

# 基于目标图像块激活的航空图像目标检测技术研究



张佳,冯婕,张骏鹏,朱潇雨

西安电子科技大学, 陕西 西安 710126

**摘要:**针对航空图像目标检测中普遍存在的图像尺寸过大、目标密集分布等关键难题,本文开创性地引入了目标图像块激活策略,不仅解决了传统方法在处理航空图像时效率低下的问题,而且显著提升了复杂场景下的精度表现,对推动航空图像目标检测技术的发展具有重要意义。现有的航空图像目标检测技术通过简单裁切图像块的方式进行处理,不仅效率低下,还因大量无效图像块的检测导致资源浪费与误检率上升。因此,本文提出了一种目标图像块激活模块(TIBAM),该模块通过引入卷积注意力机制,使检测器能够专注于包含目标物的有效图像块,从而减少在无效图像块上的资源浪费,实现了对潜在有效图像块的智能识别与处理。TIBAM的引入,为检测器带来了超过17%的推理速度提升,还实现了检测精度的稳定提升,为高效、精准的航空图像目标检测提供了创新性的解决方案。本文通过在主流的一阶段检测器Retinanet和两阶段检测器Faster-RCNN中集成TIBAM模块,以及在Visdrone航空图像数据集上进行验证,充分展示了该模块的实际应用价值与广泛适用性。

**关键词:**高效目标检测;航空图像;注意力机制;卷积神经网络

中图分类号:V219

文献标识码:A

DOI: 10.19452/j.issn1007-5453.2025.03.014

随着成像技术的飞跃,无人机与卫星拍摄的航空图像尺寸显著增长,提供了前所未有的观测精度与广度。然而,这也给航空图像目标检测带来了图像尺寸过大、小目标表观信息缺乏、目标分布密集<sup>[1]</sup>等挑战。本文针对这些难题,研发了一种既具备高精度又具备高鲁棒性的目标检测算法,该算法能够满足实时处理需求,确保在复杂多变的环境中也能快速且准确地识别目标。通过该算法提升了航空图像数据处理的能力,为城市人文环境监测、城市规划等应用提供了高效、准确的参考依据,并推动相关领域的发展。

针对航空图像尺寸过大的问题,受GPU内存容量的限制,直接将整幅航空图像输入目标检测系统中并不切实际。为解决这一问题,目前主流的检测方法采用将原始的大尺寸航空图像裁切成均匀的小尺寸图像块的做法(本文将航空图像进行分块裁切处理后得到的小尺寸图像称为图像块),然后对每个图像块逐一实施目标检测。除图像分块处理外,针对高精度航空图像,QueryDet<sup>[2]</sup>在特征图上采用创

新的级联稀疏查询(CSQ)机制,选择性地查询生成稀疏的高分辨率特征图,再输入检测头进行高效准确的目标检测,提升推理速度及检测精度。

针对航空图像小目标表观信息缺乏的问题,众多小目标检测算法的研究已转向对图像块的精细化设计与优化。董凤禹等<sup>[3]</sup>在YOLOv7基准模型的基础上进行了优化,通过在模型中间层引入广义特征金字塔网络结构,以及采用Wasserstein散度作为损失函数替代传统的交并比损失,有效地提升了该模型在空中小目标检测任务中的性能。为了提升航空图像小目标检测精度并降低定位难度,董旭彬等<sup>[4]</sup>引入SD-RPN,在不同深度的卷积层上配置了适当尺寸的滑动窗口,以有效检测各种尺度的目标,从而使生成的建议区域更加精确。为进一步解决航空图像的旋转框小目标检测,牛为华等<sup>[5]</sup>提出了基于改进YOLOv8的船舶遥感图像旋转目标检测算法,该算法通过对卷积模块进行改进,缓解了由于跨步卷积导致的图像细粒度信息丢失的问题。尽

收稿日期: 2024-07-09; 退修日期: 2024-10-17; 录用日期: 2024-12-26

基金项目: 航空科学基金(2019ZC081002)

引用格式: Zhang Jia, Feng Jie, Zhang Junpeng, et al. Research on aerial image object detection based on object image block activation[J]. Aeronautical Science & Technology, 2025, 36(03): 111-118. 张佳, 冯婕, 张骏鹏, 等. 基于目标图像块激活的航空图像目标检测技术研究[J]. 航空科学技术, 2025, 36(03): 111-118.

管这些方法都取得了不错的检测效果提升,但复杂的网络结构导致训练时间和内存开销过大。

针对航空图像目标分布密集的问题,聚类策略的引入在处理航空图像中密集且细微目标的检测任务时,展现出了卓越的效能与重要的应用价值。杨帆等<sup>[6]</sup>用迭代聚类合并候选框的思想,从骨干网络提取的特征图出发,提取聚类候选框,调整候选框和全局图像的尺寸,再进行航空图像的精细检测。为进一步解决难例挖掘问题,王毅等<sup>[7]</sup>采用均值漂移(Mean Shift)聚类算法,计算每个聚类区域的难度值,挖掘困难区域,剔除简单聚类区域,用高斯放缩函数对困难聚类放缩来减少目标尺度间的差别,提高了大尺寸航空图像的检测速度。

在对大尺寸航空图像进行深入探究的过程中,观察到目标物体往往聚集于特定区域,如行人常集中于人行道区域。因此,将图像分割成多个块后,许多图像块并不包含任何目标物体。在本文中,称包含目标物体的图像块为有效图像块,而不含的则称为无效图像块。

对于目前主流的航空图像目标检测技术来说,所有无效图像块都会被输入检测器中进行处理,这无疑会消耗大量的计算资源,并且大量无效图像块的存在也会增加误检的风险,从而降低检测的准确性。因此,一种直观的策略是引导检测器将注意力集中于有效图像块,以此实现更高效且精确的目标检测过程。

本文设计灵感源自人类视觉的注意机制,模拟人眼快速锁定图像关键区域的能力,以开发高精度、高效率的航空图像目标检测算法。近年来,众多研究致力于在卷积神经网络中嵌入注意力机制模块,如李捷<sup>[8]</sup>和赵琰<sup>[9]</sup>等在其研究中融合注意力模块都取得了很好的效果。此外,在王思宇等<sup>[10]</sup>的研究中,巧妙地融合了Swin Transformer网络架构与协调注意力机制,搭建时空交叉网络(STCNet)骨干特征提取网络,用于精准预测无人机目标的位置信息。故本文在目标图像块激活模块中引入卷积注意力机制模块<sup>[11]</sup>来更好地定位有目标物的有效图像块。这一融合不仅提升了检测器对有效图像块的定位精度,还增强了其抗干扰能力,确保在航空图像的复杂背景中也能准确捕捉目标。

本文提出的目标图像块激活模块,为航空图像目标检测领域带来了简洁而高效的解决方案。通过在骨干网络后集成此模块,实现了对图像块有效性的精准判断,有效地避免了计算资源在非目标区域的浪费,从而优化了检测效率。此外,本文所设计的模块具有良好的通用性,可集成至现有的各类检测器架构中,无论是高效的一阶段检测器还是精

确的两阶段检测器,均能显著提升其性能。通过在Visdrone航空图像数据集上的试验验证,实现了超过17%的推理速度提升,并提高了检测精度,展现了该模块在航空应用中的巨大潜力。

## 1 现有的航空图像目标检测相关技术

### 1.1 传统的航空图像目标检测数据处理架构

受限于GPU内存容量,大尺寸航空图像无法直接输入常规的目标检测器中,而将大尺寸图像降采样作为输入会丢失小物体的细节信息。为确保小目标细节信息的完整性,大尺寸的航空图像目标检测常用的方法是在数据预处理阶段将图像进行分块裁切处理<sup>[12]</sup>,将大尺寸航空图像裁切后的图像块逐一送入目标检测神经网络进行检测和识别,随后在后续处理过程中将所有检测结果进行合并和调整,以完成全图的检测<sup>[13]</sup>,其结构如图1所示。

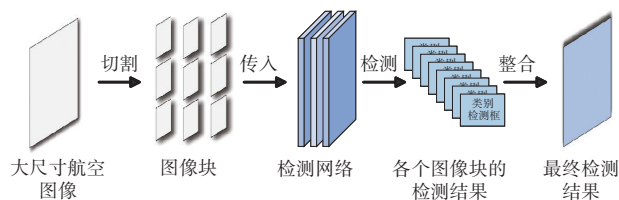


图1 传统的航空图像目标检测算法

Fig.1 Traditional aerial image object detection algorithm

虽然将图像裁切分块很大程度上缓解了目标检测过程中所带来的计算量过大,以及内存压力等问题,但该方法在裁切图像的过程中会将一些待检测目标从中截断,使得待检测目标的完整性被破坏,影响了目标的检测效果。为此,王辰等<sup>[14]</sup>使用一定的重叠区域对原始的航空图像进行切割,该方法缓解了将目标截断的概率,提升了目标的检测精度,但也导致切割图像数量上升,使得检测时间大幅度增加。

传统的航空图像目标检测方法的一个很大缺陷在于它会在所有图像块上进行穷举检测,而不判断该图像块中是否包含目标物。这不仅浪费了大量计算资源,而且会增加误检风险,降低检测准确性。

### 1.2 Focal loss 损失函数

在深度学习中,损失函数是一种用于计算模型预测结果与真实值之间差异的函数,其值是非负实数,通常表示为 $L(Y, f(x))$ ,差距程度越小,损失也就越小,从而表明该学习模型越好。

损失函数主要用于模型的训练阶段,在每个批次的训

训练数据送入模型后,通过前向传播输出预测值,接着损失函数会计算出预测值与真实值之间的差异,即损失值。获得损失值后,模型会通过反向传播算法来更新其各个参数,以最小化真实值与预测值之间的差异,从而使模型生成的预测结果更加接近真实值,最终达到学习的目的。常见的损失函数包括L1损失、L2损失和交叉熵损失<sup>[15]</sup>等。在目标图像块激活模块中,采用了焦点损失(Focal Loss)函数<sup>[16]</sup>。

Focal Loss的引入主要是为了解决目标检测中正负样本数量极不平衡的问题。Focal Loss是基于二分类交叉熵的动态缩放损失,通过一个动态缩放因子,动态地降低训练过程中易于区分样本的权重,从而将更多的重心聚焦在那些难以区分的样本上,其定义为

$$FL(p_i) = -\alpha(1-p_i)^\gamma \log(p_i) \quad (1)$$

式中, $\alpha \in [0, 1]$ 帮助处理了类别的不均衡, $(1-p_i)^\gamma$ 是调节因子, $\gamma \geq 0$ 是可调节的聚焦参数, $\gamma$ 值越大,易分类样本的损失就越小,就可以把模型的注意力投向那些难分类的样本。同时,当 $\gamma=0$ 时,这个表达式就退化成了交叉熵损失函数。

在目标图像块激活模块中引入Focal Loss,可以很好地解决正负样本数量极不平衡的问题,使网络更多地关注难区分样本,从而达到更好的检测效果。

### 1.3 卷积注意力机制模块

卷积注意力机制模块(CBAM)<sup>[11]</sup>如图2所示。巧妙地融合了空间和通道两种注意力机制,形成了两个独立的子模块:通道注意力模块(CAM)<sup>[17]</sup>和空间注意力模块(SAM)。这两种注意力机制分别关注通道和空间上的特征,使模型能够更好地理解和利用输入数据中的关键信息。

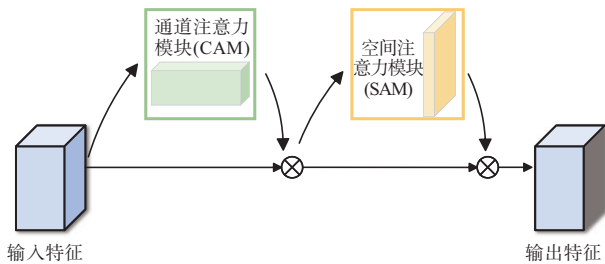


图2 卷积注意力机制结构

Fig.2 Structure of convolutional block attention module

卷积注意力机制模块实现了从通道到空间的顺序注意力结构,使模型能够更好地关注输入数据中的重要特征和信息,适应不同的网络结构和输入尺寸,从而在目标检测、图像分类、人脸识别等任务中取得出色的效果。

CAM主要关注通道维度上的特征信息。该模块利用

全局平均池化和最大池化来聚合空间信息,并将这些信息合并成一个通道注意力图。该过程有助于模型更好地理解输入数据中不同通道的重要性,从而更有效地利用特征信息。

SAM则关注空间维度上的特征信息,通过使用逐元素的加权和操作来聚合空间特征。该操作使模型在空间上关注到输入数据中的重要位置和区域,更好地捕捉和理解图像中的局部和全局特征。

总体来说,CBAM是一种强大而通用的注意力机制模块,通过灵活适应不同的网络结构和输入尺寸,以及关注通道和空间上的特征信息,为深度学习模型提供了更精确、更泛化的能力,使其在不同的任务和场景中都能表现出卓越的性能。

在目标图像块激活模块中引入CBAM,使模型能够更好地理解和利用输入数据中的关键信息,更好地定位出含有目标物的图像块,过滤掉无效图像块,从而实现更精确、更有效的航空图像目标检测。

## 2 目标图像块激活的航空图像目标检测网络

### 2.1 目标图像块激活的航空图像目标检测网络架构

本文提出了一种创新性的目标图像块激活模块,为航空图像目标检测提供了一种简单高效的解决方案,通过在骨干网络后引入目标图像块激活模块,预测每个图像块中是否包含目标物,让检测器专注于包含目标物的有效图像块,从而实现更快速的推理,获得更准确的结果。本文所设计的目标块激活的航空图像目标检测网络如图3所示,激活后的图像块中变暗的部分即为不含目标物的无效图像块,这些无效图像块不会传入后续的检测网络。

本文在两种先进的目标检测网络上构建了目标块激活的航空图像目标检测网络。图4展示了目标图像块激活模块的内部结构。具体来说,给定一张航空图像,将其划分为 $256 \times 256$ 大小的图像块。选择从骨干网络ResNet50<sup>[18]</sup>最后阶段提取的特征图(记作R50)作为目标图像块激活模块的输入。在R50上,首先应用一个包含64个滤波器的 $3 \times 3$ 步长为2卷积层,得到一个尺寸为 $16 \times 16 \times 64$ 的低维特征图。然后,将其送入一个带有128个滤波器的 $1 \times 1$ 卷积层,再送入另一个带有一个滤波器的 $1 \times 1$ 卷积层,生成大小为 $16 \times 16 \times 1$ 对象性激活图 $M$ , $M$ 在每个位置的值表示其对应网格的对象得分。

式(2)设计了一个阈值判定机制以判别图像块中是否含有目标物体。具体而言,对于给定的图像块,若其内任一



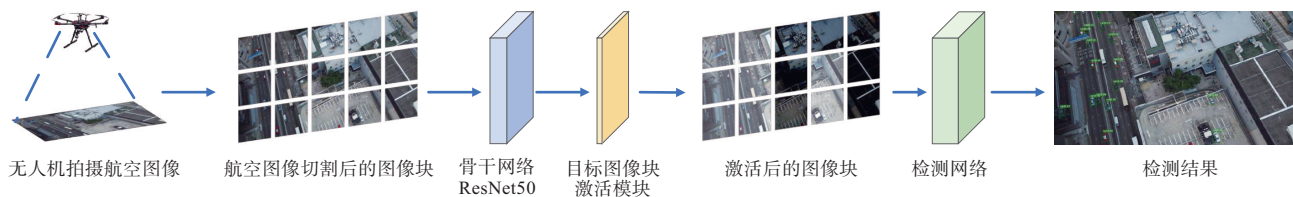


图3 目标图像块激活的航空图像目标检测网络

fig.3 aerial image object detection network activated by target image block activation module

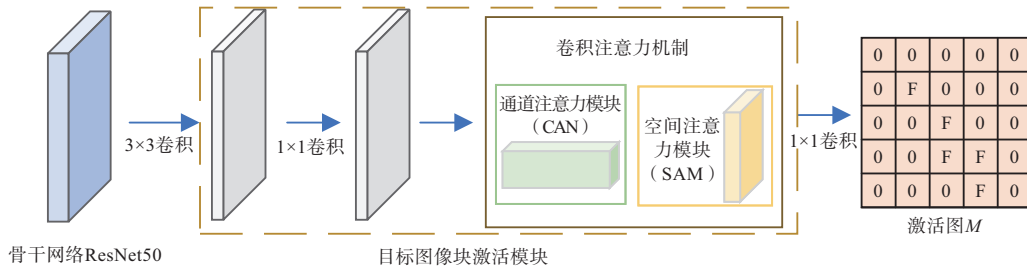


图4 目标图像块激活模块

Fig.4 Target image block activation module

网络的得分 $M$ 超越了预设的阈值 $t$ ,则该图像块被标记为有效图像块,并继续参与后续的检测流程。反之,若图像块内所有网格的得分均未达到阈值 $t$ ,则该图像块被视为无效,将其从后续检测网络的候选输入列表中删除,以此优化计算资源与提高处理效率

$$T(F, M) = \begin{cases} F, \max(M) > t \\ 0, \max(M) \leq t \end{cases} \quad (2)$$

## 2.2 损失函数

目标块激活的航空图像目标检测网络的总损失主要可以分为三部分,即目标图像块激活模块的损失、分类损失以及边界框回归损失,若为两阶段的检测器,还应加入提议生成损失。因此,一阶段检测器和两阶段检测器的总损失定义如式(3)和式(4)所示

$$L_{\text{one\_stage}} = L_{\text{class}} + L_{\text{box}} + L_{\text{TIBAM}} \quad (3)$$

$$L_{\text{two\_stage}} = L_{\text{proposal}} + L_{\text{class}} + L_{\text{box}} + L_{\text{TIBAM}} \quad (4)$$

目标图像块激活模块的损失设计如下:记载切后的图像块为 $P(i, j) \in (0, N-1)$ ,该图像块的得分为 $S(i, j)$ , $I(i, j)$ 表示真实的位置中心是否落在该 $P(i, j)$ 里,若是,则 $I(i, j) = 1$ ,否则 $I(i, j) = 0$ 。则目标图像块激活模块的损失定义如式(5)所示

$$L_{\text{TIBAM}} = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \text{FL}(I(i, j), S(i, j)) \quad (5)$$

式中,FL(\*)即在1.2损失函数小节中的Focal Loss损失函数,它可以解决检测中正负样本数量极不平衡的问题。目标图像块激活模块的损失设计较为简单,有利于快速训练。

## 3 试验与结果分析

### 3.1 数据集和试验参数

VisDrone数据集<sup>[19]</sup>是由中国天津大学机器学习与数据挖掘实验室精心策划与收集的航空视觉数据集,以其独特性和全面性在无人机视觉领域占据重要地位。该数据集的所有内容均源自无人机高空拍摄,不仅覆盖了多个城市、多样环境及广泛物体,还通过细致的标注与处理,确保了数据的高质量与可用性。VisDrone数据集共包含6471张训练集图像、548张验证集图像以及1610张测试集图像,预设了包括行人、车辆、摩托车、面包车、巴士等在内的10个类别,旨在满足目标检测、跟踪、分类及属性预测等多种视觉分析任务的需求。

本文的目标块激活的航空图像目标检测网络主要优势在于检测密集小目标,因此对原有数据集进行了筛选处理。通过将行人与车合并归类,并精选自行车、汽车及三轮车共4类小目标作为检测的重点,使数据集更加聚焦于航空图像中小目标的检测难题,旨在以更高效、更准确的方式开展航空图像小目标的相关研究。

在试验阶段,随机抽取了原VisDrone数据集的20%来进行训练和推理,其中包含1294张训练集、109张验证集和322张测试集。将原始数据剪裁为256×256大小的图像块,相邻的两个图像块之间有10个重叠像素,传入骨干网络,并在其后接入目标图像块激活模块,再将过滤后得到的有效图像块传入检测网络,得到最终检测结果。

本文的试验是基于MMDetection<sup>[20]</sup>实现的,试验部署

在单个 NVIDIA TITAN RTX 上,采用随机梯度下降策略,初始学习率设为 0.0025,在第 8 个迭代轮次和第 11 个迭代轮次之后将学习率减少 10%。使用训练集和验证集进行训练,采用平均精度(mAP)来衡量检测器的准确性,采用每秒传输帧数(FPS)来衡量检测器的推理速度,并采用召回率(Recall)来衡量被预测正确的正例占总实际正例样本的比例,以反映检测器的漏检情况。同时,采用 F1 来全面地评估模型的性能,其计算公式如下

$$F1 = \frac{2 \times (mAP \times Recall)}{mAP + Recall} \tag{6}$$

3.2 试验结果

在试验阶段,本文选择了一阶段检测网络 Retinanet<sup>[16]</sup>和两阶段检测网络 Faster-RCNN<sup>[21]</sup>,并将目标图像块激活模块分别集成到这两种先进的检测网络中,形成新的检测模型 Retinanet+TIBAM 和 Faster-RCNN+TIBAM。为了客观地评估加入目标图像块激活模块对检测效果的影响,本文对这些模型进行了详细的对比试验,并将结果整理成表 1。表 1 展示了 Retinanet 和 Faster-RCNN 在加入 TIBAM 前后的检测效果对比,mAP 值越高,表明模型的检测准确性越高、模型效果越好。通过对比试验结果,可以清晰地看到,Retinanet+TIBAM 和 Faster-RCNN+TIBAM 在 mAP 值上均有了显著的提升。这表明本文所设计的目标图像块激活模块能够充分过滤掉无效图像块,减少误检情况,有效地提高目标检测网络的准确率。

表1 传统网络与加入目标图像块激活模块后的网络精度对比  
(单位:%)

Table 1 Comparison of mAP between traditional network and network with target image block activation module (unit:%)

| 方法                | 人     | 自行车   | 汽车    | 三轮车   | mAP   |
|-------------------|-------|-------|-------|-------|-------|
| Retinanet         | 47.37 | 18.90 | 67.17 | 27.52 | 40.24 |
| Retinanet+TIBAM   | 48.36 | 20.26 | 67.58 | 25.78 | 40.50 |
| Faster-RCNN       | 54.79 | 40.97 | 77.10 | 46.81 | 54.92 |
| Faster-RCNN+TIBAM | 55.07 | 41.35 | 77.27 | 48.08 | 55.44 |
| QueryDet          | 47.63 | 19.05 | 68.42 | 26.72 | 40.46 |

在测试阶段,从裁切后的测试集中选择了 2000 张图像,以评估目标图像块激活模块在提升检测速度方面的表现。结果见表 2,与传统的检测网络相比,加入目标图像块激活模块的检测网络在速度上有了显著的提升,其中 Faster-RCNN 的检测速度提升了 17.78%,Retinanet 的检测速度提升了 30.25%。

表2 检测器的推理速度对比

Table 2 Comparison between detectors' FPS

| 方法                | FPS/(图片/s) |
|-------------------|------------|
| Retinanet         | 11.3       |
| Retinanet+TIBAM   | 16.2       |
| Faster-RCNN       | 11.1       |
| Faster-RCNN+TIBAM | 13.5       |

3.3 讨论与分析

上述试验结果与本文的设计理念完全相符,目标图像块激活模块能够有效地过滤掉那些无效图像块,从而显著减少输入检测网络的图像块数量。这使得检测网络只需要处理那些可能包含目标物的有效图像块,从而极大地提高了检测速度,同时减少了不必要的计算资源浪费。

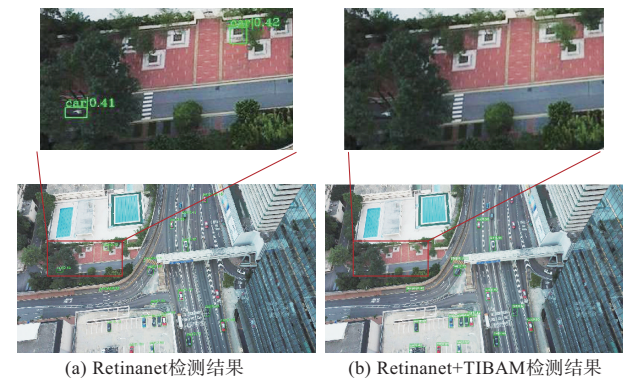


图5 Retinanet与Retinanet+TIBAM检测结果对比

Fig.5 Comparison of test results between Retinanet and Retinanet+TIBAM

与此同时,目标图像块激活模块过滤掉了大量无效图像块,减少了其后检测网络的误检情况,从而在一定程度上提高了检测精度。如图 5 中 Retinanet 的检测结果中,红框标出来的区域并不包含目标物,但存在误检情况;而在 Retinanet+TIBAM 的检测结果中,目标图像块激活模块将这一区域的无效图像块过滤掉,故有效地避免了误检情况。

图 6 详尽地展示了在引入目标图像块激活模块前后网络召回率的对比情况。具体看来,在 Retinanet+TIBAM 框架下,人的召回率显著提升,这在一定程度上实证了目标检测激活模块中嵌入的注意力机制的有效性。从全局视角看来,尽管目标图像块激活模块的融入不可避免地筛选了少数包含目标物体的图像块,从而导致了如三轮车等特定类别目标的召回率出现轻微下滑,但整体召回率指标并未因此受损。进一步细致分析发现,针对人、自行车及三轮车等小型目标,其召回率相较于汽车等大型目标展现出显著差

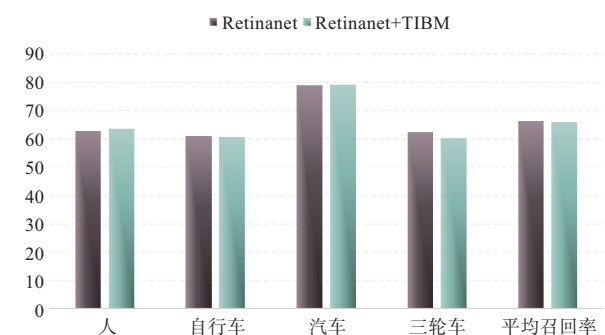


图6 传统网络与目标图像块激活网络召回率对比(单位:%)

Fig.6 Comparison of recall between traditional network and network with target image block activation module (Unit:%)

异性,此现象深刻反映了小目标检测任务所固有的严峻挑战及其技术实现的复杂性。

图7对比了集成目标图像块激活模块前后,模型在F1分数上的变化情况。试验结果显示,在Retinanet框架中引入目标图像块激活模块后,F1分数相较于原始的Retinanet模型实现了显著提升。这展现出目标图像块激活模块对于提升模型整体性能的重要作用和优势。

现有的小目标检测算法,针对小目标优化采用空洞卷积、超分辨率重建等,以提高对小目标的检测精度和效率。其中,QueryDet算法通过Retinanet框架结合CSQ机制,实现了稀疏高分辨率特征图的生成,并在检测头进行稀疏查询来提高检测整体效率。表1展示了QueryDet算法在本文处理后的VisDrone数据集上的试验结果,与QueryDet算法相比,本文提出的目标图像块激活算法在检测性能上展现出了优势。具体分析发现,QueryDet算法生成的稀疏高分辨率特征图质量尚存提升空间,而本文提出的算法在图像进入检测网前就进行了有效的图像激活,确保了输入

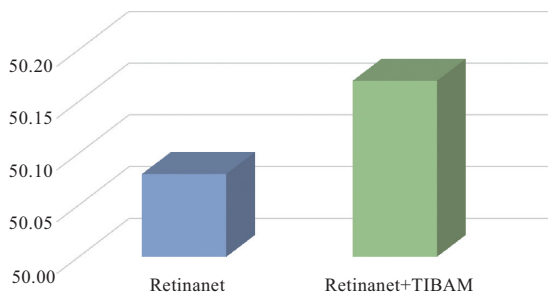


图7 传统网络与目标图像块激活网络F1对比(单位:%)

Fig.7 Comparison of F1 between traditional network and network with target image block activation module (Unit:%)

检测头的特征图质量,从而直接提升了目标检测的精度和效果。

总体而言,本目标图像块激活的航空图像目标检测网络结构简单,精度和速度提升明显。无论对于一阶段还是两阶段的检测模型,均能实现推理速度的大幅提升和检测精度的提高。特别是对于两阶段模型Faster-RCNN,该模块的引入使得推理速度提高了17.78%,同时检测精度也提升了0.52%,实现了性能的全面优化。此外,本文引入的目标图像块激活模块具有高度的兼容性,能够无缝集成至当前主流的目标检测算法体系之中,进而有效提升此类复杂算法在训练及推理阶段的执行效率。

## 4 结束语

本文在深入研究航空图像密集小目标检测面临的挑战后,提出了一种目标图像块激活模块。该模块通过精确识别并过滤不包含目标物体的无效图像块,有效缩减了输入检测网络的数据量,从而在显著提升检测效率的同时,降低了误检率并增强了检测精度。作为一个轻量级且易于集成的卷积网络设计,该模块不仅优化了大尺寸航空图像中密集小目标的检测性能,更在实际应用中凸显了其在提升处理速度与资源利用率上的重要性。本文研究的结果不仅为航空视觉领域内的密集小目标检测任务提供了高效解决方案,也为其他相似复杂场景下的目标检测任务开辟了新的研究思路与实践路径。

AST

## 参考文献

- [1] 石争浩, 仵晨伟, 李成建, 等. 航空遥感图像深度学习目标检测技术研究进展[J]. 中国图象图形学报, 2023, 28(9): 2616-2643.  
Shi Zhenghao, Wu Chenwei, Li Chengjian, et al. Research progress of deep learning object detection in aerial remote sensing images[J]. Journal of Image and Graphics, 2023, 28(9): 2616-2643. (in Chinese)
- [2] Yang Chenhongyi, Huang Zehao, Wang Naiyan. QueryDet: cascaded sparse query for accelerating high-resolution small object detection[C]. Computer Vision and Pattern Recognition, 2022: 13668-13677.
- [3] 董凤禹, 魏振忠. 面向空中小目标检测任务的YOLOv7改进模型[J]. 航空科学技术, 2023, 34(12): 111-117.  
Dong Fengyu, Wei Zhenzhong. An improved YOLOv7 model for small aerial object detection[J]. Aeronautical Science &



- Technology, 2023, 34(12): 111-117. (in Chinese)
- [4] 董旭彬, 赵清华. 改进 Mask R-CNN 在航空影像目标检测的研究应用[J]. 计算机工程与应用, 2021, 57(8): 133-144.
- Dong Xubin, Zhao Qinghua. Research and application of improved Mask R-CNN in aerial image target detection [J]. Computer Engineering and Applications, 2021, 57(8): 133-144. (in Chinese)
- [5] 牛为华, 郭迅. 基于改进 YOLOv8 的船舰遥感图像旋转目标检测算法[J]. 图学学报, 2024, 45(4): 726-735.
- Niu Weihua, Guo Xun. Rotating target detection algorithm for ship remote sensing images based on improved YOLOv8[J]. Journal of Graphics, 2024, 45(4): 726-735. (in Chinese)
- [6] Yang Fan, Fan Heng, Chu Peng, et al. Clustered object detection in aerial images[C]. International Conference on Computer Vision, 2019: 8310-8319.
- [7] Wang Yi, Yang Youlong, Zhao Xi. Object detection using clustering algorithm adaptive searching regions in aerial images [C]. European Conference on Computer Vision, 2020: 651-664.
- [8] 李捷, 周顺, 朱鑫潮, 等. 结合多通道注意力的遥感图像飞机目标检测[J]. 计算机工程与应用, 2022, 58(1): 209-217.
- Li Jie, Zhou Shun, Zhu Xinchao, et al. Aircraft target detection with Remote sensing images combined with multi-channel attention [J]. Computer Engineering and Applications, 2022, 58(1): 209-217. (in Chinese)
- [9] 赵琰, 赵凌君, 匡纲要. 基于注意力机制特征融合网络的 SAR 图像飞机目标快速检测[J]. 电子学报, 2021, 49(9): 1665-1674.
- Zhao Yan, Zhao Lingjun, Kuang Gangyao. Rapid detection of aircraft targets in SAR images based on attention mechanism feature fusion network [J]. Acta Electronica Sinica, 2021, 49(9): 1665-1674. (in Chinese)
- [10] 王思宇, 卢瑞涛, 黄攀, 等. 基于 Swin Transformer 和注意力机制的红外无人机检测算法[J]. 航空科学技术, 2024, 35(2): 39-46.
- Wang Siyu, Lu Ruitao, Huang Pan, et al. Infrared UAV detection algorithm based on Swin Transformer and attention mechanism[J]. Aeronautical Science & Technology, 2024, 35(2): 39-46. (in Chinese)
- [11] Sanghyun W, Jongchan P, Joon-Young L, et al. Cbam: convolutional block attention module[C]. European Conference on Computer Vision, 2018: 3-19.
- [12] Xie Xingxing, Cheng Gong, Wang Jiabao, et al. Oriented R-CNN for object detection[C]. International Conference on Computer Vision, 2021: 3500-3509.
- [13] Han Jiaming, Ding Jian, Li Jie, et al. Align deep features for oriented object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-11.
- [14] Wang Chen, Bai Xiao, Wang Shuai, et al. Multiscale visual attention networks for object detection in VHR remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(2): 310-314.
- [15] Zhao Hang, Orazio G, Iuri F, et al. Loss functions for image restoration with neural networks[J]. IEEE Transactions on Computational Imaging, 2016, 3(1): 47-57.
- [16] Tsung-Yi L, Priya G, Ross G, et al. Focal loss for dense object detection[C]. International Conference on Computer Vision, 2017: 2999-3007.
- [17] Hu Jie, Shen Li, Sun Gang. Squeeze-and-excitation networks [C]. Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [18] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition[C]. Computer Vision and Pattern Recognition, 2016: 770-778.
- [19] Zhu Pengfe, Du Dawei, Wen Longyin, et al. VisDrone-2019: the vision meets drone object detection in video challenge results[C]. International Conference on Computer Vision Workshops, 2019: 227-235.
- [20] Chen Kai, Wang Jiaqi, Pang Jiangmiao, et al. MMDetection: open MMLab detection toolbox and benchmark [EB/OL]. (2019-06-17). <https://arxiv.org/abs/1906.07155>.
- [21] Ren Shaoqing, He Kaiming, Ross G, 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.

## Research on Aerial Image Object Detection Based on Object Image Block Activation

Zhang Jia, Feng Jie, Zhang Junpeng, Zhu Xiaoyu

*Xidian University, Xi'an 710126, China*

**Abstract:** Aiming at the key challenges of large image size and dense distribution of targets, which are commonly found in aerial image target detection, this paper pioneers the target image block activation strategy, which not only solves the inefficiency problem of traditional methods in processing aerial images, but also significantly improves the accuracy performance in complex scenes, which is of great significance in promoting the development of aerial image target detection technology. Existing target detection techniques for aerial images are processed by simply cropping image blocks, which is not only inefficient, but also leads to waste of resources and increase of false detection rate due to the detection of a large number of invalid image blocks. Therefore, this paper proposes a target image block activation module (TIBAM), which enables the detector to focus on valid image blocks containing the target by introducing a convolutional attention mechanism, thus reducing the waste of resources on invalid image blocks and realizing the intelligent recognition and processing of potentially valid image blocks. The introduction of TIBAM brings more than 17% improvement in the inference speed for the detector, and also achieves a stable improvement in detection accuracy, providing an innovative solution for efficient and accurate aerial image target detection. This paper fully demonstrate the practical application value and wide applicability of the TIBAM module by integrating it in the mainstream one-stage detector Retinanet and two-stage detector Faster-RCNN, as well as validating it on the Visdrone aerial image dataset.

**Key Words:** efficient target detection; aerial images; attention mechanism; convolutional neural network

---

**Received:** 2024-07-09; **Revised:** 2024-10-17; **Accepted:** 2024-12-26

**Foundation item:** Aeronautical Science Foundation of China(2019ZC081002)