

基于YOLOX的遥感飞机检测算法

赵立敏, 白宗文*

延安大学物理与电子信息学院, 陕西 延安

收稿日期: 2023年2月25日; 录用日期: 2023年3月24日; 发布日期: 2023年3月31日

摘要

针对遥感飞机检测中存在背景信息复杂、目标尺寸小、特征不明显、多尺度等问题, 提出了一种基于YOLOX的遥感飞机目标检测算法。本文以YOLOX模型为基准, 在残差网络的基础上加入通道注意力和空间注意力, 使网络专注于学习图像中具体目标的结构而忽略背景图的内容细节, 从而提高了目标检测的精度, 并且对小目标的检测能力、特征不明显的目标的检测精度也有一定的提升; 同时, 用空洞空间金字塔池化(At-rous Space Pyramid Pooling, ASPP)替换空间金字塔池化(Space Pyramid Pooling, SPP), ASPP可以聚合多尺度上下文信息, 增大感受野, 提高模型检测同一类别不同尺寸的能力。实验结果显示, 在RSOD数据集上AP@0.5提高了1.17%, 在UCAS-AOD数据集上AP@0.5提高了1.03%。

关键词

遥感图像, YOLOX, 通道注意力, 空间注意力, 空间金字塔池化, 空洞空间金字塔池化

Remote Sensing Aircraft Detection Algorithm Based on YOLOX

Limin Zhao, Zongwen Bai*

School of Physics and Electronic Information, Yan'an University, Yan'an Shaanxi

Received: Feb. 25th, 2023; accepted: Mar. 24th, 2023; published: Mar. 31st, 2023

Abstract

A YOLOX-based algorithm for remote sensing aircraft target detection is proposed for the problems of complex background information, small target size, inconspicuous features, and multi-scale in remote sensing aircraft detection. In this paper, the YOLOX model is used as a benchmark, and channel attention and spatial attention are added to the residual network, so that the network fo-

*通讯作者。

cuses on learning the structure of specific targets in the image and ignores the content details of the background map, thus improving the accuracy of target detection, and the detection capability of small targets and the detection accuracy of targets with obscure features; at the same time, atrous space pyramid pooling (At The difference between ASPP and SPP lies in the fact that ASPP incorporates cavity convolution with a larger field of perception for the features extracted from the image and allows the resolution to be aggregated without too much degradation, allowing the aggregation of multi-scale contextual information and increasing the field of perception to improve the model. information, increase the perceptual field and improve the model's ability to detect different sizes of the same category. The experimental results show that on the RSOD dataset AP@0.5 increased by 1.17%, and on the UCAS-AOD dataset AP@0.5 increased by 1.03%.

Keywords

Remote Sensing Image, YOLOX, Channel Attention, Spatial Attention, Space Pyramid Pooling (SPP), At-Route Space Pyramid Pooling (ASPP)

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着遥感技术的日益发展, 通过高分辨率航空卫星采集得到的遥感图像已经在遥感图像目标检测方向提供了良好的基础[1]。遥感技术最开始应用在军事方面, 之后逐渐从军事发展到民用领域。在军事领域, 遥感图像飞机检测主要应用于军事侦察、导弹预警、武器制导和军事测绘等领域; 在民用领域, 遥感图像飞机检测主要应用在机场飞机的调度和更高效地搜寻失联飞机等方面。

目标检测一直是计算机视觉领域中的重要任务之一[2]。目标检测的具体任务就是在一张图片中确定目标对象的位置并且判断检测到的目标类别。传统的目标检测算法仅仅适用于背景信息简单、特征明显的情况, 而在实际的应用中, 待检测的目标存在背景信息复杂、目标尺度变化多样等问题, 所以传统的目标检测算法已经难以适用于当前的实际需求。

近几年, 深度学习领域已经取得了突破性的进展, 深度学习的方法已经被广泛应用在计算机视觉领域。基于深度学习的目标检测算法可以同时提取同一物体不同尺寸的特征信息, 对于背景复杂的目标也有更好的鲁棒性和泛化能力, 更加适用于实际场景中。目前, 基于深度学习的目标检测算法主要分为两大类: 第一类是基于候选框的两阶段目标检测算法, 第二类是基于回归问题的单阶段目标检测算法[3]。基于候选框的两阶段目标检测算法是由特定的算法生成一系列候选框, 再通过卷积神经网络对目标进行分类, 最后对边界框的位置进行精修, 具有代表性的算法有 R-CNN [4]等一系列算法; 基于回归问题的单阶段目标检测算法直接将目标边框定位问题转换为回归问题处理, 直接对预测的目标物体进行回归, 不用产生候选框, 经典的算法有 SSD [5]和 YOLO 等系列算法。这两类算法有各自不同的优势, 前者在检测精确度和定位精度上占优, 但后者在检测速度上更具有优势。

2. 算法基本原理

在 YOLOv5 算法提出一年多后, 2021 年 8 月, 来自旷视 Zheng Ge 等人提出高性能 YOLOX [6]检测算法。在过去两年里, 由于目标检测的主要进展在无锚点(Anchor-free)检测器[7]、先进的标签分配策略

和端到端的(NMS-free)检测器[8]等方面, 但之前提出的 YOLOv4、YOLOv5 算法还是基于 anchor 的检测器, 使用手工分配策略进行训练, 没有集成这些新的技术。矿视的研究者们将这些新技术巧妙地集成在一起, 使算法的性能得到提升。本文所采用的模型是 YOLOX-L, YOLOX-L 整体可以分为 3 部分, 其中, Backbone 部分采用主干特征提取网络 CSPDarknet 来进行浅层特征的提取, Neck 部分即采用 FPN + PAN 模块进行深层特征提取, Prediction 部分采用了三个解耦头对目标框进行预测, 得到最终的检测结果。

YOLOX 是一种端到端的单阶段目标检测算法, 通过在网络中采用残差模块和多尺度的检测方法, 在通用的公开数据集上有不错的检测效果。YOLOX 算法主要针对通用数据集, 通用数据集中的目标在整张图片中占比较大, 背景信息简单, 而遥感图像中存在飞机目标所占像素较少, 特征信息不明显, 目标排列紧密, 背景信息复杂等问题, YOLOX 算法不能很好地适用于遥感图像飞机目标检测。针对遥感图像飞机目标高精度检测的需要, 本文在基于 YOLOX 算法的基础上, 借鉴了两个方法进行改进, 第一种方法是在残差网络的基础上加入通道注意力和空间注意力, 可以使网络专注于学习图像中具体目标的结构而忽略背景图的内容细节, 从而提高了目标检测的精度, 第二种方法是把网络中的空间金字塔池化改进为空洞空间金字塔池化来融合多尺度上下文信息, 提高感受野, 增强同一物体不同尺寸的检测精度。基于以上两种方法, 可以有效地改善背景信息复杂、目标尺寸小、特征不明显、多尺度等问题。

2.1. 残差模块

残差模块的结构如图 1 所示, 残差网络可分为两个部分, 对于某一特征图, 特征图依次经过主干部分 1×1 卷积和 3×3 卷积, 学习到图像的特征信息, 残差边部分不做任何处理, 直接将主干的输入与输出结合得到最终的结果。

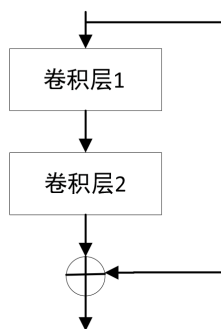


Figure 1. Residual network

图 1. 残差网络

残差网络的特点是容易优化, 并且能够增加相当的深度来提高准确率, 其内部的残差块使用了跳跃链接, 缓解了 in 深度神经网络中增加深度带来的梯度消失的问题。

2.2. 空间金字塔池化(SPP)

空间金字塔池化(SPP)[9]是由微软研究院的何凯明提出的, 空间金字塔池化模块的结构如图 2 所示, SPP 模块是由 4 个并行的分支构成, 分别是 kernel size 为 5×5 , 9×9 , 13×13 的最大池化和一个跳跃连接, 采用不同大小的池化核丰富了特征图的表达能力, 有利于待检测图像中目标尺寸大小差异大的情况。

SPP 的特点如下:

- 1) 不管输入尺寸多大, SPP 可以产生固定大小的输出;
- 2) SPP 可以使用同一图像不同尺寸作为输入, 得到同样长度的池化特征;

- 3) 训练图像尺寸的多样性比单一尺寸的训练图像更容易使得网络收敛;
- 4) 解决了卷积神经网络对图像重复特征提取的问题, 大大提高了产生候选框的速度, 且节省了计算成本。

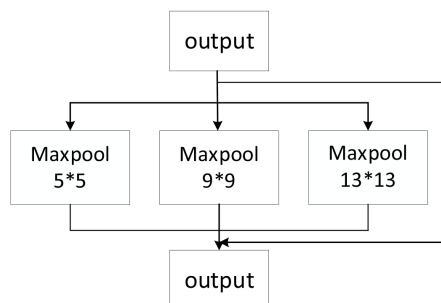


Figure 2. Spatial pyramid pooling structure
图 2. 空间金字塔池化结构

3. 算法改进

3.1. 注意力模块

CBAM [10]是一种轻量级的注意力模块, 它包括两个注意力模块, 分别为通道注意力模块和空间注意力模块, 如图 3 和图 4 所示, 它分别是在通道和空间维度上进行 Attention。

通道注意力主要是压缩了输入的特征图空间层面的内容。通道注意力模块首先是将输入的特征图分别经过一个全局最大池化和一个全局平均池化, 分别形成两个 $1 \times 1 \times C$ 的权重向量; 然后将得到的权重向量分别经过一个两层的多层感知机, 从而映射为每个通道的权重; 其次将映射后的权重进行相加, 后边接 sigmoid 激活函数输出; 最后将得到的通道权重 $1 \times 1 \times C$ 和原特征图 $H \times W \times C$ 按相应的通道进行相乘。

空间注意力主要是对通道注意力的补充, 引入的作用是挖掘最有用的内容信息。空间注意力模块, 特征图首先分别按通道经过最大池化和平均池化, 得到两个 $H \times W \times 1$ 的权重向量。通道数从 $H \times W \times C$ 变为 $H \times W \times 1$, 是对同一特征点的所有通道进行池化; 得到的两张特征图进行堆叠, 形成 $H \times W \times 2$ 的特征图空间权重; 再经过一个卷积层, 特征图的维度从 $H \times W \times 2$ 变为 $H \times W \times 1$, $H \times W \times 1$ 的特征图表示了特征图中每个点的重要性, 数值越大重要程度越高; 最后将得到的空间权重 $H \times W \times 1$ 与原特征图 $H \times W \times C$ 相乘。

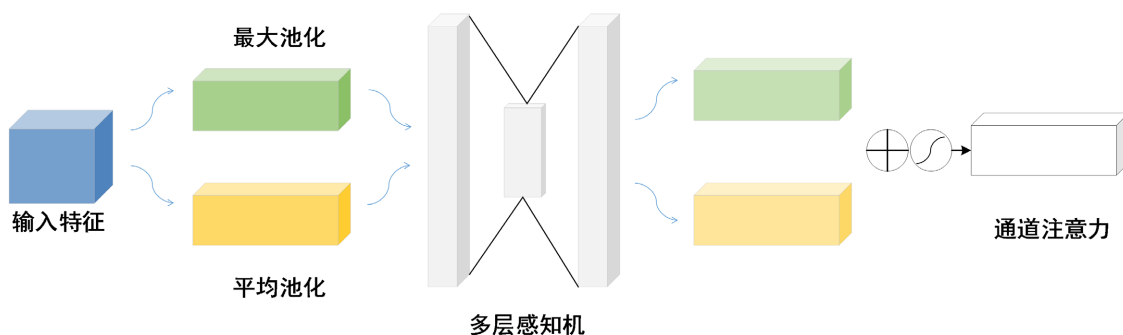


Figure 3. Channel attention
图 3. 通道注意力

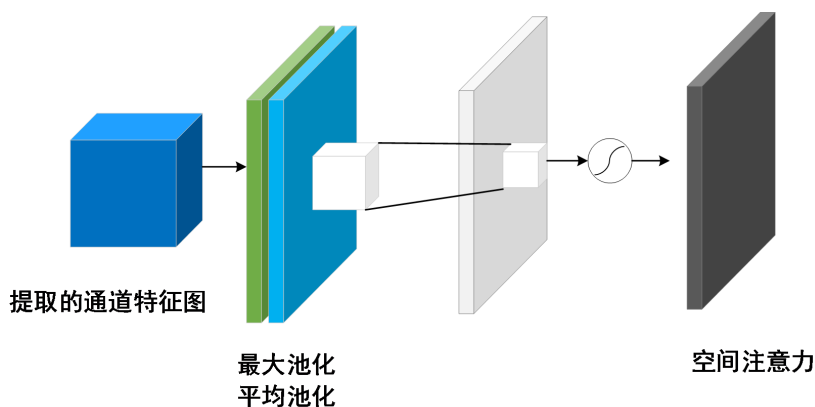


Figure 4. Spatial attention

图 4. 空间注意力

3.2. 注意力残差模块

采用注意力模块[11], 可以使网络有效地关注图像的目标信息特征, 充分地进行目标信息的提取, 并且忽略不重要的背景信息, 可以有效地提高网络分类精度的提升。因为一个标准的卷积具有一定的通道数, 每个通道都可以自适应地提出某一特定特征。

本文中采用了残差模块与注意力模块相结合的方法组成注意力残差模块, 形成了更深的网络结构, 从而使网络的分类能力有所提升。如图 5 所示, 特征图首先经过两个卷积进行特征提取, 其次再经过两个注意力模块, 最后将得到的特征与原输入图像相加。

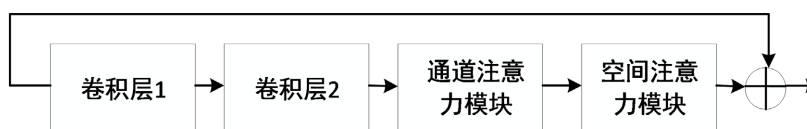


Figure 5. Attentional residuals module

图 5. 注意力残差模块

3.3. 空洞空间金字塔池化模块

在遥感飞机图像的检测任务中, 由于遥感飞机图像所处的背景信息复杂、曝光环境、目标尺寸小和飞机目标有不同的摆放角度等特点, 因此在检测飞机的时候导致提取信息不充足, 进而使遥感飞机目标的检测率下降。本文采用了 ASPP [12]模块来改进原有的 SPP 模块, ASPP 其核心思想就是采用了空洞卷积来代替了 SPP 模块中的池化操作, 采用空洞卷积有效避免了池化操作带来的局部信息与细节信息的消失, 可以有效考虑全局信息。YOLOX 模型中采用的是 SPP 模块, SPP 模块采用了三个不同大小的池化核进行池化操作, SPP 模块在进行池化操作的过程中虽然感受野扩大了, 但是图像的分辨率不断下降, 包含的信息越来越抽象, 并且图像的局部信息与细节信息会逐渐消失。本论文引入的 ASPP 模块采用了不同扩张率的空洞卷积来代替池化操作, 并且结合了全局平均池化, 整合了全局空间信息, 可以有效提取同一物体不同尺寸的特征信息。

空间金字塔池化模块(ASPP)如图 6 所示, ASPP 模块分别将输入的特征经过 5 个不同的分支, 第 1 个分支是经过一个 1×1 的卷积操作, 目的是和原始输入特征的感受野保持一致, 第 2、第 3 和第 4 个支路分别经过不同的扩张率的空洞卷积, 扩张率分别是 6、12 和 18, 并且卷积核为 3×3 , 目的是进行特征

提取, 得到不同感受野的特征图来融合多尺度信息。第 5 个分支是为了获取全局特征, 将特征做全局平均化池化后, 最后经过 1×1 卷积和归一化处理, 再进行上采样。YOLOX 模型中采用 ASPP 模块后, 模型可以学习到遥感飞机图像的不同尺寸大小的特征, 更好地降低模型的漏检率。

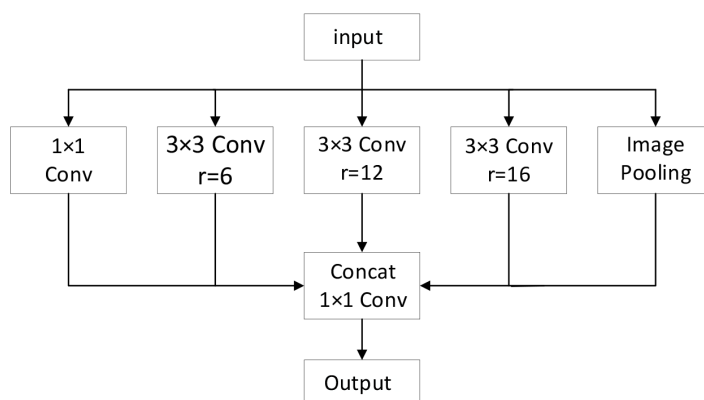


Figure 6. Spatial pyramid pooling module
图 6. 空间金字塔池化模块

4. 实验

4.1. 实验与分析

4.1.1. 实验平台

本次实验所使用的配置如下: 操作系统为 Ubuntu 16.04, 利用 2 块 v100 GPU 进行训练, 开发平台为 pycharm, 开发语言为 python, 深度学习框架为 pytorch。

4.1.2. 实验数据集

本实验采用了 RSOD [13]和 UCAS-AOD [14]两种数据集中包含飞机遥感图像作为数据集。RSOD 数据集是武汉大学公开的, UCAS-AOD 数据集是由中国科学院大学提出的遥感图像数据集。

数据集的特点(见图 7): 图片比较模糊、目标尺寸差异大、小目标多、目标和背景特征相似等特点。

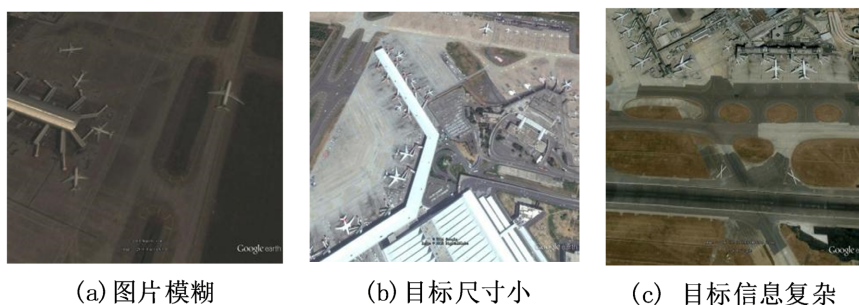


Figure 7. Characteristics of the dataset
图 7. 数据集的特点

4.1.3. 实验的评价指标

本实验是用平均正确率(AP)、F1 Score、精确度(Precision)和召回率(Recall)四个评价指标来评价该模型的性能。在目标检测算法模型中平均正确率是用来评价单一种类的指标。平均正确率(AP)是用精确度

和召回率来评判算法模型的准确度, 采用 PR 曲线下的面积来表示 AP 值。

所用公式如下列所示:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{AP} = \int_0^1 p(r) dr \quad (3)$$

$$F1_k = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

其中, TP (True Positive)为真正类, 表示一个实例本来是正例并且被判断为正类; FP (False Positive)为假正类, 表示一个实例本来是负例但被误判为正类; FN (False Negative)为假负例, 表示一个实例本来为正例但被误判为负例。

4.1.4. 消融性实验对比

为了展现算法的检测性能, 本论文在相同的实验环境下对 YOLOX 算法进行改进, 分别用 YOLOX 的算法在 RSOD 数据集和 UCAS-AOD 数据集上加入了 CBAM 模块和 ASPP 模块, 实验分别从四个指标来评判算法的性能。实验结果如表 1 和表 2 所示。

从上边两个表的数据可以看到, 论文中加入的两个模块对 YOLOX 算法检测飞机遥感图像的性能有明显的提升。在 RSOD 数据集上最终改进后的精度相比原 YOLOX 算法增加了 0.3%, 召回率增加了 3.04%, AP 值增加了 1.17%, F1-Score 增加了 0.02。在 UCAS-AOD 数据集上最终改进后的精度相比原 YOLOX 算法增加了 0.57%, 召回率增加了 1.73%, AP 值增加了 1.57%, F1-Score 增加了 0.01。

如图 8 所示, 是精度 - 召回率曲线图对比。从曲线图中的数值可以对比出, 在两个数据集上 AP 值分别提高了 1.17%、1.57%。

Table 1. Comparison of Aircraft ablation experiments in the RSOD dataset

表 1. RSOD 数据集集中的 Aircraft 消融实验对比

算法	Precision/%	Recall/%	AP/%	F1-Score
YOLOX	94.04	92.44	96.83	0.93
YOLOX + CBAM	94.05	94.16	97.68	0.94
YOLOX + CBAM + ASPP	94.34	95.48	98.00	0.95

Table 2. Comparison of Plane ablation experiments in the UCAS-AOD dataset

表 2. UCAS-AOD 数据集集中的 Plane 消融实验对比

算法	Precision/%	Recall/%	AP/%	F1-Score
YOLOX	92.96	92.50	94.90	0.93
YOLOX + CBAM	94.44	92.35	95.27	0.94
YOLOX + CBAM + ASPP	95.42	92.59	95.93	0.94

YOLOX 算法、YOLOX + CBAM 算法和 YOLOX + CBAM + ASPP 算法在测试集上的检测结果如图 3 所示。由于飞机遥感图像目标尺寸差异大、小目标多、背景信息特征相似并出现在曝光环境下, 从而[15]

造成误判、漏检等情况的出现, 如图 9 所示是 YOLOX、YOLOX + CBAM 和 YOLOX + CBAM + ASPP 算法测试对比, 本文所提出的方法对 YOLOX 算法有明显的改进。对照组 1 在 YOLOX 原算法中存在着误检和漏检的情况; 对照组的目标尺寸比较小, 在检测时由于特征信息不充足, 导致原算法的检测结果存在着漏检的情况, 改进后的算法在小目标上的检测效果有所提升; 对照组 3 的飞机目标处在曝光的环境下, 背景信息和飞机目标的特征信息极为相似, 导致原算法在曝光环境下的检测效果极差, 在改进后的算法中检测结果有所提升。

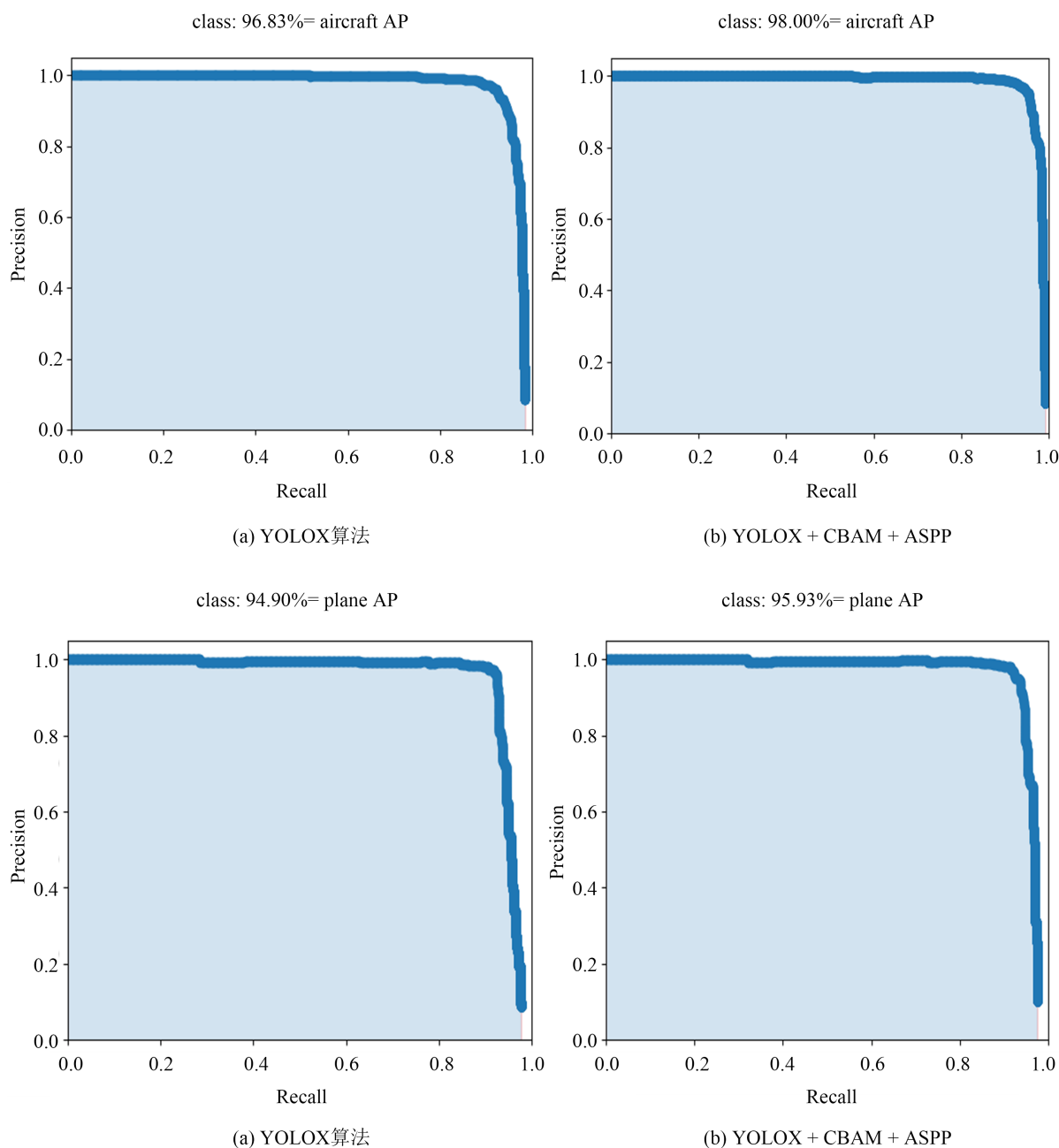


Figure 8. Comparison of precision-recall graphs
图 8. 精度 - 召回率曲线图对比

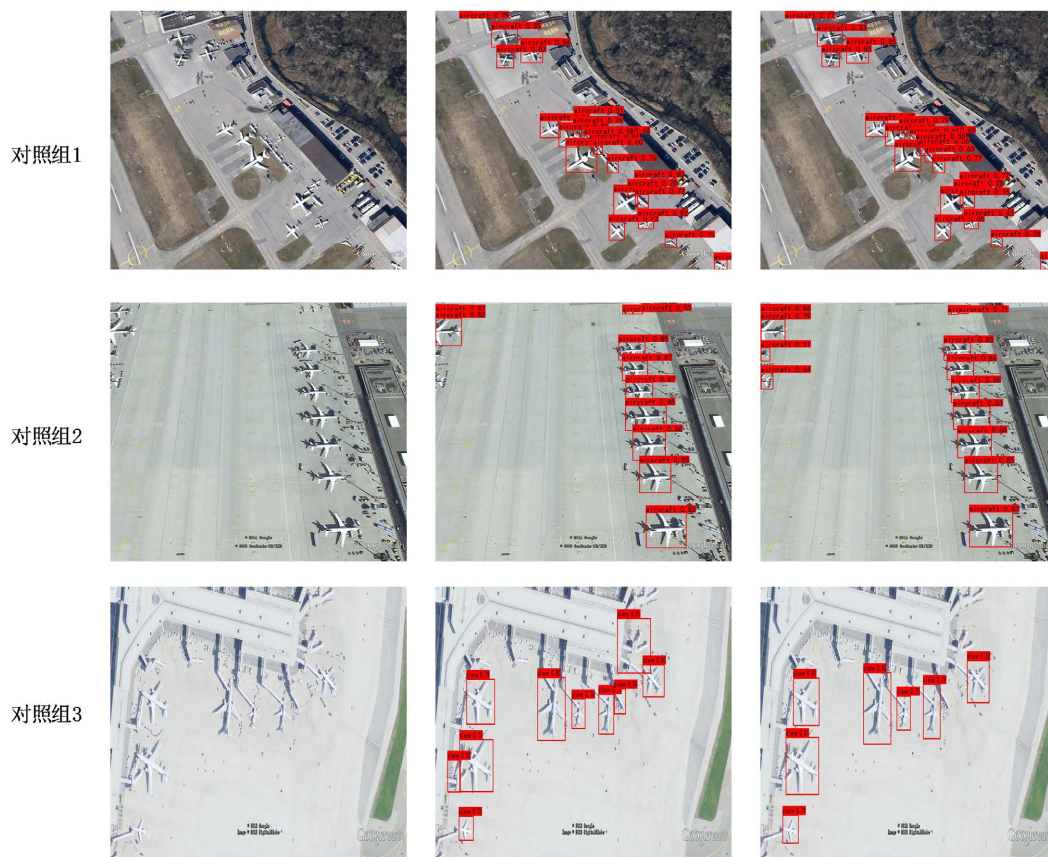


Figure 9. YOLOX YOLOX + CBAM and YOLOX + CBAM + ASPP algorithm test comparison
图 9. YOLOX YOLOX + CBAM 和 YOLOX + CBAM + ASPP 算法测试对比

5. 结论

本文是以遥感图像中飞机目标尺寸小、背景信息复杂、多尺度等问题进行研究,提出了基于 YOLOX 的改进算法。本文在原有的 YOLOX 算法框架中加入 CBAM 注意力机制模块,可以有效地提高小目标检测的精度,同时对背景信息复杂的图片可以更多关注目标信息,抑制其他无用的背景信息。本文采用空洞空间金字塔池化不仅可以有效地融合多尺度的上下文信息而且可以增大感受野。

本文所提出的改进方法在 RSOD 数据集上的检测效果有不错的效果,为了证明所改进算法的有效性和鲁棒性,在 UCAS-AOD 数据集上进行了验证,同样,检测效果也得到了提高。

基金项目

项目批准号: 62266045, 项目名称: 工业人工智能中跨媒体协同深度安全态势感知理论和应用研究。

参考文献

- [1] 侯涛, 蒋瑜. 改进 YOLOv4 在遥感飞机目标检测中的应用研究[J]. 计算机工程与应用, 2021, 57(12): 224-230.
- [2] 熊咏平. 基于 YOLO 的深度学习目标检测算法研究[D]: [硕士学位论文]. 武汉: 武汉科技大学, 2019.
- [3] 张欣, 张永强, 何斌, 李国宁. 基于 YOLOv4-tiny 的遥感图像飞机目标检测技术研究[J]. 光学技术, 2021, 47(3): 344-351. <https://doi.org/10.13741/j.cnki.11-1879/o4.2021.03.016>
- [4] Girshick, R., Donahue, J., Darrell, T., et al. (2013) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2013 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Portland, 23-28 June

- 2013, 580-587. <https://doi.org/10.1109/CVPR.2014.81>
- [5] Liu, W., Anguelov, D., Erhan, D., *et al.* (2015) SSD: Single Shot MultiBox Detector. *Computer Vision ECCV 2016 14th European Conference*, Amsterdam, 11-14 October 2016, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
- [6] Ge, Z., Liu, S., Wang, F., *et al.* (2021) YOLOX: Exceeding YOLO Series in 2021.
- [7] Tian, Z., Shen, C., Chen, H., *et al.* (2020) FCOS: Fully Convolutional One-Stage Object Detection. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 9626-9635. <https://doi.org/10.1109/ICCV.2019.00972>
- [8] Wang, J., Song, L., Li, Z., *et al.* (2020) End-to-End Object Detection with Fully Convolutional Network. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 15849-15858. <https://doi.org/10.1109/CVPR46437.2021.01559>
- [9] He, K., Zhang, X., Ren, S., *et al.* (2015) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **37**, 1904-1916. <https://doi.org/10.1109/TPAMI.2015.2389824>
- [10] Woo, S., Park, J., Lee, J.Y., *et al.* (2018) Cbam: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1
- [11] Fei, W., Jiang, M., Chen, Q., *et al.* (2017) Residual Attention Network for Image Classification. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 6450-6458.
- [12] Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L. (2018) Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected Crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**, 834-848. <https://doi.org/10.1109/TPAMI.2017.2699184>
- [13] Yang, L., *et al.* (2017) Accurate Object Localization in Remote Sensing Images Based on Convolutional Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*, **55**, 2486-2498. <https://doi.org/10.1109/TGRS.2016.2645610>
- [14] Zhu, H., Chen, X., Dai, W., *et al.* (2015) Orientation Robust Object Detection in Aerial Images Using Deep Convolutional Neural Network. *IEEE International Conference on Image Processing*, Quebec City, 27-30 September 2015, 3735-3739. <https://doi.org/10.1109/ICIP.2015.7351502>
- [15] 张港. 基于深度学习轻量化卷积神经网络的遥感图像场景分类研究[D]: [硕士学位论文]. 南京: 南京邮电大学, 2022.