智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

基于改进YOLOv5s的面向自动驾驶场景的道路目标检测算法

胡丹丹, 张忠婷

引用本文:
胡丹丹, 张忠婷. 基于改进YOLOv5s的面向自动驾驶场景的道路目标检测算法[J]. 智能系统学报, 2024, 19(3): 653-660.
HU Dandan, ZHANG Zhongting. Road target detection algorithm for autonomous driving scenarios based on improved YOLOv5s[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(3): 653-660.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202206034

DOI: 10.11992/tis.202206034
网络出版地址: https://link.cnki.net/urlid/23.1538.TP.20230913.1825.004

基于改进 YOLOv5s 的面向自动驾驶场景的道路目标检测算法

胡丹丹, 张忠婷
(中国民航大学 机器人研究所, 天津 300300)

摘　要: 在复杂道路场景中检测车辆、行人、自行车等目标时, 存在因多尺度目标及部分遮挡易造成漏检及误检等情况, 提出一种基于改进 YOLOv5s 的面向自动驾驶场景的道路目标检测算法。首先, 利用深度可分离卷积替换部分普通卷积, 减少模型的参数量以提升检测速度。其次, 在特征融合网络中引入基于感受野模块(receptive field block, RFB)改进的 RFB-s, 通过模仿人类视觉感知, 增强特征图的有效感受野区域, 提高网络特征表达能力及对目标特征的可辨识性。最后, 使用自适应空间特征融合(adaptively spatial feature fusion, ASFF)方式以提升 PANet 对多尺度特征融合的效果。实验结果表明, 在 PASCAL VOC 数据集上, 所提算法检测平均精度均值相较于 YOLOv5s 提高 1.71 个百分点, 达到 84.01%, 在满足自动驾驶汽车实时性要求的前提下, 在一定程度上减少目标检测时的误检及漏检情况, 有效提升模型在复杂驾驶场景下的检测性能。

关键词: YOLOv5s; 自动驾驶; 目标检测算法; 深度可分离卷积; 感受野模块; 自适应空间特征融合; PANet; 多尺度特征融合
中图分类号: TP391.4　　**文献标志码**: A　　**文章编号**: 1673−4785(2024)03−0653−08

中文引用格式: 胡丹丹, 张忠婷. 基于改进 YOLOv5s 的面向自动驾驶场景的道路目标检测算法 [J]. 智能系统学报, 2024, 19(3): 653–660.
英文引用格式: HU Dandan, ZHANG Zhongting. Road target detection algorithm for autonomous driving scenarios based on improved YOLOv5s[J]. CAAI transactions on intelligent systems, 2024, 19(3): 653–660.

Road target detection algorithm for autonomous driving scenarios based on improved YOLOv5s

HU Dandan, ZHANG Zhongting
(Robotics Institute, Civil Aviation University of China, Tianjin 300300, China)

Abstract: When vehicles, pedestrians, bicycles, and other targets are detected in complex road scenes, the existence of multiscale targets and partial occlusions may easily cause missed and false detections. In this paper, a road target detection algorithm is proposed based on improved YOLOv5s, orienting to autonomous driving scenarios. First, depthwise separable convolution is used to replace partial ordinary convolutions to reduce the number of parameters of the model to improve the detection speed. An improved RFB-s based on receptive field block (RFB) is introduced into the feature fusion network to enhance the effective receptive field area of the feature map, improving the network feature expression capability and the recognizability of the target features by imitating human visual perception. Finally, an adaptive spatial feature fusion method is used to enhance the effect of PANet on multiscale feature fusion. The experimental results reveal that, on the PASCAL VOC dataset, compared with YOLOv5s, the mean value of the average detection precision of the proposed algorithm is improved by 1.71%, reaching 84.01%. Under the premise of meeting the real-time requirement of autonomous driving vehicles, this algorithm has reduced false and missed detections in the target detection to a certain extent, effectively improving the detection performance of the model in complex driving scenarios.

Keywords: YOLOv5s; autonomous driving; target detection algorithm; depthwise separable convolution; receptive field block; adaptive spatial feature fusion; PANet; multiscale feature fusion

复杂道路场景中的车辆、行人、自行车等目标的检测效果直接关系到智能汽车的行驶决策, 对其进行准确定位和类别检测能够为行驶中的车辆提供信息, 保障车辆安全行驶。实时和鲁棒的

收稿日期: 2022−06−21.　**网络出版日期**: 2023−09−14.
基金项目: 中央高校基本科研业务项目 (3122022PY17, 3122017003); 天津市科技计划项目 (17ZXHLGX00120).
通信作者: 胡丹丹. E-mail: ddhu@cauc.edu.cn.

目标检测算法可以有效避免交通事故的发生，提高汽车行驶的安全性能。

计算机视觉和深度学习的快速发展，可以有效弥补基于人工提取特征进行目标检测时存在的检测精度低、易受环境干扰以及泛化能力不强等缺点，已经取得了一些显著效果。

目前基于深度学习的目标检测算法主要分为两类：1）R-CNN[1]、Fast R-CNN[2]、Faster R-CNN[3]、R-FCN[4]、Mask R-CNN[5] 等双步目标检测算法，大都先使用区域候选网络 (region proposal network, RPN) 生成一个可能包含待检测目标的候选框，再利用卷积神经网络提取特征完成对候选目标的位置和类别的预测和识别。2）SSD[6]、DSSD[7]、FSSD[8]、YOLO[9]、YOLO9000[10]、YOLOv3[11]、YOLOv4[12]、YOLOv5[13] 和 EfficientDet[14] 等单步目标检测算法，直接通过卷积神经网络提取特征来产生目标的位置和类别信息，将检测转化为回归问题，是一种端到端的目标检测算法，具有更快的检测速度。

已有一些研究工作将两类检测算法应用到自动驾驶场景中的目标检测问题中。例如，陈泽等[15] 在 Faster R-CNN 检测算法中引入基于双线性插值的对齐池化层，减少量化操作中的像素偏差。设计了基于级联的特征融合策略，实现特征复用，能够有效提高对小尺度行人的检测性能。郁强等[16] 提出一种多尺度 YOLOv3 的道路场景目标检测算法，新增两个面向小目标的特征输出模块，较好地解决小尺度目标的检测问题，同时又没有影响大目标的检测。

现有单步、双步目标检测算法在面向驾驶场景下的目标检测中已取得了一些研究进展，但是减少多尺度差异及遮挡情况下目标检测的误检及漏检仍是迫切需要解决的问题。此外，在自动驾驶场景中，目标检测算法需要对道路目标做出准确的响应且及时回传到汽车的控制系统中，即需要同时具备较高的实时性和准确性。YOLOv5s 算法在检测精度和速度两方面有较好的平衡性，但是其检测精度仍存在很大的改进空间以提高实际问题中目标检测的准确率。因此，针对上述问题，本文以 YOLOv5s 算法[13] 为基础模型进行改进，提出一种基于感受野增强和多尺度特征融合的道路目标检测算法——YOLOv5s-RFB-s-ASFF。

1　YOLOv5s-RFB-s-ASFF 检测算法应用框架

YOLOv5s-RFB-s-ASFF 算法检测过程如图 1 所示，分为离线训练阶段和在线测试阶段两部分：1）离线训练阶段将数据集的样本划分为训练样本集和测试样本集，利用 YOLOv5s 的自适应图像缩放功能将输入图像尺寸自动缩放至 640 像素×640 像素的大小；对训练样本集进行标注，送入到 YOLOv5s-RFB-s-ASFF 检测网络中进行训练，得到模型预训练权重。2）在线测试阶段，利用搭载了视觉相机的无人驾驶平台进行采集图像，使用得到的模型权重对待检测的图像进行测试验证。

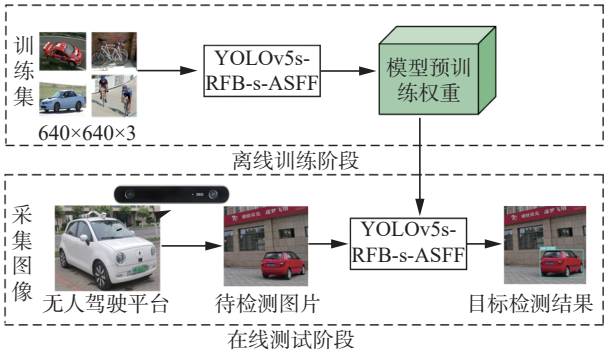


图 1　YOLOv5s-RFB-s-ASFF 算法检测过程

Fig. 1　YOLOv5-sRFB-s-ASFF algorithm detection process

2　改进 YOLOv5s 的目标检测方法

本文提出的基于感受野增强和多尺度特征自适应融合的道路目标检测方法—— YOLOv5s-RFB-s-ASFF，其网络框架如图 2 所示，由 4 部分组成：

1）输入端 (Input) 对数据进行 mosaic 数据增强、自适应锚框计算和自适应图片缩放等预处理操作。

2）主干网络 (Backbone) 主要采用 Focus 结构、跨阶段局部结构及空间金字塔池化等深度卷积操作从图像中提取不同层次的特征。

3）颈部部分 (Neck) 由特征金字塔网络和路径聚合网络组成的 PANet 网络。本文在 Neck 部分引入深度可分离卷积精简原检测网络的计算参数，以提高检测速度；嵌入改进感受野模块 RFB-s，扩大有效感受野，增强网络对道路目标的特征表达能力及对目标特征的可辨识性；在预测网络前利用自适应特征融合结构 ASFF 融合不同尺度大小的浅层特征图和深层特征图的位置和类别信息，提高检测精度。

4）预测部分 (Prediction)：主要是在不同特征图上预测不同尺寸的目标，包括计算损失函数、NMS 非极大值抑制等。损失函数由回归框预测误差 L_{cla}、置信度误差 L_{conf}、目标类别损失函数 L_{cla} 这 3 部分组成。其中 L_{loc} 采用 GIoU[17] 函数来计

算, 计算公式为

$$\mathrm{GIoU} = I_{\mathrm{oU}} - \frac{|C-(A\cup B)|}{|C|} \tag{1}$$

式中: I_{oU}表示为预测框与真实框的交并比; A表示预测框; B表示真实框; C表示A、B最小包围框, 表示预测框与真实框的并集。

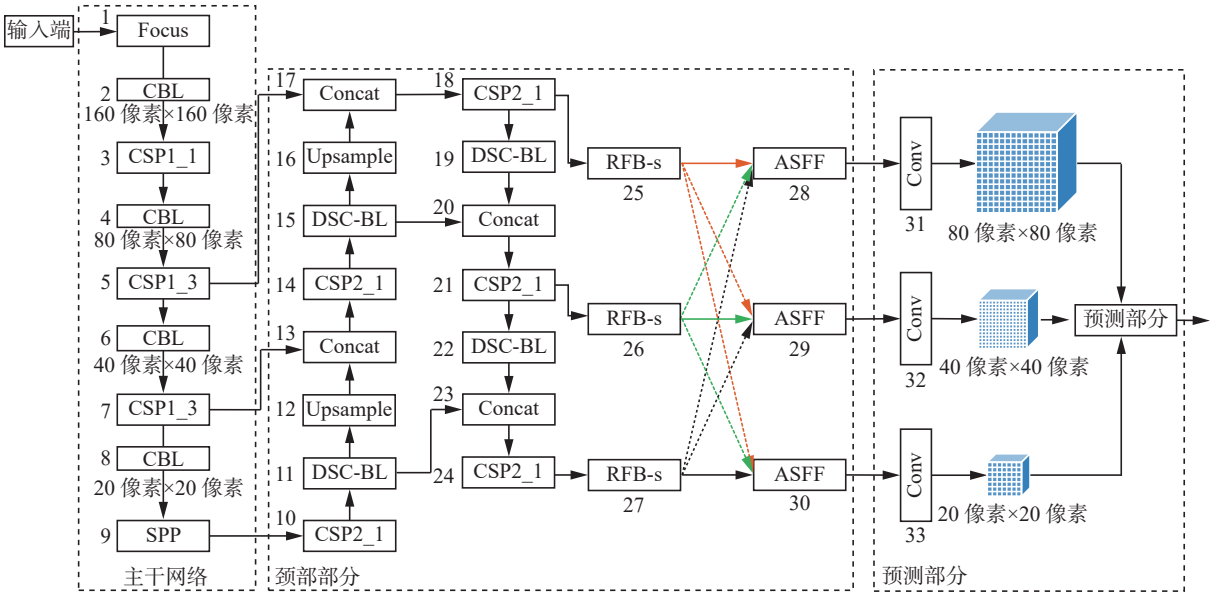


图 2　YOLOv5s-RFB-s-ASFF 算法网络框架

Fig. 2　YOLOv5s-RFB-s-ASFF algorithm network framework

2.1　深度可分离卷积

为有效降低模型参数量, 提高目标检测实时性, 使用深度可分离卷积将原始 YOLOv5s 的 Neck 部分第 11、15、19 和 22 层的普通卷积层替换为深度可分离卷积, 简化网络结构, 降低计算消耗内存。

深度可分离卷积 (depthwise separable convolution, DSC)[18] 进行卷积的过程可以分为两个步骤: 1)使用深度卷积在二维平面内对输入特征图进行逐通道的卷积, 得到与输入特征图通道相同的特征映射。2)使用尺寸大小为 1×1 的卷积对深度卷积操作生成的特征映射进行逐点卷积, 在深度方向上加权组合, 进行维度变换, 从而有效利用不同通道在相同空间位置上的特征信息。

设输入特征图尺寸为$\{D_{\mathrm{F}},D_{\mathrm{F}},M\}$, M为输入通道数, 卷积核大小为$D_{\mathrm{K}}\times D_{\mathrm{K}}$, 输出特征图尺寸为$\{D_{\mathrm{F}},D_{\mathrm{F}},N\}$, N为输出通道数。深度可分离卷积的计算量为$D_{\mathrm{K}}\cdot D_{\mathrm{K}}\cdot M\cdot D_{\mathrm{F}}\cdot D_{\mathrm{F}}+M\cdot N\cdot D_{\mathrm{F}}\cdot D_{\mathrm{F}}$, 普通卷积的计算量为$D_{\mathrm{K}}\cdot D_{\mathrm{K}}\cdot M\cdot N\cdot D_{\mathrm{F}}\cdot D_{\mathrm{F}}$。

$$\frac{D_{\mathrm{K}}\cdot D_{\mathrm{K}}\cdot M\cdot D_{\mathrm{F}}\cdot D_{\mathrm{F}}+M\cdot N\cdot D_{\mathrm{F}}\cdot D_{\mathrm{F}}}{D_{\mathrm{K}}\cdot D_{\mathrm{K}}\cdot M\cdot N\cdot D_{\mathrm{F}}\cdot D_{\mathrm{F}}}=\frac{1}{N}+\frac{1}{D_{\mathrm{K}}^2}<1 \tag{2}$$

由式(2)可以看出, 深度可分离卷积的计算量比普通卷积计算量大大减少, 占其计算量的$\frac{1}{N}+\frac{1}{D_{\mathrm{K}}^2}$, 当$D_{\mathrm{K}}$取 3 时, 即使用 3×3 卷积核时, 相比于普通卷积可以降低 8~9 倍的参数量, 加快了运算速度, 提高了检测效率, 降低了内存占用。

2.2　感受野增强模块

在颈部特征融合网络处引入基于感受野模块[19]改进的 RFB-s, 以不同大小尺寸的卷积核对特征图进行特征提取, 进一步提升网络的特征融合能力, 对有遮挡的相似物体进行更好的类别判断与区分。

如图 3 所示, 感受野模块 RFB 模拟人类视觉的感受野, 借鉴 Inception 网络[20], 使用了多分支结构。并引入空洞卷积, 在多分支结构上使用不同尺度的常规卷积和空洞卷积, 增大特征图的有效感受野区域, 增强特征分辨能力。

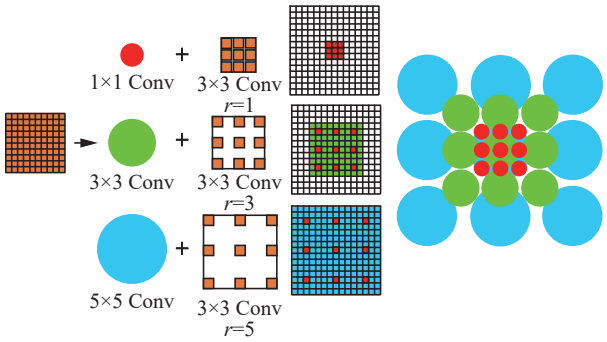


图 3　感受野模块特征提取示意

Fig. 3　Schematic diagram of feature extraction of perceptual field module

如图 4 所示, RFB 内部结构主要分为两部分。

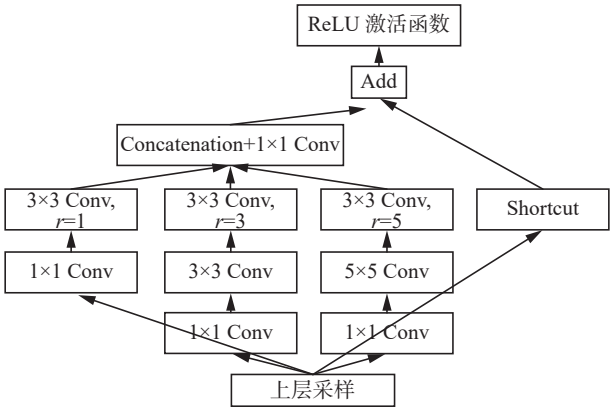


图 4　RFB 网络结构示意

Fig. 4　RFB network structure diagram

1)多分支卷积层

设计了一种多分支网络结构，包含 3 个分支，使用不同大小的卷积核，其特征提取能力优于使用相同尺寸大小卷积核的网络结构。

如图 4 所示，第 1 个分支是由 1×1 标准卷积、3×3 空洞卷积(扩张系数为 1)组成，第 2 个分支是由 1×1 标准卷积、3×3 标准卷积、3×3 空洞卷积(扩张系数为 3)组成，第 3 个分支是由 1×1 标准卷积、5×5 标准卷积、3×3 空洞卷积(扩张系数为 5)组成。此外，还借鉴了 ResNet 中的 shortcut 结构，通过直连的方式减轻深层网络的训练负担。

2)空洞卷积

空洞卷积在标准卷积层中加入了一个新的参数——扩张率 (dilation rate)，能够将卷积核扩张到规定的尺度大小，可以在参数量相同的情况，增大特征图的有效感受野，因此能够优化高分辨率图像相邻像素间的冗余信息。

如图 5 所示，基于 RFB 改进的 RFB-s 网络，用 3×3 的卷积层和代替较大的 5×5 的卷积层，用 1×3 和 3×1 的卷积层代替 3×3 的卷积层，分支变多，但卷积核的尺寸变小，减少了计算量。

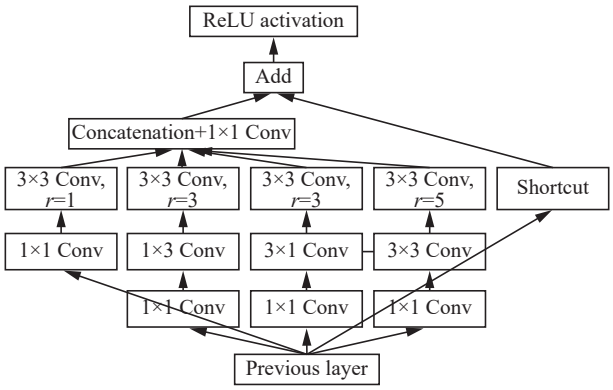


图 5　RFB-s 网络结构示意

Fig. 5　RFB-s network structure diagram

2.3　多尺度特征自适应融合

面向自动驾驶环境的目标检测，待检测的车辆、行人、自行车等目标存在尺度差异，且目标被遮挡后会出现轮廓模糊、部分特征缺失的情况，特征提取网络提取到的低级特征信息就会失效。因此需要增强网络模型对不同尺度特征的融合利用能力，从而增强网络的特征表达能力。

原始 YOLOv5s 框架采用 FPN+PAN 的结构，将特征图变换为相同尺度后，采用直接级联的方式对不同尺度的特征进行融合，对不同尺度特征的利用程度较低，容易造成网络检测精度低。

在 YOLOv5s 中加入自适应空间特征融合机制 ASFF[21]，通过给不同尺度特征分配自适应的权重参数，实现不同尺度特征的高效融合，自动学习并滤除其他层的无用信息，保留有用的信息。同时 ASFF 的融合过程是可微分的，可以通过标准的反向传播来学习并更新权重，具有实现方式简单，计算量小等优点。

在 YOLOv5s 中加入 ASFF，ASFF 结构如图 6 所示，改进后的 PANet 结构输出 3 个不同尺度的特征层分别记为 L_1、L_2 和 L_3，经过融合得到 3 个 ASFF 层分别对应 ASFF-1、ASFF-2 和 ASFF-3。

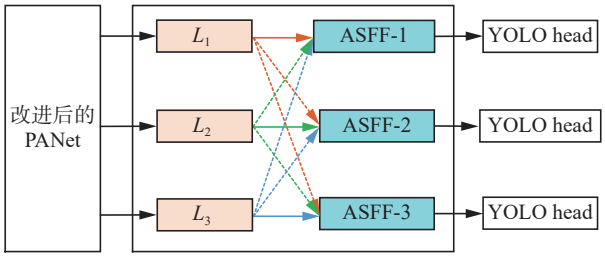


图 6　ASFF 结构

Fig. 6　ASFF structure

具体的实现步骤可以分为两步：

1) 不同尺度间的特征缩放：将输入层 $L_l(l \in \{1,2,3\})$ 尺度特征设为 X^l，将其他的特征层 $L_n(n \neq l)$ 的特征 X^n 调整到与 X^l 一样的尺度大小。

对于尺度小于给定尺度大小的特征图时，需要对该特征图进行上采样操作，使用 1×1 的点卷积将特征图的通道数压缩至与目标层 l 级相同的通道数，然后利用插值法来扩大特征图的尺寸大小。

对于尺度大于给定尺度大小的特征图时，需要对该特征图进行下采样操作，使用 3×3 的卷积核(步长为 2)进行下采样，将特征图调整为现在尺寸的 1/2，同时调整特征图的通道数和尺寸大小。使用最大池化层(步长为 2)加 3×3 的卷积层(步长为 2)进行 1/4 比例的下采样，实现特征图的尺度大小统一。

2) 自适应特征融合：通过尺度特征缩放后，

得到$\boldsymbol{X}^{1\to l}$、$\boldsymbol{X}^{2\to l}$、$\boldsymbol{X}^{3\to l}$标准化后的特征图, 沿通道方向进行拼接, 然后经过一个输出通道为 3 的 1×1 卷积层, 并利用 softmax 函数将输出调整为 [0,1] 之间, 最后得到 3 层特征图的自适应空间权重α_{ij}^{l}、β_{ij}^{l}、γ_{ij}^{l}, 每个特征图都与其对应的权重参数矩阵相乘, 然后采用对应元素值相加的方式得到最终的特征融合结果。

设$\boldsymbol{X}_{ij}^{n\to l}$表示由第$n$层的特征图调整尺度到第$l$层特征映射上$(i,j)$位置的特征向量, 对应第$l$级的特征融合结果如下:

$$\boldsymbol{y}_{ij}^{l}=\boldsymbol{X}_{ij}^{1\to l}\times\alpha_{ij}^{l}+\boldsymbol{X}_{ij}^{2\to l}\times\beta_{ij}^{l}+\boldsymbol{X}_{ij}^{3\to l}\times\gamma_{ij}^{l} \tag{3}$$

式中: $\boldsymbol{y}_{ij}^{l}$表示特征图y^{l}上(i,j)处的输出; α_{ij}^{l}、β_{ij}^{l}、$\gamma_{ij}^{l}\in[0,1]$, 是可学习的参数, 分别为在第l层特征图中学习到的权重, 满足以下关系式:

$$\alpha_{ij}^{l}+\beta_{ij}^{l}+\gamma_{ij}^{l}=1 \tag{4}$$

其中, α_{ij}^{l}、β_{ij}^{l}、γ_{ij}^{l}分别定义为

$$\alpha_{ij}^{l}=\frac{\mathrm{e}^{\lambda_{\alpha_{ij}}^{l}}}{\mathrm{e}^{\lambda_{\alpha_{ij}}^{l}}+\mathrm{e}^{\lambda_{\beta_{ij}}^{l}}+\mathrm{e}^{\lambda_{\gamma_{ij}}^{l}}} \tag{5}$$

$$\beta_{ij}^{l}=\frac{\mathrm{e}^{\lambda_{\beta_{ij}}^{l}}}{\mathrm{e}^{\lambda_{\alpha_{ij}}^{l}}+\mathrm{e}^{\lambda_{\beta_{ij}}^{l}}+\mathrm{e}^{\lambda_{\gamma_{ij}}^{l}}} \tag{6}$$

$$\gamma_{ij}^{l}=\frac{\mathrm{e}^{\lambda_{\gamma_{ij}}^{l}}}{\mathrm{e}^{\lambda_{\alpha_{ij}}^{l}}+\mathrm{e}^{\lambda_{\beta_{ij}}^{l}}+\mathrm{e}^{\lambda_{\gamma_{ij}}^{l}}} \tag{7}$$

式中: $\lambda_{\alpha_{ij}}^{l}$、$\lambda_{\beta_{ij}}^{l}$和$\lambda_{\gamma_{ij}}^{l}$由$\boldsymbol{X}^{1\to l}$、$\boldsymbol{X}^{2\to l}$、$\boldsymbol{X}^{3\to l}$分别经过 1×1 卷积得到, 然后$\lambda_{\alpha_{ij}}^{l}$、$\lambda_{\beta_{ij}}^{l}$和$\lambda_{\gamma_{ij}}^{l}$通过使用控制参数的 softmax 函数得到$\alpha_{ij}^{l}$、$\beta_{ij}^{l}$和$\gamma_{ij}^{l}$, 将其输出调整为 [0,1]。

3　实验结果与分析

3.1　实验环境及数据集

为了验证本文提出的 YOLOv5s-RFB-s-ASFF 算法的性能, 在 Pytorch 框架下对其进行训练与测试, 训练平台使用 Python 3.8.0 进行编译和测试, 处理器为 Intel(R) Core(TM) i7-10750H, 显卡为 NVIDIA GTX 1650Ti, 操作系统为 Ubuntu 18.04, 采用 11.1 版本的 CUDA, 计算机视觉库为 Python-OpenCV 4.4.0。初始学习率设为 0.01, batch size 为 4。

在 PASCAL VOC 目标检测公开数据集上进行对比实验, 将 PASCAL VOC 2007 和 PASCAL VOC 2012 的训练集进行合并, 作为 YOLOv5s-RFB-s-ASFF 模型的训练集, 测试集为 PASCAL VOC 2007 测试集部分。数据集中已经人工标注好了待检测目标的真实框的位置、大小及类别信息。数据集中的目标总共可以分为 4 大类: vehicle、household、animal、person, 细分后有 20 个类别的物体。

3.2　评价指标

为了对本文提出的 YOLOv5s-RFB-s-ASFF 目标检测算法的性能进行定量及综合的评价, 采用 mAP@0.5 和检测速度作为评价标准。

平均精度均值 (mAP) 反映目标检测精度, mAP@0.5 代表在 IoU 阈值为 0.5 时的平均精度均值。检测速度为每秒能够检测的图像帧数, 反映目标检测速度。检测结果可以分为 4 种: 真正例 TP、假正例FP、真负例TN、假负例FN。

精确率P表示网络检测的所有目标中为正类个数的比例, 召回率R表示网络检测的正类占数据集所有正类的比例, 计算公式为

$$P=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FP}}}\times100\% \tag{8}$$

$$R=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FN}}}\times100\% \tag{9}$$

平均精度P_{A}和平均精度均值P_{mA}的计算公式为

$$P_{\mathrm{A}}=\int_{0}^{1}P(R)\mathrm{d}R \tag{10}$$

$$P_{\mathrm{mA}}=\frac{1}{n}\sum_{i=1}^{n}P_{A,i} \tag{11}$$

式中: n为数据集包含的类别总数, $P_{A,i}$为第i个类别检测的平均准确率。

3.3　实验结果与分析

3.3.1　客观分析

为了验证本文提出的 YOLOv5s-RFB-s-ASFF 算法的性能, 分别统计网络的 mAP@0.5 和检测速度, 不同算法性能指标对比如表 1 所示。

表 1　不同算法性能指标对比

Table 1　Comparison of performance metrics of different algorithms

算法	输入尺寸/(像素×像素)	主干网络	mAP @0.5/%	检测速度/(f/s)
Fast R-CNN	600×1 000	VGG16	70.00	0.5
Faster R-CNN	600×1 000	VGG16	73.20	7
SSD	512×512	VGG16	79.80	19
DSSD	513×513	ResNet-101	81.50	5.5
YOLOv3	416×416	Darknet53	79.26	26.7
YOLOv4	416×416	CSP-Darknet53	83.64	66

续表 1

算法	输入尺寸/(像素×像素)	主干网络	mAP @0.5/%	检测速度/(f/s)
YOLOv5s	640×640	CSP-Darknet53	82.30	76
YOLOv5s-RFB-s-ASFF	640×640	CSP-RFB-s-ASFF	84.01	61

从表 1 的对比算法检测结果可以看出，本文提出的 YOLOv5s-RFB-s-ASFF 算法的 mAP@0.5 高于其他对比算法，相比于原始 YOLOv5s 算法的 mAP@0.5 值提高了 1.71 个百分点，RFB-s 感受野增强模块及 ASFF 机制的加入，使得模型能够更加充分利用低层次的特征信息，改善了被遮挡目标的检测效果，有效减少了漏检及误检情况，获得了较高的检测精度。由于 RFB-s 及 ASFF 模块的加入，增加了一定的计算量，采用深度可分离卷积替换部分普通卷积，使得本文算法的检测速度虽比原始算法的检测速度下降了 15 f/s，但仍保持较高的实时性，较大幅度提升了模型的性能。

3.3.2　消融实验

本文算法在 YOLOv5s 算法的基础上采用了多个改进策略，为了验证不同模块改动和不同模块组合的改进策略的有效性，设计了消融实验进行对比研究，实验结果如表 2 所示。从表 2 可以看出，深度可分离卷积替换普通卷积后，相较于原始 YOLOv5s 算法，模型 mAP@0.5 降低了 1.04 个百分点，但是检测速度明显提高；RFB-s 模块的添加，提高了算法的精度，mAP@0.5 由 81.26% 提升到 83.80%，证明了增强感受野对于特征提取的重要性；在加入 RFB-s 的基础上加入特征自适应融合模块 ASFF，mAP 提高了 0.21 个百分点，证明了 ASFF 模块可以更有效地利用浅层和高层特征之间的联系，使其获得更高的准确率。

表 2　消融实验

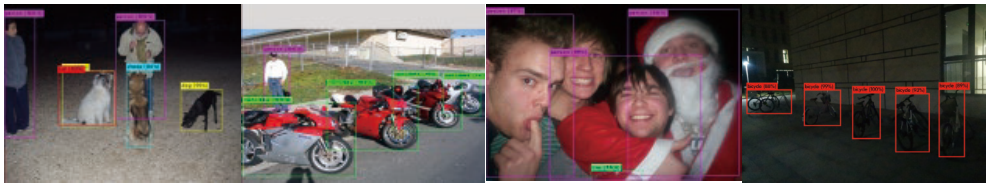
Table 2　Ablation experiments

组别	DSC	RFB-s	ASFF	mAP@0.5/%	检测速度/(f/s)
1				82.30	76
2	√			81.26	84
3		√	√	83.80	73
4		√	√	84.01	61

3.3.3　可视化分析

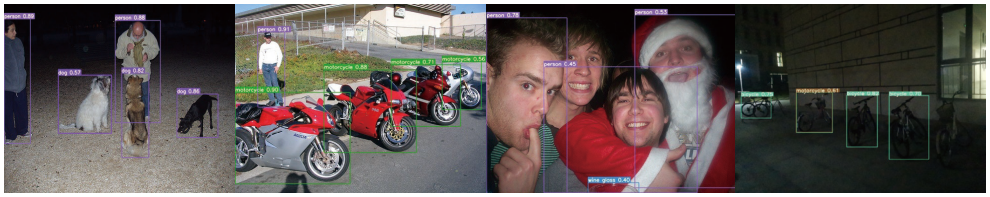
为了进一步直观的验证本文提出的 YOLOv5s-RFB-s-ASFF 算法的性能，选取具有代表性的 4 种场景对模型进行验证，分别为：多尺度且多种类待检测目标、多种类被遮挡目标、单一种类被遮挡目标、实际黑夜校园环境。实验中，将本文算法与 YOLOv3、YOLOv4、YOLOv5s 几种目标检测算法进行对比，可视化结果对比如图 7 所示，从左到右 4 组实验结果及分析如下。

(a) 原图

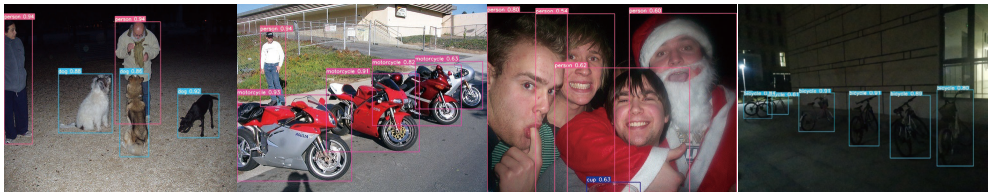


(b) YOLOv3 算法检测结果

(c) YOLOv4 算法检测结果



(d) YOLOv5s 算法检测结果



(e) YOLOv5s-RFB-s-ASFF 算法检测结果

图 7　可视化结果对比图

Fig. 7　Visualization results comparison chart

第 1 组实验: 在昏暗环境中, 存在两种类别的待检测目标。从第 1 组的实验对比结果可以看出, 所有的检测算法均可以检测出待检测目标, 但是 YOLOv3 算法存在误检现象, 将其中一只白色的狗误检为猫, 将图像中间位置的狗误检为羊, YOLOv4 与 YOLOv5s-RFB-s-ASFF 检测精度率高于 YOLOv5s。

第 2 组实验: 待检测目标为两种类别且存在严重遮挡。从第 2 组的实验对比结果可以看出, YOLOv4 算法检测结果中存在一辆摩托车漏检, 而其他 3 种算法均能够正确检测出待检测目标, YOLOv5s-RFB-s-ASFF 算法的检测精度与 YOLOv3 相当, 且高于 YOLOv5s, 由于深度可分离卷积的加入, 本文算法检测速度比 YOLOv3 快。

第 3 组实验: 单一种类且目标存在遮挡。从第 3 组实验对比结果可以看出, 改进感受野增强模块 RFB-s 的加入使本文算法能够增大有效感受野, 准确检测出所有待检测目标, 而其余算法均出现了漏检情况。

第 4 组实验: 待检测目标为实际校园环境中存在遮挡的自行车, 且图像中自行车尺度有差异。YOLOv3、YOLOv4 算法存在漏检情况, 只检测到 5 个待检测目标, YOLOv5s 只检测到 4 个待检测目标且还存在误检情况, 将其中一辆自行车误检为摩托车, 本文算法可以准确检测出所有的 6 个待检测的自行车目标。

综合而言, 相较 YOLOv3、YOLOv4、YOLOv5s 算法, 本文提出的 YOLOv5s-RFB-s-ASFF 算法目标识别误检及漏检率低, 识别精度更高, 而且实时性有保证。

4　结束语

1)提出了一种基于 YOLOv5s 改进的目标检测算法——YOLOv5s-RFB-s-ASFF 算法, 在识别精度及实时性方面表现性能较优。利用深度可分离卷积替换部分普通 3×3 卷积层, 在一定程度上减少了模型的参数量; 加入改进感受野增强模块 RFB-s, 使网络更多地捕获靠近感受野中心区域的信息, 从而生成更大感受野的特征图, 进而提高目标的检测精度; 基于 ASFF 构建多尺度特征自适应融合机制, 能更有效表达待检测目标特征, 减少了由于特征表达能力不足而造成的漏检和误检现象。

2)实际实验表明本文方法能够较好地解决多尺度特征及目标遮挡的检测问题, 改进后的算法实时检测速率相较于原始算法有所降低, 但仍满足实时性要求, 该算法可以很好地服务于驾驶场景下的目标检测任务中。下一步将制作驾驶场景下的数据集, 大幅增加目标数量, 验证 YOLOv5s-RFB-s-ASFF 算法在目标密集场景下目标检测的能力。

参考文献:

[1] 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. Columbus: ACM, 2014: 580–587.

[2] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440–1448.

[3] REN Shaoqing, HE Kaiming, 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.

[4] DAI Jifeng, LI Yi, HE Kaiming, et al. R-FCN: object detection via region-based fully convolutional networks[C]// Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona: ACM, 2016: 379–387.

[5] HE Kaiming, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2980–2988.

[6] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//European Conference on Computer Vision. Cham: Springer, 2016: 21–37.

[7] FU Chengyang, LIU Wei, RANGA A, et al. DSSD: deconvolutional single shot detector[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii: IEEE, 2017: 2881−2890.

[8] LI Zuoxin, YANG Lu, ZHOU Fuqiang. FSSD: feature fusion single shot multibox detector[EB/OL]. (2017−12−04)[2022−06−20]. http://arxiv.org/abs/1712.00960.

[9] 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. Las Vegas: IEEE, 2016: 779–788.

[10] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 6517–6525.

[11] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. (2018−04−08)[2022−06−20]. http://arxiv.org/abs/1804.02767.

[12] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[J]. (2020−04−23)[2022−06−20]. https://arxiv.org/abs/2004.10934.

[13] GLENN J. Ultralytics. YOLOv5[EB/OL]. (2020−06−03)[2021−04−15]. https://github.com/ultralytics/yolov5.

[14] TAN Mingxing, PANG Ruoming, LE Q V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 10778–10787.

[15] 陈泽, 叶学义, 钱丁炜, 等. 基于改进 Faster R-CNN 的小尺度行人检测 [J]. 计算机工程, 2020, 46(9): 226–232, 241.
CHEN Ze, YE Xueyi, QIAN Dingwei, et al. Small-scale pedestrian detection based on improved faster R-CNN[J]. Computer engineering, 2020, 46(9): 226–232,241.

[16] 郁强, 王宽, 王海. 一种多尺度 YOLOv3 的道路场景目标检测算法 [J]. 江苏大学学报(自然科学版), 2021, 42(6): 628–633,641.
YU Qiang, WANG Kuan, WANG Hai. A multi-scale YOLOv3 detection algorithm of road scene object[J]. Journal of Jiangsu University (natural science edition), 2021, 42(6): 628–633,641.

[17] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 658–666.

[18] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 1800–1807.

[19] LIU Songtao, HUANG Di, WANG Yunhong. Receptive field block net for accurate and fast object detection[C]//Computer Vision – ECCV 2018: 15th European Conference. Munich: ACM, 2018: 404–419.

[20] SZEGEDY C, LIU Wei, JIA Yangqing, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 1–9.

[21] LIU Songtao, HUANG Di, WANG Yunhong. Learning spatial fusion for single-shot object detection[EB/OL]. (2019−11−21)[2022−06−20]. http://arxiv.org/abs/1911.09516.

作者简介：

胡丹丹, 副教授, 主要研究方向为机器人环境感知、多传感器数据融合。申请发明专利 30 余项, 发表学术论文 20 余篇。E-mail: ddhu@cauc.edu.cn。

张忠婷, 硕士研究生, 主要研究方向为无人驾驶车辆环境感知, 被评为校级优秀研究生, 曾获国家励志奖学金, 华北五省(市、自治区)大学生机器人大赛类人机器人竞技体育赛(投篮)竞赛项目一等奖。E-mail: 1113276573@qq.com。