

基于 MobileNet 的多尺度感受野特征融合算法

黄裕凯, 王青旺*, 沈韬**, 朱艳, 宋健

昆明理工大学信息工程与自动化学院, 云南 昆明 650500

摘要 针对轻量化网络在目标检测中检测精度低的问题,提出了一种以 MobileNet 为基础网络的轻量级目标检测网络 MobileNet-RFB-ECA。针对目标多尺度特性,采用基于轻量化扩充感受野模块(RFB)的特征金字塔网络结构增强网络对目标多尺度特性的适应性。与此同时,针对复杂注意力模块导致计算量大的问题,在主干特征提取网络添加有效通道注意力机制模块(ECA),提高卷积神经网络的性能。实验结果表明,相较于 MobileNet,所提 MobileNet-RFB-ECA 在 PASCAL VOC 数据集和 KITTI 数据集上检测精度分别提高了 4.2 个百分点和 15.4 个百分点,模型大小分别为 50.3 MB 和 48.5 MB,平均检测速度为 34 frame·s⁻¹。

关键词 图像处理; 目标检测; 轻量化神经网络; 多尺度特征融合; MobileNet; RFB-Net 模型; 有效注意力机制

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP220628

Multi-Scale Receptive Field Feature Fusion Algorithm based on MobileNet

Huang Yukai, Wang Qingwang*, Shen Tao**, Zhu Yan, Song Jian

Faculty of Information Engineering and Automation, Kunming University of Science and Technology,
Kunming 650500, Yunnan, China

Abstract To address the problem of low target detection accuracy in lightweight networks, a lightweight target detection network MobileNet-RFB-ECA based on MobileNet is proposed. To consider the multi-scale characteristics of the target, this study proposes a feature pyramid network structure based on the lightweight extended receptive field block (RFB), which enhances the adaptability of the network to the multi-scale characteristics of the target. Moreover, owing to the large computation caused by the complex attention module, an efficient channel attention (ECA) module is added to the backbone feature extraction network to improve the performance of the convolutional neural network. Experiments reveal that compared with conventional MobileNet, the proposed method improves the detection accuracy by 4.2 percentage points and 15.4 percentage points on the PASCAL VOC and KITTI datasets, respectively. In addition, the model sizes of the proposed method are 50.3 and 48.5 MB for the aforementioned datasets, respectively, and the average detection speed achieved is 34 frame/s.

Key words image processing; object detection; lightweight neural network; multi-scale feature fusion; MobileNet; RFB-Net; efficient channel attention module

1 引言

随着计算机技术、人工智能、数字图像处理的发展,目标检测在我们生活中运用广泛。例如:在遥感图像领域,目标检测在环境监测、空中侦察、安全防护等方面发挥很大作用^[1];在医疗领域,通过分析图像上的特征来辅助病理诊断^[2];在自动驾驶领域,通过对行人、车辆、道路等物体进行检测来协助车辆自动驾

驶^[3]。目标检测技术在其他领域,如人脸检测、智能交通、军事领域等也有着广泛应用。

深度学习作为近几年的热门话题,在许多领域都展现出了它的优势,并且在光学图像领域中越来越被重视,基于深度学习的目标检测应用得越来越广泛^[4]。现阶段使用深度学习的目标检测方法主要分为一阶段检测和二阶段检测:一阶段目标检测无需生成候选框,直接生成目标边界框和分类候选框;二阶段目标检测

收稿日期: 2022-01-25; 修回日期: 2022-03-09; 录用日期: 2022-03-30; 网络首发日期: 2022-04-10

基金项目: 国家自然科学基金(61971208)、云南省中青年学术技术带头人后备人才(2019HB005)、云南省重大科技专项(202002AB080001-8)、云南省基础研究计划(202101BE070001-008)

通信作者: *786120585@qq.com; **shentao@kust.edu.cn

先生成候选框,再对候选框进行分类预测和边界调整。一阶段目标检测的代表算法主要有 YOLO^[5]、YOLO9000^[6]、SSD^[7]等,两阶段目标检测的代表算法主要有 R-CNN^[8]、Fast R-CNN^[9]等。Redmon 等^[5]提出的 YOLO 算法能够快速地进行端到端检测,但不能很好地学习特征,检测精度较差。Liu 等^[7]提出的 SSD 算法实现了对图像的高速回归检测,但不能对小目标进行很好的识别。Ren 等^[10]在 R-CNN 的基础上提出了 Faster R-CNN 算法,显著提高检测速度与检测精度,但该算法耗费的计算资源仍较大,难以部署到移动设备上。Lin 等^[11]提出了一种多尺度特征金字塔网络(FPN),FPN 将网络中上层特征语义信息和下层特征位置信息自上而下进行特征融合,能够有效提高小目标检测精度。Liu 等^[12]通过模拟人类视觉扩大网络特征的感受野,获取更大范围的先验框信息,从而提出了扩充感受野网络(RFB)。Redmon 等^[13]提出的 YOLOv3 算法加入 FPN 后,提高了对小目标的检测精度,但由于 FPN 和 YOLOv3 参数量大,其检测速度较慢,对运行设备有较高的性能要求。当前无论是一阶段还是二阶段的目标检测算法都对设备的性能有一定的要求,难以部署到嵌入式设备上。一阶段的算法能端到端地直接生成目标边界框和分类候选框,降低了模型计算量,所以轻量化网络通常是一阶段网络。

虽然目标检测在 PC 端上拥有很高的检测效率,但 PC 端体积大,难以灵活地移植到移动设备中。为了让目标检测算法顺利地运用到移动设备,能够满足实时性的同时提高目标检测精度,研究人员提出模型压缩方案来解决上述移动设备目标检测问题。目前,模型压缩的方法主要可分为以下几个方向:1)轻量化网络模型设计,从网络结构设计出发,设计检测性能高但网络参数量小的轻量化网络,如 SqueezeNet^[14]、MobileNet^[15]和 ShuffleNet^[16];2)模型剪枝,对于已经训练好的网络模型,寻找一种有效的参数重要性判断方法,裁减不重要的连接或过滤器,从而减小模型大小^[17];3)核的稀疏化,在训练过程中,对模型权重更新进行正则化诱导,使得大部分权值为零、更加稀疏,便于后续剪枝操作^[18];4)量化模型,即模型加速,主要方法有二值化网络、三值化网络、深度压缩等,量化主要是将相近值变成一个数,是一种权值共享策略^[19-21];5)迁移学习,是一种模型性能的迁移,如教师-学生网络,使用复杂但高精度的教师网络指导简单的学生网络学习,使得学生网络在轻量的基础上拥有更好的性能^[22]。

近年来,研究人员提出了很多可高效应用到目标检测上的神经网络。例如:Iandola 等^[14]提出了 SqueezeNet 模型,仅使用 AlexNet 五十分之一的参数,就能在 ImageNet 数据集上实现和 AlexNet 一样的检测精度,且该模型可移植到嵌入式设备中,实现嵌入式设备的目标检测;Howard 等^[15]针对移动设备的目标检

测问题提出了 MobileNet;张相胜等^[23]为了解决多目标跟踪的漏检率高和检测速度慢的问题,提出了一种改进的 YOLOv3 多目标跟踪算法;刘高天等^[24]针对遥感图像小目标检测精度低的问题,提出了 RFBNet 的改进算法;刘宗达等^[25]针对视频跟踪算法因目标运动快速变形导致目标丢失的问题,提出了一种基于相关滤波自适应模型与平均峰值相关能量(APCE)重检机制的视频跟踪算法。

当前算法在高性能的硬件平台上能高效检测目标,但在计算速度和内存资源受限的移动设备上,目标检测性能严重下降。针对轻量化网络在目标检测中检测精度低的问题,本文提出了一种以 MobileNet 为基础网络的轻量级目标检测网络 MobileNet-RFB-ECA。在主干特征提取网络使用 MobileNet 替换掉 VGG-16, MobileNet 只使用少量的参数就能达到高效的目标检测效果,极大减少了计算量和模型大小。进一步在主干网络加入有效注意力机制(ECA)提高主干网络的特征表达能力。为了进一步增强目标的特征表达,将浅层位置信息特征与深层语义特征相结合,在特征结合之后加入 RFB 扩大感受野,进一步提升整体网络的特征融合能力。实验结果表明,所提改进模型在保证原模型检测精度的情况下,有效提高了目标检测速度。

2 MobileNet-RFB-ECA 目标检测算法

所提算法的整体结构如图 1 所示,使用 MobileNet 替代传统的 VGG 网络,并在 MobileNet 模块中加入 ECA 注意力机制,同时在网络中加入包含 RFB 的 FPN 结构,加强对小目标的识别,进一步提出了轻量化 RFB 模块,减小模型大小。

2.1 主干网络轻量化

为提高 SSD 检测速度、降低训练难度、便于嵌入移动端,引入 MobileNet 对 SSD 骨干网络(backbone)进行改进。MobileNet 将标准卷积拆分成一个深度卷积和一个逐点卷积,其中深度卷积作用于通道,通过一一对应保证每一个通道只被一个卷积核提取特征,逐点卷积来组合深度卷积后得到的特征图,如图 2 所示。这种结构可以实现跨通道的信息整合并减少输出通道,以牺牲少量精度为代价实现 8~9 倍的计算速度提升。

标准卷积计算量为

$$D_k \times D_k \times M \times N, \quad (1)$$

深度可分离卷积计算量为

$$D_k \times D_k \times M \times 1 + M \times N \times 1, \quad (2)$$

式中: M 表示输入通道的数量; N 表示输出通道数量; D_k 表示输出通道卷积核大小。

MobileNet 主干网络中,只有第 1 层是普通卷积层,其他层都是深度可分离卷积层。通过比较 3×3 的普通卷积和深度可分离卷积的计算量可知,使用 MobileNet 能够大大降低计算量。

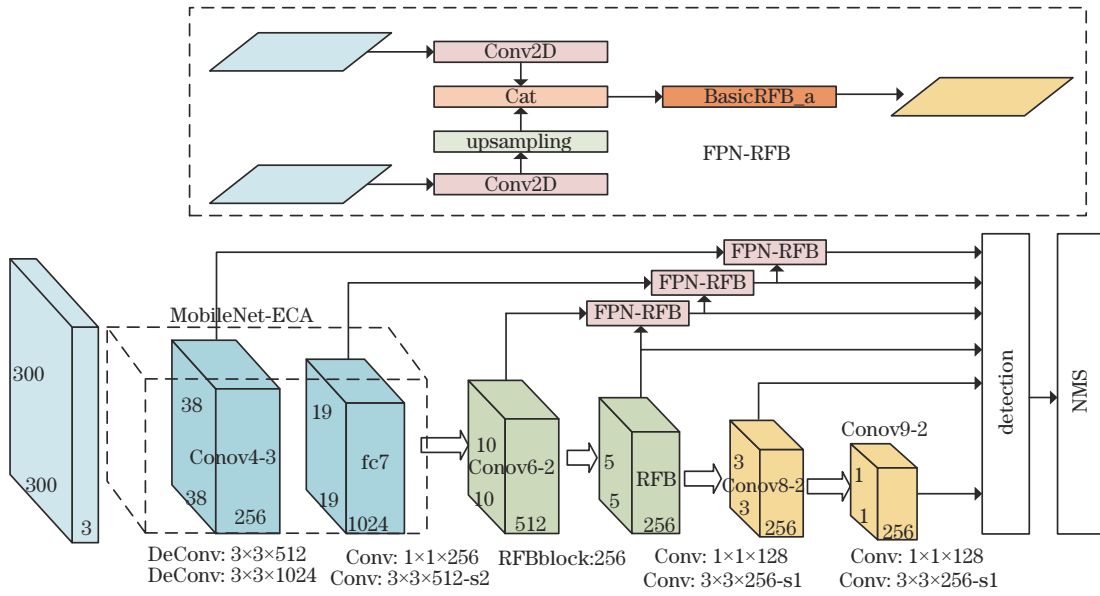


图 1 整体网络结构

Fig. 1 Overall network structure

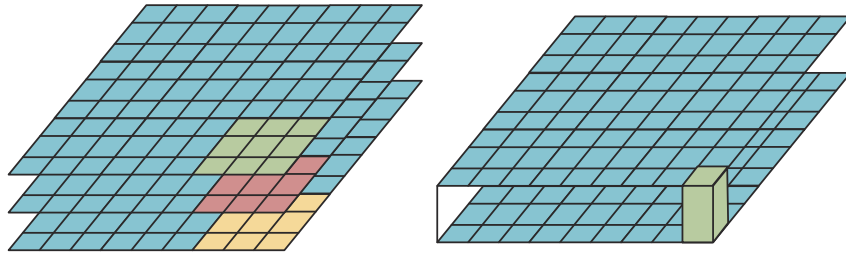


图 2 深度可分离卷积

Fig. 2 Depthwise separable convolution

2.2 浅层和深层的特征融合

FPN是在不同特征层下结合低层网络的目标位置信息和高层网络的特征语义信息的方法,如图3所示。采用最近邻方式上采样得到与前一层次尺寸相同的特征图,再经过一个卷积后与前一层次特征图相加。这种融合方法能获得更好的预测效果。

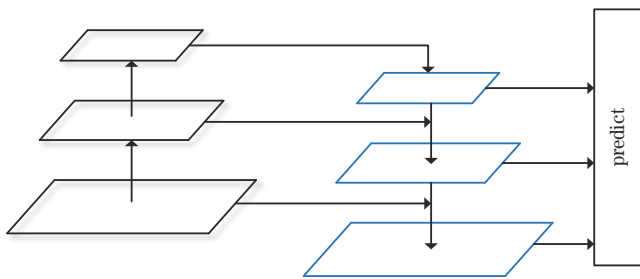


图 3 FPN 结构图

Fig. 3 Structure of FPN

结构,使用扩充感受野的特征融合方式对FPN结构进行改进,RFB通过模拟人类视觉扩大网络特征的感受野,获取更大范围的先验框信息,提高目标检测精度。RFB模块如图4所示,利用不同rate的空洞卷积进行相加,从而实现融合不同特征的目的。

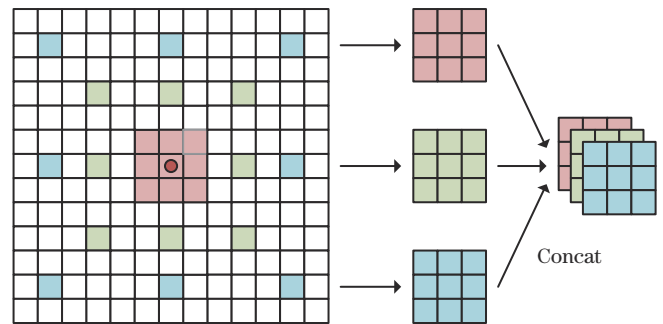


图 4 RFB 结构图

Fig. 4 Structure of RFB

2.2.1 FPN-RFB 结构

上述融合方式只是简单将浅层特征的高分辨信息与深层高语义信息进行相加直接进行预测,并没有利用到先验框周围的信息。本研究引入一种 FPN-RFB

所设计的 FPN-RFB 模块如图 5 所示。该模块是对 FPN 的一种扩展,不仅将浅层特征与深层特征融合,同时模拟人类视觉感知中的尺寸与离心率关系,使用相同大小的卷积核、不同膨胀比例的膨胀卷积进行

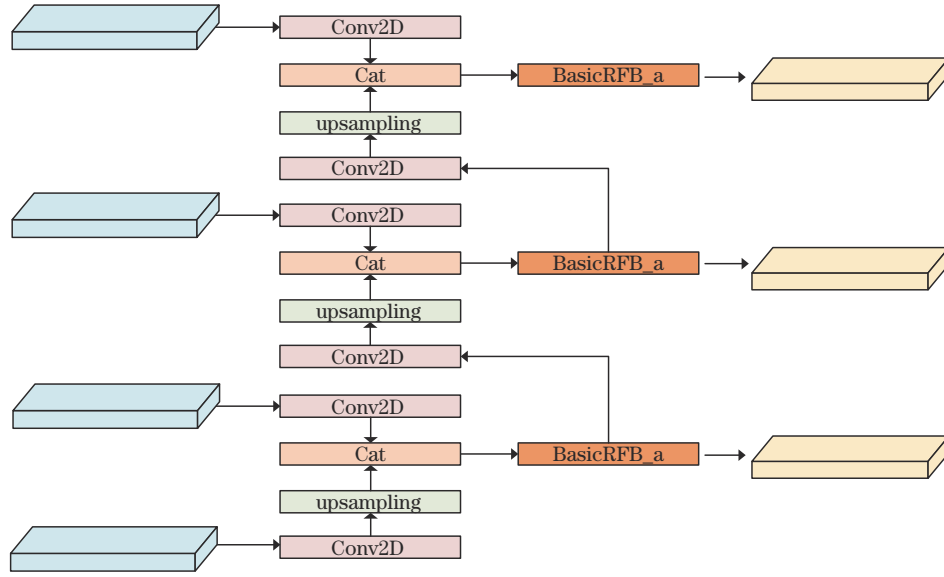


图 5 FPN-RFB 模块

Fig. 5 FPN-RFB module

不同的尺度特征提取,提高轻量化网络对小目标检测准确率。使用双线性插值算法对深层特征图进行上采样,该算法的优势在于特征图尺寸在 10×10 进行上采样时,可通过线性变化进行等比例缩放,能够得到与上层特征图大小相同的 19×19 输出特征图。主干特征提取网络采用 MobileNet,在检测阶段使用 FPN+RFB 组合模块进行检测。

2.2.2 RFB 模块的轻量化

RFB 模块是一个多分支卷积模块,如图 6(a)所示。首先,在每个分支进行 $n \times n$ 的卷积之前,为了减少通道数而先使用 1×1 卷积层。其次,在其他分支使用 $n \times 1$ 和 $1 \times n$ 的卷积层替代原来的 $n \times n$ 卷积层,第 3 分支中使用两个 3×3 卷积层来代替卷积核 5×5 卷积层,通过这种方式减少模块参数量。

为了使 RFB 模块更加轻量化,设计了轻量的 RFB 模块,如图 6(b)所示。使用空间可分离卷积替代原来卷积大小为 3×3 的卷积层,在第 3 分支中使用一个 5×5 的空间可分离卷积替代两个 3×3 卷积层。空间可分离卷积是把一个 $n \times n$ 卷积核给拆开成 $n \times 1$ 和 $1 \times n$ 的卷积核。在尺寸为 $N \times N$ 的图像上进行 $n \times n$ 卷积,传统卷积需要的计算量为

$$(N - n + 1) \times (N - n + 1) \times n \times n, \quad (3)$$

而空间可分离卷积只需要进行的计算量为

$$N \times (N - n + 1) \times n + (N - n + 1) \times (N - n + 1) \times n = (2N - n + 1) \times (N - n + 1) \times n. \quad (4)$$

空间可分离卷积和 RFB 都使用了 $1 \times n$ 和 $n \times 1$ 卷积结构,在这里使用卷积核大小为 $n \times n$ 空间可分离卷积替代 $1 \times n$ 和 $n \times 1$ 的普通卷积,并使用 5×5 的空间可分离卷积替代两个 3×3 的普通卷积。原 RFB 中 3×3 卷积核数量有 7 个,可以用空间可分离卷积替代这些普通卷积。比起传统卷积,空间可分离卷积计算

量更少。图 6(b)给出了轻量化 RFB 结构。空间可分离卷积的计算量为

$$(2N - 4) \times (N - 4) \times 5 + 5 \times (2N - 2) \times (N - 2) \times 3, \quad (5)$$

原本普通卷积的计算量为

$$7 \times (N - 2) \times (N - 2) \times 3 \times 3. \quad (6)$$

2.3 轻量注意力机制模块

在使用卷积神经网络进行特征提取的过程中,由单个卷积层直接进行特征提取不能完全获取每个先验框图像的特征,因此有研究尝试通过注意力机制来增强 CNN 中的卷积特征,更好地获取目标特征信息。注意力机制类似于人类的视觉注意。首先,快速扫描全局图像,选定需要聚焦的目标区域并进行定位。然后收集这些区域的特征信息,从而获取详细目标信息。

注意力机制主要包括两个步骤:1)确定需要注意的区域;2)对关键部位进行特征提取,获取必要信息。注意力机制通过特征图的上下文来强调需要注意的特征。通常,注意力机制用于强调通道或空间中的重要上下文信息,同时抑制与目标无关的上下文信息。例如, Hu 等^[26]提出的 squeeze-and-excitation networks (SENet)注意力模块能够通过网络中的损失函数去学习特征权重、减小无效特征权重并加强有效特征权重,从而提升网络性能。Wang 等^[27]提出的 ECA 使用一种自适应卷积核大小的方法实现不降维的局部跨通道交互。

ECA 通过每个通道的 k 个相邻信息来获取局部跨通道交互信息,如图 7 所示。ECA-Net 具体步骤如下:1)使用全局平均池化得到一个 $1 \times 1 \times C$ 的向量;2)使用一维卷积实现跨通道间的信息交互;3)在一维卷积中自适应卷积核大小,得到每个通道的权重。自适应卷积核大小的计算公式为

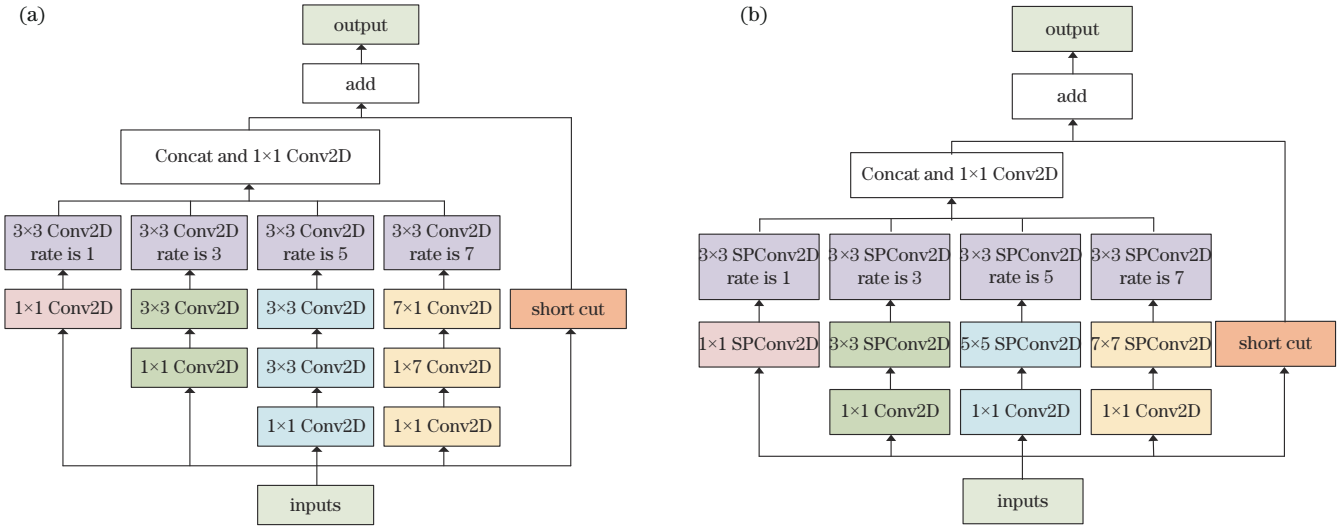


图 6 RFB 结构图。(a)原 RFB 结构图；(b)轻量 RFB 结构图

Fig. 6 Structure of RFB. (a) Original structure of RFB; (b) lightweight structure of RFB

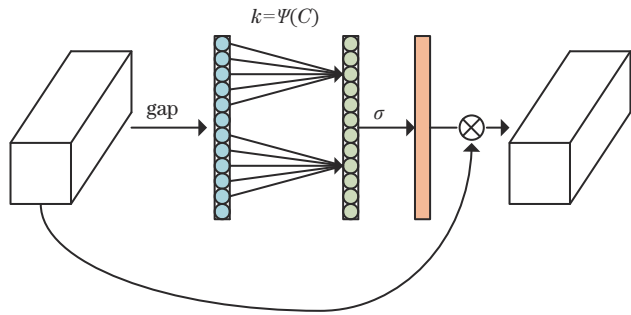


图 7 ECA 结构图

Fig. 7 Structure of ECA

$$k = \Psi(C) = \left\lfloor \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right\rfloor, \quad (7)$$

式中： $\gamma = 2$ ； $b = 1$ ； C 是通道数。本研究在特征提取网络中引入了有效的通道注意力机制。首先在骨干网络深度可分离卷积模块中引入ECA模块，如图8所示。在 1×1 卷积层、批量归一化(BN)层及ReLU置于每

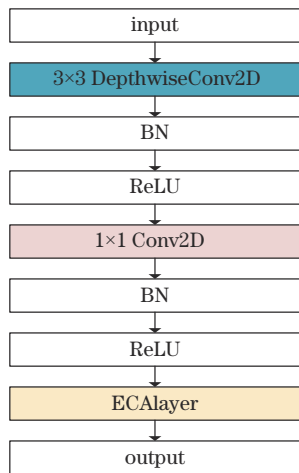


图 8 ECA 模块

Fig. 8 ECA module

次拼接特征通道后，引入ECAlayer层，以实现通道间信息交流增强网络的性能。

3 实验结果与分析

3.1 实验平台与参数设置

实验硬件平台主要包括 NVIDIA GeForce GTX 2060 显卡 (6 GB 显存)、Intel-Core i5-9400F CPU@2.90 GHz 处理器、16 GB 运行内存。软件环境包括 ubuntu 操作系统、Pycharm Community IDE、keras 2.3.1 深度学习框架、Python 3.7.7，采用 CUDA10.0、CUDNN 7.6 对 GPU 加速。

3.2 实验数据

所使用的数据集为 PASCAL VOC 数据集和 KITTI 数据集。PASCAL VOC 是世界计算机视觉挑战赛数据集，总共有 20 个类别目标，包含 16551 张训练图片和 4952 张测试图片。KITTI 是自动驾驶图像数据由车辆在驾驶过程中所拍摄的图像，包含 7481 张训练图像和 7518 张测试图像组成，总共有 7 个类别目标。

3.3 评价指标

在目标检测领域，通常用 mean average precision (mAP) 评价算法检测精度。在轻量化模型方法中，使用每秒处理图像的帧数 (FPS) 评价算法检测速度和模型占用的内存 (size) 来评价模型大小。mAP 为所有类别平均准确率 (AP) 的平均值：

$$P = \frac{N_{TP}}{N_{TotalObjection}}, \quad (8)$$

$$R_{AP} = \frac{\sum P}{N_{TotalImg}}, \quad (9)$$

$$R_{mAP} = \frac{\sum R_{AP}}{C}, \quad (10)$$

式中： N_{TP} 是当前类别正确检测次数； $N_{TotalObjection}$ 是实际

目标的总量; N_{TotalImg} 是包含当前类别在所有图像的数量; P 是准确率; C 是所有类别数。

3.4 实验结果分析

3.4.1 PASCAL VOC 实验结果分析

实验使用 VOC2007 和 VOC2012 进行训练, 在 VOC2007 的测试集上验证算法检测效果, 主要使用 MobileNet 和 SSD 算法进行对比分析, 实验结果如表 1 所示。

输入图片大小为 300×300 , 所提算法的 mAP 为 76.1%, 相较于 MobileNet 提高了 4.2 个百分点, 比原 SSD 模型低 1 个百分点, 但模型大小为 SSD 模型的一

半。对于不同尺寸目标的检测精度如表 2 所示。小目标物体主要有 bottle、plant, 从表 2 可以看出, 所提算法对于小目标精度有较大的提升。原因在于所提算法引入多尺度特征, 并且加入扩充感受野的模块, 使网络能够得到更好的目标位置信息和特征信息, 从而提高对小目标的检测效果。和同样为轻量化的 YOLOv3-Tiny 相比, YOLOv3-Tiny 在极大减小模型大小的同时提高了模型的检测速度, 但导致检测精度的严重下降。所提算法的优势在于将模型的速度与精度进行平衡, 在减小模型大小与加快检测速度的同时, 提高了检测精度。

表 1 不同算法在 VOC 数据集对比

Table 1 Comparison of performance of different algorithms in VOC

Algorithm	Backbone	Train data	Test data	Input size	mAP / %	FPS	Size / MB
Faster-RCNN	VGG-16	VOC07+12	VOC07	600×600	76.3	6	522.9
MobileNet	MobileNet	VOC07+12	VOC07	300×300	71.9	48	33.8
SSD	VGG-16	VOC07+12	VOC07	300×300	77.1	30	100.4
YOLOv3	Darknet-53	VOC07+12	VOC07	416×416	78.4	22	237.2
YOLOv3-Tiny	Darknet-19	VOC07+12	VOC07	416×416	57.3	61	23.2
Proposed algorithm	MobileNet	VOC07+12	VOC07	300×300	76.1	34	50.3

表 2 各种尺寸目标类别不同算法检测结果

Table 2 Test results of different algorithms for various target categories

Algorithm	person	car	bottle	plant	chair	mAP / %
MobileNet	73.1	81.0	34.4	45.3	49.3	71.9
SSD	78.1	86.4	41.0	48.7	56.5	77.1
YOLOv3	78.5	82.2	64.1	52.2	61.3	78.4
YOLOv3-tiny	69.1	74.7	25.0	26.9	33.4	57.3
Proposed algorithm	79.5	87.0	47.0	50.9	54.4	76.1

3.4.2 KITTI 数据集性能检测对比

所提算法在 KITTI 数据集上的检测结果如表 3 所示。由表 3 可知, 所提算法的 mAP 达到 73.5%, 比 SSD 低 1.3 个百分点, 比原模型 MobileNet 提高 15.4 个百分点, FPS 为 34。模型大小为 48.5 MB, 为 SSD 模型的一半。实验结果表明, 多尺度结合上下文特征信息和扩大感受野结合小目标周围的上下文信息, 并加入注意力机制增强特征信息, 有助于对复杂场景下

表 3 不同算法在 KITTI 数据集对比

Table 3 Comparison of performance of different algorithms in KITTI

Algorithm	Backbone	Input size	mAP / %	FPS	Size / MB
MobileNet	MobileNet	300×300	58.1	49	30.5
SSD	VGG-16	300×300	74.8	30	98.3
YOLOv3	Darknet53	416×416	75.4	22	234.8
YOLOv3-Tiny	Darknet19	416×416	47.8	61	20.1
Proposed algorithm	MobileNet	300×300	73.5	34	48.5

缺乏信息的小目标进行更好的分类和定位, 从而提升对小目标检测的准确率。

3.5 消融实验

选择 PASCAL VOC 数据集进行消融实验, 通过比较检测精度差异, 分析所提算法每个模块的性能, 在 MobileNet 模型上逐步增加 FPN、RFB、s-RFB、ECA, 消融实验结果如表 4 所示。

表 4 消融实验结果分析

Table 4 Analysis of ablation test results

FPN	RFB	s-RFB	ECA	mAP / %	FPS	Size / MB
				71.9	48	33.5
✓				74.2	42	39.8
✓	✓			76.0	33	58.3
✓		✓		75.9	34	49.8
✓		✓	✓	76.1	34	50.3

实验结果表明, 引入 FPN 后准确度提升较大, mAP 达到 74.2%, 提高了 2.3 个百分点, 模型大小为 39.8 MB, 仅增加 6.3 MB, 说明加入 FPN 后模型对小目标检测有了明显的提升。加入 FPN-RFB 模块后, mAP 为 76.0%, 提高了 1.8 个百分点, 模型大小为 58.3 MB, 增加 18.5 MB, 说明加入 RFB 模块后, 扩充感受野模块能够提高检测精度, 但模型大小也相对增加一半。使用设计的轻量 RFB 模块后, mAP 为 75.9%, 比上个实验低 0.1 个百分点, 但模型大小为 49.8 MB, 降低了 8.5 MB, 说明空间可分离卷积能够较大减少 RFB 模块的参数数量, 降低模型大小, 且准确率降低 0.1 个百分点, 在可接受范围内。ECA 增加有

效特征的关注程度,使得 mAP 提高了 0.2 个百分点,模型大小增加 0.5 MB,说明 ECA 在提升分类准确度上也有较好的效果,与所提算法结合后效果最佳。

3.6 定性分析结果

为了更直观分析所提算法的检测结果,图 9 和图 10 分别可视化了在 PASCAL VOC 数据集和 KITTI 数据集上的测试结果。从图 9 和图 10 可以看出,MobileNet 在小目标和被遮挡目标的上检测效果差,而所提算法能够更好地检测到被遮挡的目标和小目标。

图 9(a)第 2 列,MobileNet 没有检测到后面被桌

子遮挡的人和模糊的椅子,而所提算法加入了扩充感受野模块和 ECA 注意力机制,感受野模块通过模拟人类视觉的方式,增强了对周围物体的特征提取,注意力机制可以增强有效特征信息和抑制无效特征信息,从而成功检测出了被遮挡的人和椅子。从图 9 和图 10 可以看出,MobileNet 很难检测出远距离的小尺寸物体,而所提算法能够检测到更多的小尺寸目标。即所提算法可以检测更多的小目标物体,解决 MobileNet 对小目标检测效果差的问题,有效提高目标检测的准确度。

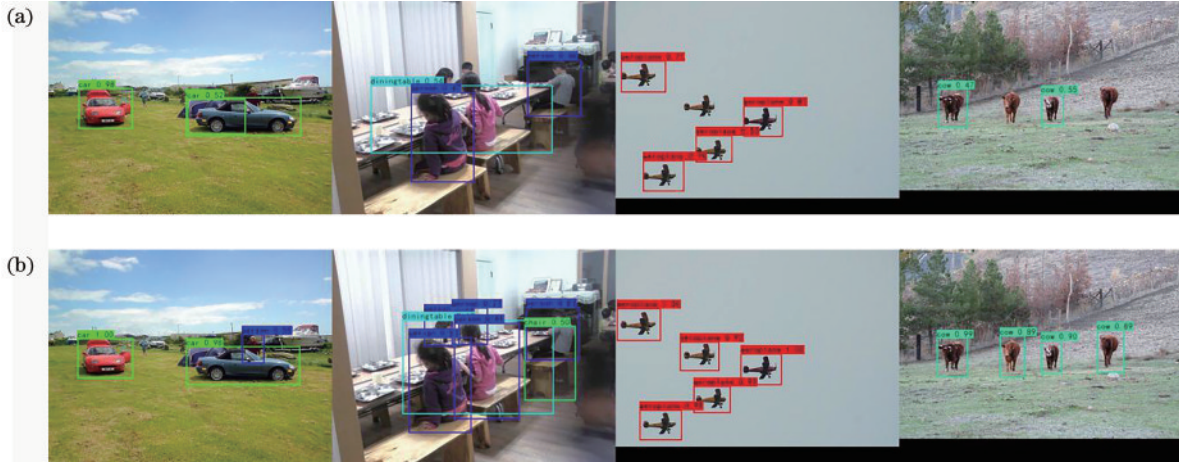


图 9 PASCAL VOC 数据集可视化结果。(a)MobileNet 检测结果;(b)所提算法检测结果

Fig. 9 Visualized results of PASCAL VOC dataset. (a) Detection result of MobileNet; (b) detection result of proposed algorithm



图 10 KITTI 数据集可视化结果。(a)MobileNet 检测结果;(b)所提算法检测结果

Fig. 10 Visualized results of KITTI dataset. (a) Detection result of MobileNet; (b) detection result of proposed algorithm

4 结 论

针对嵌入式设备目标检测准确率低及设备资源有限的问题,提出了一种以轻量级网络 MobileNet 为基础网络的特征融合改进方案。首先,设计了一种包含 FPN 结构的 RFB 模块,在融合上下文特征信息的同时扩大感受野,在小目标的检测上提高准确率。其次,设计了轻量化的 RFB 模块,使用空间可分离卷积代替了

里面 3×3 大小卷积核,减少计算量。最后,ECA 模块的加入解决了为获得更好的性能而增加计算负担的问题,提高了各种 CNN 架构的性能。在 PASCAL VOC 数据集和 KITTI 数据集上的实验结果表明,相比 MobileNet 算法,所提算法的检测精度为分别为 76.1% 和 73.5%,提高了 4.2 个百分点和 15.4 个百分点,模型大小分别为 50.3 MB 和 48.5 MB,平均 FPS 为 34,可满足目标检测中实时性的要求。

参 考 文 献

- [1] 汪鹏, 刘瑞, 辛雪静, 等. 基于残差网络的光学遥感图像场景分类算法[J]. 激光与光电子学进展, 2021, 58(2): 0210001.
Wang P, Liu R, Xin X J, et al. Scene classification of optical remote sensing images based on residual networks [J]. Laser & Optoelectronics Progress, 2021, 58(2): 0210001.
- [2] 李阳曦, 胡成全, 马龙飞, 等. 智能化精准光学诊疗技术研究进展[J]. 中国激光, 2021, 48(15): 1507002.
Li Y X, Hu C Q, Ma L F, et al. Research progress in intelligent and precise optical diagnosis and treatment technology[J]. Chinese Journal of Lasers, 2021, 48(15): 1507002.
- [3] Ju J G, Xing J S. Moving object detection based on smoothing three frame difference method fused with RPCA[J]. Multimedia Tools and Applications, 2019, 78(21): 29937-29951.
- [4] 王一同, 周宏强, 闫景道, 等. 基于深度学习算法的计算光学研究进展[J]. 中国激光, 2021, 48(19): 1918004.
Wang Y T, Zhou H Q, Yan J X, et al. Advances in computational optics based on deep learning[J]. Chinese Journal of Lasers, 2021, 48(19): 1918004.
- [5] 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.
- [6] 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. New York: IEEE Press, 2017: 6517-6525.
- [7] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[EB/OL]. (2015-12-08) [2021-02-03]. <https://arxiv.org/abs/1512.02325>.
- [8] 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.
- [9] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1440-1448.
- [10] 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.
- [11] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [12] Liu S T, Huang D, Wang Y H. Receptive field block net for accurate and fast object detection[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer Vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11215: 404-419.
- [13] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08) [2021-02-05]. <https://arxiv.org/abs/1804.02767>.
- [14] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50× fewer parameters and <0.5 MB model size[EB/OL]. (2016-02-24) [2021-02-05]. <https://arxiv.org/abs/1602.07360>.
- [15] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17) [2021-02-05]. <https://arxiv.org/abs/1704.04861>.
- [16] 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.
- [17] Li H, Kadav A, Durdanovic I, et al. Pruning filters for efficient ConvNets[EB/OL]. (2016-08-31) [2021-05-02]. <https://arxiv.org/abs/1608.08710>.
- [18] Wen W, Wu C P, Wang Y D, et al. Learning structured sparsity in deep neural networks[C]//NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems, December 5-10, 2016, Barcelona, Spain. New York: ACM Press, 2016: 2082-2090.
- [19] Courbariaux M, Hubara I, Soudry D, et al. Binarized neural networks: training deep neural networks with weights and activations constrained to +1 or -1[EB/OL]. (2016-02-09) [2021-02-05]. <https://arxiv.org/abs/1602.02830>.
- [20] Li F F, Zhang B, Liu B. Ternary weight networks[EB/OL]. (2016-05-16) [2021-04-02]. <https://arxiv.org/abs/1605.04711>.
- [21] Han S, Mao H Z, Dally W J. Deep compression: compressing deep neural networks with pruning, trained quantization and Huffman coding[EB/OL]. (2015-10-01) [2021-02-04]. <https://arxiv.org/abs/1510.00149>.
- [22] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network[J]. Computer Science, 2015, 14(7): 38-39.
- [23] 张相胜, 沈庆. 基于改进YOLOv3的多目标跟踪算法研究[J]. 激光与光电子学进展, 2021, 58(16): 1610004.
Zhang X S, Shen Q. Multitarget tracking algorithm based on an improved YOLOv3 algorithm[J]. Laser & Optoelectronics Progress, 2021, 58(16): 1610004.
- [24] 刘高天, 段锦, 范祺, 等. 基于改进RFBNet算法的遥感图像目标检测[J]. 吉林大学学报(理学版), 2021, 59(5): 1188-1198.
Liu G T, Duan J, Fan Q, et al. Target detection for remote sensing image based on improved RFBNet algorithm [J]. Journal of Jilin University (Science Edition), 2021, 59(5): 1188-1198.
- [25] 刘宗达, 董立泉, 赵跃进, 等. 视频中快速运动目标的

- 自适应模型跟踪算法 [J]. 光学学报, 2021, 41(18): 1815001.
- Liu Z D, Dong L Q, Zhao Y J, et al. Adaptive model tracking algorithm for fast-moving targets in video[J]. Acta Optica Sinica, 2021, 41(18): 1815001.
- [26] 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.
- [27] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.