



激光与光电子学进展
Laser & Optoelectronics Progress
ISSN 1006-4125, CN 31-1690/TN

《激光与光电子学进展》网络首发论文

题目: 基于 YOLOv5s 的自动扶梯乘客异常行为实时检测算法
作者: 王源鹏, 万海斌, 黄凯, 迟兆展, 张金旗, 黄智星
收稿日期: 2023-05-30
网络首发日期: 2023-08-23
引用格式: 王源鹏, 万海斌, 黄凯, 迟兆展, 张金旗, 黄智星. 基于 YOLOv5s 的自动扶梯乘客异常行为实时检测算法[J/OL]. 激光与光电子学进展.
<https://link.cnki.net/urlid/31.1690.TN.20230821.1432.102>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

基于 YOLOv5s 的自动扶梯乘客异常行为实时检测算法

王源鹏¹，万海斌^{1*}，黄凯¹，迟兆展²，张金旗¹，黄智星¹

¹广西大学计算机与电子信息学院，广西 南宁 530004

²广西大学机械工程学院，广西，南宁，530004

摘要 乘客乘坐自动扶梯时的不规范行为是造成自动扶梯人身安全事故的重要原因。为了实时检测乘客异常行为，本文提出一种基于 YOLOv5s 的轻量化的自动扶梯乘客异常行为实时检测算法 YOLO-STE。首先在主干网络引入轻量化 ShuffleNetV2 网络，以降低主干网络的参数量和计算量；其次在骨干网络中的最后一层引入基于 Transformer 编码的 C3TR 模块，以更好地提取丰富的全局信息和融合不同尺度的特征；最后在 YOLOv5s 的特征融合网络中嵌入 SE 注意力机制，以更好地关注主要信息从而提高模型精度。本文自建数据集并进行实验，实验结果表明，相比原 YOLOv5s，改进的算法的 mAP 高出 2%，达到了 96.1%，模型大小减少了 70.8%。并且在 Jetson Nano 硬件上部署测试所得，改进后的算法前传耗时比原 YOLOv5s 模型缩短了 39.9%。通过对比所得改进后的算法能更好地实现对自动扶梯乘客异常行为的实时检测，从而可以更好地保障乘客安全乘梯。

关键词 目标检测；轻量化；YOLOv5s；ShuffleNetV2；C3TR 模块；注意力机制

中图分类号 TP391.4；X705 **文献标志码** A

Real-time Detection of Abnormal Behavior of Escalator Passengers Based on YOLOv5s

Wang Yuanpeng¹, Wan Haibin^{1*}, Huang Kai¹, Chi Zhaozhan², Zhang Jinqi¹,
Huang Zhixing¹

¹*School of Compute, Electrics and Information, Guangxi University, Nanning ,
Guangxi, 530004, China*

²*School of Mechanical Engineering, Guangxi University, Nanning , Guangxi, 530004,
China*

Abstract The irregular behavior of passengers riding escalators is an important cause of escalator personal safety accidents. In order to detect passengers' abnormal behavior in real time, this paper proposes a lightweight escalator passenger abnormal behavior real-time detection algorithm YOLO-STE based on YOLOv5s. Firstly, a lightweight ShuffleNetV2 network is introduced in the backbone network to reduce the number of parameters and computation of the backbone network; secondly, a Transformer-based Secondly, the C3TR module based on Transformer encoding is introduced in the last layer of the backbone network to better extract rich global information and fuse features at different scales; finally, the SE attention mechanism is embedded in the feature fusion network of YOLOv5s to better focus on the main information and thus improve the model accuracy. In this paper, we build our own

基金项目：国家自然科学基金(62171145)、广西大学生创新训练项目(202210593061)

通信作者：*E-mail: hbwan@gxu.edu.cn

dataset and conduct experiments, and the experimental results show that compared with the original YOLOv5s, the mAP of the improved algorithm is 2% higher, reaching 96.1%, and the model size is reduced by 70.8%. Moreover, the improved algorithm's forward propagation time is 39.9% shorter than the original YOLOv5s model when deployed and tested on Jetson Nano hardware. By comparison, the improved algorithm can better achieve real-time detection of abnormal behavior of escalator passengers, which can better ensure the safety of passengers riding the escalator.

Key words object detection; lightweight; YOLOv5s; shufflenetV2; C3TR module; attention mechanism

1 引言

目前自动扶梯广泛应用在商场、医院、地铁站和火车站等公共场所,给乘客出行带来了巨大的便捷。然而,由于乘客乘坐自动扶梯时的不规范行为,造成自动扶梯人身安全事故频发,如乘客携带超大件行李乘坐自动扶梯,不规范使用婴儿车乘坐自动扶梯等。自动扶梯人身安全事故的发生往往给乘客造成严重心理阴影和身体创伤^[1]。目前对自动扶梯上乘客的乘梯行为普遍缺乏检测,多是事故发生后才进行补救。为了从源头上防范化解安全风险,有必要对自动扶梯上的乘客异常行为进行实时检测。近年来随着深度学习技术的不断发展和应用,以及嵌入式设备性能不断增强,为减少或避免自动扶梯人身安全事故的发生提供了解决方案^[2]。

目标检测算法的发展历程可以分为传统算法和深度学习算法两个阶段^[3]。传统算法主要使用手工设计的特征和分类器进行目标检测,包括基于滑动窗口的检测算法、基于区域的检测算法等。这些算法具有较好的可解释性和计算效率,但是在复杂场景下的性能有限,鲁棒性和泛化能力较差。基于深度学习的目标检测算法凭借其优秀的检测性能成为近年来目标检测研究的主流,它使用神经网络对图像进行端到端的学习和处理^[4]。常见的深度学习目标检测算法包括两种检测算法^[5]: 1) 基于 two-stage 的检测算法,如 R-CNN^[6]、Fast R-CNN^[7]、Faster R-CNN^[8]等,此类算法分两阶段执行,先获得候选区域,后进行区域内目标位置的预测和类别的识别,检测精度一般较高,但是检测速度慢,效率低^[9]; 2) 基于 one-stage 的检测算法,如 SSD^[10]、YOLO 系列^[11-13]等,此类算法通过目标检测网络直接预测目标的定位与分类,检测速度更快,效率更高。因此本文选择 YOLOv5s 算法作为网络基础框架,对其进行轻量化改进,部署到高性能低功耗的嵌入式设备 Jetson Nano 中,实时检测乘客乘梯行为。本算法可以检测到乘客摔倒、携带大件物体、不规范推轮椅或婴儿车乘梯等异常危险行为,根

据检测结果可以及时采取相应措施避免事故发生，保障乘客的乘梯安全。实验结果表明，改进后的算法比原 YOLOv5s 具有更高的精度和更快的检测速度，并且适合在资源受限的嵌入式设备进行部署。

2 YOLOv5s 原理

YOLOv5s 是一种基于单阶段目标检测的算法^[5]，由 Ultralytics 团队在 2020 年提出，有 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x 四个版本，其中 YOLOv5s 网络是 YOLOv5 系列中深度最小、特征图的宽度最小的网络。图 1 为 YOLOv5s 的网络结构图。

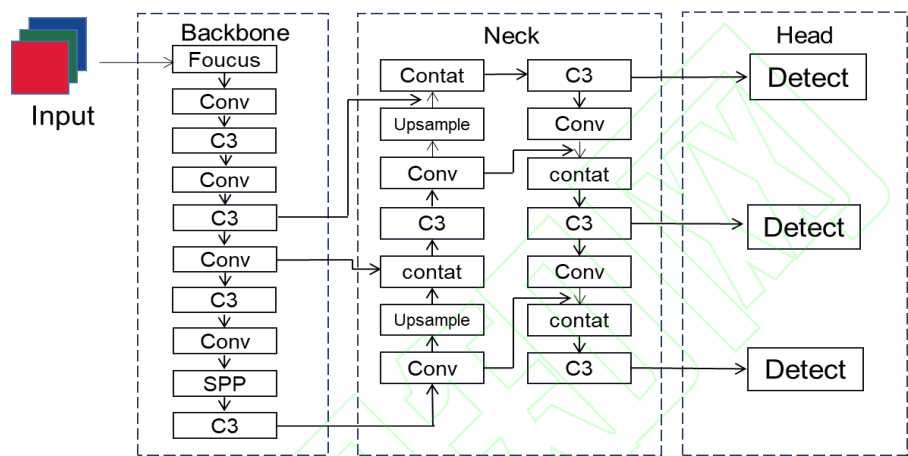


图 1 YOLOv5s 的网络结构

Fig. 1 YOLOv5s network structure

它的整个架构可以分为四个部分：输入端，Backbone，Neck，Head。下面是各个部分的作用：

输入端：主要包含 Mosaic 数据增强、自适应锚框计算和自适应图片缩放。可以提高模型的泛化能力和鲁棒性。

Backbone：它是 YOLOv5s 的主干网络，它通过卷积提取输入图像的特征。YOLOv5s 采用 CSPDarknet53 作为 Backbone，它是一种卷积神经网络，使用残差连接来减少训练时间，并采用跨阶段连接来提高特征表示能力。

Neck：用于加强 Backbone 提取的特征。YOLOv5 的 Neck 网络沿用了 FPN+PAN 的结构，FPN 自上而下把深层的语义特征传到浅层，增强多个尺度上的语义表达。PAN 自下而上把浅层的定位信息传导到深层，增强多个尺度上的定位能力。

Head：用于在输入图像中检测目标并输出它们的位置和类别。采用了 YOLOv3 的思路，分别预测目标的类别、置信度和位置信息。

3 改进的轻量化网络

3.1 基于 ShuffleNetV2 的轻量化特征提取网络

YOLOv5 的主干提取网络采用 CSPDarknet53 网络提取特征，虽然 CSPDarknet53 检测性能优秀，但网络结构复杂且参数量大，难以在算力有限的边缘设备上部署。近年来，研究人员提出了许多适合在移动设备和嵌入式设备中部署的轻量化网络，其中 ShuffleNetV2^[14]能高效地进行特征提取，有效降低模型的参数和计算量，本文基于此优势，使用 ShuffleNetV2 作为改进网络的特征提取网络。

ShuffleNetV2 继承了 ShuffleNetV1^[15] 的深度可分离卷积（Depthwise Separable Convolution）和通道混洗（Channel Shuffle），并提出了通道划分（Channel Split）。ShuffleNetV2 有两个基本单元，分别是基本单元(a)和下采样单元(b)，具体结构如图2所示。

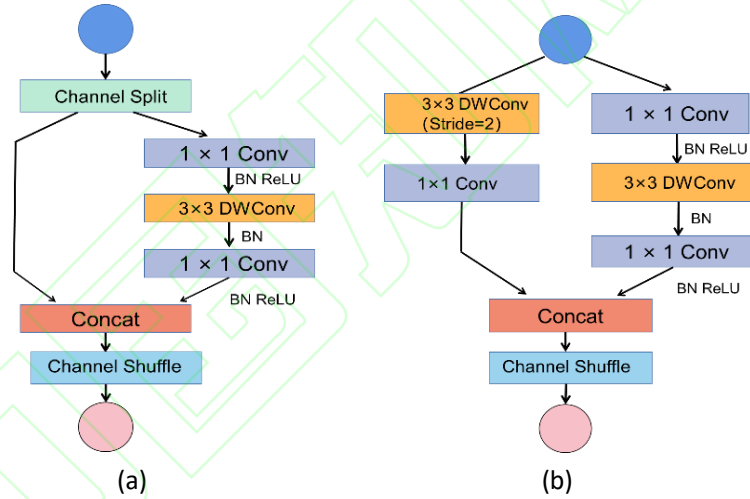


图 2 ShuffleNetV2 基本单元。(a) S_Block1;(b) S_Block2

Fig. 2 ShuffleNetV2 Unit. (a) S_Block1;(b) S_Block2

当特征图输入图 2 (a) 所示的基本单元时，首先按通道数随机且平均地拆分为左右两个分支，右分支中的 1×1 的常规卷积对通道数进行调整，让后面 3×3 的深度可分离卷积可以更好地处理特征图，与常规卷积相比，深度可分离卷积能显著减少参数量和计算量，加速网络的训练和推理速度。再经过 1×1 的常规卷积调整输出特征图的通道数，经历三次卷积后右分支特征图通道数不变，内存访问量降低。右分支输出的特征图与左分支直接向下传递的特征图进行拼接，经过通道混洗（Channel Shuffle），实现通道之间的信息交互和消息整合，提高网络的表达能力。

当特征图输入图 2 (b) 所示的下采样单元时，主要对输入的特征图进行下采样操作，将

特征图尺寸减半，同时增加网络的感受野，进一步减少网络的计算量和参数量，同时通道数会加倍，提高网络的特征提取能力和表达能力。最后经过通道混洗，使得不同通道之间的信息得到充分地交流和利用。

3.2 基于 Transformer 的 C3TR 编码模块

Transformer 结构最初被设计出来主要是为了自然语言处理(NLP)任务，基于其出色的性能，Transformer 结构在机器翻译、文本分类、语音识别，目标检测等领域也得到了广泛的应用，并表现出优异的性能^[16]。Transformer 编码器主要包括两个层归一化 (layer normalization, LN)，多头自注意力机制 (multi-head self-attention, MSA) 和多层感知机 (multi-layer perceptron, MLP)。Transformer 编码器结构如图 3 所示。

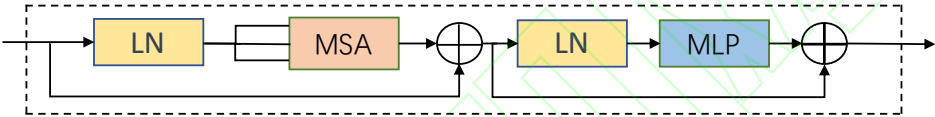


图 3 Transformer 编码器结构

Fig. 3 Transformer encoder structure

Transformer 编码器结构首先会将输入特征图的每个通道在 LN 层进行归一化处理，将图像数据限制在一定范围内，保证图像的每个通道在整体上具有相似的尺度分布，提高模型的收敛性。接着，MSA 对归一化后输入图像的不同位置进行关注并计算与其他位置的相关权重，生成相应加权向量，帮助模型捕捉序列中重要信息和依赖关系。MSA 输出的特征图与开始所输入 Transformer 编码器结构的特征图进行相加，并再次经过 LN 层进行归一化处理。MLP 层对 LN 层的输出特征图每个位置的特征向量进行非线性变换和映射，增强特征表达能力。最后，Transformer 编码器将 MLP 的输出与第一次残差连接的输出相加，并传递结果到下一结构。

在本文中，将 Transformer encoder 模块嵌入到 C3 模块当中构成 C3TR 模块，并添加到 Backbone 最后一层。与原始 C3 模块相比，C3TR 模块能够更好地学习图像的空间关系和上下文信息，提取丰富的全局信息，提高模型的检测精度。图 4 是 C3TR 模块的结构图。

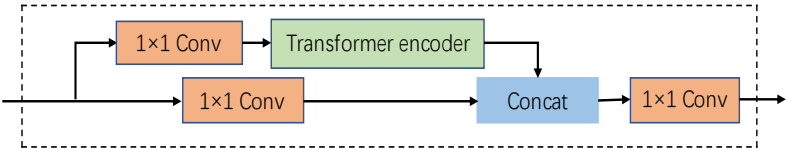


图 4 C3TR 模块结构

Fig. 4 C3TR block structure

3.3 基于 SE 通道注意力机制的特征融合网络

安装有自动扶梯的公共场所场景复杂,存在大量冗余信息。为了强化图像重要特征,让网络更多地去关注扶梯上的乘客的行为特征,抑制无用特征,本文特征融合网络在 YOLOv5s 的 FPN+PAN 结构基础上,加入 SE^[17] (Squeeze-and-Excitation) 通道注意力模块。经实验证明,在 Neck 网络加入了 SE 通道注意力模块的模型,能够有效利用通道之间的依赖关系,利用通道信息引导模型对特征进行有区分度的加权学习,可在一定程度上提高检测精度。同时 SE 模块结构相对简单,计算量较小,以略微增加的模型的复杂度和计算量换取准确率的极大提升。SE 注意力机制主要包括两个操作,即压缩 (Squeeze) 和激励 (Excitation)。如图 5 是 SE 的网络结构图。其中, \mathbf{X} 为输入特征图,其高度为 H' , 宽度为 W' , 通道数为 C' , 经过 \mathbf{F}_{tr} 卷积操作得到高度为 H , 宽度为 W , 通道数为 C 的特征图 \mathbf{U} 。对特征图 \mathbf{U} 进行 $\mathbf{F}_{sq}(\cdot)$ 压缩 (Squeeze) 操作, $\mathbf{F}_{ex}(\cdot, W)$ 激励 (Excitation) 操作和 $\mathbf{F}_{scale}(\cdot, \cdot)$ 重标定 (Scale) 操作, 得到标定特征通道权重的新特征图 $\tilde{\mathbf{X}}$ 。

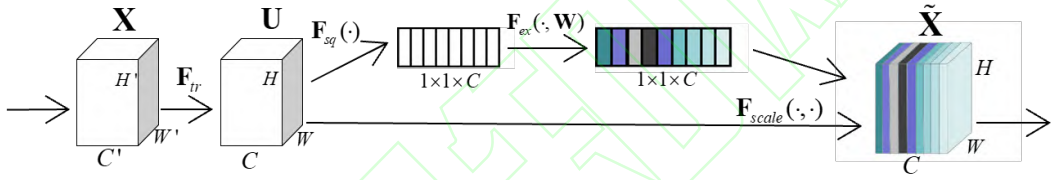


图 5 SE 注意力机制结构

Fig. 5 Structure of SE attention mechanism

$\mathbf{F}_{sq}(\cdot)$ 操作通过全局平均池化 (Global Average Pooling), 将每个通道上的空间特征编码为一个全局特征, 有效解决了由于卷积没有全局的感受野而导致难以提取通道之间的关系特征这一问题, 并增大了感受野, 原理公式如公式 (1) 所示。

$$z_c = \mathbf{F}_{sq}(\mathbf{u}_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (1)$$

式中, z_c 表示通道 c 的全局特征值, \mathbf{u}_c 为通道 c 的特征矩阵, W 为特征图宽度大小, H 为特征图高度大小, $u_c(i, j)$ 为标量, 表示通道 c 在 (i, j) 点处的特征值。

$\mathbf{F}_{ex}(\cdot, W)$ 激励操作利用压缩后的信息和通道间的信息依赖, 将上一步得到的特征图经过两个全连接层进行降维和升维, 有利于全局感知和自适应调整通道权重, 最后选择使用 Sigmoid 函数激活, 原理公式如公式 (2) 所示。

$$\mathbf{s} = \mathbf{F}_{ex}(\mathbf{z}, \mathbf{W}) = \sigma(\mathbf{g}(\mathbf{z}, \mathbf{W})) = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{z})) \quad (2)$$

式中, \mathbf{s} 为激励权重向量, \mathbf{z} 为上一步得到的全局特征向量, σ 指 Sigmoid 激活函数, \mathbf{W}_1

表示第一个全连接层的权重矩阵， \mathbf{W}_2 表示第二个全连接层的权重矩阵， δ 为 ReLU 激活函数。

$\mathbf{F}_{scale}(\cdot, \cdot)$ 重标定操作将特征图与激励权重向量相乘进行加权，赋予重要通道较大的权重，赋予不重要通道较小的权重，实现通道维度上原始特征的重标定。

$$\tilde{\mathbf{x}}_c = \mathbf{F}_{scale}(\mathbf{u}_c, \mathbf{s}_c) = \mathbf{s}_c \mathbf{u}_c \quad (3)$$

式中， $\tilde{\mathbf{x}}_c$ 表示重标定后 c 通道上的特征矩阵， $\mathbf{X} = [\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_c]$ 为输出结果的集合，即输出特征图 \mathbf{X} 。 \mathbf{u}_c 为通道 c 的特征矩阵， \mathbf{s}_c 为标量，表示通道 c 的权重大小， $\mathbf{F}_{scale}(\mathbf{u}_c, \mathbf{s}_c)$ 表示通道特征与通道权重的乘积。

3.4 改进后的总体网络结构

本文提出了一种基于 YOLOv5s 的自动扶梯乘客异常行为实时检测算法 YOLO-STE。该算法 YOLOv5s 的基础上进行了以下三方面的改进：1) 引入 ShuffleNetV2 轻量化特征提取网络；2) 在主干网络的最后一层加入基于 Transformer 编码的 C3TR 模块；3) 在特征融合网络嵌入 SE 注意力机制。改进后的 YOLO-STE 的网络结构如图 6 所示。

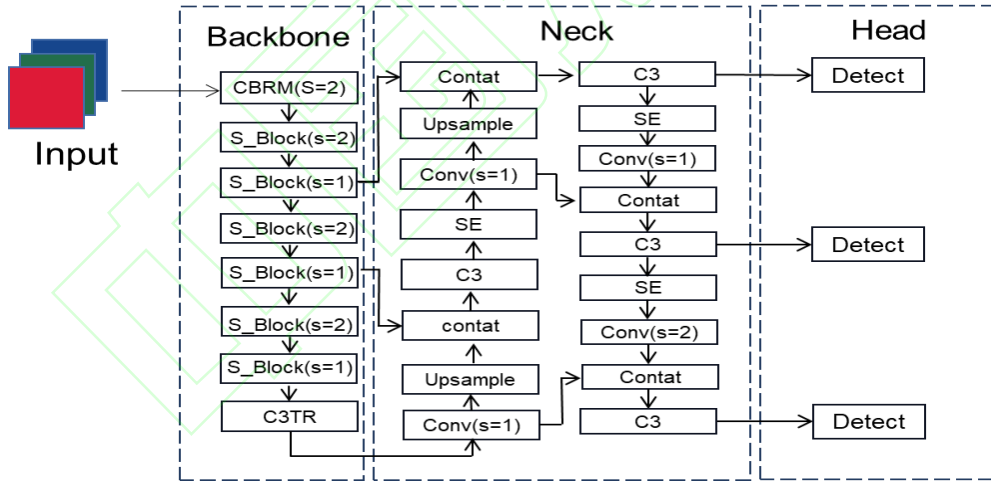


图 6 YOLO-STE 的网络结构

Fig. 6 The network structure of YOLO-STE

4 实验与结果分析

4.1 实验数据集

由于缺乏公开的关于乘客乘梯行为检测的数据集，因此本文所需要的数据集通过网络搜

集和手机拍摄来获得，包括以下五个类别：摔倒(Down)、站立(Up)、行李箱(Suitcase)、婴儿车(Stroller)、轮椅(Wheelchair)，使用 python 对图片进行随机改变亮度，随机剪切以及添加高斯噪声总共获得图片 19723 张图像，按照 8:2 将数据集划分为训练集和验测试集，其中训练集包括 15778 张图像，测试集包括 3945 张图像。训练集中的部分图像如图 7 所示。使用标注工具 LabelImg 对每幅图像进行标注，各类别标签数量如表 1 所示。



图 7 训练集中的部分图片

Fig. 7 Partial images in the training set

表 1 各类别标签数量

Table 1 The number of labels for each category unit: page			
Classes	Tarin	Test	Total
Up	4184	1046	5230
Down	3928	982	4910
Suitcase	2840	710	3550
Stroller	2586	646	3232
Wheelchair	2402	601	3003

4.2 实验环境

本实验训练环境采用 Windows 11 操作系统，硬件采用 NVIDIA RTX3060，AMD Ryzen 7 5800H with Radeon Graphics 处理器，语言为 Python3.8，加速环境为 CUDA11.6，深度学习框架为 Pytorch。

网络模型部分重要训练参数设置如下：输入图片尺寸为 640×640，训练轮次为 300，批尺寸为 16，学习率为 0.01，余弦退火超参数为 0.15，使用随机梯度下降优化器 SGD (Stochastic Gradient Descent)，优化器学习率动量为 0.937，权重衰减系数为 0.0003。

4.3 评估标准

本文对模型训练的评价指标主要从以下角度分析：精准率 (Precision, P)，召回率 (Recall, R)，精度 (Average Precision, AP)，全类平均精度值 (mean Average Precision, mAP)，模型参数量 (Parameters, Param)，模型权重大小 (Weights)，前传耗时 (FP time) 等。

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$AP = \int_0^1 PdR \quad (6)$$

$$mAP = \frac{1}{n} \sum_{i=0}^n AP_i \quad (7)$$

式中， TP 为正样本被检测正确的数量； FP 为负样本被检测为正样本的数量； FN 为背景被错误检测为正样本的数量。 FP time 为从一张图像从输入到输出检测结果所用的总时间，包括前预处理耗时（Preprocessing time）、网络前向传播耗时（Inference time）、非极大值抑制的后处理耗时（NMS time）。

4.4 消融实验

为了验证本文提出的轻量化检测算法的有效性，使用自制数据集进行训练，进行 4 组消融实验，消融实验的结果如表 2 所示。

表 2 消融实验
Table 2 Ablation experiments

Model	P(%)	R(%)	mAP_0.5(%)	Parameters/ 10^6	FLOPs(G)	Weights(MB)
YOLOv5s	93.6	89.7	94.2	7.07	16.5	13.6
ShuffleNetV2	92.4	89.3	93.1	1.55	3.7	3.75
ShuffNetV2 +C3TR	93.8	91.4	94.0	1.64	3.9	3.91
ShuffNetV2 +SE	94.2	90.2	94.8	1.62	3.8	3.85
ShuffNetV2+ C3TR+SE(YOLO-STE)	94.7	93.7	96.1	1.71	3.9	3.97

由消融实验结果可知，使用 ShuffleNetV2 替换 YOLOv5s 中的 CSPDarknet53 特征提取网络，模型的参数量下降了 78.1%，模型大小减少了 72.4%，实现特征提取网络轻量化的同时牺牲了一定的精度，mAP_0.5 下降了 1.2%。将 C3TR 模块添加到 Backbone 的最后一层，模型的 mAP_0.5 上升为 94.0%，模型大小略微增加了 4.3%。在 Neck 网络中添加 SE 注意力机制，使得模型的 mAP_0.5 提升到 94.8%。最后结合三种改进方法，在轻量化的基础上，在特征提取网络最后一层加入 C3TR 模块，在特征融合网络添加 SE 注意力机制，使得 YOLO-STE 模型的大小相比于原 YOLO5s 下降了 70.8%，参数量下降了 75.8%，mAP_0.5 提升了 2%。

消融实验验证了本文提出的改进算法不仅能提高检测精度，而且实现了模型的轻量化，加快了推理的速度，满足了实时性检测的要求。

4.5 对比实验

为了进一步测试改进模型的效果，本文所提算法 YOLO-STE 与其他主流的目标检测算法如 Fast R-CNN (ResNet50)、YOLOv3、YOLOv4，YOLOv5s 进行对比实验。对比实验的结果如表 3 所示。

表 3 与常见模型的对比实验

Table 3 Comparison experiments with common models

Models	P(%)	R(%)	mAP_0.5(%)	Parameters/10 ⁶	FLOPs(G)	Weights(MB)
Fast R-CNN	74.32	85.62	80.65	22.48	303.6	108.33
YOLOv3	89.87	76.25	88.67	61.55	155.3	234.68
YOLOv4	88.56	79.63	90.62	64.36	134.6	244.53
YOLOv5s	93.6	89.7	94.2	7.07	16.5	13.6
YOLO-STE	94.7	93.7	96.1	1.38	3.3	3.42

由对比实验结果可知，改进后的算法模型大小分别比 YOLOv3 和 YOLOv4 减少了 98.5%和 98.6%，极大程度地压缩了模型的大小，mAP_0.5 分别提高了 8.4%和 6.0%，同时实现了检测精度的提升。与原始模型 YOLOv5s 相比，模型大小下降了 74.9%，mAP_0.5 提高了 2%。图 8 展示了上述算法在测试数据集中随机选取的一张图片上的检测结果，由图可见，本文所提算法 YOLO-STE 在测试中具有最高的检测精度。总的来说，改进后的算法模型比常见的主流模型体积更小，精度更高，更适合在嵌入式设备中进行部署。

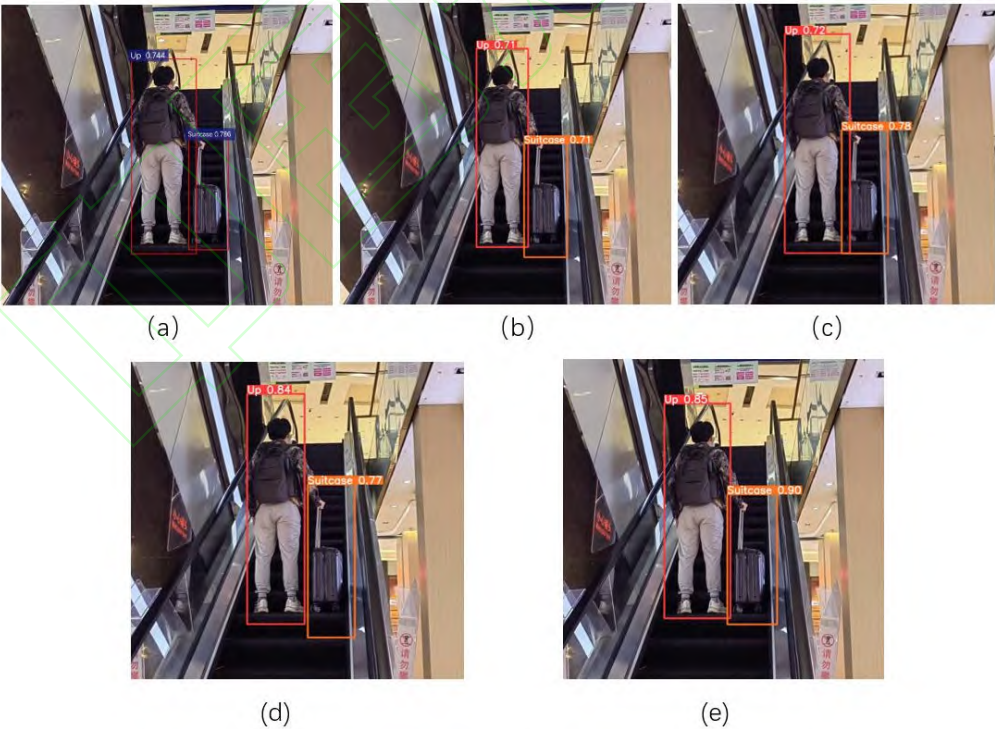


图 8 不同算法检测结果对比。(a) Fast R-CNN 检测结果；(b) YOLOv3 检测结果 (c) YOLOv4 检测结果；(d) YOLOv5 检测结果；(e) YOLO-STE 检测结果

Fig.8 Comparison of detection results of different algorithms. (a) Fast R-CNN detection results; (b) YOL

Ov3 detection results; (c)YOLOv4 detection results; (d) YOLOv5 detection results; (e) YOLO-STE detection results

5 部署实验

Jetson Nano 是由 Nvidia 推出的面向边缘计算场景的嵌入式开发者套件，可以完成图像分类，目标检测等任务，运行功率低至 5W。表 4 为 Jetson Nano 的具体配置。

表 4 Jetson Nano 的具体配置

Table 4 The specific configuration of Jetson Nano	
Hardware and Software Platform	Configuration
Operating system	Ubuntu18.04
CPU	4-core ARM® Cortex®-A57 MPCore
GPU	NVIDIA Maxwell™ with 128 NVIDIA CUDA® core
Graphic memory	4 GB 64 bit LPDDR4
CUDA	10.2
Framework	Pytorch
Programming Language	Python3.6

为了验证改进后的 YOLOv5 算法模型在嵌入式设备中能够胜任检测任务，本文改进后的模型和原 YOLOv5s 模型分别部署到边缘计算设备 Jetson Nano 上进行测试，选取前传耗时 (FP time)作为实验部署实验的评价指标，部署实验的结果如表 5 所示。

表 5 在 Jetson Nano 的对比实验

Table 5 Comparison experiments at Jetson Nano				
Model	Processing(ms)	Inference(ms)	NMS(ms)	FP(ms)
YOLOv5s	2.6	178.5	4.5	185.6
YOLO-STE	2.6	103.2	5.7	111.5

由对比实验可知，改进后的网络模型的 FP 比原模型的缩短了 39.9%，在保证检测精度的条件下，大幅提升了检测速度，实现了模型的轻量化，可以将该模型部署到边缘设备进行实时检测。实际检测结果示例图如下图 9 所示。

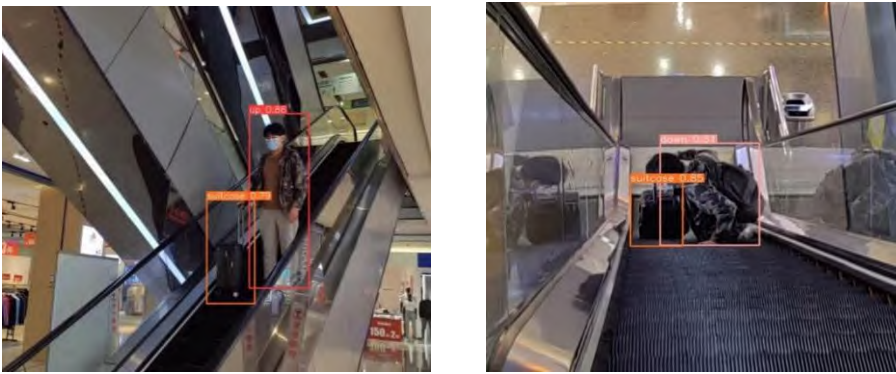


图 9 实际检测结果示例图

Fig. 9 Example figure of actual detection results

6 结 论

本文针对目前对自动扶梯上的乘客异常行为普遍缺乏检测这一问题,提出了一种基于 YOLOv5s 的轻量化自动扶梯乘客异常行为实时检测算法 YOLO-STE,该算法包括采用轻量化的 ShuffleNetV2 特征提取网络,该操作显著降低了模型的大小和复杂程度,适合在嵌入式设备中部署。在特征提取网络的最后一层加入基于 Transformer 编码的 C3TR 模块以及在特征融合网络中加入 SE 注意力机制来提升模型的精度,弥补由于轻量化带来的精度损失。实验及部署测试表明,改进后的模型平均准确率达到了 96.1%,与 YOLOv5s 模型相比,提升了 2%,参数量下降了 75.8%,模型大小减少了 70.8%,检测速度提升了 39.9%,满足实际部署中轻量化和实时性的要求,可实现对自动扶梯上的乘客异常行为进行实时检测,防范化解乘梯安全隐患,在实际生活中具有重要意义。

参考文献

- [1] 马爱萍. 自动扶梯事故频发原因分析及对策探讨[J]. 科技信息, 2013(25): 258-259. Aiping Ma. Analysis of the causes of the frequent occurrence of escalator accidents and countermeasures[J]. Technology Information, 2013(25): 258-259.
- [2] 吉训生, 滕彬. 基于深度神经网络的扶梯异常行为检测[J]. 激光与光电子学进展, 2020, 57(06): 140-149. Xunsheng Ji, Bin Teng. Detection of Abnormal Escalator Behavior Based on Deep Neural Network[J]. Laser & Optoelectronics Progress, 2020, 57(6): 061010.
- [3] Hoeser T, Kuenzer C. Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends[J]. Remote Sensing, 2020, 12(10): 1667.
- [4] 赵菲, 邓英捷. 融合多异构滤波器的轻型弱小目标检测网络[J]. 光学学报, 2023, 43(9): 0915001. Fei Zhao, Yingjie Deng. Light Dim Small Target Detection Network with Multi-Heterogeneous Filters[J]. Acta Optica Sinica, 2023, 43(9): 0915001.
- [5] 罗安能, 万海斌, 司志巍, et al. 基于改进 YOLOv5s 的可回收垃圾的检测算法[J]. 激光与光电子学进展, 2023, 60(10): 1010010. Anneng Luo, Haibin Wan, Zhiwei Si, Tuanfa Qin. Detection Algorithm of Recyclable Garbage Based on Improved YOLOv5s[J]. Laser & Optoelectronics Progress, 2023, 60(10): 1010010.
- [6] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014: 580-587.
- [7] Girshick R, Ieee. Fast R-CNN[C]. IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [8] 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.

- [9] 张寅, 朱桂熠, 施天俊, et al. 基于特征融合与注意力的遥感图像小目标检测[J]. 光学学报, 2022, 42(24): 2415001. Yin Zhang, Guiyi Zhu, Tianjun Shi, Kun Zhang, Junhua Yan. Small Object Detection in Remote Sensing Images Based on Feature Fusion and Attention[J]. Acta Optica Sinica, 2022, 42(24): 2415001.
- [10] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[C]. 14th European Conference on Computer Vision (ECCV), 2016: 21-37.
- [11] 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), 2016: 779-788.
- [12] Redmon J, Farhadi A, leee. YOLO9000: Better, Faster, Stronger[C]. 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6517-6525.
- [13] Ma N N, Zhang X Y, Zheng H T, et al. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design[C]. 15th European Conference on Computer Vision (ECCV), 2018: 122-138.
- [14] Zhang X, Zhou X Y, Lin M X, et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices[C]. 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 6848-6856.
- [15] Am A, Id O, B B, et al. - Vision Transformers in Image Restoration: A Survey[J]. - Sensors (Basel). 2023 Feb 21;23(5):2385.
- [16] J H, L S, S A, et al. - Squeeze-and-Excitation Networks[J]. - IEEE Trans Pattern Anal Mach Intell. 2020 Aug;42(8):2011-2023. doi:(- 1939-3539 (Electronic)): - 2011-2023.

网络首发:

标题: 基于YOLOv5s的自动扶梯乘客异常行为实时检测算法

作者: 王源鹏, 万海斌, 黄凯, 迟兆展, 张金旗, 黄智星

收稿日期: 2023-05-30

录用日期: 2023-07-24

DOI: 10.3788/LOP231408

引用格式:

王源鹏, 万海斌, 黄凯, 迟兆展, 张金旗, 黄智星. 基于YOLOv5s的自动扶梯乘客异常行为实时检测算法[J]. 激光与光电子学进展, 2024, 61(08): 0812004.

网络首发文章内容与正式出版的有细微差别, 请以正式出版文件为准!

您感兴趣的其他相关论文:

大口径半环形高次非球面检测与轻量化分析

崔金龙 胡明勇 毕亚超 封志伟 白茜 陈光宇 徐剑锋
合肥工业大学仪器科学与光电工程学院, 安徽 合肥 230009
激光与光电子学进展, 2023, 60(7): 0722003

基于改进DeepLabV3+的遥感影像语义分割方法

苏志鹏 李景文 姜建武 陆妍玲 朱明
桂林理工大学测绘地理信息学院, 广西 桂林 541004
激光与光电子学进展, 2023, 60(6): 0628003

基于改进YOLOv4的遥感图像目标检测方法

肖振久 杨玥莹 孔祥旭
辽宁工程技术大学软件学院, 辽宁 葫芦岛 125105
激光与光电子学进展, 2023, 60(6): 0628009

自动色阶与双向特征融合的水下目标检测算法

杨婷 高武奇 王鹏 李晓艳 吕志刚 邱若海
西安工业大学兵器科学与技术学院, 陕西 西安 710021
激光与光电子学进展, 2023, 60(6): 0610012

改进YOLOv5s算法的地铁场景行人目标检测

张秀再 邱野 张晨
南京信息工程大学电子与信息工程学院, 江苏 南京 210044
激光与光电子学进展, 2023, 60(6): 0610013