

DOI: 10.13382/j.jemi.B2205809

UAST-RCNN: 遮挡行人的目标检测算法^{*}

刘毅 于畅洋 李国燕 潘玉恒

(天津城建大学计算机与信息工程学院 天津 300384)

摘要:为解决行人检测中对遮挡行人的检测度低,漏检率较高的问题,提出一种基于注意力机制的 UAST-RCNN 网络,其在 Faster-RCNN 网络的基础上进行改进。首先,选用 Swin-Transformer 作为骨干网络,通过采用一种窗口多头自注意力机制提升全局感受野;然后,通过层级重采样模块,改进特征金字塔提升特征样本的质量,并且引入渐进式焦点损失函数平衡正负样本;最后,在实验预处理阶段采用改进的数据预处理扩充 City Persons 数据集进行多尺度训练。实验结果表明该算法对比原模型在遮挡行人检测上有了明显提升,其中在检测精度(AP)提升了 6.3%,漏检率(MR)下降了 4.1%。验证了所提算法在行人检测的可行性,可满足遮挡行人场景的检测要求。

关键词: Faster-RCNN; 行人检测; 特征金字塔; Swin-Transformer

中图分类号: TP391; TP181; TN99 **文献标识码:** A **国家标准学科分类代码:** 520.2060

UAST-RCNN: Object detection algorithm for blocking pedestrians

Liu Yi Yu Changyang Li Guoyan Pan Yuheng

(School of Computer and Information Engineering, Tianjin Chengjian University, Tianjin 300384, China)

Abstract: In order to solve the problem of low detection degree and high missed detection rate of occluded pedestrians in pedestrian detection, an attention mechanism based UAST-RCNN network is proposed, which is improved on the basis of Faster-RCNN network. Firstly, Swin-Transformer is selected as the backbone network to improve the global receptive field by using a window multi-head self-attention mechanism. Then, the feature pyramid is improved for the quality of feature samples through the hierarchical resampling module, and a progressive focus loss function is introduced to balance positive and negative samples. Finally, in the preprocessing stage of the experiment, the improved data preprocessing was used to extend the City Persons dataset for multi-scale training. The experimental results show that the algorithm has a significant improvement in the detection of occluded pedestrians compared with the original model, in which the AP is increased by 6.3%, and the MR (miss rate) is decreased by 4.1%. The feasibility of the proposed algorithm in pedestrian detection is verified, and it can meet the detection requirements of occluded pedestrian scene.

Keywords: Faster-RCNN; pedestrian detection; feature pyramid; Swin-Transformer

0 引言

在计算机视觉中,行人检测是其重要的研究任务。在行人检测中行人的定位主要通过图像处理和机器学习的方法实现,相对精确的预测出行人的定位。构建行人检测系统首先需要精确的行人检测算法以便于实现后续如行人追踪、行人重识别、检索等图像的智能分析^[1]。随

着卷积神经网络算法的出现,通用场景下行人检测算法已能取得相对良好的检测效果,但是在复杂场景下,如密集行人遮挡、小目标行人检测等的检测性能仍需要提升。遮挡问题可分为类内遮挡和非目标物体遮挡。类内遮挡即高密度行人之间的相互遮挡,产生相互检测框的抑制影响目标的阈值^[2]。非目标物体遮挡是指在实际道路中,车辆、建筑物等非行人障碍物对行人的遮挡,使得行人特征损失对检测框的形成造成困难。且行人尺度不

收稿日期:2022-09-09 Received Date: 2022-09-09

^{*} 基金项目:国家自然科学基金(61876131)、天津市教委科研计划项目(2016CJ12)资助

同,行人因服饰体态等特点使得行人外观差异较大,同一场景下的行人可能会有多种姿态,这都给行人检测带来了新的问题^[3]。最早最完善的行人检测算法是方向梯度直方图(histogram of oriented gradient, HOG)^[4-5]在 CVPR2005 上由法国的研究人员 Dalal 和 Triggs 最先提出。提取特征时应用梯度算子,分类器选择支持向量机(support vector machine, SVM)进行分类,其将检测问题转化为分类问题也为后续的研究提供了思路。随着卷积神经网络的飞速发展,行人检测算法也更进一步,基于卷积神经网络并进行区域划分而设计出的区域卷积神经网络(region-convolution neural network, R-CNN)^[6-8]算法使行人检测算法的研究获得了新方法。针对行人遮挡问题,Zhang 等^[9]提出 OR-CNN 算法,通过改进原算法损失函数以及对候选框进行再分类的操作,减少了遮挡行人误检框的出现;李奇^[10]提出的 YOLOv3 目标检测算法平衡了检测实时性和精度,特征提取网络在 ResNet^[11]的基础上改进并融入了特征金字塔网络(feature pyramid networks, FPN)^[12]结构,提升网络在多尺度的特征提取能力,因此在处理遮挡问题上有着突出的效果。虽然近年来卷积神经网络的出现使行人检测有了性能上的提升,但其在遮挡度高的复杂场景下检测效果一般。

综上所述,现有在行人检测领域中对遮挡行人的检

测精度低,漏检率较高的问题。针对此问题本文提出一种基于注意力机制的 UAST-RCNN 网络,此网络选用 Swin-Transformer^[13]替换 Faster-RCNN 的骨干网络,并优化特征金字塔中各层的特征提取,最后使用渐进式焦点损失策略解决对特征提取过程中正负样本不均衡,从而进一步解决遮挡场景下检测精度低,漏检率较高的问题。

1 行人检测网络

1.1 Faster-RCNN

Faster-RCNN 因其较高的准确度成为目标检测领域流行的网络模型。该算法以 Fast-RCNN 为基础提出了 RPN 候选框生成算法,解决了目标定位难,生成候选框耗时长等问题。算法的检测精度对行人检测十分重要,所以本文选用 Faster-RCNN 算法进行改进。Faster-RCNN 算法原理如图 1 所示,通过由卷积神经网络组成的骨干网络(Backbone)提取特征,提取特征后通过特征金字塔进行特征融合并输入 RPN 生成候选框区域。根据生成的候选框获得相对应的感兴趣区域(region of interests, RoI)并输入 ROI 检测头生成特征层,之后进行分类与回归,生成边框类别和边框位置。

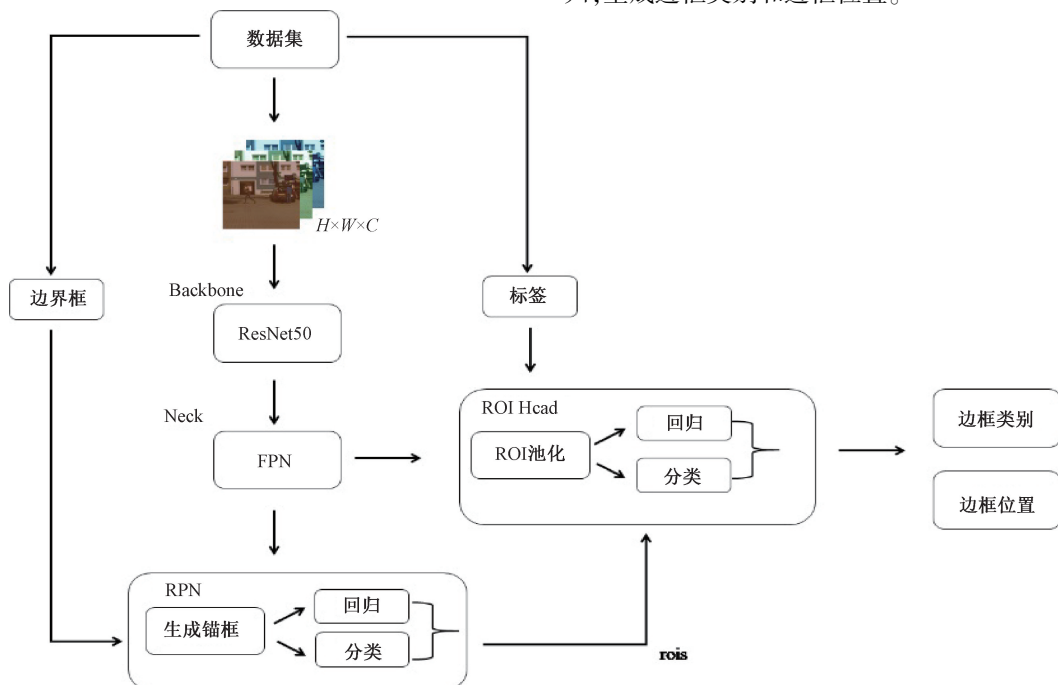


图 1 Faster-RCNN 网络结构图

Fig. 1 Faster-RCNN network structure diagram

本文将基于注意力机制的 Swin-Transformer 应用于 Faster-RCNN 模型作为骨干网络,并加入多级优化范例模块进行渐进式挖掘,优化深层特征提取,提升模型对行人

类内遮挡的识别度,在训练时采用改进自动数据增强的训练策略,得到一种提高遮挡场景下行人检测精度的网络模型。

1.2 Swin-Transformer

以卷积神经网络为主的特征提取模块在计算机视觉中取得了突破性的进展。但最近的研究表明基于注意力机制的框架也可以在特征提取任务上表现优异,本文使用 Swin-Transformer 作为骨干网络^[14],用于提取图像特征。Swin-Transformer 是微软最近提出基于 Vision-Transformer 框架改进的模型。此模型具有注意力机制关注全局信息建模的能力,且通过跨窗口移位增加了各个窗口间的连接性,使此网络模型提升对相邻窗口之间的特征信息交互,增加了对完整图片的特征提取率^[15-16]。主体网络结构如图 2 所示。

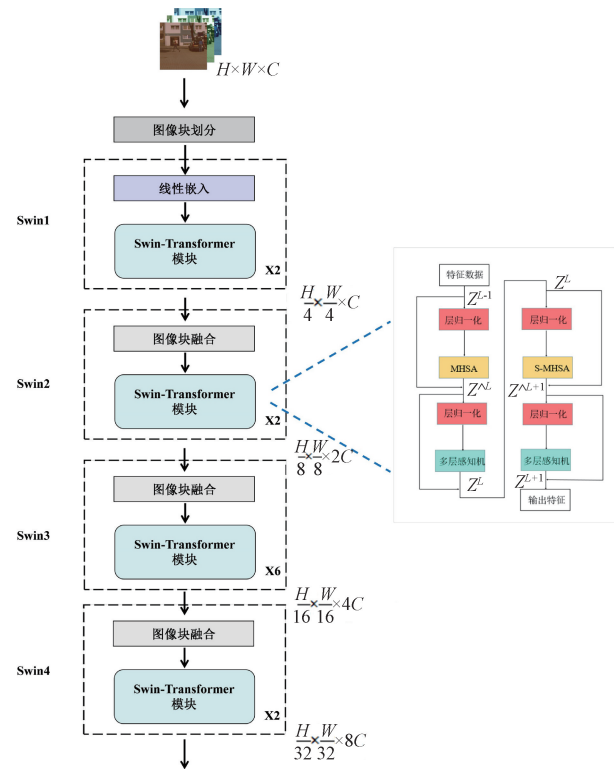


图 2 骨干网络结构图
Fig. 2 Backbone network structure diagram

本文选取了体量最小的 Swin-Transformer-tiny,其整体架构可分为 4 层,其中每层对应连续的 Swin-Transformer 模块。第 1、2、4 层为 2 块,第 3 层为 6 块。每一层都会相应地缩小输入图片的分辨率,扩大整体网络感受野。此网络需要将输入的图片进行裁剪,裁剪后进行线性嵌入并输入 Swin-Transformer 模块中,后 3 层再输入之前需要进行图块的融合降低图片分辨率,起到 CNN 中池化层的效果。每组 Swin-Transformer 模块是由两个不同的结构组成,第 1 个结构包括多头窗口自注意力机制 (multi head -self attention, MSA)、层归一化模块 (layer normaliazation, LN) 和多层感知机 (multi layer perceptron, MLP) 组成。通过残差连接彼此进行信息交互,之后特征

信息进入第 2 个相似的结构,不同点在于将多头窗口自注意力机制替换为滑窗多头自注意力机制,先对特征数据进行窗口滑动再进行计算,实现了多窗口的信息交互。

但多头自注意力机制的引入也带来了参数量增大的问题, Swin-Transformer 引入窗口多头自注意力机制 (window -multi head self attention, W-MSA) 将注意力的计算限制在相邻的窗口内,并采用了一种滑窗多头自注意力机制 (shifted window -multi head self attention, SW-MSA) 增强相邻窗口间信息交互。如图 3 所示,在第 L 层根据图片的高宽比例将图片划分成窗口并在窗口内执行自注意力运算,并在 L+1 层进行滑动窗口,在此图片上滑动划分出新的窗口执行自注意力运算。新划分的窗口包含了 L 层相邻窗口的边界,实现了跨窗口通信提供了窗口间的信息交互。

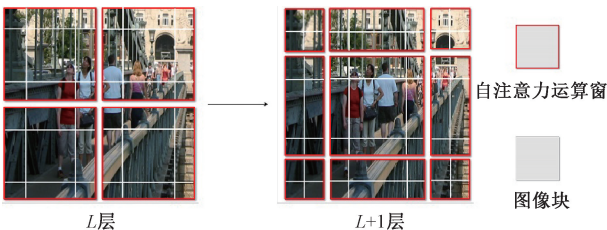


图 3 窗口多头自注意力运算机制
Fig. 3 Window multi-head self-attention operation mechanism

在进行滑动窗口多头自注意力时生成了重新划分的第 2 批窗口。为避免窗口数量的增多造成的模型运算量增加,滑动窗口多头自注意力中添加了循环移位操作来解决此问题如图 4 所示。通过将输入图片进行旋转左移合成一个新的待处理窗口,包含了不相邻窗口的信息,并添加了掩码操作将自注意力的运算维持在子窗口中进行,此操作避免了窗口数的增加,有效减少了模型运算量。

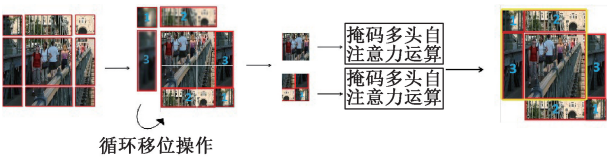


图 4 滑动窗口计算
Fig. 4 Sliding window calculation

窗口与滑动窗口多头自注意力的引入降低了多头自注意力的运算复杂度公式 (式 (1)、(2)),为多头自注意力与窗口多头自注意力的复杂度对比公式。

$$\Omega(MSA) = 4HWC^2 + 2(HW)^2C \tag{1}$$

$$\Omega(W-MSA) = 4HWC^2 + 2M^2HWC \tag{2}$$

式中: H 为输入图片的高, W 为输入图片的宽, C 为输入图片维度, M 为窗口数。图像块数取决于输入图像的高

宽,因此算法需要的窗口数(一般为 7)远小于图像块数所 W-MSA 的计算复杂度和图像长宽呈线性关系,明显得出窗口多头自注意力的复杂度优于多头自注意力。

1.3 多级优化范例模块

为减少特征金字塔在逐级特征提取时产生的特征遗漏现象,本文在特征金字塔各级别的特征提取方面进行了进一步优化,采用多级优化范例模块^[17]。此模块包括两个部分:1)通过设立一个独立的重采样模式,应用各级重新采样来监控每一级特征金字塔。2)采用了一种渐进式的焦点损失函数,通过预测每级特征金字塔的正样本逐步调整单级的突出局部损失。这样的设计使模型在各级特征提取中得到补充,去除无关的背景元素,提升网络对遮挡行人的识别率,也在一定程度上减少了训练时间。

1) 层级重采样

层级重采样不同于其他多级检测器中用迭代重采样机制进行优化来解决不平衡现象,此种机制可以排除低质量特征样本(正负样本的比例相差过大)并且防止特征样本质量不匹配。式(3)为用各级重采样的平均值定义的总分类损失:

$$\text{Loss}_{cls} = \frac{1}{L} \sum_{l=1}^L \text{Loss}_l(Z_l, Y_l) \quad (3)$$

其中, L 为此一阶段网用于预测的特征金字塔层数, Z_l 为第 l_1 层的预测结果, Y_l 是此目标分配策略的标签, Loss_l 为此 l_1 层的总损失,此层级重采样损失策略避免了提取特征时的特定样本不均衡,使层级特征金字塔更加灵活,减少了障碍物对人体检测的影响。

2) 渐进式焦点损失函数

针对特征金字塔水平的不平衡性和多样性,且不同层级中的应用中超参数的不同^[18]。此网络中不同于传统的焦点损失函数,本文提出了一个超参数动态覆盖调整策略并使用渐进式焦点损失函数,可以根据特征金字塔的不同层级的收敛情况动态调整攫取特征的强度。式(4)为二分类中的 SIGMOID 焦点损失函数。

$$\text{FL}(z_i, y_i) = \begin{cases} -\alpha(1-z_i)^\beta \log(z_i), & y_i = 1 \\ - (1-\alpha) z_i^\beta \log(1-z_i), & y_i = 0 \end{cases} \quad (4)$$

式中:参数 α 是便于调节的恒定值,保持着动态采样过程中正负样本的梯度平衡,而超参数 β 通过动态采样的方法缓解由难易样本不平衡导致的预测概率^[19]。此模型使用的调整策略可以自动调整每层特征金字塔的收敛情况如式(5)所示。

$$\text{PFL}(z_i, y_i) = \begin{cases} -\alpha_*(1-z_i)^{\beta_*} \log(z_i), & y_i = 1 \\ - (1-\alpha_*) z_i^{\beta_*} \log(1-z_i), & y_i = 0 \end{cases} \quad (5)$$

此损失函数主要调整了超参数 α 和 β, β_* 的设计考虑到每层正样本的独立性,并且 α_* 的设计遵循 β_* 的调节如式(6)、(7)所示。

$$\beta_* = -\log\left(\frac{1}{n_{pos}} \sum_{i=1}^n y_i \cdot z_i\right) \quad (6)$$

$$\alpha_* = w/\beta_* \quad (7)$$

此调节模式保持正负样本之间的平衡, β_* 的作用类似于交叉熵损失且很自然的反映了特征金字塔的水平收敛情况, y_i 指特定层特征金字塔中一个样本所分配的标签,并且 z_i 为相应的概率,所以 $y_i \cdot z_i$ 表示正样本的概率,负样本定义为 0, n 为每层特征金字塔的总样本数, n_{pos} 为每层特征金字塔的正样本总数。公式中 α_* 的调整遵循 β_* 且通常情况下 w 为常数,表示 α_* 和 β_* 的负相关性,也有利于层级别的重新采样。这种动态调整策略不会增加额外的计算复杂度并且提升模型对行人预测的准确率。改进后的网络结构如图 5 所示。

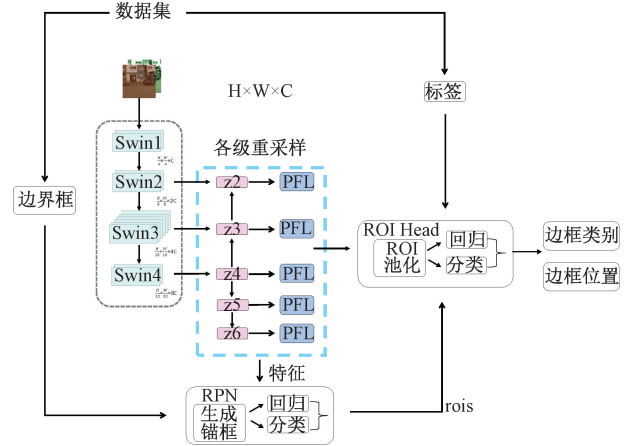


图 5 整体网络结构图

Fig. 5 Overall network structure diagram

1.4 改进的自动数据增强策略

Swin-Transformer 的训练需要大量的数据集进行补充,本文引入并改进了一种基于搜索的数据增强策略,此策略由两部分组成:1)此策略需要创建数据增强的搜索空间并直接在一些数据集上评估特定策略的质量。2)每一种策略由多个子策略组成,在每一个批次训练中会随机将子策略应用到图像中,子策略由两个图像处理方法组成,如裁剪、噪声、或剪切等。策略的评价标准由每种操作的使用性质得出^[20]。本文针对行人检测扩充了此策略的搜索空间,加入了新的数据增强方法 HSV 变换,随机对比度,随机裁剪等如图 6 所示

使用上图方法将数据增强策略扩充到搜索空间,之后通过自动学习得到最适合此 CityPersons 数据集的最佳的数据增强策略并应用于此数据集。

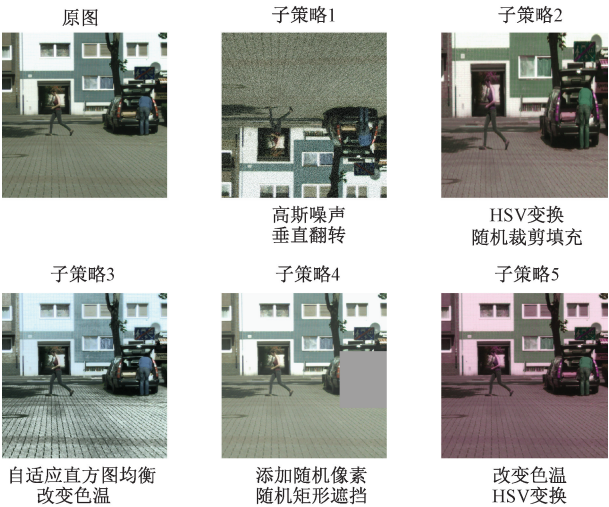


图 6 数据增强策略

Fig. 6 Data enhancement Strategy

2 实 验

2.1 行人检测数据集

为验证本实验行人检测算法,本文采用 CityPersons 数据集^[21]。其中包含来自 50 个不同城市的街道场景并且每幅图像中行人的平均数量为 7,且包含 3 种标注标签分别是行人框、可见行人框、语义分割图。此数据集分为训练集、验证集、测试集 3 部分。其中有 2 975 张图片用于行人检测的训练,500 张图片用于训练过程中的验证,1 575 张图片用于训练完成后的性能测试。CityPersons 数据集提供算法评估的接口针对通用行人和遮挡行人有不同的评估标准,可用其测试子集测试人体可见比例。

2.2 实验环境

为满足此实验要求本文实验平台配置如表 1 所示。

表 1 实验软硬件环境配置

Table 1 Experiment software and hardware environment configuration

名称	相关配置
操作系统	Windows 10
中央处理器(CPU)/GHz	Intel(R) Core(TM) i7-7700 ,3. 6
内存/GB	16
GPU	NVIDIA GeForce GTX 1080
GPU 加速器	CUDA10. 2, CUDNN7. 6
深度学习框架	Pytorch

2.3 模型参数设置

本实验在深度学习框架 Pytorch 框架上运行并使用 mmdetection 目标检测工具包其超参数设置为 samples_per_gpu = 1, workers_per_gpu = 1, 因显卡算力有限故设置

每个 GPU 读取图像数量为 1,读取数据时每个 GPU 分配的线程数为 1。其是流行的开源目标检测库,其中包括多种目标检测算法,有利于算法的对比实验。此训练采用 AdamW 优化算法进行优化^[22],初始学习率为 0. 000 1,权重衰减(weight_decay)为 0. 05。Epoch 设定为 150 次,并在第 16, 25, 75 次时降低学习率。

2.4 目标检测评价标准:

本文选择平均精度值(average precision, AP)、召回率(Recall)、漏检率(miss rate, MR)、实时检测速度(FPS)作为此实验的评价指标。AP 的定义公式如下式所示,由召回率与准确率算出。

$$AP = \int_0^1 P(R) dR$$
 (8)

召回率(Recall)的定义如下式所示,表示测试数据集全部目标中被检测到的数量所占比例。

$$Recall = \frac{TP}{TP + FN}$$
 (9)

漏检率(MR)的定义如下式所示,表示测试数据集全部目标漏检数量所占比例,其数值越低表示检测效果越好。

$$MR = 1 - Recall = \frac{FN}{TP + FN}$$
 (10)

上述 3 个公式中 TP 表示预测为正样本且为正确结果, FN 表示预测为负样本但实际为正确结果,该指标对应误检率。FPS 指画面每秒传输的帧数,其数值越大表示检测结果越流畅。

2.5 实验结果与分析

1) 消融实验

本文与 Faster-RCNN 在相同实验环境下和相同数据集上进行目标检测实验。实验首先对比不同骨干网络在行人检测网络中的检测效果,其实验结果如表 2 所示。表中 AP₅₀₋₉₅, AP₅₀, AP₇₅, 分别代表在计算时物体检测框和真实框 IOU 值为 50%~95%之间,和大于 50%, 75% 的平均精度。

表 2 骨干网络消融实验表

Table 2 Experimental table of backbone network ablation

网络模型	骨干网络	AP ₅₀₋₉₅	AP ₅₀	AP ₇₅
Faster-RCNN	ResNet50	34. 2	55. 6	37. 3
Faster-RCNN	Swin-Tiny	36. 8	56. 2	38. 7

由表 2 可以看出更换骨干网络后的 Faster-RCNN 网络在检测精度上有了明显优势,特别在 AP₅₀₋₉₅ 指标提升了 2. 4%。表明其在物体检测框与真实框重合度提升明显。接着采用修改后的特征提取模块对原网络进行消融实验,结果如表 3 所示。

表 3 多级优化模块消融实验表

Table 3 Experimental table of multistage optimized module ablation

网络模型	多级优化 FPN	AP ₅₀₋₉₅	AP ₅₀	AP ₇₅
Faster-RCNN	×	34. 2	55. 6	37. 3
Faster-RCNN	√	37. 5	56. 8	38. 9

由表 3 可看出多级优化模块的加入有效地提升了网络的预测精度,表明层级重采样的提取模式增强了对行人特征的提取。最后使用本文所改进的 UAST-RCNN 与经典的目标检测网络对比,使用相同的实验环境与数据集进行对比实验。为验证其他改进 Faster-R-CNN 网络与本文网络效果,本文选用 Cascade R-CNN 作为双阶段检测模型的代表,并选用 YOLOv3 与 ssd300 作为单阶段检测模型的代表,在相同的实验条件和数据集的实验数据如图 7 所示。

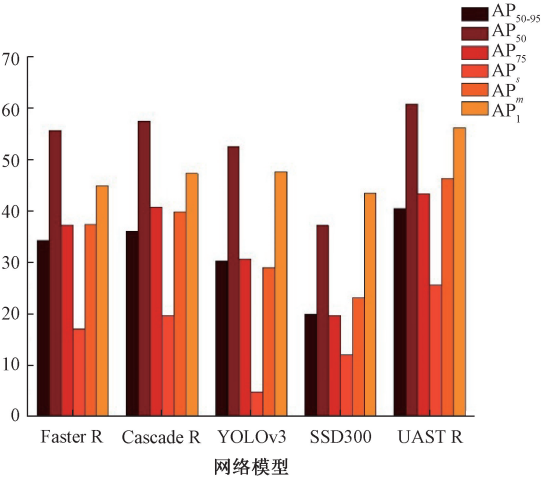


图 7 网络模型对比
Fig. 7 Network model comparison

根据图 7 得到的数据可以看出改进后的 UAST-RCNN 模型在 AP_s 和 AP_m 指标上有较大提升。对比此模型与其他基于 Faster-RCNN 改进的模型中,在中小目标的准确度上有了明显的提高,但其在 AP_l 大尺度目标的提升不够突出,表 4 为与其他网络在其他参数对比表。

表 4 网络模型对比表

Table 4 Network model comparison table

网络模型	Backbone	AP	Param	Flops	FPS
Faster-RCNN	ResNet50	34. 2	41. 5 M	194. 3 G	20
Cascade-RCNN	ResNet50	36. 1	68. 9 M	195. 1 G	18
YOLOv3	Darknet53	29. 0	61. 6 M	193. 9 G	33
SSD300	VGG16	27. 9	24. 8 M	345. 0 G	38
UAST-RCNN	Swin-Tiny	40. 5	47. 4 M	150. 8 G	13

根据表 4 消融实验数据可以看出,且模型所需的参数量也有一定的增加。检测速度没有明显的下降,且对

比单阶段检测模型虽然检测速度仍存在一定的差距,但可在基本满足实时性要求的检测速度下,模型的准确率明显优于其他模型。

2) 遮挡行人对比实验

为进一步探究本文算法在遮挡行人场景下的检测性能,同时也为证明骨干网络及多级优化模块在此场景下的可行性,本文设置了遮挡行人对比实验。实验同样选用 City Persons 作为测试集。其中 48. 8% 的行人相互遮挡的 IOU 高于 0. 1, 26. 4% 的行人相互遮挡的 IOU 高于 0. 3。本文应用漏检率 MR 作为评价指标,并且用 AP_s 辅助验证合理性。控制每张图片的合理锚框数,输出阈值合理的锚框,遮挡行人的对比实验结果如表 5 所示。

表 5 遮挡行人对比实验

Table 5 A comparative experiment of shielding pedestrians

网络模型	MR (漏检率)/%	AP _s
Faster-RCNN	16. 8	17. 1
Cascade R-CNN	15. 4	19. 7
Yolov3	23. 2	4. 8
SSD300	28. 9	12. 1
UAST-RCNN	12. 7	25. 6

由表 5 可以看出本文所提出 UAST-RCNN 算法在遮挡行人场景下的漏检率明显低于原 Faster-RCNN 算法,同时检测精度也高于其他改进算法和单阶段检测算法。证明此骨干网络及多级优化模块在遮挡行人场景下的可行性以及有效性。图 8 为原算法和改进算法在遮挡行人场景下的检测效果图。

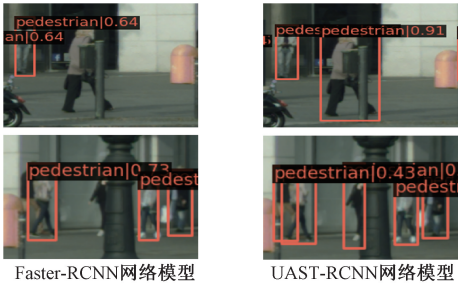


图 8 检测效果对比
Fig. 8 Comparison of detection effects

对比细节可以看出 UAST-RCNN 相对于 ST 检测出类内遮挡和非目标物体(路灯)对行人的遮挡,对目标行人的敏感度更高。对比所框目标的置信度也可以看出 UAST-RCNN 算法的置信度会更高。

3) 测试效果验证

为了较直观地区分优化后的 Faster-RCNN 算法与原始算法在通用行人场景下的检测效果之差,从 City Persons 数据集的测试集中随机抽取两张图像在相同实

验环境下进行测试对比,其检测结果如图 9 所示。其中,图 9(a)和(c)为原始 Faster-RCNN 算法的检测结果,而图 9(b)和(d)为优化后的本文算法的检测结果。由图 9 可以看出,优化后的算法相较于原始算法,降低了遮挡人群场景下的漏检率,也在一定程度上提升了通用行人场景下目标检测的精度。

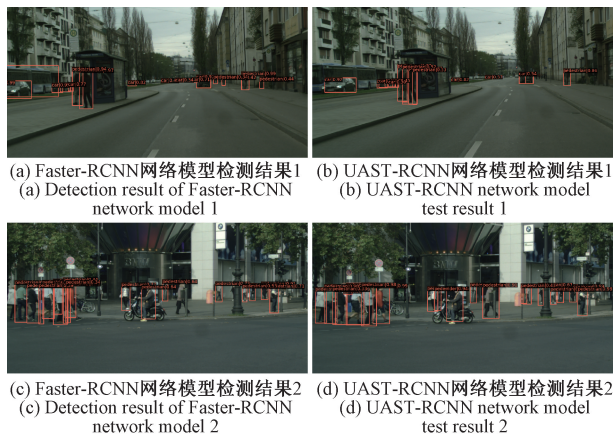


图 9 测试结果对比

Fig. 9 Comparison of test results

3 结 论

本文针对行人检测任务,将 Faster-RCNN 的骨干网络替换为 Swin-Transformer,并通过对特征金字塔的层级重采样避免了提取特征时的特定样本不均衡,又引入了渐进式焦点损失加速模型收敛,提升对遮挡行人和小目标行人的检测效果。最后提出了一种改进的数据增强策略,解决 Swin-Transformer 需要较大数据集的训练难题,减少了训练难度,提升了整体模型的 AP,并降低模型在遮挡场景行人检测下的误检率。未来的本研究将会考虑优化 Swin-Transformer 系列算法的骨干网络结构,并针对特征金字塔进行进一步的改进,并在提高算法 AP 的基础上适当减少模型大小和训练时长,更好地适应实际工程中的应用。

参考文献

- [1] 肖雨晴. 目标检测算法在交通场景中的应用综述[J]. 计算机工程与应用 2021, 57(6): 30-41.
XIAO Y Q. Research on application of object detection algorithm in traffic scene [J]. Computer Engineering and Applications, 2021, 57(6): 30-41.
- [2] 曹红燕, 沈小林, 刘长明, 等. 改进的 YOLOv3 的红外目标检测算法[J]. 电子测量与仪器学报, 2020, 34(8): 188-194.
CAO H Y, SHEN X L, LIU CH M, et al. Improved infrared target detection algorithm of YOLOv3 [J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(8): 188-194.
- [3] SHAO S, ZHAO Z, LI B, et al. CrowdHuman: A benchmark for detecting human in a crowd [J]. ArXiv Preprint, ArXiv:1805.00123, 2018.
- [4] 姬晓飞. D-S 理论多分类器融合的光学遥感图像多目标识别[J]. 电子测量与仪器学报, 2020, 34(5): 127-132.
JI X F. D-S theory based multi-classifier fusion optical remote sensing image target recognition [J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(5): 127-132.
- [5] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models [J]. IEEE Transactions on Software Engineering, 2010, 32(9): 1627-1645.
- [6] GIRSHICK R. Fast R-CNN [C]. 2015 IEEE International Conference on Computer Vision (ICCV), IEEE, 2015: 1440-1448.
- [7] 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, IEEE, 2014: 580-587.
- [8] 张德磊, 宋晓宁, 於东军. 基于统一划分的特征自适应行人再识别方法[J]. 南京理工大学学报, 2020, 44(3): 266-271.
ZHANG D L, SONG X N, YU D J. Feature adaptive person re-identification method based on unified partition [J]. Journal of Nanjing University of Science and Technology, 2020, 44(3): 266-271.
- [9] ZHANG S, WEN L, BIAN X, et al. Occlusion-aware R-CNN: Detecting pedestrians in a crowd [J]. ArXiv Preprint, ArXiv: 1807.08407, 2018.
- [10] 李奇. 基于深度学习的一阶目标检测算法应用研究[D]. 长春: 吉林大学, 2019.
LI Q. Application of first-order target detection algorithm based on deep learning [D]. Changchun: Jilin University, 2019.
- [11] HAN X. Modified cascade RCNN based on contextual information for vehicle detection [J]. Sens Imaging, 2021, 22(1): 19.
- [12] WANG Z, LU J, LIN R, et al. Correlated and Individual multi-modal deep learning for RGB-D object recognition [J]. ArXiv Preprint, ArXiv:1604.01655, 2016.
- [13] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted

- windows[J]. ArXiv Preprint, ArXiv:2103.14030,2021.
- [14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. ArXiv Preprint, ArXiv:1706.03762,2017.
- [15] DONG X, BAO J, CHEN D, et al. CSWin transformer: A general vision transformer backbone with cross-shaped windows[J]. ArXiv Preprint, ArXiv:2107.00652,2021.
- [16] YUAN L, HOU Q, JIANG Z, et al. VOLO: Vision outlooker for visual recognition [J]. ArXiv Preprint, ArXiv:2106.13112,2021.
- [17] WU B, YANG Y, YANG D, et al. Progressive hard-case mining across pyramid levels in object detection[J]. ArXiv Preprint, ArXiv:2109.07217,2021.
- [18] LI B, YAO Y, TAN J, et al. Equalized focal loss for dense long-tailed object detection[J]. ArXiv Preprint, ArXiv:2201.02593,2022.
- [19] YANG J, LI C, ZHANG P, et al. Focal self-attention for local-global interactions in vision transformers[J]. ArXiv Preprint, ArXiv:2107.00641,2021.
- [20] BENJUMEA A, TEETI I, CUZZOLIN F, et al. YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles[J]. ArXiv Preprint, ArXiv:2112.11798,2021.
- [21] 叶正喆, 苍岩. 基于卷积神经网络的行人检测方法[J]. 应用科技, 2022,49(2):55-62.
- YE ZH ZH, CANG Y. A pedestrian detection method based on convolutional neural network [J]. Applied Science and Technology, 2022,49(2):55-62.
- [22] LIU Z, HU H, LIN Y, et al. Swin transformer V2: Scaling up capacity and resolution[J]. ArXiv Preprint, ArXiv:2111.09883,2021.

作者简介



刘毅,1999 年于中国农业大学获硕士学位,2009 年于东北大学获博士学位,现为天津城建大学教授,硕士生导师,主要研究方向为计算机控制理论与仿真等。

E-mail:liuyi@tcu.edu.cn

Liu Yi, received his M. Sc. degree from China Agricultural University in 1999, and Ph. D. degree from Northeastern University in 2009. degree. Now he is a professor

and master's supervisor at Tianjin Chengjian University. His research interests include computer control theory and simulation, etc



于畅洋,2020 年于天津科技大学获得学士学位,现为天津城建大学计算机与信息工程学院硕士研究生,主要研究方向为计算机视觉、目标检测等。

E-mail:969882566@qq.com

Yu Changyang received his B. Sc. degree from Tianjin University of Science and Technology in 2020. Now he is a M. Sc. candidate at the School of Computer and Information Engineering, Tianjin Chengjian University. His main research interests include computer vision, object detection, etc.



李国燕(通信作者),2006 年于河北师范大学获得学士学位,2009 年于河北工业大学获得硕士学位,2013 年于河北工业大学获得博士学位,现为天津城建大学副教授,主要研究方向为下一代网络技术、智能信息处理等。

E-mail: ligy@tcu.edu.cn

Li Guoyan (Corresponding author) received her B. Sc. degree from Hebei Normal University in 2006, M. Sc. degree from Hebei University of Technology in 2009 and Ph. D. degree from Hebei University of Technology in 2013, respectively. Now she is an associate professor in Tianjin Chengjian University. Her main research interests include next generation network technology and intelligent information processing, etc.



潘玉恒,2002 年于曲阜师范大学获得学士学位,2005 年于哈尔滨工程大学获得硕士学位 2017 年于天津大学获得博士学位 现为天津城建大学副教授,主要研究方向为光纤传感及其信号处理等。

E-mail:panyuheng@tju.edu.cn

Pan Yuheng received her B. Sc. degree from Qufu Normal University in 2002, M. Sc. degree from Harbin Engineering University in 2005 and Ph. D. degree from Tianjin University in 2017. Now she is an associate professor at Tianjin Chengjian University. Her main research interests include optical fiber sensing, signal processing, etc.