

# 控制与决策

Control and Decision

AF-RetinaNet:一种基于自适应融合与特征细化的微小行人检测算法

邵香迎, 郭颖, 王友伟

引用本文:

邵香迎, 郭颖, 王友伟. AF-RetinaNet:一种基于自适应融合与特征细化的微小行人检测算法[J]. 控制与决策, 2024, 39(3): 939–946.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2022.0933>

---

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### [Anchor-free的尺度自适应行人检测算法](#)

Anchor-free scale adaptive pedestrian detection algorithm

控制与决策. 2021, 36(2): 295–302 <https://doi.org/10.13195/j.kzyjc.2020.0124>

#### [多目标小尺度车辆目标检测方法](#)

Multi-target and small-scale vehicle target detection method

控制与决策. 2021, 36(11): 2707–2712 <https://doi.org/10.13195/j.kzyjc.2020.0635>

#### [复杂背景下全景视频运动小目标检测算法](#)

Panoramic video motion small target detection algorithm in complex background

控制与决策. 2021, 36(1): 249–256 <https://doi.org/10.13195/j.kzyjc.2019.0686>

#### [基于多尺度特征表示的行人再识别](#)

Multi-scale feature representation for person re-identification

控制与决策. 2021, 36(12): 3015–3022 <https://doi.org/10.13195/j.kzyjc.2020.0952>

#### [基于多层次特征的机械臂单阶段抓取位姿检测](#)

Single-stage grasp pose detection of manipulator based on multi-level features

控制与决策. 2021, 36(8): 1815–1824 <https://doi.org/10.13195/j.kzyjc.2019.1840>

# AF-RetinaNet: 一种基于自适应融合与特征细化的微小行人检测算法

邵香迎<sup>1,2</sup>, 郭颖<sup>1,2†</sup>, 王友伟<sup>1,2</sup>

(1. 南京信息工程大学江苏省大气环境与装备技术协同创新中心, 南京 210044;  
2. 南京信息工程大学自动化学院, 南京 210044)

**摘要:** 为解决目前目标检测算法在微小行人的识别与定位过程中准确率较低的问题, 提高微小行人检测能力, 提出一种基于自适应融合与特征细化的微小行人检测算法 AF-RetinaNet. 首先, 将特征增强模块与 ResNet 相结合构建特征提取网络, 采用并行结构获得增强特征; 其次, 使用上下文自适应学习模块, 通过获得目标上下文的特征信息, 从而关注相似特征的差异性, 缓解误检问题; 最后, 构造具有图像超分思想的特征细化模块, 对目标特征信息进行放大重构, 优化小目标的特征表达能力, 缓解漏检问题. 在 TinyPerson 数据集上, AF-RetinaNet 算法的检测精度达到 56.78%, 漏检率达到 85.38%. 与基于 RetinaNet 算法的研究基准相比, 检测精度提高 5.57%, 漏检率降低 3.67%. 实验结果表明, 该模型能有效提高对微小行人的检测和识别精度.

**关键词:** 机器视觉; 微小行人; 小目标检测; 特征增强; 上下文信息; 特征细化

中图分类号: TP391.4 文献标志码: A

DOI: 10.13195/j.kzyjc.2022.0933

引用格式: 邵香迎, 郭颖, 王友伟. AF-RetinaNet: 一种基于自适应融合与特征细化的微小行人检测算法[J]. 控制与决策, 2024, 39(3): 939-946.

## AF-RetinaNet: A tiny person detection algorithm based on adaptive fusion and feature refinement

SHAO Xiang-ying<sup>1,2</sup>, GUO Ying<sup>1,2†</sup>, WANG You-wei<sup>1,2</sup>

(1. Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET), Nanjing University of Information Science & Technology, Nanjing 210044, China; 2. School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044, China)

**Abstract:** In order to solve the problem of low accuracy of current target detection algorithms in the process of tiny person recognition and location, and improve the detection ability of tiny person, this paper proposes a tiny person detection algorithm AF-RetinaNet based on adaptive fusion and feature refinement. Firstly, the algorithm combines the feature enhancement module with ResNet to build a feature extraction network and uses a parallel structure to obtain enhanced features. Secondly, the context adaptive learning module is used to obtain the feature information of the target context, so as to pay attention to the differences of similar features and alleviate the problem of false detection. Finally, the feature refinement module with the idea of image super-resolution is constructed to enlarge and reconstruct the target feature information, optimize the feature expression ability of small targets and alleviate the problem of missed detection. On the TinyPerson dataset, the average precision of the AF-RetinaNet algorithm reaches 56.78%, and the missed rate reaches 85.38%. Compared with the research benchmark based on the RetinaNet algorithm, the average precision is improved by 5.57%, and the missed rate is reduced by 3.67%. The experimental results show that the model can effectively improve the accuracy of tiny person detection and recognition.

**Keywords:** machine vision; tiny person; small target detection; feature enhancement; context information; feature refinement

## 0 引言

目标检测是计算机视觉领域重要的研究方向之一<sup>[1]</sup>. 早期目标检测主要是通过传统方法实现, 利

用 HOG<sup>[2]</sup>、Haar<sup>[3]</sup> 等方法进行手工设计特征, 结合 SVM 等分类器进行目标检测. 但是传统算法存在提取目标特征准确性、稳定性较差等弊端, 不能很好应

收稿日期: 2022-05-25; 录用日期: 2022-11-10.

基金项目: 国家自然科学基金项目(61971229).

†通讯作者. E-mail: yguo@nuist.edu.cn.

用于实际情况.近年来目标检测任务大多使用基于深度学习的方法,如RCNN<sup>[4]</sup>、Fast RCNN<sup>[5]</sup>、Faster RCNN<sup>[6]</sup>、YOLO<sup>[7]</sup>、SSD<sup>[8]</sup>、RetinaNet<sup>[9]</sup>等检测算法.

虽然目标检测技术已经被广泛应用,但是在小目标上(通常定义为 $32 \times 32$ 像素以下的目标)的检测仍然面临许多挑战.小目标物体在图像中覆盖的区域较小,像素值通常在几像素到几十像素之间,由于其分辨率低,涵盖特征信息少,缺乏特征表示能力,导致小目标存在难以检测的问题.此外,真实场景是错综复杂的,通常会存在光照变化、目标密集、遮挡、尺度不一等问题,使得检测变得更加困难.

针对小目标检测问题,近年来,相继出现许多优化算法.如文献[10]基于运动的识别方式可以有效避免在进行行人识别的过程中,某些局部纹理变化所引发的对应用环境不适的问题.文献[11]提出微小行人目标研究基准,针对微小行人提出尺度匹配(scale match, SM)算法,为微小目标提供了一个新的研究方向.文献[12]认为特征金字塔网络(feature pyramid network, FPN)<sup>[13]</sup>相邻层之间的自上而下连接对微小物体检测存在一定的影响,进而提出融合因子来控制深层传递给浅层的信息.当使用适当的融合因子配置FPN时,网络能够在微小目标数据集上实现显著的性能增益.文献[14]提出一个用于微小行人检测的尺度选择金字塔网络,利用注意力热图、尺度增强模块、尺度选择模块来优化微小目标检测的效果.文献[15]针对现有目标检测算法对于小目标检测精度低的问题,提出一种全局与局部图像特征自适应融合的一阶段小目标检测算法.文献[16]指出车辆目标检测是智能交通系统中的重要环节,针对传统车辆目标检测方法效率低、小目标检测效果不好、漏检率高等问题,提出一种基于改进YOLOv3网络的车辆目标检测算法.文献[17]提出一种基于深度学习的弱小目标检测方法,该方法基于语义分割任务,利用全卷积递归网络学习复杂背景下弱小目标的特征,并在网络中使用了残差学习和递归操作,具有加速网络优化、模型参数少、深度递归监督和特征重用等特点.文献[18]提出一种新的注意力结构,用于降低数据集规模分布对FPN特征融合过程的影响.文献[19]提出一种基于高斯感受野的标签分配策略,通过直接测量高斯感受野与真实框的相似度进行样本分配,从而提高微小目标检测效果.

为了提高对微小行人目标的检测性能,本文提出一种新的检测算法(AF-RetinaNet),通过引入特征增强模块、上下文自适应学习模块、特征细化模块,获

得丰富语义信息及位置信息,便于更好地分类与定位.本文的工作主要体现在以下几个方面:

1) 设计特征增强模块(feature enhancement module, FEM),通过与ResNet<sup>[20]</sup>网络构成双分支,对原始信息进行并行提取,获得增强特征,在根源上对目标信息进行补充,缓解小目标信息表示微弱或丢失的问题,提高其检测效果.

2) 设计上下文自适应学习模块(context adaptive learning module, CALM),通过学习目标上下文信息,给予相似特征差异性更多关注,缓解误检问题.

3) 设计特征细化模块(feature refinement module, FRM),利用嵌入超分思想对富含小目标信息的浅层特征进行信息细化,获取区域中更多目标细节信息,缓解漏检情况.

## 1 本文方法

本文提出的AF-RetinaNet算法整体结构如图1所示.

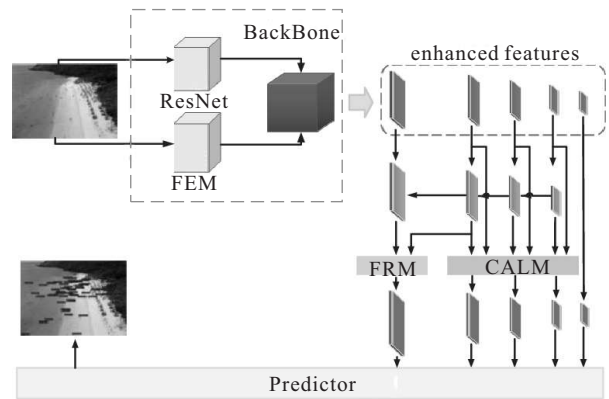


图1 网络模型整体结构

该模型通过使用双分支结构形成主干网络,其中:分支1为ResNet 50网络,分支2为特征增强模块(FEM).利用主干网络对目标数据进行提取,获得增强特征,处理微小目标信息随着网络层次加深导致表示微弱或丢失问题.然后,采用上下文自适应学习模块(CALM),利用特征上下文信息,强化模型辨别相似特征的能力,帮助模型更容易区分前景与背景问题.为进一步加强对微小行人的检测性能,优化目标漏检情况,引入特征细化模块(FRM),采用图像超分对富含小目标信息的浅层特征进行精细学习,获得高分辨率特征.将获得的特征送入检测器进行检测,有效提升模型对微小行人的检测效果.

### 1.1 特征增强模块

卷积神经网络的浅层特征包含丰富的纹理、边缘等细节空间信息,有利于对物体进行定位,但缺乏

全局语义信息,不利于物体的分类. 而深层特征具有丰富的语义信息,但丢失了许多空间细节信息,不利于小目标的检测<sup>[21]</sup>. RetinaNet检测算法采用ResNet作为其主干网络,提取目标的特征信息;然后通过FPN的自顶向下和横向连接操作融合相邻层的语义

信息,获得丰富的目标特征,但仍存在小目标信息表示微弱或者丢失的情况. 鉴于上述问题,本文提出特征增强模块. 通过采用并行分支结构,对原始图像信息进行并行提取,以此增强小目标特征信息,优化小目标检测效果. 图2为特征增强模块结构图.

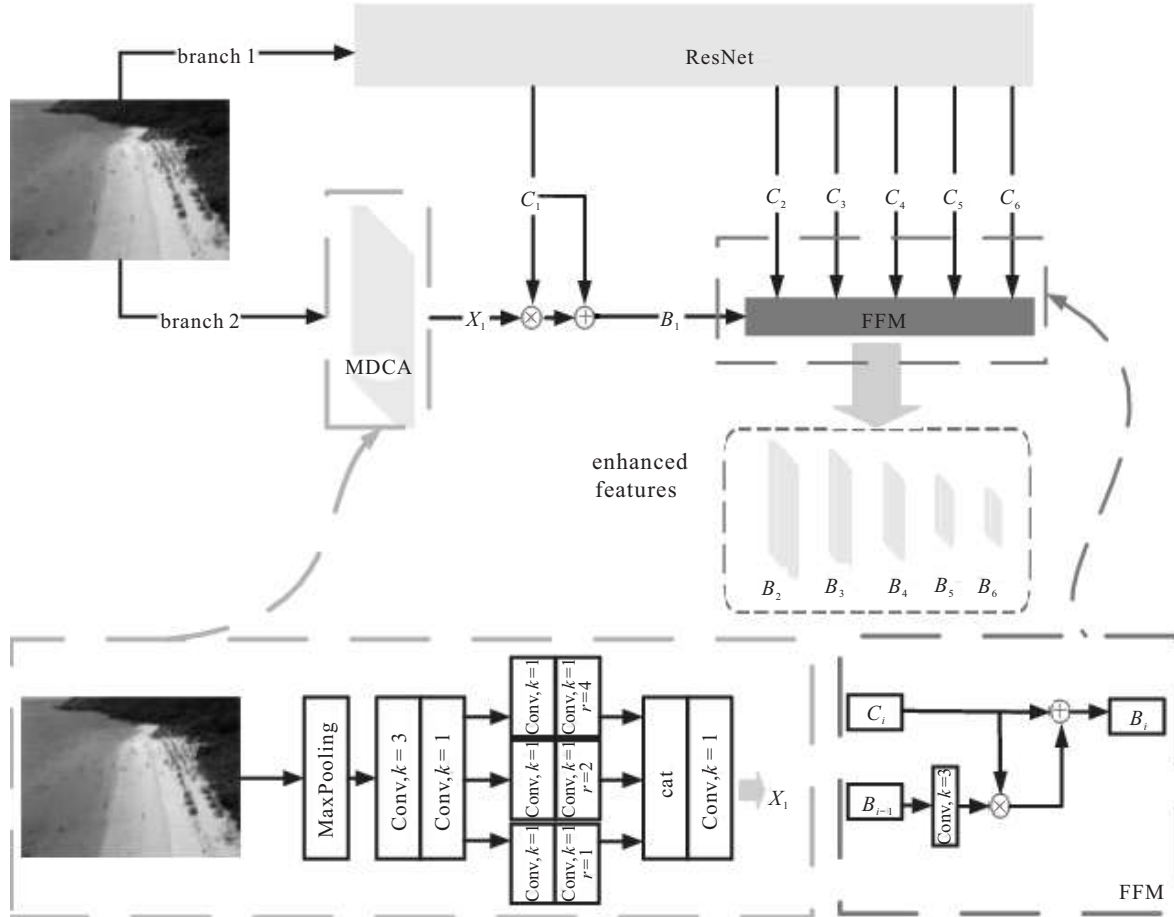


图2 特征增强模块结构

本文对输入图像采用并行分支<sup>[22]</sup>结构,分支1为ResNet网络结构,分支2为FEM网络结构. 输入图像经过图2所示的多空洞卷积分支聚合模块(multi-dilation convolutional aggregation, MDCA),采用最大池化和多空洞卷积操作,获得特征 $X_1$ . 为了获得更浅层小目标信息,本文对 $C_1$ 与 $X_1$ 进行元素点积、相加操作,得到 $B_1$ . 将 $B_{i-1}$ 与 $C_i$  ( $i = 2, 3, 4, 5$ )依次送入特征融合模块(feature fusion module, FFM)中,利用 $3 \times 3$ 卷积操作缩小特征图 $B_{i-1}$ 的尺寸,最终获得增强特征 $B_2 \sim B_5$ . 为了进一步获得丰富的深层次语义信息,对 $C_5$ 层进行池化下采样得到 $C_6$ ,将 $B_5$ 与 $C_6$ 送入FFM模块中,得到特征图 $B_6$ .

$$B_i = \begin{cases} X_1 C_i + C_i, & i = 1; \\ \text{Conv}(B_{i-1}) C_i + C_i, & i \in \{2, 6\}. \end{cases} \quad (1)$$

特征增强模块的工作原理如式(1)所示. MDCA表示多空洞卷积分支聚合模块,Conv为 $3 \times 3$ 卷积、BN、ReLU操作;  $X_1$ 为原始输入图像经过MDCA得到的新输入图像;  $C_i$ 表示经过ResNet网络提取到的特征图;  $B_i$ 表示增强特征.

### 1.2 上下文自适应学习模块

通过特征增强模块获得的增强特征,为AF-RetinaNet模型提供了更加丰富的特征信息;增强特征送入FPN后,经过自顶向下和横向连接操作,只获得两个尺寸特征图之间的联系,不同尺度间的语义信息联系过于单薄<sup>[23]</sup>. 为增强多尺度语义信息之间的联系,辨别相似特征,减少模型误检情况,本文提出上下文自适应学习模块.

首先对输入特征进行Resize操作,使其与 $D_i$ 尺寸相同. 对于上采样,采用双线性插值操作提高特征分辨率;对于下采样,采用stride为2的卷积进行实

现. 为增强多尺度语义信息之间的联系, 引入了自适应学习方法<sup>[24]</sup>, 对于不同特征图的同一位置像素信息, 采用 $1 \times 1$ 卷积、BN、LeakyReLU获得对应权重信息. 为了减少计算量的同时降低通道数, 可对其进行cat拼接、 $1 \times 1$ 卷积操作, 最终实现每个特征图的每个像素在通道上权重共享. 采用Softmax操作将权重 $W$ 的取值保持在0-1之间. 对不同特征图的每个像素信息进行加权求和, 得到特征图 $P'_i$ . 最终与 $M_i$ 进行融合, 获得边缘位置信息, 优化误检问题. 整体结构如图3所示.

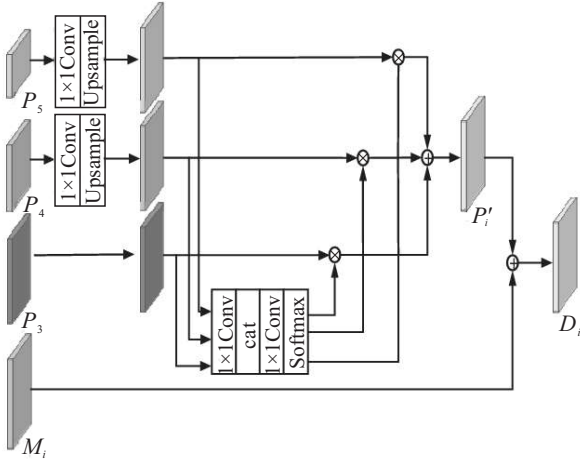


图3 上下文自适应学习模块结构

$$W_i = \frac{\exp(W(F_i))}{\sum_i \exp(W(F_i))}, \quad (2)$$

$$D_i = \sum_i W_i F_i + M_i. \quad (3)$$

其中:  $F_i$  为特征图  $P_i$  经过尺度缩放后的特征图;  $W(\cdot)$  表示卷积、concatenation 操作;  $W_i$  表示归一化的权重;  $M_i$  表示网络FPN横向连接获得的特征图;  $D_i$  表示模块的最终输出特征图;  $i = 3, 4, 5$ .

### 1.3 特征细化模块

为缓解小目标检测效果差的问题, 在目标检测中引入一些超分辨率方法. 早期的实践<sup>[25-26]</sup> 直接对输入图像进行超分辨率处理, 但后续网络中特征提取的计算成本较高. 为解决高分辨率带来的计算成本问题, 通过嵌入超分<sup>[27]</sup> 思想来提高对小目标的检测效果. 而本文为学习更多精细语义信息和可靠区域纹理, 直接对浅层的特征进行学习, 得到具有丰富区域细节的超分特征来解耦小目标检测. 同时, 为进一步减小计算成本并未对最浅层特征  $P_1$  进行输出并检测. 基于此, 设计特征细化模块, 整体结构如图4所示.

通过局部残差模块和全局残差连接获得低分辨率特征图  $P_3$  的主要语义内容. 局部残差连接模块 (Resblock 模块) 如图4右边框所示, 由2个Conv块组

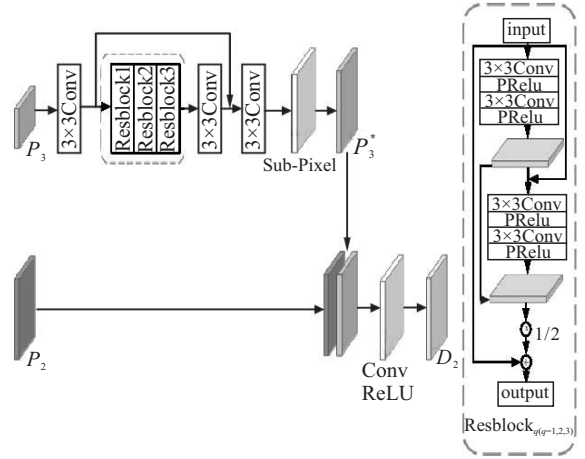


图4 特征细化模块结构

成, 每个Conv块由2个 $3 \times 3$ 卷积、PReLU激活函数构成. 全局残差连接与残差函数作用相似, 随着网络层数的加深, 可能会出现梯度消失的情况, 加入全局残差连接, 能够有效解决此问题, 避免模型陷入权值更新缓慢或更新停滞状态. 将特征图进行Sub-Pixel卷积操作生成高分辨率特征  $P_3^*$ . Sub-Pixel卷积操作使特征图的宽度和高度转变为原来的2倍, 通道数变换为原来的1/4, 即  $P_3 \in R^{H \times W \times C}$  转变为  $P_3^* \in R^{2H \times 2W \times 1/4C}$ . 将  $P_3^*$  与  $P_2$  通过拼接操作进行特征重构, 获得特征图  $D_2$ , 从而加强浅层小目标的特征表达能力, 提高检测精度.

### 1.4 损失函数

损失函数由定位损失函数和分类损失函数两部分组成, 即

$$\text{Loss} = \frac{1}{N_{\text{pos}}} \left( \sum_i L_{\text{loc}}^i(t^u, v) + \sum_j L_{\text{cls}}^j(p_t) \right). \quad (4)$$

其中:  $N_{\text{pos}}$  表示锚框正样本数量,  $i$  表示所有的正样本,  $j$  表示所有样本.  $L_{\text{loc}}$  表示定位损失函数,  $L_{\text{cls}}$  表示分类损失函数. 模型的定位损失函数是预测框与真实框之间的  $\text{smooth}_{L_1}$  损失, 即

$$L_{\text{loc}}(t^u, v) = \sum_{i \in \{x, y, w, h\}} \text{smooth}_{L_1}(t_i^u - v_i). \quad (5)$$

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2, & |x| < 1; \\ |x| - 0.5, & \text{otherwise.} \end{cases} \quad (6)$$

其中:  $t^u = (t_x^u, t_y^u, t_w^u, t_h^u)$  表示预测框的位置参数,  $v = (v_x, v_y, v_w, v_h)$  表示真实框的位置参数,  $x, y, w, h$  分别表示框的坐标、宽和高.

分类损失采用如下Focal Loss函数:

$$L_{\text{cls}}(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t). \quad (7)$$

$$p_t = \begin{cases} p, & y = 1; \\ 1 - p, & \text{otherwise.} \end{cases} \quad (8)$$

$$\alpha_t = \begin{cases} \alpha, & y = 1; \\ 1 - \alpha, & \text{otherwise.} \end{cases} \quad (9)$$

其中: 权重因子  $\alpha$  取值为 0.25;  $(1 - p_t)^\gamma$  为调节因子,  $\gamma$  取值为 2.0, 是可调节的聚焦参数,  $p$  表示预测样本属于 1 的概率。

## 2 实验结果与分析

### 2.1 实验数据集

为了验证本文方法的有效性, 在目标检测任务的 TinyPerson 数据集<sup>[11]</sup>上进行实验验证及分析。TinyPerson 数据集主要针对微小行人进行检测, 包含 1 610 张图片, 其中 794 张图片用于训练, 816 张图片用于测试。数据集中的目标大部分来自自然场景, 目标密集, 具有较多的小目标物体, 评估标准也更加严格, 要求算法具有更精确的检测能力。

### 2.2 实验设置与评价指标

本文实验使用的 Linux 系统版本为 Ubuntu 16.04, GPU 为 Nvidia RTX2080Ti, 显存为 12 G。实验使用 Pytorch 深度学习框架对模型进行搭建、训练和测试, CUDA 版本为 10.2, Python 版本为 3.7。主干网络分支 1 的权重在 ImageNet 上进行预训练。网络模型整体在 COCO 数据集<sup>[28]</sup>上进行预训练, COCO 数据集经过 SM 算法<sup>[11]</sup>, 使得 COCO 中的目标尺寸与 TinyPerson 数据集尺寸相似, 可解决因目标数据集少、模型复杂造成的过拟合现象。整体模型在 COCO 数据集上进行训练时, 初始学习率设置为 0.01, batchsize 设置为 2, 设置最大训练 epoch 为 12 个, 在训练到第 8 个 epoch 和第 11 个 epoch 时, 分别将学习率减少为原来的 1/10。采用随机梯度下降算法来优化网络的权重, 其动量设置为 0.9, 衰减为 0.000 1。在 TinyPerson 数据集上进行训练时, 输入图像的尺寸为 640×512, 初始学习率为 0.005, batchsize 设置为 4, 设置最大训练 epoch 为 12 个, 在训练到第 8 个 epoch 和第 11 个 epoch 时, 分别将学习率减少为原来的 1/10。

采用平均精度 (average precision, AP) 和漏检率 (miss rate, MR) 作为评价指标, AP 值越高, 检测器预测的类别与位置效果越好; 反之, MR 值越低, 目标漏检的越少, 检测效果越好;  $AP_{50}^{tiny}$  指的是在区分正负样本的区域交并比 (intersection over union, IOU) 阈值选取在 0.5 的基础上, 检测器获取的微小行人 (像素在 1~20 之间) 检测精度; tiny 1、tiny 2、tiny 3 分别表示检测的小行人像素在 1~8、8~12、12~20 之间。  $AP_{25}^{tiny}$  与  $AP_{75}^{tiny}$  分别表示在区分正负样本的区域交并比 (intersection over union, IOU) 阈值选取在

0.25 和 0.75 的基础上, 检测器获取的微小行人目标检测精度。  $AP_{50}^{small}$  指的是在区分正负样本的区域交并比 (intersection over union, IOU) 阈值选取在 0.5 的基础上, 检测器获取的小目标 (像素在 20~32 之间) 检测精度。同理, MR 表示微小行人在不同区间的漏检率。

### 2.3 实验结果分析

本节比较了本文提出的 AF-RetinaNet 算法与近些年基于卷积神经网络的目标检测算法的性能, 展现了本文提出的目标检测网络模型在微小行人目标检测方面的良好性能。本文的研究基准是 RetinaNet 经过 SM 算法所得。

为了研究其在 AF-RetinaNet 中的贡献, 采用不同主干网络进行预测。如表 1 所示, 使用 ResNet 50 的 AF-RetinaNet 算法, 获得检测精度为 55.36%, 比使用 ResNext 101-32x4d 和 ResNet 101 分别提高了 0.32%、0.86%。结果表明, 更深层次的主干网络对微小行人的特征提取存在弱化情况, 不利于微小目标检测。因此, 本文的工作均基于 ResNet 50 进行实现。\* 表示研究基准。

表 1 主干网络实验结果对比

method	Backbone	$AP_{50}^{tiny}$	$AP_{50}^{small}$
RetinaNet*	ResNet 101	50.81	68.76
	ResNext 101-32x4d	50.86	67.70
	ResNet 50	51.21	69.04
AF-RetinaNet	ResNet 101	54.50	70.57
	ResNext 101-32x4d	55.04	70.09
	ResNet 50	55.36	71.95

为了解特征增强模块如何有助于训练 AF-RetinaNet, 首先针对不同的融合方式进行训练, 其次对该模块的输出层数进行训练。如表 2 所示, 采用元素点乘相加 (element point multiplication and addition, EPMA) 融合方式, 获得检测精度为 54.25%, 比元素点乘 (element point multiplication, EPM) 融合方式提高了 0.56%。在 EPMA 融合方式基础上, 添加增强特征  $B_6$  层输出, 获得检测精度为 54.50%; 添加特征  $B_1$ , 获得检测精度为 54.74%。结果表明, 使用 EPMA 融合方式以及输出  $B_1$  和  $B_6$ , 特征增强模块对 AF-RetinaNet 性能贡献最佳。

表 2 特征增强模块实验结果对比

method	$AP_{50}^{tiny}$	$AP_{25}^{tiny}$	$AP_{75}^{tiny}$
	51.21	71.42	6.16
EPM	53.69	73.14	7.11
EPMA	54.25	73.10	7.43
EPMA+ $B_6$	54.50	73.28	7.77
EPMA+ $B_6$ + $B_1$	54.74	73.62	7.57

在这一部分,将通过逐步应用各个模块来分析网络模型中每个模块的效果.实验结果如表3所示,√表示模型含有该模块.由表3中可以看出,FEM,CALM,FRM均能提升AF-RetinaNet对小目标的检测能力.添加FEM模块,通过对目标进行双重提取,在根本上对目标信息进行补充丰富,可以发现在各像素间的小目标检测精度均有所提升.其中检测精度 $AP_{50}^{tiny}$ 值为54.74%,与基准相比提高了3.53%,表明该模块能有效缓解小目标表示微弱的情况.添加CALM模块,针对不同尺度、同一像素间的上下文信息进行学习,可缓解误检情况,增强检测效果.其中检测精度 $AP_{50}^{tiny}$ 值为53.71%,与基准相比提高了2.50%.添加FRM模块,主要针对浅层的微小目标进行精细检测,经过特征放大,可加强更多微弱目标的特征表达能力.结合表3可以看出,其检测精度 $AP_{50}^{tiny}$ 值为54.19%,与基准相比提高2.98%.虽然在

两两融合之后,检测效果会出现不同程度上的抑制,但是当三者均存在时,能够缓解抑制情况.

表3 AF-RetinaNet算法每个模块的影响效果

FEM	CALM	FRM	$AP_{50}^{tiny}$	$AP_{50}^{tiny1}$	$AP_{50}^{tiny2}$	$AP_{50}^{tiny3}$	$AP_{50}^{small}$
			51.21	30.61	57.06	63.58	69.04
√			54.74	33.55	60.90	66.68	69.80
	√		53.71	32.50	59.54	65.78	71.55
		√	54.19	34.07	60.43	65.57	69.70
√	√		54.39	32.99	61.14	65.98	70.76
√		√	54.82	33.68	61.86	66.04	70.45
	√	√	54.75	34.37	60.82	66.40	71.58
√	√	√	55.36	34.30	62.35	66.60	71.95

在TinyPerson数据集上将AF-RetinaNet与近年主流的目标检测算法进行比较.作为检测任务的评价标准,AP值越高,检测器性能越好;反之,MR值越低,检测器性能越好.表4和表5展示了不同检测方法

表4 不同检测器在TinyPerson数据集上的AP值

method	$AP_{50}^{tiny}$	$AP_{50}^{tiny1}$	$AP_{50}^{tiny2}$	$AP_{50}^{tiny3}$	$AP_{50}^{small}$	$AP_{25}^{tiny}$	$AP_{75}^{tiny}$
FCOS <sup>[29]</sup>	17.90	2.88	12.95	31.15	40.54	41.95	1.50
SSD512 <sup>[8]</sup>	34.00	13.54	35.16	48.73	57.14	61.21	2.52
RetinaNet <sup>[9]</sup>	33.53	12.24	38.79	47.38	48.26	61.51	2.28
Grid RCNN <sup>[30]</sup>	47.14	30.65	52.21	57.21	62.48	68.89	6.38
Faster RCNNt-FPN <sup>[6]</sup>	47.35	30.25	51.98	58.95	63.18	63.18	5.83
NAS-FPN <sup>[31]</sup>	37.75	26.71	40.69	45.33	52.63	66.24	3.10
Swin-T <sup>[32]</sup>	40.52	31.92	41.67	47.06	52.53	59.42	4.24
RetinaNet-MSM <sup>[11]</sup>	49.59	31.63	56.01	60.78	63.38	71.24	6.16
RetinaNet-SM with S- $\alpha$ <sup>[12]</sup>	52.56	33.90	58.00	63.72	65.69	73.09	6.64
RetinaNet-AFM- $\alpha$ <sup>[18]</sup>	48.81	28.20	55.54	59.69	61.87	71.90	5.70
Faster RCNN-AFM- $\alpha$ <sup>[18]</sup>	50.35	33.82	55.43	61.48	65.59	71.15	6.80
Faster RCNN-RFLA <sup>#</sup> <sup>[19]</sup>	49.99	30.20	53.40	63.67	69.32	72.53	5.36
FCOS-RFLA <sup>#</sup> <sup>[19]</sup>	37.32	15.41	37.72	52.74	62.84	64.04	3.90
RetinaNet*	51.21	30.61	57.06	63.58	69.04	71.42	6.16
AF-RetinaNet	55.36	34.30	62.35	66.60	71.95	74.24	8.04
AF-RetinaNet <sup>+</sup>	56.78	36.36	63.29	68.77	72.91	75.53	8.20

表5 不同检测器在TinyPerson数据集上的MR值

method	$MR_{50}^{tiny}$	$MR_{50}^{tiny1}$	$MR_{50}^{tiny2}$	$MR_{50}^{tiny3}$	$MR_{50}^{small}$	$MR_{25}^{tiny}$	$MR_{75}^{tiny}$
FCOS <sup>[29]</sup>	96.28	99.23	96.56	91.67	84.16	90.34	99.56
SSD512 <sup>[8]</sup>	93.56	94.55	90.42	85.54	76.79	82.80	99.23
RetinaNet <sup>[9]</sup>	92.66	94.52	88.24	86.52	82.84	81.95	99.13
Grid RCNN <sup>[30]</sup>	87.96	88.31	82.79	79.55	73.16	78.27	98.21
Faster RCNNt-FPN <sup>[6]</sup>	87.57	87.86	82.02	78.78	72.56	76.59	98.39
NAS-FPN <sup>[31]</sup>	92.41	90.37	87.41	87.50	81.78	77.79	99.20
Swin-T <sup>[32]</sup>	89.91	87.20	85.44	85.31	80.28	82.36	98.89
RetinaNet-MSM <sup>[11]</sup>	88.39	87.80	79.23	79.77	72.18	76.25	98.57
RetinaNet-SM with S- $\alpha$ <sup>[12]</sup>	87.00	87.62	79.47	77.39	69.25	74.72	98.41
RetinaNet-AFM- $\alpha$ <sup>[18]</sup>	87.42	89.43	81.03	79.36	72.46	74.37	98.48
Faster RCNN-AFM- $\alpha$ <sup>[18]</sup>	85.00	87.69	79.90	75.53	68.55	73.66	98.13
Faster RCNN-RFLA <sup>#</sup> <sup>[19]</sup>	87.76	89.92	81.93	78.79	68.44	75.38	98.60
FCOS-RFLA <sup>#</sup> <sup>[19]</sup>	91.46	95.02	88.51	83.67	73.63	80.99	98.87
RetinaNet*	89.05	90.54	83.36	80.35	69.81	79.23	98.60
AF-RetinaNet	86.08	88.26	79.00	76.63	65.09	73.83	98.27
AF-RetinaNet <sup>+</sup>	85.38	87.68	78.20	75.00	64.42	72.35	98.25

在AP与MR之间的性能. 其中\*表示研究基准, +表示多尺度训练, #表示结果经过实验所得. AF-RetinaNet较RetinaNet\*精度提升4.15%,漏检率降低2.97%. 在进行多尺度训练时,较RetinaNet\*精度提升5.57%,漏检率降低3.67%.

为了进一步展示AF-RetinaNet对微小行人的检测效果,给出AF-RetinaNet与基准网络RetinaNet在

TinyPerson数据集上的图片检测效果对比,如图5所示. 图5为在不同光照影响、目标存在遮挡、前背景目标相似的环境下,小目标的检测结果. 可以看出,基于RetinaNet研究的基准算法对小目标、前背景相似目标、密集目标的检测效果较差,存在目标漏检、误检的情况. 而AF-RetinaNet算法可以较好地检测出图像中的行人目标,缓解漏检、误检情况.

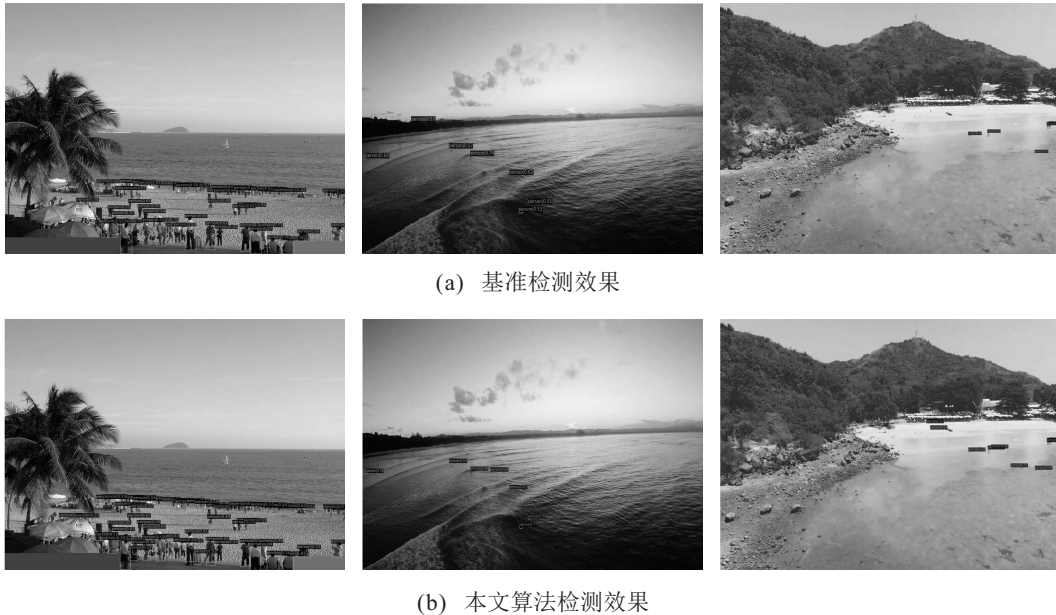


图5 检测效果对比

### 3 结论

为了解决微小行人在图像中所占区域小、分辨率低、涵盖信息少、容易被误检和漏检等问题,本文基于RetinaNet算法进行改进,提出一种基于自适应融合与特征细化的微小行人检测算法AF-RetinaNet. 通过引入特征增强模块,补充目标特征信息,以及自适应融合目标的上下文信息,减少了误检情况;使用特征细化模块,采用嵌入超分来增强微小行人特征的表达能力,缓解漏检问题的同时减少了计算成本. 将本文的模型算法在TinyPerson数据集上进行实验验证,结果具有优化目标检测的效果,表明了该算法的有效性. 在以后的研究工作中,将进一步优化网络结构,对数据集进行处理,提高算法检测性能,从而获得更高的准确率.

#### 参考文献(References)

- [1] Liang F T, Zhou Y, Chen X, et al. Review of target detection technology based on deep learning[C]. Proceedings of the 5th International Conference on Control Engineering and Artificial Intelligence. Sanya, 2021: 132-135.
- [2] Kato T, Relator R, Ngouy H, et al. Segmental HOG: New

descriptor for glomerulus detection in kidney microscopy image[J]. BMC Bioinformatics, 2015, 16(1): 316.

- [3] 王庆伟, 应自炉. 一种基于Haar-Like T特征的人脸检测算法[J]. 模式识别与人工智能, 2015, 28(1): 35-41. (Wang Q W, Ying Z L. A face detection algorithm based on haar-like T features[J]. Pattern Recognition and Artificial Intelligence, 2015, 28(1): 35-41.)
- [4] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, 2014: 580-587.
- [5] Girshick R. Fast R-CNN[C]. 2015 IEEE International Conference on Computer Vision. Santiago, 2016: 1440-1448.
- [6] 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.
- [7] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 779-788.
- [8] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot MultiBox detector[C]. Computer Vision—ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.

- [9] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]. 2017 IEEE International Conference on Computer Vision. Venice, 2017: 2999-3007.
- [10] 贲晔, 徐森, 王科俊. 行人步态的特征表达及识别综述[J]. 模式识别与人工智能, 2012, 25(1): 71-81. (Ben X Y, Xu S, Wang K J. Review on pedestrian gait feature expression and recognition[J]. Pattern Recognition and Artificial Intelligence, 2012, 25(1): 71-81.)
- [11] Yu X H, Gong Y Q, Jiang N, et al. Scale match for tiny person detection[C]. 2020 IEEE Winter Conference on Applications of Computer Vision. Snowmass, 2020: 1246-1254.
- [12] Gong Y Q, Yu X H, Ding Y, et al. Effective fusion factor in FPN for tiny object detection[C]. 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa, 2021: 1159-1167.
- [13] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 936-944.
- [14] Hong M B, Li S W, Yang Y C, et al. SSPNet: Scale selection pyramid network for tiny person detection from UAV images[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.
- [15] 赵亮, 刘世鹏. 全局与局部图像特征自适应融合的小目标检测算法[J]. 控制与决策, 2023, 38(4): 935-943. (Zhao L, Liu S P. Small target detection algorithm based on adaptive fusion of global and local image features[J]. Control and Decision, 2023, 38(4): 935-943.)
- [16] 柳长源, 王琪, 毕晓君. 多目标小尺度车辆目标检测方法[J]. 控制与决策, 2021, 36(11): 2707-2712. (Liu C Y, Wang Q, Bi X J. Multi-target and small-scale vehicle target detection method[J]. Control and Decision, 2021, 36(11): 2707-2712.)
- [17] 杨其利, 周炳红, 郑伟, 等. 基于全卷积递归网络的弱小目标检测方法[J]. 光学学报, 2020, 40(13): 49-61. (Yang Q L, Zhou B H, Zheng W, et al. Dim and small target detection based on fully convolutional recursive network[J]. Acta Optica Sinica, 2020, 40(13): 49-61.)
- [18] Li Y C, Zhou S L, Chen H. Attention-based fusion factor in FPN for object detection[J]. Applied Intelligence, 2022, 52(13): 15547-15556.
- [19] Xu C, Wang J W, Yang W, et al. RFLA: Gaussian receptive field based label assignment for tiny object detection[J/OL]. 2022, arXiv: 2208.08738.
- [20] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 770-778.
- [21] 谭红臣, 李淑华, 刘彬, 等. 特征增强的SSD算法及其在目标检测中的应用[J]. 计算机辅助设计与图形学学报, 2019, 31(4): 573-579. (Tan H C, Li S H, Liu B, et al. Feature enhancement SSD for object detection[J]. Journal of Computer-Aided Design & Computer Graphics, 2019, 31(4): 573-579.)
- [22] Nie J, Anwer R M, Cholakkal H, et al. Enriched feature guided refinement network for object detection[C]. 2019 IEEE/CVF International Conference on Computer Vision. Seoul, 2020: 9536-9545.
- [23] 陈鸿坤, 罗会兰. 多尺度语义信息融合的目标检测[J]. 电子与信息学报, 2021, 43(7): 2087-2095. (Chen H K, Luo H L. Multi-scale semantic information fusion for object detection[J]. Journal of Electronics & Information Technology, 2021, 43(7): 2087-2095.)
- [24] Liu S T, Huang D, Wang Y H. Learning spatial fusion for single-shot object detection[J/OL]. 2019: arXiv: 1911.09516.
- [25] Haris M, Shakhnarovich G, Ukita N. Task-driven super resolution: Object detection in low-resolution images[C]. Communications in Computer and Information Science. Cham: Springer International Publishing, 2021: 387-395.
- [26] Bai Y C, Zhang Y Q, Ding M L, et al. SOD-MTGAN: Small object detection via multi-task generative adversarial network[C]. Computer Vision—ECCV 2018. Cham: Springer International Publishing, 2018: 210-226.
- [27] Deng C F, Wang M M, Liu L, et al. Extended feature pyramid network for small object detection[J]. IEEE Transactions on Multimedia, 2022, 24: 1968-1979.
- [28] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common objects in context[C]. Computer Vision—ECCV 2014. Cham: Springer International Publishing, 2014: 740-755.
- [29] Tian Z, Shen C H, Chen H, et al. FCOS: Fully convolutional one-stage object detection[C]. 2019 IEEE/CVF International Conference on Computer Vision. Seoul, 2020: 9626-9635.
- [30] Lu X, Li B Y, Yue Y X, et al. Grid R-CNN[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, 2020: 7355-7364.
- [31] Ghiasi G, Lin T Y, Le Q V. NAS-FPN: Learning scalable feature pyramid architecture for object detection[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, 2020: 7029-7038.
- [32] Liu Z, Lin Y T, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. 2021 IEEE/CVF International Conference on Computer Vision. Montreal, 2022: 9992-10002.

## 作者简介

邵香迎(1997—), 女, 硕士生, 从事计算机视觉、目标检测等研究, E-mail: 2639407478@qq.com;

郭颖(1985—), 男, 副教授, 博士, 从事计算机视觉、微弱信号检测等研究, E-mail: yguo@nuist.edu.cn;

王友伟(1998—), 男, 硕士生, 从事计算机视觉、目标检测等研究, E-mail: 2638901386@qq.com.