计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
62(2): 397−407, 2025

基于改进 SSD 算法的地铁场景小行人目标检测

张秀再[1,2] 邱 野[1] 沈 涛[1]

[1](南京信息工程大学电子与信息工程学院 南京 210044)

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

(zxzhering@163.com)

Small Pedestrian Target Detection in Subway Scene Based on Improved SSD Algorithm

Zhang Xiuzai[1,2], Qiu Ye[1], and Shen Tao[1]

[1] (*School of Electronic and Information Engineering*, *Nanjing University of Information Science & Technology*, *Nanjing* 210044)

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

Abstract In the subway scene, small pedestrian targets contain less feature information due to their low resolution, and it is still challenging for object detectors to detect such objects at this stage. SSD target detection algorithm uses the multi-scale detection head of the pyramid network, which can improve the pedestrian target detection performance to a certain extent, but it still has certain limitations in small pedestrian target detection application in complex environments such as subways. In view of the above problems, we propose an improved SSD algorithm to enhance the detection effect of small pedestrian targets in subway scenes, construct a dataset of pedestrian targets in subway scenes, mark the corresponding labels, and perform data preprocessing operations at the same time. In this study, a pyramid feature enhancement module is added to the feature extraction network, and the multi-branch residual unit, sub-pixel convolution and feature pyramid are combined to obtain image multi-scale and multi-receptive field fusion features. We use the context information fusion module to fuse the low-level features of the image with the context features to generate an extended feature layer for detecting small pedestrian targets, and design an Anchor-free dynamic positive and negative sample allocation strategy to generate optimal positive samples for small pedestrian targets. A dynamic positive and negative sample allocation strategy based on Anchor-free is designed to generate optimal positive samples for small pedestrian targets. The experimental results show that the proposed improved SSD algorithm can effectively improve the performance of small pedestrian target detection in subway scenes, and the effect of small pedestrian target detection with severe occlusion is more obvious.

Key words small pedestrian object detection; SSD algorithm; attention mechanism; subpixel convolution; multi-scale feature fusion

摘 要 在地铁场景中,小行人目标由于分辨率低,包含特征信息较少,现阶段目标检测器对此类目标的检测仍具有挑战性. SSD 目标检测算法利用金字塔网络的多尺度检测头,能一定程度提高行人目标检测性能,但将其应用于地铁等复杂环境中实现小行人目标检测仍具有一定局限性. 针对上述问题,提出一种

收稿日期: 2023-02-06; **修回日期**: 2024-05-13

基金项目: 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项目(BK20141004); 国家自然科学基金青年科学基金项目(11504176, 61601230)

This work was supported by the Natural Science Foundation of Jiangsu Provincial Higher Education Institutions (13KJA510001), the Natural Science Foundation of Jiangsu Province for Young Scientists (BK20141004), and the National Natural Science Foundation of China for Young Scientists (11504176, 61601230).

改进 SSD 算法以加强地铁场景中小行人目标检测效果. 通过构建地铁场景行人目标数据集, 标注相应标签, 同时进行数据预处理操作; 在特征提取网络中加入金字塔特征加强模块, 将多分支残差单元、亚像素卷积和特征金字塔相结合获得图像多尺度、多感受野融合特征; 利用上下文信息融合模块将图像低层特征与上下文特征相融合, 生成扩展特征层用于检测小行人目标; 设计一种基于 Anchor-free 的动态正负样本分配策略, 为小行人目标生成最优正样本. 实验结果表明: 提出的改进 SSD 算法能有效提高地铁场景小行人目标检测性能, 对遮挡严重的小行人目标检测, 效果提升更为明显.

关键词 小行人目标检测; SSD 算法; 注意力机制; 亚像素卷积; 多尺度特征融合

中图法分类号 TP391

DOI: 10.7544/issn1000-1239.202330069 **CSTR:** 32373.14.issn1000-1239.202330069

随着深度学习的蓬勃发展, 目标检测技术已发展得相当成熟. 小目标检测一直是计算机视觉任务的热点, 也是目标检测方向发展的瓶颈. 现有的目标检测器对大尺度和中等尺度目标的检测效果表现较为出色, 但对于小目标的检测效果较差. 实际检测场景中, 光照变化和目标不同程度遮挡等外界因素的存在, 严重阻碍了检测器对小目标特征信息的提取, 而小目标自身存在特征信息不足的缺陷, 造成小目标检测漏检率和错检率较高.

1 相关工作

目标检测技术是目标识别和目标定位的结合, 要求检测器不仅准确识别出目标具体类别, 而且同时能精确定位目标具体位置. 目前主流目标检测算法可分为 2 类: Two-stage 算法和 One-stage 算法. Two-stage 算法是将目标识别和定位分为 2 步完成, 此类算法对各类目标的漏检率和错检率相对较低, 但实时性较差, 其中代表性算法有 R-CNN[1](region convolutional neural networks)算法和 Fast R-CNN[2] 算法等. One-stage 算法可有效解决 Two-stage 算法的问题, 满足高实时性的同时准确率也基本达到 Two-stage 算法水平, 其代表性算法有 YOLO[3] (you only look once)系列算法和 SSD[4](single shot multibox detector)算法等.

为解决多场景中小目标检测精度较低、漏检率和错检率较高等问题, 戚玲珑等人 [5] 设计出一种融合注意力模块的改进 YOLOv7 小目标检测算法, 通过提高模型对小目标敏感度和减少小目标特征损失等方法, 进一步降低小目标的漏检率. 李子豪等人[6] 提出一种基于自适应协同注意力机制的目标检测算法, 对模型主干网络和检测头等模块进行轻量化设计, 引出大尺度目标检测头的密集锚框提升小目标检测精度. 李亚东等人[7] 针对小目标漏检率较高和图像特征融合不充分问题, 提出一种多注意力特征加权融合的小目标检测算法, 利用注意力机制聚焦小目标的区域特征, 降低小目标漏检率. 上述各类改进算法能一定程度提高对应环境小目标检测精度, 但对于遮挡严重的小目标, 其检测效果仍然不佳.

SSD 算法作为 One-stage 目标检测算法, 对目标实现端到端检测, 且推理速度和准确度均有较高指标. YOLO 系列算法在实现全连接层后再对多尺度目标进行检测, SSD 算法则利用卷积神经网络直接对多尺度目标进行检测, 利用目标低层特征图和高层特征图分别检测小目标和中等及大尺度目标. YOLO 系列算法缺点是小目标检测性能不佳, 反之, SSD 算法网络结构更适合检测小目标, 且定位更为准确. 因此, 本文选取 SSD 算法作为基础算法并进行改进, 以提高算法对地铁小行人目标的检测效果.

改进 SSD 算法在特征提取网络后加入金字塔特征加强模块[8](pyramid feature enhancement module, PFEM), 首先利用双维度注意力模块[9](two-dimensional attention module, TDCAM)辅助模型关注行人目标空间维度和通道维度等重要特征信息, 聚焦可视化行人特征; 其次使用亚像素卷积代替上采样操作实现图像特征低分辨率到高分辨的转化; 再次通过多分支残差单元[10](multi-branch residual unit, MBRU), 加强模型融合图像多尺度特征能力; 接着模型加入上下文特征融合模块(context feature fusion module, CFFM), 为小行人目标拓展额外检测层; 最后, 模型应用基于 Anchor-free[11] 的动态正负样本分配策略(dynamic positive and negative sample allocation strategy, DAS), 为目标真实框分配最优正样本.

2 原理与方法

2.1 SSD 算法

SSD 算法主要由特征提取模块、特征加强模块和预测模块 3 部分组成. 特征提取模块负责提取输入

图像的低层特征，特征加强模块负责提取图像的高层特征并对其进行多尺度融合，预测模块负责输出特征映射图的位置信息和类别信息. SSD 算法的核心在于利用不同层次的特征图进行多尺度目标检测，并将检测结果进行综合以选出目标最佳预测框. SSD 算法模型如图 1 所示.

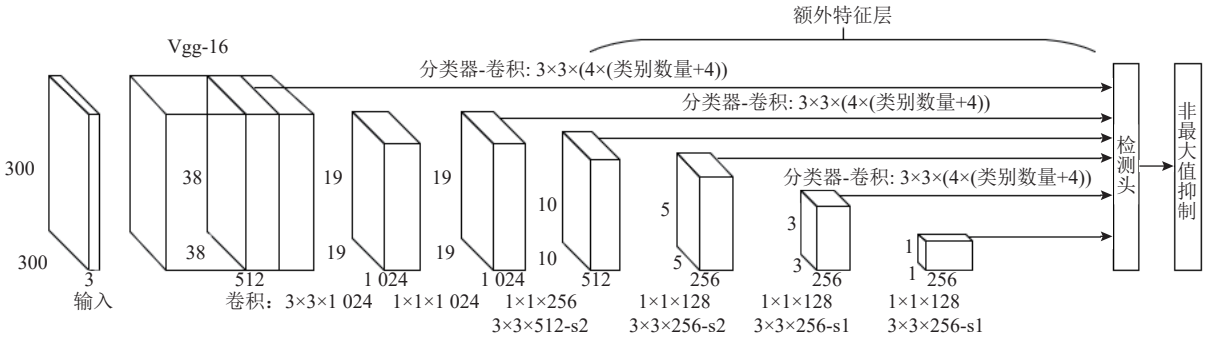


Fig. 1 SSD algorithm model

图 1 SSD 算法模型

SSD 算法的主干网络结构是 Vgg-16，其中 2 个全连接层改成卷积层，并增加 4 个卷积层构成特征金字塔网络. SSD 算法中 6 个不同层次的特征图，经过 2 个 3×3 的卷积操作后生成 6 个多尺度目标检测头. 检测头的输出内容包括目标分类和目标预测框回归，输出分类信息用置信度表示，每个预测框生成 C+1 个置信度（C 为目标类别数），另一个输出目标预测框回归位置信息，每个预测框生成 4 个坐标值（x, y, w, h）. 最后，SSD 算法使用非最大值抑制算法对所有检测层的目标预测框进行筛选，生成单个目标最佳预测框.

2.2 改进 SSD 算法

本文以 SSD 算法为基础，对原始模型进行优化，模型改进方法为：1）为避免常规上采样对特征图产生无效信息[12]，模型采用亚像素卷积实现特征图从低分辨率到高分辨率转换. 2）模型添加双维度注意力模块，从空间维度和通道维度聚焦小行人目标特征，协助模型准确定位包含丰富目标信息量的特征通道和空间位置，并抑制无效背景特征. 3）小行人目标分辨率较低，包含有用特征信息不足，设计出一种上下文特征融合模块，将图像高层特征与低层特征有效融合，利用上下文信息生成面向小目标检测的扩展特征层. 4）设计一种基于 Anchor-free 的动态正负样本分配策略，为小行人目标分配最优正样本，通过改变模型正样本分配机制，加速模型收敛. 改进 SSD 算法模型如图 2 所示.

1）金字塔特征加强模块

常规卷积操作中，图像特征所映射的空间信息具体位置由卷积核大小控制，这种“滤波器”学习方式具有一定局限性. 单尺寸卷积核构成的特征提取

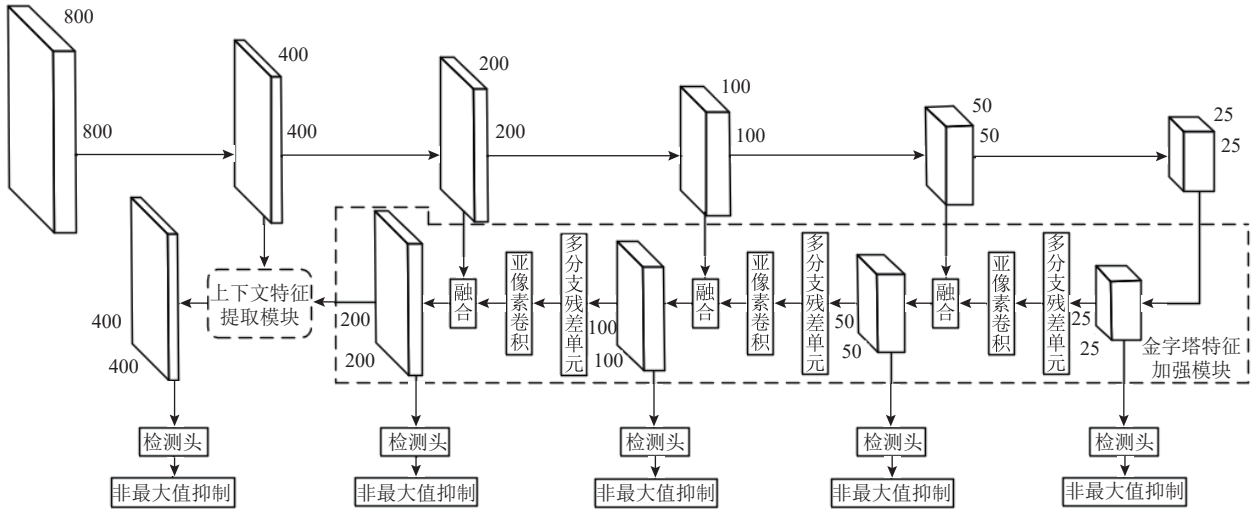


Fig. 2 Improved SSD algorithm model

图 2 改进 SSD 算法模型

层缺乏丰富的感受野,模型无法捕获充分的全局语义信息,从而造成特征图的辨识度降低.

为捕获图像多尺度特征,同时增强主干网络对多尺度特征的表达能力和泛化能力,本文提出一种多分支残差单元.与增加卷积层获得多尺度特征方式不同,该单元在像素层次上来表征图像多尺度特性,同时增加了单个卷积层的感受野,以获得图像多感受野的融合特征.多分支残差单元利用多分支卷积组来代替原始3×3卷积,采用类似“残差”的层次化结构相连,在细粒度层次上利用图像多尺度特征.残差单元结构和多分支残差单元结构分别如图3和图4所示.

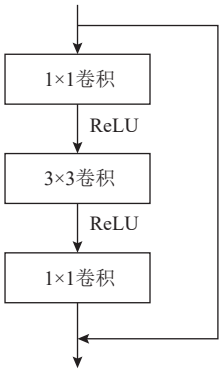


Fig. 3 Residual unit structure diagram

图3 残差单元结构图

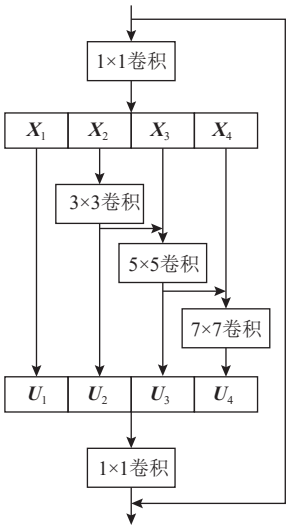


Fig. 4 Multi-branch residual unit structure diagram

图4 多分支残差单元结构图

在多分支残差单元结构中,输入特征图经过1×1卷积后被划分为4个分支特征图$\boldsymbol{X}_i$.每个分支特征图保持宽度和高度相同,但通道数为输入特征图的1/4.除特征图$\boldsymbol{X}_1$直接与输出特征图$\boldsymbol{U}_1$直连,其余分支特征图$\boldsymbol{X}_i(i\neq1)$都经过一组不同尺寸卷积核进行采样,然后将卷积采样结果与特征图$\boldsymbol{X}_{i+1}$进行融合,最后经过一个新尺寸卷积核输出特征图$\boldsymbol{U}_{i+1}$.

多分支残差单元结构上比常规残差单元增加了多感受野的残差分支,在维度方面添加一个重要参数*scale*,即分支特征图数量,*scale*的增加比采用多层卷积方式获得图像多尺度特征更有效.本文设*scale*=4.某种程度上,上述2类残差单元在提取图像空间特征上具有关联性,多分支残差单元可以在卷积神经网络架构中实现即插即用,该单元通过分层残差连接[13],感受野对于细粒度级别上的变化更加敏感,帮助模型有效捕获图像细节和全局特性.

为提高特征图的分辨率,常规方法有上采样、上池化和反卷积等方式.上述方式是将图片设定预期尺寸后,使用双线性插值等方式对其余扩展点进行补充来完成上采样过程.插值的方式可能造成特征图存在相当数量的无效信息,导致图像边缘信息出现模糊现象.因此,本文使用亚像素卷积(sub-pixel convolution, SPC)来对特征图进行上采样.亚像素卷积原理如图5所示.

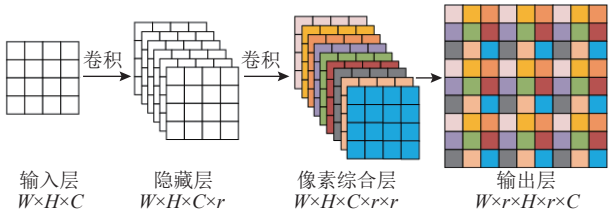


Fig. 5 Principle diagram of sub-pixel convolution

图5 亚像素卷积原理图

亚像素卷积去除大量卷积计算,是常规卷积的一种简化形式,主要是利用卷积和不同通道的像素重组而输出高分辨率特征图.亚像素卷积原理为:第1组特征图为输入层,第2组特征图为隐藏层,主要负责对输入图片进行特征提取,由输入层经过步长为1、卷积核大小为3×3×r的常规卷积操作输出,其中r为上采样倍数.第3组和第4组为像素重组层和输出层,像素重组层是按照一定规则将特征值进行重新排列组合.亚像素卷积实现了图像从低分辨率到高分辨率的转变,完成图像放大的插值函数隐含在卷积操作中,此外所有的卷积运算过程都在低分辨率图像层面上进行,有效提高图像从低分辨率到高分辨率的转变效率.

图像经过主干网络完成特征提取后生成多级不同尺度特征图,深层特征图包含丰富抽象语义信息,但由于其分辨率较低,无法准确定位目标位置信息.浅层特征图保留的语义信息相对较少,但由于其分

辨率较高，能准确定位目标位置信息. 本文提出一种金字塔特征加强模块来实现不同层次的特征融合[14]，该模块将特征信息通过“自顶向下”的方式进行传播，浅层特征可以结合深层特征抽象语义信息进行优化，进而生成适合检测小行人目标的检测头分支.

金字塔特征加强模块将深层特征图进行上采样操作，然后与浅层特征图进行融合，保证图像浅层特征也保留丰富的抽象语义信息. 图像深层特征经过亚像素卷积进行上采样与低层特征融合后，会造成特征信息的非连续性，导致原始特征图语义信息失真. 金字塔特征加强模块在深层特征和浅层特征融合后加入多分支残差单元，利用特征再提取来保证融合特征信息的稳定性. 金字塔特征融合模块结构如图 6 所示.

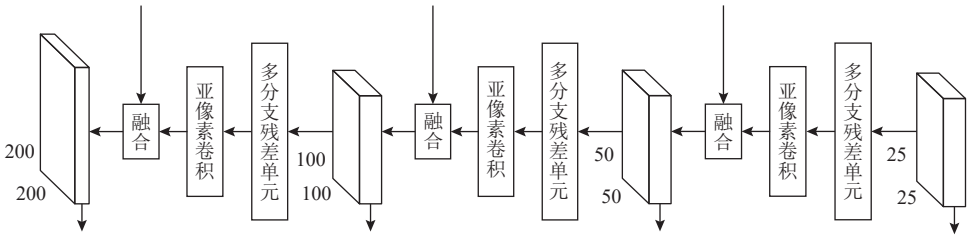


Fig. 6 Pyramid feature enhancement module structure diagram

图 6 金字塔特征加强模块结构图

2）上下文信息融合模块

常规卷积运算是将输入图像的通道信息和空间信息融合后进行特征提取，模型无法关注图像不同维度的重要特征[15]. 双维度卷积注意力模块是一种连接空间维度和通道维度的注意力机制，通过独立的空间维度和通道维度来推理注意力权值，输入特征图与注意力权值相乘进行特征优化，加强模型自适应学习图像重要特征能力，进而提高小行人目标检测准确性. 双维度卷积注意力模块结构及其原理分别如图 7 和图 8 所示.

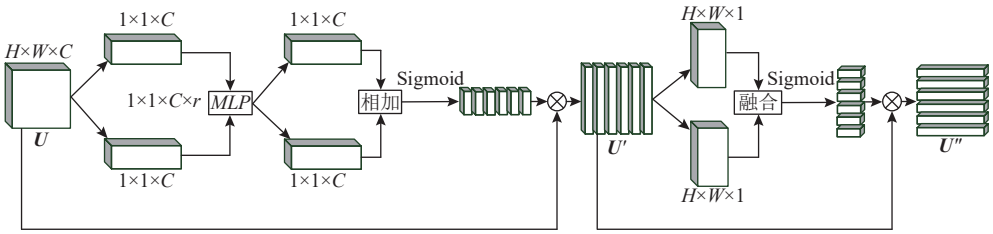


Fig. 7 Two-dimensional convolutional attention module structure

图 7 双维度卷积注意力模块结构

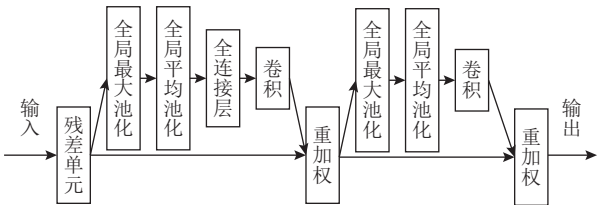


Fig. 8 Principle diagram of the two-dimensional convolutional attention module

图 8 双维度卷积注意力模块原理图

双维度卷积注意力模块主要由通道注意力加权和空间注意力加权 2 部分组成，可灵活应用于多种网络中. 第 1 步实现输入特征图 $\boldsymbol{U}$ 通道维度加权，其具体流程为：特征图 $\boldsymbol{U}$ 经过空间维度全局最大池化和全局平均池化，分别生成 2 个尺度为 1×1×C 的特征图，然后特征值逐元素相加，以聚合特征映射的空间信息. 2 个特征图共享全连接层，压缩特征图 $\boldsymbol{U}$ 的空间信息，帮助网络分析特征图中不同通道信息，强化模型对小行人目标重要通道特征的学习能力. 最后，经过一个 Sigmoid 激活函数生成通道注意力权值，特征图 $\boldsymbol{U}$ 与权重值逐元素相乘输出通道加权特征图 $\boldsymbol{U}'$. 第 2 步实现特征图 $\boldsymbol{U}'$空间维度加权，其具体流程为：特征图 $\boldsymbol{U}'$经过通道维度全局最大池化和全局平均池化，以聚合特征映射的通道信息，分别生成 2 个尺度为 H×W×1 的特征图，2 个特征图进行特征融合操作后经过 Conv+Sigmoid 操作生成空间注意力权值. 最后，特征图 $\boldsymbol{U}'$与该权值逐元素相乘生成输入特征图 $\boldsymbol{U}''$. 通道权重矩阵 $\boldsymbol{M}_c(\boldsymbol{U})$ 和空间权重矩阵 $\boldsymbol{M}_s(\boldsymbol{U})$ 定义可由式(1)(2)所示：

$$\boldsymbol{M}_{\mathrm{c}}(\boldsymbol{U})=\sigma\left(MLP\left(\begin{bmatrix}AvgPool(\boldsymbol{U})\\MaxPool(\boldsymbol{U})\end{bmatrix}\right)\right),\tag{1}$$

$$\boldsymbol{M}_{\mathrm{s}}(\boldsymbol{U})=\sigma\left(f^{3\times3}\left(\begin{bmatrix}AvgPool(\boldsymbol{U})\\MaxPool(\boldsymbol{U})\end{bmatrix}\right)\right),\tag{2}$$

其中 MLP 为全连接层，σ 为 Sigmoid 激活函数，$f^{3\times3}$ 为 3×3 卷积操作，$AvgPool$ 为全局平均池化，$MaxPool$ 为全局最大池化，$\boldsymbol{U}$ 为输入特征图.

SSD 算法中，主干网络主要通过卷积和池化等操作提取图像特征，这类重复操作能一定程度保留图像的重要语义信息，但可能丢失有助于模型区分背景和对象区域的低级特征信息. 此外，单层特征图由于感受野的局限性仅能捕获固定量的图像特征信息[16]. 实际检测中，小行人目标自身分辨率较低，包含有用特征信息较少，且无效背景特征信息对其检测影响较大. 因此，本文设计出一种基于注意力机制的上下文信息融合模块，目的是将图像低层特征和上下文特征相融合生成额外的特征层，用于检测小行人目标. 为使模型更加专注特征图中小行人目标，在特征融合前对辅助特征图应用注意力机制，帮助模型在通道和空间维度方面聚焦小行人目标和关注上下文信息. 上下文信息融合模块结构如图 9 所示.

上下文信息融合模块主要由 2 个分支组成，特征图 $\boldsymbol{U}_2$ 是为目标特征图 $\boldsymbol{U}_1$ 提供上下文信息的辅助特征层. 特征图 $\boldsymbol{U}_2$ 连接双维度卷积注意力模块，捕获可视化小行人目标区域细节，同时去除图像冗余背景信息，生成特征图 $\boldsymbol{U}_2'$. 特征图 $\boldsymbol{U}_1$ 首先经过一个信息采样模块提取图像重要语义内容，然后利用亚像素卷积进行上采样，输出特征图 $\boldsymbol{U}_1'$，其分辨率提升 1 倍，最后将特征图 $\boldsymbol{U}_2'$ 与特征图 $\boldsymbol{U}_1'$ 相融合. 残差模块是对特征图 $\boldsymbol{U}_2'$ 和特征图 $\boldsymbol{U}_1'$ 所提供不同层次的目标信息进行再提取，选择可信区域纹理特征辅助模型进行小目标检测. “Shortcut”连接方式是将残差模块提取的区域纹理特征和具有超分辨率内容特征图 $\boldsymbol{U}_1'$ 有效结合[17]，进而生成面向小目标检测的扩展特征层 $\boldsymbol{U}_1''$.

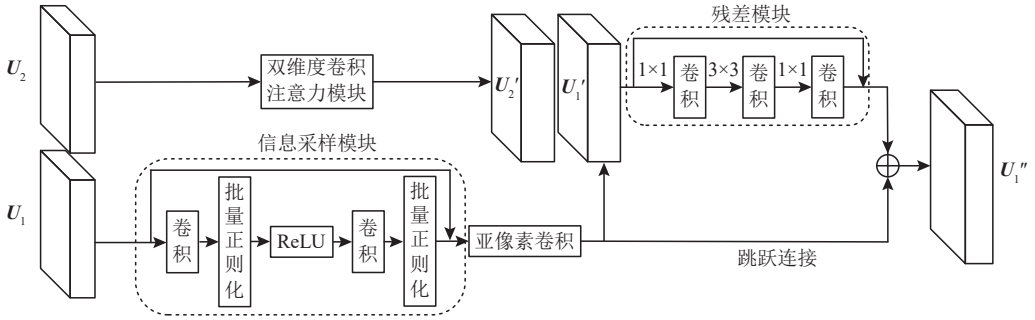


Fig. 9 Context information fusion module structure

图 9 上下文信息融合模块模块结构

3)基于 Anchor-free 的动态正负样本分配策略

在 Anchor-based 的目标检测模型中，通常会采用 K-means 聚类算法并结合遗传算法，得到一组预先设定尺寸和形状的锚框. 实际训练过程中，模型会以目标真实框为基准使用预设定的锚框作为训练对象，模型对目标预测时会以某种规则在待检测图片上生成一系列锚框，而锚框位置表示可能存在物体的候选区域，相当于模型对目标预测框初始化. 假设锚框位置存在检测目标，模型会在锚框的基础上对其进行长宽比微调，生成能够准确描述物体具体位置的预测框，模型同样需要预测出微调幅度.

基于 Anchor-free 的目标检测模型放弃锚框训练策略，但同样以目标真实框为目标，以特征图像素点为样本进行回归训练，假设某像素点在目标真实框内则对应样本概率为 1，该点设为正样本像素点，反之为 0 时设为负样本像素点. 对于正样本像素点，模型会输出该点与目标真实框之间的垂直和水平距离来表示预测框. 假设特征图某像素点与目标真实框有一定关联度，则该像素点的训练回归目标垂直距离 t, b 和水平距离 l, r 定义可由式(3)~(6)所示：

$$t=y-y_0,\tag{3}$$

$$b=y_1-y,\tag{4}$$

$$l=x-x_0,\tag{5}$$

$$r=x_1-x,\tag{6}$$

其中 (x_0, y_0) 和 (x_1, y_1) 分别表示目标真实框的左上角和右下角坐标值；(x, y) 为正样本像素点坐标.

在 Anchor-based 的目标检测模型中，正负样本是根据锚框与真实框的交并比 (iou) 值进行划分的，大于设定阈值的锚框为正样本，低于阈值的锚框为负样本. 对于小目标而言，锚框的微小偏差会造成与真实框的 iou 显著变化. 假设在某个特征图域内，锚框

B 相对于真实框 A 沿对角线方向平移 3 个像素，大尺度目标的 *iou* 从 100% 下降至 71.8%，中尺度目标的 *iou* 从 100% 下降至 56.6%，然而小目标的 *iou* 从 100% 下降至 32.6%. 当对角线平移像素设为 6 个点时，该下降趋势对于小目标更加明显，小目标的 *iou* 仅为 8.7%，反映出小目标对锚框扰动的容忍度更低. 实际工作中，锚框的预设定无法完全符合小目标尺寸，小目标预测框与真实框在同等量级下偏差较大，造成小目标正负样本严重不平衡，模型对小目标的召回率降低. 因此，选择合适正样本对检测精度提升至关重要.

为解决小行人目标正负样本不匹配问题，本文提出一种基于 Anchor-free 的动态正负样本分配策略，该策略可实现模型动态匹配不同小目标的正样本数量，并自动定位检测目标真实框的特征层. 通过算法所产生正样本在定位精准和分类精度方面实现与目标真实框最大程度拟合，然后模型将该正样本集合所对应的特征图像素传入模型参与损失计算及反传，可一定程度提高小行人目标检测精度和召回率.

该策略主要流程如下：①算法首先提取正样本候选区域所生成预测框的位置信息、前背景目标和类别等信息. ②利用上述预测框集合构建预测框与真实框的代价矩阵和 *iou* 矩阵. ③将生成的 *iou* 矩阵元素值按从大到小排序，并将 *iou* 值进行相加取整，得到分配正样本数量 N. ④将代价矩阵元素值按从小到大排序，选取前 N 个预测框作为正样本，其余预测框为负样本. 边界框损失 L_{reg}、类别损失 L_{cls} 和代价矩阵 $\boldsymbol{C}$ 中元素值 C_{ij} 的定义可由式(7)(8)(9)所示：

$$L_{\mathrm{reg}} = -\mathrm{lb}(iou(B_{\mathrm{gt}}, B_{\mathrm{pred}})), \tag{7}$$

$$L_{\mathrm{cls}} = -\sum_{i=1}^{n}(P_i \mathrm{lb}(P_i) + (1-P_i)\mathrm{lb}(1-P_i)), \tag{8}$$

$$C_{ij} = L_{\mathrm{reg}} + \lambda L_{\mathrm{cls}}, \tag{9}$$

其中 B_{gt} 为目标真实框，B_{pred} 为目标预测框，$iou(B_{\mathrm{gt}}, B_{\mathrm{pred}})$ 为真实框与预测框的交并比，P_i 为判决行人目标的概率，λ 为系数，本文取 λ=3.

3 实验数据与分析

3.1 实验环境

使用 Tensorflow 深度学习框架进行网络模型部署，核心硬件配置为 Intel Core i7 9700K 处理器和显卡类型为 NVIDIA GeForce RTX2080Ti(13.4TFLOPS)，软件环境为 Python 语言编写，操作环境为 Windows 10 系统.

3.2 实验训练设计

训练数据集一部分为某地铁站监控视频采样图片，另一部分为开源数据集 CUHK Occlusion Dataset 和 INRIA Person Dataset，两者图片数量混合比例设为 6∶4[18]. 实验数据采集多场景各类行人目标，如密集行人目标、不同程度遮挡行人目标和各种尺度行人目标等. 对地铁站监控视频抽帧采样的图片用 Labelimg 软件进行手动标注，对于遮挡严重的行人目标按照“人头”标注“person”类，标注后的标签是一系列可扩展标记语言(extensible markup language, XML)格式. 实验各部分数据集比例设为 6∶2∶2，其中训练集 4 800 张，验证集和测试集各 1 600 张，一共选取 8 000 张图片，共有 56 935 条标签，平均每张图片包含 7 个行人标签.

模型训练输入图片尺寸为 800×800，训练 epochs 设为 300，冻结训练阶段 epochs 设为 140，非冻结阶段 epochs 设为 160，batchsize 设为 8. 模型使用 Adam 优化器对学习率进行优化，最大学习率设为 10^{-3}，最小学习率设为 10^{-5}. 模型使用权重衰减策略防止过拟合，权重衰减值设为 5×10^{-4}. 模型训练使用 Early-stopping 策略，检测出验证损失趋于平稳则自动终止训练，模型则达到基本收敛.

3.3 实验评估指标

本文使用目标检测通用评估指标对改进 SSD 算法在地铁场景行人检测的性能进行定量分析，分别为精确率(precision, P)、召回率(recall, R)、平均准确率(average precision, AP)和检测速度(frames per second, FPS).

精确率 P 定义为所有检测出的目标检测准确率，可由式(10)表示：

$$P = TP/(TP+FP), \tag{10}$$

其中 TP 为正样本预测为正例的数量，FP 为负样本预测为正例的数量.

召回率 R 定义为所有正样本中的检测准确率，可由式(11)所示：

$$R = TP/(TP+FN), \tag{11}$$

其中 FN 为正样本预测为负例的数量.

平均准确率 AP 定义为不同召回率下精确率的均值，可由式(12)所示：

$$AP = \int_0^1 P(R)\mathrm{d}R. \tag{12}$$

对目标检测器的实际部署过程，模型检测准确性和实时性都需满足一定要求. 模型检测速度由 FPS 来衡量，FPS 定义为模型每秒检测图片数量.

3.4 数据分析

为验证改进 SSD 算法的检测性能,将改进 SSD 算法与 YOLOX 算法[19]、YOLO-v5 算法、PP-YOLO 算法、SSD 算法和 Fast R-CNN 算法来对小行人目标检测结果进行定量分析. 6 种算法对数据集验证部分检测结果的指标数值对比如表 1 所示.

Table 1 Comparison of Detection Indicators of the Six Different Models

表 1 六种不同模型的检测指标对比

算法	*P*/%	*R*/%	*AP*/%	*FPS*/fps
改进 SSD(本文)	**91.1**	**89.9**	**90.3**	37.4
YOLOX	89.6	88.3	88.6	**43.4**
YOLO-v5	86.3	87.1	84.4	36.7
PP-YOLO	85.8	83.2	81.2	42.6
SSD	88.7	89.2	88.1	40.6
Fast R-CNN	90.1	87.5	88.4	30.5

注:黑体数值表示最优值.

由表 1 可知, Fast R-CNN 算法在精确率和平均准确率方面都高于 YOLOX 算法、YOLO-v5 算法、PP-YOLO 算法和 SSD 算法,但 *FPS* 值大幅度下降. 分析其原因是 Fast R-CNN 属于 Two-stage 算法,将目标检测任务分为目标识别和定位 2 步实现,此方式能一定程度提高检测准确性,缺点是该模型在检测时产生几千个候选框,导致模型进行大量计算而不能满足实时性,造成 *FPS* 值相对偏低.

PP-YOLO 算法较于其他算法对小行人检测效果表现最差,因为 PP-YOLO 算法是基于 YOLO-v5 算法的一种优化模型,其引入空间金字塔池化(spatial pyramid pooling, SPP)来加强各类行人的目标特征提取,能一定程度提高模型对行人检测的精确性,但对于小行人目标模型的漏检率较高. 该算法插入可变形卷积来实现轻量化设计, PP-YOLO 算法在 *FPS* 方面表现较好.

提出的改进 SSD 算法模型的精确率达到 91.1%,召回率达到 89.9%,平均准确率达到 90.3%. *FPS* 值相比于其他算法不是最优表现,但模型总体表现性能优于其他算法. 改进 SSD 算法与原始 SSD 算法相比,精确率提高 2.4 个百分点,召回率提高 0.7 个百分点,平均准确率提高 2.2 个百分点. 与 YOLO 系列表现最佳的 YOLOX 算法相比,精确率提高 1.5 个百分点,召回率提高 1.6 个百分点,平均准确率提高 1.7 个百分点. 分析原因主要有改进 SSD 算法使用上下文特征融合模块加强小行人目标特征,注意力机制使得模型能够聚焦可视化行人目标,一定程度提高模型预测的准确性. 同时,改进 SSD 算法加入基于 Anchor-free 的动态正样本分配策略为小行人目标真实框分配最优正样本,而 YOLO 系列算法和 SSD 算法使用阈值判决法分配正负样本,可能会造成训练正样本不足,影响模型的检测召回率和精确率.

为验证设计模块的性能,以 SSD 算法为基准模型,设计消融实验进行对比验证,评价指标为 *AP* 和 *FPS*,实验结果如表 2 所示.

Table 2 Comparison of Detection Indicators of the Different Models

表 2 不同模型的检测指标对比

序号	算法组成	*AP*/%	*FPS*/fps
1	SSD	88.1	40.6
2	SSD+DAS	88.4	39.9
3	SSD+PFEM	88.6	38.1
4	SSD+CFFM	89.1	38.9
5	SSD+ DAS+PFEM	89.4	38.4
6	SSD+ DAS+CFFM	89.7	39.4
7	SSD+PFEM+CFFM	90.1	37.1
8	SSD+ DAS+PFEM+CFFM	90.3	37.4

选取每组消融实验最优结果进行分析,在表 2 中对比方法 2 和方法 5 可知,利用 PFEM 对主干网络提取特征进行多尺度融合,改进 SSD 算法相对于 SSD 算法对小行人目标检测 *AP* 大约提高 1.0 个百分点, *FPS* 大约降低 1.5 fps. 对比方法 2 和方法 6 可知,算法运用 CFFM 生成扩展特征层用于检测小行人目标,小行人目标检测 *AP* 大约提高 0.5 个百分点, *FPS* 大约降低 0.5 fps. 对比方法 3 和方法 5 可知,在对模型正负样本分配过程中, DAS 相较于原始阈值分配正负样本法,小行人目标检测 *AP* 大约提高 0.8 个百分点, *FPS* 基本不变. 由数据可知,上述所设计模块运用于地铁场景小行人检测中,模型检测 *AP* 都实现了进一步提升,但 *FPS* 有小幅度降低,分析原因是改进 SSD 算法相较于 SSD 算法在结构上增加了一系列模块,网络模型复杂度提升和计算量增大,一定程度影响模型 *FPS*.

为定性分析改进 SSD 算法的检测性能,选用 SSD 算法、YOLO-v5 算法、YOLOX 算法和改进 SSD 算法对地铁场景不同尺度行人目标进行检测,检测结果如图 10 所示. 其中图 10(a)为待检测的原始图像,图 10(b)~(e)为各类算法检测效果图.

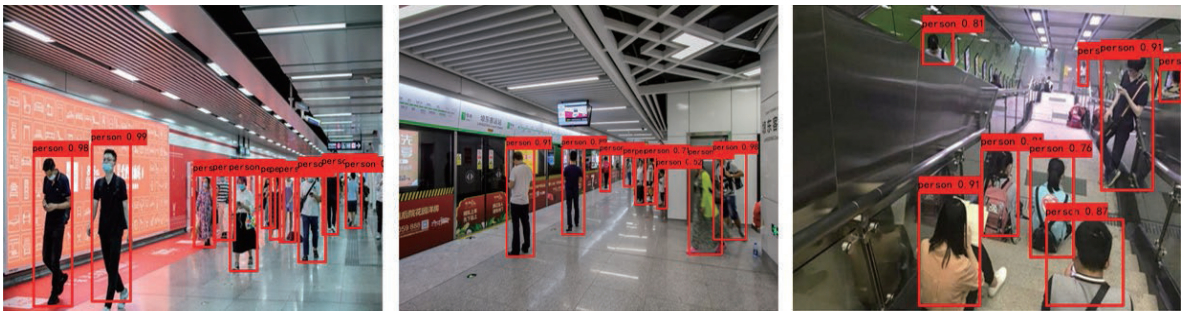
由图 10(b)可以看出,改进 SSD 算法对地铁场景



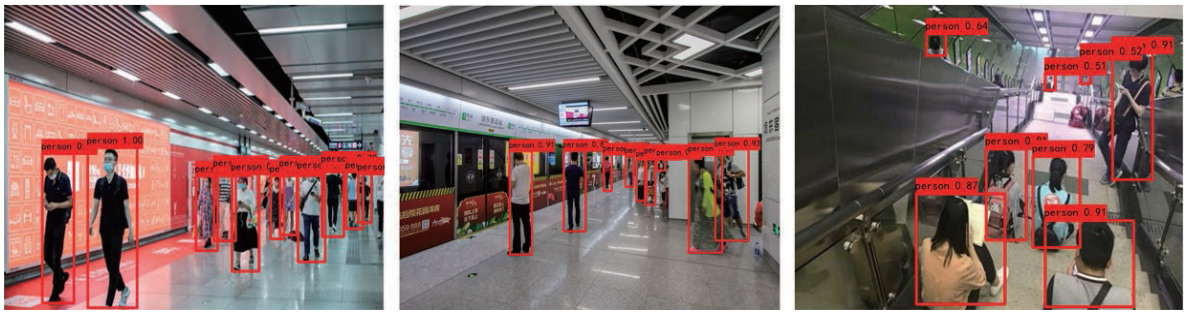
(a) 原始图像



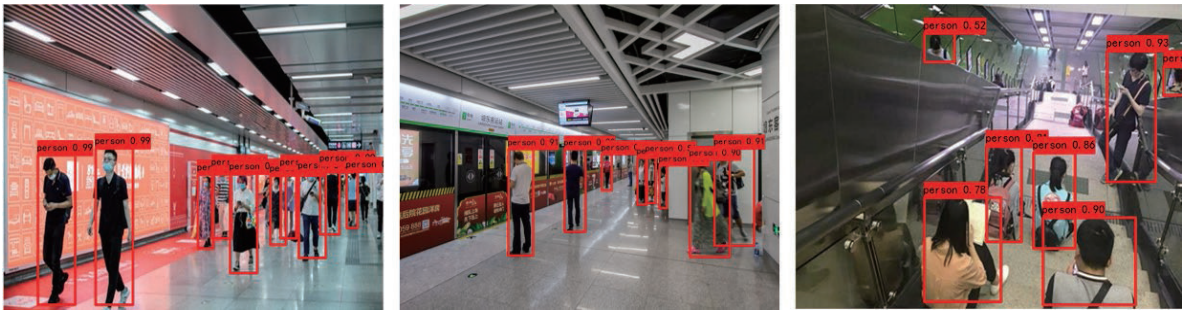
(b) 改进SSD检测效果



(c) YOLOX检测效果



(d) SSD检测效果



(e) YOLO-v5检测效果

Fig. 10 Pedestrian detection effects of different models

图 10 不同模型行人检测效果

小行人目标检测在置信度、目标定位和检测数量方面都优于其他算法. 对于不同程度的遮挡小行人目标和不同姿态的小行人目标, 改进 SSD 算法几乎都能精准定位其具体位置, 小行人目标检测的置信度平均在 0.60 以上. 在对大尺度和中等尺度行人目标检测中, 改进 SSD 算法同样具有较高置信度和精确定位框.

由图 10(d)可以看出, SSD 算法对于大尺度和中等尺度行人目标检测也表现良好, 但对于一些小行人目标, SSD 算法不能完全检测出此类目标, 对于已检测出的小行人目标, 模型所输出的定位框与目标真实位置有些偏移. 由图 10(c)和图 10(e)可以看出, YOLOX 算法和 YOLO-v5 算法几乎能检测出全部大尺度和中等尺度行人目标, 而对于一些遮挡行人目标和小行人目标其检测效果有所下降, 对于模糊小行人目标, 这 2 种算法几乎检测不出. 因此, 本文提出的改进 SSD 算法对于地铁场景小行人目标的检测效果优于其他对比算法.

4 结 论

本文提出一种改进 SSD 算法用于加强地铁场景小行人目标检测, 模型加入金字塔特征加强模块, 运用亚像素卷积实现图像低分辨率到高分辨转变以实现上采样操作, 同时利用多分支残差单元增大卷积层局部感受野, 实现对输出特征图多尺度融合和拼接, 捕获图像全局有效特征. 模型加入上下文特征融合模块, 生成面向小行人目标检测的拓展特征层, 通过双维度卷积注意力机制使模型对地铁小行人目标特征更加关注. 利用基于 Anchor-free 的动态正负样本分配策略为小行人目标分配最优训练正样本, 以提高模型检测召回率. 实验结果分析表明, 改进 SSD 算法可有效增强地铁场景小行人目标检测效果, 小行人目标检测的召回率和平均准确率相对于 SSD 算法分别提高 0.7 个百分点和 1.2 个百分点. 实际地铁场景测试中, 改进模型对遮挡严重的小行人目标检测漏检率仍较高, 同时对小行人目标检测有一定概率错检现象. 未来将进一步优化网络模型来提高模型性能.

作者贡献声明: 张秀再指导与修改论文; 邱野改进 SSD 算法的设计与实现模型, 以及撰写论文; 沈涛负责相关文献的收集和论文的校对.

参 考 文 献

[1] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proc of the 27th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 580−587

[2] Girshick R. Fast R-CNN[C]//Proc of the 15th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2015: 1440−1448

[3] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proc of the 29th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 779−788

[4] Liu Wei, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]//Proc of the 14th European Conf on Computer Vision. Berlin: Springer, 2016: 21−37

[5] Qi Linglong, Gao Jianling. Small object detection based on improved YOLOv7[J]. Computer Engineering, 2023, 49(1): 41−48 (in Chinese)
(戚玲珑, 高建瓴. 基于改进 YOLOv7 的小目标检测[J]. 计算机工程, 2023, 49(1): 41−48)

[6] Li Zihao, Wang Zhengping, He Yuntao. Aerial photography dense small target detection algorithm based on adaptive cooperative attention mechanism[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(8): 327944 (in Chinese)
(李子豪, 王正平, 贺云涛. 基于自适应协同注意力机制的航拍密集小目标检测算法[J]. 航空学报, 2023, 44(8): 327944)

[7] Li Yadong, Ma Xing, Mu Chunyang, et al. Improved the small target detection method of bearing defects in YOLOX network[J]. Computer Engineering and Applications, 2023, 59(1): 100−107 (in Chinese)
(李亚东, 马行, 穆春阳. 改进 YOLOX 网络的轴承缺陷小目标检测方法[J]. 计算机工程与应用, 2023, 59(1): 100−107)

[8] Yi Lin, Dollar P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proc of the 30th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 2117−21255

[9] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proc of the 16th European Conf on Computer Vision. Berlin: Springer, 2018: 3−19

[10] Szegedy C, Liu Wei, Jia Yangqing, et al. Going deeper with convolutions[C]//Proc of the 28th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 1−9

[11] Yu Donghang, Xu Qing, Guo Haitao, et al. Anchor-free arbitrary-oriented object detector using box boundary-aware vector[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15(1): 2535−2545

[12] Cui Lisha, Ma Rui. MDSSD: Multi-scale deconvolutional single shot detector for small objects[J]. arXiv preprint, arXiv: 1805.07009, 2018

[13] Krishna H, Jawahar C V. Improving small object detection[C]//Proc of the 4th IEEE Asian Conf on Pattern Recognition. Piscataway, NJ:

IEEE, 2017: 340−345

[14] Brauwers G, Frasincar F. A general survey on attention mechanisms in deep learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(4): 3279−3298

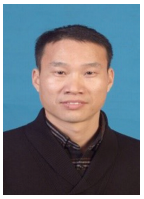
[15] Hou Qibin, Zhou Daquan, Feng Jiashi. Coordinate attention for efficient mobile network design[C]//Proc of the 34th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 13713−13722

[16] Fu Chengyang, Liu Wei. DSSD: Deconvolutional single shot detector[J]. arXiv preprint, arXiv: 1701.06659, 2017

[17] Kisantal M, Wojna Z, Murawski J, et al. Augmentation for small object detection[J]. arXiv preprint, arXiv: 1902.07296, 2019

[18] Zhao Gewei, Xu Shengquan. DL-MAML: An innovative model for automatically identifying butterfly species[J]. Journal of Computer Research and Development, 2024, 61(3): 674−684 (in Chinese)
（赵戈伟, 许升全. DL-MAML: 一种新的蝴蝶物种自动识别模型[J]. 计算机研究与发展, 2024, 61(3): 674−684）

[19] Benjumea A, Teeti I, Cuzzolin F, et al. YOLO-Z: Improving small object detection in YOLO-v5 for autonomous vehicle[J]. arXiv preprint, arXiv: 2112.11798, 2021



Zhang Xiuzai, born in 1979. PhD. His main research interests include meteorological communication technology and security, and machine learning.

张秀再, 1979 年生. 博士. 主要研究方向为气象通信技术与安全、机器学习.



Qiu Ye, born in 1995. Master. His main research interests include image processing and object detection.

邱　野, 1995 年生. 硕士. 主要研究方向为图像处理、目标检测.



Shen Tao, born in 2000. Master. His main research interests include deep learning and object detection.

沈　涛, 2000 年生. 硕士. 主要研究方向为深度学习、目标检测.