计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2022,31(12):251−258 [doi: 10.15888/j.cnki.csa.008827]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于尺度感知的多路径特征融合目标检测①

潘　浩[1], 郑　华[1,2,3,4], 陈清俊[1], 廖晓琦[1], 王泓楷[1]

[1](福建师范大学 光电与信息工程学院, 福州 350007)
[2](福建师范大学 医学光电科学与技术教育部重点实验室, 福州 350007)
[3](福建师范大学 福建省光子技术重点实验室, 福州 350007)
[4](福建师范大学 福建省光电传感应用工程技术研究中心, 福州 350007)
通信作者: 郑　华, E-mail: hzheng@fjnu.edu.cn

摘　要: 在通用的目标检测算法中, 目标多变的尺度和特征融合利用一直是限制目标检测任务的难题. 针对上述问题, 首先文中提出了多路径特征融合模块, 模块采用跨尺度跨路径特征融合的方法, 强化输入输出特征之间的联系, 缓解了特征信息在传递时的稀释问题. 同时, 文中通过改进注意力模型提出了尺度感知模块, 该模块能根据目标的尺度自行地选择感受野大小, 从而使模型易于识别多尺度目标. 将尺度感知模块嵌入到多路径特征融合模块中, 使模型的特征提取和利用能力均得到提升. 经实验验证, 文中提出的算法在数据集 PASCAL VOC 和 MS COCO 上的平均检测精度分别达到了 82.2% 和 38.0%, 相比基线 FPN Faster RCNN 分别提升了 1.3% 和 0.6%, 其中对小尺度目标的检测效果提升最为显著.

关键词: 目标检测; 特征融合; 注意力机制; 尺度感知; 卷积神经网络

引用格式: 潘浩,郑华,陈清俊,廖晓琦,王泓楷.基于尺度感知的多路径特征融合目标检测.计算机系统应用,2022,31(12):251–258. http://www.c-s-a.org.cn/1003-3254/8827.html

Multi-path Feature Fusion Object Detection Based on Scale-aware

PAN Hao[1], ZHENG Hua[1,2,3,4], CHEN Qing-Jun[1], LIAO Xiao-Qi[1], WANG Hong-Kai[1]

[1](College of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou 350007, China)
[2](Key Laboratory of Optoelectronic Science and Technology for Medicine (Ministry of Education), Fujian Normal University, Fuzhou 350007, China)
[3](Fujian Provincial Key Laboratory of Photonic Technology, Fujian Normal University, Fuzhou 350007, China)
[4](Fujian Provincial Engineering Technology Research Center of Photoelectric Sensing Application, Fujian Normal University, Fuzhou 350007, China)

Abstract: The variable scales of objects and the use of feature fusion have been the challenges for popular object detection algorithms. Considering the problems, this study proposes a multi-path feature fusion module, which strengthens the connection between input and output features and alleviates the dilution of feature information in transmission by adopting cross-scale and cross-path feature fusion. Meanwhile, the study also proposes a scale-aware module by refining the attention model, which allows the model to easily recognize multi-scale objects by selecting the size of the receptive field corresponding to the scale of the objects independently. After the scale-aware module is embedded into the multi-path feature fusion module, the feature extraction and utilization abilities of the model are improved. The experimental results reveal that the proposed method achieves 82.2 mAP and 38.0 AP on PASCAL VOC and MS COCO datasets, respectively, an improvement of 1.3 mAP and 0.6 AP over the baseline FPN Faster RCNN, respectively, with the most significant improvement in detection of small-scale objects.

Key words: object detection; feature fusion; attention mechanism; scale aware; convolutional neural networks

① 基金项目: 福建省高校产学合作项目 (2021H6025)
收稿时间: 2022-03-25; 修改时间: 2022-04-22; 采用时间: 2022-04-29; csa 在线出版时间: 2022-07-29

1 引言

近年来, 深度学习方法在图像分类、目标检测和文本识别等计算机视觉领域获得了显著的成功[1−3]. 卷积神经网络通过自底向上构建多层卷积和下采样, 用来获得不同分辨率的特征表示, 在计算机视觉任务模型构建中已然成为一种范例. 在目标检测任务中, 由于不同目标的尺度多变, 在不同环境中的同一类别目标尺度不一, 如何识别多尺度目标一直是目标检测模型的一个难题, 为了提升目标检测的效果, 研究人员在目标检测模型的多个方面进行改进.

在各类计算机视觉任务中, 通过构建特征金字塔形式的结构逐渐成为一种通用模式. 其中最具代表性的结构为 Lin 等人提出的 FPN[4], FPN 通过增加自上而下的路径和横向连接来增强底层特征图的语义信息. 后续基于 FPN 改进的 PAFPN[5] 和 BiFPN[6] 等均通过增加额外的信息传递路径来丰富各阶段特征图的信息内容; 为了更充分的传递信息, FPG[7] 设计了深度多层的网络结构. 不同于上述手工设计特征融合路径的方式, NASFPN[8] 采用神经结构搜索的方法让网络自行寻找最优的特征融合方式. DFP[9] 和 BFP[10] 考虑到不同阶段特征的信息鸿沟, 将各阶段特征图先进行融合再来传递全局信息. 然而多数改进的 FPN 模型往往具有更长的信息传递路径, 这极大地稀释了 FPN 基础结构中传递的语义信息. 针对长路径特征融合带来的信息稀释问题, 本文提出多路径特征融合模块 (multi-path feature fusion module, MPFFM), 用来增强输入阶段到输出阶段特征图间信息的传递, 通过缩短信息传递路径来缓解信息稀释.

为了获得更好的特征表示, 注意力机制广泛应用于各类深度学习任务中[11−14]. SENet[11] 通过建立特征图通道的权重函数学习各个通道的重要性, CBAM[12] 在 SENet[11] 的基础上在特征图空间维度引入注意力, 从而进一步提升特征表示能力. Li 等人提出的 SKNet[13] 通过建立多路具有不同感受野特征图的权重函数, 使卷积神经网络能够自行选择不同感受野大小来适应多尺度目标. 但其结构中含有多层全连接网络来生成注意力权重, 这占用了大量参数, 并且采用了空洞卷积来提取不同尺度的特征, 这可能会导致特征信息的不连续, 影响目标的定位效果. 本文通过改进 SKNet 结构并将其引入到目标检测模型中, 称为尺度感知模块 (scale-aware module, SAM), 相比 SKNet, SAM 在减小参数量的同时增加了相邻通道间信息的传递, 通过建立输入输出特征的残差连接防止特征信息被过度处理, 来进一步完善模块的特征表征能力. 本文将 SAM 嵌入到 MPFFM 中, 在特征图的通道中增加尺度感知信息, 来增强模型对不同尺度目标的识别能力.

本文运用多路径特征融合模块, 通过建立跨尺度跨路径连接, 来增强输入特征到输出特征的信息传递; 并改进了 SKNet[13] 的结构, 在避免信息丢失的同时增加了特征图通道的信息交互, 通过建立残差连接缓解特征信息被过度处理的弊端; 最后通过在公开的基准数据集 PASCAL VOC[14] 和 MS COCO[15] 上进行测试, 相对于基线方法, 本文提出的尺度感知的多路径特征融合目标检测算法取得了更优的结果.

2 尺度感知的多路径特征融合模型

2.1 网络结构

本文提出的尺度感知多路径特征融合模型结构如图 1 所示.

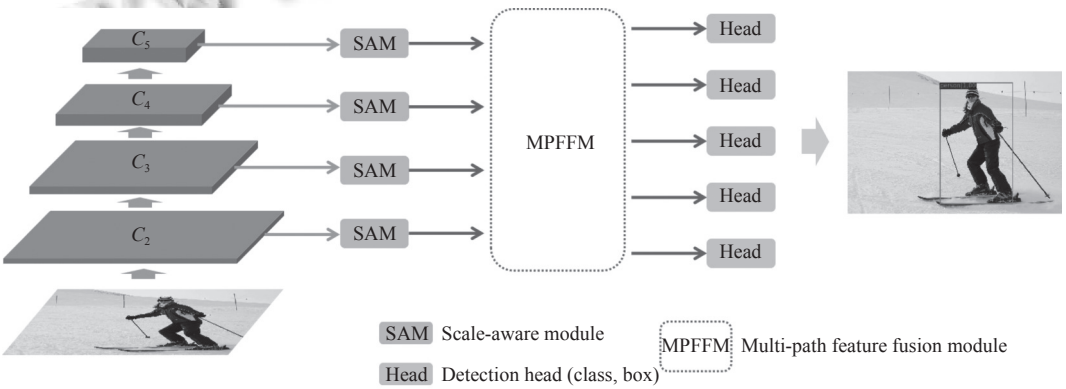


图 1 尺度感知多路径特征融合模型

整体的网络结构主要包含了 4 个部分: 骨干网络、尺度感知模块 (SAM)、多路径特征融合模块 (MPFFM) 和检测头模块. 模型在训练和评估时其他部分的结构设计如区域建议网络、检测头模块和损失函数等均与模型 Faster RCNN[16] 保持一致. 输入的图片在经骨干网络提取特征后, 取出不同尺度的 4 个阶段特征图来构建特征金字塔, 由于每个阶段特征图内的目标尺度同样具有差异, 因此各阶段特征图首先送入尺度感知模块实现多尺度特征提取, 让每个阶段特征图自行选择感受野大小来适应不同尺度的目标, 然后再进行下游的特征融合部分. 由卷积神经网络的性质可知, 浅层特征图拥有高分辨率和强定位信息, 深层特征图具有低分辨率和强语义信息, 为了使深层和浅层特征图信息之间能够相互补充, 本文提出了多路径特征融合模块, 强化输入特征图到输出特征图的语义信息传递, 同时多路径的设计使深层特征图的语义信息和浅层特征图的定位信息在整个模块中得到充分融合.

2.2 多路径特征融合模块 MPFFM

图 2 展示了几种不同的特征金字塔结构, 对比各子图可以发现, 虽然 BiFPN[6] 对 PAFPN[5] 结构进行了有效的改进, 但由于其结构中输入特征到输出特征的映射路径繁杂, 基于 FPN[4] 的自上而下的语义增强路径到输出特征图时被稀释, 淹没在众多信息流中. 因此本文提出了多路径特征融合模块 (MPFFM), 如图 2(d) 所示, 由 3 部分组成: 自上而下路径、残差跨尺度连接和自下而上路径, 通过在输入特征图与输出特征图之间建立跨尺度连接, 来增强不同尺度特征图之间的信息传递, 缓解特征信息在传递时的稀释问题, 从而突出输入的本征特征的重要性. MPFFM 中 $\{C_2, C_3, C_4, C_5\}$ 是骨干网络如 ResNet[17] 的 4 个阶段的输出特征图, 特征图 $\{P_2, P_3, P_4, P_5\}$ 由维度为 1×1×256 的卷积层将特征图 $\{C_2, C_3, C_4, C_5\}$ 通道统一为 256, 特征层 P_6 由 P_5 经步长为 2 尺寸为 3×3 的卷积层下采样得到.

自上而下路径得到的特征图 M_3、M_4、M_5 和 N_2 由式 (1) 计算:

$$\begin{cases} M_5 = Conv(P_5 + Conv(f_u(P_6))) \\ M_i = Conv(P_i + Conv(f_u(M_{i+1}))), i \in \{3,4\} \\ N_2 = Conv(P_2 + Conv(f_u(M_3))) \end{cases} \tag{1}$$

残差跨尺度连接和自下而上路径获得的特征图 N_3、N_4、N_5 和 N_6 由式 (2) 计算:

$$\begin{cases} N_i = Conv\begin{pmatrix} P_i + M_i + f_d(N_{i-1}) + \\ Conv(f_u(P_{i+1})) \end{pmatrix}, i \in \{3,4\} \\ N_5 = Conv(P_5 + M_5 + f_d(N_4)) \\ N_6 = Conv(P_6 + f_d(N_5)) \end{cases} \tag{2}$$

各层特征图的维度由式 (3) 表示:

$$P_i, M_i, N_i \in \mathbb{R}^{C \times \frac{H}{2^i} \times \frac{W}{2^i}} \tag{3}$$

在式 (1) 和式 (2) 中, $f_u(\cdot)$ 为上采样函数, 采用尺度因子为 2 最近邻插值的方法, $f_d(\cdot)$ 为下采样函数, 通过步长为 2 尺寸为 3×3 的卷积层实现, *Conv* 为保持特征图输入输出分辨率不变的 3×3 卷积运算. 每次上采样之后会紧跟一层卷积用来特征对齐, 同时跨尺度特征相加时的卷积层用来消除混叠效应. 式 (3) 中参数 C、H 和 W 分别表示为特征图的通道数、高和宽, 参数 i 代表特征图的不同阶段, 从而对应着特征图不同的尺度.

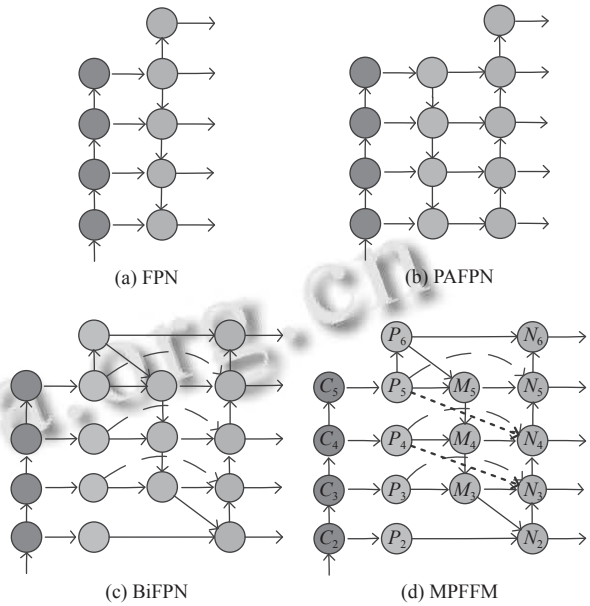


图 2 各类特征金字塔结构对比

2.3 尺度感知模块 SAM

对于目标的多尺度问题, RFB[18] 通过组合不同感受野大小的特征图能够有效地适应多尺度目标, 但这些特征图进行融合时往往具有相同的权重, 这样的设计可能造成很多冗余的信息被保留, 导致模型不能辨别不同大小感受野特征图各自的重要性. SKNet[13] 通过建立单个通道上不同感受野特征图的权重函数, 让卷积神经网络自行判断不同感受野各自的重要性, 将

其嵌入在骨干网络中用来增强模型多尺度特征提取能力, 在图像分类任务中获得了提升.

为了增强目标检测模型对不同尺度目标的感知能力, 本文通过改进 SKNet 结构使其适用于目标检测任务, 使模型只需要增加少量的参数就能捕获到不同尺度的信息. 如图 3 所示, 本文提出的尺度感知模块 (SAM) 不同于 SKNet 使用多路空洞卷积来获取多尺度特征, 而是采用卷积层级联的形式获取多尺度特征, 从不同的节点输出以实现输出不同感受野大小的特征图, 如尺寸为 5×5 卷积核的感受野可由两个级联形式的 3×3 卷积核代替, 这样的设计实现了参数共享并且避免了使用空洞卷积可能带来的信息不连续等问题. 图 3 中输入特征图向量 $X \in \mathbb{R}^{C\times H\times W}$ 和向量 $O_3 \in \mathbb{R}^{C\times H\times W}$ 之间、向量 O_3 和向量 $O_5 \in \mathbb{R}^{C\times H\times W}$ 之间均经过一层 3×3 卷积和激活函数 ReLU, 以此来获取具有不同感受野的特征向量 O_3 和 O_5, 将向量 O_3 和 O_5 按元素相加得到向量 O. 图 3 中 $f_g(\cdot)$ 为全局平均池化函数, 如式 (4) 所示:

$$G^c = f_g(O^c) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} O^c(i,j),\ c\in\{1,2,\cdots,C\} \tag{4}$$

其中, G^c 代表向量 $G \in \mathbb{R}^{C\times 1\times 1}$ 的第 c 位元素的取值, O^c 代表特征向量 $O \in \mathbb{R}^{C\times H\times W}$ 的第 c 层通道的特征图. 受 ECANet[19] 启发, 为了避免在维度变换时通过压缩通道来减小参数量而带来的信息丢失, 本文在通道权重向量 $G \in \mathbb{R}^{C\times 1\times 1}$ 和向量 $E \in \mathbb{R}^{C\times 1\times 1}$ 之间采用一维卷积代替原结构中的第一层全连接网络, 在减少参数量的同时避免了因压缩通道带来的信息丢失, 并且增加了相邻通道间的信息交互, 增强特征利用, 向量 G 到 E 的变换如式 (5) 所示:

$$E = f_e(G) = \sigma(Conv1d(G)) \tag{5}$$

其中, $Conv1d$ 本文采用尺寸为 1×5 的一维卷积, σ 采用激活函数 ReLU 增加模型非线性拟合能力. 函数 $f_c(\cdot)$ 为全连接层, 本文采用尺寸为 $1\times 1\times mC$ 的卷积层来升高维度, 得到向量 $H \in \mathbb{R}^{mC\times 1\times 1}$, 其中 m 为向量 X 和向量 O 之间级联的 3×3 卷积层的层数, 图 3 仅画出了 m 取 2 时两层 3×3 卷积的特殊情况, 输出只包含 3×3 和 5×5 感受野大小的特征图 O_3 和 O_5.

图 3 中将向量 $H \in \mathbb{R}^{2C\times 1\times 1}$ 按通道分为向量 $a \in \mathbb{R}^{C\times 1\times 1}$ 和向量 $b \in \mathbb{R}^{C\times 1\times 1}$, a 和 b 先逐元素成对送入 Softmax 函数进行权重重新分配, 如式 (6) 所示:

$$\begin{cases} a_i = \dfrac{e^{a_i}}{e^{a_i}+e^{b_i}},\ i\in\{1,2,\cdots,C\} \\ b_i = \dfrac{e^{b_i}}{e^{a_i}+e^{b_i}},\ i\in\{1,2,\cdots,C\} \end{cases} \tag{6}$$

经权重分配后的向量 a 和 b 再经广播机制后维度变为 $a \in \mathbb{R}^{C\times H\times W}$ 和 $b \in \mathbb{R}^{C\times H\times W}$, 得到的 a 和 b 分别代表特征图 O_3 和 O_5 的权重, 再与 O_3 和 O_5 逐元素相乘后得到通道带关联权重具有不同感受野大小的特征图 $\tilde{O}_3 \in \mathbb{R}^{C\times H\times W}$ 和 $\tilde{O}_5 \in \mathbb{R}^{C\times H\times W}$, 将特征图 $\tilde{O}_3$ 和 $\tilde{O}_5$ 按元素相加, 相加后特征图的每层通道均受到两种感受野的影响, 从而实现模型自行选择不同大小感受野的能力. 本文还在输入特征图 $X \in \mathbb{R}^{C\times H\times W}$ 和输出特征图 $Y \in \mathbb{R}^{C\times H\times W}$ 之间增加了残差连接, 用来防止特定通道的信息在权重分配时被过度放大或者抑制. 本文的 SAM 算法实现流程如图 4 所示.

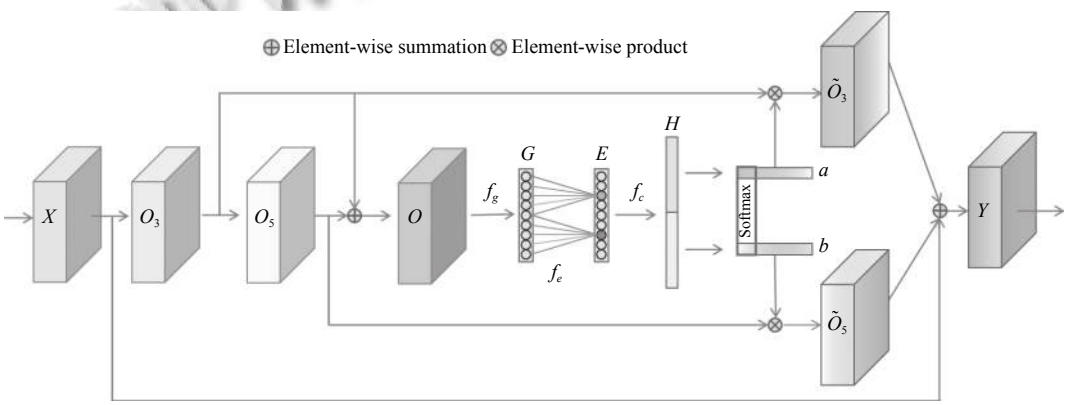


图 3 尺度感知模块

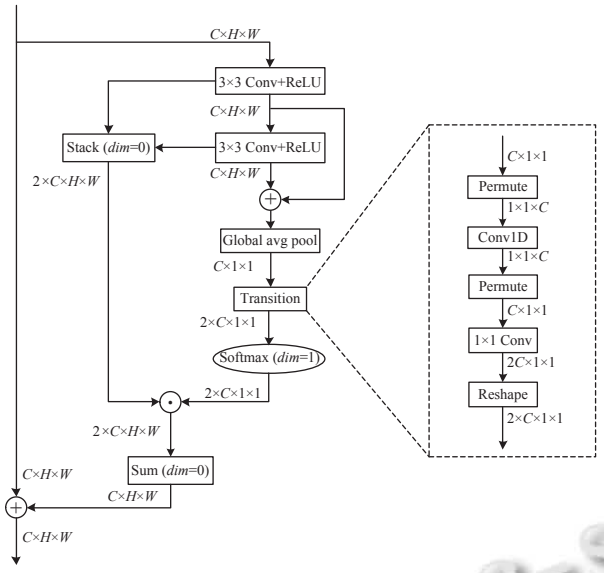


图 4 SAM 算法流程图

3 实验及分析

本节给出了本文的模型分别在目标检测数据集 PASCAL VOC 和 MS COCO 上的评估结果. PASCAL VOC 数据集包括 20 种类别, 数据的训练集和测试集的划分根据通用做法[16], 在 VOC2007 和 VOC2012 训练集的联合集上对所有模型进行训练, 并在 VOC2007 测试集上进行评估, 模型评估采用平均精度 (average precision, AP) 和均值平均精度 (mean average precision, mAP) 作为精度性能评估指标. MS COCO 数据集包含 80 种类别, 其中含有数量为 118k 张图片 (train-2017) 用于训练和 5k 张图片 (val-2017) 用于验证, 性能评估指标 AP_{50} 和 AP_{75} 分别为交并比为 0.5 和 0.75 时的平均精度, AP 为综合的性能评估指标, 并且对大、中、小 3 种尺寸目标的识别能力分别进行了评估. 同时本节给出了本文方法与其他基于 FPN 的改进方法的对比结果, 并通过消融实验验证了各模块的有效性.

3.1 实验细节

为了公平对比实验结果, 本文实验是基于 PyTorch 框架在开源模型库 mmdetection 上实现, 使用模型库中公开的骨干网络, 并采用预训练模型微调, 在单卡 GeForce GTX 1 080 Ti 上进行训练. 在训练 PASCAL VOC 数据集时, 本文采用 SGD 作为优化器, momentum 为 0.9, weight_decay 为 0.000 1, batch_size 为 2, 初始学习率为 0.001 25, 总训练轮数为 12, 训练到第 9 轮时学习率乘以 0.1. 对于训练 MS COCO 数据集, 初始学习率为 0.002 5, 在第 8 和第 11 轮时学习率均乘以 0.1, 其他的超参数均与训练 PASCAL VOC 数据集时保持一致. 在训练过程中采用随机翻转等数据增强技术, 以增强模型的鲁棒性和泛化能力.

3.2 消融实验

为了分析本文提出的两种模块的有效性, 本节设计了消融实验. 所有评估过程均在数据集 PASCAL VOC2007 test 上进行, 基线采用主干网络为 ResNet-50-FPN 的目标检测框架 Faster RCNN. 实验结果如表 1 所示, 为了详细对比各模块对检测结果的提升效果, 基线模型在仅加入尺度感知模块 (SAM) 时, 检测精度提升了 0.5%, 仅将多路径特征融合模块 (MPFFM) 代替 FPN 时, 模型检测精度提升了 0.7%. 同时应用两个模块时, 检测精度相对于基线提升了 1.3%, 而且两个模块组合使用时获得了额外的提升, 验证了尺度感知模块和多路径特征融合模块的有效性.

表 1 PASCAL VOC2007 test 上的消融试验

Detector	FPN	MPFFM	SAM	mAP (%)
Faster RCNN	√	×	×	80.9
	√	×	√	81.4
	×	√	×	81.6
	×	√	√	**82.2**

3.3 实验结果

表 2 详细对比了改进前后模型在数据集 PASCAL VOC2007 test 上对 20 类目标的检测结果, “*”表示本文复现的 FPN 和 BiFPN 结构. 为公平对比, 本文复现的 BiFPN 和 MPFFM 均未设置权重, 同时均使用普通卷积. 在相同的标准下, 本文提出的 MPFFM 相比 BiFPN 和 FPN 精度分别提升了 0.4% 和 0.7%, 明通过增加额外的跨尺度连接可以有效提升目标检测效果, 也表明了由骨干网络输入的特征图信息对相邻尺度输出阶段特征图的重要性. 同时将本文提出的 SAM 与 SKNet 进行了比较, 在均使用 MPFFM 结构时, 模型的检测精度分别达到了 82.2% 和 81.8%, 表明改进的注意力模型 SAM 具有更优秀的性能. 从表 2 的最后一行可以看出, 本文的方法在多个类别目标的识别精度上都能取得最优的结果.

表 3 对比了本文方法与主流算法在 MS COCO 上的检测结果, 同时对各种方法的参数量进行了比较, 并且在骨干网络为 ResNet-50 和 ResNet-101 时分别进行了实验验证, 表中“*”表示在 mmdetection 中复现的结

果. 通过对比可以发现, 相对于基线方法, 当使用 ResNet-50 作为骨干网络时, 本文的方法在引入少量参数 (+5.1M) 的同时精度得到有效提升 (+0.6%), 在各个尺度目标上的检测精度均有提升, 特别是对小目的精度提升最明显 (+1.4%). 当骨干网络为 ResNet-101 时, 虽然模型检测精度仅比基线提升 0.4%, 但对小目标的检测效果依旧能获得更高的精度提升 (+1.6%). 综合对比各尺度目标检测精度的提升效果可以看出, 本文方法比基线方法更能适应目标尺度的变化, 并且在使用轻量的骨干网络时, 本文算法相对于其他主流的方法具有更高的效益, 同时对小目标的检测始终保持着较高精度.

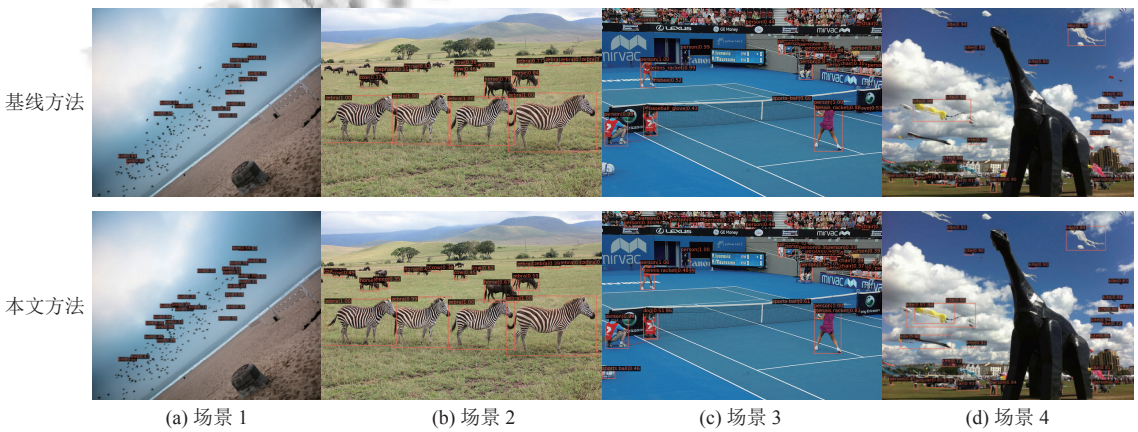
图 5 展示了本文算法与基线方法进行目标检测的可视化对比图, 本文从 MS COCO 数据集中选出了 4 张极具挑战性的图片进行对比实验. 在图 5(a) 和图 5(b) 中, 本文方法相比基线方法能够识别出更多的小目标, 这得益于尺度感知模块的设计, 使得模型能自行选择感受野大小, 能够对小尺寸感受野分配更大的权重. 对比图 5(c) 中的结果, 对于部分遮挡的目标, 如图中观众席坐着的观众和左下角被遮挡的目标均能够被识别出来, 而且比基线方法识别出更多, 说明本文的模型能充分利用环境中的语义信息进行推测, 被遮挡目标的特征信息经多路径特征融合模块后, 得到深层特征的语义信息补充从而能被成功识别. 对比图 5(d) 结果, 本文检测出来的目标具有更高的分类得分而且定位精度更高, 这得益于本文的多路径特征融合模块的设计, 能够充分传递语义信息和定位信息.

表 2 基于 Faster RCNN 在 PASCAL VOC2007 test 上 20 类的检测结果 (%)

Pyramid	mAP	aero	bick	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
FPN[4]*	80.9	86.2	86.3	83.7	71.5	52.4	85.1	88.5	89.1	65.7	86.5	76.0	88.1	87.5	85.7	85.9	56.7	83.4	77.4	85.7	77.2
BiFPN[6]*	81.2	86.5	**87.7**	**85.0**	70.4	70.1	87.1	88.4	89.3	66.3	86.9	76.1	88.8	84.0	84.6	85.9	56.7	83.5	78.8	86.3	78.1
MPFFM	81.6	86.2	87.0	80.0	71.1	72.3	85.9	88.8	**89.8**	**67.8**	87.1	76.7	89.3	88.2	85.4	86.1	58.7	83.9	79.7	85.9	82.4
SKNet[13] +MPFFM	81.8	86.0	87.5	79.7	**74.7**	**73.6**	**87.2**	**88.9**	89.7	66.7	87.3	**76.8**	**89.4**	88.7	86.0	**86.4**	55.7	85.7	79.4	86.5	79.4
SAM + MPFFM	**82.2**	**86.6**	86.5	**85.0**	72.2	72.8	85.6	88.8	89.7	67.2	**88.3**	75.8	88.7	**88.9**	**86.7**	86.2	**59.3**	**85.8**	**79.6**	**86.9**	**82.8**

表 3 COCO 2017-val 上的检测结果对比

Method	Pyramid	Backbone	#params (M)	AP (%)	AP_{50} (%)	AP_{75} (%)	AP_S (%)	AP_M (%)	AP_L (%)
Faster RCNN[16]*	FPN[4]	ResNet-50	41.2	37.4	58.1	40.4	21.2	41.0	48.1
Faster RCNN[16]*	PAFPN[5]		52.2	37.5	58.6	40.8	21.5	41.0	48.6
RetinaNet[20]*	FPN		37.7	36.5	55.4	39.1	20.4	40.3	48.1
RetinaNet[20]	CE-FPN[21]		65.0	37.5	57.3	40.2	21.6	41.2	48.7
YOLOF[22]	—		44.1	37.7	56.9	40.6	19.1	**42.5**	**53.2**
本文	SAM+MPFFM		46.6	**38.0**	**59.0**	**41.5**	**22.6**	41.3	49.1
Faster RCNN[16]*	FPN	ResNet-101	60.2	39.3	60.0	43.3	22.1	43.5	51.3
RetinaNet[20]*	FPN		56.6	38.5	57.6	41.0	21.7	42.8	50.4
YOLOF[22]	—		63.2	**39.8**	59.4	42.9	20.5	**44.5**	**54.9**
本文	SAM+MPFFM		65.5	**39.7**	**60.3**	**43.4**	**23.7**	43.7	51.8



(a) 场景 1 (b) 场景 2 (c) 场景 3 (d) 场景 4

图 5 基线方法与本文方法的可视化对比

图 6 为模型改进前后的训练损失曲线图, 可以发现, 与原 Faster RCNN FPN 模型相比, 本文改进后的模型在整个训练过程的前段和后段始终保持着更低的损失, 表明本文的方法相较于原方法具有更好的收敛效果.

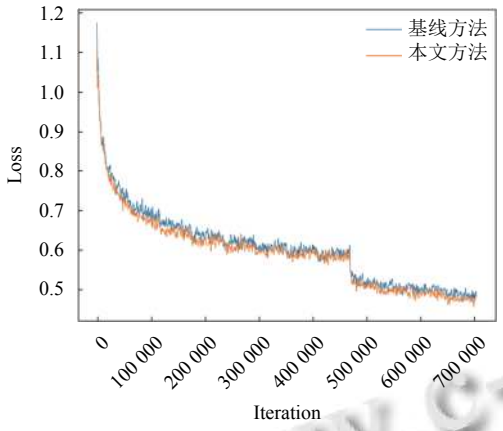


图 6 模型训练的 Loss 对比图

4 结论

本文提出了一种基于尺度感知的多路径特征融合的目标检测算法, 其中包含了多路径特征融合模块和尺度感知模块两个主要部分. 多路径特征融合模块通过增加输入输出特征图之间的跨尺度连接, 增强了各尺度特征图之间的信息传递, 有效地缓解了信息在长路径传递时的稀释问题. 同时通过改进注意力模型而提出的尺度感知模块, 使模型在进行非线性变换时, 减少了信息的丢失, 加强模型的特征利用能力. 将尺度感知模块嵌入到多路径特征融合模块中, 能够有效增强模型的尺度鲁棒性和特征复用能力. 通过在通用的目标检测任务数据集上验证本文算法, 本文方法与基线方法相比在多种类别和各尺度大小目标上的检测精度均获得了提升, 同时本文方法与其他同类方法相比具有更高的效益, 并且对小尺度目标识别保持优异的性能, 验证了本文模型的有效性.

参考文献

1 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc., 2012. 1097–1105.

2 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556, 2014.

3 LeCun Y, Bottou L, Bengio Y, *et al*. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278–2324. [doi: 10.1109/5.726791]

4 Lin TY, Dollár P, Girshick R, *et al*. Feature pyramid networks for object detection. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 936–944.

5 Liu S, Qi L, Qin HF, *et al*. Path aggregation network for instance segmentation. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8759–8768.

6 Tan MX, Pang RM, Le QV. Efficientdet: Scalable and efficient object detection. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 10778–10787.

7 Chen K, Cao YH, Loy CC, *et al*. Feature pyramid grids. arXiv:2004.03580, 2020.

8 Ghiasi G, Lin TY, Le QV. NAS-FPN: Learning scalable feature pyramid architecture for object detection. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 7029–7038.

9 Kong T, Sun FC, Tan CB, *et al*. Deep feature pyramid reconfiguration for object detection. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 172–188.

10 Pang JM, Chen K, Shi JP, *et al*. Libra R-CNN: Towards balanced learning for object detection. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 821–830.

11 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141.

12 Woo S, Park J, Lee JY, *et al*. CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 3–19.

13 Li X, Wang WH, Hu XL, *et al*. Selective kernel networks. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 510–519.

14 Everingham M, van Gool L, Williams CKI, *et al*. The PASCAL visual object classes (VOC) challenge. International Journal of Computer Vision, 2010, 88(2): 303–338. [doi:

10.1007/s11263-009-0275-4]

15 Lin TY, Maire M, Belongie S, *et al*. Microsoft COCO: Common objects in context. 13th European Conference on Computer Vision. Zurich: Springer, 2014. 740–755.

16 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2015. 91–99.

17 He KM, Zhang XY, Ren SQ, *et al*. Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778.

18 Liu ST, Huang D, Wang YH. Receptive field block net for accurate and fast object detection. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 404–419.

19 Wang QL, Wu BG, Zhu PF, *et al*. ECA-Net: Efficient channel attention for deep convolutional neural networks. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020. 11531–11539.

20 Lin TY, Goyal P, Girshick R, *et al*. Focal loss for dense object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318–327. [doi: 10.1109/TPAMI.2018.2858826]

21 Luo YH, Cao X, Zhang JT, *et al*. CE-FPN: Enhancing channel information for object detection. Multimedia Tools and Applications, 2022. [doi: 10.1007/s11042-022-11940-1]

22 Chen Q, Wang YM, Yang T, *et al*. You only look one-level feature. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13034–13043.

(校对责编: 牛欣悦)