

基于平滑标签和多路聚合网络的遥感飞机检测

李克文¹, 张宝华^{1,3*}, 吕晓琪^{2,3}, 谷宇^{1,3}, 王月明^{1,3}, 刘新^{1,3}, 任彦¹, 李建军^{1,3}, 张明^{1,3}

¹内蒙古科技大学信息工程学院, 内蒙古 包头 014010;

²内蒙古工业大学信息工程学院, 内蒙古 呼和浩特 010051;

³内蒙古模式识别与智能图像处理重点实验室, 内蒙古 包头 014010

摘要 针对遥感图像背景复杂、飞机目标尺寸变化大的问题,提出了基于平滑标签和多路聚合网络的遥感飞机检测方法。考虑到遥感图像中飞机目标辨识难度大,利用联合注意力机制捕获目标区域,缩小搜索范围。再使用改进的路径聚合网络提取主干网络中的4个特征层,可以有效提取浅层特征信息,将各层特征归一化后进行融合,预测目标的位置。为了避免训练模型过度依赖预测标签,造成过拟合,在网络中使用平滑标签技术,减小类内距离,有效提高训练模型的泛化能力。通过在两个公开数据集RSOD和HRRSD上进行大量实验,验证了提出方法的有效性。实验结果表明在RSOD数据集中,提出方法的平均精确率为0.967,在HRRSD数据集中平均精确率为0.993,与相关算法对比,检测精度有明显提升。

关键词 图像处理; 遥感图像; 平滑标签; 多路聚合网络; 注意力机制

中图分类号 TP751.1

文献标志码 A

DOI: 10.3788/LOP202259.1210011

Remote Sensing Aircraft Detection Based on Smooth Label and Multipath Aggregation Network

Li Kewen¹, Zhang Baohua^{1,3*}, Lv Xiaoqi^{2,3}, Gu Yu^{1,3}, Wang Yueming^{1,3}, Liu Xin^{1,3}, Ren Yan¹,
Li Jianjun^{1,3}, Zhang Ming^{1,3}

¹*School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou 014010, Inner Mongolia, China;*

²*School of Information Engineering, Mongolia Industrial University, Huhehaote 010051, Inner Mongolia, China;*

³*Inner Mongolia Key Laboratory of Pattern Recognition and Intelligent Image Processing, Baotou 014010, Inner Mongolia, China*

Abstract In order to solve the problem of complex background in remote sensing images and the large variation of aircraft target size, a new algorithm for remote sensing aircraft detection based on the smooth label and multipath aggregation network is proposed. Considering the difficulty of aircraft target identification in remote sensing images, an associative attention mechanism is used to capture the target area and narrow the search range. Then, the improved path aggregation network is used to extract the four feature layers in the backbone network, so as to effectively extract the shallow feature information. When the features of each layer are normalized, they are fused to

收稿日期: 2021-04-22; 修回日期: 2021-05-20; 录用日期: 2021-06-11

基金项目: 国家自然科学基金(61962046, 62001255, 61841204)、内蒙古杰青培育项目(2018JQ02)、内蒙古科技计划项目(202001)、内蒙古草原英才、内蒙古自治区自然科学基金(2019MS06003, 2018MS06018)、教育部“春晖计划”合作科研项目(教外司留1383号)、内蒙古自治区高等学校科学技术研究项目(NJZY145)

通信作者: *zbh_wj2004@imust.cn

predict the position of the target. In order to avoid the training model relying too much on the prediction labels, resulting in over fitting, technology for the smooth label is used in the network to reduce the inter-class distance, which effectively improves the generalization ability of the training model. The effectiveness of the proposed algorithm is verified by a large number of experiments on two public data sets RSOD and HRRSD. The experimental results show that the average accuracy in the RSOD data set and the HRRSD data set is 0.967 and 0.993 respectively. Compared with the related algorithms, the detection accuracy of the proposed algorithm has been significantly improved.

Key words image processing; remote sensing image; smooth label; multipath aggregation network; attention mechanism

1 引言

卫星和无人机等对地遥感观测技术发展迅速,遥感图像数据的分辨率不断提高,图像涵盖的目标种类更丰富。飞机遥感图像检测在军事和民用上都存在巨大的应用价值^[1],通过获取持续的机场飞机动态信息,有利于观测飞机变化情况,对于战略决策和机场调度都具有重要意义。高分辨率的遥感图像具有丰富的纹理信息和细节特征,可以直观地表达物体的外部信息,但其丰富的特征减少了目标对比度,并且图像中的飞机目标受到光照、阴影、遮挡和天气等因素的影响,干扰目标检测。因此,飞机目标的精确检测仍然面临很多挑战^[2]。

汪亚妮等^[3]提出一种基于注意力和特征融合的单阶段遥感图像目标检测模型,在检测分支上加入注意力分支然后再将两个分支进行特征融合,优点是能获得细节信息和语义信息更丰富的特征图,存在问题是对密集目标检测效果不理想;徐俊峰等^[4]提出一种自适应的多特征融合变化检测与深度学习相结合的遥感飞机目标检测方法,加入自适应直方图统计方法获取变化和不变化的样本,然后再提取图像的光谱、边缘和纹理特征,完成多特征融合,优点是实现了飞机变化目标的检测,存在问题检测结果可能出现飞机特征不完整的现象;Huang等^[5]提出一种跨尺度特征融合金字塔网络,在网络中引入跨尺度模块从特征中提取足够全面的语义信息,进行多尺度特征融合,优点是通过对多层次、多尺度的特征融合,解决了遥感图像中目标尺寸差异大的问题,存在问题是对复杂背景下的检测性能不理想;Qiu等^[6]提出一种自适应纵横比多尺度网络,用于遥感图像中不同大小和宽高比的目标检测,优点是减轻了训练过程中宽高比变化引起的物体形状扭曲,存

在问题是会产生一些嘈杂的边界框,这些边界框没有紧紧围绕在对象周围;Chen等^[7]提出一种基于场景长下文特征金字塔网络,能够加强目标和场景之间的关系,解决目标大小变化带来的问题,优点是保证多尺度目标检测的有效性,为小目标提取更丰富的信息,存在问题是对密集目标检测效果不理想。

本文所用检测策略是学习图像中的飞机目标特征,在测试时通过搜索整张图像中的飞机目标进行定位和检测。首先在主干网络中加入融合注意力机制搜索目标区域,抑制背景信息干扰,再使用多路聚合网络提取目标特征。在此基础上,加入平滑标签技术对标签提供一个惩罚机制,提高模型泛化能力,有效防止模型过拟合,提高检测精度。最后使用数据增强技术对目标图像进行旋转、缩放、亮度等调整,提高了网络算法的鲁棒性。

2 基本原理

2.1 主干网络

本文采用YOLOv4算法进行遥感飞机检测,其骨干网络为CSPDarknet-53,跨阶段部分网络(CSP)算法可以增强网络学习能力^[8,9],将其底层目标的特征映射分成两个部分处理,一部分经过密集块和过渡层连接到下一层,另一部分和传输的特征映射相结合再连接到下一阶段。图1为总体网络结构。其中上半部分为CSPDarknet-53主干网络,下半部分左侧为YOLO head输出部分,右侧为多路聚合网络部分。

图中 $608 \times 608 \times 3$ 表示尺寸大小为608 pixel \times 608 pixel的RGB这3个维度的图像,Mish为激活函数,Resblock为残差块,CBAM为融合注意力机制,SPP为空间金字塔池化层,Concat为融合策略,Conv为卷积层,up sampling为上采样,down sampling为下采样,YOLO head为预测输出。

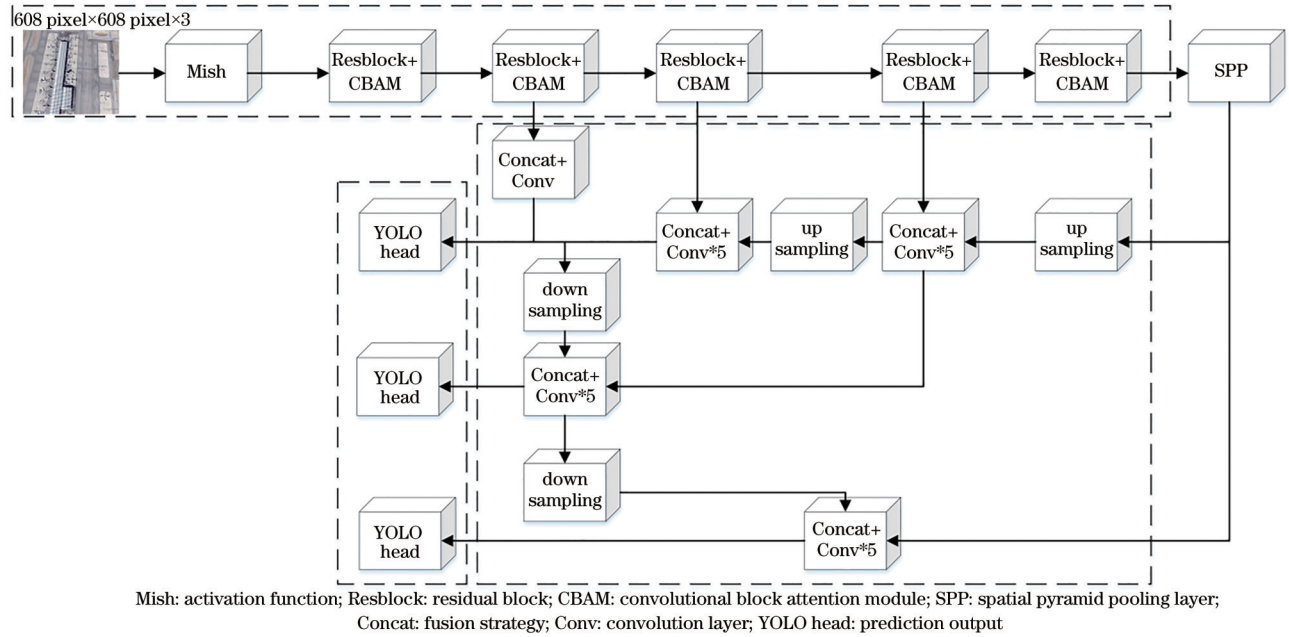


图1 总体网络结构

Fig. 1 Overall network structure

2.2 融合注意力机制

遥感图像的成像范围较广,且其背景复杂、目标较多,如何有效提取其显著性特征区域,抑制复杂背景,对于遥感目标检测起着关键性的作用,本文在YOLOv4主干网络中的每一个残差块引入融合注意力机制(CBAM)^[10],用于关注目标区域。

融合注意力是由通道注意力和空间注意力顺序排列组成。首先使用通道注意力将飞机目标的特征图在空间维度进行压缩处理,得到一个一维矢量,压缩时使用平均池化和最大池化聚合特征映射的空间信息,然后将其送到共享网络,对输入飞机特征图的空间维数进行压缩,再将各元素求和合并,以此生成通道注意力图^[10]。其表达式为

$$M_c(F) = \sigma \left\{ \text{MLP} \left[\text{AvgPool}(F) \right] + \text{MLP} \left[\text{MaxPool}(F) \right] \right\} = \sigma \left\{ W_1 \left[W_0(F_{\text{avg}}^c) \right] + W_1 \left[W_0(F_{\text{max}}^c) \right] \right\}, \quad (1)$$

式中: σ 为激活函数;MLP为多层感知机;AvgPool(F)为平均池化;MaxPool(F)为最大池化; F_{avg}^c 为平均池化特征; F_{max}^c 为最大池化特征; W_0 为 $\frac{C}{r} * C$ 的矩阵, W_1 为 $C * \frac{C}{r}$ 的矩阵,其中 C 为通道数, r 为缩减率。

与通道注意力差异在于,空间注意力更关注的是目标的位置,首先在通道维上应用平均池化和最大池化得到两个不同的特征,将两者相连接实现一个有效的特征表达,然后在级联的特征图上使用卷积层生成空间注意力图^[10]。其表达式为

$$M_s(F) = \sigma \left\{ f^{7*7} \left[\text{AvgPool}(F); \text{MaxPool}(F) \right] \right\} = \sigma \left[f^{7*7} (F_{\text{avg}}^s; F_{\text{max}}^s) \right], \quad (2)$$

式中: f^{7*7} 为7*7的卷积运算; F_{avg}^s 为平均池化特征;

F_{max}^s 为最大池化特征。

2.3 多路聚合网络

网络中信息的流通过程很重要,因此YOLOv4采用了空间金字塔池化层(SPP)+路径聚合网络(PANet)的结构^[11]。SPP结构可以增加模型感受野,有效分离出上下文的显著特征。图2为多路聚合网络结构图,本文改进PANet,提取主干网络中4个特征层信息,首先对其中一个特征层做上采样操作,经过一个5*5卷积层后,再做一次上采样操作,进行输出。然后对经过上采样的特征做下采样处理,同时将第一个特征层进行一次卷积后与其相连接,最后进行输出。原方法只是提取3个特征层信息,不能有效提取浅层特征信息,本文提出方法提取了4个特征层信息,可以有效提取浅层特征信息,更准确地定位飞机目标。

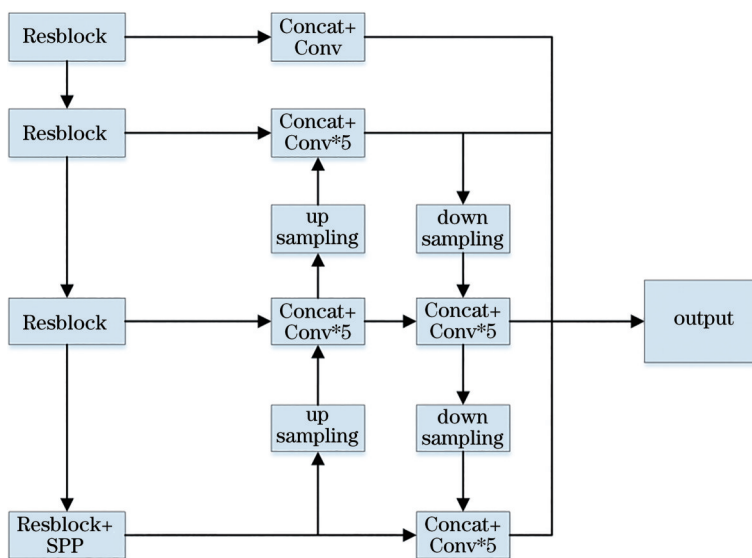


图 2 多路聚合网络结构图

Fig. 2 Structure diagram of multipath aggregation network

2.4 平滑标签

在遥感图像的飞机目标检测中加入平滑标签^[12-13],可以有效避免模型训练过程中过度依赖标签的情况,对分类准确度做了惩罚,使所训练模型的泛化能力得到显著提高,避免出现过拟合。 Y 为平滑标签后得到的一位有效编码^[13],其表达式为

$$Y = Y_i * (1.0 - V_{LS}) + \frac{V_{LS}}{N}, \quad (3)$$

式中: Y_i 为一位有效编码; V_{LS} 为平滑标签设定值; N 为分类数。

本文主要检测飞机目标,那么其分类只有两个:一个是飞机目标;一个不是飞机目标,分别用1和0表示,即一位有效编码 Y_i 的值为[1,0],经过实验证明将平滑标签值设为0.005时得到的模型最优,因此平滑标签后得到的一位有效编码的 Y 值为[0.9975,0.0025]。

对平滑标签取值实验结果对比如表1所示,表

表 1 平滑标签取值结果对比

Table 1 Results comparison of different smoothed label values

V_{LS}	R_{AP}
0	0.933
0.003	0.949
0.005	0.956
0.007	0.936
0.01	0.922

中 V_{LS} 表示平滑标签的取值, R_{AP} 表示平均精确率。

由表中数据可以看出, V_{LS} 为0.005时, R_{AP} 最大,得到的模型最优。

2.5 数据增强

遥感图像中飞机目标位置是固定的,每次训练都是同样的信息,但实际测试时则是多样化的。本文引入数据增强算法改变图像的亮度、饱和度、尺寸大小、位置变换等使图像更加丰富。图3为数据原图像和增强后图像,增强后图像是由原图像水平

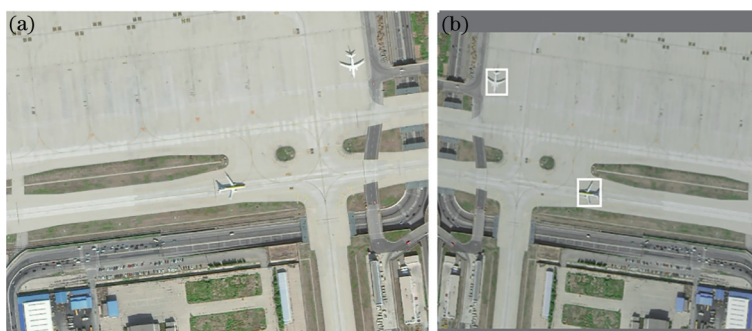


图 3 数据增强对比。(a)原图像;(b)增强后图像

Fig. 3 Data enhancement comparison. (a) Original image; (b) enhanced image

翻转得到,使原飞机目标位置发生变化,增强后的图像与原图像相比亮度变暗,同时对图像进行缩小处理,虽然增强后图像亮度、位置等发生了变化,但本质上仍是飞机。

3 分析与讨论

实验环境:实验是基于 Pytorch 框架搭建,使用 NVIDIA DGX Station 服务器的 Linux 系统 GPU 为 Tesla V100 版本,CUDA10.1 版本。

本文使用的遥感数据集是由武汉大学标注的 RSOD^[14]和中国科学院西安光学精密机械研究所标注的 HRRSD^[15]数据集,其中 RSOD 数据集共 446 张图像,4993 个飞机目标,HRRSD 数据集包括 13 个类别,共 21763 张图像,选取其中飞机目标图像共 1505 张。

实验时,对于 RSOD 数据集,按照 9:1 的比例随机划分数据集,即选取 90% 的图像训练 10% 的图像进行测试,对于 HRRSD 数据集则是按照 8:2 的比例划分。训练时,使用 coco 数据集模型为预训练模型,迭代周期(epoch)设为 120,首先对前 60 个 epoch 进行冻结训练,批尺寸(batch size)设置为 4,学习率为 0.001,然后再对后 60 个 epoch 进行解冻训练,解冻训练 batch size 设置为 2,学习率为 0.0001。

实验的评价标准主要有精确率^[5-6]、召回率^[5-6]、调和平均^[5]以及平均精确率^[5-6],计算公式为

$$P = \frac{T_p}{T_p + F_p}, \quad (4)$$

$$R = \frac{T_p}{T_p + F_n}, \quad (5)$$

$$F_1 = \frac{2PR}{P + R}, \quad (6)$$

$$R_{AP} = \int_0^1 PRdR, \quad (7)$$

式中: P 为精确率; R 为召回率; T_p 为预测目标是正样本、实际是正样本的个数; F_p 为预测目标是正样本、实际是负样本的个数; F_n 为预测目标是负样本、实际是正样本的个数; F_1 为 P 和 R 的调和平均; R_{AP} 为平均精确率。

提出算法在 RSOD 数据集中测试集包括 552 架飞机,测试结果对比如表 2 所示,表中 FPS 为每秒帧传输帧数。

由表 2 中结果可以看出,提出方法的召回率 R 为 0.918, F_1 为 0.94, 平均精确率 R_{AP} 为 0.967, 均高

表 2 RSOD 测试结果对比

Table 2 Comparison of RSOD test results

Algorithm	P	R	F_1	R_{AP}	FPS
SSD	0.979	0.699	0.82	0.890	44.3
YOLOv4	0.952	0.873	0.91	0.933	34.2
YOLOv4-mobilenet	0.948	0.862	0.90	0.930	59.7
Proposed	0.967	0.918	0.94	0.967	31.8

于对比方法,其中提出算法的 R_{AP} 比 SSD 算法高 0.077, 比 YOLOv4 算法高 0.034, 比 YOLOv4-mobilenet 算法高 0.037。

在 HRRSD 数据集中测试集包括 984 架飞机,测试结果对比如表 3 所示。

表 3 HRRSD 测试结果对比

Table 3 Comparison of HRRSD test results

Algorithm	P	R	F_1	R_{AP}	FPS
SSD	0.989	0.905	0.95	0.987	44.3
YOLOv4	0.992	0.986	0.99	0.990	35.5
YOLOv4-mobilenet	0.991	0.983	0.99	0.988	62.9
Proposed	0.991	0.990	0.99	0.993	31.2

由表 3 中结果可以看出,提出方法的召回率 R 为 0.990, 平均精确率 R_{AP} 为 0.993, 均高于对比方法,其中提出算法的 R_{AP} 比 SSD 算法高 0.006, 比 YOLOv4 算法高 0.003, 比 YOLOv4-mobilenet 算法高 0.005。

图 4 为可视化结果对比,通过对 RSOD 数据集图像中飞机目标使用不同算法进行测试对比。

图中 P 为检测到飞机的精确率,可以看出,SSD 算法在该图像中存在 2 个漏检和 3 个误检现象,故其平均精确率相对较低,YOLOv4 方法存在 1 个漏检现象,YOLOv4-mobilenet 存在 2 个漏检现象,而提出算法不存在漏检和误检现象,因此可以证明提出算法的优越性。

为了证明提出算法应用各部分模块在遥感飞机检测精度提升方面的有效性,对 RSOD 数据集做消融实验如表 4 所示,表中 SL 为平滑标签,MPANet 为多路聚合网络,CBAM 为融合注意力机制。

从表 4 结果可以看出,原方法平均精确率为 0.933, 单独增加 SL 平均精确率为 0.956, 单独增加 MPANet 平均精确率为 0.953, 单独增加 CBAM 平均精确率为 0.953, 同时增加 SL、MPANet 和 CBAM 平均精确率为 0.967, 相比原算法平均精确率提升了 0.034, 因此可以证明提出算法的优越性。

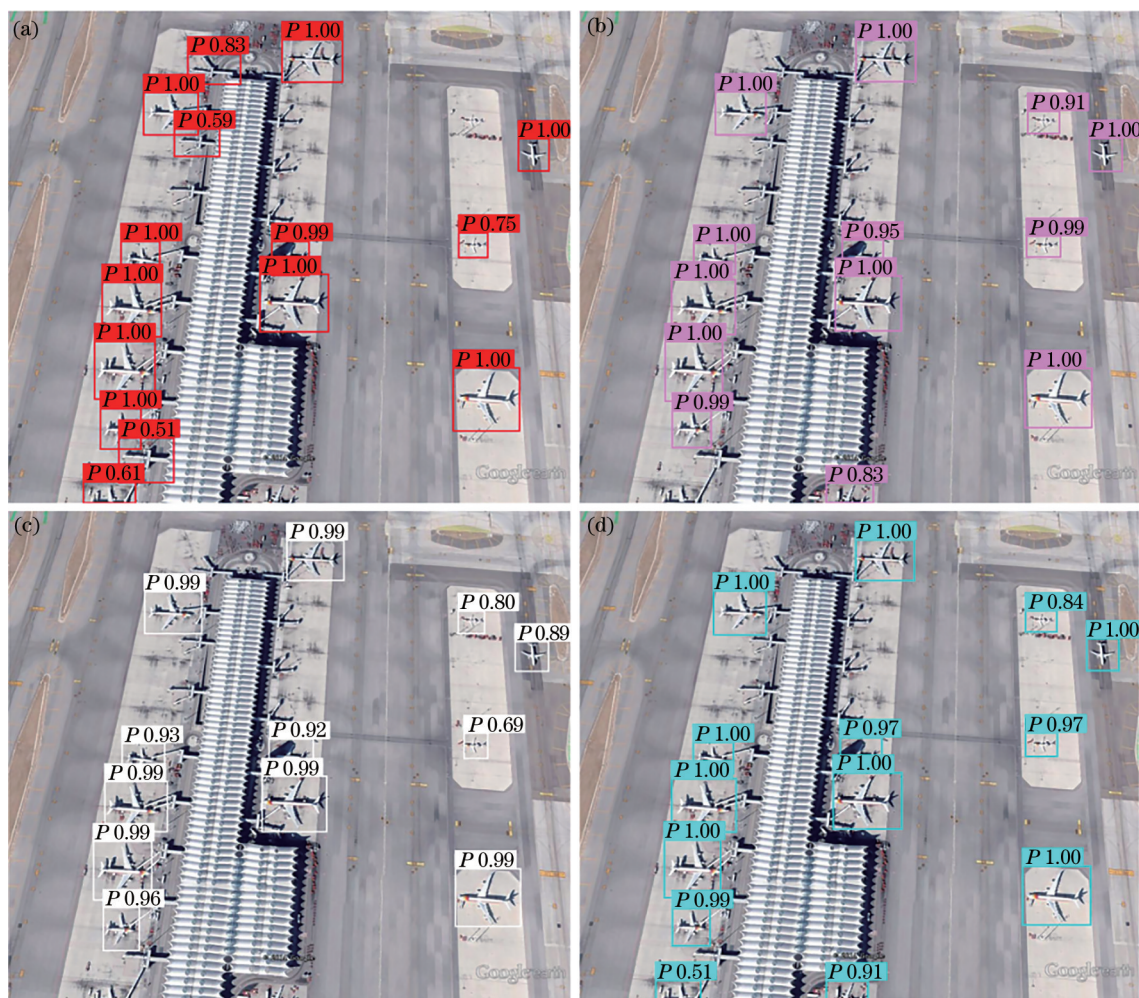


图 4 不同算法可视化结果对比。(a) SSD算法;(b) YOLOv4算法;(c) YOLOv4-mobilenet算法;(d) 提出算法
 Fig. 4 Comparison of visual results of different algorithms. (a) SSD algorithm; (b) YOLOv4 algorithm; (c) YOLOv4-mobilenet algorithm; (d) proposed algorithm

表 4 消融实验

Table 4 Ablation experiment

Algorithm	P	R	F_1	R_{AP}
YOLOv4	0.952	0.873	0.91	0.933
YOLOv4+SL	0.966	0.896	0.93	0.956
YOLOv4+MPANet	0.974	0.893	0.93	0.953
YOLOv4+CBAM	0.963	0.900	0.93	0.953
YOLOv4+SL+MPANet+CBAM	0.967	0.918	0.94	0.967

4 结 论

在 YOLOv4 网络的基础上,提出了基于平滑标签技术和多路径聚合网络的遥感飞机检测方法,首先在主干网络中的每一个残差块中加入融合注意力机制,关注目标区域。然后改进路径聚合网络在原网络的基础上多提取一层特征,将提取的特征层与其他 3

个特征层相连接做融合操作,可以更准确地对飞机进行定位。最后在算法中加入平滑标签技术避免训练模型过拟合,加入数据增强技术使图像更加丰富。检测结果相比于其他算法,有效提高了检测的召回率和平均精确率。该算法不足之处是相对于其他算法速度较慢,后续将使用轻量型网络提高检测速度,并加入注意力增强算法增强网络中的注意力机制。

参 考 文 献

- [1] 戴媛, 易本顺, 肖进胜, 等. 基于改进旋转区域生成网络的遥感图像目标检测[J]. 光学学报, 2020, 40(1): 0111020.
Dai Y, Yi B S, Xiao J S, et al. Object detection of remote sensing image based on improved rotation region proposal network[J]. Acta Optica Sinica, 2020, 40(1): 0111020.
- [2] 董永峰, 仇长涛, 汪鹏, 等. 基于深度学习的光学遥感图像飞机检测算法[J]. 激光与光电子学进展, 2020, 57(4): 041007.
Dong Y F, Zhang C T, Wang P, et al. Airplane detection of optical remote sensing images based on deep learning[J]. Laser & Optoelectronics Progress, 2020, 57(4): 041007.
- [3] 汪亚妮, 汪西莉. 基于注意力和特征融合的遥感图像目标检测模型[J]. 激光与光电子学进展, 2021, 58(2): 0228003.
Wang Y N, Wang X L. Remote sensing image target detection model based on attention and feature fusion[J]. Laser & Optoelectronics Progress, 2021, 58(2): 0228003.
- [4] 徐俊峰, 张保明, 余东行, 等. 多特征融合的高分辨率遥感影像飞机目标变化检测[J]. 遥感学报, 2020, 24(1): 37-52.
Xu J F, Zhang B M, Yu D H, et al. Aircraft target change detection for high-resolution remote sensing images using multi-feature fusion[J]. Journal of Remote Sensing, 2020, 24(1): 37-52.
- [5] Huang W, Li G Y, Chen Q Q, et al. CF2PN: a cross-scale feature fusion pyramid network based remote sensing target detection[J]. Remote Sensing, 2021, 13(5): 847.
- [6] Qiu H Q, Li H L, Wu Q B, et al. A2RMNet: adaptively aspect ratio multi-scale network for object detection in remote sensing images[J]. Remote Sensing, 2019, 11(13): 1594.
- [7] Chen C Y, Gong W G, Chen Y L, et al. Object detection in remote sensing images based on a scene-contextual feature pyramid network[J]. Remote Sensing, 2019, 11(3): 339.
- [8] Wang C Y, Liao H Y M, Wu Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 390-391.
- [9] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23) [2021-04-05]. <https://arxiv.org/abs/2004.10934>.
- [10] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [11] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8759-8768.
- [12] Yang X, Yan J C. Arbitrary-oriented object detection with circular smooth label[M]//Vedaldi A, Bischof H, Brox T, et al. Computer vision-ECCV 2020. Lecture notes in computer science. Cham: Springer, 2020, 12353: 677-694.
- [13] Müller R, Kornblith S, Hinton G E. When does label smoothing help?[EB/OL]. (2019-06-06) [2021-05-06]. <https://arxiv.org/abs/1906.02629>.
- [14] Long Y, Gong Y P, Xiao Z F, et al. Accurate object localization in remote sensing images based on convolutional neural networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(5): 2486-2498.
- [15] Zhang Y L, Yuan Y, Feng Y C, et al. Hierarchical and robust convolutional neural network for very high-resolution remote sensing object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(8): 5535-5548.