基于特征通道和空间联合注意机制的遮挡行人检测方法

陈 勇*① 刘 曦① 刘焕淋②

①(重庆邮电大学工业物联网与网络化控制教育部重点实验室 重庆 400065)

②(重庆邮电大学通信与信息工程学院 重庆 400065)

摘 要：遮挡是行人检测任务中导致漏检发生的主要原因之一，对检测器性能造成了不利影响。为了增强检测器对于遮挡行人目标的检测能力，该文提出一种基于特征引导注意机制的单级行人检测方法。首先，设计一种特征引导注意模块，在保持特征通道间的关联性的同时保留了特征图的空间信息，引导模型关注遮挡目标可视区域；然后，通过注意模块融合浅层和深层特征，从而提取到行人的高层语义特征；最后，将行人检测作为一种高层语义特征检测问题，通过激活图的形式预测得到行人位置和尺度，并生成最终的预测边界框，避免了基于先验框的预测方式所带来的额外参数设置。所提方法在CityPersons数据集上进行了测试，并在Caltech数据集上进行了跨数据集实验。结果表明该方法对于遮挡目标检测准确度优于其他对比算法。同时该方法实现了较快的检测速度，取得了检测准确度和速度的平衡。

关键词：遮挡行人检测；单级检测器；注意机制

中图分类号：TN911.73; TP391.41 **文献标识码**：A **文章编号**：1009-5896(2020)06-1486-08

DOI: 10.11999/JEIT190606

Occluded Pedestrian Detection Based on Joint Attention Mechanism of Channel-wise and Spatial Information

CHEN Yong① LIU Xi① LIU Huanlin②

①(*Key Laboratory of Industrial Internet of Things & Network Control, Ministry of Education, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)

②(*School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)

Abstract: Pedestrian detector performance is damaged because occlusion often leads to missed detection. In order to improve the detector's ability to detect pedestrian, a single-stage detector based on feature-guided attention mechanism is proposed. Firstly, a feature attention module is designed, which preserves the association between the feature channels while retaining spatial information, and guides the model to focus on visible region. Secondly, the attention module is used to fuse shallow and deep features, then high-level semantic features of pedestrians are extracted. Finally, pedestrian detection is treated as a high-level semantic feature detection problem. Pedestrian location and scale are obtained through heat map prediction, then the final prediction bounding box is generated. This way, the proposed method avoids the extra parameter settings of the traditional anchor-based method. Experiments show that the proposed method is superior to other comparison algorithms for the accuracy of occlusion target detection on CityPersons and Caltech pedestrian database. At the same time, the proposed method achieves a faster detection speed and a better balance between detection accuracy and speed.

Key words: Occluded pedestrian detection; Single-stage detector; Attention mechanism

收稿日期：2019-08-09；改回日期：2020-02-18；网络出版：2020-03-13

*通信作者： 陈勇 chenyong@cqupt.edu.cn

基金项目：国家自然科学基金(51977021)

Foundation Item: The National Natural Science Foundation of China (51977021)

1　引言

行人检测是计算机视觉应用中的重要任务，在自动驾驶、视频监控和机器人等领域有广泛应用。然而，遮挡往往导致漏检发生，从而降低检测器准确性。因此对遮挡问题进行研究具有非常重要的现实意义[1]。

传统行人检测器通常采用基于手工特征设计和机器学习的方法，如文献[2]采用的基于自适应增强算法(adaboost)和支持向量机(Support Vector Machine, SVM)的方法，以及文献[3]所提出的基于后验方向梯度直方图特征(Histogram of Oriented Gradient, HOG)的多姿态行人检测方法。而随着深度学习技术不断发展，出现了大量基于深度神经网络的行人检测方法。主流基于深度学习的行人检测器通常采用基于更快的区域卷积神经网络(Faster Region-based Convolutional Neural Networks, Faster R-CNN)[4]的两级检测结构，如文献[5,6]分别通过多尺度网络改进Fast[7]和Faster R-CNN，以此来处理尺度变化问题并取得了有竞争力的性能。单级检测器方面，文献[8]提出了渐近定位拟合策略提升目标定位准确度，而文献[9]则将行人检测问题作为语义特征检测问题加以处理。虽然行人检测任务的准确度不断被刷新，但受现实场景中遮挡问题的影响，检测器漏检情况仍很普遍[10]。

针对行人遮挡问题，文献[11]提出了一种基于遮挡感知的池化单元用于取代Fast R-CNN的池化层。文献[12]构建了一个统一的框架来同时处理特征提取、形变、遮挡和分类问题，然而基于部件检测方法通常时间开销较大。文献[13]充分利用了背景信息来提升对遮挡目标的检测能力。文献[14]通过整合行人头-肩特征的方式来识别遮挡行人。文献[15]则从特征关注的角度设计了多种注意机制，引导检测器更多关注遮挡行人可视部分，所提出的方法虽然对于遮挡行人的检测能力有显著提升，但是仍存在以下问题：首先基于Faster R-CNN的两阶段网络结构虽然可以帮助检测器取得更高的检测准确度，但由于复杂度高而缺乏实时性；其次，该方法采用了基于先验框的检测策略，然而先验框通常难以取得最优的设置和分配方式[16]；最后，该方法仅对特征通道间的相关性建模而忽略了特征的空间信息，使得网络难以准确定位感兴趣区域。

针对上述问题，本文通过提出的注意网络对浅层和深层特征进行融合，引导模型更多关注遮挡目标可视区域；并通过改进的解析网络将行人检测问题简化为高层语义特征检测问题，采用激活图的方式预测得到行人的位置和尺度，避免了额外参数设置。实验表明，所提方法在公开数据集上对遮挡行人检测的准确度优于目前主流方法，并取得了较快的检测速度。

2　算法原理设计

图1为本文所构建的行人检测模型，它由特征提取网络和行人解析网络两部分组成。输入图像经特征提取网络提取高层语义特征，并通过特征引导注意模块进行特征融合；解析网络在所获取高层语义抽象的基础上分别预测行人位置、高度和偏移激活图，得到预测边界框。

2.1　特征提取网络

特征提取网络是在特征金字塔网络(Feature Pyramid Networks, FPN)[17]的基础上构建的，包括基础网络和注意网络。由于ResNet50在图像分类等视觉任务中的性能表现优异，因此将其作为基础网络。ResNet50可以分为5级，每级相对于输入图像的下采样率为$s=2^l$, $l\in\{1,2,3,4,5\}$表示级数，如图1所示，每级的输出特征图表示为C_l。为了充分利用浅层特征图的位置信息和深层特征的语义信息，将浅层和深层特征图通过所提出的特征引导注意网络进行特征融合。其具体描述为：首先通过卷积核大小为1×1的卷积层将$C3$和$C4$特征通道数降低为256以减少计算量；然后将经过双线性插值上采样2倍后的主干网络特征图(即$P4$和$P3$)分别输入到引导注意模块中进行融合。

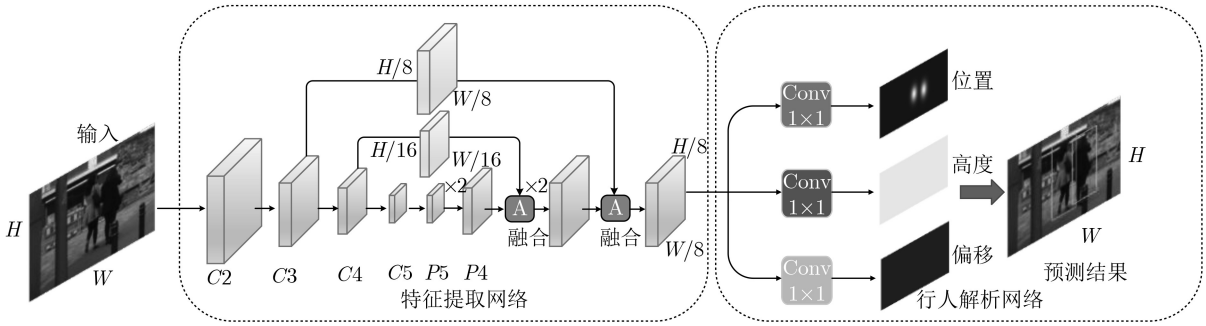

图 1 模型总体结构

2.2 注意网络

卷积网络的不同特征通道会对行人目标特定区域产生响应[15]，即通过不同的特征通道可以描述目标的遮挡形式，并将遮挡形式n定义为

$$\mathrm{occl}(n)=[v_0p_0v_1p_1\cdots v_kp_k] \tag{1}$$

其中，p_i表示行人目标的不同区域，$v_i\in\{0,1\}$，$i\in[0,k]$用于指示行人部分区域是否可见。

由于传统CNN通道权重通常是固定且相等的，限制了网络对于不同遮挡形式的表达能力。文献[15]对每个通道的权重进行重新标定，使得表达遮挡目标可视区域的特征通道对于最终卷积特征有更大的贡献，从而在背景中突出遮挡目标。通道重加权的过程可以表示为

$$\boldsymbol{F}_{\mathrm{occl}}(n)=\boldsymbol{\Omega}_n\boldsymbol{F}_{\mathrm{chn}} \tag{2}$$

其中，$\boldsymbol{F}_{\mathrm{chn}}$是通道特征，而$\boldsymbol{\Omega}_n$是对应遮挡形式$n$的通道权重向量，注意模块的任务之一就是通过学习得到注意向量$\boldsymbol{\Omega}_n$，并最终通过$\boldsymbol{\Omega}_n$对特征通道进行重加权，使得网络能自适应地表达不同遮挡形式。

然而现有文献，如文献[16]仅考虑了通道间的联系而忽略了空间信息对于特征图的重要性。由于特征图的空间信息有利于网络定位感兴趣目标区域，对此文献[18]将特征通道注意机制和空间注意机制用于图像描述任务。同样，文献[19]将空间注意机制用于目标检测任务，从而引导网络突出对当前任务有用的特征。在参考上述文献的基础上，将特征空间信息用于行人检测任务当中，从而突出遮挡行人目标区域，并构建了空间注意模块来实现这一点。空间注意模块通过统计特征图的空间信息得到空间注意图用于对输入特征进行重新激活，从而引导网络关注遮挡行人目标并抑制背景干扰。

如图2所示，注意网络由通道注意和空间注意两个子模块组成。注意网络的输入分别为浅层和深层卷积层引出的两个特征图(如$C4$和$P4$)。网络首先将输入特征在通道维度进行连接得到$\boldsymbol{F}\in\mathbb{R}^{H\times W\times C}$，再将$\boldsymbol{F}$输入通道注意模块和空间注意模块进行特征融合。

综上所述，利用注意模块建模特征通道间的相关性以及特征图的空间信息，网络不仅能增强相关区域的特征表示，还能获取感兴趣区域的位置信息，在充分利用有用的特征来处理行人遮挡问题的同时，还抑制了无用信息，为提高遮挡行人检测的准确度提供了理论依据。

2.2.1 特征通道注意模块

图3为本文所提特征通道注意模块。对于输入特征图$\boldsymbol{F}$，首先通过全局平均池化和最大池化操作获取每个特征通道的全局信息，分别构成通道描述符$z_{\mathrm{chn}}^{\mathrm{avg}}$和$z_{\mathrm{chn}}^{\mathrm{max}}$；再通过两个全连接层FC1和FC2获取特征通道注意向量$\boldsymbol{\Omega}_{\mathrm{chn}}\in\mathbb{R}^{1\times1\times C}$，通过学习的方式让网络自动表征不同样本的遮挡形式，具体步骤由式(3)表示。

$$\boldsymbol{\Omega}_{\mathrm{chn}}=\sigma\left(\boldsymbol{W}_2\left(\delta\left(\boldsymbol{W}_1z_{\mathrm{chn}}^{\mathrm{avg}}\right)\right)+\boldsymbol{W}_2\left(\delta\left(\boldsymbol{W}_1z_{\mathrm{chn}}^{\mathrm{max}}\right)\right)\right) \tag{3}$$

其中，σ表示sigmoid函数，δ表示ReLU函数，$\boldsymbol{W}_1\in\mathbb{R}^{C/r\times C}$和$\boldsymbol{W}_2\in\mathbb{R}^{C\times C/r}$分别表示两个全连接层参数，$r$是降维比例，最后用$\boldsymbol{\Omega}_{\mathrm{chn}}$对输入特征$\boldsymbol{F}$的进行逐通道加权得到$\boldsymbol{F}'$

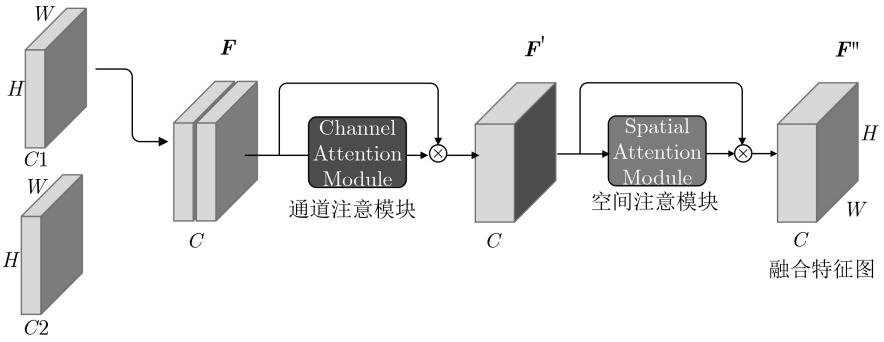

图 2 注意模块总体结构

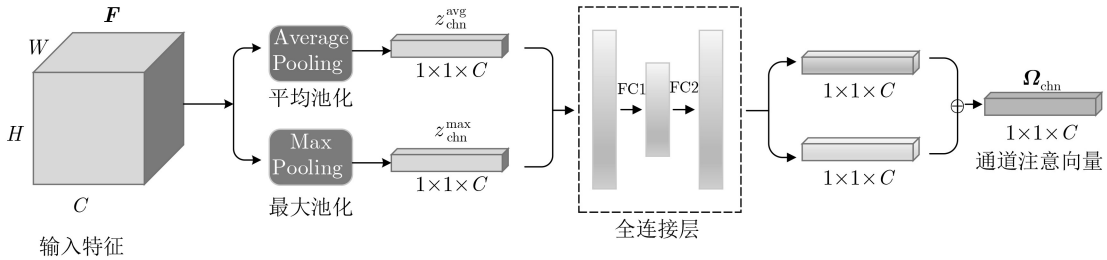

图 3 特征通道注意模块结构

$$\boldsymbol{F}' = \boldsymbol{\Omega}_{\text{chn}} \cdot \boldsymbol{F} \tag{4}$$

其中，·表示逐通道相乘。

2.2.2 空间注意模块

由于遮挡行人目标的有用信息通常被背景所掩盖，在经过通道注意模块增强遮挡目标的特征表达的同时，网络也需要对有用信息的空间位置进行确定。与通道注意机制不同，空间注意机制主要用于突出特征图中与当前任务有关联的区域，即引导网络关注遮挡目标的可视区域。

图4为本文所构建的空间注意模块。首先在通道维度上对输入特征图$\boldsymbol{F}'$进行最大池化操作得到特征图$\boldsymbol{F}'_{\text{max}} \in \mathbb{R}^{H\times W\times 1}$，用来统计特征图的空间信息；再将特征图输入一个3×3的卷积层f_{conv}并通过sigmoid函数输出得到空间注意图$\boldsymbol{M}_{sp} \in \mathbb{R}^{H\times W\times 1}$

$$\boldsymbol{M}_{\text{sp}} = \sigma(f_{\text{conv}}(\boldsymbol{F}'_{\text{max}})) \tag{5}$$

其中，σ表示sigmoid函数，最后用空间注意图$\boldsymbol{M}_{\text{sp}}$重新激活输入$\boldsymbol{F}'$得到最终的特征图$\boldsymbol{F}''$

$$\boldsymbol{F}'' = \boldsymbol{M}_{\text{sp}} \odot \boldsymbol{F}' \tag{6}$$

其中，⊙表示特征图逐元素相乘。

2.3 行人解析网络

将行人检测任务视为一种高层语义特征检测问题，在获得语义特征基础上，通过行人解析网络得到最终的预测边界框，解析网络结构如图5所示。

参照文献[9]，本文对行人位置、高度及位置偏移量进行预测，通过简单的几何换算得到边界框的大小。具体而言，在得到行人的预测高度h后，通过边界框长宽比$a = 0.41$，即可计算得到边界框宽度$w = h \cdot a$。

特征提取网络的输出特征图为$\boldsymbol{F}_{\text{final}} \in \mathbb{R}^{H/s\times W/s\times C}$，通过并联的3个1×1卷积层分别预测3个激活图，对应行人中心位置$\boldsymbol{H}_{\text{center}} \in \mathbb{R}^{H/s\times W/s}$、高度$\boldsymbol{H}_{\text{height}} \in \mathbb{R}^{H/s\times W/s}$和位置偏移$\boldsymbol{H}_{\text{offset}} \in \mathbb{R}^{H/s\times W/s}$，其中，$s$是输出激活图相对于输入图像的下采样率。通过预测激活图的方式避免了传统方法所采用的先验框限制，实现了更加灵活的检测。

2.3.1 位置预测

行人目标的位置预测通过位置激活图$\boldsymbol{H}_{\text{center}}$实现。本文将位置预测问题简化为一个二分类问题，设行人目标中心在特征图$\boldsymbol{F}_{\text{final}}$上的位置为$(x_c, y_c)$，将目标中心像素作为正样本，其他位置作为负样本，通过交叉熵损失函数来优化训练位置预测分支。训练真值$\boldsymbol{H}^{\text{gt}}_{\text{center}}$由一个2D高斯函数$G$[20]生成，对于任意位置$(i, j)$处的真值可通过计算式(7)得到

$$\left.\begin{aligned} \boldsymbol{H}^{\text{gt}}_{\text{center}}(i,j) &= \max\left(G\left(i,j;x_c,y_c,\sigma_w,\sigma_h\right),0\right) \\ G\left(i,j;x_c,y_c,\sigma_w,\sigma_h\right) &= \mathrm{e}^{-\left((i-x_c)^2/2\sigma_w^2+(j-y_c)^2/2\sigma_h^2\right)} \end{aligned}\right\} \tag{7}$$

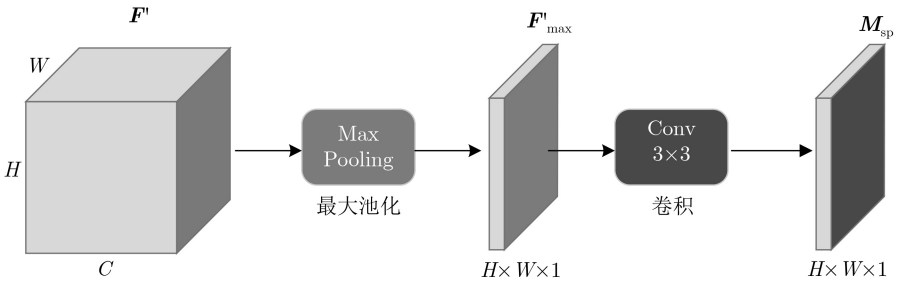

图 4 空间关注模块

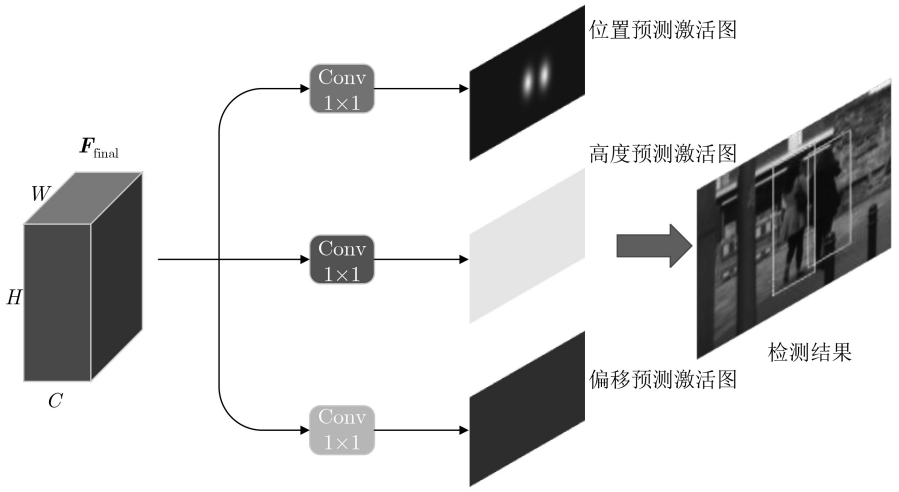

图 5 行人解析网络

其中，(x_c, y_c)为目标的中心位置，σ_w和σ_h分别为目标的宽和高的标准差。为了缓解训练过程中正负样本类别失衡问题，采用Focal Loss将中心位置预测损失函数定义为

$$L_{\text{center}} = -\frac{1}{N}\sum_{i=1}^{W/s}\sum_{j=1}^{H/s} \begin{cases} (1-p_{i,j})^{\alpha}\lg(p_{i,j}) & ,\boldsymbol{H}^{\text{gt}}_{\text{center}}(i,j)=1 \\ \left(1-\boldsymbol{H}^{\text{gt}}_{\text{center}}(i,j)\right)^{\beta} p_{i,j}^{\alpha}\lg(1-p_{i,j}), & \text{其他} \end{cases} \tag{8}$$

其中，$p_{i,j}$表示预测激活图中(i,j)处为目标中心的预测分数，N是图片中目标的个数，α和β是平衡因子，本文设置$\alpha=2$, $\beta=4$。

2.3.2 高度预测

设行人k在高度激活图中的位置为(x_k, y_k)，对应真值为$\boldsymbol{H}^{\text{gt}}_{\text{height}}(x_k, y_k)$，由式(9)得到

$$\boldsymbol{H}^{\text{gt}}_{\text{height}}(x_k, y_k) = \lg(h_k) \tag{9}$$

其中，h_k表示行人目标k的高度。本文将(x_k, y_k)半径r范围内的真值设置为$\lg(h_k)$, r根据目标的宽度设定，设置$r=0.5w_k$，使用smooth L1作为损失函数

$$L_{\text{height}} = \frac{1}{N}\sum_{k=1}^{N}\text{SmoothL}1\left(\hat{h}_k, \boldsymbol{H}^{\text{gt}}_{\text{height}}(x_k, y_k)\right) \tag{10}$$

其中，$\hat{h}_k$为激活图中目标k的预测高度，N是图片中目标的个数。

2.3.3 偏移预测

由于卷积网络通常是一个下采样的过程，输入图像上的位置(x, y)映射在激活图上的位置可以表示为$(x/s, y/s)$，其中s为网络的下采样率。当把激活图上的位置重新映射回输入图像上时会产生误差，尤其影响较小目标的预测结果。为缓解这一问题，通过预测中心位置的偏移$\hat{o}_k$来修正目标的位置预测，相应的真值o_k为

$$\boldsymbol{H}^{\text{gt}}_{\text{offset}}(x_k, y_k) = \left(x_k/s - \lfloor x_k/s \rfloor, y_k/s - \lfloor y_k/s \rfloor\right) \tag{11}$$

其中，(x_k, y_k)是目标k的中心位置。在训练阶段，同样采用smooth L1作为损失函数

$$L_{\text{offset}} = \frac{1}{N}\sum_{k=1}^{N}\text{SmoothL}1\left(\hat{o}_k, \boldsymbol{H}^{\text{gt}}_{\text{offset}}(x_k, y_k)\right) \tag{12}$$

最后，通过多任务损失函数联合优化训练网络进行训练

$$L = \lambda_{\text{c}}L_{\text{center}} + \lambda_{\text{h}}L_{\text{height}} + \lambda_{\text{o}}L_{\text{offset}} \tag{13}$$

其中，λ_{c}, λ_{h}和λ_{o}为权重因子，参照文献[9]分别设置为0.01, 1和0.1。

3 实验结果与分析

3.1 实验设计

3.1.1 实验环境

算法基于Pytorch深度学习框架实现，工作站配备为：64 GB内存、Intel Xeon E5 CPU和两块Nvidia GTX1080Ti GPU，它主要用于网络训练，在测试阶段仅使用1块GPU。

3.1.2 数据集的选择

CityPersons[21]和Caltech[22]为所选用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2048×1024。由于CityPersons包含大量被遮挡的行人图片，对此，选其作为所提方法的验证和对比实验的数据集；Caltech包含约35万个行人样本，其中标准测试集由4024张分辨率为640×680城市道路场景图片组成，为验证所提方法的泛化性，相关测试在Caltech数据集上进行。

3.1.3 评价指标的选择

采用Caltech的标准评价指标进行性能评估，即为每张图片的误检率(False Positives Per Image, FPPI)介于$[10^{-2}, 10^{0}]$间的对数平均漏检率(log-average Miss Rate, MR)，记为MR^{-2}。MR^{-2}值越低说明检测器性能越好。由于本文主要关注遮挡行人问题，按照目标可视范围比例v的不同，实验条件设置如表1所示。

3.1.4 参数的设置

选择预训练的ResNet50作为主干网络，采用Adam算法优化训练模型，初始学习率设置为2×10^{-4}，训练批次大小为4张图片，总共训练250轮。训练和测试图片尺寸分别为1280×640和2048×1024。采用非极大值抑制算法(Non-Maximum Suppression, NMS)滤除冗余预测结果，其中交并比(Intersection over Union, IoU)阈值设置为0.5，仅保留目标置信度分数大于0.1的预测结果。

3.2 注意网络的验证

为了验证特征引导注意网络的有效性，将去除注意网络的检测器作为测试基准(Baseline)，将文献[16]作为对比方法。Baseline采用与FPN一致的特征融合方式来构建模型，CA表示通道注意模块，SA表示空间注意模块，实验中分别对比添加各模块后的检测器性能，验证实验条件设置如表1所示，其测试结果如表2所示。

表 1 验证实验条件设置

R (Reasonable)	HO (Heavy Occlusion)	R+HO (Reasonable+Heavy Occlusion)
$v\in[0.65,\infty)$	$v\in[0.20,0.65]$	$v\in[0.20,\infty)$

表 2　注意网络验证结果MR^{-2}(%)

方法	R	HO	R+HO
文献[16]	16.0	56.7	38.2
Baseline	12.1	41.1	38.1
Baseline+CA	11.8	39.2	37.8
Baseline+CA+SA	**11.6**	**38.5**	**37.3**

与基准模型对比，在添加注意模块后，检测器对于遮挡目标的MR^{-2}具有明显下降，说明所提注意机制能有效地引导检测器关注遮挡目标。与文献[16]对比，所提方法在所有评价标准下，MR^{-2}都得到显著的降低，特别是在HO评价标准下具有明显优势，其MR^{-2}下降了18.2%，表明所提方法对于遮挡目标检测的有效性。

为了更直观了解注意模块对检测器性能的影响，图6给出可视化的位置预测激活图$\boldsymbol{H}_{\text{center}}$。其中，图6(a)是一幅遮挡行人图像，分别将其输入Baseline和Baseline+CA+SA得到位置预测热度图6(b)和图6(c)。通过观察可以发现图6(c)中的特征响应更接近目标可视区域，而图6(b)中还存在背景干扰。这也证明了注意模块可以引导网络更多关注遮挡目标的可视部分，同时还降低了背景噪声对检测性能的影响。

3.3　对比实验

为了进一步验证所提方法的性能，将本文方法与其他行人检测方法进行对比。实验条件设置：CityPersons数据集所提供的标准评估条件，评价指标为MR^{-2}。对比方法选取OR-CNN[11], Faster-RCNN[21]和RepLoss[23]等主流方法，ALFNet[8], CSP[9]和TLL[20]等领先方法以及最新方法：CAFL[13]和PedJointNet[14]。表3给出了实验结果。在Heavy评价标准下，本文方法取得了MR^{-2}为47.6%，优于对比方法。

在检测效率方面，本文方法对CityPersons数据集中分辨率为1024×2048的输入图像的检测速度为0.22 s，实现了速度和准确度较好的平衡。若检测更小分辨率的输入图像，本文方法的检测速度将进一步提升，这一点在3.4节的泛化性实验中得到了验证。

3.4　泛化性实验

为了验证所提方法的泛化性能，将所提方法在CityPersons训练集上进行训练，在Caltech的Heavy测试子集上采用修正后的标签[10]进行了跨数据集实验。其中Heavy子集由高度大于50像素且可视范围为[0.20, 0.65]的行人目标组成。

如图7所示，所提方法取得的MR^{-2}为61.41%，

(a) 输入图像　　(b) Baseline输出位置热图　(c) Baseline+CA+SA输出位置热图

图 6 可视化位置预测热图

表 3　CityPersons数据集测试结果MR^{-2}(%)

方法	主干网络	Reasonable	Heavy	Partial	Bare	测试时间(s)
OR-CNN[11]	VGG-16	12.8	55.7	15.3	**6.7**	–
FasterRCNN[21]	VGG-16	15.4	–	–	–	–
ALFNet[8]	ResNet-50	12.0	51.9	11.4	8.4	0.27
CSP[9]	ResNet-50	**11.0**	49.3	10.4	7.3	0.33
CAFL[13]	ResNet-50	11.4	50.4	12.1	7.6	–
PedJointNet[14]	ResNet-50	13.5	52.1	–	–	–
TLL[20]	ResNet-50	15.5	53.6	17.2	10.0	–
RepLoss[23]	ResNet-50	13.2	56.9	16.8	7.6	–
本文方法	ResNet-50	11.6	**47.6**	**9.8**	7.5	**0.22**

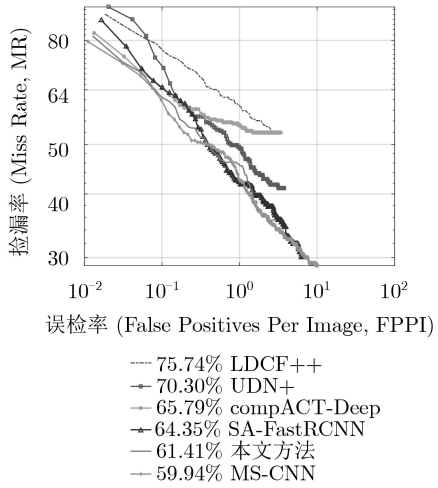

图 7 Caltech跨数据库实验

略高于MS-CNN(59.94%)，但优于其他对比方法，特别是优于主流行人检测方法：SA-FastRCNN[5](64.4%)和UDN+[12](70.3%)等，充分说明所提方法具有较强的泛化性能。同时，在Caltech测试集上的检测速度为0.028s/img，满足了实时检测的要求。

4 结论

针对行人检测普遍面临的遮挡问题，本文构建了一种特征引导注意网络用于特征融合，通过建模特征通道的关联性，使得网络自适应地表达不同的遮挡形式，并结合特征的空间信息引导网络关注遮挡目标的可视区域。在所构建的注意网络的基础上，设计了一种单级行人检测器，将行人检测问题简化为语义特征检测问题，避免了先验框所带来的限制。在CityPersons数据集上的实验结果表明所提方法对于遮挡目标检测具有优势，通过在Caltech上的跨数据集实验，验证了所提方法具有较好的泛化性能，同时体现出所提方法具有较好的实时性。

参 考 文 献

[1] 张功国, 吴建, 易亿, 等. 基于集成卷积神经网络的交通标志识别[J]. 重庆邮电大学学报: 自然科学版, 2019, 31(4): 571–577. doi: 10.3979/j.issn.1673-825X.2019.04.019.
ZHANG Gongguo, WU Jian, YI Yi, *et al*. Traffic sign recognition based on ensemble convolutional neural network[J]. *Journal of Chongqing University of Posts and Telecommunications*: *Natural Science Edition*, 2019, 31(4): 571–577. doi: 10.3979/j.issn.1673-825X.2019.04.019.

[2] 种衍文, 匡湖林, 李清泉. 一种基于多特征和机器学习的分级行人检测方法[J]. 自动化学报, 2012, 38(3): 375–381. doi: 10.3724/SP.J.1004.2012.00375.
CHONG Yanwen, KUANG Hulin, and LI Qingquan. Two-stage pedestrian detection based on multiple features and machine learning[J]. *Acta Automatica Sinica*, 2012, 38(3): 375–381. doi: 10.3724/SP.J.1004.2012.00375.

[3] 刘威, 段成伟, 遇冰, 等. 基于后验HOG特征的多姿态行人检测[J]. 电子学报, 2015, 43(2): 217–224. doi: 10.3969/j.issn.0372-2112.2015.02.002.
LIU Wei, DUAN Chengwei, YU Bing, *et al*. Multi-pose pedestrian detection based on posterior HOG feature[J]. *Acta Electronica Sinica*, 2015, 43(2): 217–224. doi: 10.3969/j.issn.0372-2112.2015.02.002.

[4] REN Shaoqing, HE Kaiming, GIRSHICK R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks[C]. 2015 Advances in Neural Information Processing Systems, Montreal, Canada, 2015: 91–99.

[5] LI Jianan, LIANG Xiaodan, SHEN Shengmei, *et al*. Scale-aware fast R-CNN for pedestrian detection[J]. *IEEE Transactions on Multimedia*, 2018, 20(4): 985–996. doi: 10.1109/TMM.2017.2759508.

[6] 王进, 陈知良, 李航, 等. 一种基于增量式超网络的多标签分类方法[J]. 重庆邮电大学学报: 自然科学版, 2019, 31(4): 538–549. doi: 10.3979/j.issn.1673-825X.2019.04.015.
WANG Jin, CHEN Zhiliang, LI Hang, *et al*. Hierarchical multi-label classification using incremental hypernetwork[J]. *Journal of Chongqing University of Posts and Telecommunications*: *Natural Science Edition*, 2019, 31(4): 538–549. doi: 10.3979/j.issn.1673-825X.2019.04.015.

[7] GIRSHICK R. Fast R-CNN[C]. 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1440–1448.

[8] LIU Wei, LIAO Shengcai, HU Weidong, *et al*. Learning efficient single-stage pedestrian detectors by asymptotic localization fitting[C]. The 15th European Conference on Computer Vision, Munich, Germany, 2018: 618–634.

[9] LIU Wei, LIAO Shengcai, REN Weiqiang, *et al*. High-level semantic feature detection: A new perspective for pedestrian detection[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 5182–5191.

[10] ZHANG Shanshan, BENENSON R, OMRAN M, *et al*. How far are we from solving pedestrian detection?[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 1259–1267.

[11] ZHANG Shifeng, WEN Longyin, BIAN Xiao, *et al*. Occlusion-aware R-CNN: detecting pedestrians in a crowd[C]. The 15th European Conference on Computer Vision, Munich, Germany, 2018: 637–653.

[12] OUYANG Wanli, ZHOU Hui, LI Hongsheng, *et al*. Jointly learning deep features, deformable parts, occlusion and classification for pedestrian detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(8): 1874–1887. doi: 10.1109/TPAMI.2017.2738645.

[13] FEI Chi, LIU Bin, CHEN Zhu, *et al*. Learning pixel-level

and instance-level context-aware features for pedestrian detection in crowds[J]. *IEEE Access*, 2019, 7: 94944–94953. doi: 10.1109/ACCESS.2019.2928879.

[14] LIN C Y, XIE Hongxia, and ZHENG Hua. PedJointNet: Joint head-shoulder and full body deep network for pedestrian detection[J]. *IEEE Access*, 2019, 7: 47687–47697. doi: 10.1109/ACCESS.2019.2910201.

[15] ZHANG Shanshan, YANG Jian, and SCHIELE B. Occluded pedestrian detection through guided attention in CNNs[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 6995–7003.

[16] ZHU Chenchen, HE Yihui, and SAVVIDES M. Feature selective anchor-free module for single-shot object detection[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 840–849.

[17] 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, USA, 2017: 936–944.

[18] CHEN Long, ZHANG Hanwang, XIAO Jun, *et al*. SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 6298–6306.

[19] WOO S, PARK J, LEE J Y, *et al*. Cbam: Convolutional block attention module[C]. The 15th European Conference on Computer Vision, Munich, Germany, 2018: 3–19.

[20] SONG Tao, SUN Leiyu, XIE Di, *et al*. Small-scale pedestrian detection based on topological line localization and temporal feature aggregation[C]. The 15th European Conference on Computer Vision, Munich, Germany, 2018: 536–551.

[21] ZHANG Shanshan, BENENSON R, and SCHIELE B. Citypersons: A diverse dataset for pedestrian detection[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 4457–4465.

[22] DOLLAR P, WOJEK C, SCHIELE B, *et al*. Pedestrian detection: An evaluation of the state of the art[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(4): 743–761. doi: 10.1109/TPAMI.2011.155.

[23] WANG Xinlong, XIAO Tete, JIANG Yuning, *et al*. Repulsion loss: Detecting pedestrians in a crowd[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 7774–7783.

陈　勇：男，1963年生，博士，教授，研究方向为图像处理.

刘　曦：男，1993年生，硕士生，研究方向为行人目标检测.

刘焕淋：女，1970年生，博士，教授，研究方向为信号处理等方面的研究.