

基于注意力与特征融合的光学遥感图像飞机目标检测

兰旭婷, 郭中华, 李昌昊

引用本文:

兰旭婷, 郭中华, 李昌昊. 基于注意力与特征融合的光学遥感图像飞机目标检测[J]. 液晶与显示, 2021, 36(11): 1506–1515.

LAN Xu-ting, GUO Zhong-hua, LI Chang-hao. Attention and feature fusion for aircraft target detection in optical remote sensing images[J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(11): 1506–1515.

在线阅读 View online: <https://doi.org/10.37188/CJLCD.2021-0088>

您可能感兴趣的其他文章

Articles you may be interested in

一种高精度的卷积神经网络安全帽检测方法

High precision detection method of safety helmet based on convolution neural network

液晶与显示. 2021, 36(7): 1018–1026 <https://doi.org/10.37188/CJLCD.2020-0309>

基于注意力掩模融合的目标检测算法

Object detection algorithm based on attention mask fusion

液晶与显示. 2019, 34(8): 825–833 <https://doi.org/10.3788/YJYXS20193408.0825>

改进的多尺度火焰检测方法

Improved multi-scale flame detection method

液晶与显示. 2021, 36(5): 751–759 <https://doi.org/10.37188/CJLCD.2020-0221>

基于改进Faster R-CNN的交通标志检测算法

Traffic sign detection algorithm based on improved Faster R-CNN

液晶与显示. 2021, 36(3): 484–492 <https://doi.org/10.37188/CJLCD.2020-0195>

融合注意力机制的移动端人像分割网络

Mobile-based portrait segmentation network with attention mechanism

液晶与显示. 2020, 35(6): 547–554 <https://doi.org/10.3788/YJYXS20203506.0547>

文章编号:1007-2780(2021)11-1506-10

基于注意力与特征融合的光学遥感 图像飞机目标检测

兰旭婷¹, 郭中华^{1,2*}, 李昌昊¹

(1. 宁夏大学 物理与电子电气工程学院, 宁夏 银川 750021;

2. 宁夏沙漠信息智能感知重点实验室, 宁夏 银川 750021)

摘要: 光学遥感图像受背景复杂度和语义信息量大等影响, 在检测精度和效率上仍然存在一定的不足。本文提出了以 Resnet50 为基础特征提取的 SSD300 网络模型, 加入注意力机制 CBAM 模块和特征融合 FPN 模块, 采用 Soft-NMS 策略选取最终预测框, 对遥感图像飞机目标进行更加有效地检测。最后, 在 2 150 张飞机遥感图像数据集上进行训练, 当 IoU 为 0.5 和 0.75 时平均精度 MAP 达到 92.54% 和 63.44%, 较改进前的算法模型提升了 5.04% 和 11.38%, 检测速度达到 24.1 FPS。实验结果表明, 该方法可以有效提高物体的检测能力, 以及快速、准确地检测机场区域内的飞机物体, 有效降低了飞机物体的漏检率, 提高了检测精度和速度。

关键词: 遥感图像; 目标检测; 注意力机制; 特征融合

中图分类号: TP391.41 **文献标识码:** A **doi:** 10.37188/CJLCD.2021-0088

Attention and feature fusion for aircraft target detection in optical remote sensing images

LAN Xu-ting¹, GUO Zhong-hua^{1,2*}, LI Chang-hao¹

(1. School of Physics and Electronic and Electrical Engineering, Ningxia University,
Yinchuan 750021, China;

2. Ningxia Key Laboratory of Desert Information Intelligent Perception,
Yinchuan 750021, China)

Abstract: Optical remote sensing images are affected by the complexity of the background and the large amount of semantic information, and there are still certain deficiencies in detection accuracy and efficiency. This paper proposes an SSD300 network model based on Resnet50 for feature extraction, adding the attention mechanism CBAM module and feature fusion FPN module, and adopting the Soft-NMS strategy to select the final prediction frame to detect aircraft targets in remote sensing images more effectively. Finally, training on 2 150 aircraft remote sensing image data sets, when the IoU

收稿日期: 2021-04-02; **修订日期:** 2021-04-29.

基金项目: 宁夏自然科学基金(No. 2020AAC03026); 宁夏大学研究生创新研究项目(No. GIP2020075)

Supported by Natural Science Foundation of Ningxia Hui Autonomous Region(No. 2020AAC03026);
Ningxia University Graduate Innovation Research Project(No. GIP2020075)

* 通信联系人, E-mail: guozhh@nxu.edu.cn

is 0.5 and 0.75, the average accuracy MAP reaches 92.54% and 63.44%, which are 5.04% and 11.38% higher than the previous algorithm model, and the detection speed reaches 13.4 FPS. Experimental results show that this method can effectively improve the detection ability of objects and quickly and accurately detect aircraft objects in the airport area, effectively reducing the missed detection rate of aircraft objects and improving detection accuracy and speed.

Key words: remote sensing image; target detection; attention mechanism; feature fusion

1 引言

高分辨率光学遥感图像有着丰富的空间信息以及清晰的结构和纹理信息,可以对遥感目标如飞机等进行精确的检测识别^[1]。光学遥感图像的目标检测与识别表示在遥感图像中确定是否包含一个或多个目标物体,并确定目标在图像中的位置与类别^[2]。由于光学遥感图像中的飞机目标类型、尺寸、停放位置等不确定性因素,以及高分辨率光学遥感图像易受光照、天气环境等因素的影响,再加之背景复杂程度大,给高分辨率图像中的目标检测带来了较大的挑战。

为了解决这一挑战,很多学者不断提出了一些传统的新方法。例如,Cheng 等人^[3]提出 HOG 算子提取目标的梯度方向,并通过将训练样本与感兴趣区域匹配来获得飞机目标。但飞机梯度方向特征不是唯一的,导致初步处理中出现大量虚假目标,增加了后续处理的负担。Liu 等人^[4]通过基于大量样本计算平均模板来匹配目标,并根据目标与模板的相似性来识别目标。但是这种方法也需要大量的训练样本,平均模板不能匹配所有的目标。Arandjelovi^[5]提出了一种使用边界描述符为平滑和无纹理图像进行对象匹配的方法,该描述符包括采样边界法线方向的轮廓。等距关键点是通过采样图像边界曲线获得的。然后,利用一个关键点周围的局部轮廓的边界法线来描述该点。该方法在许多经典的目标图像集上表现出很强的鲁棒性。但局部描述对遥感图像中普遍存在的局部噪声和不正确的形状分量更加敏感。Kahaki 等^[6]提出了一种新的基于三重特征提取方法的匹配技术和相异性指标,根据测量两个兴趣点之间的路径,提取特征之间的相异性。该方法可以很好地检测目标图像中的对应点,还可以预测目标图像中的遗漏点。不仅依赖于局部点的信息,而且还利用这些点之间的空间信息。但对

拍摄到的高分辨率光学遥感图像来说,在采集点间的特征上有一定的时间和精度限制。

近年来,遥感图像领域广泛应用了深度学习的方法,尤其在目标检测任务上该方法明显优于传统的固定模型的方法。现阶段,比较成熟的有基于候选区域的目标检测器(两阶段)和单次目标检测器(单阶段)这两类目标检测算法。基于 R-CNN (Region with CNN feature)^[7]、Fast-RCNN^[8]与 Faster-RCNN^[9]的两阶段目标检测算法,采用 RPN 网络提取候选框的方式,能在一定程度上提高精度,但由于每个候选框的位置都需要精修导致检测速度过慢。Joseph Redmon^[10]等人提出了一种新的目标检测方法 YOLO,该方法利用整个图作为网络的输入,直接在输出层回归 bounding box 的位置和其所属类别,实现端对端的模型优化。但该方法每个方格只能预测一类物体,对于小目标检测困难。YOLO 检测速度相比 Faster-RCNN 提升很大,但同时也带来了检测精度下降的问题。YOLOv2^[11]和 YOLOv3^[12]是 YOLO 的后续改进方法。YOLOv2 借鉴了 Faster-RCNN 的思想,引入了 anchor 机制,比其他模型在各种检测数据集上的速度更快,但由于 YOLO 将图像分割成 $S \times S$ 的网格,导致精度不高的问题仍然存在。SSD (Single shot Multibox Detector)^[13]算法,结合了 YOLO 中的回归思想和 Faster-RCNN 中的 Anchor 机制,并使用多尺度的特征图进行回归,既保持了 YOLO 速度快的特点,也保证了 Faster RCNN 精度高的优势^[14]。该方法在自然图像中检测精度以及速度上都有一定的提升效果,但在纹理信息弱、背景复杂的光学遥感图像飞机目标的检测中,效果并不是很理想。

因此,为了解决光学遥感图像在复杂环境中小飞机目标检测精度及效率低的问题,本文以 SSD300 为基本算法,参考 Resnet50^[15]建立了光学遥感图像数据库,对网络模型进行了改进和优

化。最终有效提高了弱小物体的表达能力,提高了检测精度和速度,具有一定的通用性。

2 SSD 检测网络

2.1 SSD 算法

SSD 算法是一种目标检测算法,相比于 YOLO 在检测精度上有更大的提升。SSD 的算法框架图如图 1 所示。SSD 共有 11 个 block,相比于 VGG16^[16] 网络改变了第 5 个 block 的第 4 层,并将 VGG16 中的第 6、7、8 卷积层去掉,将 FC7 改成卷积层 Conv7,并添加额外的 Conv8、Conv9、Conv10、Conv11 卷积层用于目标检测。其中:Conv4_3、FC7 位于浅层但携带的细节信息更完整,用于小目标物体检测。对于 Conv8_2、Conv9_2、Conv10_2、Conv11_2,随着特征层的增高,感受野也逐渐增大,因此这些层含有更多的语义信息,用来检测大目标物体。

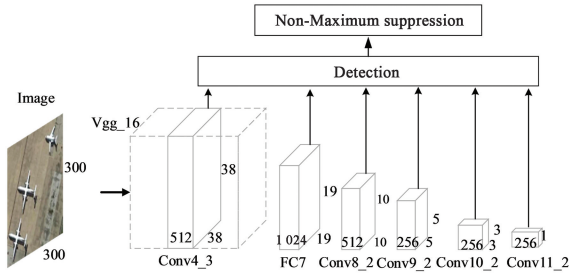


图 1 SSD_VGG16 网络结构

Fig.1 SSD_VGG16 Network structure

2.2 Resnet50+SSD 网络结构

在 SSD 网络结构中是以 VGG16 作为特征提取的骨干网络,而在本文中采用 Resnet50 替换原来的特征提取骨干网络,如图 2 所示。对于 Resnet50 模型,只保留了前 4 层用于特征提取,舍弃了 Conv5_x 及以后的层,并额外添加了 5 个层,并将生成更多预测特征层。

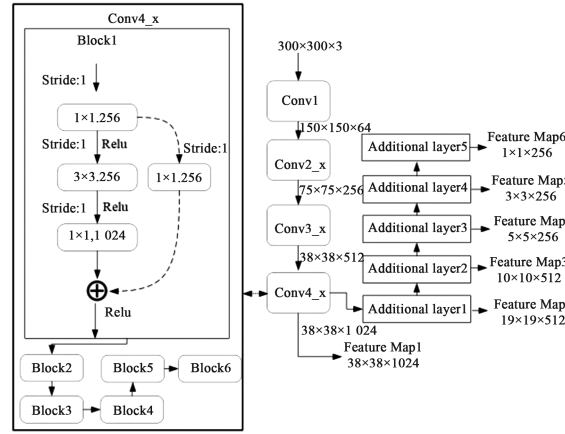


图 2 Resnet50+SSD 网络结构

Fig.2 Resnet50+SSD network structure

特征信息进行融合的网络,具体网络结构如图 3 所示。该结构表示:自底向上,自顶而下,横向连接。通过高层特征进行上采样与经过卷积操作后的低层特征进行自顶向下的连接,大幅度提升了小物体检测的性能。本文将从 6 个检测层分别进行特征融合,如图 4 所示。

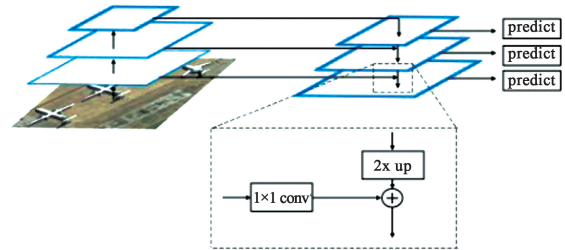


图 3 特征金字塔网络

Fig.3 Feature pyramid networks

3 改进 SSD 网络结构

3.1 FPN 特征融合

在神经网络结构中由于小目标本身具有的像素信息非常少,卷积操作容易导致小目标信息的丢失,因此,采用特征金字塔网络 (Feature Pyramid Networks),它是一种将多个检测层的

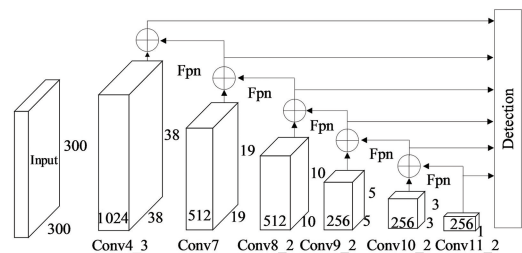


图 4 带有特征融合的网络结构

Fig.4 Network structure with feature fusion

3.2 增加通道与空间注意力机制模块

神经网络中每经过不同的卷积核都会产生不同的通道信息,而这些通道所包含的信息量以及重要程度也不同。对于原先的网络,特征层会默

认特征层的每个通道上的信息重要性是相同的。故在改进的算法中,引入了 CBAM^[17] (Convolutional-Block-Attention Module),即通道注意机制模块(Channel Attention Module)和空间注意力模块(Spatial Attention Module),如图 5 所示。本文在进行 FPN 之后的特征检测层层上分别加入双通道注意力模块 CBAM,不仅可以很好地集中需要的信息,忽略与检测目标无关的边界信息,还可以提高网络的特征表达能力。

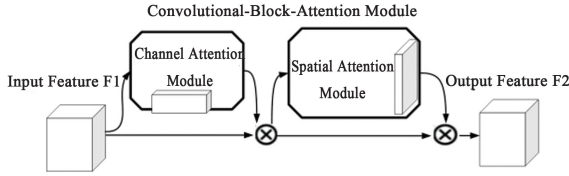


图 5 卷积块注意力模块

Fig.5 Convolution block attention module

3.3 Soft-NMS

非极大值抑制 (Non-maximum suppression, NMS)^[18],顾名思义就是抑制一些不是极大值的值。在网络检测与回归过程中,NMS 具体计算如式(1)所示。

$$S_i = \begin{cases} S_i, \text{iou}(M, b_i) < N_t \\ 0, \text{iou}(M, b_i) \geq N_t \end{cases}, \quad (1)$$

式中: S_i 、 M 、 b_i 表示预测框; N_t 表示人为设定的阈值; iou 表示两个预测框之间交集与并集的比值。

在上述的 NMS 算法当中,如图 6 所示,一方面若存在两个重叠区域较大的目标检测框,NMS 会去除得分较低的蓝框而保留较高分的红框,此时物体的检测召回率就比较低;另一方面,传统的 NMS 的阈值也不容易确定,过大容易产生误检。而 Soft-non maximum suppression(Soft-NMS)^[19] 很好地解决了这个问题,两种改进形式。如公式(2)、(3)所示。

$$S_i = \begin{cases} S_i, \text{iou}(M, b_i) < N_t \\ S_i (1 - \text{iou}(M, b_i)), \text{iou}(M, b_i) \geq N_t \end{cases}, \quad (2)$$

$$S_i = S_i e^{-\frac{\text{iou}(M, b_i)^2}{\sigma}}, \forall b_i \in D. \quad (3)$$

3.4 调整特征层上的先验框比例

在原先 SSD300 模型当中,每个特征层上的先验框的宽高比为 $1:1$ 、 $1:2$ 、 $1:3$ 、 $2:1$ 、 $3:1$

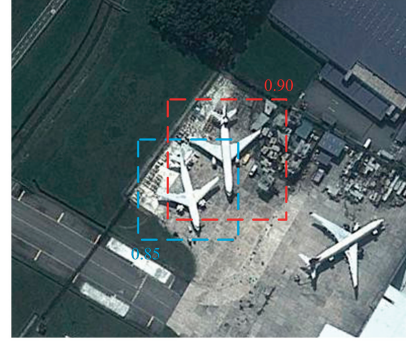


图 6 重叠度超过设定阈值参照图

Fig.6 Degree of overlap exceeds the set threshold

这几种形式。在多目标的数据集如 Pascal VOC^[20] 等数据集物体类别较多,但对于遥感图像的飞机目标来说,此时的先验框比例不再适合。因此按照飞机的尺寸大小将原有的先验框比例调整为 $1:1$ 、 $1:2/5$ 、 $1:3/5$ 、 $2/5:1$ 、 $3/5:1$ 。

3.5 改进后的网络结构

改进后的网络结构如图 7 所示:先将不同尺度的特征层进行融合(FPN)得到更多有效的特征图信息,再经过双注意力模块后得到更具有判断能力的特征图,最后将获取到含有丰富语义信息的 6 个特征图送入检测层。

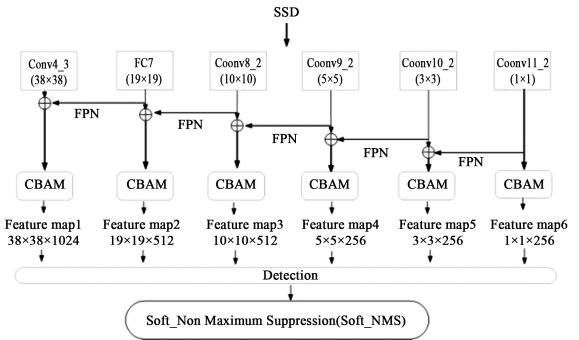


图 7 改进的网络结构

Fig.7 Improved network structure

3.6 损失函数的选择

对于目标检测回归损失函数,我们依然延续 SSD300 算法中的 Smooth L1 Loss 对改进后的算法进行优化处理。在预测层中预测了每个特征图上的点的矩形框位置信息和目标物体的类别信息。即该点的损失函数就等于矩形框位置损失与物体类别损失之和。具体可以表示为:

$$L(x, c, l, g) = \frac{1}{N} (L_{\text{conf}}(x, c) + \alpha L_{\text{loc}}(x, l, g)), \quad (4)$$

其中: N 为样本数, 若 N 为 0, 则将损失设为 0; x 为当前预测框的类别信息; c 代表预测类别信息; l 为预测框的位置信息; g 代表真实标注框的值; L_{conf} 为类别损失函数, L_{loc} 位置损失函数; 而参数 α 用于调整 L_{conf} 与 L_{loc} 之间的比例, 一般默认 $\alpha=1$ 。损失函数具体计算公式如式(5)、(6)所示:

$$L_{\text{loc}}(x, l, g) = \sum_{i \in \text{Pos}} \sum_{m \in \{cx, cy, w, h\}} \cdot x_{ij}^k \cdot \text{smooth}_{L1}(l_i^m - \hat{g}_j^m), \quad (5)$$

$$L_{\text{conf}}(x, c) = - \sum_{i \in \text{Pos}} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in \text{Neg}} \log(\hat{c}_i^0),$$

$$\text{where } \hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)}, \quad (6)$$

其中: x_{ij}^k 代表第 i 个预测框与第 j 个真实框关于类别 k 是否匹配, 若 x_{ij}^k 值为 0 表示不匹配, 若 x_{ij}^k 值为 1 则表示类别匹配; l_i^m 表示预测框; $\hat{g}_j^m$ 表示真实框; $\hat{c}_i^0$ 表示预测框无物体且类别为背景预测框的概率, 预测为背景的概率越大则顺势越小; $\hat{c}_i^p$ 根据 Softmax 函数计算的概率值, 概率越大损失越小。

4 实验结果与分析

4.1 实验条件

为评估本文所改进的模型算法的检测性能, 公开数据集 NWPU-RESISC45 和武汉大学团队含有飞机信息的遥感图像共 2 150 张, 其中按照 7:3 生成 1 505 张为训练集, 645 张为测试集, 并按照 Pascal VOC 数据集格式进行标注。网络输入前的图像尺寸为 256×256 , $1\,280 \times 659$, $1\,372 \times 941$, 经过网络输入后均为 300×300 。并且整个



图 8 飞机样本例图

Fig.8 Aircraft sample illustration

实验在 Ubuntu16.04 操作系统、Pytorch1.6 环境下完成, 并使用单个 NVIDIA GeForce RTX2080Ti 型号的 GPU 加速训练模型。

具体数据集的分配如表 1 所示, 数据集样本例图如图 8 所示。

表 1 数据集的信息统计

Tab.1 Information statistics of the data set

数据库	图像	尺寸
Train Set	1 505	300×300
Test Set	645	300×300

4.2 实验流程及训练参数

基于双通道的注意力机制及锚框比例缩放为适用飞机形状比例的原来的 1/5 模型, 首先在以 Resnet50 作为 SSD 特征提取骨干网络结构上, 用上述标注好的数据集进行训练。批处理操作 Batch size 设置为 4, 动量因子系数设定为 0.9, 权重衰减系数设为 0.000 5, 初始学习率设为 0.000 5。

4.3 评价指标

本文为更好地评估改进后的网络模型, 采用目标检测领域中通用的平均精度均值 (Mean Average Precision, MAP)^[21] 和每秒帧率 (Frame Per Second, FPS) 作为本次算法改进的评价标准。其中平均精度的定义为: $AP = \int_0^1 p(r) dr$, 即为召回率 (R) 和精度 (P) 所绘制的曲线与坐标轴所围成的面积, 所有类别的 AP 进行平均得到 MAP。其中, 被正确识别为飞机目标的样本称为 TP, 没有被正确识别出飞机目标的样本称为 FP, 将飞机的背景正确判别为背景的样本称为 TN, 将飞机目标误判为背景的样本称为 FN。则召回率与精度的计算公式如式(7)、(8)所示:

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (8)$$

4.4 算法对比实验

设定了 10 组实验与未改进的模型进行对照: 第 1 组为最初网络结构实验结果; 剩余组是在开始的网络基础上分别加入不同的改进内容后训练的结果, 具体数据对比如表 2 所示。

表 2 不同检测方法结果对比

Tab.2 Comparison of results of different detection methods

方 法	网络结构	MAP@0.5/%	MAP@0.75/%
无	Resenet50_SSD	87.50	52.06
CBAM	Resenet50_SSD	87.42	53.63
先验框比例	Resenet50_SSD	88.81	53.51
Soft-NMS	Resenet50_SSD	88.65	53.73
CBAM+先验框比例	Resenet50_SSD	88.73	54.93
CBAM+Soft-NMS	Resenet50_SSD	88.07	54.17
特征融合	Resenet50_SSD	91.89	62.08
CBAM+特征融合	Resenet50_SSD	91.65	61.28
CBAM+先验框比例+Soft-NMS	Resenet50_SSD	88.71	54.28
CBAM+先验框比例+特征融合	Resenet50_SSD	91.69	62.07
CBAM+先验框比例+Soft-NMS+特征融合	Resenet50_SSD	92.54	63.44

实验结果表明,加入 CBAM、Soft-NMS、特征融合及调整先验框后的模型检测效果明显优于以 resnet50 为骨干网络的 SSD 模型,在 IoU(Intersection over Union)值为 0.5 的情况下检测效果达到 5.04%,IoU 为 0.75 时达到 11.38%。对于单独加入注意力机制,在 IoU 为 0.5 时下降 0.08%,IoU 为 0.75 时提升 1.57%;采用 Soft-NMS 算法在 IoU 为 0.5 时提升 1.15%,IoU 为 0.75 时提升 1.67%;调整先验框比例,在 IoU 为 0.5 和 0.75 时分别提升 1.31%、1.45%;特征金字塔的融合方式效果分别提升 4.39%、10.02%。很明显,特征层间的信息融合大幅提高了物体检测精度。

4.5 算法模型对比实验与结果分析

图 9 所示为 SSD300、YOLOV3、Faster-RCNN 以及改进 SSD 四种算法分别在遥感图像飞机目标数据集训练 1 000 次的精度均值对比

图。表 3 为这 4 种网络模型训练 1 000 次后的具体精度均值对比。

如图 10 所示,10(a)为原始图像,均来自不同于训练集的数据集 Google earth 光学遥感图像;

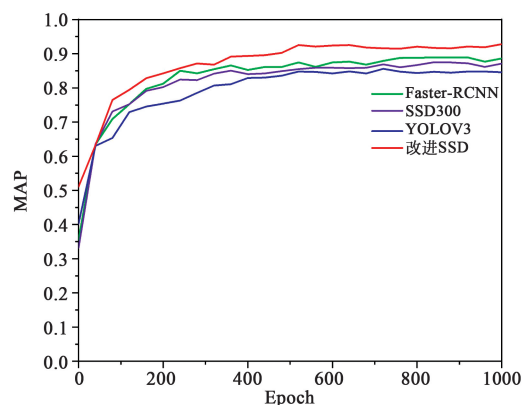


图 9 4 种目标检测算法模型的检测精度对比

Fig.9 Comparison of detection accuracy of four target detection algorithm models

表 3 不同模型精度结果对比

Tab.3 Comparison of accuracy results of different models

Method	Backbone	MAP@0.5/%	MAP@0.75/%	FPS/(frame · s ⁻¹)
YOLOV3	DarkNet-53	85.58	40.19	21.8
Faster-RCNN	VGG_16	88.9	59.91	4.3
SSD300	VGG_16	87.5	52.06	24.6
改进 SSD	ResNet50	92.54	62.11	24.1



(a) 原始图像
(a) Original image



(b) SSD300 模型
(b) SSD300 model



(c) YOLOV3 模型
(c) YOLOV3 model



(d) Faster-RCNN 模型
(d) Faster-RCNN model



(e) 改进 SSD 模型
(e) Improve SSD model

图 10 不同模型实验检测结果对比
Fig.10 Comparison of test results of different models

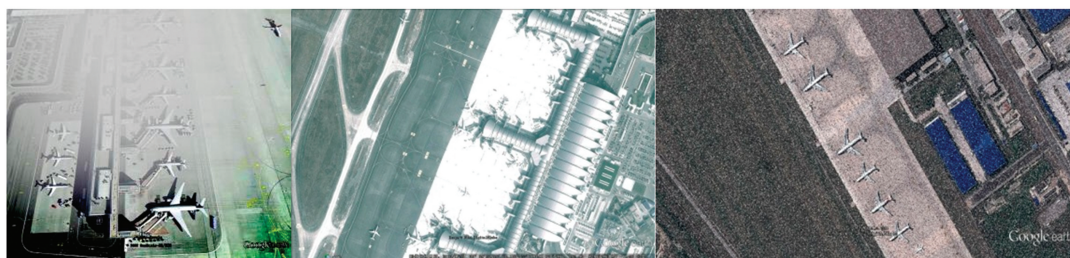
10(b)为 SSD300 目标检测算法检测结果图像;10(c)为 YOLOV3 目标检测算法检测结果图像;10(d)为 Faster-RCNN 目标检测算法检测结果图像;10(e)为改进 SSD 目标检测算法检测结果图像。

在图 10(a)中,有 3 幅具有代表性的图。其中:左图为环境复杂度高,中间的图为多尺度、小目标聚集,右图为环境复杂度高并且小目标较多。通过以上 4 种算法模型检测结果可以得出,在环境复杂度高、小目标聚集的情况下,除了 Faster-RCNN 与改进的 SSD 模型可以很好地检测出所有的飞机目标,其余两个模型都存在漏检问题;在多尺度小目标聚集的情况下,只有改进的 SSD 模型存在漏检一个飞机目标,其余 3 个模型出现大量检测目标丢失,但 Faster-RCNN 检测结果优于 SSD300 和 YOLOV3 模型;在小目标多、环境非常复杂的条件下,更加体现出了改进模型的优势。其中 SSD300 和 YOLOV3 漏

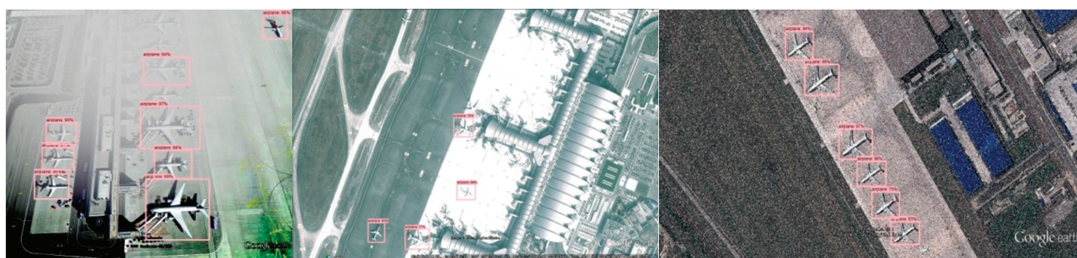
检情况明显,并且检测精度不是很高。对于 Faster-RCNN 与改进的模型,都检测出了飞机目标,但 Faster-RCNN 的检测速度却远低于改进的模型检测速度。因此,对比以上结果发现,对背景复杂度高、小目标繁多的遥感图像单目标实现高精度的实时检测,改进后的模型很大程度上优于 Faster-RCNN、SSD300 和 YOLOV3 模型。

4.6 算法扩展实验与结果分析

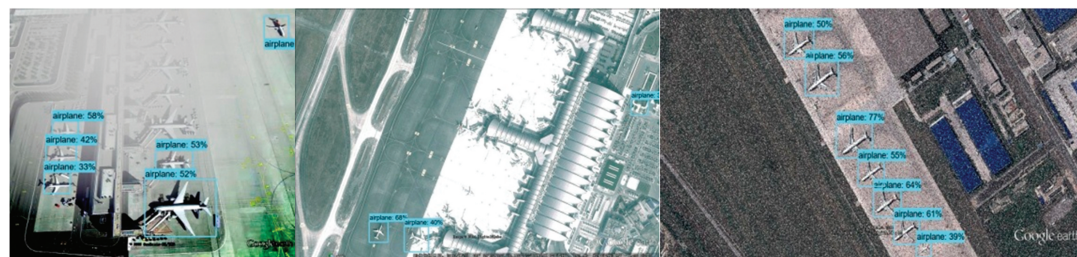
为证明改进 SSD 目标算法的可行性及通用性,本文设计了扩展实验,分别对雾天条件下的遥感图像、曝光环境下的遥感图像和噪声污染下的遥感图像进行检测,如图 11 所示。实验证明,改进后的 SSD 目标检测算法模型不仅精度最高而且可以实时检测,并且适用于不同复杂程度环境下的光学遥感图像飞机目标的检测,具有一定的通用性。



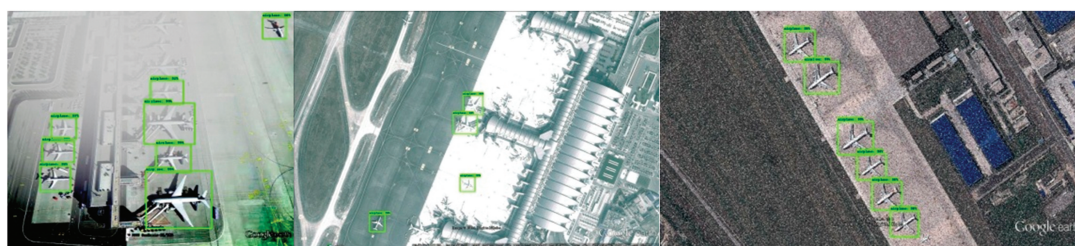
(a) 原始图片
(a) Original image



(b) SSD300 模型
(b) SSD300 model



(c) YOLOV3 模型
(c) YOLOV3 model



(d) Faster-RCNN 模型

(d) Faster-RCNN model



(e) 改进 SSD 模型

(e) Improve SSD model

图 11 复杂环境下实验对比检测图

Fig.11 Experimental comparison test in complex environment

5 结 论

本文针对经典的 SSD300 算法在光学遥感图像中存在飞机目标检测精度差以及检测速度慢的问题,提出了嵌入注意力机制模块、FPN 特征融合、优化先验框的宽高比例以及利用 Soft-NMS 算法筛选出更加有效的检测框的方案。从算法改进的实验结果分析来看,第一,在不同卷积层引入通道、空间注意力机制后的模型检测效率优于 SSD 算法,并且针对单一目标检测效果有 1.5% 的提升;第二,加入特征金字塔,将不同尺度的特征进行融合后增强了特征信息,得到更加具有识

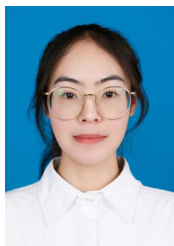
别能力的特征图,精度明显提升。第三,利用 Soft-NMS 较于传统 NMS 算法保留了实际检测到的候选框,提升了召回率。最后,将所有改进的方法融合训练后精度达到 92.54%,检测速度达到 24.1 FPS。可见,改进的网络模型达到了实时检测以及高精度识别飞机目标。为判断改进模型能否在复杂环境下对光学遥感图像飞机目标的检测适用性,做了拓展实验。实验结果表明,改进后的 SSD 模型在不同环境下对飞机的检测有很好的效果,并且应用广泛。

参 考 文 献:

- [1] 祝文韬,谢宝蓉,王琰,等.光学遥感图像中的飞机目标检测技术研究综述[J].计算机科学,2020,47(S2):165-171,182.
ZHU W T, XIE B R, WANG Y, *et al.* A review of aircraft target detection technology in optical remote sensing images [J]. *Computer Science*, 2020, 47(S2): 165-171, 182. (in Chinese)
- [2] 李文斌,何冉.基于 FCN 与 CNN 的遥感影像飞机目标检测方法[J/OL].计算机工程:1-11[2021-02-01].<https://www.doc88.com/p-71161845887317.html>.
LI W B, HE R. Aircraft detection models of remote sensing images based on FCN and CNN [J/OL]. *Computer Engineering*: 1-11[2021-02-01]. <https://www.doc88.com/p-71161845887317.html>. (in Chinese)
- [3] CHENG G, HAN J W, GUO L, *et al.* Object detection in remote sensing imagery using a discriminatively trained

- mixture model [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2013, 85: 32-43.
- [4] LIU G, SUN X, FU K, *et al.* Aircraft recognition in high-resolution satellite images using coarse-to-fine shape prior [J]. *IEEE Geoscience and Remote Sensing Letters*, 2013, 10(3): 573-577.
- [5] ARANDJELOVI O. Object matching using boundary descriptors [C]//*Proceedings of the British Machine Vision Conference*. Surrey, UK: BMVA Press, 2012.
- [6] KAHAKI S M M, NORDIN M J, ASHTARI A H, *et al.* Deformation invariant image matching based on dissimilarity of spatial features [J]. *Neurocomputing*, 2016, 175: 1009-1018.
- [7] GIRSHICK R, DONAHUE J, DARRELL T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation [C]//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus: IEEE, 2014: 580-587.
- [8] GIRSHICK R. Fast R-CNN [C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago: IEEE, 2015: 1440-1448.
- [9] REN S Q, HE K M, GIRSHICK R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks [C]//*Proceedings of the 28th International Conference on Neural Information Processing Systems*. Montreal: MIT Press, 2015: 91-99.
- [10] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You only look once: unified, real-time object detection [C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 779-788.
- [11] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017: 7263-7271.
- [12] REDMON J, FARHADI A. YOLOv3: an incremental improvement [J]. *arXiv*:1804.02767, 2018.
- [13] LIU W, ANGUELOV D, ERHAN D, *et al.* SSD: single shot MultiBox detector [C]//*Proceedings of the 14th European Conference on Computer Vision*. Amsterdam, The Netherlands: Springer, 2016: 21-37.
- [14] 董潇潇,何小海,吴晓红,等.基于注意力掩模融合的目标检测算法[J].*液晶与显示*,2019,34(8):825-833.
- DONG X X, HE X H, WU X H, *et al.* Object detection algorithm based on attention mask fusion [J]. *Chinese Journal of Liquid Crystals and Displays*, 2019, 34(8): 825-833. (in Chinese)
- [15] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 770-778.
- [16] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2019-06-01]. <https://arxiv.org/pdf/1409.1556.pdf>.
- [17] WOO S, PARK J, LEE J Y, *et al.* CBAM: convolutional block attention module [C]//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: Springer, 2018: 3-19.
- [18] REN S Q, HE K M, GIRSHICK R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [19] BODLA N, SINGH B, CHELLAPPA R, *et al.* Soft-NMS—improving object detection with one line of code [C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017.
- [20] EVERINGHAM M, ESLAMI S M A, VAN GOOL L, *et al.* The PASCAL visual object classes challenge: a retrospective [J]. *International Journal of Computer Vision*, 2015, 111(1): 98-136.
- [21] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [DB/OL]. (2015-05-02)[2019-12-01]. <https://arxiv.org/abs/1502.03167>.

作者简介:



兰旭婷(1995—),女,宁夏银川人,硕士研究生,2019年于西北民族大学获得学士学位,主要从事深度学习与目标检测方面的研究。E-mail:1742192370@qq.com



郭中华(1973—),男,宁夏银川人,博士,教授,2010年于西北工业大学获得博士学位,主要从事图像处理、机器视觉方面的研究。E-mail:guozhh@nxu.edu.cn