matrix of YOLOv5s

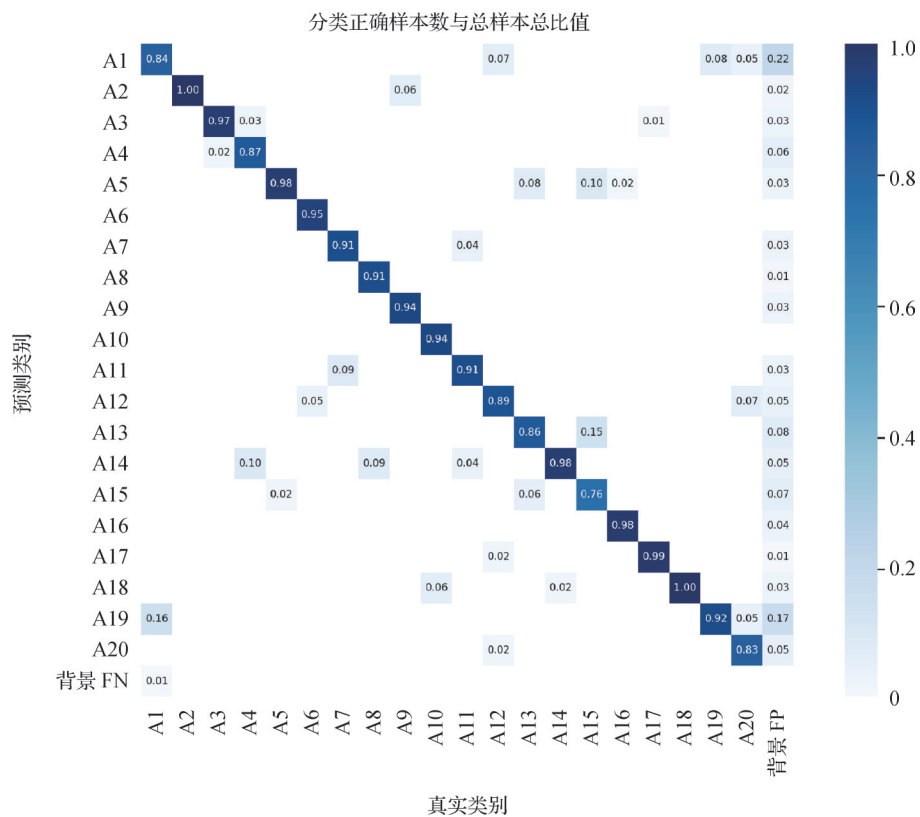


图10 YOLOv8s的混淆矩阵可视化
Fig. 10 Visualization of confusion matrix of YOLOv8s

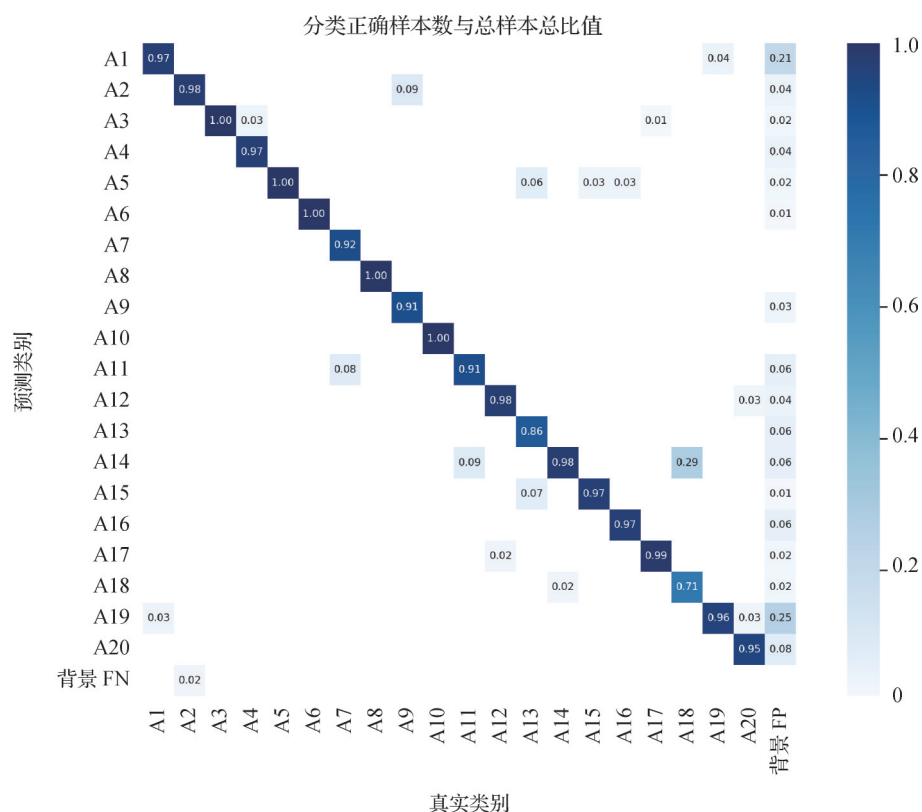


图11 改进YOLOv5s的混淆矩阵可视化

Fig. 11 Visualization of confusion matrix of improved YOLOv5s

于MAR20中特征相似的飞机,改进YOLOv5s在本文3种算法中具有最好的细粒度识别性能。

2.5 识别效果样例

实际的识别效果对比如图12—图14所示,图中方形检测框上的标注分别为类别名和其置信度(置信度在目标检测中指的是模型对预测结果的自信程度)。

图12(a)和图13(a)显示YOLOv5s和YOLOv8s

对A13有错检情况,分别将A13错检为A9和A15。图12(b)和图13(b)显示YOLOv5s和YOLOv8s对A5分别有置信度不高、漏检情况。图12(c)和图13(c)显示,在目标较小的情况下,YOLOv5s和YOLOv8s均不能对A1、A19进行很好的区分,YOLOv5s将部分A19误分类为A1,YOLOv8s将A1误分类为A16。

相比之下,图14显示上述问题均没有在改进

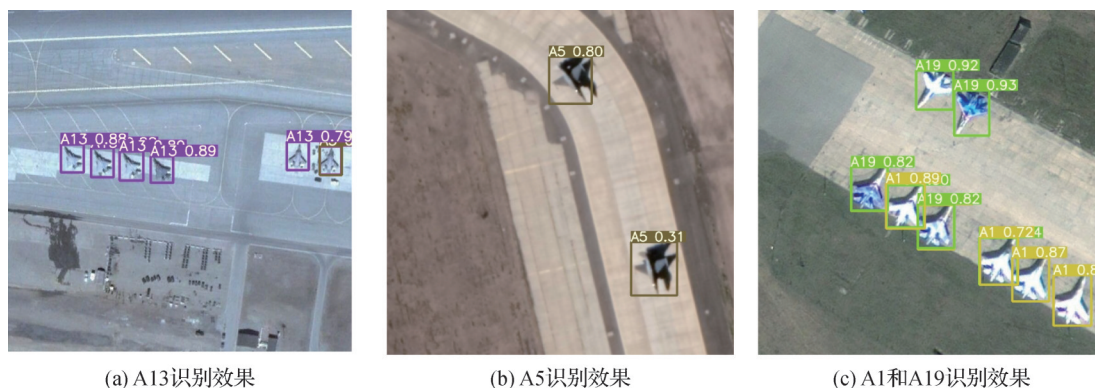


图12 YOLOv5s的实际识别效果

Fig. 12 The actual recognition renderings of YOLOv5s

((a) the recognition of A13; (b) the recognition of A5; (c) the recognition of A1 and A19)



图 13 YOLOv8s 的实际识别效果

Fig. 13 The actual recognition renderings of YOLOv8s

((a) the recognition of A13; (b) the recognition of A5; (c) the recognition of A1 and A19)



图 14 改进 YOLOv5s 的实际识别效果

Fig. 14 The actual recognition renderings of improved YOLOv5s

((a) the recognition of A13; (b) the recognition of A5; (c) the recognition of A1 and A19)

YOLOv5s 上出现,改进 YOLOv5s 的检测结果全部正确,实验效果图体现出本文模型对数据集 MAR20 具有良好的识别效果。

3 结 论

为解决遥感飞机影像识别过程中目标尺寸差异大、类内相似度高等问题,本文在 YOLOv5s 的基础上,融合了边界框高斯建模与特征聚合分发机制,提出了一种改进算法。具体地,使用 NWD 与 IoU 相结合的方式改进了 YOLOv5s 的损失函数,提高了对不同尺寸目标的检测效果,进而提高模型的检测精度。同时针对不同种类飞机特征相似、子类间难以区分的问题,利用 Gold-YOLO 中的 gather-and-distribute 特征聚合模块,增强 YOLOv5s 网络提取细粒度特征的能力。将改进的 YOLOv5s 与 YOLOv5s、YOLOv8s、Gold-YOLO 和 Faster R-CNN 进行对比,本文的模型

检测精度最佳。

为了在不使模型精度下降的前提下提高模型的图像处理速度,并尽可能减少算力资源的消耗以便未来实现轻量化部署,后续考虑使用 C3_DSConv 网络替代 YOLOv5s 检测部分的 C3 网络,提高模型检测速度并做到轻量化。

参考文献 (References)

Gai R L, Chen N and Yuan H. 2023. A detection algorithm for cherry fruits based on the improved YOLO-v4 model. *Neural Computing and Applications*, 35(19): 13895-13906 [DOI: 10.1007/s00521-021-06029-z]

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Hua Y R, Zhang Z, Long S and Zhang Q L. 2020. Remote sensing

- image target detection based on improved YOLO algorithm. *Electronic Measurement Technology*, 43(24): 87-92 (化嫣然, 张卓, 龙赛, 张青林. 2020. 基于改进YOLO算法的遥感图像目标检测. *电子测量技术*, 43(24): 87-92) [DOI: 10.19651/j.cnki.emt.2005268]
- Huang J, Jiang Z G, Zhang H P and Yao Y. 2017. Ship object detection in remote sensing images using convolutional neural networks. *Journal of Beijing University of Aeronautics and Astronautics*, 43(9): 1841-1848 (黄洁, 姜志国, 张浩鹏, 姚远. 2017. 基于卷积神经网络的遥感图像舰船目标检测. *北京航空航天大学学报*, 43(9): 1841-1848) [DOI: 10.13700/j.bh.1001-5965.2016.0755]
- Li X, Liu S N, Yang Z and Wang K K. 2021. Multi-target detection of transmission lines based on improved cascade R-CNN. *Journal of Electronic Measurement and Instrumentation*, 35(10): 24-32 (李鑫, 刘帅男, 杨桢, 王珂珂. 2021. 基于改进Cascade R-CNN的输电线路多目标检测. *电子测量与仪器学报*, 35(10): 24-32) [DOI: 10.13382/j.jemi.B2104058]
- Lin T Y, Dollár P, Girshick R, He K M, Hariharan B and Belongie S. 2017. Feature pyramid networks for object detection [EB/OL]. [2017-04-19]. <http://arxiv.org/pdf/1612.03144.pdf>
- Liu Q L, Jiao B L and Liu L. 2009. On remote sensing image classification method based on improved BP neural network model. *Electronics Optics and Control*, 16(8): 65-67 (刘钦龙, 焦斌亮, 刘立. 2009. 基于改进的BP神经网络模型的遥感图像分类方法研究. *电光与控制*, 16(8): 65-67) [DOI: 10.3969/j.issn.1671-637X.2009.08.016]
- Liu S, Qi L, Qin H F, Shi J P and Jia J Y. 2018. Path aggregation network for instance segmentation [EB/OL]. [2023-12-19]. <http://arxiv.org/pdf/1803.01534.pdf>
- Liu Y, Fu Z Y and Zheng F B. 2015. Review on high resolution remote sensing image classification and recognition. *Journal of Geo-Information Science*, 17(9): 1080-1091 (刘扬, 付征叶, 郑逢斌. 2015. 高分辨率遥感影像目标分类与识别研究进展. *地球信息科学学报*, 17(9): 1080-1091) [DOI: 10.3724/SP.J.1047.2015.01080]
- Minaee S, Boykov Y Y, Porikli F, Plaza A J, Kehtarnavaz N and Terzopoulos D. 2022. Image segmentation using deep learning: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7): 3523-3542 [DOI: 10.1109/TPAMI.2021.3059968]
- Oh M and Park H M. 2011. Blind source separation based on independent vector analysis using feed-forward network. *Neurocomputing*, 74(17): 3713-3715 [DOI: 10.1016/j.neucom.2011.06.008]
- Pan X Y, Jia N X, Mu Y Z and Gao X R. 2023. Survey of small object detection. *Journal of Image and Graphics*, 28(9): 2587-2615 (潘晓英, 贾凝心, 穆元震, 高炫蓉. 2023. 小目标检测研究综述. *中国图象图形学报*, 28(9): 2587-2615) [DOI: 10.11834/jig.220455]
- Pang Y W, Zhang K, Yuan Y and Wang K Q. 2014. Distributed object detection with linear SVMs. *IEEE Transactions on Cybernetics*, 44(11): 2122-2133 [DOI: 10.1109/TCYB.2014.2301453]
- Redmon J, Divvala S, Girshick R and Farhadi A. 2016. You only look once: unified, real-time object detection//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE: 779-788 [DOI: 10.1109/cvpr.2016.91]
- Shelhamer E, Long J and Darrell T. 2017. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4): 640-651 [DOI: 10.1109/TPAMI.2016.2572683]
- Shi T J, Gong J N, Jiang S K, Zhi X Y, Bao G Z, Sun Y and Zhang W. 2023. Complex optical remote-sensing aircraft detection dataset and benchmark. *IEEE Transactions on Geoscience and Remote Sensing*, 61: #5612309 [DOI: 10.1109/TGRS.2023.3283137]
- Shi W X, Tan D L and Bao S L. 2020. Feature enhancement SSD algorithm and its application in remote sensing images target detection. *Acta Photonica Sinica*, 49(1): #0128002 (史文旭, 谭代伦, 鲍胜利. 2020. 特征增强SSD算法及其在遥感目标检测中的应用. *光子学报*, 49(1): #0128002) [DOI: 10.3788/gzxb20204901.0128002]
- Shi Z H, Wu C W, Li C J, You Z Z, Wang Q and Ma C C. 2023. Object detection techniques based on deep learning for aerial remote sensing images: a survey. *Journal of Image and Graphics*, 28(9): 2616-2643 (石争浩, 仵晨伟, 李成建, 尤珍臻, 王泉, 马城城. 2023. 航空遥感图像深度学习目标检测技术研究进展. *中国图象图形学报*, 28(9): 2616-2643) [DOI: 10.11834/jig.221085]
- Sugiarto B, Prakasa E, Wardoyo R, Damayanti R, Krisdianto N, Dewi L M, Pardede H F and Rianto Y. 2017. Wood identification based on histogram of oriented gradient (HOG) feature and support vector machine (SVM) classifier//*Proceedings of the 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*. Yogyakarta, Indonesia: IEEE: 337-341 [DOI: 10.1109/ICITISEE.2017.8285523]
- Šulc M and Matas J. 2017. Fine-grained recognition of plants from images. *Plant Methods*, 13: #115 [DOI: 10.1186/s13007-017-0265-4]
- Wang C C, He W, Nie Y, Guo J Y, Liu C J, Han K and Wang Y H. 2023. Gold-YOLO: efficient object detector via gather-and-distribute mechanism [EB/OL]. [2023-12-19]. <https://arxiv.org/pdf/2309.11331.pdf>
- Wang J W, Xu C, Yang W and Yu L. 2022. A normalized Gaussian Wasserstein distance for tiny object detection [EB/OL]. [2023-12-19]. <http://arxiv.org/pdf/2110.13389.pdf>
- Yang Z M and Song W. 2023. Selecting and fusing coarse-and-fine granularity features for fine-grained image recognition. *Journal of Image and Graphics*, 28(7): 2081-2092 (阳治民, 宋威. 2023. 选择并融合粗细粒度特征的细粒度图像识别. *中国图象图形学报*, 28(7): 2081-2092) [DOI: 10.11834/jig.220052]
- Yilmaz A, Javed O and Shah M. 2006. Object tracking: a survey. *ACM Computing Surveys (CSUR)*, 38(4): #13-es [DOI: 10.1145/1177352.1177355]

- Yuan X, Cheng G, Li G, Dai W, Yin W X, Feng Y C, Yao X W, Huang Z L, Sun X and Han J W. 2023. Progress in small object detection for remote sensing images. *Journal of Image and Graphics*, 28(6): 1662-1684 (袁翔, 程焱, 李戈, 戴威, 尹文昕, 冯瑛超, 姚西文, 黄钟铃, 孙显, 韩军伟. 2023. 遥感影像小目标检测研究进展. *中国图象图形学报*, 28(6): 1662-1684) [DOI: 10.11834/jig.221202]
- Zhang X S and Gao T. 2020. Multi-head attention model for aspect level sentiment analysis. *Journal of Intelligent and Fuzzy Systems*, 38(1): 89-96 [DOI: 10.3233/JIFS-179383]
- Zhao L Q and Li S Y. 2020. Object detection algorithm based on improved YOLOv3. *Electronics*, 9(3): #537 [DOI: 10.3390/electronics9030537]
- Zheng Z H, Wang P, Liu W, Li J Z, Ye R G and Ren D W. 2020. Distance-IoU loss: faster and better learning for bounding box regression//*Proceedings of the 34th AAAI Conference on Artificial Intelligence*. New York, USA: AAAI: 12993-13000 [DOI: 10.1609/aaai.v34i07.6999]
- Zhu X K, Lyu S C, Wang X and Zhao Q. 2021. TPH-YOLOv5:

improved YOLOv5 based on Transformer prediction head for object detection on drone-captured scenarios//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops*. Montreal, Canada: IEEE: 2778-2788 [DOI: 10.1109/iccwv54120.2021.00312]

作者简介

王晓燕,女,副教授,主要研究方向为图像处理及数学在工程中的应用。E-mail:wangxy252@163.com

李杰,通信作者,男,教授,主要研究方向为智能无人车辆设计与控制。E-mail:lijie1@bucea.edu.cn

梁文辉,女,硕士研究生,主要研究方向为遥感图像处理和识别。E-mail:13513678536@163.com

牟建宏,男,硕士研究生,主要研究方向为计算机视觉。E-mail:jianhongmu@163.com

王禧钰,女,硕士研究生,主要研究方向为图像处理。E-mail:22311081203007@bwu.edu.cn