Computer Science and Application 计算机科学与应用, 2023, 13(6), 1199-1207
Published Online June 2023 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2023.136117

Hans汉斯

基于YOLOv5的改进密集行人检测算法

丛笑含，李士心*，陈范凯，孟　玥

天津职业技术师范大学电子工程学院，天津

收稿日期：2023年5月12日；录用日期：2023年6月9日；发布日期：2023年6月19日

摘　要

针对密集行人检测中行人之间高度遮挡重叠所带来的精度低和漏检高的问题，提出一种密集行人检测方法：Serried-YOLOv5。实验基于YOLOv5s，首先在网络特征融合阶段引入注意力机制，添加1个SE模块提高对有用信息定位的精度；然后使用Soft-NMS代替原有的NMS，保留IOU中等，但置信度较高的框，防止漏检。实验结果表明：Dense-YOLOv5相比原YOLOv5在CrowdHuman数据集上，在保证实时性的前提下，FPS提高了9.091；AP提高了1.5%；召回率Recall提高了5%，检测平均精度均值mAP$_{0.5}$提升了1.5%，证明了Serried-YOLOv5方法在密集行人检测中的有效性。

关键词

YOLOv5，密集行人检测，SE，Soft-NMS，深度学习，注意力机制

An Improved Dense Pedestrian Detection Algorithm Based on YOLOv5

Xiaohan Cong, Shixin Li*, Fankai Chen, Yue Meng

School of Electronic Engineering, Tianjin University of Technology and Education, Tianjin

Received: May 12th, 2023; accepted: Jun. 9th, 2023; published: Jun. 19th, 2023

Abstract

In order to solve the problems of low accuracy and high missed detection caused by high occlusion overlap between pedestrians in dense pedestrian detection, a dense pedestrian detection method is proposed: Serried-YOLOv5. The experiment is based on YOLOv5s. Firstly, the attention mechanism is introduced in the network feature fusion stage, and an SE module is added to improve the accuracy of locating useful information. Then replace the original NMS with Soft-NMS,

*通讯作者。

文章引用: 丛笑含, 李士心, 陈范凯, 孟玥. 基于YOLOv5的改进密集行人检测算法[J]. 计算机科学与应用, 2023, 13(6): 1199-1207. DOI: 10.12677/csa.2023.136117

and reserve the enclosure with a medium IOU but high confidence to prevent missing checks. Experimental results show that compared with original YOLOv5 in CrowdHuman dataset, Dense-YOLOv5 has an FPS increase of 9.091 under the premise of ensuring real-time performance. AP increased 1.5 percent; The Recall rate is increased by 5%, and the average accuracy of $mAP_{0.5}$ is increased by 1.5%, which proves the effectiveness of Serried-YOLOv5 method in dense pedestrian detection.

Keywords

YOLOv5, Dense Pedestrian Detection, SE, Soft-NMS, Deep Learning, Attention Mechanism

1. 引言

随着智慧城市和智能交通系统的快速发展，对行人检测的需求显著增加。行人检测是在给定的图像或视频中，准确地检测并定位行人的位置。在许多实际应用场景中，如智能监控、自动驾驶、人机交互等，行人检测都具有极为重要的作用。近几年发生了几起大型人群聚集引发的踩踏事件，这对我们的社会造成了不小的伤害和损失。为了确保公共安全并能够及时处置突发事件，我们需要在聚集场所进行实时传输画面，对人群聚集数量、密度以及人流方向等多种数据进行采集与分析。密集人群检测技术的研究和发展一直是计算机视觉领域的热点和难点问题之一，其难度在于行人在图像中的表现形式非常多样化且背景复杂、遮挡现象普遍存在。随着深度学习技术的不断发展和普及，深度学习方法在行人检测任务中表现出了优异的性能，成为当前行人检测领域主流的技术之一。在本文中，我们提出了一种新的密集行人检测方法，该方法考虑了行人彼此之间的接近性，并利用上下文信息来提高准确性。我们的方法在几个具有挑战性的数据集上具有显著优势，证明了其在现实场景中的有效性。

行人检测发展大致如下：2005 年，Viola-Jones 算法被应用于行人检测，并取得了一定的成功。但是，在人群密集的场景中，该算法的表现不佳[1]。2008 年，以人头为特征的行人检测算法被提出，该算法将人头看作行人的重要部分，并通过检测头部位置来判断行人位置。这种算法适用于密集人群场景，但是在低头、带帽等情况下会受到影响。2012 年，使用姿态模型的行人检测算法被提出，该算法通过对人体姿态的建模来判断行人位置[2]。这种算法在处理复杂情况和姿态变化较大的场景中表现优异，但是对于部分遮挡情况下的行人检测仍有瓶颈。2013 年，基于深度神经网络的行人检测算法开始发展，并取得了明显进展[3]-[8]。该算法通过大量的数据训练深度神经网络模型，实现对行人的准确识别和定位。这种算法的优势主要在于准确率高、鲁棒性强。

针对上述问题，本文对 YOLOv5s-6.0 模型进行改进。实验基于 YOLOv5s，首先在网络特征融合阶段引入注意力机制，添加 1 个 SE 模块提高对有用信息定位的精度；然后使用 Soft-NMS 代替原有的 NMS，保留 IOU 中等，但置信度较高的框，防止漏检。实验结果表明：Serried-YOLOv5 相比原 YOLOv5 在 CrowdHuman 数据集上，在保证实时性的前提下，FPS 提高了 9.091；AP 提高了 1.5%；召回率 Recall 提高了 5%，检测平均精度均值 $mAP_{0.5}$ 提升了 1.5%，证明了 Serried-YOLOv5 方法在密集行人检测中的有效性。改进后的 Serried-YOLOv5 网络在保证实时性的同时，提高了密集行人场景下网络的平均精确度并降低了漏检率。

2. YOLOv5 算法

2.1. YOLOv5 概述

YOLOv5 是 Ultralytics 团队在 2020 年提出来的 One Stage 目标检测算法，YOLOv5 采用了一种新的模型结构，将原来的模型简化了，并加强了模型的鲁棒性、速度和精度。相对于其他目标检测算法，YOLOv5 具有响应速度快、精确度高等优势，并且可以在较低的硬件条件下进行物体检测任务。YOLOv5 有 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x 总计 4 个模型中，这几个模型的结构基本一样，不同的是模型深度和模型宽度这两个参数。YOLOv5s 网络是 YOLOv5 系列中深度最小，特征图的宽度最小的网络。其他的三种都是在此基础上不断加深、加宽。用户可以根据不同的需求进行应用，极大的满足不同行业的需求。YOLOv5 包含 4 个模块：输入端、特征提取层(Backbone)、特征融合层(Neck)、检测层(Head)。YOLOv5s 结构图如图 1 所示。

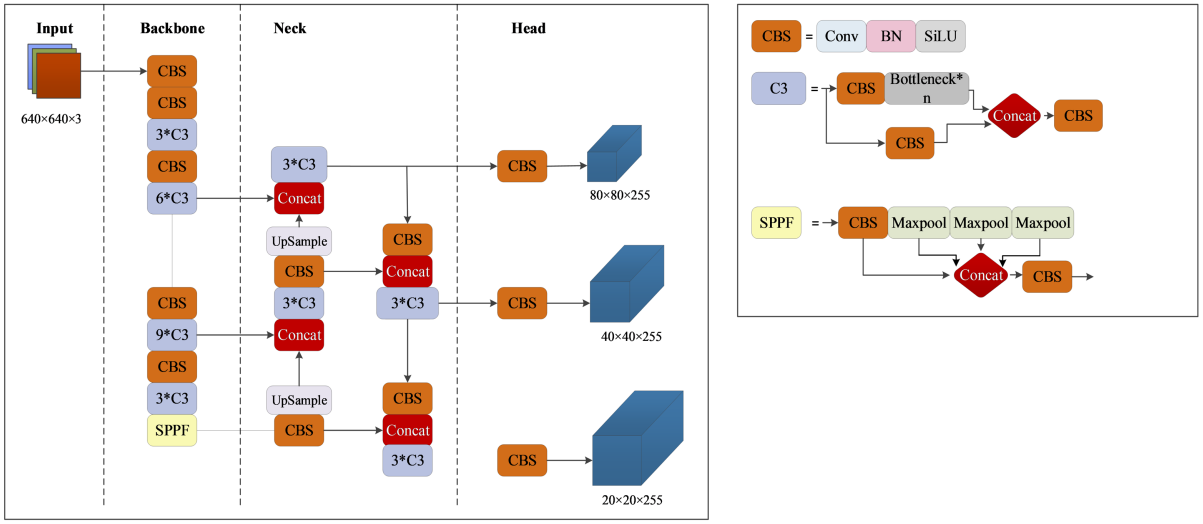

Figure 1. YOLOv5s-6.0 structure diagram
图 1. YOLOv5s-6.0 结构图

2.2. 输入端

YOLOv5 的输入端采用了 Mosaic 数据增强，该算法将多张图片按照一定比例组合成一张图片，使模型在更小的范围内识别目标。自适应锚框计算在 YOLOv3、YOLOv4 中，训练不同的数据集时，是使用单独的脚本进行初始锚框的计算，但在 YOLOv5 中，是将此功能嵌入到整个训练代码里中。所以在每次训练开始之前，它都会根据不同的数据集来自适应计算 anchor。YOLOv5 还使用了自适应图片缩放，简单的改进使得计算量大大减少，推理速度得到了 37%的提升。

2.3. 特征提取层 Backbone

主干网络能从图像中提取到特征，在 YOLOv5 中主要使用 C3 模块和 SPPF 模块。其中使用 C3 模块能减少模型计算量和提高推理速度，使用 SPPF 模块能对同一个特征图进行多尺度特征提取，有利于提升模型的精度。C3 模块其包含了 3 个标准卷积层以及多个瓶颈模块(bottleneck)，增强了算法的学习能力，并且能够在保持算法检测精度的同时实现轻量化。SPPF 模块原理和空间金字塔池化(spatial pyramid pooling, SPP)基本一致，但池化核选用不一样。SPP 在 YOLOv5 中默认使用 4 个池化核，分别是：5 × 5、

9 × 9、13 × 13 和 1 × 1。SPPF 在 YOLOv5 中默认使用两个池化核，分别是：5 × 5 和 1 × 1。SPPF 在提取特征时速度会更快。

2.4. 特征融合层 Neck

YOLOv5 的 Neck 和 YOLOv4 中一样，都采用由特征金字塔(feature pyramid network, FPN)和路径聚合结构(path aggregation network, PAN)组成。FPN 同时使用低层特征高分辨率和高层特征的语义信息，在网络中自上而下传递语义信息。PAN 是自下而上传递定位信息，使低层信息更容易传播到顶层，YOLOv5 在 YOLOv4 的基础上做了一些改进操作：YOLOv4 的 Neck 结构中，采用的都是普通的卷积操作，而 YOLOv5 的 Neck 中，采用 CSPNet 设计的 CSP2 结构，从而加强了网络特征融合能力。

2.5. 检测层 Head

检测层 Head 中，YOLOv5 采用了非极大值抑制(NMS)将交并比(intersection over union, IoU)超过设定阈值的重叠预测框丢弃。输出 20 × 20、40 × 40、80 × 80 共 3 个不同尺寸的特征图，依次检测图片中的大目标、中目标、小目标。

3. 改进 YOLOv5

改进后模型简称 Serried-YOLOv5。在原始模型的主干网络添加压缩与激励模块(squeeze and excitation, SE)，提升模型对关键特征的提取能力，从而提高目标检测的精度；然后使用 Soft-NMS 代替原有的 NMS，保留 IOU 中等，但置信度较高的框，防止漏检。

3.1. 主干网络(Backbone)改进

在自动驾驶场景下进行目标检测[9]，由于复杂的环境会使模型学习到较多背景特征，这不利于目标区域的特征学习，进而影响目标检测的精度[10] [11] [12]。原始 YOLOv5 骨干网络 3 × 3 的 Conv 模块采取的是卷积核加激活函数直接连接的设计，对于非密集场景下行人检测的特征提取往往有不错的效果，但是对于密集遮挡现象，往往很难有效地进行特征提取。为此引入压缩与激励模块(简称 SE 模块)并对其进行修改。

SE 模块主要包含压缩(squeeze)和激励(excitation)两部分。SE 模块对输入的特征信息先经过压缩操作，然后经过激励操作，最终得到模块的输出。它能使模型更加关注目标区域的通道特征，而抑制不重要的通道特征。SE 模块显式地建模特征通道之间的相互依赖关系，通过学习的方式获取每个 channel 的重要程度，然后依照这个重要程度来对各个通道上的特征进行加权，从而突出重要特征，抑制不重要的特征。简单说就是训练一组权重，对各个 channel 的特征图加权。

本质上，SE 模块是在 channel 维度上做 attention 或者 gating 操作，这种注意力机制让模型可以更加关注重要的 channel 的特征。SE 模块可以轻松的移植到其他网络架构，能够以轻微的计算性能损失带来极大的准确率提升。

本文将主干网络中的 SPPF 模块前的 C3 结构改成了一个 SE 模块。改进后的主干网络，如图 2 所示。主干网络加入的 SE 模块能使模型更加关注目标区域的通道特征，并能抑制不重要的通道特征，提升模型对关键特征的提取能力，从而提高目标检测的精度。

3.2. NMS 改进

NMS (Non-Maximum Suppression)的基本思想是：利用得分高的边框抑制得分低且重叠度(IOU)高的边框。虽然传统的 NMS 简单有效，但在某些特殊要求场景下，存在以下缺陷：简单地过滤掉得分低且重叠度高的边框可能会导致漏检问题，比如密集和拥挤情况下，容易漏检，降低召回率；NMS 阈值无法确

定，如果阈值较低，则很容易将本属于两个物体的预测框抑制掉一个，造成漏检。而阈值过高，很可能会在两个物体之间多了一个错误预测框，造成误检；将得分作为衡量标准，某些情况下，得分高的边框不一定位置更准，此衡量标准待考虑；执行速度，NMS 的实现存在较多循环判断步骤，擅长并行化处理的 GPU 执行效率不高。

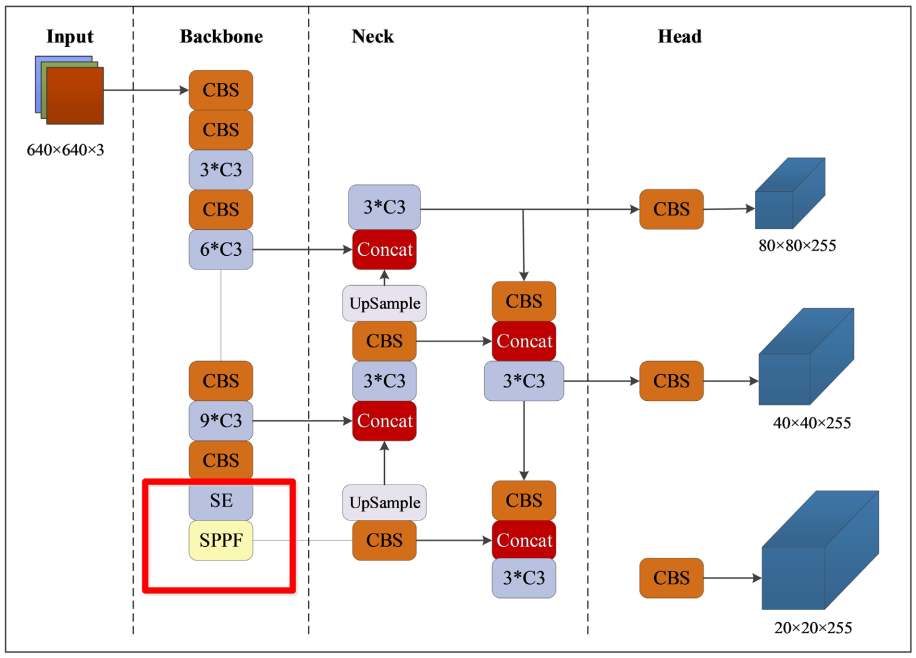

Figure 2. Improved partial backbone network structure
图 2. 改进后的部分骨干网络结构

Soft-NMS [13] [14] [15]解决方案是对 IOU 大于阈值的边框，Soft-NMS 采取得分惩罚机制，降低该边框的得分，即使用一个与 IOU 正相关的惩罚函数对得分进行惩罚。当邻居检测框 b 与当前框 M 有大的 IOU 时，它更应该被抑制，因此分数更低。而远处的框不受影响。Soft-NMS 完整算法：

Input : B= $\{b_1, .., b_N\}$,S= $\{s_1, .., s_N\}$, N_t

B is the list of initial detection boxes

S contains corresponding detection scores

N_t is the NMS threshold

begin

D←{}

while B ‡ empty do

m←argmax S

M←b_m

D←DUM;B←B-M

for b_i, in B do

if iou(M, b_i) ≥ $N_i$then

B←B- b_i ;S←S- s_i;

end NMS

$S_i \leftarrow s_i\ f(iou(M, b_i))$ Soft-NMS

end

end

return D_i S

end

于是在不改变网络结构的前提下，在 utils/general 中增加一个 Soft-NMS 模块，进行训练，原因是传统的 NMS 在出现较为密集时，本身属于两个物体的边框，其中得分较低的框就很有可能被抑制掉，从而降低了模型的召回率，Soft-NMS 加入可以减少边框的漏检、提高召回率。

4. 实验结果与分析

4.1. 实验环境

本文实验环境为：11th Gen Intel(R) Core(TM) i7-11700@2.50GHz，内存为 16 G；显卡为 NVIDIA RTX A4000；操作系统为 Windows 10，64 位；深度学习框架为 pytorch 1.11.0；编程语言为 python；Cuda11.3。

4.2. 实验数据集

CrowdHuman 数据[16]集相比于传统的行人检测数据集 Caltech [17]和 Citypersons [18]，行人更加密集和拥挤，并且所涉及场景更加广泛。其中 CrowdHuman 数据集包含 15,000 张训练图像、4370 张用于验证的图像和 5000 张用于测试的图像，其中训练集约有 340 K 人类实例，平均每张图片包含 23 个人类实例。考虑到 CrowdHuman 数据集没有开放测试集的标注，本文挑选 CrowdHuman 数据集中训练集和验证集 3000 和 500 张图片。CrowdHuman 数据集中遮挡小目标较多，可视化结果如图 3 所示。

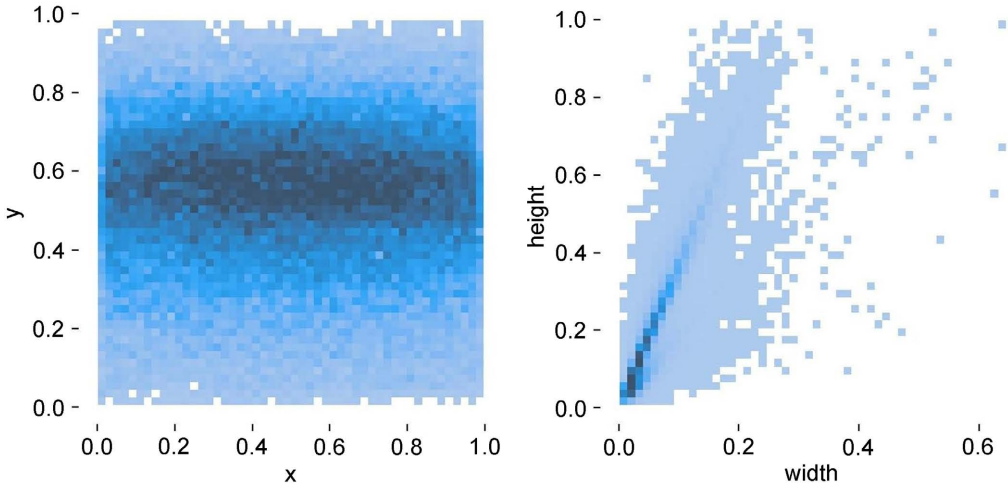

Figure 3. Dataset target visualization results
图 3. 数据集目标可视化结果

4.3. 训练模型

模型参数 depth_multiple 表示模型深度倍数，设置其值为 0.33，参数 width_multiple 表示模型宽度倍数，设置其值为 0.50。改进后的检测模型在 CrowdHuman 目标检测数据集上训练，模型具体训练参数：输入尺寸 640 * 640；批次大小为 16；训练轮数为 300 次。

4.4. 评价指标

本文实验采用 3 个指标，包括类别的平均精度(average precision, AP)；平均精度均值(mean average precision, mAP)、帧率(frames per second, FPS)和召回率(recall, R)。

计算公式如式(1)~(4)所示：

$$\mathrm{AP}=\int_0^1 P(R)\mathrm{d}R \tag{1}$$

$$\mathrm{mAP}=\frac{1}{N}\sum_{i=1}^{N}\mathrm{AP}_i \tag{2}$$

$$\mathrm{FPS}=\frac{\mathrm{FramNum}}{\mathrm{ElapsedTime}} \tag{3}$$

$$R=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{4}$$

式中：TP (true positive)代表的是预测框中预测为真实际也是真的例子，FN (false positive)代表的是预测框预测为假实际为真的例子；AP 指的是 $P(R)$ (precision-recall)曲线所围成的面积大小。FPS 指的是检测器每秒钟检测图片的个数，即检测图片数量与检测时间的比值。

4.5. 消融实验与改进实验

为了验证改进算法对 YOLOv5 各个模块改进优化效果，在 CrowdHuman 数据集上设置了 1 组消融实验。消融实验包括 2 个改进模块的对比，首先使用 Soft-NMS 代替原有的 NMS，保留 IOU 中等，但置信度较高的框，防止漏检；然后在网络特征融合阶段引入注意力机制，添加 1 个 SE 模块提高对有用信息定位的精度，其中√表示加入此模块。

Table 1. mAP ablation experimental results
表 1. mAP 消融实验结果

SE	Soft-NMS	$\mathrm{mAP}_{0.5}$/%	$\mathrm{mAP}_{0.95}$/%
		66.3	39
√		67.4	39.2
	√	67.7	39.1
√	√	67.85 (↑1.55)	38.9

由表 1 可知，在 CrowdHuman 数据集下，在加入第 1 个模块 SE 后，$\mathrm{mAP}_{0.5}$ 上升了 1.1%；Soft-NMS 替换了传统 NMS，$\mathrm{mAP}_{0.5}$ 上升 1.4%。经过 2 个模块改进后，相比于原始 YOLOv5s，$\mathrm{mAP}_{0.5}$ 上升 1.55%。虽然 $\mathrm{mAP}_{0.95}$ 下降了 0.1%，但是就密集人群重叠问题上，检测框 IOU ≥ 0.5，该检测改进有提升。

由表 2 可知，在 CrowdHuman 数据集下，在加入第 1 个模块 SE 后，AP 上升了 0.9%；Soft-NMS 替换了传统 NMS，AP 上升 1.4%。经过 2 个模块改进后，相比于原始 YOLOv5s，AP 上升 1.5%。速度 FPS 上升了 9.091，召回率 Recall 提升了 5%。但是就密集人群实时检测来讲，这符合密集场景下的实时性要求。

YOLOv5s 与改进后的 Serried-YOLOv5 算法检测效果部分对比图如图 4(a)和图 4(b)所示。白色箭头指出部分为提升部分。

Table 2. Ablation results of other detection indicators
表 2. 其他检测指标消融实验结果

SE	Soft-NMS	AP/%	Recall/%	FPS
		66.3	78	66.667
√		67.2	80	74.627
	√	67.7	81	71.942
√	√	67.8 (↑1.5)	83 (↑5)	75.758 (↑9.091)

(a) YOLOv5s (b) Serried-YOLOv5

Figure 4. Partial comparison of algorithm detection effect
图 4. 算法检测效果部分对比图

5. 结论

本文在 YOLOv5s-6.0 的基础上提出了一种改进的密集行人检测方法[19] [20]。经实验数据证明，本文提出的 Serried-YOLOv5 算法能够很好地解决密集场景下的实时行人检测任务。在原始模型的主干网络添加压缩与激励模块(squeeze and excitation, SE)，提升模型对关键特征的提取能力，从而提高目标检测的精度；然后使用 Soft-NMS 代替原有的 NMS，保留 IOU 中等，但置信度较高的框，防止漏检。此外，在研究时发现，正负样本比例不平衡引起的漏检误检问题是制约密集行人检测算法的一个很大难题。之后，将继续着眼于优化算法网络结构，提升网络精度，将尝试对正负样本比例和分配问题进一步优化，对检测层后处理阶段进一步进行提高，在保证实时的前提下，继续提高 AP 大小，降低漏检率。

参考文献

[1] 赵子露. 基于深度学习的商品图片检测算法[D]: [硕士学位论文]. 石家庄: 河北经贸大学, 2022. https://doi.org/10.27106/d.cnki.ghbju.2022.000265

[2] 肖伶. 基于机器视觉的行人检测算法研究与实现[D]: [硕士学位论文]. 贵阳: 贵州大学, 2019.

[3] Ivars, N., Roberts, K., Anatolijs, Z., Laura, L. and Artis, D. (2022) Dataset of Annotated Virtual Detection Line for Road Traffic Monitoring. *Data*, **7**, 40. https://doi.org/10.3390/data7040040

[4] Guo, Z.X., Liao, W.Z., Xiao, Y.F., Veelaert, P. and Philips, W. (2021) Weak Segmentation Supervised Deep Neural Networks for Pedestrian Detection. *Pattern Recognition*, **119**, Article ID: 108066. https://doi.org/10.1016/j.patcog.2021.108063

[5] 高宗, 李少波, 陈济楠, 李政杰. 基于 YOLO 网络的行人检测方法[J]. 计算机工程, 2018, 44(5): 215-219+226. https://doi.org/10.19678/j.issn.1000-3428.0046885

[6] 刘善良, 王士华, 史宝周, 姜鹏, 袁勇超, 李振凯, 亓昭敏. 基于 GWO-SVM 的行人跌倒检测方法[J]. 计算技术与自动化, 2023, 42(1): 84-90. https://doi.org/10.16339/j.cnki.jsjsyzdh.202301015

[7] 张超, 王亮. 基于边缘计算的行人检测算法研究[J]. 现代信息科技, 2023, 7(6): 81-84.

https://doi.org/10.19850/j.cnki.2096-4706.2023.06.021

[8] Roszyk, K., Nowicki, M.R. and Skrzypczyński, P. (2022) Adopting the YOLOv4 Architecture for Low-Latency Multispectral Pedestrian Detection in Autonomous Driving. *Sensors*, **22**, Article No. 1082. https://doi.org/10.3390/s22031082

[9] 李颀, 王娇, 邓耀辉. 基于遮挡感知的行人检测与跟踪算法[J]. 传感器与微系统, 2023, 42(4): 126-130. https://doi.org/10.13873/J.1000-9787(2023)04-0126-05

[10] 高莹. 智能交通系统中的计算机视觉技术应用[J]. 石河子科技, 2023(2): 71-72.

[11] 曹海涛, 邓小颖, 张梦, 张剑云, 贺翔, 朱金荣. 改进 YoloV5 的行人检测算法[J]. 计算机辅助工程, 2023, 32(1): 53-59. https://doi.org/10.13340/j.cae.2023.01.010

[12] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. 2018 *IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-22 June 2018, 7132-7141. https://doi.org/10.1109/CVPR.2018.00745

[13] Hendrie, M., Mushkin, H., Lombeyda, S. and Davidoff, S. (2022) JPL/Caltech ArtCenter. *Information Design Journal*, **27**, 76-84. https://doi.org/10.1075/idj.22009.hen

[14] Xie, J., Pang, Y.W., Cholakkal, H., Anwer, R., Khan, F. and Shao, L. (2021) PSC-Net: Learning Part Spatial Co-Occurrence for Occluded Pedestrian Detection. *Science China* (*Information Sciences*), **64**, 31-43. https://doi.org/10.1007/s11432-020-2969-8

[15] 李恒超, 刘香莲, 刘鹏, 冯斌. 基于多尺度感知的密集人群计数网络[J/OL]. 西南交通大学学报: 1-9. http://kns.cnki.net/kcms/detail/51.1277.U.20230316.1101.004.html, 2023-04-25.

[16] 韩晶, 王希畅, 吕学强, 张凯. SGNet: 融合多特征的密集人群计数网络[J]. 计算机工程与设计, 2022, 43(11): 3001-3007. https://doi.org/10.16208/j.issn1000-7024.2022.11.001

[17] 李华, 孔娇, 乔峥元, 白乐. 风险感知对城市景区密集人群不安全行为的影响研究[J/OL]. 安全与环境学报: 1-10. http://kns.cnki.net/kcms/detail/11.4537.X.20220617.1422.002.html, 2023-04-25.

[18] 张译. 基于深度学习的密集人群计数算法研究[D]: [硕士学位论文]. 无锡: 江南大学, 2022. https://doi.org/10.27169/d.cnki.gwqgu.2022.001512

[19] 潘昊, 刘翔, 赵静文, 张星. 一种面向密集场景的轻量化人群检测网络[J/OL]. 电子科技: 1-9. https://doi.org/10.16180/j.cnki.issn1007-7820.2023.08.006, 2023-04-25.

[20] 罗富章. 车站自动检票闸机控制系统研究[D]: [硕士学位论文]. 南昌: 东华理工大学, 2022. https://doi.org/10.27145/d.cnki.ghddc.2022.000366