

基于感受野和特征增强的遥感图像实时检测

张廓¹, 陈章进^{1,2*}, 乔栋¹, 张岩¹

¹上海大学微电子研究与开发中心, 上海 200444;

²上海大学现代教育技术中心, 上海 200444

摘要 遥感图像检测在计算机视觉领域中是一项具有挑战性的任务, 目前的遥感图像检测方法在提升精度的基础上忽略了速度, 而在资源勘测和海上救援等需要实时检测的场景中, 提高检测速度也是必不可少的。基于此, 提出了一种轻量化的目标检测网络, 以实现检测精度和速度的权衡。基于 YOLOv4 进行设计, 将原主干网络替换为剪枝后的 MobileNetV2, 将特征提取部分的普通卷积计算替换为深度可分离卷积以大幅度减小模型计算量, 同时嵌入感受野增强模块和注意力机制模块以提升模型检测精度。在遥感图像数据集上的实验结果表明: 所提网络的平均准确率达到 89.80%, 检测速度达到 33.4 frame/s; 与 YOLOv4 相比, 平均准确率仅下降了 1.48 个百分点, 但检测速度提升了近 1.5 倍; 与 YOLOv4-Tiny 相比, 平均准确率提升了 9.05 个百分点。所提模型实现了速度和精度的权衡, 权重仅为 44 MB, 易于部署, 且能满足实时检测场景需要。

关键词 遥感; 目标检测; 遥感图像; 轻量化网络; 感受野增强; 注意力机制

中图分类号 TP753

文献标志码 A

DOI: 10.3788/LOP212698

Real-Time Image Detection via Remote Sensing Based on Receptive Field and Feature Enhancement

Zhang Kuo¹, Chen Zhangjin^{1,2*}, Qiao Dong¹, Zhang Yan¹

¹Microelectronics Research and Development Center, Shanghai University, Shanghai 200444, China;

²Modern Educational Technology Center, Shanghai University, Shanghai 200444, China

Abstract Object detection in remote sensing images is a challenging task in the field of computer vision. Existing remote sensing image detection methods ignore the speed with the aim of improving the accuracy; however, it is also essential to increase the detection speed in real-time detection scenes, such as in resource surveys and maritime rescue. This paper proposes a lightweight target detection network to realize a trade-off between detection accuracy and speed. The design replaces the original backbone network of YOLOv4 with the pruned MobileNetV2. In addition, the ordinary convolution calculation of the feature extraction method is replaced by deep separable convolution to considerably reduce the computational complexity of the model. Finally, the receptive field enhancement module and attention mechanism module are embedded to improve the detection accuracy of the model. The experimental results on the images in the dataset containing the remote sensing images show that the accuracy, as measured by the mean average precision, is 89.80%; further, the model detection speed is 33.4 frame/s. Compared with YOLOv4, the accuracy only decreases by 1.48 percentage points, but the detection speed increases by nearly 1.5 times. Compared with the YOLOv4-Tiny algorithm, the average accuracy is 9.05 percentage points higher. The proposed model successfully meets the trade-off requirements of speed and accuracy. The weight of the model is only 44 MB, which makes it easy to deploy and indicates that it meets the requirements of real-time detection scenarios.

Key words remote sensing; target detection; remote sensing image; lightweight network; receptive field enhancement; attention mechanism

1 引言

遥感技术的快速发展极大地提高了遥感图像的数量和质量,现有的目标检测算法可以用来检测各种地面目标,如飞机、操场、船舶、建筑物等^[1]。遥感卫星技术的发展使得遥感图像成像技术更加发达,如何提取出高分辨率遥感图像的深度语义的特征信息已经成为了当前学者研究的学术焦点。遥感图像检测在农业、地质勘查、安防、城市规划等领域应用广泛。然而,对于农作物灾害检测、地理资源勘测、海上救援、道路交通分析等一些需要满足实时遥感图像监测的场景,检测速度和精度的权衡仍是一个巨大的挑战。

传统的遥感图像检测算法基于手工特征^[2]或纹理特征^[3]等分类器进行检测,这种检测模型在检测速度及精度上都很难满足目前时代发展的需求。随着深度学习的不断发展,已经有许多成熟的基于卷积神经网络(CNN)的目标检测方法,如以 Faster-RCNN^[4]、Mask R-CNN^[5]为代表的两阶段算法和以 SSD^[6]、YOLOv3^[7]、YOLOv4^[8]为代表的单阶段算法等。两阶段 RCNN 系列目标检测算法凭借高定位精度在目标检测领域中具有很大的优势,但其检测速度较慢;单阶段 SSD 算法的检测速度略高于 YOLOv3 和 YOLOv4 目标检测算法,但检测精度较低,在高分辨率的遥感图像检测场景中很难检测出复杂目标及小目标。综合考虑,选用 YOLO 系列的目标检测算法为宜,YOLOv4 作为 YOLOv3 的改进版,模型的检测速度和精度都得到了提升,故本文以 YOLOv4 作为参考展开进一步的研究。但 YOLOv4 算法的卷积特征提取器主要为普通 3×3 卷积,使用普通卷积提取图像特征时存在以下问题:1) 在浅层的特征图上检测目标时,由于浅层特征图感受野较小,语义特征较弱,对于一些小目标的检测,仅使用浅层特征图检测会增加物体检测的难度;2) 在普通卷积计算过程中,同时考虑图像的区域和通道信息,造成过多的冗余计算,在复杂的网络模型中难以保证实时性,模型部署也有一定的困难;3) 轻量化的 YOLOv4-Tiny 模型的检测实时性高,权重较小且易于部署,但在高分辨率的遥感图像检测场景中的精确度较低。

针对以上问题,国内外的科研工作者们进行了一些相关的工作:Chen 等^[9]在遥感图像船舶识别任务中,利用多层恒等映射的空洞卷积和注意力机制(DAM)融合的方法改进 YOLOv3 网络,提高了特征提取的感受野,在精度上有所提高,但 DAM 结构中会产生过多的模型参数,不利于设备部署;谢俊章等^[10]为了提高密集小目标的特征信息提取能力,减少了特征提取网络的卷积层数,速度上比原 YOLOv4 模型有一些提高,

但仍未达到实时检测的目的;张欣等^[11]在遥感飞机单目标检测任务中,将 YOLOv4-Tiny 的激活函数替换为 Mish,以更好传递梯度流,并添加空间金字塔池化结构增强感受野,在速度上超越了 YOLOv4,其网络结构简单,对于单目标检测具有一定的优势,但在多目标检测任务中仍存在准确率和召回率低等问题;此外,也有一些提高感受野和增进速度的算法^[12-13],但从整体框架而言,计算量过于冗余,对于一些算力较小的设备来说难以满足实时性要求。

结合上述研究的优势与不足,本文提出了一种基于感受野和特征增强的轻量网络模型。所提模型以 YOLOv4 为基础,将其主干网络替代为 MobileNetV2^[14],在网络的有效特征层中引入感受野增强模块和空间与通道注意力机制模块,在扩大感受野的同时增强提取复杂语义特征信息的能力;为进一步减少特征提取网络的计算冗余,改进 PANet 结构,将 3×3 普通卷积和下采样结构替换为深度可分离卷积以减少模型参数,并提高模型的实时性。

2 所提方法

2.1 YOLO 目标检测原理

YOLOv4 的网络结构如图 1 所示:YOLOv4 以宽高为 416×416 的图片作为输入层;主干网络为 CSPDarkNet-53,以 Mish 为激活函数加强梯度流的传输;分支结构使用空间池化结构^[15]和路径聚合网络^[16],空间池化结构用于加强深层特征图的感受野,路径聚合网络通过多次上采样及下采样操作对不同特征信息进行反复聚合再分别提取,提取出的 3 个有效特征层为 19×19 、 38×38 、 76×76 ,分别用来检测小目标、中目标及大目标;最后,Head 结构对提取的 3 个特征图进行解码,预测计算 3 个边界框,寻找目标的位置、尺寸、不同类别的概率张量,并对每个边界框进行分类预测,每个边界框对应 $5 + \text{classes}$ 个值,其中 classes 表示分类总数,5 表示预测边界框的信息:中心坐标 (x, y) 、宽 (w) 、高 (h) 及置信度(confidence)。计算预测框与真实框的交并比(IOU),使用非极大值抑制(NMS)抑制冗余的预测框,并选出得分最优的预测框,给出相应的置信度。为进一步提高预测的准确性,更好实施目标定位,采用了基于 K-means 的聚类算法,对训练集的真实框(ground-truth)进行聚类,以找到合适的先验框。

与传统的两阶段 RCNN 模型^[17]结构相比,YOLOv4 采用回归的思想,使用单阶段网络实现了分类预测与位置定位的任务。此外,与基于候选框预测的方式不同,YOLOv4 利用整张图片训练模型,可更好地实现多目标定位和识别。

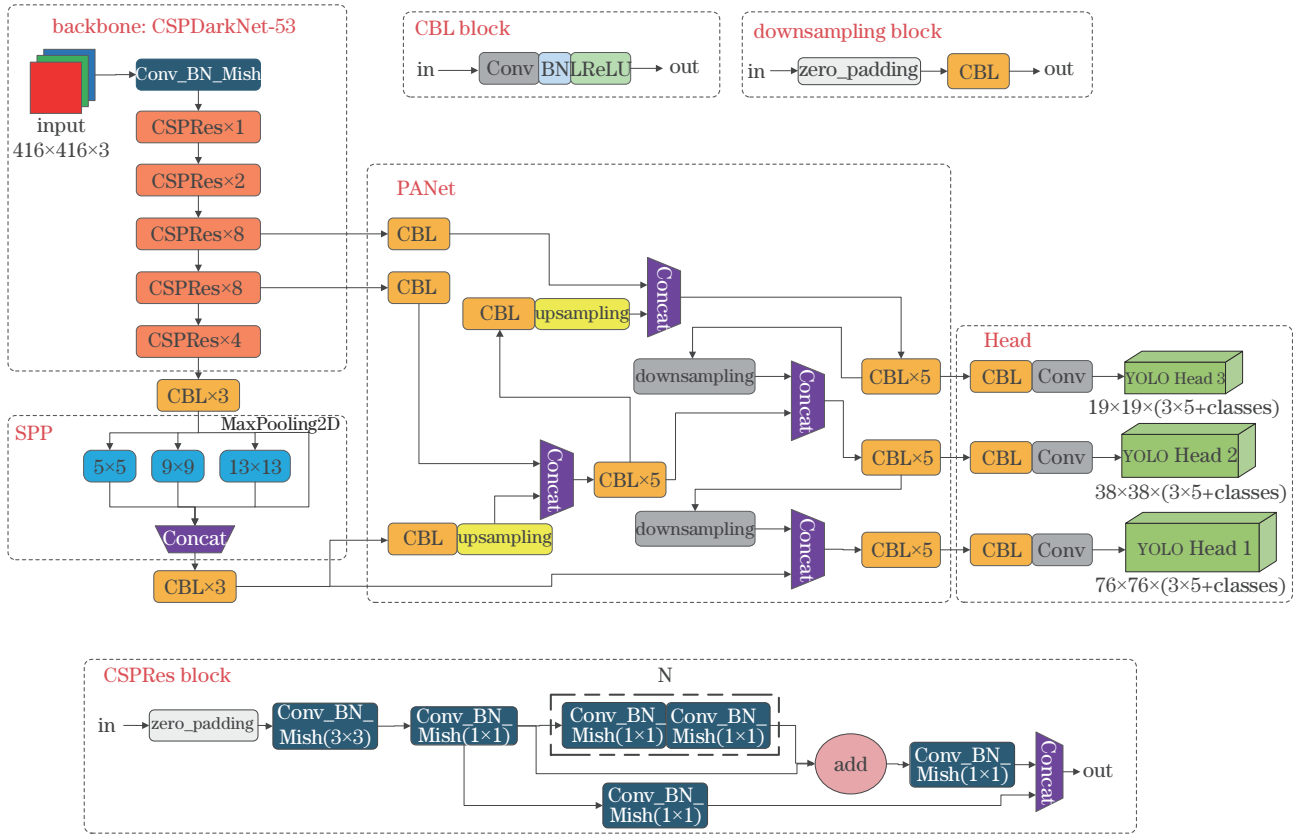


图 1 YOLOv4 网络结构图

Fig. 1 YOLOv4 network structure diagram

2.2 网络结构设计

YOLOv4 虽然在一些多目标分类检测数据集中的精度有很大提升,但是其忽略了实时性,资源有限的小型边缘设备无法满足实时性需求。针对此问题,本研究对原有的网络结构进行轻量化改进,通过替换主干网络、使用深度可分离卷积结构替代普通卷积来提升模型检测速度,为减少精度损失,嵌入感受野增强模块和注意力机制模块。

2.2.1 主干网络改进

以轻量化的 MobileNetV2 代替原 CSPDarkNet-53 作为主干网络。MobileNetV2 主要通过基于残差结构的深度可分离卷积(Bottleneck)来提取相应的特征,该结构如图 2 所示,主通道部分利用 1x1 卷积升维、3x3 深度可分离卷积提取特征、再利用 1x1 卷积降维,残差部分直接和输出部分相接,升维部分和特征提取均使用归一化处理(BN)和 ReLU6 激活函数,降维部分仅进行归一化处理。归一化处理的作用在于加快模型收敛,ReLU6 激活函数将 ReLU 激活函数的最大输出限制为 6,即

$$f(x) = \min[\max(0, x), 6]. \quad (1)$$

通过 ReLU6 函数的幅度限制,在减少模型参数的同时提高模型的数值分辨率,同时过滤掉负值,增强模型整体泛化能力。

MobileNetV2 的主要计算参数分布于 Bottleneck

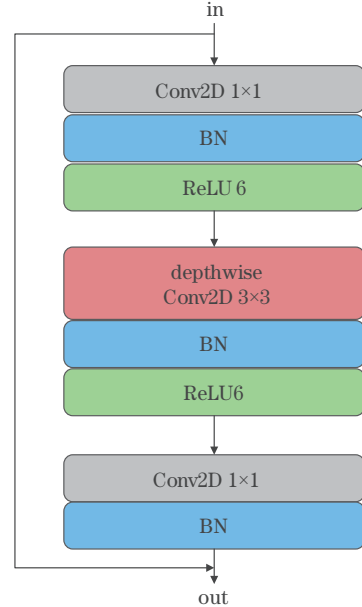


图 2 Bottleneck 结构

Fig. 2 Bottleneck structure

结构中,其最后的 1x1 卷积及平均池化核主要用于分类输出。与分类网络不同,目标检测网络结构复杂,且在进行特征提取时需要多特征层进行聚合,故主干部分以 416x416 为输入,尺寸为 (52, 52, 32)、(26, 26, 96)、(13, 13, 320) 的 3 个特征图用于路径聚合网络中的有效提取,同时剪裁掉不必要的最后三层结构。剪

枝后的 MobileNetV2 结构描述如表 1 所示,其中 Module 为不同网络层模块描述, T 表示扩张维度的倍数, Channel 表示通道数, Num 表示模块使用次数, Stride 表示步长, Output 表示每个模块输出特征图尺寸,同时也表示下一模块的输入, M1、M2 和 M3 分别表示提取的 3 个特征图。

表 1 剪枝后的 MobileNetV2 结构描述

Table 1 Description of MobileNetV2 structure after pruning

Module	T	Channel	Num	Stride	Output	Name
Conv2D		32	1	2	$208 \times 208 \times 32$	
Bottleneck	1	16	1	1	$208 \times 208 \times 16$	
Bottleneck	6	24	2	2	$104 \times 104 \times 24$	
Bottleneck	6	32	3	2	$52 \times 52 \times 32$	M1
Bottleneck	6	64	4	2	$52 \times 52 \times 64$	
Bottleneck	6	96	3	1	$26 \times 26 \times 96$	M2
Bottleneck	6	160	3	2	$13 \times 13 \times 160$	
Bottleneck	6	320	1	1	$13 \times 13 \times 320$	M3

2.2.2 特征提取结构改进

在 YOLOv4 网络中,加强特征提取网络使用路径聚合网络,该结构中的 3×3 卷积直接影响网络的后续

计算,如果直接应用于所提模型中,产生的模型权重较大,检测实时性难以保证。针对 PANet 结构的改进如图 3 所示,改进后的结构命名为 PANet-lite。在该结构中,将原结构中的 3×3 普通卷积和下采样结构替换为 1×1 卷积和 3×3 深度可分离卷积交替结构,并将激活函数替换为 ReLU6。

2.2.3 感受野增强模块

在使用特征提取网络提取主干网络中的特征图时,由于遥感图像的分辨率较大,且背景较为复杂,未经过处理的特征图中特征的感受野较小,语义信息不足,因此需要使用 RFBNet 模块^[18]增强特征图的感受野来提升精度。考虑到最后一层的特征图已经由空间金字塔池化(SPP)结构扩大感受野,故所提模型重点针对两层浅层特征图进行感受野增强。感受野增强模块如图 4 所示,其中 k 表示先前特征层的通道数,该结构共包括 4 条支路,其中 L_4 为基于恒等映射的残差支路, L_1 、 L_2 和 L_3 由 1×1 、 3×3 的普通卷积外加不同膨胀率(rate)的 3×3 空洞卷积级联形成,其膨胀率分别为 1、3、5。每个卷积层添加归一化处理加快模型的收敛速度,同时为了进一步加强模块消融于模型的泛化能力,将激活函数替换为 ReLU6。

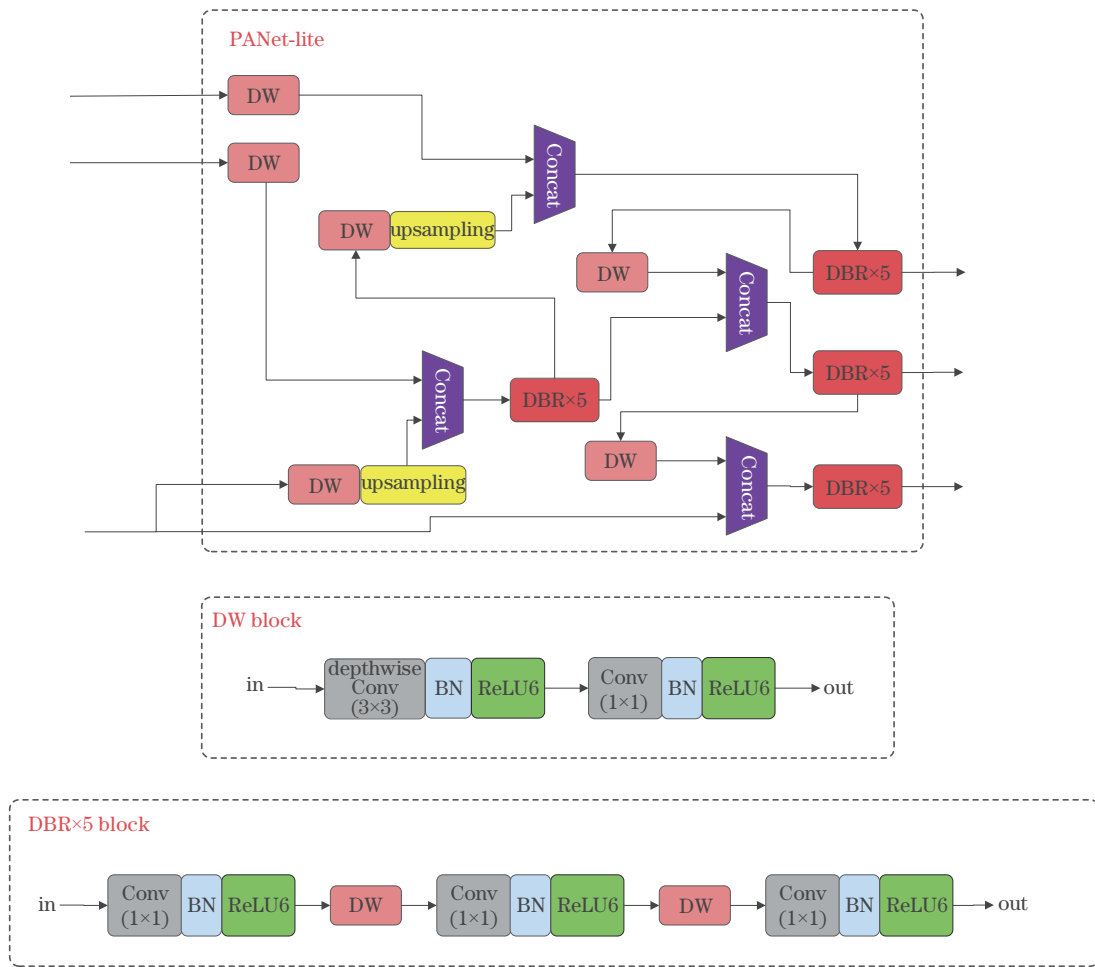


图 3 PANet-lite 结构设计

Fig. 3 Structure design of PANet-lite

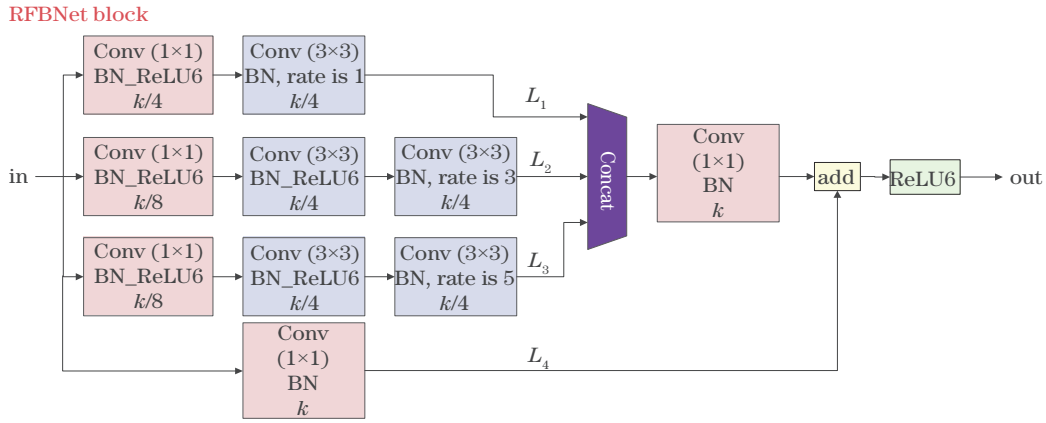


图 4 感受野增强模块

Fig. 4 Receptive field enhancement module

该感受野增强模块中的计算可描述为

$$L_1 = D_1[P(X_{\frac{k}{4}})], \quad (2)$$

$$L_2 = D_3[C \cdot P(X_{\frac{k}{8}})], \quad (3)$$

$$L_3 = D_5[C \cdot P(X_{\frac{k}{8}})], \quad (4)$$

$$L_4 = P(X_{\frac{k}{4}}), \quad (5)$$

$$Z = \zeta[\delta(L_1, L_2, L_3) + L_4]. \quad (6)$$

式中: X 表示输入特征图, 下标为通道数; P 表示 1×1 卷积、归一化处理层和 ReLU6 激活函数的非线性组合; C 表示 3×3 卷积、归一化处理层和 ReLU6 激活函数的非线性组合; D 表示 3×3 空洞卷积、归一化处理层和 ReLU6 激活函数的非线性组合, 下标表示膨胀率, 分别为 1、3、5; δ 表示通道整合操作; ζ 表示 ReLU6 激活函数; Z 为最终输出的特征图。

通过感受野增强模块的处理, 网络对小目标及复杂目标的敏感能力得到提升, 网络的检测精度也有所

提高。

2.2.4 注意力机制模块

Convolutional block attention module (CBAM) 注意力机制是一个简单而有效的前馈卷积神经网络注意力机制模块, 同时也是一种轻量级的通用模块, 它可以无缝集成到任何卷积神经网络模型架构中, 提高模型的代表能力^[19]。CBAM 主要分为两个部分, 通道注意力机制 (channel attention) 子模块和空间注意力机制 (spatial attention) 子模块, 每个模块之间相互独立。

通道注意力机制结构如图 5 所示。通道注意力机制将输入的特征图在空间维度上进行压缩, 即压缩特征图高和宽, 从而得到一个一维矢量后再进行操作, 平均池化和最大池化可用来聚合特征映射的空间信息, 之后送到一个可共享的感知神经网络 (MLP), 压缩输入特征图的空间维度, 并逐元素求和, 以提取通道注意力图, 最后与原特征图逐元素相乘获得通道注意力机制特征。

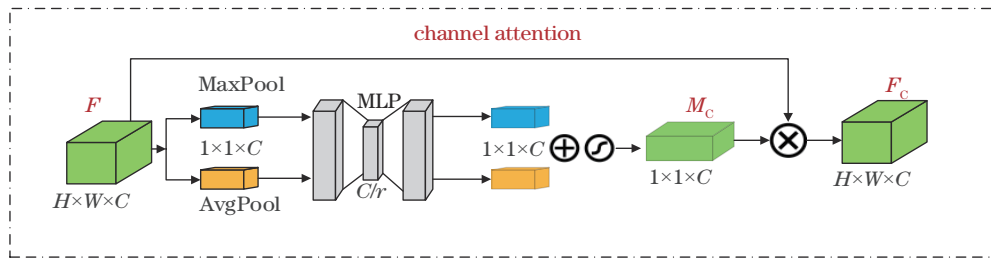


图 5 通道注意力机制结构

Fig. 5 Channel attention mechanism structure

通道注意力机制计算过程可表示为

$$M_c = \sigma \left\{ \text{MLP} \left[\text{AvgPool}(F) \right] + \text{MLP} \left[\text{MaxPool}(F) \right] \right\}, \quad (7)$$

$$F_c = M_c \otimes F, \quad (8)$$

式中: F 、 M_c 、 F_c 分别代表输入特征、通道注意力机制提取特征、 M_c 与 F 逐元素相乘后的通道注意力机制特征; σ 代表 Sigmoid 激活函数; AvgPool 和 MaxPool 表示平均池化核和最大池化核; \otimes 代表逐元素相乘。

将通道注意力机制获取的特征 F_c 输入空间注意力机制网络。空间注意力机制网络结构如图 6 所示。空间注意力机制对通道进行压缩, 在通道维度分别进行了平均池化和最大池化, 将特征沿通道合并得到一个二维的特征图, 最后同样采用逐元素相乘方法获得空间注意力机制特征。

空间注意力机制计算方法可描述为

$$M_s = \sigma \left\{ f^{7 \times 7} \left[\text{AvgPool}(F_c); \text{MaxPool}(F_c) \right] \right\}, \quad (9)$$

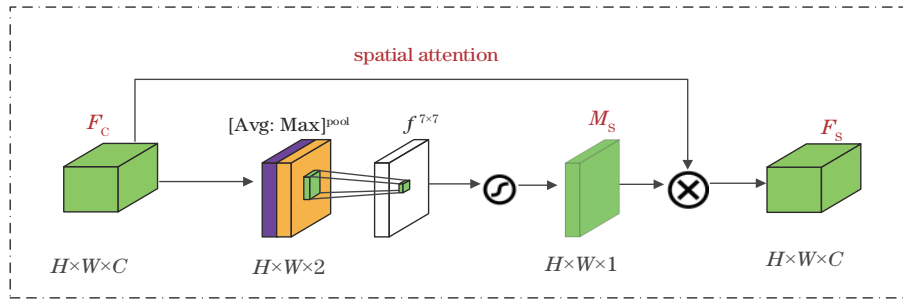


图 6 空间注意力机制结构

Fig. 6 Spatial attention mechanism structure

$$F_s = M_s \otimes F, \quad (10)$$

式中： F_s 、 M_s 分别代表空间注意力机制提取特征、 M_s 与 F_s 逐元素相乘后的空间注意力特征，也即最终特征； $f^{7 \times 7}$ 表示使用 7×7 卷积核进行通道降维。

2.2.5 改进后的网络结构

改进后的网络结构如图 7 所示，其中 Bottleneck 表示主干网络的残差深度可分离卷积结构，改进后的网络命名为 LM2-YOLOv4。新网络通过改进主干网络，选用图 2 中的 M1、M2 和 M3 作为主干网络的有效特征提取层，M3 层经过空间金字塔池化结构（记为 M3'）与 M1 和 M2 共同作为特征提取网络中的 3 个有效特征层。M1 和 M2 层通过消融 RFBNet 结构增强了模型的

感受野，可获得更多的特征语义信息，进而提升了检测召回率。3 个特征层均消融了 CBAM 注意力机制结构，在特征提取时更多关注通道与空间特征信息。经消融后的特征经过图 4 中的 PANet-lite 结构前，使用 1×1 卷积对通道进行调整，调整后的特征尺寸从上到下依次为 (52, 52, 128)、(26, 26, 256)、(13, 13, 512)，之后再行上采样与反复聚合操作使得 YOLO Head 可用于模型检测。为进一步降低计算量，将空间金字塔池化结构及检测头部分的 3×3 普通卷积替换为 3×3 深度可分离卷积。改进后的模型与 YOLOv4 网络相比，以降低部分精度为前提，计算量大幅度降低，速度有显著提升，符合轻量化网络设计标准。

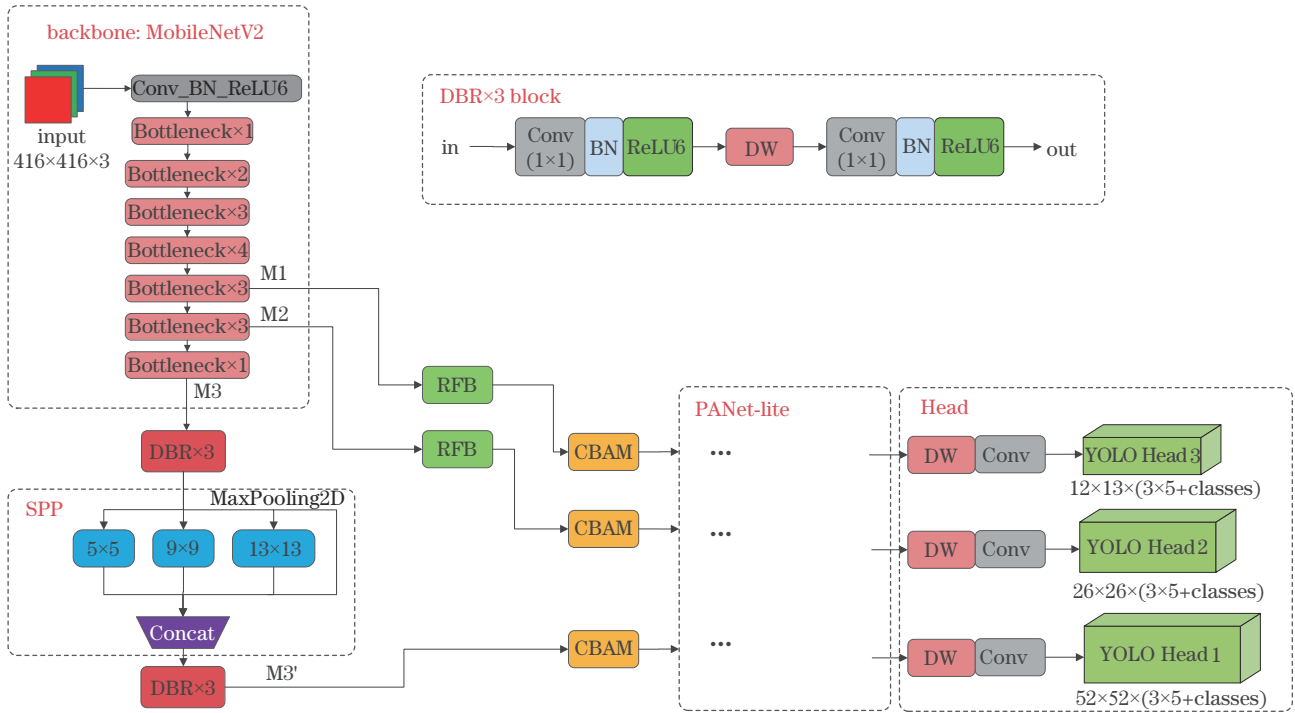


图 7 改进后的网络结构

Fig. 7 Improved network structure

3 实验过程与结果分析

3.1 实验环境及数据集准备

实验在 Windows 10 操作系统下运行，CPU 平台

为 Intel i5-10400F，GPU 平台为 RTX2060，显存为 6 GB，使用 Python 编程语言，深度学习框架为 tensorflow-gpu 2.2.0，使用 Cuda 和 Cudnn 加速网络模型。

采用的数据集是公开的 RSOD 遥感图像数据集^[20],原数据集共包含 4 类目标:飞机(aircraft)、操场(playground)、立交桥(overpass)、油罐(oiltank)。考虑到立交桥类特征复杂及数据集中未包含海上目标,故从 LEVIR 数据集^[21]中选取了具有海上船舶(ship)目标的图片,并从 DIOR 数据集^[22]中选取一些清晰的立交桥图片,使用 Labelimg 软件对目标进行标注,最终数据集共包括 1814 张图片和 9068 个真实框实例。在该数据集中,选用 1160 张图片作为训练集,291 张作为验证集,363 张图片作为测试集,使用 K-means 聚类算法产生 (8, 8)、(12, 12)、(15, 16)、(20, 20)、(25, 28)、(36, 38)、(48, 56)、(79, 112)、(164, 182) 等 9 个不同的锚点框作为先验框。在模型训练中,采用缩放、亮度调整、加噪等随机方法对训练集进行数据增强处理,将初始学习率设置为 0.001, batch-size 为 4, 使用 Adam 优化器优化,迭代 300 个 epoch,最终损失值降为 1.8 左右,模型基本收敛,最终模型的训练集与验证集的损失曲线如图 8 所示。

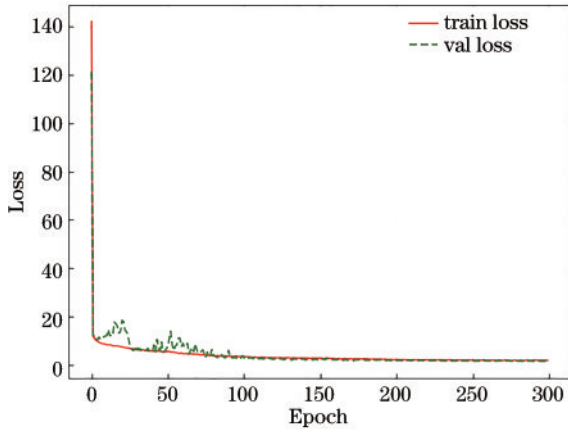


图 8 所提模型训练损失曲线
Fig. 8 Training loss curve of proposed model

3.2 评价指标

为定量评估所提网络模型的检测性能,选取 PASCAL VOC 数据集^[23]中的评价指标平均精度均值(mAP)来评价模型性能:

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}, \quad (11)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}}, \quad (12)$$

$$R_{AP} = \sum_i^N \int_0^1 P(R) dR, \quad (13)$$

$$R_{mAP} = \frac{R_{AP}}{N}, \quad (14)$$

式中: P 、 R 分别表示单一类别查准率(precision)和召回率(recall); N_{TP} 表示真正例的数量, N_{FN} 表示假反例的数量, N_{FP} 表示假正例的数量; R_{AP} 表示单一类别的查准率-召回率曲线(P - R)下方形成的面积积分; N 表

示类别数。同时以在计算机摄像头每秒传输的帧数衡量模型的检测速度。

3.3 模块消融对比实验

在消融实验中,以未经消融感受野增强模块和 CBAM 注意力机制模块的网络作为基准(Baseline),在 PANet-lite 结构前消融 SE 注意力机制^[24]、ECA 注意力机制^[25]、CBAM 注意力机制、感受野增强模块、同时消融感受野增强模块和 CBAM 注意力机制模块的模型进行对比实验。消融实验后的实验结果如表 2 所示。

从表 2 可以看出:嵌入不同的注意力机制模块,模型精度均有所提升,其中 CBAM 注意力机制较 SE、ECA 注意力机制分别提升了 1.81 个百分点和 0.96 个百分点,即模型同时关注空间特征和通道特征更有利于提升特征判别性;嵌入感受野增强模块后,模型的 mAP 提升了 1.12 个百分点;同时嵌入 RFB 和 CBAM 模块后, mAP 提升了 4.5 个百分点,即增强特征图的感受野使得遥感图像中的复杂语义得到充分提取,且模型参数量仅提升 1.97 MB。

表 2 消融实验对比结果
Table 2 Ablation experiment comparison results

Parameter	Baseline	SE	ECA	CBAM	RFB	RFB+ CBAM
mAP / %	85.30	86.46	87.31	88.27	86.42	89.80
Weight / MB	41.64	41.83	41.66	42.03	43.22	43.61

3.4 算法对比实验

在算法对比实验中,将双阶段的 Faster-RCNN、单阶段的 SSD、YOLOv3、YOLOv4 及 YOLOv4-Tiny 和 LM2-YOLOv4 算法进行比较,比较各个模型的权重大小、精度和检测速度,比较结果如表 3 所示。

从表 3 可以看出,除 YOLOv4 模型外, LM2-YOLOv4 算法模型的 mAP 均高于其他传统算法模型,达到了 89.80%,与 YOLOv4 模型相比, LM2-YOLOv4 算法的模型权重显著降低,仅有约 44 MB,检测精度与其仅相差 1.48 个百分点,但检测速度较 YOLOv4 却有明显的提升,达到了 33.4 frame/s。与 YOLOv4-Tiny 模型相比, LM2-YOLOv4 的 mAP 值提升了 9.05 个百分点,速度仅降低了 1.8 frame/s,可满足实时性遥感图像检测场景需要。即所提目标检测模型在检测速度和精度之间达到了相对权衡的检测效果。

图 9 为 YOLOv4、YOLOv4-Tiny、LM2-YOLOv4 等 3 种算法在测试集中的实际检验效果对比,每一列从上到下的类别依次为飞机、油罐、立交桥、操场、船舶。从检测效果中可以明显看出:YOLOv4-Tiny 由于精度较低,对飞机、舰船等小目标的漏检率较大; LM2-YOLOv4 模型较 YOLOv4 模型相比漏检现象较少,对小目标的检测精度相差不大,单在立交桥这一复杂目

表 3 所提模型与其他模型实验结果对比

Table 3 Comparison of experimental results between proposed model and other models

Model	Backbone	Input size	mAP / %	Weight / MB	Detection speed / (frame · s ⁻¹)
Faster-RCNN	ResNet50	600 × 600	88.16	111	4.8
SSD	VGG16	300 × 300	74.79	95	25.2
YOLOv3	DarkNet-53	416 × 416	85.08	241	20.2
YOLOv4	CSPDarkNet-53	416 × 416	91.28	251	22.5
YOLOv4-Tiny	CSPDarkNet-Tiny	416 × 416	80.75	23	35.2
LM2-YOLOv4	MobileNetV2	416 × 416	89.80	44	33.4

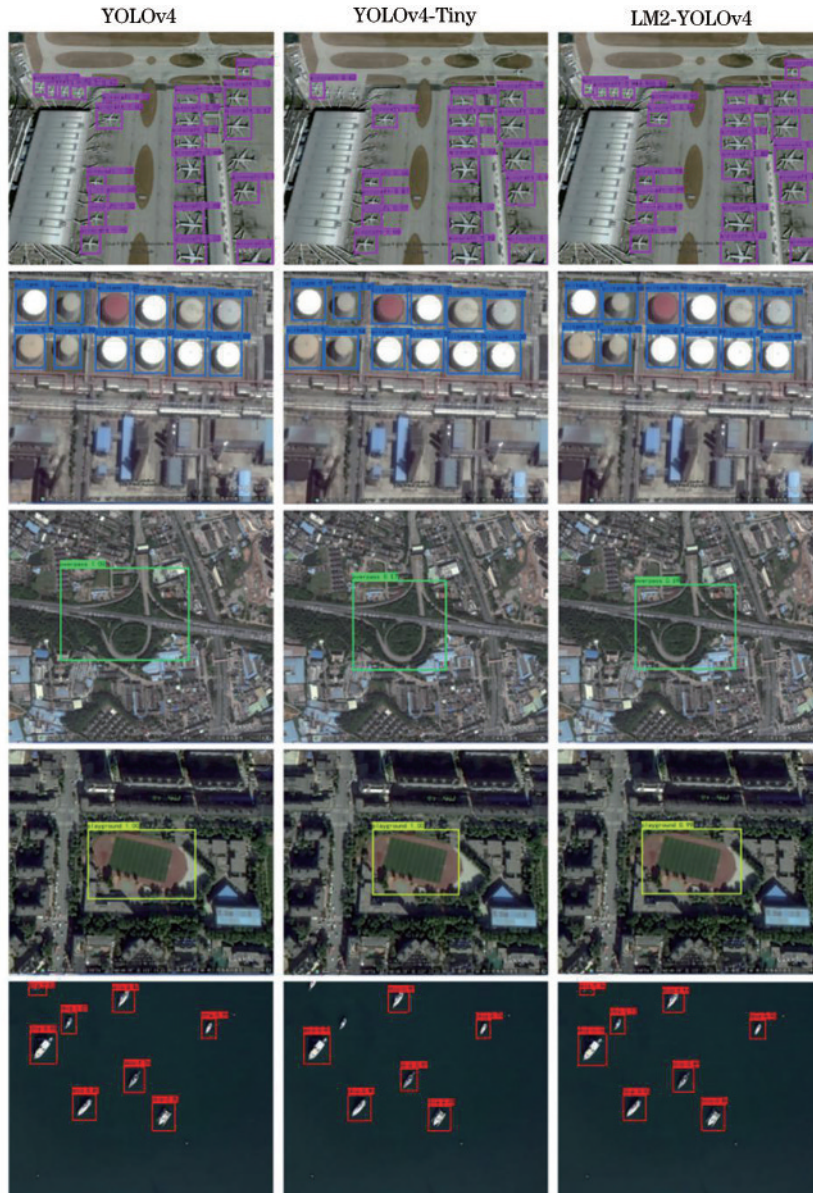


图 9 检测效果对比图

Fig. 9 Detecting results comparison

标中精度较低;而所提模型的检验效果较好。结合表 3 的分析可知,所提模型通过牺牲小部分精度来大幅度提升检测速度,可更好在实时检测需求较高的场景中使用。

4 结 论

提出了一种平衡检测速度和精度的检测模型 LM2-YOLOv4。该模型通过修改主干网络和使用深

度可分离卷积降低模型参数、提高检测速度,通过增强感受野和注意力特征减少模型的精度损失。实验结果表明,所提模型精度达到了 89.80%,速度达到了 33.4 frame/s,与 YOLOv4 和 YOLOv4-Tiny 相比,该模型达到速度和精度的相对平衡。所提模型较小,可更好地部署于具有轻量 AI 算力嵌入式平台等移动设备中。在未来工作中,可进一步改进网络,如将 32 位浮点数模型量化为 8 位整型数模型,从而更好部署于现场可编程门阵列等可编程逻辑器件设备中。

参 考 文 献

- [1] 李晓斌, 江碧涛, 杨渊博, 等. 光学遥感图像目标检测技术综述[J]. 航天返回与遥感, 2019, 40(4): 95-104.
Li X B, Jiang B T, Yang Y B, et al. A survey on object detection technology in optical remote sensing images[J]. Spacecraft Recovery & Remote Sensing, 2019, 40(4): 95-104.
- [2] Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with discriminatively trained part-based models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645.
- [3] Zhu C R, Zhou H, Wang R S, et al. A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features[J]. IEEE Transactions on Geoscience and Remote Sensing, 2010, 48(9): 3446-3456.
- [4] 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.
- [5] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 386-397.
- [6] 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.
- [7] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2020-04-02]. <https://arxiv.org/abs/1804.02767>.
- [8] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-02-03]. <https://arxiv.org/abs/2004.10934>.
- [9] Chen L Q, Shi W X, Deng D X. Improved YOLOv3 based on attention mechanism for fast and accurate ship detection in optical remote sensing images[J]. Remote Sensing, 2021, 13(4): 660.
- [10] 谢俊章, 彭辉, 唐健峰, 等. 改进 YOLOv4 的密集遥感目标检测[J]. 计算机工程与应用, 2021, 57(22): 247-256.
Xie J Z, Peng H, Tang J F, et al. Improved YOLOv4 for dense remote sensing target detection[J]. Computer Engineering and Applications, 2021, 57(22): 247-256.
- [11] 张欣, 张永强, 何斌, 等. 基于 YOLOv4-tiny 的遥感图像飞机目标检测技术研究[J]. 光学技术, 2021, 47(3): 344-351.
- [12] Zhang X, Zhang Y Q, He B, et al. Research on remote sensing image aircraft target detection technology based on YOLOv4-tiny[J]. Optical Technique, 2021, 47(3): 344-351.
- [12] 王伟锋, 金杰, 陈景明. 基于感受野的快速小目标检测算法[J]. 激光与光电子学进展, 2020, 57(2): 021501.
Wang W F, Jin J, Chen J M. Rapid detection algorithm for small objects based on receptive field block[J]. Laser & Optoelectronics Progress, 2020, 57(2): 021501.
- [13] 孙佳, 郭大波, 杨甜甜, 等. 基于改进的 YOLOv3 网络的实时目标检测[J]. 激光与光电子学进展, 2020, 57(22): 221505.
Sun J, Guo D B, Yang T T, et al. Real-time object detection based on improved YOLOv3 network[J]. Laser & Optoelectronics Progress, 2020, 57(22): 221505.
- [14] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[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: 4510-4520.
- [15] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [16] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation[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: 8759-8768.
- [17] 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.
- [18] Liu S, Di H, Wang Y. 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: 385-400.
- [19] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [20] Long Y, Gong Y P, Xiao Z F, et al. Accurate object localization in remote sensing images based on convolutional neural networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(5): 2486-2498.
- [21] Zou Z X, Shi Z W. Random access memories: a new paradigm for target detection in high resolution aerial remote sensing images[J]. IEEE Transactions on Image Processing, 2018, 27(3): 1100-1111.
- [22] Li K, Wan G, Cheng G, et al. Object detection in optical remote sensing images: a survey and a new

- benchmark[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 296-307.
- [23] Everingham M, Gool L, Williams C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.
- [24] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [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: 7132-7141.
- [25] 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.