第 60 卷 第 6 期/2023 年 3 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

改进YOLOv5s算法的地铁场景行人目标检测

张秀再[1,2*]，邱野[1]，张晨[1]

[1]南京信息工程大学电子与信息工程学院，江苏 南京 210044；

[2]南京信息工程大学江苏省大气环境与装备技术协同创新中心，江苏 南京 210044

摘要 地铁场景行人目标存在大小不一、不同程度遮挡以及环境过暗导致目标模糊等问题，很大程度影响了行人目标检测的准确性。针对上述问题，本研究提出了一种改进YOLOv5s目标检测算法以增强地铁场景行人目标检测的效果。构建地铁场景行人数据集，标注对应标签，进行数据预处理操作。本研究在特征提取模块中加入深度残差收缩网络，将残差网络、注意力机制和软阈值化函数相结合以增强有用特征信道，削弱冗余特征信道；利用改进空洞空间金字塔池化模块，在不丢失图像信息的前提下获得多尺度、多感受野的融合特征，有效捕获图像全局上下文信息；设计了一种改进非极大值抑制算法，对目标预测框进行后处理，保留检测目标最优预测框。实验结果表明：提出的改进YOLOv5s算法能有效提高地铁场景行人目标检测的精度，尤其对小行人目标和密集行人目标的检测，效果提升更为显著。

关键词 行人目标检测；YOLOv5s；注意力机制；改进空洞空间金字塔池化

中图分类号 TP393 **文献标志码** A **DOI**：10.3788/LOP213000

Pedestrian Target Detection in Subway Scene Using Improved YOLOv5s Algorithm

Zhang Xiuzai[1,2*], Qiu Ye[1], Zhang Chen[1]

[1]*School of Electronic and Information Engineering, Nanjing University of Information Science & Technology, Nanjing* 210044, *Jiangsu*, *China*;

[2]*Jiangsu Province Atmospheric Environment and Equipment Technology Collaborative Innovation Center, Nanjing University of Information Science & Technology, Nanjing* 210044, *Jiangsu*, *China*

Abstract Pedestrian targets in subway scenes pose problems such as varying sizes, different degrees of occlusion, and blurred images caused by dark environments, which adversely affect the accuracy of pedestrian target detection. To address these problems, this study proposes an improved YOLOv5s target detection algorithm to improve the accuracy of pedestrian target detection in subway scene video signals. The pedestrian dataset of a subway scene is constructed, the corresponding labels are marked, and the data preprocessing operation is performed. Moreover, a deep residual shrinkage network is added to the feature extraction module, and the residual network, attention mechanism, and soft thresholding function are combined to enhance the useful feature channel and weaken the redundant feature channel. The fusion features of multiscale and multireceptive fields of the image are obtained using the improved atrous spatial pyramid pooling module without losing image information, and the global context information of the image is effectively captured. The improved non-maximum suppression algorithm is designed to postprocess the target prediction frame and retain the optimal prediction frame of the detection target. The experimental results demonstrate that the improved YOLOv5s algorithm proposed in this study can effectively improve the accuracy of pedestrian target detection in subway scene video signals, particularly for small and dense pedestrian target scenes.

Key words pedestrian target detection; YOLOv5s; attention mechanism; improved atrous spatial pyramid pooling

1 引 言

随着我国城市现代化建设步伐不断加快，地铁轨道交通系统逐步转向高服务性和高效性，给市民出行带来极大方便，也积极响应国家“绿色出行，低碳生活”的号召[1]。与传统交通工具不同，地铁多建设于地下，

收稿日期：2021-11-18；**修回日期**：2021-12-16；**录用日期**：2022-01-24；**网络首发日期**：2022-02-08

基金项目：国家自然科学基金青年科学基金(11504176，61601230)、江苏省高校自然科学研究项目(13KJA510001)、江苏省自然科学基金青年基金(BK20141004)

通信作者：*zxzhering@163.com

面对危险事故，人员疏散和救援相对困难，这就要求地铁内部视频监控分析系统和安检系统必须满足高实时性和高准确性。近年来，深度学习领域得到突破性发展，目标检测技术愈发成熟，将其应用于地铁安检和视频监控分析领域有着重要的现实意义。

目标检测是计算机视觉领域的核心任务之一，其目的是找出图像或视频中待检的目标，同时标注出对应的位置和类别。深度学习取得蓬勃发展后，目标检测算法主要集中在2个方向：Two stage算法，如Girshick[2]提出的区域卷积神经网络算法（Region-convolutional neural networks，R-CNN）；One stage算法，如Redmon团队[3]提出YOLO（You only look once）系列算法，Liu等[4]提出单阶段多预测框检测器（Single shot multiBox detector，SSD）算法等。Two stage算法需要进行两步检测：1）检测生成需检物体的候选框；2）对输入图片进行像素级的目标检测。One stage算法会根据网络所提取的特征直接预测目标的位置和类别，其检测速度比Two stage算法更快。尽管目标检测算法的各项性能不断优化，但将其应用到复杂地铁场景行人目标检测中，算法的漏检率和误检率仍较高。

为解决真实场景遮挡行人目标和小行人目标检测精度较低等问题，邹梓吟等[5]提出了一种融合注意力机制的遮挡行人检测算法，模型聚焦可视化行人目标，进一步提升模型检测精度。李经宇等[6]提出了一种SE-YOLOv3-SPP^{+}-PAN算法，对不同尺寸行人目标进行检测，其准确率和召回率有明显提升。Bodla等[7]通过修改目标预测框分数策略，降低预测框分数来增加预测框个数，优化了模型对遮挡目标的检测性能。董小伟等[8]为解决小行人目标检测率较低等问题，提出了一种多尺度加权特征融合网络，并对小行人目标采用过采样增强算法，检测精度得到提高。尽管这些算法能在一定程度上提高特定环境行人目标检测的精度，但对于遮挡严重的小行人目标，模型提取的特征包含大量冗余背景信息，不能聚焦行人目标区域，检测精度仍然不佳。

YOLOv5算法作为YOLO系列中最新算法，模型满足轻量化设计的思路，易于实际环境部署，同时在检测精度和速度等方面优于其他YOLO系列算法[9]。YOLOv5算法有4种模型结构，YOLOv5s网络是同系列中深度最小、特征图宽度最小的网络。因此，本研究基于YOLOv5s算法设计了改进YOLOv5s算法，并将其应用在地铁场景行人检测任务中。改进YOLOv5s算法在特征提取模块中加入深度残差收缩网络[10]，以关注有用的特征信道，聚焦可视化行人特征；通过加入改进空洞空间金字塔池化模块[11]，加强模型融合图像多尺度特征能力；模型应用改进非极大值抑制算法（SDIOU-NMS），保留目标最优预测框[12]。通过实验数据可知，改进YOLOv5s算法综合性能优于其他算法，更适合应用于地铁场景中的行人目标检测。

2 原理与方法

2.1 YOLOv5s算法

YOLOv5s算法主要由4个部分构成：输入模块（input）、主干特征提取模块（backbone）、特征加强模块（neck）、检测模块（head），YOLOv5s算法模型如图1所示。

YOLOv5s算法将输入图像划分为$N\times N$个单元格，单元格用来检测中心坐标位于网格内的目标。单元格预测P个边界框，每个边界框包含5种信息。因此，单张图片输入模型的最终预测值为一个$N\times N\times(P\times 5+C)$的张量（$C$为类别数），模型一共预测$N\times N\times P$个边界框。在利用YOLOv5s算法进行目标检测时，算法会设置置信度阈值（通常设置为0.5），首先将预测边界框置信度小于该阈值的框筛除，初步筛选后模型保留置信度相对较高的预测框；然后再利用非极大值抑制（Non-maximum suppression，NMS）算法过滤同一目标的多个预测框，保留该目标的最优预测框。YOLOv5s算法检测原理如图2所示。

2.2 改进YOLOv5s算法

本研究以YOLOv5s算法为基础，对原始模型进行优化，模型改进方法为：1）为加强模型从含噪信号中提取有用特征的能力，模型引入软阈值化函数；2）为消除背景干扰，使得模型能够聚焦行人目标区域，模型加入设计的自适应注意力模块（D-SE），选择性增强包含目标信息量最大的特征通道，并抑制无用特征；3）小行人目标和遮挡行人目标分辨率较低，包含有用特征信息不充足，将空间金字塔池化模块（Spatial pyramid pooling，SPP）与空洞卷积（Atrous convolution，AC）相结合设计出改进空洞空间金字塔池化模块（S-ASPP），通过扩大卷积核的感受野，捕获图像多尺度融合特征，加强模型对此类行人目标的检测能力；4）设计一种SDIOU-NMS，对目标预测框进行后处理，通过改变算法对预测框抑制机制，提高模型对遮挡行人目标的召回率。改进YOLOv5s模型如图3所示。

2.2.1 深度残差收缩网络

1）软阈值化函数。软阈值化函数在图像信号降噪中起着重要作用，将图像特征绝对值低于某个阈值的特征删除，同时将高于某个阈值的特征绝对值向坐标原点“收缩”。该函数的导数值只有0或1，和ReLU激活函数导数性质相同，这一性质可有效预防模型在训练过程中出现梯度弥散和梯度爆炸问题。此外，阈值的设定需符合3个条件：① 阈值必须为正值；② 阈值不能大于图像特征最大值；③ 每个特征层有独立的阈值。软阈值化函数y和导函数$\mathrm{d}y/\mathrm{d}x$可分别表示为

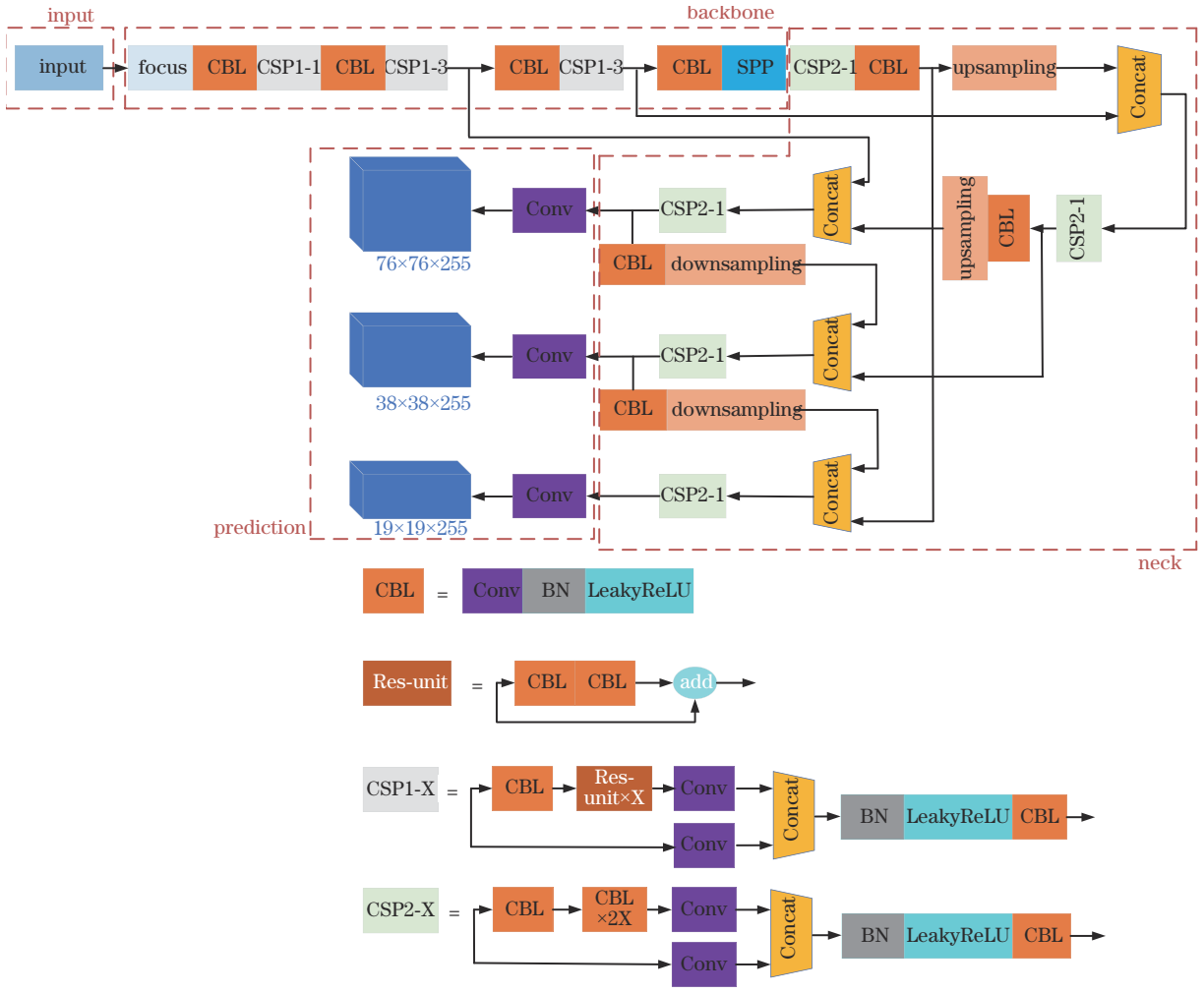

图 1　YOLOv5s 算法模型

Fig. 1　YOLOv5s algorithm model

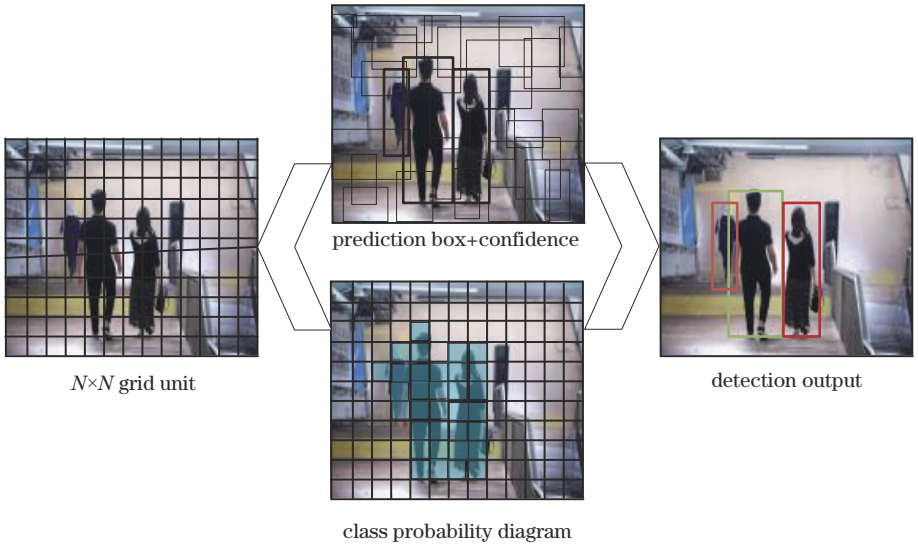

图 2　YOLOv5s 算法检测原理图

Fig. 2　Schematic diagram of YOLOv5s algorithm detection

$$y=\begin{cases}x-\tau, & x>\tau\\ 0, & -\tau\leqslant x\leqslant\tau\\ x+\tau, & x<-\tau\end{cases}, \tag{1}$$

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\begin{cases}1, & x>\tau\\ 0, & -\tau\leqslant x\leqslant\tau\\ 1, & x<-\tau\end{cases}, \tag{2}$$

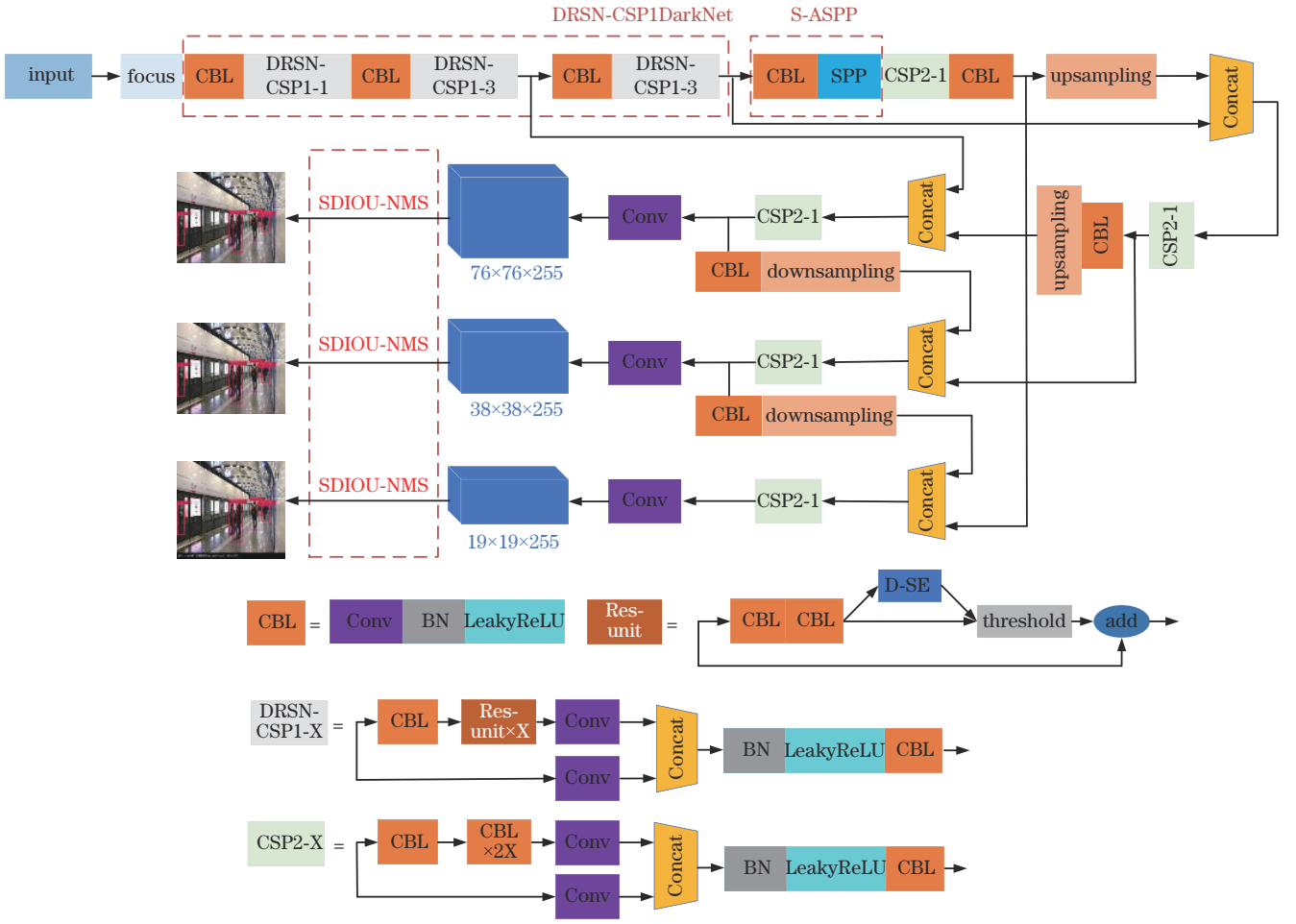

图 3　改进 YOLOv5s 算法模型

Fig. 3　Improved algorithm YOLOv5s model

式中：x 为变量值；τ 为阈值。

2）D-SE 注意力模块。卷积神经网络中的卷积操作主要是提高局部感受野，以获得多尺度的空间信息。常规卷积操作基本是对输入特征图的全部通道进行融合[13]，网络无法聚焦重要特征通道。引入 D-SE 注意力模块，特征图经过一系列卷积操作后将会预测一组权重值，然后对各个通道特征进行加权，模型能够自动提取包含目标特征信息量最大的通道。D-SE 注意力模块和结构图分别如图 4、图 5 所示。

D-SE 注意力模块主要包括 squeeze、excitation 和 scale 3 个操作，可以灵活应用到各种卷积映射。如图 4 所示，特征图 X 经过 Conv 3×3 和 Conv 5×5 并行卷积操作后，输出相加得到特征图 U，特征图 U 增加了原始输入特征图的多尺度信息，能够建立更有效的空间信息图。D-SE 注意力模块具有高度对称性，上下分支共用 squeeze 和 excitation 操作。特征图 U 通过聚合不同感受野特征信息引导通道注意力，有助于网络筛选辨别度最高的特征。

由于局部空间卷积操作的有限性，特征图 U 无法利用有效信息来获取不同通道之间的关系。squeeze

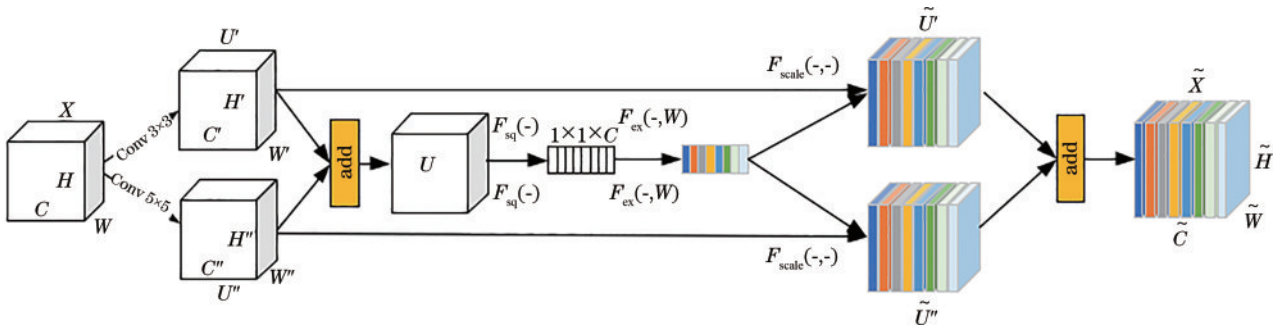

图 4　D-SE 注意力模块

Fig. 4　D-SE attention module

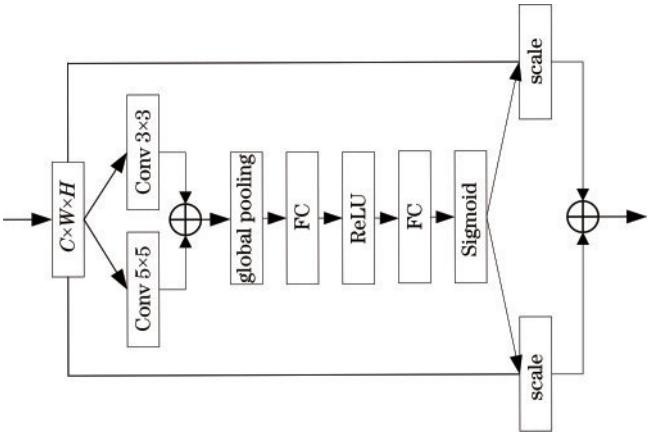

图 5　D-SE 注意力模块结构图

Fig. 5　Structure diagram of D-SE attention module

操作对特征图 U 进行全局平均池化，单通道层元素被压缩为 1×1 标量，输出形状为 1×1×C 特征图，代表特征通道上全局空间信息。excitation 操作是将 squeeze 输出的 1×1×C 特征图连接 FC-ReLU-FC-Sigmoid 层，输出 C 个范围为 0～1 之间的权值标量，经过 2 个全连接层（Fully connected layer，FC）有利于网络解析行人特征图中不同通道信息，进而强化模型对可视化行人特征的学习能力。scale 操作将权值标量与特征图 $\tilde{U}'$ 和 $\tilde{U}''$ 对应每个通道逐元素相乘后，将上下分支输出特征进行逐元素相加得到输出特征图 $\tilde{X}$。加入 D-SE 注意力模块，当输入不同尺度特征图时，模型可以自适应调节行人目标接受域大小，以突出特征图中与行人目标相关的区域，提高模型的检测精度。

深度残差收缩网络是残差网络（Residual network，ResNet）的一种改进[14]，由残差模块、D-SE 注意力机制和软阈值化函数结合构成，其结构如图 6 所示。基本原理是将 D-SE 注意力机制和 ResNet 相结合，在模型训练学习中自动设置软阈值函数的阈值，对不重要的特征通过软阈值函数将其置为零，提高网络模型提取有用特征的能力。

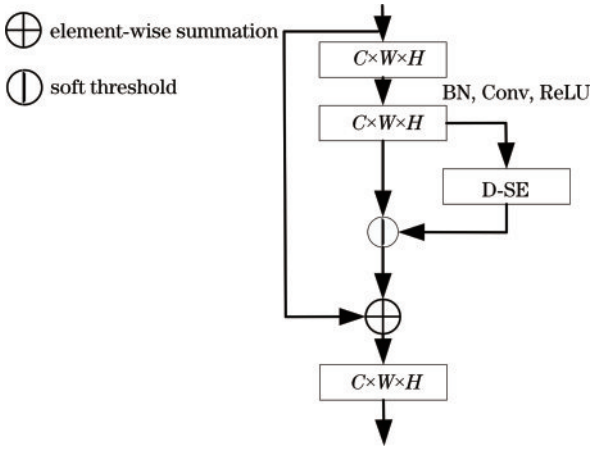

图 6　深度残差收缩网络结构图

Fig. 6　Structure diagram of depth residual shrinkage network

图像噪声会使图像模糊，影响系统对图像信息的判断。在地铁场景中视频监控系统和安检系统采集图像分析过程中会受到工作环境、电子元器件等影响，引入各种噪声，对密集目标和小目标的检测产生更为严重的影响。深度残差收缩网络使模型“注意”重要行人目标特征信息，同时将不重要的特征（相当于噪声信号）通过软阈值函数置零，网络可以拟合相对重要的语义信息，从而提高地铁场景中行人目标检测的准确性。

2.2.2　改进空洞空间金字塔池化模块

常规的卷积操作扩大感受野主要有增大卷积核大小和增加网络层数，然而，这两种方式可能会造成特征空间信息丢失、计算量过多、网络模型复杂难以训练等问题[15]。空洞卷积能在保持图像特征信息不丢失且不增加参数量的前提下，来改变感受野大小。例如，一个扩张率为 2，卷积核为 3×3 的空洞卷积，本层的感受野为 5×5，上一层的感受野为 7×7。本层感受野 η 和上层感受野 F_r 分别为

$$\eta=(k-1)\times r+1, \tag{3}$$

$$F_r=(2^{r+1}-1)\times(2^{r+1}-1), \tag{4}$$

式中：r 为扩张率；k 为卷积核大小。当 $r=1$ 时，该卷积为常规卷积。

为加强网络对可视化行人目标特征的学习，同时消除背景干扰，本研究提出一种 S-ASPP 模块对特征进行融合[16]，引用空洞卷积一方面通过利用不同感受野卷积操作提取图像特征；另一方面不改变图像分辨率仍能精确定位目标。S-ASPP 结构如图 7 所示。

S-ASPP 模块将输入特征图通道数 C 降低为原来的 1/4，减小网络模型计算量，然后再利用空间金字塔结构对输入特征图采样。本研究采用 4 种 1、3、5、7 扩张率的空洞卷积对特征图进行采样，以聚合行人目标多尺度上下文信息。S-ASPP 模块将不同扩张率的空洞卷积输出特征图级联相加，再将每个分支输出特征图进行融合。

特征图经过空洞卷积操作得到某一层的输出后，由于相邻元素缺少相关性，会造成上一层图像特征局部信息丢失。S-ASPP 模块通过多分支、多感觉野的卷积操作共享特征图信息，网络学习丰富的图像特征，有效避免特征信息丢失和“网格效应”[17]。S-ASPP 模块添加“shortcut”支路，通过平均池化-卷积操作，再与不同扩张率的融合特征进行相加，使输出特征图中包含行人目标信息更加丰富。

2.2.3　SDIOU-NMS 算法

目标检测中，模型对输入图像进行推理检测会对同一目标产生多个预测框，实际应用中模型对单个目标只需保留一个最优预测框。NMS 算法是常用的目标预测框后处理算法，用于消除同一目标冗余的预测框。

NMS 算法[18]原理：设 A 为模型预测的 N 个预测框

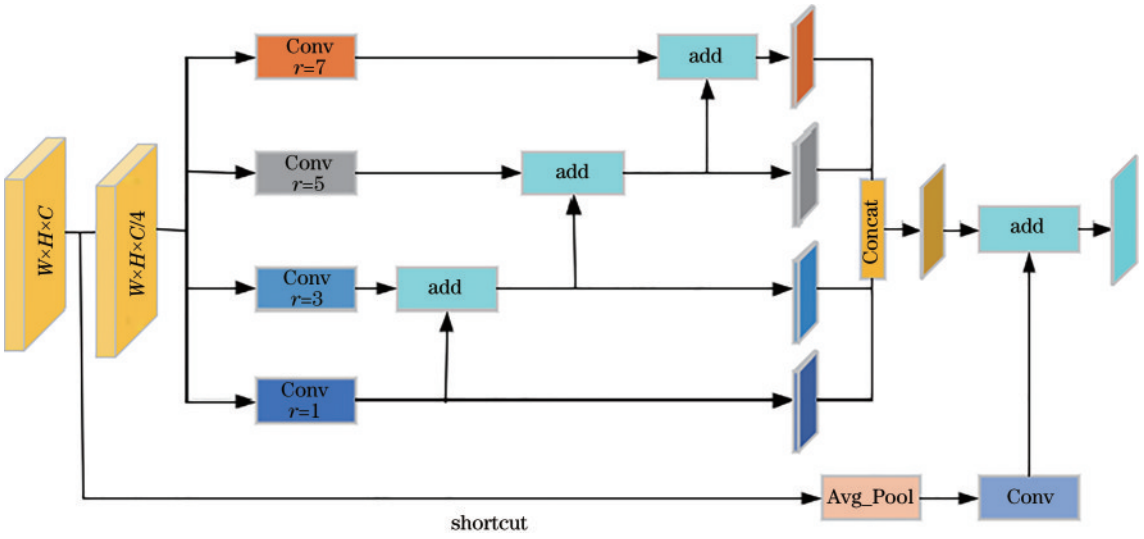

图 7　S-ASPP 结构图

Fig. 7　Structure diagram of S-ASPP

的集合($A=\{a_1, a_2, \cdots, a_N\}$),集合 B 为 A 中 N 个预测框对应的置信度($B=\{b_1, b_2, \cdots, b_N\}$),$N_t$ 为 NMS 所设置的阈值,C 为存放最优框的空集合($C=\varnothing$)。具体流程为:首先,将集合 A 中 N 个预测框按置信度由高到低进行排序,并选出置信度最高的预测框 n_i,同时将预测框 n_i 移到集合 C;然后,遍历集合 A 中剩余预测框 a_j(不含 n_i),分别与预测框 n_i 计算两框的交并比 f_{IoU},将 $f_{IoU} \geqslant N_t$ 的预测框判定为与预测框 n_i 重叠度较高,将其从集合 A 剔除,更新集合 A 中的预测框 a_j 和集合 B 中预测框对应的得分 b_j;最后,回到第一步进行迭代,直到集合 A 为空集,集合 C 则为筛选出的最优预测框集合。集合 A 更新预测框对应的得分 b_j,表示为

$$b_j=\begin{cases} b_j, & f_{IoU}(n_i, a_j)<N_t \\ 0, & f_{IoU}(n_i, a_j)\geqslant N_t \end{cases}, i\neq j。 \tag{5}$$

DIOU-NMS 算法利用 D_{IoU} 计算方式代替 f_{IoU} 作为 NMS 抑制准则,算法原理与传统的 NMS 相同。运用 DIOU-NMS 算法更新预测框分数 b_j 和 D_{IoU}、$R_{D_{IoU}}(n_i, a_j)$ 式,分别表示为

$$b_j=\begin{cases} b_j, & D_{IoU}(n_i, a_j)<N_t \\ 0, & D_{IoU}(n_i, a_j)\geqslant N_t \end{cases}, i\neq j, \tag{6}$$

$$D_{IoU}=f_{IoU}(n_i, a_j)-R_{D_{IoU}}(n_i, a_j), i\neq j, \tag{7}$$

$$R_{D_{IoU}}(n_i, a_j)=d^2/c^2, i\neq j, \tag{8}$$

式中:$f_{IoU}(n_i, a_j)$ 为两个框的交并比;d^2 为两框中心点距离的平方;c^2 为两框最小闭包矩形的对角线距离平方。由式(5)、式(6)可知,传统 NMS 算法和 DIOU-NMS 算法的抑制原理都是 f_{IoU}、$D_{IoU} \geqslant N_t$ 时,直接将预测框的得分置为零。此外,传统 NMS 算法和 DIOU-NMS 算法只是将预测框得分作为是否抑制的衡量标准,并将得分简单作为预测框的置信度,但某些情况下得分高的预测框位置不一定准确。

在地铁行人目标检测中,当目标出现密集和严重遮挡的情况时,不同目标的预测框重叠度较高,传统 NMS 算法和 DIOU-NMS 算法判断预测框得分较高则可能被抑制,会大大降低模型的召回率。为此,本研究设计了一种 SDIOU-NMS 算法,该算法利用一种衰减机制对预测框得分进行优化,通过降低预测框得分而非直接置零,使得一些高分预测框可能作为正确预测框被保留,改进预测框分数 b_j,可表示为

$$b_j=\begin{cases} b_j, D_{IoU}(n_i, a_j)<N_t \\ b_j\exp\left[-\dfrac{D_{IoU}(n_i, \alpha_j)^2}{\sigma}\right], D_{IoU}(n_i, a_j)\geqslant N_t \end{cases}, i\neq j, \forall a_j\notin C, \tag{9}$$

式中:σ 为调整因子(本研究取 0.3)。由式(9)可知,SDIOU-NMS 是通过一个权重函数与预测框置信度相乘策略来更新预测框分数,得分衰减程度与重叠程度成正比。

SDIOU-NMS 算法思想:当 $D_{iou}(n_i, a_j)\geqslant N_t$ 时,算法引入一个衰减机制来降低预测框的分数,算法会根据 $D_{iou}(n_i, a_j)$ 大小来调整分数降低的程度,最后得到一个新加权的预测框分数 b_j。待算法按照流程遍历完全部预测框后,对模型保留的预测框,算法将再设置一个阈值决定是否保留该预测框。SDIOU-NMS 算法主要针对遮挡目标中出现预测框分数较高而会被错误抑制的情况,通过降低该预测框分数并对其重计算判决,这在一定程度上可防止预测框的误删,进而提高模型预测准确率。

3 实验数据与分析

3.1 实验环境

本研究使用 TensorFlow 深度学习框架进行网络模型部署，代码基于 Windows 10 操作系统使用 Python 语言编写，所用 CPU 为 Intel Core i7 9700K 处理器，显卡类型为 NVIDIA GeForce RTX2080Ti(11 G)。

3.2 实验数据集

实验数据集一部分为地铁监控视频抽帧图片，另一部分为“人头”作为“person”类别标记的 CrowdHuman 数据集。实验数据采集地铁大多数场景环境的行人目标，如密集行人目标、不同程度遮挡行人目标等。对地铁监控视频抽帧采样的图片用 Labelimg 软件进行手动标注，标注后的标签是一系列可扩展标记语言(XML)格式。标注过程中对于遮挡严重的行人目标按照“人头”标注“person”类。实验按照 6∶2∶2 比例，训练集 4800 张，验证集和测试集各 1600 张，一共选取 8000 张图片，共有 41935 条标签，平均每张图片包含大约 5 个行人标签。

3.3 实验训练设计

模型训练输入图像尺寸设置为 608×608，训练 epoch 为 300，训练损失进行归一化设置，标签平滑设置为 0.01。采用冻结训练方法加快模型训练速度，冻结模型层数为 150 层，冻结 epoch 为 50。解冻前模型 batch-size 设置为 8，最大学习率为 10^{-3}；解冻后 batch-size 设置为 2，最大学习率为 10^{-4}。模型训练使用 Early-Stopping，模型验证损失(Validation-loss)多次不下降自动结束训练，模型则基本收敛。

3.4 实验评估指标

本研究使用目标检测通用评估指标来定量分析改进 YOLOv5s 算法在地铁场景行人检测的性能，分别为精确率 P(Precision)、召回率 R(Recall)、P-R 曲线、平均精确率 f_{AP}(Average precision)、验证损失(Validation-loss)、检测速度(Frames per second，FPS)。在目标检测中，通常把样本 $f_{iou}\geqslant$置信度阈值分为正例，样本 $f_{iou}<$置信度阈值分为负例。本研究将置信度阈值设置为 0.5。

所有检测出的目标检测精确率 P 可表示为

$$P=\frac{f_{TP}}{f_{TP}+f_{FP}},\tag{10}$$

所有正样本中检测召回率 R 可表示为

$$R=\frac{f_{TP}}{f_{TP}+f_{FN}},\tag{11}$$

不同召回率下精确率的均值 f_{AP} 可表示为

$$f_{AP}=\int_0^1 P\cdot R\mathrm{d}R,\tag{12}$$

式中：f_{TP} 为正样本预测为正类个数；f_{FP} 为负样本预测为正类个数；f_{FN} 为正样本预测为负类个数。

在模型训练过程中，Validation-loss 曲线能够直接反映模型训练的好坏程度——网络欠拟合或过拟合，为此选择增加模型网络层、损失函数正则化、Dropout 等方法解决模型训练中可能出现的问题。

目标检测实际应用中，网络模型需较高的准确率，其检测速度也要满足实时性。网络的检测速度由 FPS 来衡量，FPS 定义为模型每秒检测图片数量。

3.5 数据分析

为验证改进模块的性能，本研究以 YOLOv5s 原始模型为基准方法，设计一系列消融实验进行对比验证，使用 f_{AP}、FPS 两种定量评价指标，实验结果如表 1 所示。

表 1 不同模块的性能比较

Table 1 Performance comparison of different modules

Method	Backbone	Neck	NMS	f_{AP} /%	FPS /(frame·s^{-1})
1	CSPDarkNet	SPP+PAN	DIOU-NMS	87.6	35.6
2	CSPDarkNet	SPP+PAN	SDIOU-NMS	88.2	33.9
3	DRSN-CSPDarkNet	SPP+PAN	DIOU-NMS	88.6	33.1
4	DRSN-CSPDarkNet	SPP+PAN	SDIOU-NMS	89.1	31.1
5	CSPDarkNet	S-ASPP+PAN	DIOU-NMS	87.8	35.1
6	CSPDarkNet	S-ASPP+PAN	SDIOU-NMS	88.7	33.4
7	DRSN-CSPDarkNet	S-ASPP+PAN	DIOU-NMS	88.6	33.8
8	DRSN-CSPDarkNet	S-ASPP+PAN	SDIOU-NMS	89.6	30.4

本研究选取每组消融实验最优结果进行分析，表 1 中：对比方法 1 和方法 3 可知，在模型主干网络方面，DRSN-CSPDarkNet 相对于 YOLOv5s 中 CSPDarkNet 行人目标检测大约提高 1.0 个百分点，模型 FPS 降低大约 2.5 frame/s。对比方法 2 和方法 6 可知，在模型特征加强网络中，算法运用 S-ASPP 模块融合图像的多尺度特征相对于 YOLOv5s 算法的 SPP 模块，行人目标检测大约提高 0.5 个百分点，模型 FPS 基本不变。对比方法 7 和方法 8 可知，在对预测框后处理过程中，SDIOU-NMS 算法相较于 DIOU-NMS 算法，行人目标检测提高大约 1.0 个百分点，但 FPS 降低大约 3.4 frame/s。由数据分析可知，本研究所提出的模

块运用在地铁场景行人检测中，网络检测精度都得到一定程度的提升，但FPS有着小幅度降低，分析原因是YOLOv5s算法新增加一系列模块，网络模型复杂度提升，同时对预测框后处理过程中，SDIOU-NMS算法保留一些可能被剔除的预测框并对其重新判决，在一定程度上影响网络检测速度。

为体现改进YOLOv5s算法的网络性能，本研究将改进YOLOv5s算法与YOLOv5s算法、YOLOv4算法、YOLOv3算法[19]、YOLOv3-Tiny算法和R-CNN算法进行地铁不同场景行人检测结果定量分析。6种模型对测试集检测得出的评价指标数值如表2所示。

表2　不同模型检测指标对比

Table 2　Comparison of detection indexes of different models

Method	P /%	R /%	f_{AP} /%	FPS /(frame·s^{-1})
Improved YOLOv5s	**89.7**	**88.9**	**89.6**	30.4
YOLOv5s	89.1	87.2	87.6	35.6
YOLOv4	86.4	86.1	85.6	32.9
YOLOv3	85.8	83.2	81.2	26.4
YOLOv3-Tiny	84.1	81.1	83.1	**37.4**
R-CNN	80.1	82.1	79.8	20.4

由表2可知，R-CNN算法相比其他算法对行人目标检测效果最差，因为该方法在特征提取网络中没有对高层特征进行多尺度融合，对不同尺度的行人目标检测鲁棒性较差。同时，模型对单张输入图片检测产生几千个候选框，导致卷积神经网络做大量前向计算，不能够满足目标检测所需的实时性，FPS相对较低。

YOLOv3-Tiny算法召回率R相对较低，因为YOLOv3-Tiny算法在YOLOv3算法基础上去除一些特征层，模型只保留两个独立的预测分支，这对小行人目标的检测影响较大，但YOLOv3-Tiny算法相比其他方法模型较轻，FPS则相对偏高。

YOLOv3算法、YOLOv4算法和YOLOv5s算法的精确率P、召回率R和平均精确率f_{AP}与改进YOLOv5s算法相比都有一定程度下降，分析原因是主要有改进YOLOv5s算法添加D-SE注意力模块，模型训练中能够关注图像重要特征，在一定程度上提高了模型预测准确性。同时，YOLOv3、YOLOv4和YOLOv5s这3种算法在目标检测后处理过程中，NMS算法处理多余预测框条件比较苛刻，可能造成误检或漏检，影响模型检测精度和召回率。从检测速度方面来看，改进YOLOv5s算法由于添加一系列模块，模型复杂度增高且计算量增加，造成FPS相对偏低。

本研究提出的改进YOLOv5s算法模型P达到89.7%，R达到88.9%，f_{AP}达到89.6%。虽然FPS相对偏低，但其综合性能优于其他算法。算法结构精简，未来将继续优化算法的检测精度。

本研究选取改进YOLOv5s算法与YOLOv5s算法Validation-loss曲线来对模型训练阶段进行分析，如图8所示。由图8可知，大约在200个epoch后两个模型训练逐渐达到收敛。网络模型收敛时，改进YOLOv5s算法比YOLOv5s算法Validation-loss值偏低，这表明改进YOLOv5s算法对本研究所使用的数据集有更好的泛化性。

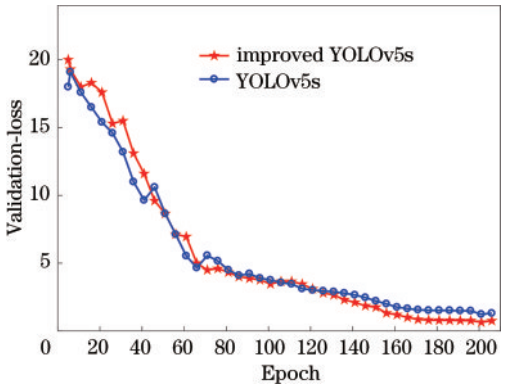

图8　Validation-loss对比

Fig. 8　Comparison of Validation-loss

此外，本研究将YOLOv3算法、YOLOv4算法、YOLOv5s算法和改进YOLOv5s算法用于对实际地铁场景行人目标进行检测，检测结果如图9所示。图9(a)～图9(c)为待检测的原始图像，分别代表3种不同遮挡程度和不同尺寸的行人目标场景。

由检测效果图9(d)、图9(g)、图9(j)、图9(m)可知，YOLOv3算法、YOLOv4算法、YOLOv5s算法和改进YOLOv5s算法应用在遮挡度较低的场景中，4种算法对大尺寸行人目标检测效果都较好，能精确检测出行人目标的位置，但改进YOLOv5s算法和YOLOv5s算法检测出目标的置信度略高于其他两种方法。

由图9(e)、图9(h)、图9(k)、图9(n)可知，在行人目标数量相对较多且遮挡度适中的场景中，改进YOLOv5s算法仍能较为准确地检测出未遮挡行人目标，对于遮挡的行人目标，算法仍能较好地检测出准确位置，但目标置信度降低。YOLOv5s算法对于此类行人目标检测效果也表现良好，但在检测框置信度方面，较改进YOLOv5s算法偏低。YOLOv3算法、YOLOv4算法对此类场景行人检测效果则大幅度下降，遮挡的行人目标和中等尺寸的行人目标漏检率增高，且YOLOv3算法不能精确定位检测目标的位置。

由图9(f)、图9(i)、图9(l)、图9(o)可知，在密集且遮挡严重的行人目标场景中，YOLOv3算法、YOLOv4算法和YOLOv5s算法几乎检测不出遮挡严重的行人目标和小行人目标，改进YOLOv5s算法虽不能完全检测图像中行人目标，但对遮挡严重的小行人目标总体仍有较好的检测结果。

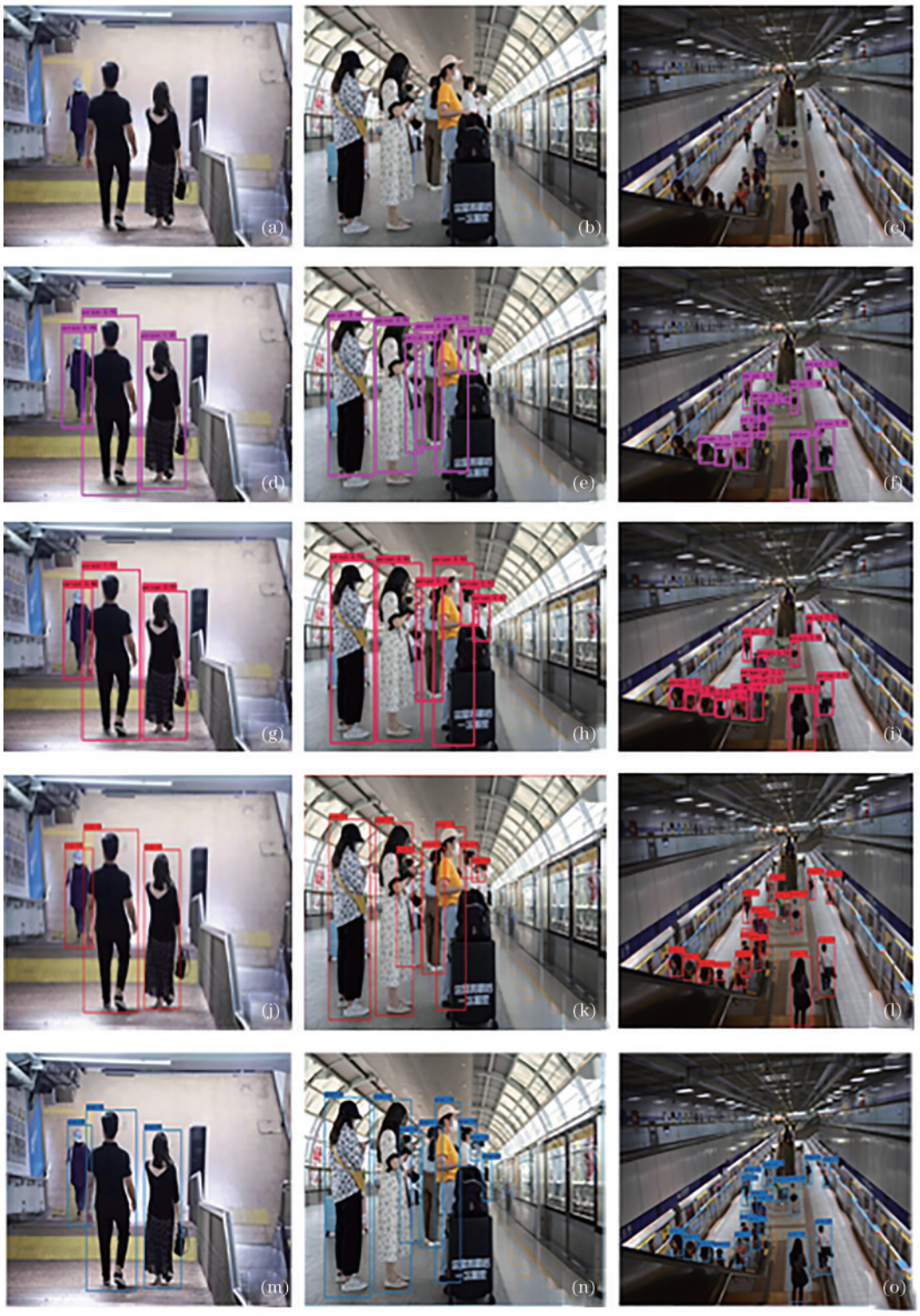

图 9　不同模型行人检测效果。(a)～(c)原始图像;(d)～(f) YOLOv3 检测效果;(g)～(i) YOLOv4 检测效果;(j)～(l) YOLOv5s 检测效果;(m)～(o) 改进 YOLOv5s 检测效果

Fig. 9　Pedestrian detection effects of different models. (a)-(c) Original image; (d)-(f) YOLOv3 detection effect; (g)-(i) YOLOv4 detection effect; (j)-(l) YOLOv5s detection effect; (m)-(o) improved YOLOv5s detection effect

4　结　　论

本研究提出一种改进 YOLOv5s 算法并将其应用到地铁场景行人目标检测中,模型加入深度残差收缩网络,将软阈值化函数作为可训练模块嵌入到深度残差网络中,同时加入 D-SE 注意力机制模块,使网络对地铁行人目标特征更加关注,提高了模型对地铁场景中遮挡行人目标的检测精度。通过添加改进空洞空间金字塔池化模块 S-ASPP 增大局部感受野,对输出进行多尺度融合和拼接,捕获全局有效特征。利用

SDIOU-NMS算法对目标预测框后处理，有效保留遮挡目标的预测框并对其重计算判决，进而提高网络模型对遮挡目标检测的召回率。实验结果分析表明：改进YOLOv5s算法可有效提高地铁场景行人目标的召回率和平均精确率，相对于YOLOv5s算法分别提高1.7个百分点和2.0个百分点；在实际场景测试中，模型对遮挡严重的小行人目标检测漏检率和错检率仍较高，未来将进一步优化网络模型来提高模型性能。

参 考 文 献

[1] 车志富，苗振江，王梦思. 地铁视频监控系统中的行人检测研究与应用[J]. 现代城市轨道交通，2010(2): 31-33，36，80.
Che Z F，Miao Z J，Wang M S. Investigation and application of pedestrian detection in metro video monitoring system[J]. Modern Urban Transit，2010(2): 31-33，36，80.

[2] Girshick R，Donahue J，Darrell T，et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition，June 23-28，2014，Columbus，OH，USA. New York: IEEE Press，2014: 580-587.

[3] 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，June 27-30，2016，Las Vegas，NV，USA. New York: IEEE Press，2016: 779-788.

[4] 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.

[5] 邹梓吟，盖绍彦，达飞鹏，等. 基于注意力机制的遮挡行人检测算法[J]. 光学学报，2021，41(15): 1515001.
Zou Z Y，Gai S Y，Da F P，et al. Occluded pedestrian detection algorithm based on attention mechanism[J]. Acta Optica Sinica，2021，41(15): 1515001.

[6] 李经宇，杨静，孔斌，等. 基于注意力机制的多尺度车辆行人检测算法[J]. 光学 精密工程，2021，29(6): 1448-1458.
Li J Y，Yang J，Kong B，et al. Multi-scale vehicle and pedestrian detection algorithm based on attention mechanism [J]. Optics and Precision Engineering，2021，29(6): 1448-1458.

[7] Bodla N，Singh B，Chellappa R，et al. Soft-NMS: improving object detection with one line of code[C]// 2017 IEEE International Conference on Computer Vision，October 22-29，2017，Venice，Italy. New York: IEEE Press，2017: 5562-5570.

[8] 董小伟，韩悦，张正，等. 基于多尺度加权特征融合网络的地铁行人目标检测算法[J]. 电子与信息学报，2021，43(7): 2113-2120.
Dong X W，Han Y，Zhang Z，et al. Metro pedestrian detection algorithm based on multi-scale weighted feature fusion network[J]. Journal of Electronics & Information Technology，2021，43(7): 2113-2120.

[9] Bochkovskiy A，Wang C Y，Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-02-05]. https://arxiv.org/abs/2004.10934.

[10] Zhao M H，Zhong S S，Fu X Y，et al. Deep residual shrinkage networks for fault diagnosis[J]. IEEE Transactions on Industrial Informatics，2020，16(7): 4681-4690.

[11] Mehta S，Rastegari M，Shapiro L，et al. ESPNetv2: a light-weight，power efficient，and general purpose convolutional neural network[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)，June 15-20，2019，Long Beach，CA，USA. New York: IEEE Press，2019: 9182-9192.

[12] 王凤随，王启胜，陈金刚，等. 基于注意力机制和Soft-NMS的改进Faster R-CNN目标检测算法[J]. 激光与光电子学进展，2021，58(24): 2420001.
Wang F S，Wang Q S，Chen J G，et al. Improved faster R-CNN target detection algorithm based on attention mechanism and soft-NMS[J]. Laser & Optoelectronics Progress，2021，58(24): 2420001.

[13] Chen Y B，Wang H，Han Z. Improved YOLO model with multi-feature fully convolutional network for object detection[J]. Proceedings of SPIE，2020，11526: 1152607.

[14] 王天昊. 一种改进的CNN-ResNet深度学习神经网络及应用[D]. 长春：吉林大学，2020: 14-20.
Wang T H. An improved CNN-ResNet deep learning neural network and its application[D]. Changchun: Jilin University，2020: 14-20.

[15] 杨耘，李龙威，高思岩，等. 基于YOLOv3网络训练优化的高分辨率遥感影像目标检测[J]. 激光与光电子学进展，2021，58(16): 1601002.
Yang Y，Li L W，Gao S Y，et al. Objects detection from high-resolution remote sensing imagery using training-optimized YOLOv3 network[J]. Laser & Optoelectronics Progress，2021，58(16): 1601002.

[16] Zhao L，Zhang X F. Object detector based on enhanced multi-scale feature fusion pyramid network[C]//2021 IEEE 5th Advanced Information Technology，Electronic and Automation Control Conference (IAEAC)，March 12-14，2021，Chongqing，China. New York: IEEE Press，2021: 289-293.

[17] Lu X Y，Zhong Y F，Zheng Z，et al. GAMSNet: Globally aware road detection network with multi-scale residual learning[J]. ISPRS Journal of Photogrammetry and Remote Sensing，2021，175: 340-352.

[18] Liu S T，Huang D，Wang Y H. Adaptive NMS: refining pedestrian detection in a crowd[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)，June 15-20，2019，Long Beach，CA，USA. New York: IEEE Press，2019: 6452-6461.

[19] Redmon J，Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08) [2021-02-05]. https://arxiv.org/abs/1804.02767.