

# 基于弱显著图的实时热红外图像行人检测

李传东<sup>1,2</sup>, 徐望明<sup>1,2,3</sup>, 伍世虔<sup>1,2</sup>

(1. 武汉科技大学信息科学与工程学院, 湖北 武汉 430081; 2. 武汉科技大学机器人与智能系统研究院, 湖北 武汉 430081;  
3. 武汉科技大学教育部冶金自动化与检测技术工程研究中心, 湖北 武汉 430081)

**摘要:** 针对现有热红外图像行人检测方法在精度和速度方面存在的问题, 提出一种基于弱显著图的实时行人检测方法。该方法以轻量级LFFD (Light and Fast Face Detector) 网络为基础, 由两级改进网络即SD-LFFD (Saliency Detection-LFFD) 和SF-LFFD (Saliency Fusion-LFFD) 组成, 首先以热红外图像作为输入经SD-LFFD网络产生初步行人检测结果和行人区域弱显著图, 接着将该弱显著图与原热红外图像结合“点亮”潜在行人区域并经SF-LFFD网络产生新的行人检测结果, 最后将两级改进网络的行人检测结果融合得到最终结果。在数据集CVC-09和CVC-14上实验结果表明, 该方法与现有轻量级神经网络相比行人检测的平均精确率有大幅提升, 且在有限硬件资源下可实现实时检测。

**关键词:** 热红外图像; 行人检测; 显著性检测; 实时检测

中图分类号: TP391.41 文献标识码: A 文章编号: 1001-8891(2021)07-0658-07

## Real-Time Pedestrian Detection Based on the Weak Saliency Map in Thermal Infrared Images

LI Chuandong<sup>1,2</sup>, XU Wangming<sup>1,2,3</sup>, WU Shiqian<sup>1,2</sup>

(1. School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China;  
2. Institute of Robotics and Intelligent Systems, Wuhan University of Science and Technology, Wuhan 430081, China;  
3. Engineering Research Center for Metallurgical Automation and Detecting Technology of Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China)

**Abstract:** To address the low precision and speed of existing pedestrian detection methods for thermal infrared images, a real-time pedestrian detection method based on a weak saliency map is herein proposed. The proposed method comprises two improved networks, namely, SD-LFFD and SF-LFFD, which use lightweight LFFD as the basic network. First, the thermal infrared image is input into the SD-LFFD to produce the preliminary pedestrian detection results and a weak saliency map indicating the pedestrian regions. Then, the weak saliency map and the original thermal infrared image are combined to highlight the potential pedestrian regions and generate new results using the SF-LFFD. Finally, the pedestrian detection results obtained by the two improved networks are integrated to obtain the final results. The experimental results on the CVC-09 and CVC-14 datasets indicate that the proposed method significantly improves the average precision (AP) of pedestrian detection compared with that of existing lightweight neural networks, and that it achieves real-time detection with limited hardware resources.

**Key words:** thermal infrared image, pedestrian detection, saliency detection, real-time detection

## 0 引言

自动行人检测技术被广泛应用于车载安全系统、视频监控系统等计算机视觉任务中。基于可见光图像

的行人检测算法在光照不足、不均时效果较差, 而基于热红外图像的行人检测算法因其热辐射成像原理受光照条件影响小, 适合全天候工作, 近年来得以广泛关注且成为研究热点。传统的热红外行人检测算法

收稿日期: 2020-09-24; 修订日期: 2020-11-03.

作者简介: 李传东 (1995-), 男, 湖北黄冈人, 硕士研究生, 研究方向为图像处理、目标检测及深度学习。

通信作者: 徐望明 (1979-), 男, 湖北武汉人, 博士, 高级工程师, 研究方向为图像处理与模式识别。E-mail: xuwangming@wust.edu.cn.

基金项目: 国家自然科学基金 (61775172); 湖北省教育厅科研计划资助项目 (D20191104); 教育部产学研合作协同育人项目 (201902303039)。

主要通过提取人工特征并结合分类器实现,如:刘峰等<sup>[1]</sup>提出了一种多特征级联方法,利用感兴趣区域长宽比特征和头部 Haar 特征组成初级分类器,再通过 HOG (Histogram of Oriented Gradient) 与 SVM (Support Vector Machine) 结合的方式得到最终检测结果;Cai 等<sup>[2]</sup>提出一种对显著图提取局部强度差异直方图特征并使用 SVM 分类器得到行人检测结果;还有些方法通过提取感兴趣区域或热点图,将其转化为特征向量送入分类器进行判别<sup>[3-5]</sup>。这类传统方法由于依赖于特征设计,存在鲁棒性不强、准确率较低等缺点。随着深度学习的发展,使用深度卷积神经网络 (Deep Convolutional Neural Network, DCNN) 来解决行人检测问题成为当前主流方法。DCNN 能够自学习更可靠、表达能力更强的图像特征,使得行人检测方法泛化能力更强、检测精度更高,如李慕锴等<sup>[6]</sup>通过将 YOLOv3 与 SENet 两种深度网络进行结合,提高了检测精度。

不过,因夜晚环境与人体温度相差较大,包含行人的热红外图像会呈现较明显的人体边缘,有利于算法检测,而在温差较小的白天人体目标成像不够明显,会导致算法检测效果差。Liu 等<sup>[7]</sup>和 Wagner 等<sup>[8]</sup>提出多光谱融合的行人检测方法,将可见光与热红外图像同时送入深度网络,利用两者优劣势互补提高了检测精度,但多光谱融合方法需要使用多个传感器,成本较高,且不同光谱图像很难完全对准。Ghose 等<sup>[9]</sup>以热红外图像作为深度网络输入,结合强显著图检测来缓解温差较小时行人与背景对比度低的问题,但显著图中出现漏检时,行人会被视作背景而忽略,且该方法使用复杂的显著性检测网络,以像素级标注作为显著图标签进行训练,费时较长。此外,上述方法中均采用了较为复杂的 Faster R-CNN 或 YOLO 系列网络,在实际应用中实现实时检测需要依赖昂贵的硬件资源。

针对以上问题,本文提出一种基于弱显著图的实时热红外行人检测方法,采用弱标注方式训练显著性检测网络,将显著性检测网络和目标检测网络产生的结果进行融合从而提高行人检测准确率;同时,本文方法以轻量级单目标检测网络 LFFD<sup>[10]</sup>为基础进行改进,使之在硬件资源有限的情况下也能实时工作。

## 1 LFFD 网络简述

LFFD 是由 He 等提出的一种实用轻量级单目标

检测网络<sup>[10]</sup>,最初针对人脸检测任务设计,但能方便地扩展到行人检测、头部检测、车辆检测等任务中。作为一种无锚框 (Anchor-free) 网络, LFFD 利用感受野作为天然锚框,特征图中的每个像素点都具有相对应大小的感受野,当前像素点的感受野大小就是锚框大小。文中提出浅层网络的有效感受野要比小目标尺度大,这样可充分利用周围特征对目标检测的贡献;由于深层网络的感受野较大,适合检测大目标,同时大目标本身有足够的特征信息辅助目标判别,因此有效感受野与目标比例无需太大。基于这些观点,通过精心设计感受野大小,可将锚框大小覆盖目标检测的各个尺度,同时能在一定程度上缓解小目标检测问题。LFFD 网络结构简单,便于在各个平台上部署,且运行速度较快。

本文方法使用的基础网络就是用来解决行人检测任务的 LFFD,其结构如图 1 所示,包含由 20 个卷积层 C1~C20 组成的骨干网络 (backbone),并包含 4 个不同的输出分支 (out branch),负责检测不同尺度大小的目标,高层网络具有较大的感受野,适合用来预测较大目标,因此输出目标的大小从 out branch1 到 out branch4 依次递增。

LFFD 在速度与精度间有较好的平衡,运行速度快,但与复杂的深度学习网络相比提取的图像特征表达能力有限,影响了检测精度,因此本文方法使用两级改进的 LFFD,增强网络提取特征的表达力,在提高检测精度的同时保证实时性。

## 2 行人检测方法原理及实现

### 2.1 行人检测方法流程

本文提出的基于弱显著图的实时热红外图像行人检测方法的工作流程如图 2 所示,使用了两级改进的 LFFD 网络。为便于表述,本文将原始 LFFD 网络简称为 ORI-LFFD (Original LFFD),其输入为热红外图像,输出为目标检测结果,包括位置信息、类别信息和置信度。两级 LFFD 都是在 ORI-LFFD 基础上进行改进,前一级增加目标显著性检测功能,简称为 SD-LFFD,其输入为热红外图像,输出为目标位置信息、类别信息和置信度以及弱显著图;后一级为融合上述弱显著图信息进行进一步检测的 LFFD 网络,简称为 SF-LFFD,其输入为弱显著图与原始热红外图像,输出为目标位置信息、类别信息和置信度。

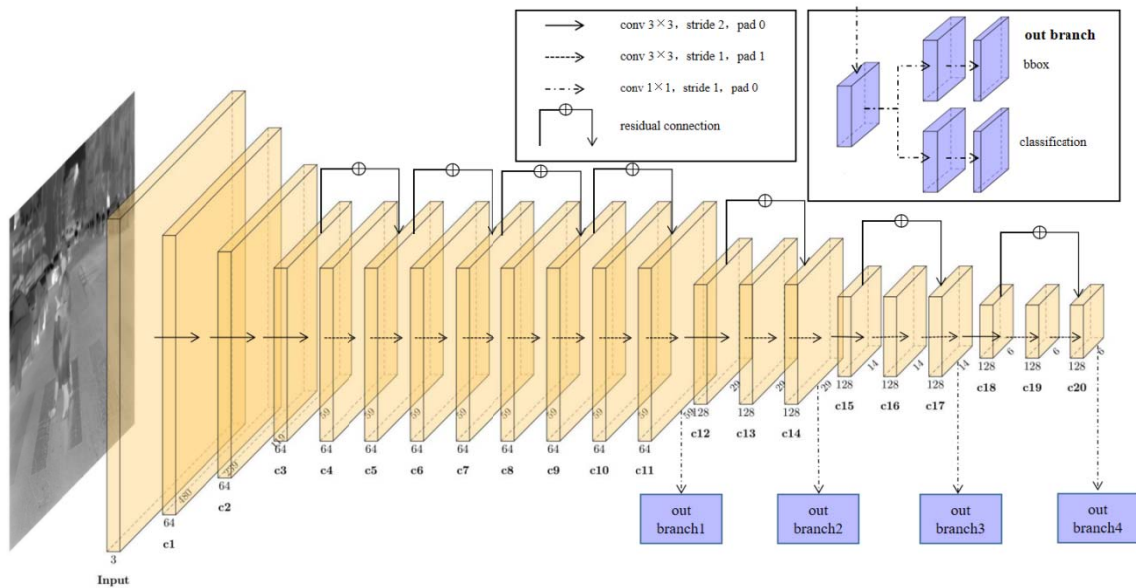


图 1 LFFD 网络结构图

Fig.1 The network structure of LFFD

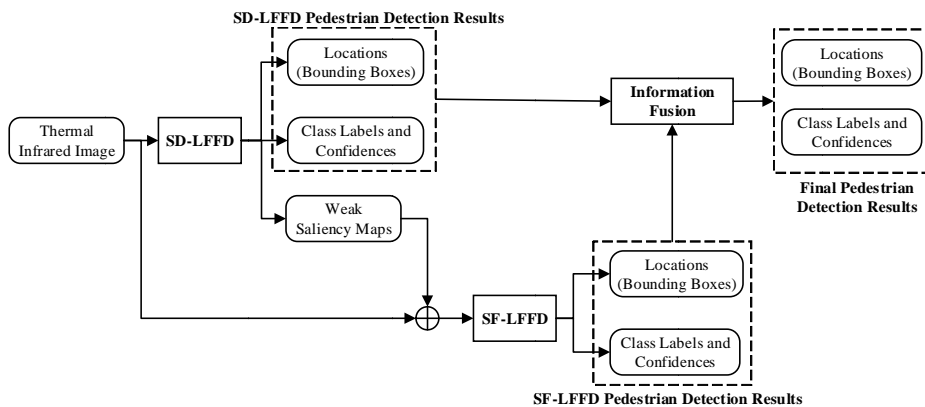


图 2 本文方法的工作流程

Fig.2 Flowchart of the proposed method

本文方法的工作流程可概括为 3 个步骤:

1) 将原始热红外图像送入 SD-LFFD 网络, 产生初步行人检测结果, 同时产生行人区域弱显著图。训练 SD-LFFD 网络时, 只需将已标记的行人区域作为显著性检测的标签, 故易于实现。

2) 将产生的弱显著图与原始热红外图像进行结合送入 SF-LFFD 中再一次进行检测, 产生新的行人检测结果。为保持 LFFD 网络输入通道数不变, 本文输入的原始热红外图像是 RGB 格式, 但 3 个通道像素值相同 (其实质为灰度图像), 因此该步骤中取其中两通道与 SD-LFFD 产生的弱显著图组成新的三通道图像, 再送入 SF-LFFD 中进行预测得到新的行人检测结果。该做法的优点在于能够“点亮”潜在在行人区域, 使神经网络能聚焦在可能的行人区域并抑制背景信息, 从而凸显行人区域。

3) 将两级改进的 LFFD 网络即 SD-LFFD 和 SF-LFFD 产生的行人检测结果融合得到最终结果, 即行人的位置信息、类别信息和置信度。使用两级改进的 LFFD 网络, 相当于将 ORI-LFFD 网络结构加深, 增强了网络信息处理和特征表达能力。

## 2.2 弱显著性检测网络 SD-LFFD

SD-LFFD 网络主要由两部分组成: ①目标检测部分, 与 ORI-LFFD 结构相同, 主要用来生成目标位置信息、类别信息和置信度; ②目标显著性检测部分, 在 ORI-LFFD 基础上进行改造, 主要用来产生弱显著图, 实现对热红外图像中行人区域的粗略增强。

为追求速度, 本文方法不使用像素级强显著图检测方法<sup>[11]</sup>, 只简单地将网络多个输出层的特征图连接后输出即可。在深度网络中, 从低层的特征层进行输出, 纹理信息丰富, 能较好地定位图像重要区域, 但

缺乏语义信息, 难以准确区分行人和背景, 会导致在热红外图像中的汽车、路灯等高亮区域也被输出为行人; 而从高层的特征层进行输出, 则正好相反。因此将尺度小、语义信息充足的高层特征和尺度大、缺乏语义信息的低层特征连接, 实现了不同尺度特征的融合, 可改善检测效果。具体做法是: 在 ORI-LFFD 的

网络结构中的 C11、C14、C17 和 C20 四个输出分支处插入卷积层和上采样层, 将得到的特征图在通道维度上连接, 经过一个  $1 \times 1$  的卷积层改变通道数, 最后经过  $\text{sigmoid}$  激活函数进行输出, 对输出的特征图利用双线性插值的方式进行缩放即可得到最终显著图。SD-LFFD 中显著性检测网络结构图如图 3 所示。

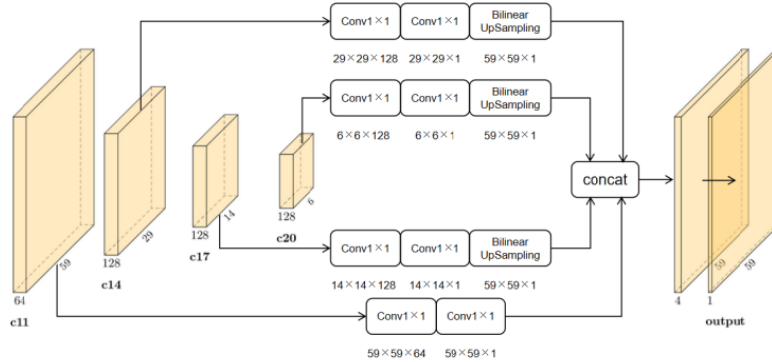


图 3 显著性检测网络结构

Fig.3 The structure of saliency detection network

SD-LFFD 的损失函数为:

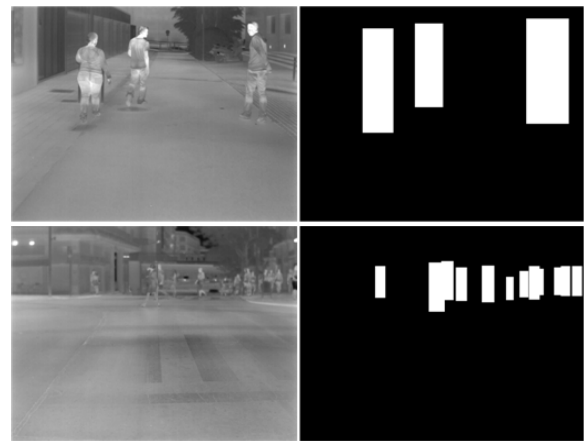
$$L = \sum_{i=0}^B \sum_{j=0}^S L_c(c_{ij}, \hat{c}_{ij}) + \sum_{i=0}^B \sum_{j=0}^S L_r(t_{ij}, \hat{t}_{ij}) + \sum_{k=0}^S L_s(p_k, \hat{p}_k) \quad (1)$$

式中:  $i$  表示第  $i$  个输出分支;  $j$  表示第  $j$  个像素点;  $S$  表示当前输出分支面积  $S=w \times h$ 。第 1 项是分类损失函数  $L_c$ , 使用交叉熵损失函数, 当第  $i$  个输出分支第  $j$  个像素点落入真实框中, 则  $c_{ij}=1$ , 否则  $c_{ij}=0$ ; 第 2 项是回归损失函数  $L_r$ , 使用  $L_2$  损失函数,  $t_{ij}$  表示的是当前像素点感受野对应的坐标位置与真实框坐标位置的相对位移; 第 3 项是显著性检测部分的损失函数  $L_s$ , 使用交叉熵损失函数,  $k$  表示第  $k$  个像素点,  $p$  表示显著图标签, 行人区域  $p_k=1$ , 背景区域  $p_k=0$ 。

训练 SD-LFFD 网络时, 显著图标注方式如图 4 所示, 其中(a)为原图, (b)为标签。由于热红外行人检测的像素级标注数据集很少, 标注过程繁琐、成本高, 且本文方法无需高精度显著性检测结果, 只需粗略的显著图便可达到突出行人潜在区域的目的, 因此本文方法通过简化显著性检测网络结构和弱标注的方式来产生弱显著图, 提高检测效率。

由于 SD-LFFD 网络中显著性检测部分与目标检测部分的输出形式和对于特征图的利用方式不同, 会导致二者对于行人区域的输出不一致。对于显著图中检测到行人区域而目标检测网络发生漏检时, 可通过显著图与热红外图像结合, 送入下一级 LFFD 网络即 SF-LFFD 再次检测, 对其进行弥补; 虽然本文方法仅使用弱显著图来突出行人, 但 SF-LFFD 的输入图像中仍包含了两通道不变的热红外图像信息, 故使用弱

显著图相当于增强了原始红外图像中的行人区域, 有利于 SF-LFFD 进行更准确地检测。对于目标检测网络检测到行人区域, 而显著图中发生漏检时, 行人作为背景被抑制导致 SF-LFFD 出现漏检, 则可通过对 SD-LFFD 与 SF-LFFD 产生的行人检测结果进行融合实现弥补。



(a) 原图 (b) 标签  
(a) Original image (b) Label

图 4 显著图标签

Fig.4 Saliency map label

### 2.3 两级 LFFD 网络的检测结果融合

本文将两级 LFFD 网络即 SD-LFFD 和 SF-LFFD 产生的行人检测结果进行融合, 实现两种方法的互补, 以得到更准确的结果。将 SD-LFFD 产生的置信度和位置信息分别表示为  $C^{\text{SD-LFFD}}$  和  $B^{\text{SD-LFFD}}$ , 将 SF-LFFD 产生的置信度和位置信息分别表示为  $C^{\text{SF-LFFD}}$  和  $B^{\text{SF-LFFD}}$ , 则最终输出的置信度和位置信息

$C_{out}$  和  $B^{out}$  为:

$$\begin{cases} C^{out} = w_C^{SD-LFFD} C^{SD-LFFD} + w_C^{SF-LFFD} C^{SF-LFFD} \\ B^{out} = w_B^{SD-LFFD} B^{SD-LFFD} + w_B^{SF-LFFD} B^{SF-LFFD} \end{cases} \quad (2)$$

确定参数  $w_C^{SD-LFFD}$ ,  $w_C^{SF-LFFD}$ ,  $w_B^{SD-LFFD}$ ,  $w_B^{SF-LFFD}$  时, 由于 SF-LFFD 的输入结合了 SD-LFFD 产生的显著图和原始热红外图像, 故 SF-LFFD 的检测结果要优于 SD-LFFD, 从而可设置  $w^{SF-LFFD} > w^{SD-LFFD}$ , 且  $w^{SF-LFFD} + w^{SD-LFFD} = 1$ 。本文基于上述原则尝试了多组值取得的效果后, 取  $w_C^{SD-LFFD} = 0.3$ ,  $w_C^{SF-LFFD} = 0.7$ ,  $w_B^{SD-LFFD} = 0.2$ ,  $w_B^{SF-LFFD} = 0.8$ 。

### 3 实验研究

#### 3.1 实验数据集及训练参数设置

为了验证本文方法的有效性, 采用 CVC-09<sup>[12]</sup>和 CVC-14<sup>[13]</sup>两个典型的热红外图像行人数据集进行网络训练和测试。训练时将白天和夜晚的训练集合并, 测试时将白天与夜晚的测试集分开, 以便比较两种不同环境下的行人检测效果。两个数据集的样本分布如表 1 所示。为提升网络的鲁棒性, 采用了随机裁剪、随机水平翻转等操作进行数据增强。

表 1 CVC-09 和 CVC-14 数据集的样本分布

Table 1 The distribution of samples in CVC-09 and CVC-14

Dataset	Day		Night	
	Train set	Test set	Train set	Test set
CVC-09	4225	2882	3201	2883
CVC-14	3695	707	3390	727

由于原 CVC-09 数据集在训练集中将骑行的人标注成行人类别, 但在测试集中却未进行标注, 为了保持统一性, 本文将 CVC-09 测试集中骑行的人重新标注为行人类别, 且所有实验都是在重新标注的数据集上进行的。

网络训练时, 本文使用 MXNet 框架, NVIDIA GTX 1080 GPU, CUDA 版本 10.0, cuDNN 版本 7.6.5, 使用 SGD 梯度下降法, 初始学习率 (learning rate)

为 0.1, 动量 (momentum) 为 0.9, 批大小 (batch size) 为 16。为避免正负样本严重不均衡, 采用正负样本比例 1:5 的方式限制负样本数量来进行反向传播, 保证训练过程更快更稳定。整个训练过程迭代 200000 次, 学习率设置为动态调整, 在第 60000 和第 120000 次迭代时将学习率降低 90%。

#### 3.2 实验结果及分析

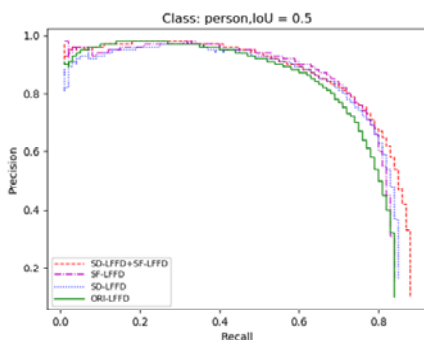
##### 1) 行人检测准确率比较

本文使用 P-R (Precision-Recall) 曲线作为评价指标。P-R 曲线刻画了精确率 (Precision) 和召回率 (Recall) 之间的关系。精确率和召回率分别定义为:

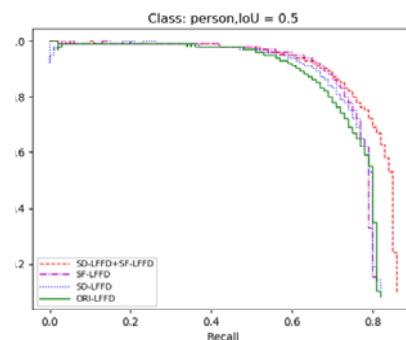
$$\text{Precision} = \frac{TP}{TP+FP}, \quad \text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

式中: TP 为正样本被正确预测的数量; FP 为负样本被预测为正样本的数量; FN 为正样本被预测为负样本的数量。将所有样本按预测为正样本的置信度从大到小排序, 根据设置的 IoU 阈值 (本文取 0.5), 计算 Precision 和 Recall 值, 并以 Recall 为横坐标、Precision 为纵坐标绘图即得到 P-R 曲线。P-R 曲线下方面积即为平均精确率 (Average Precision, AP), AP 值越高, 效果越好。

本文行人检测实验结果的 P-R 曲线如图 5 所示, 比较了 ORI-LFFD、SD-LFFD、SF-LFFD 以及本文方法 (即 SD-LFFD+SF-LFFD) 得到的检测结果, 其中图 5(a)和(b)分别为 CVC-09 数据集白天和夜晚的测试结果, 图 5(c)和(d)分别为 CVC-14 数据集白天和夜晚的测试结果。在 4 幅图中本文方法 (即 SD-LFFD+SF-LFFD) 得到的 P-R 曲线均优于其它 3 种独立网络 (即 ORI-LFFD、SD-LFFD 或 SF-LFFD), 可见本文方法对两级改进的 LFFD 网络的行人检测结果进行融合能有效提升最终的行人检测精度。表 2 列出了不同网络行人检测的 AP 值, 其中 Day、Night 和 Total 分别代表数据集中白天、夜晚和整体数据集 3 种测试场景。



(a) CVC-09 Day



(b) CVC-09 Night

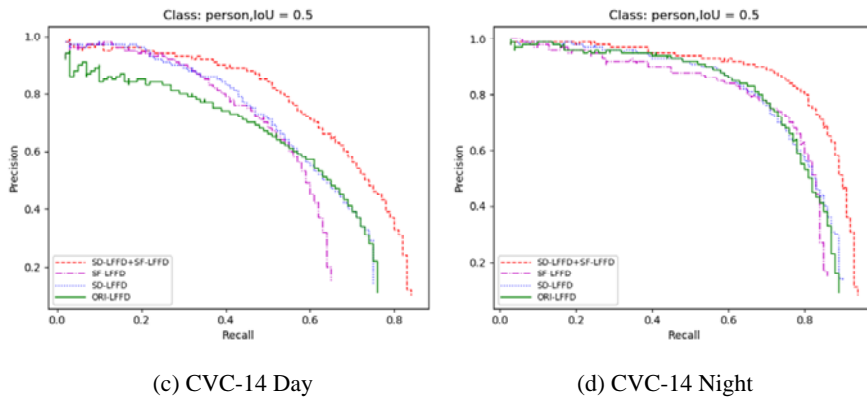


图5 两个数据集上行人检测结果的 P-R 曲线

Fig.5 The P-R curves of pedestrian detection results on two datasets

表2 行人检测 AP 值比较

Table 2 AP comparison for pedestrian detection

Dataset	TestScenario	AP(IoU=0.5)			
		ORI-LFFD	SD-LFFD	SF-LFFD	SD-LFFD+SF-LFFD
CVC-09	Day	74.15	73.25	76.05	78.46
	Night	74.70	75.54	75.81	79.85
	Total	73.82	74.01	75.52	78.74
CVC-14	Day	53.94	57.93	64.81	66.76
	Night	75.70	76.17	83.61	83.94
	Total	63.45	66.06	73.21	74.46

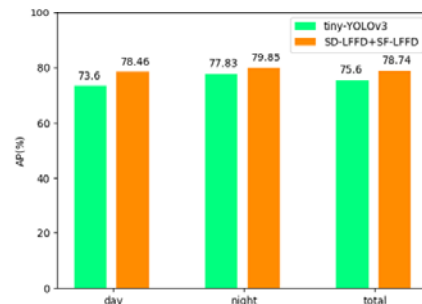
相比于原始 LFFD 网络 (即 ORI-LFFD), 本文方法 (即 SD-LFFD+SF-LFFD) 在 CVC-09 数据集上整体检测效果提升了近 5%, 在 CVC-14 数据集上提升了近 11%。由于白天人体与环境温差较夜晚要小, 白天的检测效果往往要比夜晚差, 在使用本文方法后白天和夜晚的检测精度都有提升, 且对白天的提升更为明显, 尤其在 CVC-14 数据集上提高了近 13%, 因此本文方法能够在一定程度上缓解热红外图像在白天检测效果较差的问题。此外, SF-LFFD 的 AP 值在不同的数据集和不同的测试场景中均要优于 ORI-LFFD 和 SD-LFFD, 可以体现出本文方法中使用弱显著图对于提升 SF-LFFD 网络的目标检测效果是有帮助的。

2) 与典型轻量级检测网络的对比

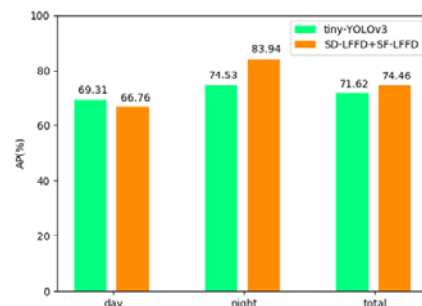
进一步在相同条件下通过实验将本文改进的轻量级网络与典型的轻量级目标检测网络 Tiny-YOLOv3<sup>[14]</sup>进行了对比, 在上述两个数据库上的行人检测 AP 值对比情况如图 6 所示。

可见, 在 CVC-09 数据集中, 对于 Day、Night

和 Total 不同的测试场景下, 本文方法 (SD-LFFD+SF-LFFD) 的 AP 值均要高于 Tiny-YOLOv3; 在 CVC-14 数据集中, 在 Day 的测试场景下, 本文方法 AP 值稍低, 但在 Night 的测试场景中本文方法的 AP 值要高出约 10%, 在 Total 的测试场景下本文方法表现更佳。由此可体现出本文方法在同等的轻量级目标检测网络中具有一定的精度优势。



(a) CVC-09



(b) CVC-14

图6 本文方法与 Tiny-YOLO v3 方法的 AP 值对比

Fig.6 AP comparison between the proposed method and Tiny-YoLov3 method

3) 行人检测速度测试及比较

测试阶段, 本文使用 MXNet 框架, NVIDIA GTX 960M 4G GPU, CUDA 版本 9.0, cuDNN 版本 7.4.1, 输入图像分辨率为 640×480。除测试本文方法外, 还

对 Tiny-YOLOv3 进行速度测试，与本文方法进行对比，结果如表 3 所示。

表 3 行人检测的速度对比

Table 3 Speed comparison for pedestrian detection

Method	Model	Frame rate	Inference
	size/M	/fps	speed/ms
Tiny-YOLOv3	33.99	18.31	54.61
SD-LFFD+SF-LFFD	14.45	31.25	32

可见，相较于 Tiny-YOLO v3，本文方法 (SD-LFFD+SF-LFFD) 用更小的模型取得了更快的速度，处理帧率约为 31 fps，说明在硬件资源有限的情况下本文方法能实时工作，这得益于使用简单易实现的目标弱显著性检测算法和改进的轻量级 LFFD 网络。

#### 4 结束语

本文提出了一种基于弱显著图的实时热红外图像行人检测方法，主要贡献在于：①设计了一种针对行人的弱显著性检测网络结构，在产生行人检测结果的同时能够产生行人的弱显著图；②将检测到的弱显著图与原始热红外图像相结合，使得深度网络能更加关注行人潜在区域，从而改善在人体与背景温差较小的白天时热红外图像行人检测效果较差的问题；③对两级改进的 LFFD 网络产生的行人检测结果进行融合，提升了算法整体的检测精度。实验结果表明，本文方法中使用弱显著图和两级改进的轻量级 LFFD 网络进行行人检测，不仅有效提升了检测准确率，而且实现了在硬件资源有限的情况下的实时检测。

#### 参考文献:

[1] 刘峰, 王思博, 王向军, 等. 多特征级联的低能见度环境红外行人检测方法[J]. *红外与激光工程*, 2018, **47**(6): 137-144.  
LIU Feng, WANG Sibao, WANG Xiangjun, et al. Infrared pedestrian detection method in low visibility environment based on multi feature association[J]. *Infrared and Laser Engineering*, 2018, **47**(6): 137-144.

[2] CAI Y, LIU Z, WANG H, et al. Saliency-based pedestrian detection in far infrared images[J]. *IEEE Access*, 2017, **5**: 5013-5019.

[3] Ko B C, Kim D Y, Nam J Y. Detecting humans using luminance saliency in thermal images[J]. *Optics Letters*, 2012, **37**(20): 4350-4352.

[4] MA Y, WU X, YU G, et al. Pedestrian detection and tracking from low-resolution unmanned aerial vehicle thermal imagery[J]. *Sensors*, 2016, **16**(4): 446.

[5] Jeon E S, Choi J S, Lee J H, et al. Human detection based on the generation of a background image by using a far-infrared light camera[J]. *Sensors*, 2015, **15**(3): 6763-6788.

[6] 李慕锴, 张涛, 崔文楠. 基于 YOLOv3 的红外行人小目标检测技术研究[J]. *红外技术*, 2020, **42**(2): 176-181.  
LI Mukai, ZHANG Tao, CUI Wennan. Research of Infrared Small Pedestrian Target Detection Based on YOLOv3[J]. *Infrared Technology*, 2020, **42**(2): 176-181.

[7] LIU J, ZHANG S, WANG S, et al. Multispectral deep neural networks for pedestrian detection[J/OL]. *arXiv preprint*, 2016, <https://arxiv.org/pdf/1611.02644.pdf>.

[8] Wagner J, Fischer V, Herman M, et al. Multispectral pedestrian detection using deep fusion convolutional neural networks[C]//*24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, 2016: 509-514.

[9] Ghose D, Desai S M, Bhattacharya S, et al. Pedestrian Detection in Thermal Images using Saliency Maps[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019: 988-997.

[10] HE Y, XU D, WU L, et al. LFFD: A Light and Fast Face Detector for Edge Devices[J/OL]. *arXiv preprint*, 2019, <https://arxiv.org/abs/1904.10633v1>.

[11] HOU Q, CHENG M M, HU X, et al. Deeply supervised salient object detection with short connections[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 3203-3212.

[12] Socarrás Y, Ramos S, Vázquez D, et al. Adapting pedestrian detection from synthetic to far infrared images[C]//*ICCV-Workshop Visual Domain Adaptation and Dataset Bias*, 2013: 1-3.

[13] González A, Fang Z, Socarras Y, et al. Pedestrian detection at day/night time with visible and fir cameras: A comparison[J]. *Sensors*, 2016, **16**(6): 820.

[14] Redmon J, Farhadi A. Yolov3: An incremental improvement[J/OL]. *arXiv preprint*, 2018, <https://arxiv.org/pdf/1804.02767.pdf>.