激光与光电子学进展 先进成像

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 为了实时检测乘客的异常行为，提出一种基于YOLOv5s算法的轻量化自动扶梯乘客异常行为实时检测算法YOLO-STE。首先在主干网络中引入轻量化ShuffleNetV2网络，以减少主干网络的参数量和计算量；其次在骨干网络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(Squeeze-and-excitation)注意力机制，以更好地关注主要信息，从而提高模型精度。自建数据集并进行实验，实验结果表明，相比于原YOLOv5s，改进算法的全类平均精度值(mAP)高出1.9百分点，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%。通过对比改进前后的算法，后者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乘梯安全。

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 **DOI**：10.3788/LOP231408

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

Wang Yuanpeng[1], Wan Haibin[1*], Huang Kai[1], Chi Zhaozhan[2], Zhang Jinqi[1], Huang Zhixing[1]
[1]*School of Computer, Electronics and Information, Guangxi University, Nanning* 530004, *Guangxi, China*;
[2]*School of Mechanical Engineering, Guangxi University, Nanning* 530004, *Guangxi, China*

Abstract To detect passengers' abnormal behavior in real time, we propose a lightweight escalator passenger' abnormal behavior real-time detection algorithm, YOLO-STE, based on YOLOv5s. First, a lightweight ShuffleNetV2 network was introduced in the backbone network to reduce the number of parameters and its computation. Second, a C3TR module based on Transformer encoding was introduced in the last layer of the backbone network to better extract rich global information and fuse features at different scales. Finally, an SE (Squeeze-and-excitation) attention mechanism was embedded in the feature fusion network of YOLOv5s to better focus on the main information and improve the model accuracy. We developed our dataset and conducted experiments. The experimental results demonstrate that compared with the original YOLOv5s, the mean Average Precision (mAP) of the improved algorithm is 1.9 percentage points 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 that of the original YOLOv5s model when deployed and tested on the Jetson Nano hardware. Compared with the original YOLOv5s model, 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

目前自动扶梯被广泛应用于商场、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，乘客乘坐自动扶梯时的不规范行为，使相关的人身安全事故频频发生，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。在自动扶梯上发生的人身安全事故往往会给乘客造成心理阴

收稿日期：2023-05-30；**修回日期**：2023-06-24；**录用日期**：2023-07-24；**网络首发日期**：2023-08-18
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：*hbwan@gxu.edu.cn

影和身体创伤[1]。目前对自动扶梯上乘客的乘梯行为普遍疏于检测，大多是事故发生后才进行补救。为了从源头防范与化解风险，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，深度学习技术得到了长足的发展与应用，伴随着嵌入式设备的性能不断增强，该技术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较高的计算效率，但是在复杂场景下的性能有限，其鲁棒性较弱且泛化能力较差。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方向，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2种[5]：1）基于two-stage的检测算法，如区域卷积神经网络[6]（R-CNN）、快速区域卷积神经网络[7]（Fast R-CNN）、更快区域卷积神经网络[8]（Faster R-CNN）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识别，检测精度通常较高，但是检测速度慢，效率低[9]；2）基于one-stage的检测算法，如单发多框架检测[10]（SSD）、基于深度学习的目标检测[11-13]（YOLO）算法系列等，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，并在高性能低功耗的嵌入式设备Jetson Nano中进行部署，用于实时检测乘客的乘梯行为。所提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算法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中进行部署。

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等4个版本，其中YOLOv5s网络是YOLOv5系列中深度最小、特征图宽度最小的网络。图1为YOLOv5s的网络结构图。图1中：C3的全称为Concentrated-comprehensive convolution，即集中综合卷积；SPP全称为Spatial pyramid pooling，即空间金字塔池化。

YOLOv5s的整个架构可以分为4个部分：Input，Backbone，Neck，Head。下面是各个部分的作用。

1）Input：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。

2）Backbone：它是YOLOv5s的主干网络，通过卷积提取输入图像的特征。YOLOv5s采用CSPDarknet53作为Backbone，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。

3）Neck：用于加强Backbone提取的特征。YOLOv5的Neck网络沿用了FPN+PAN（特征金字塔网络和路径聚合网络）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，提高多个尺度上的定位能力。

4）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。

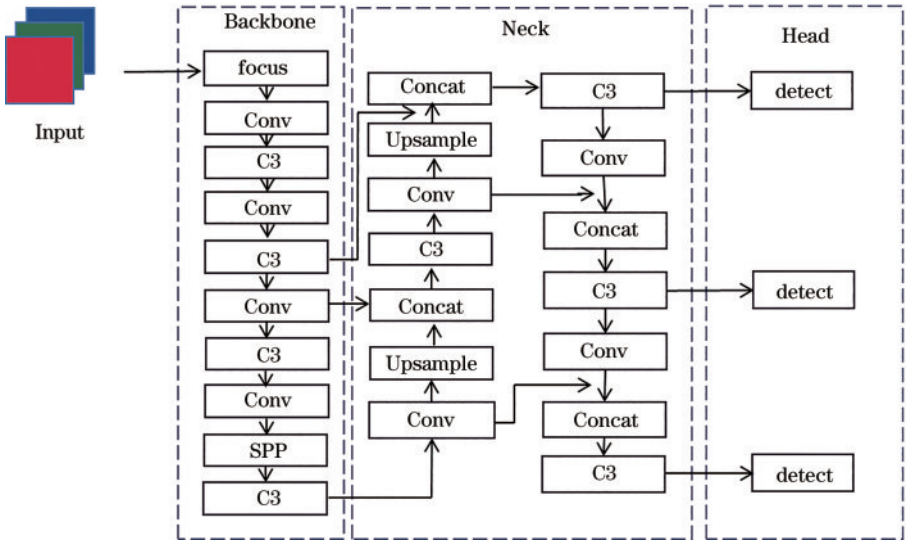

图1　YOLOv5s的网络结构

Fig. 1　YOLOv5s network structure

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 的主干采用 CSPDarknet53 网络提取特征，虽然 CSPDarknet53 网络的检测性能优秀，但该网络结构复杂且参数量大，难以在算力有限的边缘设备上进行部署。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，有效减少了模型的参数量和计算量，基于此优势，将 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。在图 2 中，DWConv 表示深度可分离卷积，Conv 表示常规卷积，BN 表示批标准化，ReLU 表示一种激活函数。

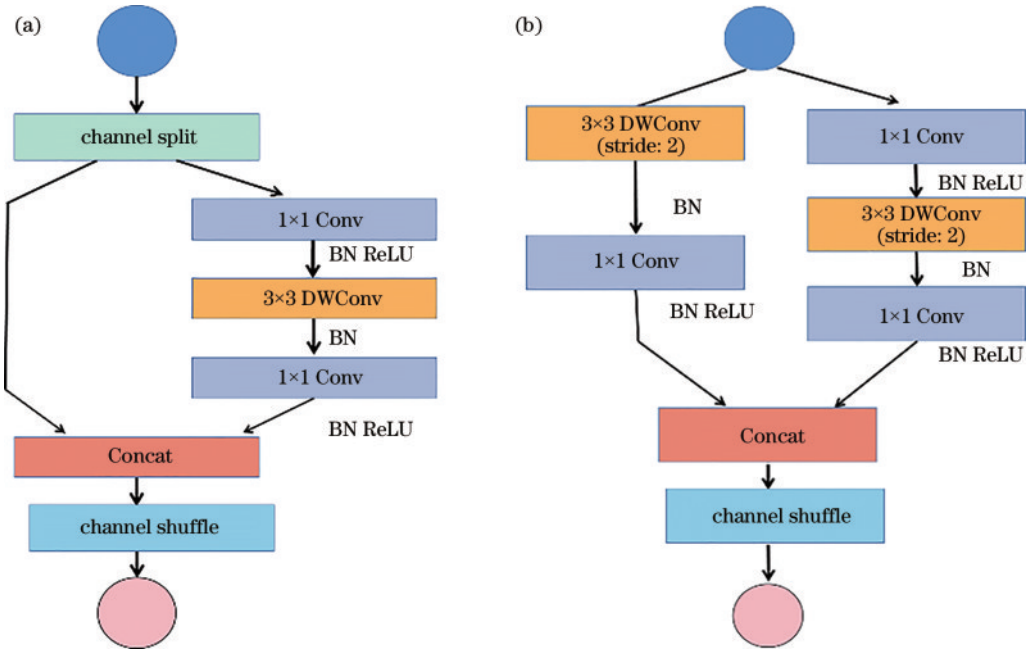

图 2　ShuffleNetV2 基本单元。（a）S_Block1；（b）S_Block2

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 2 个分支，右侧分支中的 1×1 的常规卷积先对通道数进行调整，以便于后面 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，提高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，在经历 3 次卷积后，右分支的特征图通道数不变，内存访问量减少。将右分支输出的特征图与左侧分支直接向下传递的特征图进行拼接，经过通道混洗，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的交流和利用。

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 结构也被广泛应用于机器翻译、文本分类、语音识别、目标检测等领域，并表现出优异的效果[16]。Transformer 编码器主要包括两个层归一化（LN）、多头自注意力机制（MSA）和多层感知机（MLP）。Transformer 编码器结构如图 3 所示。

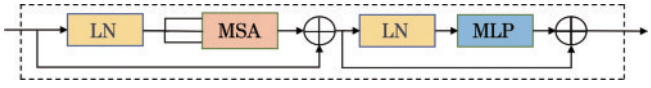

图 3　Transformer 编码器结构

Fig. 3　Transformer encoder structure

Transformer 编码器结构首先会将特征图输入到每个通道中，经过 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。其次，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，生成相应的加权向量，帮助模型捕捉序列中的重要信息和依赖关系。将 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。然后，

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，增强其特征表达能力。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。

在所提模型中，将Transformer编码器模块嵌入到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，进而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。

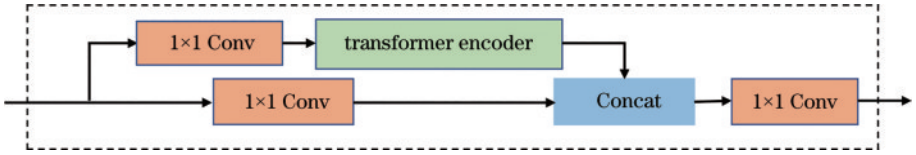

图4　C3TR模块结构

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的场景复杂，存在大量冗余信息。为了强化图像的重要特征，让所提网络更多地去关注扶梯上乘客的行为特征，抑制无用特征，所提特征融合网络在YOLOv5s的FPN＋PAN结构基础上，加入SE[16](Squeeze-and-excitation)通道注意力模块。经实验证明，在Neck网络加入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模型检测精度。同时SE模块结构相对简单，计算量较小，可以略微增加的模型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)。图5是SE的网络结构图。其中，$\boldsymbol{X}$为输入特征图，其高度为H'、宽度为W'、通道数为C'，经过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$\boldsymbol{U}$。对特征图$\boldsymbol{U}$进行$\mathrm{F}_{\mathrm{sq}}(\cdot)$压缩操作、$\mathrm{F}_{\mathrm{ex}}(\cdot,\boldsymbol{W})$激励操作和$\mathrm{F}_{\mathrm{scale}}(\cdot,\cdot)$重标定(Scale)操作，得到标定特征通道权重的新特征图$\tilde{\boldsymbol{X}}$。

$\mathrm{F}_{\mathrm{sq}}(\cdot)$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，有效解决了卷积操作因没有全局的感受野导致难以提取通道之间的关系特征这一问题，并扩大了感受野，原理公式可表示为

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

式中：z_c表示第c个通道的全局特征值；$\boldsymbol{u}_c$为第c个通道的特征矩阵；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，表示第c个通道在点(i,j)处的特征值。

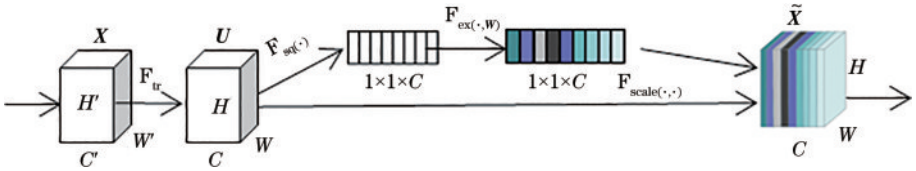

图5　SE注意力机制结构

Fig. 5　Structure of SE attention mechanism

$\mathrm{F}_{\mathrm{ex}}(\cdot,\boldsymbol{W})$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处理，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，其中$\boldsymbol{W}$为全连接层的权重矩阵。原理公式可表示为

$$\boldsymbol{s}=\mathrm{F}_{\mathrm{ex}}(\boldsymbol{z},\boldsymbol{W})=\sigma\left[\mathrm{g}(\boldsymbol{z},\boldsymbol{W})\right]=\sigma\left[\boldsymbol{W}_2\delta(\boldsymbol{W}_1\boldsymbol{z})\right],\tag{2}$$

式中：$\boldsymbol{s}$为激励权重向量；$\boldsymbol{z}$为上一步得到的全局特征向量；$\sigma(\cdot)$为Sigmoid激活函数；$\boldsymbol{W}_1$为第一个全连接层的权重矩阵；$\boldsymbol{W}_2$为第二个全连接层的权重矩阵；δ为ReLU激活函数。

$\mathrm{F}_{\mathrm{scale}}(\cdot,\cdot)$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，实现在通道维度上对原始特征的重标定。

$$\tilde{\boldsymbol{x}}_c=\mathrm{F}_{\mathrm{scale}}(\boldsymbol{u}_c,s_c)=s_c\boldsymbol{u}_c。\tag{3}$$

式中：$\tilde{\boldsymbol{x}}_c$为重标定后通道c上的特征矩阵；$\tilde{\boldsymbol{X}}=\{\tilde{\boldsymbol{x}}_1,\tilde{\boldsymbol{x}}_2,\cdots,\tilde{\boldsymbol{x}}_c\}$为输出结果的集合，即输出特征图$\tilde{\boldsymbol{X}}$；$\boldsymbol{u}_c$为通道$c$的特征矩阵；$s_c$为标量，表示通道$c$的权重大小；$\mathrm{F}_{\mathrm{scale}}(\boldsymbol{u}_c,s_c)$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

所提算法在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。

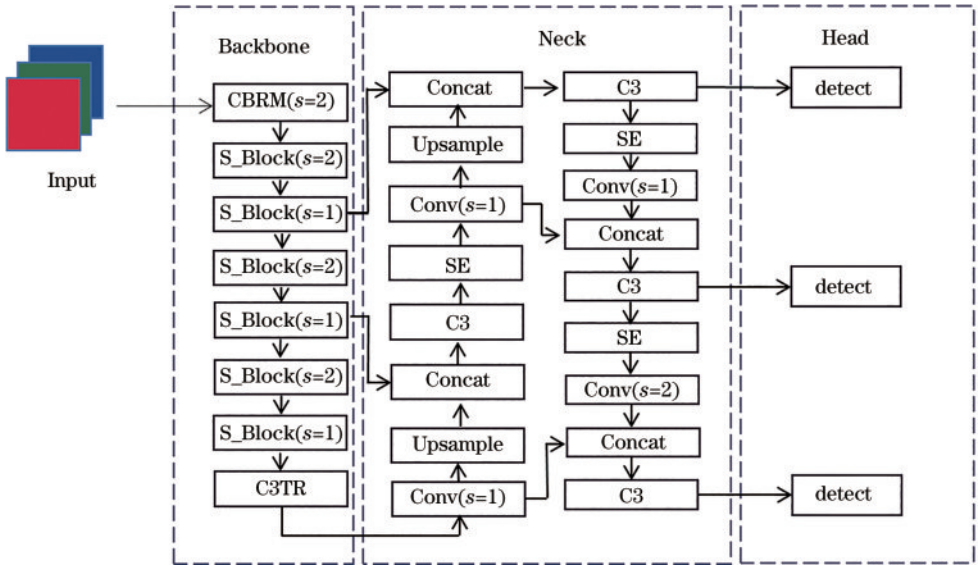

图 6　YOLO-STE 的网络结构

Fig. 6　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，因此所需要的数据集通过网络爬虫和手机拍摄来获得，包括以下 5 个类别：摔倒（Down）、站立（Up）、行李箱（Suitcase）、婴儿车（Stroller）、轮椅（Wheelchair）。使用 Python 随机改变图片亮度，随机剪切以及添加高斯噪声，总共获得 19723 幅图像，按照 8∶2 将数据集划分为训练集和测试集，其中训练集包括 15778 幅图像，测试集包括 3945 幅图像。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。

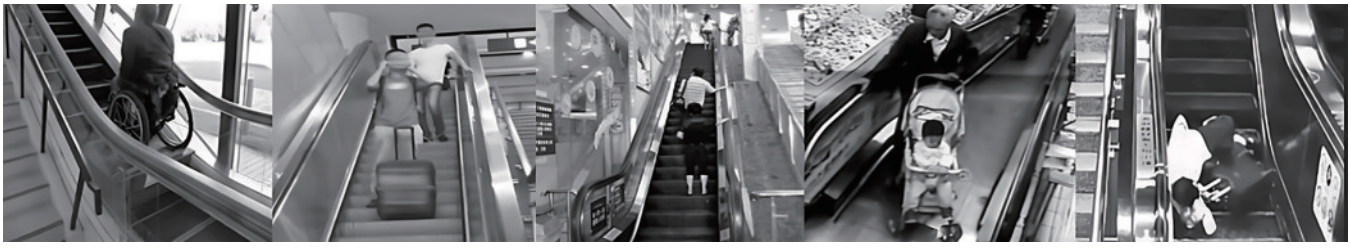

图 7　训练集中的部分图片

Fig. 7　Partial images in training set

表 1　各类别标签数量

Table 1　Number of labels for each category　unit: frame

Class	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 pixel×640 pixel，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

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（AP；P_{AP}），全类平均精度值（mAP；P_{mAP}），模型参数量（Params；N_{Params}），模型计算量（FLOPs；N_{FLOPs}），模型权重大小（Weights；w），前传耗时（FP time；t_{FP}）等。

$$P=\frac{N_{TP}}{N_{TP}+N_{FP}},\tag{4}$$

$$R=\frac{N_{TP}}{N_{TP}+N_{FN}},\tag{5}$$

$$P_{AP}=\int_0^1 P\mathrm{d}R, \tag{6}$$

$$P_{mAP}=\frac{1}{n}\sum_{i=0}^{n}P_{AP\,i}, \tag{7}$$

式中：N_{TP}为正样本被检测正确的数量；N_{FP}为负样本被检测为正样本的数量；N_{FN}为背景被错误检测为正样本的数量。t_{FP}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（Non-maximum suppression time，NMS time）。

4.4　消融实验

为了验证所提的轻量化检测算法的有效性，使用自制数据集对所提算法进行训练，进行4组消融实验，消融实验结果如表2所示。

表2　消融实验

Table 2　Ablation experiments

Model	P /%	R /%	$P_{mAP\_0.5}$ /%	N_{Params} /10^6	N_{FLOPs} /G	w /MB
YOLOv5s	93.6	89.7	94.2	7.07	16.5	13.60
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

由表2可知，使用ShuffleNetV2替换YOLOv5s中的CSPDarknet53特征提取网络，模型的参数量减少了78.1%，模型大小减少了72.4%，在实现特征提取网络轻量化的同时牺牲了一定的检测精度，mAP_0.5下降了1.1百分点。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注意力机制模块，使得模型的mAP_0.5提升到94.8%。最后结合3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注意力机制模块，使得YOLO-STE相比于原YOLO5s模型的大小减少了70.8%，参数量减少了75.8%，mAP_0.5提升了1.9百分点。

消融实验验证了所提的改进算法不仅能提高检测精度，还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，将所提算法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

Model	P /%	R /%	$P_{mAP\_0.5}$ / %	N_{Params} / 10^6	N_{FLOPs} / G	w / 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	93.60	89.70	94.20	7.07	16.5	13.60
YOLO-STE	94.7	93.70	96.10	1.71	3.9	3.97

由表3可知，改进后的算法模型大小分别比YOLOv3和YOLOv4减少了98.3%和98.4%，极大程度地压缩了模型的大小，mAP_0.5分别提高了7.43、5.48百分点，同时提升了检测精度。与原始模型YOLOv5s相比，模型大小减少了70.8%，mAP_0.5提高了1.9百分点。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8可知，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)YOLOv5s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 result; (b) YOLOv3 detection result; (c) YOLOv4 detection result; (d) YOLOv5s detection result; (e) YOLO-STE detection result

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 5 W。表 4 为 Jetson Nano 的具体配置。

表 4　Jetson Nano 的具体配置

Table 4　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否胜任检测任务，将改进后的模型和原 YOLOv5s 模型分别部署到边缘计算设备 Jetson Nano 上进行测试，选取前传耗时作为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	t_{FP} /ms
YOLOv5s	2.6	178.5	4.5	185.6
YOLO-STE	2.6	103.2	5.7	111.5

由表 5 可知，改进后的网络模型的 t_{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。

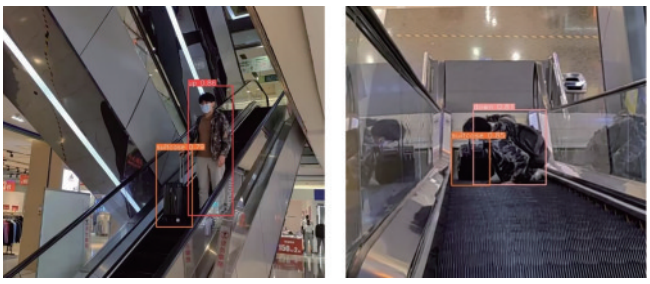

图 9　实际检测结果示例图

Fig. 9　Example figure of actual detection results

6　结　　论

针对目前对自动扶梯上的乘客异常行为普遍疏于检测这一问题，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，该算法采用轻量化的 ShuffleNetV2 作为特征提取网络，该操作显著减少了模型的大小及降低了其复杂程度，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 注意力机制模块来提升模型的精度，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 1.9 百分点，参数量减少了 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，从而防范与化解乘梯安全隐患，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.
Ma A P. Escalator frequent accident reason analysis and countermeasure discussion[J]. Science & 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(6): 061010.
Ji X S, Teng B. 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.
Zhao F, Deng Y J. 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, 等. 基于改进 YOLOv5s 的可回收垃圾检测算法[J]. 激光与光电子学进展, 2023, 60(10): 1010010.
Luo A N, Wan H B, Si Z W, et al. 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]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: ACM Press, 2014: 580-587.

[7] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-

13, 2015, Santiago, Chile. New York: IEEE Press, 2016: 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, 等. 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.
Zhang Y, Zhu G Y, Shi T J, et al. 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[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 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), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.

[12] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.

[13] Ma N N, Zhang X Y, Zheng H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design [M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11218: 122-138.

[14] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 6848-6856.

[15] Ali A M, Benjdira B, Koubaa A, et al. Vision transformers in image restoration: a survey[J]. Sensors, 2023, 23(5): 2385.

[16] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.