·开发研究与工程应用· 文章编号：1000-3428(2021)10-0276-07 文献标志码：A 中图分类号：TP391

基于CenterNet的实时行人检测模型

姜建勇，吴 云，龙慧云，黄自萌，蓝 林
（贵州大学 计算机科学与技术学院，贵阳 550025）

摘 要：针对传统目标检测模型不能同时兼顾检测速度和准确度的问题，提出一种新的PD-CenterNet模型。在CenterNet的基础上对网络结构和损失函数进行改进，在网络结构的上采路径中，设计基于注意力机制的特征融合模块，对低级特征和高级特性进行融合，在损失函数中通过设计α、γ、δ 3个影响因子来提高正样本与降低负样本的损失，以平衡正负样本的损失。实验结果表明，相比CenterNet模型，该模型在网络结构和损失函数上的准确度分别提高5.1%、9.81%。

关键词：PD-CenterNet网络；实时检测；行人检测；样本不平衡；损失函数；特征融合

开放科学（资源服务）标志码（OSID）：

中文引用格式：姜建勇，吴云，龙慧云，等. 基于CenterNet的实时行人检测模型[J]. 计算机工程，2021，47(10)：276-282.

英文引用格式：JIANG J Y，WU Y，LONG H Y，et al. CenterNet-Based real-time pedestrian detection model[J]. Computer Engineering，2021，47(10)：276-282.

CenterNet-Based Real-Time Pedestrian Detection Model

JIANG Jianyong，WU Yun，LONG Huiyun，HUANG Zimeng，LAN Lin
（College of Computer Science and Technology，Guizhou University，Guiyang 550025，China）

【Abstract】Generally，the speed gain of traditional target detection models comes at the cost of accuracy，and vice versa. To address the problem，a new pedestrian detection model，PD-CenterNet，is proposed based on CenterNet by improving its network structure and loss function. In terms of network structure，a feature fusion module based on attention mechanism is given in the up-sampling path to fuse low-level features and high-level features. In terms of the loss function，three factors α、γ and δ are designed to increase the loss of positive samples and reduce the loss of negative samples，balancing the loss of the samples. Experimental results show that compared with the CenterNet model，the proposed model improves the accuracy of network structure by 5.1% and the accuracy of the loss function by 9.81%.

【Key words】PD-CenterNet；real-time detection；pedestrian detection；sample imbalance；loss function；feature fusion
DOI：10. 19678/j. issn. 1000-3428. 0059043

0 概述

行人检测作为计算机视觉领域的研究热点[1]，在车辆高级驾驶辅助系统、视频监控、安全检查以及反恐防暴等方面有着重要应用。在过去的几十年中，研究人员针对行人检测问题做了大量研究并取得一系列成果。行人检测方法主要分为基于人工设计特征和基于神经网络特征的2种检测方法。

传统的检测器多数使用HOG方法进行检测，如文献[2]通过改进HOG并且联合使用SVM进行行人检测。在传统的检测方法中，需要人工手动去提取图像特征，使得检测模型存在可扩展性、泛化能力差以及计算速度慢等问题。

随着在机器视觉中使用深度学习，研究人员开始寻找深度学习方案来解决目标检测问题。R-CNN[3]提出结合深度学习的方法解决对象检测的问题，后续很多两阶段的方法都是基于R-CNN[3]去构建的，如Fast-RCNN[4]、Faster-RCNN[5]等。在行人检测上，如文献[6]使用更为快速的Faster-RCNN[5]进行行人检测，在Caltech数据集上比其他传统方法更为准确和快速，文献[7]使用单阶段式网络YOLO[8-10]进行行人检测，在INRIA数据集上取得比

基金项目：国家自然科学基金(61741124)；贵州省科技计划项目(5781)。
作者简介：姜建勇(1996—)，男，硕士研究生，主研方向为深度学习、目标检测；吴 云(通信作者)、龙慧云，副教授、博士；黄自萌、蓝 林，硕士研究生。
收稿日期：2020-07-20 **修回日期：**2020-09-16 **E-mail：**364912908@qq.com

传统方法更好的准确度。虽然Faster-RCNN在R-CNN的基础上改进了很多组件，使得在准确度和速度上有了很大的提升，速度能够达到20 frame/s，但运用在实时检测中效果还不是很理想。而单阶段网络YOLO[8-10]在速度上很快，但是准确度距离Faster-RCNN相差较大。

CenterNet[11]与R-CNN[3]相比，不需要区域建议网络以及ROI等重要组件，而与YOLO[8]和SSD[12]相比，则无需预先去设定Anchor的大小，并且CenterNet在推理阶段不需要NMS（Non-Maximum Suppression）[13]，因此在速度上有很大的提升。为了能够平衡检测的速度和准确度，本文基于CenterNet，提出PD-CenterNet改进模型对行人进行检测。该模型通过融合低级语义信息来减少细节性信息在下采样过程中丢失的问题，并设计一个新的损失函数来解决正负样本不平衡的问题。

1　PD-CenterNet模型设计

目标检测往往在图像上将目标以轴对称的框形式框出，成功的目标检测器都是先穷举出潜在目标位置，然后对该位置进行分类，但这种做法浪费时间、低效、需要额外的后处理。PD-CenterNet在构建模型时将目标作为一个点，即目标BBox（Bounding Box）的中心点，检测器能够通过特征图的相对位置来估计出BBox的中心点和尺寸。本文设计的PD-CenterNet主要由网络结构、Anchor和损失函数3个部分构成，下文将围绕这3个方面进行介绍。

1.1　网络结构设计

PD-CenterNet由主干网络、上采路径和网络顶端2个卷积组成，如图1（b）所示。由于在目标检测中，感受野的大小在很大程度上直接影响检测的效果，因此本文通过主干网络来获取一个1/32的特征图，编码高层的语义信息。其中向下箭头表示下采样过程，向上箭头表示上采样过程。

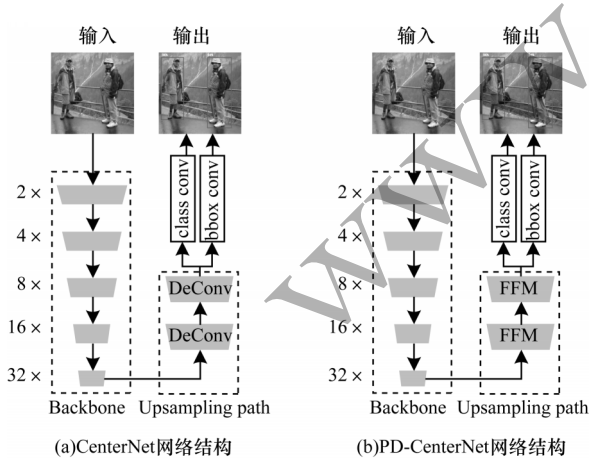


图1　CenterNet和PD-CenterNet网络结构

Fig.1　Network structure of CenterNet and PD-CenterNet

在网络结构上，使用主干网络和上采路径的设计方式能够为整个模型带来很好的可扩展性，通过主干网络可切换为MobileNet[14]、ResNet18[15]、Xception[16]、ShuffleNet[17]等轻量级网络即可获得更快的速度，也可切换至ResNet101[15]、GoogLeNet[18]、DenseNet[19]等较大的网络来获取更高的准确度。

在上采路径上，BiSeNet[20]指出在特征表示的层面上，低层和高层的特征表示不同，仅以通道来连接低层和高层特征，则就会带来很多噪音，所以本文设计了一个特征融合模块（FFM）来融合低层丰富的空间信息和高层的语义信息，如图2所示。在特征融合模块中，首先将高层特征进行上采至低层特征图一致的大小，然后按通道进行连接，后面紧接一个深度可分离卷积来学习通道上每一层的表示，最后使用类似于SENet[21]的通道特征注意力机制，把相连接的特征使用全局平均池化为一个特征向量，并学习出一个权重向量，然后对先前的特征进行加权，增强了低层和高层特征融合之后的表示能力，同时也减少了特征融合之后带来的噪音。

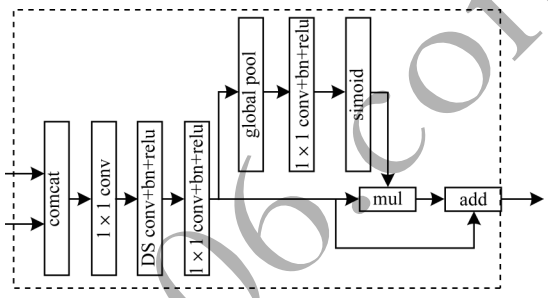


图2　特征融合模块

Fig.2　Feature fusion module

在网络顶端分别使用2个卷积用来预测类别置信度、BBox的位置和尺寸信息。

通过对网络结构的改进，使得模型的网络参数和计算量大幅减少。尤其将MobileNet[14]作为主干网络时使得模型的参数减少了50%，降低了计算量，在以ResNet[15]作为主干网络时参数量和计算量也得到了很大的降低，具体的信息如表1所示。

表1　模型参数及数量

Table 1　Model parameters and quantity

模型	参数/10^6	Multi-Adds/10^9
CenterNet + MobileNet	10.6	17.7
CenterNet + ResNet18	16.4	19.3
CenterNet + ResNet50	35.0	26.0
CenterNet + ResNet101	54.0	32.9
PD-CenterNet + MobileNet	5.3	2.1
PD-CenterNet + ResNet18	13.9	11.5
PD-CenterNet + ResNet50	27.6	19.9
PD-CenterNet + ResNet101	46.6	26.8

1.2　Anchor设计

在训练时获得网络输出的类别置信度和BBox后，需要用BBox去匹配GT BBox，在这个过程中，BBox即为Anchor BBox。标准的Anchor设计了在

低分辨率特征图中一系列固定的BBox，通过计算交并比（IoU）来判断是否为正样本，若交并比大于0.7则标记为正样本，若小于0.3则标记为负样本。而本文则将对象的中心点在低分辨率上所对应BBox作为正样本，其他没有包含对象中心点的标记为负样本，并且一个中心点仅检测一个对象，如图3所示。

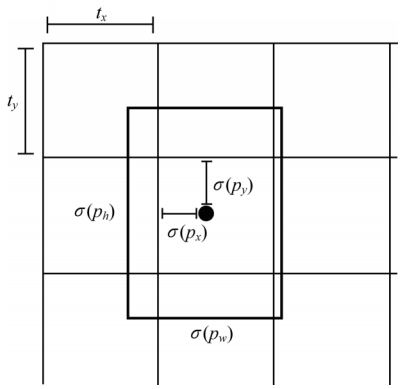


图3 PD-CenterNet Anchor的设计

Fig.3 Design of PD-CenterNet Anchor

基于图3所示的设计，网络只需要预测在某一个单元格内的偏移即可。对于网络输出的每一个BBox为(p_x,p_y,p_w,p_h)，其中：(p_x,p_y)为BBox的中心点位置；(p_w,p_h)为BBox的宽和高。如果该单元格对应于图像单元格左上角的坐标为(t_x,t_y)，则最后预测的BBox为(b_x,b_y,b_w,b_h)，计算公式如下：

$$b_x=\sigma(p_x)+t_x \tag{1}$$

$$b_y=\sigma(p_y)+t_y \tag{2}$$

$$b_w=\sigma(p_w) \tag{3}$$

$$b_h=\sigma(p_h) \tag{4}$$

1.3 损失函数设计

PD-CenterNet由BBox的中心点、BBox尺寸和BBox类别置信度3个部分损失构成，计算公式如下：

$$L_{\text{det}}=L_k+L_{\text{size}}+L_{\text{off}} \tag{5}$$

$$L_{\text{size}}=|\hat{b}_w-b_w|+|\hat{b}_h-b_h| \tag{6}$$

$$L_{\text{off}}=|\hat{b}_x-b_x|+|\hat{b}_y-b_y| \tag{7}$$

其中：L_{det}为总损失；L_k为BBox类别置信度损失；L_{size}为BBox尺寸损失；L_k为置信度损失；L_{off}为中心点损失。由于模型的输入是300×300，因此最终得到一个75×75的一个特征图，并且模型一个特征点仅预测一个对象，极端情况下会出现1∶5 625的正负样本极度不平衡，所以，本文设计一个损失函数来解决这个问题。

在类别置信度损失中，本文设计α、γ、δ 3个影响因子提高正样本的损失和减小负样本的损失以解决正负样本不平衡的问题，定义如式(8)~式(10)所示：

$$L_k=\alpha_1 L_{\text{neg}}+\alpha_2 L_{\text{pos}} \tag{8}$$

$$L_{\text{neg}}=-(1-\hat{y})^{\gamma}{}_1\times\log_a(\hat{y}+\delta) \tag{9}$$

$$L_{\text{pos}}=-(1-\hat{y})^{\gamma}{}_2\times\log_a(\hat{y}) \tag{10}$$

在负样本损失中通过设置δ和$\gamma_1$2个因子来减小负样本的损失，定义见式(9)，在正样本损失中通过γ_2进行调节，定义见式(10)，最后通过α因子来控制正样本和负样本损失所占的比例。通过对损失函数中的α_1、α_2、γ_1、γ_2、δ使用网格搜索得到最佳的一组参数，如表2所示。

表2 损失函数参数值

Table 2 Loss function parameter value

参数	参数值
α_1	0.25
α_2	1.00
γ_1	3.00
γ_2	1.50
δ	0.20

2 模型实现

2.1 网络结构

本文对残差网络（ResNet[15]）进行修改以适应PD-CenterNet。选取ResNet[15]网络中“layer2”“layer3”和“layer4”的输出分别作为“8×”“16×”和“32×”的特征图，然后通过特征融合模块来对这3个特征图进行融合，接着在融合后“8×”倍的特征图上通过反卷积上采到“4×”，最后通过网络顶端的2个卷积来进行类别置信度和BBox预测。

MobileNet专注于移动端或者嵌入式设备中的轻量级CNN网络，它使得推理速度能够得到极大的提高。因此，修改MobileNet[12]作为本文模型的主干网络，选取MobileNet[12]中第7层、第14层和最后一个卷积层的输出分别作为“8×”“16×”和“32×”的特征图，然后通过上采路径和2个卷积来进行类别置信度和BBox预测。

2.2 Anchor选择

根据Anchor的设计，使得一个特征点仅能预测一个对象，如果一张图像中有超过一个对象的中心点重叠，则导致模型存在漏检。而由于在行人检测中存在很多对象会存在中心点一致的问题，因此在Anchor选择时，如果BBox被占用，则选择离中心点最近的BBox来预测对象，这样就避免了中心点重复的问题，算法过程如下：

通过网络的输出得出$B(W\times H,4)$，其中B的宽和高在本文中为(75,75)，第三维分别为BBox的中心点位置(x,y)和大小(w,h)。对于GT(Ground Truth)中的每一个G^i，计算其在特征图大小为(W,H)的中心点位置(x,y)，然后计算出(x,y)在B上的索引c（见算法1第5行）。如果c不在正样本集合A中，则将c添加到集

合A中，否则计算出离c最近的一个点并添加到集合A中。最后通过构建一个D集合，其中D的区间为[0，$W\times H$]，此时正样本Anchor为$B[A]$，负样本Anchor为$B[D-A]$。Anchor选择算法如算法1所示。

算法1　Anchor选择算法

输入

B is a set of bounding boxes, shape is [$W\times H$, 4]

G is a set of ground-truth boxes on the image, shape is [N, 4]

S is the size of the input image

输出

P_{pos} is the BBox of the positive sample

P_{neg} is the BBox of the negative sample

```
initialize the index of positive samples as A
compute the downsampling multiple of the feature map: r = S_width/W
for each level i in [1, N] do
    compute the position of the BBox on the x-axis of the feature map: x = ⌊G_x^i/r⌋
    compute the position of the BBox on the y-axis of the feature map: y = ⌊G_y^i/r⌋
    c = y × W + x
    if c in A do
        find unused points around c: c = FindPoint(x, y, c)
    end if
    A = A U c
end for
D = range(0, W × H)
P_pos = B[A]
P_neg = B[D-A]
return P_pos, P_neg
```

2.3　模型训练

训练使用分辨率为300像素×300像素的图像作为输入，对于输入的图像使用随机翻转、随机缩放（0.5~1.5的比例）、随机裁剪和色彩抖动做数据增强，使用学习率为1e-4的Adam[22]作为优化器，使用批大小为4训练80轮，其中前5轮使用线性学习率进行预热，后75轮使用余弦退火算法来对学习率进行衰减，如图4所示。

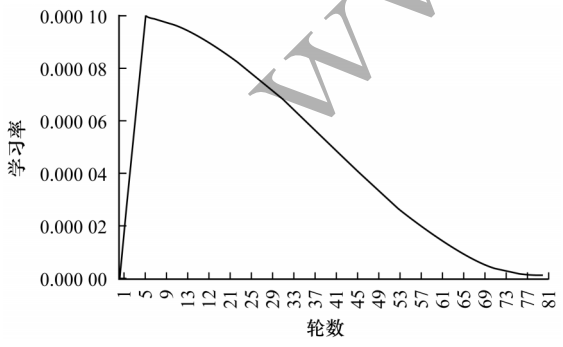


图4　学习率衰减

Fig.4　Decline of learning rate

2.4　推理阶段

在推理阶段不使用数据增强，将输入的图片进行等比例的缩放，其余的地方用0进行填充至300像素×300像素的输入。目前大多数的目标检测模型都依赖于NMS[13]对模型最后输入的BBox进行筛选，由于NMS的时间复杂度为$O(n^2)$，因此使得在后处理阶段变得更加耗时。结合本文模型的设计，本文设计一种时间复杂度为$O(n)$的后处理算法，算法过程如下：

在模型的推理和模型的训练中规定类别置信度第一个位置的值为背景的置信度，在推理时设置一个置信度阈值d用来过滤置信度低的BBox。当获得网络输出的预测框B和类别置信度C时，则需要遍历每一个B_i所对应的类别置信度C_i，如果C_i的最大值位置为0或者对象的置信度低于设置好的置信度阈值，那么就忽略该预测框，否则保存B_i和C_i作为最好的输出，最后将面积特别小的预测框移除，以及将预测框超出图像范围的区域裁剪掉，最终的输出即为对行人的检测以及置信度，具体的实现如算法2所示。

算法2　PD-CenterNet中BBox过滤算法

输入

B is a set of bounding boxes, shape is [N, 4]

L is the number of categories

C is class confidence, where C[i][0] is the background category, shape is [N, L]

d is the threshold of confidence

a is the threshold of the area

S is the size of the input image

输出

P is the BBox of the positive sample

T is the confidence of the positive sample

```
for each level i in [1, N] do
    compute the index of the maximum value of C_i: m = argmax(C_i)
    calculate the maximum value of C_i: v = max(C_i)
    if m ! = 0 and v >= d do
        P = PUB[i]
        T = TUC[i]
    end if
end for
remove small BBoxes: P, T = RemoveSmallBoxes(P, T, a)
clip BBox out of range of image: P = ClipBoxes (P, S)
return P, T
```

3　实验

3.1　数据集和环境

实验使用INRIA行人数据集，它是当前使用较为广泛的静态行人检测数据集[23]，具有拍摄条件多样化、背景复杂、存在人体遮挡以及光线强度变化大等情况。

实验使用的深度学习框架为PyTorch，模型中的主干网络ResNet和MobileNet均来自于torchvision的实现，实验使用的GPU为Tesla P100 16G型号。

3.2 实验过程

本文通过比较改进后的主干网络和损失函数进行实验。实验采用不同的IoU阈值计算平均精度(AP)去评价预测的结果，IoU阈值的选取分别为0.50~0.95(AP)、0.5(AP50)和0.75(AP75)。

实验将原有的CenterNet、改进后的PD-CenterNet和损失函数进行对比，在实验中使用轻量级主干网络MobileNet和较大的ResNet作为主干网络进行不同的实验。

第1组实验使用原有的CenterNet和Focal Loss损失函数，并在MobileNet和ResNet上进行实验，如表3所示。

表3 CenterNet网络和Focal loss函数

Table 3 CenterNet network and Focal Loss function %

网络模型	AP	AP50	AP75
CenterNet + MobileNet	30.01	67.23	19.75
CenterNet + ResNet18	32.76	75.13	22.67
CenterNet + ResNet50	38.20	77.05	27.92
CenterNet + ResNet101	37.69	80.43	29.84

第2组实验使用CenterNet来预测BBox和置信度，在训练时使用改进后的损失函数来计算损失，如表4所示。

表4 CenterNet网络和改进的损失函数

Table 4 CenterNet network and improved loss function %

网络模型	AP	AP50	AP75
CenterNet + MobileNet	31.12	70.81	20.46
CenterNet + ResNet18	34.82	80.29	25.48
CenterNet + ResNet50	38.05	81.31	28.54
CenterNet + ResNet101	39.98	81.50	27.46

第3组实验使用改进后的预测网络PD-CenterNet来预测BBox和置信度，在训练时使用Focal Loss来计算损失，如表5所示。

表5 PD-CenterNet网络和Focal loss函数

Table 5 PD-CenterNet letwork and Focal Loss function %

网络模型	AP	AP50	AP75
PD-CenterNet + MobileNet	31.24	72.33	20.85
PD-CenterNet + ResNet18	40.85	76.67	38.55
PD-CenterNet + ResNet50	38.10	78.16	31.46
PD-CenterNet + ResNet101	41.58	81.14	36.68

第4组实验使用改进后的预测网络PD-CenterNet来预测BBox和置信度，在训练时使用改进后的损失函数来计算损失，如表6所示。

表6 PD-CenterNet网络和改进后的损失函数

Table 6 PD-CenterNet network and improved loss function %

网络模型	AP	AP50	AP75
PD-CenterNet + MobileNet	32.64	75.24	24.27
PD-CenterNet + ResNet18	43.29	84.94	43.02
PD-CenterNet + ResNet50	43.88	84.08	42.47
PD-CenterNet + ResNet101	45.16	84.70	45.90

3.3 结果分析

本文改进的PD-CenterNet相对于CenterNet在准确度上有了明显的提升，最高提升了9.81%，如图5所示。改进后的网络结构由于使用了特征融合模块，所以在AP50准确度上提高了5.1%(见表3、表5)，在AP准确度上提高了1.23%，在AP75准确度提高了0.39%；改进后的损失函数也相对于Focal Loss函数[24]在使用ResNet18作为主干网络时，AP50准确度也提高了5.16%(见表3、表4)，同时在AP和AP75准确度上也都有着明显的提升。最终改进网络结构和损失函数在使用ResNet18作为主干网络时提升最大，在AP50准确度提升了9.81%(见表3、表6)，在AP准确度上提高了8.09%，在AP75准确度上提高了20.35%。

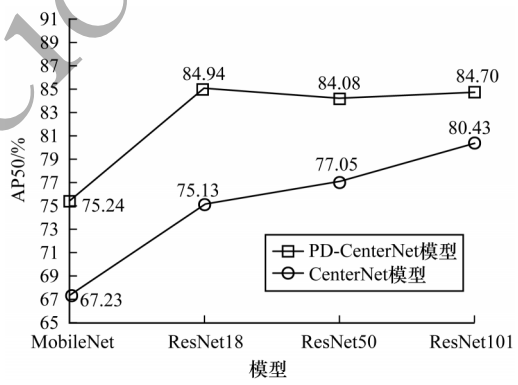


图5 模型准确度比较

Fig.5 Comparison of model accuracy

ResNet在所有主干网络中平均精度最高，但相对于MobileNet作为主干网络在速度上相对较慢。ResNet18速度在ResNet系统中最快，达到136 frame/s(图6)，在AP50准确度上与ResNet50和ResNet101持平，但PD-CenterNet在使用ResNet101时AP和AP75最高。实验结果表明，主干网络越大，则AP和AP75就越好，因此在对行人进行检测时也就更为准确。

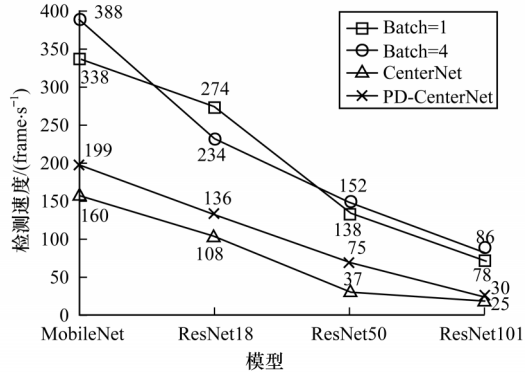
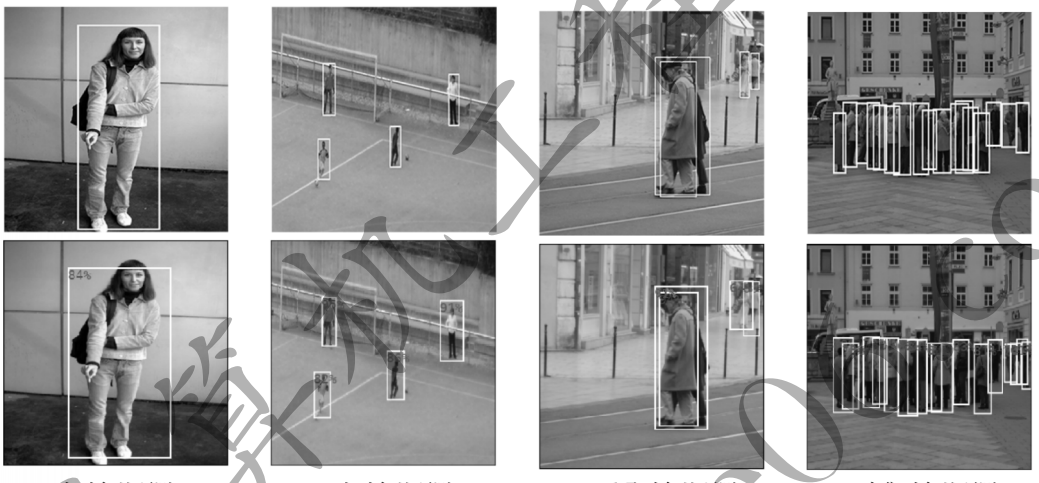


图6　模型检测速度比较

Fig.6　Comparison of mode detectl speeds

MobileNet相对于ResNet在速度上优势更为明显，能够达到199 frame/s（图6），但是在准确度上低于ResNet，而改进后的模型准确度相比ResNet提高了8.01%。

从上述实验结果可以看出，在网络结构中使用特征融合改进后的PD-CenterNet，AP、AP50以及AP75都具有较好的表现；改进后的损失函数也表现出了良好的结果。最终的检测结果如图7所示，其中，第1行为GT（Ground Truth），第2行为模型预测的结果，在检测结果中设置的置信度阈值为0.5。从图7可以看出，本文模型在大对象、小对象、对象重叠、密集对象等场景下仍然能获得较好的结果。



(a)大对象的预测　(b)小对象的预测　(c)重叠对象的预测　(d)密集对象的预测

图7　行人检测结果

Fig.7　Pedestrian detection results

4　结束语

为平衡行人检测速度和准确度的问题，本文提出一种基于CenterNet的行人检测模型PD-CenterNet。在速度上设计一个轻量级特征融合模块来减小网络结构中上采路径的计算量，并在推理时降低后处理程序的时间复杂度。在准确度上设计特征融合模块对低层和高层特征进行融合，并在损失函数中设计α、γ、δ 3个因子来改善正负样本不平衡的问题。实验结果表明，该模型对行人检测的AP50准确度为84.94%，检测速度达到136 frame/s。本文在网络设计时仅使用了2个特征融合模块，在小对象上的IoU不是最优，因此提高小对象的IoU将是下一步的主要工作。

参考文献

[1] 邢浩强，杜志岐，苏波. 基于改ssd的行人检测方法[J]. 计算机工程，2018，44(11)：228-238.
XING H Q，DU Z Q，SU B. Pedestrian detection method based on modified SSD[J]. Computer Engineering，2018，44(11)：228-238. (in Chinese)

[2] 田仙仙，鲍泓，徐成. 一种改进hog特征的行人检测算法[J]. 计算机科学，2014，41(9)：320-324.
TIAN X X，BAO H，XU C. Improved HOG algorithm of pedestrian detection[J]. Computer Science，2014，41(9)：320-324. (in Chinese)

[3] GIRSHICK R，DONAHUE J. Rich feature hierarchies for accurate object detection and semantic segmentation[EB/OL]. [2020-06-10]. https://arxiv. org/abs/1311. 2524.

[4] GIRSHICK R. Fast R-CNN[C]//Proceedings of ICCV'15. Santiago，USA：IEEE Press，2015：1440-1448.

[5] 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.

[6] 陈恩加，唐向宏，傅博文. 遗漏负样本挖掘的行人检测方法[J]. 计算机辅助设计与图形学学报，2019，31(2)：332-339.
CHEN E J，TANG X H，FU B W. Pedestrian detection based on escaped negative samples mining[J]. Computer & Digital Engineering，2019，31(2)：332-339. (in Chinese)

[7] 高宗，李少波，陈济楠，等. 基于YOLO网络的行人检测方法[J]. 计算机工程，2018，44(5)：215-226.
GAO Z，LI S B，CHEN J N，et al. Pedestrian detection method based on YOLO network[J]. Computer Engineering，2018，44(5)：215-226. (in Chinese)

[8] REDMON J，DIVVALA S，GIRSHICK R，et al. You only look once：unified，real time object detection [C]//

Proceedings of CVPR'16. Las Vegas, USA: IEEE Press, 2016: 779-788.
[9] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//Proceedings of IEEE CVPR'17. Honolulu, USA: IEEE Press, 2017: 6517-5425.
[10] REDMON J, FARHADI J. Yolov3: an in-cremental improvement[EB/OL]. [2020-06-10]. https://arxiv. org/abs/1804. 02767.
[11] ZHOU X Y, WANG D Q, KRAHENBUHL P. Centernet: objects as points[EB/OL]. [2020-06-10]. https://arxiv. org/abs/1904. 07850.
[12] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//Proceedings of ECCV'16. Amsterdam, Netherlands: Springer, 2016: 21-37.
[13] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-nms improving object detection with one line of code [C]// Proceedings of ICCV'17. Venice, Italy: [s. n.], 2017: 5562-5570.
[14] SANDLER M, HOWARD A, ZHU M L, et al. MobileNet v2: inverted residuals and linear bottlenecks[C]//Proceedings of CVPR'18. Salt Lake City, USA: IEEE Press, 2018: 4510-4520.
[15] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//Proceedings of CVPR' 16. Las Vegas, USA: IEEE Press, 2016: 770-778.
[16] CHOLLER F. Xception: deep learning with depthwise separable convolutions [C]//Proceedings of CVPR' 17. Honolulu, USA: IEEE Press, 2017: 1800-1807.
[17] ZHANG X Y, ZHOU X Y, LIN M X, et al. Shufflenet: an extremely efficient convolutional neural network for mobile devices[C]//Proceedings of CVPR' 18. Salt Lake City, USA: IEE Press, 2018: 6848-6856.
[18] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]//Proceedings of CVPR'15. Boston, USA: IEEE Press, 2015: 1-9.
[19] HUANG G, LIU Z, VAN DER M, et al. Densely connected convolutional networks[C]//Proceedings of CVPR' 17. Honolulu, USA: IEEE Press, 2017: 2261-2269.
[20] YU C Q, WANG J B, PENG C, et al. BISENET: bilateral segmentation network for real-time semantic segmentation [C]//Proceedings of ECCV' 18. Munich, Germany: Springer, 2018: 325-341.
[21] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of CVPR'18. Salt Lake City, USA: IEEE Press, 2018: 7132-7141.
[22] KINGMA D P, BA J. Adam: a method for stochastic optimization[EB/OL]. [2020-06-10]. https://arxiv. org/abs/1412. 6980v8.
[23] 谢林江, 季桂树, 彭清, 等. 改进的卷积神经网络在行人检测中的应用[J]. 计算机科学与探索, 2018, 12(5): 708-718.
XIE L J, JI G S, PENG Q, et al. Application of preprocessing convolutional neural network in pedestrian detection[J]. Journal of Frontiers of Computer Science and Technology, 2018, 12(5): 708-718. (in Chinese)
[24] LIN, T Y, GOYAL P. Focal loss for dense object detection [C]//Proceedings of ICCV'17. Venice, Italy: [s. n.], 2017: 318-327.

编辑 索书志