第 41 卷 第 15 期/2021 年 8 月/光学学报 研究论文

光学学报

基于注意力机制的遮挡行人检测算法

邹梓吟[1,2]，盖绍彦[1,2*]，达飞鹏[1,2,3]，李昱[1,2]

[1]东南大学自动化学院，江苏 南京 210096；

[2]东南大学复杂工程系统测量与控制教育部重点实验室，江苏 南京 210096；

[3]东南大学深圳研究院，广东 深圳 518063

摘要 针对真实场景中因行人相互遮挡难以被精确检测的情况，提出一种基于注意力机制的特征提取增强检测算法。首先，通过添加注意力模块学习特征通道间关系和特征图空间信息，增强对行人目标可视区域的特征提取。其次根据行人数据的实际尺寸，采用 k-means++算法对行人标注进行聚类，确定锚框(anchor)大小及比例。利用距离交并比损失函数(DIOULoss)设计检测器的损失函数，使得检测框的回归更关注候选框与真实框的交并比与两框的中心距离。最后使用新设计的非极大值抑制算法(DSoft-NMS)保留更精确的预测框。所提方法在 CityPersons 和 WiderPerson 数据集上进行了实验，结果表明该方法在遮挡行人检测方面具有更高的检测精度，同时网络结构简单，方便后续研究。

关键词 机器视觉；遮挡行人检测；注意力机制；k-means 聚类；交并比

中图分类号 TP391.4 **文献标志码** A **doi**：10.3788/AOS202141.1515001

Occluded Pedestrian Detection Algorithm Based on Attention Mechanism

Zou Ziyin[1,2], Gai Shaoyan[1,2*], Da Feipeng[1,2,3], Li Yu[1,2]

[1] *School of Automation, Southeast University, Nanjing, Jiangsu* 210096, *China*;

[2] *Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Southeast University, Nanjing, Jiangsu* 210096, *China*;

[3] *Shenzhen Research Institute, Southeast University, Shenzhen, Guangdong* 518063, *China*

Abstract In light of the situation that it is difficult to accurately detect pedestrians in real scenes due to mutual occlusion, a feature extraction enhanced detection algorithm based on attention mechanism is proposed. Firstly, attention modules are added to learn the relationship between feature channels and the spatial information of feature maps, so as to enhance feature extraction in the visual area of pedestrian targets. Secondly, according to the actual size of pedestrian data, the k-means++ algorithm is used to cluster pedestrian labels, so as to determine the size and proportion of anchors. Distance-intersection over union loss function (DIOULoss) is used to design the loss function of the detector, so that the regression of the detection box pays more attention to the intersection over union between the candidate box and the real box, as well as the center distance between the two boxes. Finally, a new non-maximum suppression algorithm (DSoft-NMS) is presented to preserve more accurate prediction boxes. The proposed method has been tested on CityPersons and WiderPerson datasets, and the results show that the proposed method with a simple network structure has higher detection accuracy in occluded pedestrian detection, which is convenient for subsequent research.

Key words machine vision; occluded pedestrian detection; attention mechanism; k-means clustering; intersection

收稿日期：2021-01-13；**修回日期**：2021-02-06；**录用日期**：2021-03-08

基金项目：国家自然科学基金(51475092)、江苏省自然科学基金(BK20181269)、江苏省前沿引领技术基础研究专项(BK20192004C)、深圳市科技创新委员会(JCYJ20180306174455080)

通信作者：* qxxymm@163.com

over union
OCIS codes 150.0155; 150.1135; 100.4996

1 引 言

随着计算机视觉技术的不断发展，目标检测算法在自动驾驶领域取得了重大进展[1]。行人检测是目标检测的重要组成部分，在自动驾驶等领域有着广阔的应用前景。虽然行人检测已经在室内等背景固定的场景中取得了不错的效果，但是在室外等复杂场景下依然面临姿态各异、运动模糊及不同程度遮挡等问题，亟需探究和解决。

传统的行人检测器通常采用手工设计特征及机器学习的方法，Wu 等[2]提出“小边”特征和 Sabzmeydani 等[3]提出 Shapelet 特征，且都取得很好的效果。然而实际复杂场景下，手工设计的特征往往不能取得很好的效果。随着深度学习技术的蓬勃发展，现阶段的行人检测器通常采用基于卷积神经网络（CNN）的结构，如 Faster R-CNN[4-6]、YOLO[7-9]、SSD[10]等主流方法，自动学习目标深层次的语义信息[11]，挖掘目标在数据中隐含的统计规律和本质特征[12]，以获得更好的检测结果。虽然行人检测的精度不断提升，但面对现实中行人尺度变化大、类行人轮廓物体的干扰及严重的遮挡状况，检测器依然存在误检和漏检的情况。

为解决遮挡行人检测问题，Zhang 等[13]提出 CityPersons 数据集、Shao 等[14]提出 CrowdHuman 数据集和 Zhang 等[15]提出 WiderPerson 数据集为后续的研究提供了数据支持。Zhang 等[16]提出了一种新的聚合损失函数，通过预测人体各部分提高整体预测结果的准确性。Wang 等[17]提出的 Reploss 在拥挤场景中达到更高的定位精度。Bodla 等[18]通过降低预测框分数增加预测框个数，以更好地进行互遮挡目标的检测。Liu 等[19]对密度进行了预测，根据不同区域的密集程度设置非极大值抑制（NMS）阈值，保留低置信度候选框。Fei 等[20]通过对背景信息的利用提升遮挡行人的检测精度，Lin 等[21]通过对行人头、肩的定位增强定位效果。Wu 等[22]利用相邻帧图像寻找无遮挡或少遮挡目标对当前图像中的遮挡行人进行辅助识别，Hou 等[23]综合考虑多个相机以减小遮挡对检测系统的影响。虽然这些方法在一定程度上提升了行人检测的精度，但是使用预设 anchor 会增加候选框的回归时间，并且提取的行人特征包含大量的遮挡背景信息，导致网络难以关注重要区域，同时引入过多输入会增加数据标注难度。

针对以上问题，本文提出了一种基于注意力机制的遮挡行人检测算法。为了提升网络对遮挡行人可视区域的特征学习，对数据集中的行人区域进行随机遮挡，以增强模型鲁棒性。通过在骨架网络中增加改进的注意力模块，提升网络对行人可视区域的特征提取。通过设计严格回归策略的区域候选网络（RPN）自适应地生成 anchor 参数，降低了模型的训练难度，并采用基于距离交并比的回归（D-regress）得到更精确的位置信息。最后设计基于距离交并比的非极大值抑制（DSoft-NMS）算法，保留更准确的候选框，避免误删遮挡导致的有效预测框。实验结果表明所提方法在公开数据集上取得了更好的实验结果。

2 基本原理设计

图 1 为本文所构建的行人检测模型的主体架构，由特征提取网络和行人检测网络两部分组成，其中 i-th 代表第 i 个模块。特征提取网络采用特征金字塔网络（FPN）[24]，考虑 ResNet50 对特征提取的效果好，将其用作基础网络。ResNet50 分为 5 层，每层相对于输入图像的下采样率为 2^l，其中 l 表示层数，$l\in\{1,2,3,4,5\}$。如图 1 所示，在主干网络的第二层和第四层分别添加不同内容的第一层注意力模块和第二层注意力模块，进行特征提取。FPN 结合浅层空间信息和高层语义信息，生成多尺度的特征图，将其传入后续的行人检测网络。行人检测网络在获取高层语义信息和浅层纹理信息的基础上用 RPN 和 D-regress 对行人位置进行预测，经后处理（DSoft-NMS）算法保留最终的预测结果，得到预测边界框。

2.1 基于注意力机制的特征提取算法

为了使网络更加关注行人可视区域的特征，本文受 CBAM（Convolutional Block Attention Module）[25]和 SENet（Squeeze-and-Excitation Net）[26]的启发，采用通道注意力机制增大可视区域特征通道的权重，并通过添加空间注意力模块来关注可视区域的空间位置，提升模型的识别能力和鲁棒性。图 2(a)为 SENet 模块，本文提出的注意力模块如图2(b)所示，包含通道注意模块和空间注意模

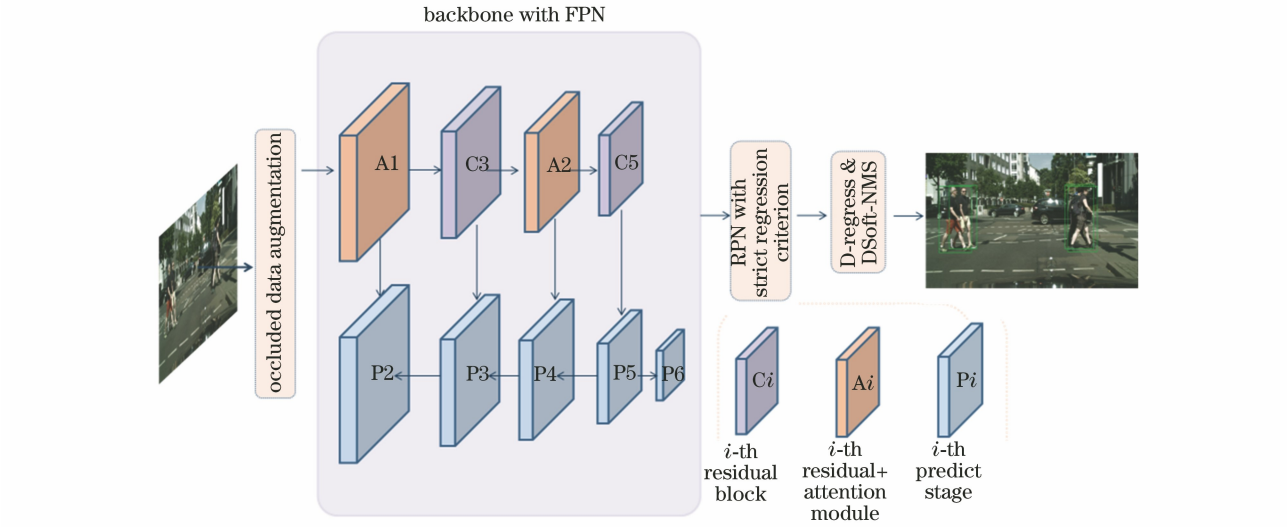

图 1　模型总体架构

Fig. 1　Overall architecture of model

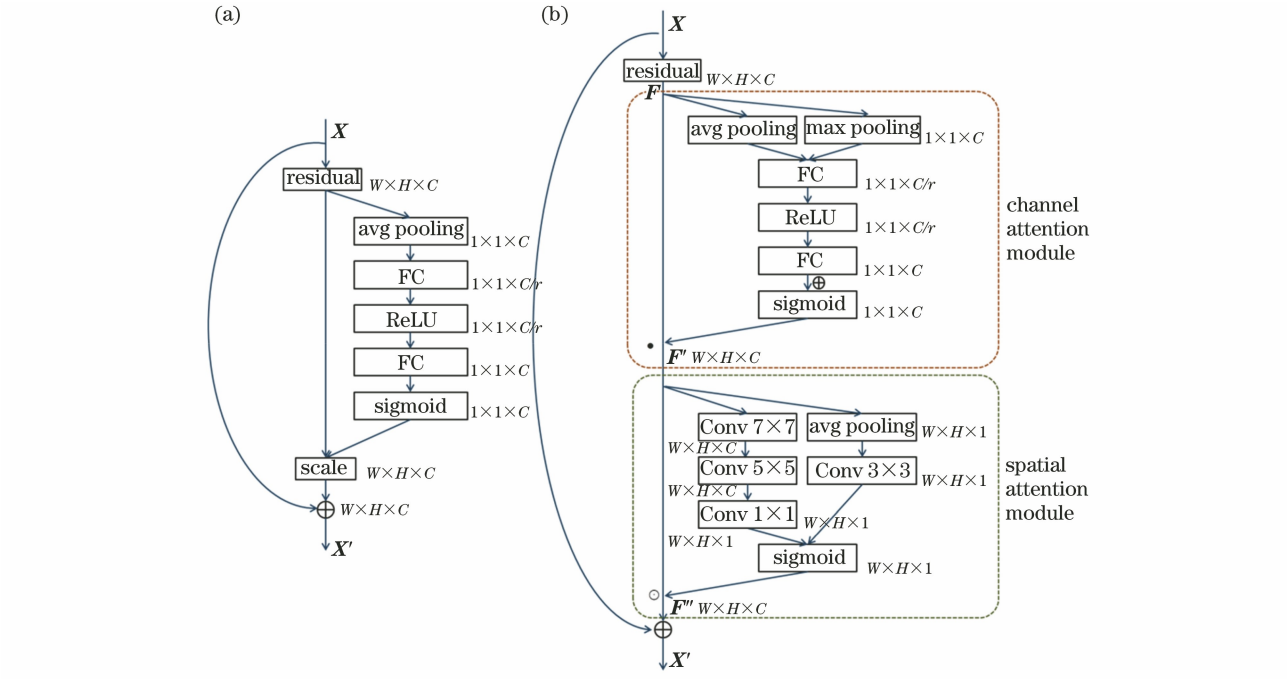

图 2　注意力模块。(a) SENet 模块;(b)所提模块

Fig. 2　Attention module. (a) SENet module; (b) our module

块两部分。图 2 中 X 为 SENet 模块的输入特征图，X'为 SENet 模块的输出特征图，X''为本文模块的输出特征图。输入注意力模块的特征图为 $F\in\mathbf{R}^{H\times W\times C}$，经过通道注意模块和空间注意模块得到输出特征图 F''，$F''\in\mathbf{R}^{H\times W\times C}$，其中，$H$ 为特征图高度，W 为特征图宽度，C 为特征通道数。通过对未遮挡区域的特征学习增强并抑制背景干扰，提升检测器的检测精度。

对于输入的特征图 F 分别采用平均池化(Avg pooling)和最大池化(Max pooling)聚合潜在信息并引导通道注意力，新增的最大池化有助于筛选辨识度最高的特征，提供更好的非线性。经过两层全连接层(FC)学习通道之间的关联性，并对各通道特征权重进行重分配，这有助于学习可视区域对应的通道信息[27]，更好地对遮挡行人特征进行学习。

本文在已有的通道注意模块上进行了修改，增加了对特征图的空间信息学习，以突出特征图中目标相关的区域，这有助于对遮挡行人的有效信息进行空间表述。在第一层注意力模块中，考虑到特征图 F'更大的尺寸，为更有效地利用上下文信息，构

建更有效的空间信息图，使用空洞卷积增大感受野。针对输入的特征图 F' 使用感受野大小为 7×7 和 5×5 的空洞卷积增大感受野，聚合上下文信息，最后利用 1×1 的卷积层进行降维，并通过 sigmoid 增加非线性特征，得到特征图 $M_s \in \mathbf{R}^{H\times W\times 1}$，其计算式为

$$M_s=\sigma\{f_{\text{conv}}^{1\times1}\{f_{\text{conv}}^{5\times5}[f_{\text{conv}}^{7\times7}(F')]\}\}, \tag{1}$$

式中：σ 为 sigmoid 函数；$f_{\text{conv}}^{1\times1}$、$f_{\text{conv}}^{5\times5}$、$f_{\text{conv}}^{7\times7}$ 分别表示感受野大小为 1×1、5×5、7×7 的卷积层。

在第二层注意力模块中，为获取更多的语义信息，使用平均池化统计特征图的空间信息，并使用 3×3 的卷积层连接得到空间注意特征图 $M_s \in \mathbf{R}^{H\times W\times 1}$，其计算过程为

$$M_s=\sigma\{f_{\text{conv}}^{3\times3}[\text{AvgPool}(F')]\}, \tag{2}$$

式中：AvgPool 为平均池化；$f_{\text{conv}}^{3\times3}$ 为 3×3 卷积层。使用空间注意特征图 M_s 激活输入特征图 F'，得到最终的特征图 F''：

$$F''=M_s\odot F', \tag{3}$$

式中：⊙为特征图逐元素相乘。

2.2 严格回归策略的 RPN

为提升模型在密集场景下对行人位置的回归，增强对遮挡行人的检测，本文提出了严格回归策略的 RPN 架构。首先根据行人数据集生成尺度匹配的 anchor，提升模型对遮挡行人的表达。其次使用距离交并比损失替代传统的范数损失作为回归函数，增强预选框与真实框的匹配程度。

2.2.1 k-means++自适应生成 anchor

anchor 机制是 RPN 的核心，用于生成不同尺度的预选框，以供后续的类别判断及定位。预设的 anchors 大小及比例依照 VOC 数据集聚类得出，和行人尺度存在明显差异。本文使用 k-means++聚类算法自适应地生成 anchor 参数，k-means++基于 k-means，可选择更远的聚类中心，以提升聚类效果。

行人检测主要关注预选框与真实框之间的交并比(IOU)，越高的 IOU 越能反映预选框与真实框的相近程度。k-means 默认的欧氏距离对标注框尺寸较敏感，本文重新设计了一种基于 IOU 的距离度量方式，该方式更加关注框与框之间的重叠状况，在框与框交并比较高时，生成更远的距离，这使得不同聚类组之间的差异性增大，生成更匹配的 anchor 尺度。其距离表达式为

$$\text{IOU}(a,b)=\frac{|a\cap b|}{|a\cup b|}, \tag{4}$$

$$d_i=\beta\sqrt{1-\text{IOU}(b_{\text{bbx}},c_{\text{cluster},i})}, \tag{5}$$

式中：d_i 为边界框和第 i 个聚类中心的距离；$|a\cap b|$ 为 a 框和 b 框相交的区域；$|a\cup b|$ 为 a 框和 b 框总的区域；b_{bbx} 和 $c_{\text{cluster},i}$ 分别为边界框和第 i 个聚类中心；β 为可变系数，实验中取 1.4。

由(5)式可知，本文使用的距离度量公式增加了 IOU 的影响，改善了欧氏距离对尺寸敏感的现象，可以生成更准确的聚类结果。

2.2.2 DIOU 损失函数

预选框的回归发生在两处：1)RPN 处的正样本回归；2)最终的候选框分类及位置回归。基准模型采用的 smooth L_1 损失会因为预选框 4 个坐标分量间存在相关性无法很好地学习和调参，并会对尺度不一致的目标产生不同的响应。本文使用 D-regress 解决这个问题。

D-regress 采用 DIOULoss(可用 L 表示)[28]作为回归损失函数，该函数满足度量的所有特性，且不受目标尺度大小影响，在 IOU 损失的基础上更好地反映预测框和真实框之间的对齐程度，并通过引入了预测框和真实框中心距离的差距，使得预测框能更好地和真实框匹配，更好地匹配常规认知下的行人目标定位。采用的 DIOU 损失函数为

$$L=1-\text{IOU}(B,B^{\text{gt}})+\frac{d^2(B,B^{\text{gt}})}{c^2(B,B^{\text{gt}})}, \tag{6}$$

式中：$d(B,B^{\text{gt}})$ 为预测框 B 的中心点和真实框 B^{gt} 的中心点间的欧氏距离；$c(B,B^{\text{gt}})$ 为两框最小覆盖矩形的对角线长度。

DIOULoss 通过添加预测框和真实框中心距离的度量，提高了预测框与真实框之间的对齐程度，最终提升了预测框位置的准确程度。

2.3 基于距离交并比的非极大值抑制算法

NMS 算法是目标检测常用的后处理算法，通过 NMS 在预测的所有预选框中筛选出置信度最高的结果，完成检测。传统的 NMS 思想是保留置信度最高的预选框，并将 IOU 大于阈值的其他预选框删除。

本文针对传统 NMS 算法在遮挡情景下会产生大量的错误抑制的情况进行了思考，提出了 DSoft-NMS 的处理策略。仅将预选框与真实框的 IOU 作为判断依据会导致密集场景下预测置信度较低的框被抑制。图 3 为两个相互遮挡严重的行人目标预测结果。由于框与框之间较高的 IOU，置信度较低的框将会被抑制，因此无法正确预测出图中所有行人的位置，出现严重的漏检情况。图 3(a)中两框的 IOU 为 0.59，图 3(b)中两框的 IOU 为 0.95。根据

传统 NMS 算法的结果，图 3(a)和图 3(b)中仅有一个预选框会被保留。本文在传统的 NMS 中引入对预选框间中心距离的判断(DIOU)。在新的条件下，图 3(a)的判定分数为 0.47，在相同的阈值下，两个预选框都得以保留，这提高了模型对行人目标的预测能力。图 3(b)的分数为 0.88，依照传统的 NMS 算法，置信度较低的预选框将被删除。本文提出用惩罚函数进行优化的方法，DIOU 越高的预测框，惩罚越重；DIOU 越低的框，惩罚越轻。这样可保留更多预测框，用于后续筛选。其抑制过程为

$$S_i=\begin{cases}S_i, & \mathrm{IOU}(M,b_i)-\dfrac{d^2(b_i,M)}{c^2(b_i,M)}<N_t\\ S_i\left[1-\mathrm{IOU}(M,b_i)+\dfrac{d^2(b_i,M)}{c^2(b_i,M)}\right], & \mathrm{IOU}(M,b_i)-\dfrac{d^2(b_i,M)}{c^2(b_i,M)}\geqslant N_t\end{cases},\tag{7}$$

式中：$d(b_i,M)$为预测框 b_i 和置信度最高的框 M 中心点的欧氏距离；$c(b_i,M)$为两框最小覆盖矩形的对角线长度；N_t 为比较阈值。

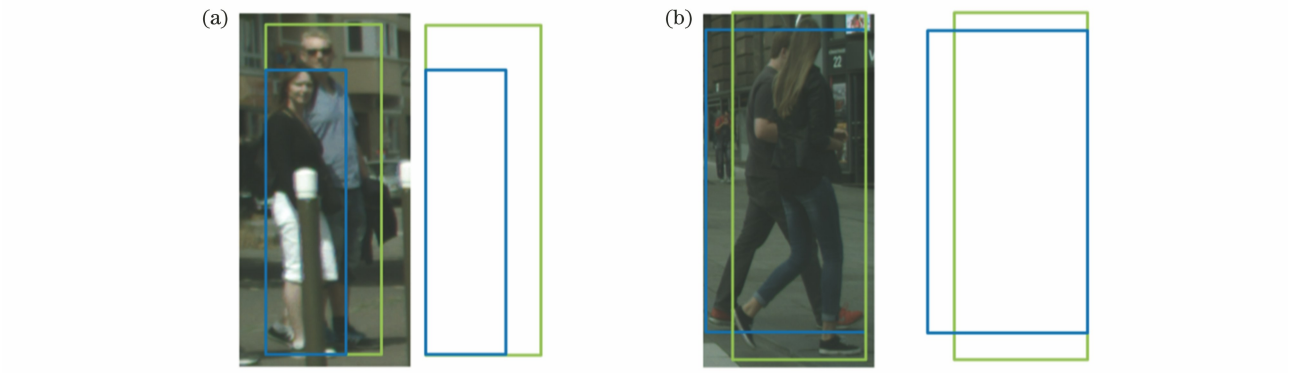

图 3　目标预测示意图。(a)预测结果 1；(b)预测结果 2

Fig. 3　Diagrams of target prediction. (a) Prediction result 1; (b) prediction result 2

在实际场景中，当近处行人将较远处行人遮挡时，较大的 IOU 导致远处行人的预测框被抑制，无法进行正确的预测。优化的惩罚函数增加了对预测框中心距离的考虑，保留了遮挡行人的预选框，并降低遮挡严重的预测框的置信度，进行重筛选，这在一定程度上避免预选框的误删，提高了召回率，增强了模型的预测能力。

3　实验结果与分析

为了验证本文所提模型的有效性，在 CityPersons 和 WiderPerson 数据集上进行对比实验，检验模型的检测效果。

实验条件：操作系统为 Windows Server 2019，深度学习框架为 Pytorch 1.4.0，CPU 为 Intel Core i7 8700 K，内存为 16 GB，GPU 为 NVIDIA GeForce GTX 1070。

数据集选择：CityPersons 和 WiderPerson 为所选用的两个行人检测数据集。其中，CityPersons 训练集包含 2975 张城市道路场景图片，测试集由 500 张图片组成，图片分辨率为 2048 pixel×1024 pixel。由于 CityPersons 被广泛实验及应用，选其作为所提方法的测试数据集，遵循 Reploss 及后续研究对 CityPersons 数据集的实验，依照遮挡情况对数据集进行了划分，划分情况如表 1 所示。WiderPerson 数据集是一个户外行人检测数据集，包含行人、骑自行车的人、部分可见人、人群和忽略区域 5 种标注，将前 4 种作为行人标注，训练图片为 8000 张，测试图片为 1000 张。为验证所提方法在遮挡严重的场景下依然有效，相关测试在 WiderPerson 数据集上进行。两个数据集存在行人两两相互遮挡的情况如表 2 所示，表 2 中 # person/img 代表每张图含有的行人总数量。可以看出 WiderPerson 数据集提供了更密集的行人场景，有助于验证检测器对两两相互遮挡的样本的检测结果。

表 1　CityPersons 数据集遮挡状况划分

Table 1　Occlusion status partition of CityPersons dataset

Type	Bare	Partial	Reasonable	Heavy
Occlusion	[0,0.10]	(0.10,0.35)	[0,0.35)	[0.35,∞)

遮挡数据增强：针对 CityPersons 数据集中遮挡行人比例较小的情况，提出对训练数据进行随机遮挡的方式，构建一个对遮挡更鲁棒的模型。针对 CityPersons 中给出的行人可见部分和全身部分的标注，随机选取可见部分在 80%以上的行人进行遮挡处理，除了头部区域以外，选择 15%的身体部分进行遮挡处理。为了保证数据选取的随机性，每个符合条件的标注框有 50%的可能被遮挡。在图片中随机添加遮挡可以增强模型的鲁棒性，有助于提升对遮挡行人判断的准确性，同时在预测阶段不会带来额外的计算成本。

表 2 数据集互相遮挡的行人数量

Table 2 Number of pedestrians of mutual occluding datasets

Dataset	CityPersons	WiderPerson
♯person/img	6.47	28.87
IOU is larger than 0.3	0.96	9.21
IOU is larger than 0.5	0.32	2.15
IOU is larger than 0.7	0.08	0.24

评价指标：在行人数相对较少的 CityPersons 数据集上采用对假阳性敏感的虚警率在[0.01，1.00]之间的对数平均漏检率(M_{MR}^{-2})作为评价指标。M_{MR}^{-2} 值越低，说明检测器的性能越好。在行人密集的数据集 WiderPerson 上采用 M_{MR}^{-2} 的方法会得到较多的假阳性结果，这些结果不能很好地反映检测器的检测性能，采用 IOU 阈值为 0.5 的平均精确度(AP)以及召回率(Recall)作为评价指标，AP 和召回率越高，检测效果越好。

训练细节：选择预训练的基于 ResNet50 的 FPN 作为主干网络，采用随机梯度下降(SGD)算法优化训练模型，初始学习率设置为 1.25×10^{-4}，衰减系数设置为 0.001。训练批次大小为 2，总共训练 320 轮。采用 DSoft-NMS 滤除冗余结果，阈值设置为 0.5。

3.1 CityPersons 数据集上的对比实验

为验证各模块的有效性，以原有的双阶段 Faster RCNN＋ResNet50＋FPN 作为基准方法(baseline)，利用多尺度输入图像对所有模块的有效性进行测试。其中 AM 表示注意力模块，RRPN 为严格回归策略的 RPN 模块，DN 为使用 D-regress 的 Fast RCNN 和 DSoft-NMS 算法，对比添加各模块后的检测器性能，测试结果如表 3 所示。

表 3 各添加模块的实验结果 M_{MR}^{-2}

Table 3 Results for each added module M_{MR}^{-2} %

Method	Reasonable	Heavy	Partial	Bare
Baseline	15.7	56.3	18.8	9.7
Baseline＋AM	13.6	52.8	15.3	7.5
Baseline＋AM＋RRPN	13.1	52.1	14.5	7.3
Baseline＋AM＋RRPN＋DN	12.7	51.9	13.2	6.6

与 baseline 相比，所添加的各个模块对于遮挡目标的 M_{MR}^{-2} 都有明显的下降，说明本文所提方法适用于遮挡行人检测任务，特别是在 Heavy 子集的评价标准下具有明显的优势，其 M_{MR}^{-2} 下降了 4.4%，这表明了针对严重遮挡场景下的行人检测的有效性。

为了验证本文方法的性能，将其与其他行人检测方法进行对比，评价指标为 M_{MR}^{-2}。对比方法选取 Faster-RCNN[6]，OR-CNN[16]，RepLoss[17]，TLL[29]，Adaptive-NMS[19]，PedJointNet[21] 等主流方法以及最新方法。表 4 给出了实验结果，本文方法在遮挡较小的 Partial 和 Bare 子集的评价标准下仅次于最优方法，在严重遮挡的 Heavy 子集的评价标准下取得了 M_{MR}^{-2} 为 51.9%的检测结果，优于对比方法，这证明了在严重遮挡情况下本文所提方法的优越性。

表 4 CityPersons 数据集测试结果 M_{MR}^{-2}

Table 4 Experimental results of M_{MR}^{-2} on CityPersons %

Method	Backbone	Reasonable	Heavy	Partial	Bare
Faster-RCNN	VGG-16	15.4	—	—	—
WiderPerson	VGG-16	11.1	—	—	—
OR-CNN	VGG-16	12.8	55.7	15.3	6.7
TLL	ResNet-50	15.5	53.6	17.2	10.0
TLL＋MRF	ResNet-50	14.4	52.0	15.9	9.2
PedJointNet	ResNet-50	13.5	52.1	—	—
RepLoss	ResNet-50	13.2	56.9	16.8	7.6
Adaptive-NMS	ResNet-50	10.8	54.0	11.4	6.2
Our method	ResNet-50	12.7	51.9	13.2	6.6

在检测速度方面，针对分辨率为 1024 pixel×2048 pixel 的输入图像，baseline 和本文方法针对每张图片的检测速度分别为 0.20 s 和 0.32 s。本文方法在检测速度小幅降低的情况下，检测效果提升明显，实现了速度和精确度较好的平衡，可以满足实时检测的任务需求。

3.2 WiderPerson 数据集上的实验

为了检测本文所提模型在行人密集数据集 WiderPerson 上的检测效果，分别使用 baseline 和所提网络对测试集中 1000 张图片进行测试，检测结果如图 4 所示。

图 4 中用虚线框标出了漏检(false negative)样本，并指出了误检(false postive)样本。通过对比图 4(c)、(g)可以发现，基准网络在检测行人时会将背景误当作目标，而本文方法可以有效提取目标特征，避免背景信息的干扰。通过对比图 4(a)、(e)和图 4(b)、(f)可以发现 baseline 对行人的检测存在漏检的问题，而本文方法可以检测出漏检目标。通过对比图 4(a)、(e)和图 4(d)、(h)可以发现，针对遮挡行人存在的大量误检情况，本文方法对假阳性的检测结果有明显的抑制效果，使得检测结果更加精确有效。

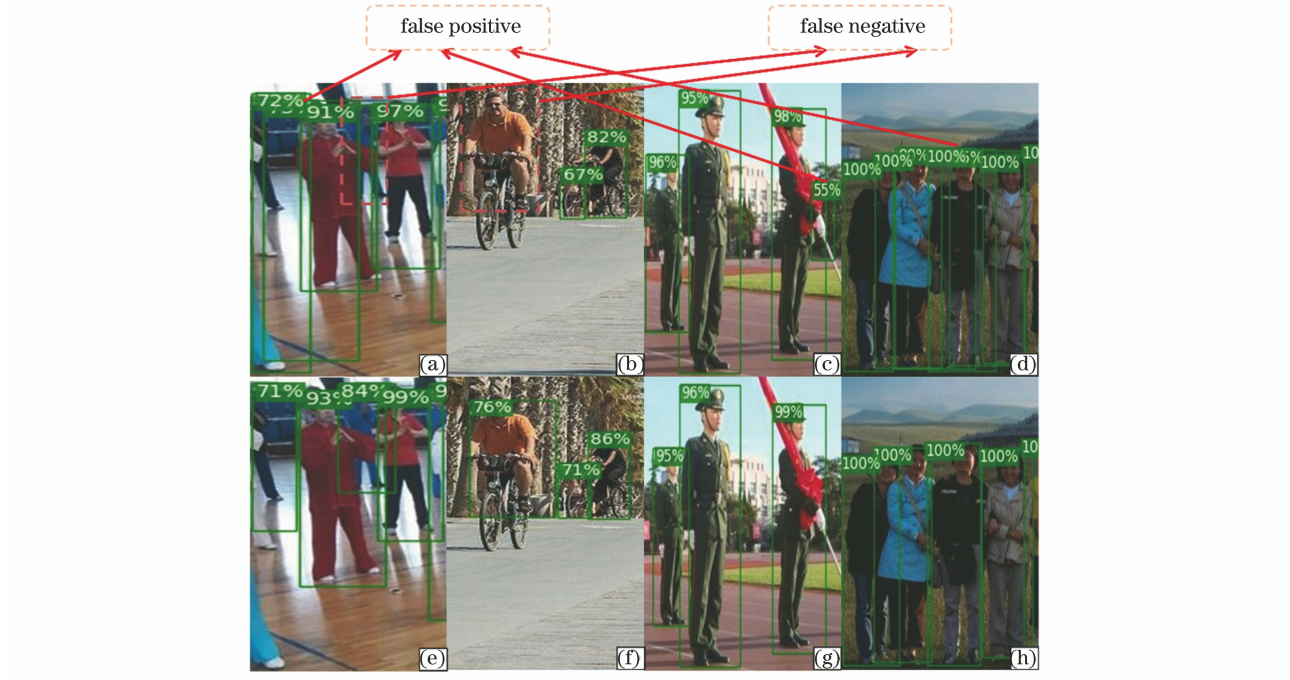

图 4　行人检测对比结果。(a)～(d) Baseline 检测效果；(e)～(h)本文所提模型的检测结果

Fig. 4　Comparison of detection results of pedestrians. (a)-(d) Detection results of baseline; (e)-(h) detection results of our method

为了精确评估模型在遮挡严重的数据集上的检测效果，还原真实场景下严重遮挡行人的检测结果，在 WiderPerson 数据集上进行了测试，所提方法的平均精确度达到 85.12%，召回率达到 89.94%，相比 baseline 分别提升了 0.69%和 0.73%，结果如表 5 所示。

表 5　WiderPerson 数据集上的实验结果

Table 5　Experimental results on WiderPerson

Method	AP /%	Recall /%
Baseline	84.43	89.21
Our method	85.12	89.94

4　结　　论

针对自然场景下普遍存在的行人遮挡检测问题，构建了基于注意力机制的行人检测网络，使得网络更好地应对遮挡行人的检测，更加关注可见区域的信息。利用聚类算法自适应地设置 anchor 大小及比例，使得模型更容易训练。采用的回归损失函数帮助网络更加精确地定位目标，设计的 NMS 算法减少了漏检样本量。在 CityPersons 数据集上的实验结果表明所提算法对遮挡行人的检测具有较好的效果，在 WiderPerson 数据集上的实验结果表明所提算法对遮挡严重的行人场景也有较好的检测效果，所提算法在一定程度上完成了遮挡行人检测的任务要求。所提算法的检测置信度依赖行人上半身区域，下一步将对上半身遮挡的行人检测进行研究。

参 考 文 献

[1] Sun Y C, Pan S G, Zhao T, et al. Traffic light detection based on optimized YOLOv3 algorithm[J]. Acta Optica Sinica, 2020, 40(12): 1215001.
孙迎春, 潘树国, 赵涛, 等. 基于优化 YOLOv3 算法的交通灯检测[J]. 光学学报, 2020, 40(12): 1215001.

[2] Wu B, Nevatia R. Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors[C]//Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, October 17-21, 2005, Beijing, China. New York: IEEE Press, 2005: 90-97.

[3] Sabzmeydani P, Mori G. Detecting pedestrians by learning shapelet features[C]//2007 IEEE Conference on Computer Vision and Pattern Recognition, June 17-22, 2007, Minneapolis, MN, USA. New York: IEEE Press, 2007: 1-8.

[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, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.

[5] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 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 (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.

[8] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.

[9] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2019-09-22]. https://arxiv.org/abs/1804.02767.

[10] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.

[11] Ju M R, Luo J N, Wang Z B, et al. Multi-scale target detection algorithm based on attention mechanism[J]. Acta Optica Sinica, 2020, 40(13): 1315002.
鞠默然, 罗江宁, 王仲博, 等. 融合注意力机制的多尺度目标检测算法[J]. 光学学报, 2020, 40(13): 1315002.

[12] Zhao B, Wang C P, Fu Q, et al. Multi-scale infrared pedestrian detection based on deep attention mechanism[J]. Acta Optica Sinica, 2020, 40(5): 0504001.
赵斌, 王春平, 付强, 等. 基于深度注意力机制的多尺度红外行人检测[J]. 光学学报, 2020, 40(5): 0504001.

[13] Zhang S S, Benenson R, Schiele B. CityPersons: a diverse dataset for pedestrian detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 4457-4465.

[14] Shao S, Zhao Z J, Li B X, et al. CrowdHuman: a benchmark for detecting human in a crowd[EB/OL]. (2018-04-30)[2018-05-30]. https://arxiv.org/abs/1805.00123.

[15] Zhang S F, Xie Y L, Wan J, et al. WiderPerson: a diverse dataset for dense pedestrian detection in the wild[J]. IEEE Transactions on Multimedia, 2020, 22(2): 380-393.

[16] Zhang S F, Wen L Y, Bian X, et al. Occlusion-aware R-CNN: detecting pedestrians in a crowd[M]// Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11207: 657-674.

[17] Wang X L, Xiao T T, Jiang Y N, et al. Repulsion loss: detecting pedestrians in a crowd[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: 7774-7783.

[18] Bodla N, Singh B, Chellappa R, et al. Soft-NMS: improving object detection with one line of code[C]// 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 5562-5570.

[19] Liu S T, Huang D, Wang Y H. Adaptive NMS: refining pedestrian detection in a crowd[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019,

Long Beach, CA, USA. New York: IEEE Press, 2019: 6452-6461.

[20] Fei C, Liu B, Chen Z, et al. Learning pixel-level and instance-level context-aware features for pedestrian detection in crowds[J]. IEEE Access, 2019, 7: 94944-94953.

[21] Lin C Y, Xie H X, Zheng H. PedJointNet: joint head-shoulder and full body deep network for pedestrian detection[J]. IEEE Access, 2019, 7: 47687-47697.

[22] Wu J L, Zhou C L, Yang M, et al. Temporal-context enhanced detection of heavily occluded pedestrians[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 13427-13436.

[23] Hou Y Z, Zheng L, Gould S. Multiview detection with feature perspective transformation[M]//Vedaldi A, Bischof H, Brox T, et al. Computer vision-ECCV 2020. Lecture notes in computer science. Cham: Springer, 2020, 12352: 1-18.

[24] 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 (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.

[25] 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.

[26] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[27] Zhang S S, Yang J, Schiele B. Occluded pedestrian detection through guided attention in CNNs[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: 6995-7003.

[28] Zheng Z H, Wang P, Liu W, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

[29] Song T, Sun L Y, Xie D, et al. Small-scale pedestrian detection based on topological line localization and temporal feature aggregation[M]// Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 554-569.