

DOI:10.19651/j.cnki.emt.2107598

基于改进特征金字塔的目标检测*

史晨晨¹ 张长伦^{1,2} 何强¹ 王恒友¹

(1.北京建筑大学理学院 北京 100044; 2.北京建筑大学北京未来城市设计高精尖创新中心 北京 100044)

摘要: 特征金字塔网络(FPN)已经成为目标检测中提取多尺度特征的有效框架。然而,FPN存在着由于通道减少导致语义信息丢失、高层特征只包含单尺度的上下文信息和具有语义差别的不同层特征直接融合造成混叠效应等问题。针对上述问题,提出了基于注意力增强指导的特征金字塔网络。该模型由通道特征增强模块、上下文增强模块和注意力指导特征融合模块3个部分组成。具体来说,通道特征增强模块通过建模特征之间的依赖关系减轻由于通道减少造成的信息损失,上下文增强模块利用不同级别特征进行上下文信息提取以增强高层特征,注意力指导特征融合模块利用注意力机制指导相邻层特征学习来增进彼此语义信息的一致性。将Faster R-CNN和Mask R-CNN目标检测器中的FPN替换为本文模型并在不同的数据集上进行实验,实验结果表明,改进后的Faster R-CNN检测器在PASCAL VOC和MS COCO数据集上的平均精度分别提高1.5%和1%,改进后的Mask R-CNN检测器在MS COCO数据集上也分别将Mask AP和Box AP的性能提升了0.8%和1.1%。

关键词: 特征金字塔;目标检测;注意力;特征增强;特征融合

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Object detection based on improved feature pyramid

Shi Chenchen¹ Zhang Changlun^{1,2} He Qiang¹ Wang Hengyou¹

(1. Science School, Beijing University of Civil Engineering and Architecture, Beijing 100044, China; 2. Beijing Advanced Innovation Center for Future Urban Design, Beijing University of Civil Engineering and Architecture, Beijing 100044, China)

Abstract: Feature pyramid network (FPN) has become an effective framework for extracting multi-scale features in object detection. However, FPN has problems such as loss of semantic information due to channel reduction, high-level features only contain single-scale context information, and the direct fusion of different layer features with semantic differences resulting in aliasing effects. In response to the above problems, proposes a feature pyramid network based on attention enhancement guidance, which is composed of channel feature enhancement module, context enhancement module and attention guidance fusion module. Specifically, the channel feature enhancement module reduces the information loss caused by channel reduction by modeling the dependency between the features, the context enhancement module uses different levels of features to extract context information to enhance high-level features, and the attention guidance feature fusion module uses the attention mechanism to guide the feature learning of adjacent layers to enhance the consistency of semantic information with each other. This paper replaces the FPN in the Faster R-CNN and Mask R-CNN object detectors with AEGFPN and performs experiments on different data sets, which experimental results show that the average accuracy of the improved Faster R-CNN detector on the PASCAL VOC and MS COCO datasets is increased by 1.5% and 1%, respectively, and the improved Mask R-CNN detector also improves the performance of Mask AP and Box AP by 0.8% and 1.1% on the MS COCO data set.

Keywords: feature pyramid; object detection; attention; feature enhancement; feature fusion

0 引言

目标检测是计算机视觉领域的一项基本任务,随着深

度学习的不断发展目标检测如今已成为该领域中的一个研究热点,并被广泛应用于多个领域,如人体姿态检测^[1]、行人检测^[2]、安全预警与检测^[3-4]等。

收稿日期:2021-08-15

* 基金项目:国家自然科学基金(62072024)、北京建筑大学北京未来城市设计高精尖创新中心项目(UDC2017033322,UDC2019033324)、北京建筑大学市属高校基本科研业务费专项资金(X20084,ZF17061)资助

随着卷积神经网络(convolutional neural network, CNN)^[5]的发展,近年来开发出了许多具有显著性能的检测器,这些目标检测器一般可分为两类。第 1 类是一阶段的检测算法,这类算法是端到端的来检测物体,直接在特征图上实现定位和分类,效率更高,但精确度相对较低,代表算法有 YOLO v3^[6]、SSD^[7]、DSSD^[8]、RetinaNet^[9]等。第 2 类算法是两阶段检测算法,这类算法对特征图对应的原图预设多种锚框,在输入的图像上生成一系列的候选区域,然后利用分类器和回归器对感兴趣区域进行细化,精度高,但效率相对较低,代表算法有 R-CNN^[10]、Fast R-CNN^[11]、Faster R-CNN^[12]、Mask R-CNN^[13]等。

无论是一阶段检测算法还是两阶段算法,为解决多尺度处理问题,广泛采用了特征金字塔网络(feature pyramid network, FPN)^[14]。FPN 网络使用自上而下的融合路径,将语义信息从高层次传播到低层次来构建多尺度特征,进而提高检测器的性能。但是 FPN 这种简单的设计会影响特征金字塔进一步学习更多的区别特征。其主要有如下几点缺陷。首先,在特征提取中,使用 1×1 卷积的横向连接将来自骨干网络的特征 C_i 提取为有相同通道维度的特征 F_i ,在这个特征提取过程中由于通道维度减少会造成严重的信息丢失(例如 C_5 层从 2 048 到 256)。其次,在自上而下的路径中,最高层特征并未接受来自其他层的特征信息,只包含单尺度的上下文信息,缺乏其他尺度的特征信息。最后,在跨尺度特征融合时不同层特征的语义信息存在着语义差别,然而 FPN 并没有考虑语义信息的差别,直接把高层特征上采样后直接与下一层直接相加得到融合特征,限制了特征学习和表示。针对 FPN 以上存在的问题,近年来也有一些工作对其进行不断改进。PANet^[15]在原有的结构上额外添加了自下而上的路径,以进一步增加深层的低级信息。NAS-FPN^[16]采用神经网络结构搜索来学习所有跨尺度连接之间的更好融合。EfficientDet^[17]提出了一种加权双向的 FPN 来实现简单高效的特征融合。这些方

法大部分在特征融合过程进行不同的改进,在一定程度上缓解这些问题。

另外,近年来注意力机制和上下文信息提取等工作被广泛应用于各种计算机视觉任务,进一步促进了计算机视觉检测算法的性能提升。在注意力方面,Non-Local^[18]提出用于获得两个位置之间的依赖关系,解决卷积操作在每一层获得的感受野有限问题。SENet^[19]提出通道加权注意力机制,通过加权解决不同通道的重要程度。后来,又有许多学者提出了含通道和空间双注意力机制如 CBAM^[20]、DANet^[21]等。受此启发,PAN^[22]和 DFN^[23]将注意力应用于特征金字塔网络,用于指导不同层特征间的学习。同样,在上下文信息提取方面,ASPP^[24]利用上下文信息来增强特征学习和表达,AC-FPN^[25]从多个感受野中探索大量上下文信息用于增强特征金字塔网络的高层特征等。

本文通过引入基于注意力增强机制指导特征金字塔网络学习,进而提高了基于特征金字塔的目标检测器的检测与分割性能,并通过在 PASCAL VOC 和 MS COCO 数据集上用不同类型的检测器对本文模型进行验证和评估。实验表明,本方案不仅可以在不同数据集上带来检测精度的提升,也可以对不同框架的检测器带来性能的提升。

1 基本原理

在本节中,将介绍基于注意力增强指导的特征金字塔网络(attention enhancement guidance feature pyramid network, AEGFPN)。AEGFPN 相对于 FPN 增加了 3 个组件:通道特征增强(channel feature enhancement, CFE)、上下文增强(context enhancement, CE)和注意力指导特征融合(attention guidance fusion, AGF)。在 AEGFPN 中按照 FPN(如图 1(a)所示)的网络架构将骨干网络的输出特征表示为 $\{C_2, C_3, C_4, C_5\}$,融合前的特征提取记为 $\{F_2, F_3, F_4, F_5\}$,金字塔多尺度特征记为 $\{P_2, P_3, P_4, P_5\}$ 。AEGFPN 整体框架如图 1(b)所示。

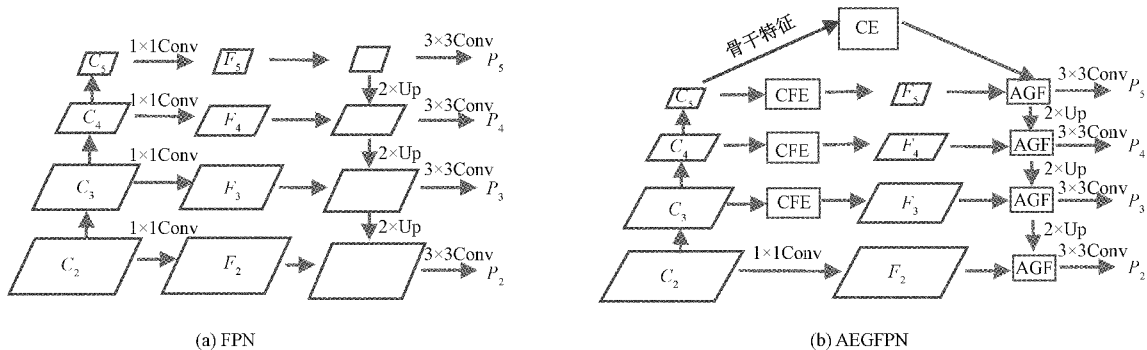


图 1 FPN 与 AEGFPN 网络架构

1.1 通道增强(CFE)

特征金字塔网络 FPN 在构建自上而下的路径前,首先会将骨干网络的输出特征 C_i 进行通道维度降维,通过 1×1 的卷积生成具有相同通道的特征层 F_i 。这

样做的目的是为了减少网络的计算复杂度,但是也会带来特征的信息损失,尤其是在高层特征中这样的直接降维会造成高层特征语义信息更多的损失,一定程度上影响高层特征表达。从而限制特征金字塔网络更好的学

习多尺度特征。因此,本文受注意力机制 Non-Local 和 DANet 关于学习特征的全局依赖关系来增强特征的表达的启发,提出了通道增强模块(Δ FE),替换原来的 1×1 卷积,通过对通道之间的相互依赖性进行建模,用

来学习特征 C_i 和特征 F_i 之间的通道全局特征依赖关系,增进通道之间的相关性进而增强缩减后的特征表达,减轻由通道减少带来的特征损失, Δ FE 详细网络结构如图 2 所示。

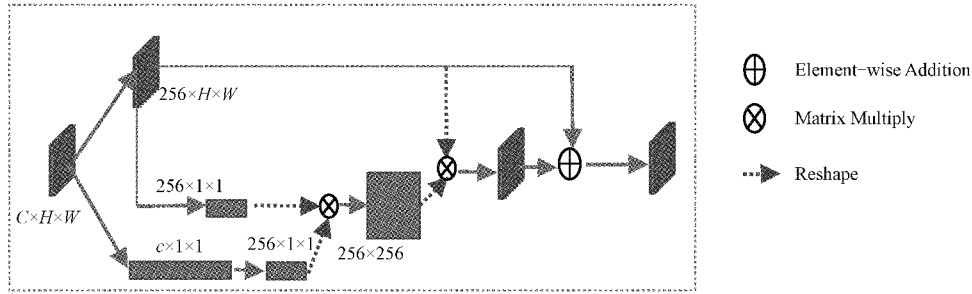


图 2 通道增强网络结构图

在 AFE 中,给定输入特征 C_i , 利用 3×3 卷积对特征 C_i 进行通道降维得到中间特征 F_i^1 , 在这里我们利用 3×3 卷积替换原 FPN 中的 1×1 卷积以更好的学习特征。然后,将 C_i 和 F_i^1 进行全局最大池化,得到二者特征通道上的全局特征,并将 C_i 全局最大池化后的特征利用 1×1 卷积将通道数降维到 256 维,两者 Reshape 相乘得到通道间依赖关系 M_i , 并作用于特征 F_i^1 , 得到捕获特征 F_i 和特征 C_i 之间的通道全局相关性新特征,再通过一个残差连接生成最终增强的中间特征 F_i , 其用公式表示如式(1)所示。另外,在 FPN 中只有 C_3, C_4, C_5 层有通道的减少,因此可以只将通道增强模块应用在这 3 层上, C_2 层特征提取保持不变。

$$F_i = \varphi(C_i) + M_i \varphi(C_i) \quad (1)$$

$$M_i = \phi(G(C_i)) \cdot G(\varphi(C_i))$$

其中, φ 代表 3×3 的卷积,用于将特征 C_i 的通道维度降到 256; M_i 表示学习的通道间依赖关系 ($M_i \in R^{256 \times 256}$); G 是全局最大池化操作; ϕ 是 1×1 卷积; F_i 表示最终增强的特征。

1.2 上下文增强(CE)

在特征金字塔网络 FPN 中, F_5 层特征是由 C_5 层特征经过 1×1 卷积进行通道降维所得到,特征通道的减少会导致高层特征信息丢失,同时在 FPN 自上而下的传播路径中 F_5 层只通过向下的路线传播特征到低层,但并未收到来自其他层特征融合,导致 F_5 层只包含单尺度的上下文信息。针对此问题提出了上下文特征增强模块(CE)来丰富 F_5 层上下文信息并对 F_5 层特征进行增强。CE 结构如图 3 所示。

在 CE 中,首先将来自骨干的特征 $C_i (i=2,3,4)$ 使用卷积网络上采样到 C_5 特征尺寸大小并将特征通道维度降维到 256 维,同时将 C_5 的特征通道数维度降维到 256 维,再将得到的所有特征进行按通道拼接,得到包含多尺度的上下文新特征,再将其送入到空洞空间卷积池化金字塔(ASPP)算法中,以不同采样率的空洞卷积并行采样,相当

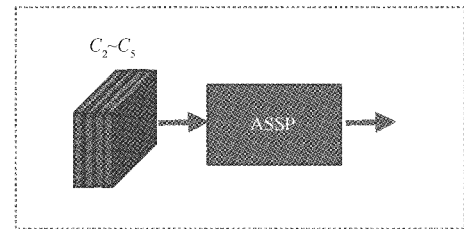


图 3 上下文增强网络结构

于以多个比例捕捉图像丰富的上下文。然后,把最终生成的带有多尺度的丰富上下文特征作用于 F_5 层,来丰富 F_5 层特征信息并增强 F_5 层特征信息。最后,通过自上而下的路径传播到其他特征层中,从而增强整个特征金字塔网络不同尺度的特征层。

1.3 注意力指导融合(AGF)

在注意力指导融合模块(AGF)中,利用通道注意力机制来学习两个向量 S_1 和 S_2 , 用于加权调整相邻层的特征,让相邻层特征的语义信息趋于一致性,避免 FPN 中忽略不同层特征之间的语义差距,直接对高层特征插值后与低层信息相加融合而造成的融合特征学习不充分问题,从而更好地指导相邻层特征融合,AGF 结构如图 4 所示。

在 AGF 中,首先将相邻层的高层特征 F_i 进行上采样到低层特征 F_{i-1} 大小记为 F'_i , 将特征 F'_i 与 F_{i-1} 进行按通道拼接,生成新的特征 $N_i (256 \times H \times W)$, 该特征包含有两层特征的语义信息用于学习相邻层特征的权重指导系数。然后,对 N_i 使用 1×1 的卷积做通道降维并用 Softmax 对特征进行增强得到新的特征图 $M_i (1 \times H \times W)$, 再将 M_i 和 N_i 进行 Reshape 并相乘,最后将得到的一维特征送到一个 SENet 网络中。在这里 SENet 最终会生成两个 C 维通道的注意力记为 S_1 和 S_2 , 分别作用特征 F'_i 和 F_{i-1} , 指导相邻特征的特征选择,以实现校正相邻的特征信息作用,然后把校正后的特征相加融合得到更好的融合特征 P_{i-1} , 其用公式表示如式(2)所示。

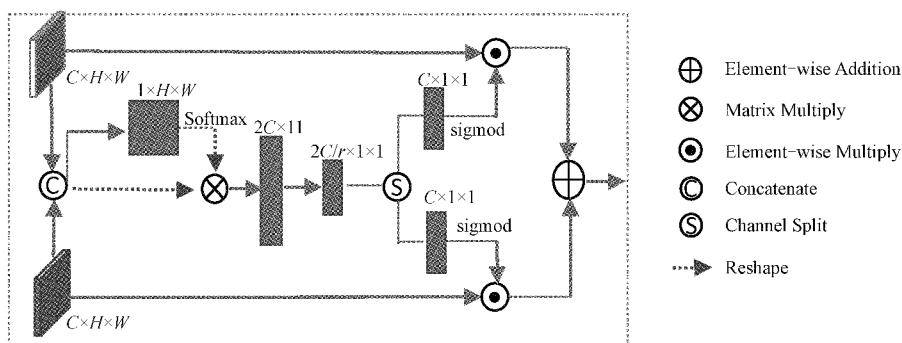


图 4 注意力指导融合网络结构

$$\begin{aligned}
 M_i &= \text{Softmax}(\varphi_1(N_i)) \\
 N_i &= [I(F_i), F_{i-1}] \\
 [S_1, S_2] &= \sigma(\phi_i(\text{Relu}(\varphi_2(N_i M_i))) \\
 P_{i-1} &= S_1 F_{i-1} + S_2 F_i
 \end{aligned} \quad (2)$$

其中, I 代表上采样操作, φ_1, φ_2 为 1×1 卷积, ϕ_i 表示有两个不同的 1×1 卷积, σ 表示 sigmoid 激活函数。

2 实验分析

2.1 实验平台及参数设置

本文实验在 Ubuntu18.04 环境下完成。配置 GPU 为 Nvidia RTX2080ti, 显存为 11 GB。Python 版本为 3.7.10, CUDA 版本为 10.1, 实验环境选择基于 PyTorch 1.6 的 MMDetection^[26] 开源目标检测工具箱, 所有预训练模型均由官方提供。另外由于 MMDetection 的官方模型结果都是基于多 GPU 实现的, 因此, 本文在一块 GPU 下重新实现了所有用到的模型, 结果显示在单卡下复现的精度均高于官方提供的精度, 所以为了公平的比较论文中精度提升对比都以在单卡下进行衡量。

2.2 数据集与评价指标

PASCAL VOC 数据集和 MS COCO 数据集都是计算机视觉领域的经典数据集。PASCAL VOC 数据集包含 20 个类别, 有 VOC-2007 和 VOC-2012 两个版本, 本文使用 VOC-2007 和 VOC-2012 的 trainval 数据用于训练模型共 16 551 张图片, VOC-2007 的 test 数据对模型进行测试共 4 952 张图片, PASCAL VOC 数据集的评价指标为均值平均精度 (mAP) 和平均精度 (AP)。MS COCO 数据集包含 80 个类别, 由 115 000 张用于训练的图像 (train-2017) 和 5 000 张用于验证的图像 (val-2017) 组成, 在测试图像中还有 20 000 张没有公开标签的图片。本文在 train-2017 上对所用模型进行训练, 在 val-2017 上报告研究结果, 并将最终结果提交给服务器在 test-dev 进行评估比较。另外, 为了综合评价算法的性能, MS COCO 数据集根据交并比 (IOU) 和目标尺寸对评价指标均值平均精度 (mAP) 进行进一步划分。根据 IOU 阈值的不同, 划分为 AP_{50} (IOU 阈值为 0.5 的 AP), AP_{75} (IOU 阈值为 0.75 的 AP)。根据目标尺寸的不同, 划分为 AP_s (检测小目标的 AP), AP_m (检

测中等目标的 AP), AP_l (检测大目标的 AP)。

2.3 实验结果与分析

为了验证本文算法的有效性, 选择了不同的数据集和不同的算法来进行实验与评估。首先, 使用 PASCAL VOC 和 MS COCO 这两个不同的数据集在 Faster R-CNN 上来评估 AEGFPN 模型在不同数据集上的效果。另外, 为了进一步验证 AEGFPN 的通用性和有效性, 本文还在 MS COCO 数据集上使用实例分割模型 Mask R-CNN 来评估 AEGFPN 模型对不同检测算法效果。

本文先基于 Faster R-CNN 在 PASCAL VOC 和 MS COCO 数据集上验证 AEGFPN 模型有效性, 在实验中统一使用 ResNet-50 作为 Faster R-CNN 的骨干网络, 且训练过程中批大小 (batchsize) 都设为 2。两个数据集的其他参数设置如下, 首先, 在 PASCAL VOC 数据集上, 迭代次数设为 4 个周期 (epoch), 学习率遵循 MMDetection 对应的缩放准则初始学习率设为 0.001 25 (8GPU 对应 $lr = 0.01$, 1GPU 则为 $lr = 0.001 25$), 在迭代 3 个周期后将学习率降低 1/10, 其他的超参数均是 MMDetection 的默认设置; 其次, 在 MS COCO 数据集上, 迭代次数设为 12 个周期, 学习率仍遵循 MMDetection 缩放准则初始学习率设为 0.002 5 (COCO 数据集上 8GPU 对应 $lr = 0.02$), 在迭代到第 8 个周期和第 11 个周期时的学习率分别降低 1/10, 其他的超参数也遵循 MMDetection 的默认设置。实验结果如表 1~3 所示。

表 1 Faster R-CNN 在 PASCAL VOC 数据集上的 mAP 值

算法	mAP/%
Faster R-CNN+FPN	80.5
Faster R-CNN+AEGFPN	82.0

表 1 和 2 为 Faster R-CNN 在 PASCAL VOC 数据集上的检测结果, 分别对应 mAP 值和各类 AP 值。从表 1 可以看到 Faster R-CNN 的 mAP 值由 80.5% 提升到 82%, 有 1.5% 的提升。表 2 是在 PASCAL VOC 数据集上各类的 AP 值, 可以看到除了 boat、bottle 和 dog 这 3 类有稍微的降低, 其他 17 个类别的 AP 值均有提升。这也从整体说

表 2 Faster R-CNN 在 PASCAL VOC 数据集上不同类别的 AP 值 %

类别	Faster R-CNN+	Faster R-CNN+
	FPN	AEGFPN
aeroplane	86.1	87.9
bicycle	86.7	88.3
bird	83.3	86.0
boat	73.6	72.5
bottle	72.8	71.4
bus	85.9	86.4
car	88.3	88.8
cat	89.4	89.4
chair	65.5	67.8
cow	85.8	88.8
diningtable	72.0	76.3
dog	88.0	87.8
horse	86.5	87.9
motorbike	85.0	86.4
person	85.8	86.1
pottedplant	53.4	54.7
sheep	82.8	86.0
sofa	77.2	78.4
train	84.3	86.4
tvmonitor	77.2	83.5

明在 PASCAL VOC 数据集上 AEGFPN 模型相对于 FPN 模型在 FasterR-CNN 上是有效的,能够提升检测器的性能。

表 3 为 Faster R-CNN 在 MS COCO 数据集上的检测结果,可以看到 AEGFPN 模型加入后整体 AP 值由 37.6% 提升到 38.6%,提升了 1%。同时,对应 AP_{50} 、 AP_{75} 、 AP_s 、 AP_m 、 AP_l 分别提升了 1.6%、1.5%、1.5%、1.3%、1.4%。尤其,从大中小物体 AP 值可以看出,在 AEGFPN 中其对大中小目标都是有效的能带来一致的提升。这也反映出,AEGFPN 相比于原始的 FPN 在 MS COCO 数据集上是有效的。同时,从对表 1~3 结果来看,AEGFPN 模型在 Faster R-CNN 上可以对不同数据集的检测精度显示出一致的提升。

表 3 Faster R-CNN 在 MS COCO 数据集上的 AP 值 %

算法	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
Faster R-CNN+FPN	37.6	58.5	40.6	21.3	41.3	48.5
Faster R-CNN+AEGFPN	38.6	60.1	41.7	22.8	42.6	49.9

图 5 和 6 为 Faster R-CNN 使用 FPN 和 AEGFPN 网络结构在 MS COCO 数据集上的一些可视化检测结果对比。从图 5 和 6 对比可知,基于注意力增强指导的特征金字塔网络不仅可以解决漏检和错检的问题,而且对目标的定位更准确。

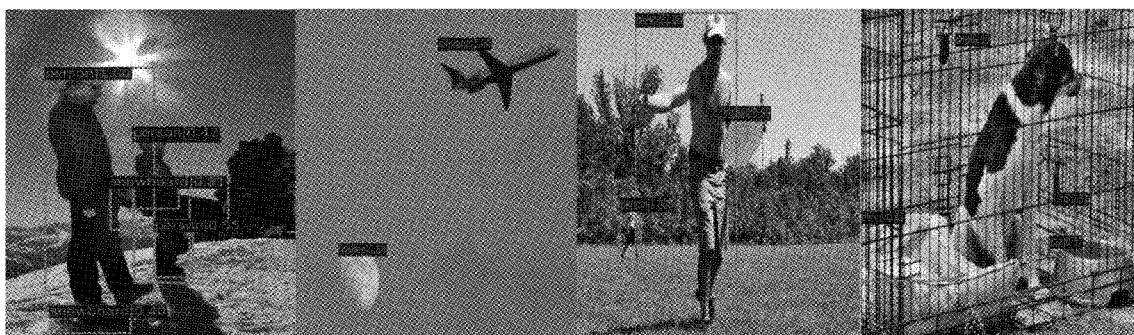


图 5 Faster R-CNN+FPN 的检测结果

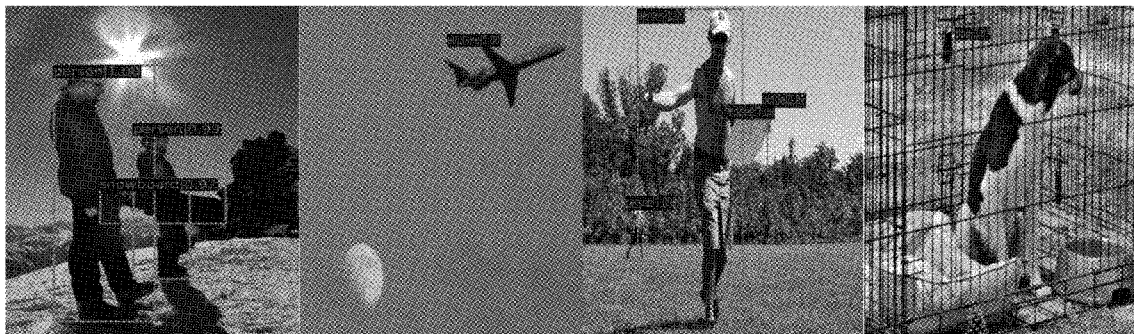


图 6 Faster R-CNN+AEGFPN 的检测结果

为了进一步验证 AEGFPN 模型的有效性,本文用 MS COCO 数据集上在实例分割模型 Mask R-CNN 进一步来评估 AEGFPN 模型效果。在实验中使用 ResNet-50 作为 Mask R-CNN 的骨干网络,训练过程中批大小(batchsize)设为 2,迭代次数设为 12 个周期,学习率遵循循

MMDetection 缩放准则初始学习率设为 0.002 5(8GPU 对应 $lr=0.01$,1GPU 则为 $lr=0.0025$),在迭代到第 8 个周期和第 11 个周期时的学习率分别降低 1/10,其他的超参数也遵循 MMDetection 的默认设置。实验结果如表 4 所示。

表 4 Mask R-CNN 在 MS COCO 数据集上的 AP 值

%

算法	AP^{seg}	AP_{30}^{seg}	AP_{75}^{seg}	AP_s^{seg}	AP_m^{seg}	AP_l^{seg}	AP^{bb}	AP_{30}^{bb}	AP_{75}^{bb}	AP_s^{bb}	AP_m^{bb}	AP_l^{bb}
MaskR-CNN+FPN	34.7	55.7	37.1	16.5	37.3	50.5	38.3	58.9	41.8	22.3	41.6	49.6
MaskR-CNN+AEGFPN	35.5	57.3	37.6	17.1	38.1	52.5	39.4	60.8	42.7	23.5	42.9	51.2

由表 4 中 Mask R-CNN 在 MS COCO 数据集上的检测结果,可以看到 AEGFPN 模型加入后无论是分割精度还是框的精度都有所提升,其中,Mask R-CNN 在分割上平均精度达到 35.5% 提升 0.8%,同时在分割上大中小目标的 AP_s 、 AP_m 、 AP_l 精度分别提升了 0.6%、0.8%、2%,从这个结果可以看到 AEGFPN 模型在大物体上的分割精度提升明显,这也说明了本文提出的 AEGFPN 模型对高层特征增强的有效性。另外,Mask R-CNN 在检测框上的平均精度也达到 39.4% 提升 1.1%,检测框在大中小目标的精度 AP_s 、

AP_m 、 AP_l 分别提升了 1.2%、1.3%、1.6%。通过对比可以发现在 Mask R-CNN 上框的提升精度高于分割精度,说明所提出的模型在目标检测上表现更好,整体上可以反映出 AEGFPN 在 Mask R-CNN 上是有效的。图 7 和 8 为 Mask R-CNN 使用 FPN 和 AEGFPN 网络结构在 MS COCO 数据集上的一些可视化检测结果对比。最后,结合表 3 和 4 的结果,可以得出所提出的 AEGFPN 在 MS COCO2017 数据集上可以对不同类型检测器的 AP 值带来一致性的提升,这也说明 AEGFPN 网络具有较强的泛化能力。

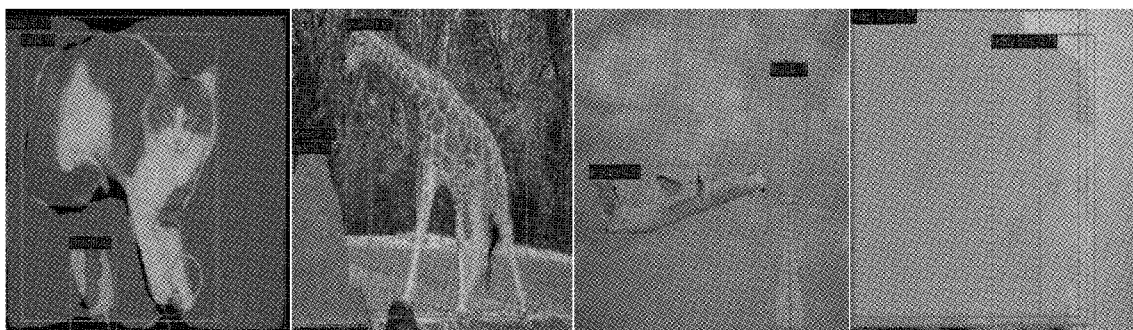


图 7 Mask R-CNN+FPN 的检测结果

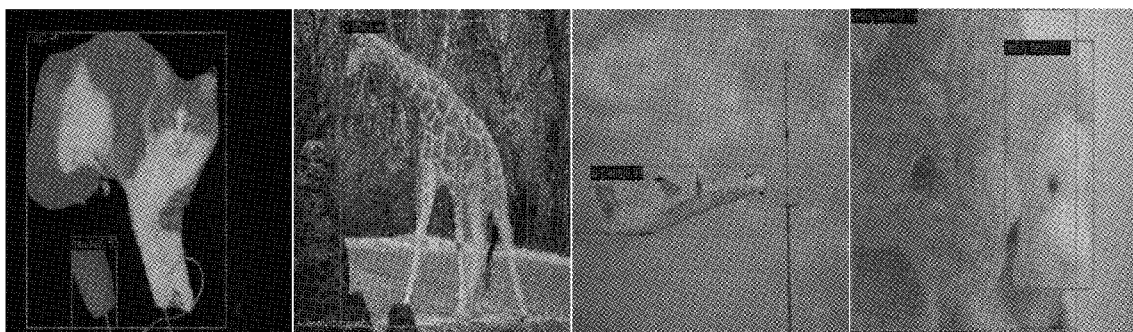


图 8 Mask R-CNN+AEGFPN 的检测结果

3 结 论

针对特征金字塔(FPN)中存在的通道信息损失和不同层之间语义信息差距的问题,本文提出一种基于注意力增强指导的特征金字塔网络,该算法通过构造 3 个有效模块分别用于增强和丰富高层特征信息以及指导相邻层特

征进行融合。本文算法在 PASCAL VOC 和 MSCOCO 数据集上使用不同类型检测器进行实验,结果表明基于 AEGFPN 网络不仅在不同数据集上表现出一致的有效性,也能够使不同类型的检测器获得更高的检测精度。但在实验过程中,将 AEGFPN 应用于实例分割网络 Mask R-CNN 上时,分割精度上出现了大物体的精度提升高于

中小物体,在后续的研究中,应考虑如何在实例分割上针对分割精度取得更好更均衡的提升,使模型具有更好的泛化能力。

参考文献

- [1] CHEN Y, WANG Z, PENG Y, et al. Cascaded pyramid network for multi-person pose estimation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018; 7103-7112.
- [2] 宋荣,周大可,杨欣.基于特征融合的尺度感知行人检测[J].电子测量技术,2020,43(5):116-123.
- [3] 高洁,王战红,刘纲.基于FSSD的微光烟雾检测方法[J].电子测量技术,2021,44(5):123-128.
- [4] 张业宝,徐晓龙.基于改进SSD的安全帽佩戴检测方法[J].电子测量技术,2020,43(19):80-84.
- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [J]. Advances in Neural Information Processing Systems, 2012, 25: 1097-1105.
- [6] REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. ArXiv Preprint,2018, ArXiv:1804.02767.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C]. European Conference on Computer Vision, Springer, Cham, 2016; 21-37.
- [8] FU C Y, LIU W, RANGA A, et al. Dssd: Deconvolutional single shot detector [J]. ArXiv Preprint,2017, ArXiv:1701.06659.
- [9] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017; 2980-2988.
- [10] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014; 580-587.
- [11] GIRSHICK R. Fast r-cnn [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015; 1440-1448.
- [12] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 28: 91-99.
- [13] HE K, GKIOXARI G, P DOLLÁR, et al. Mask R-CNN [C]. IEEE, 2017;2980-2988.
- [14] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017; 2117-2125.
- [15] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018; 8759-8768.
- [16] GHIASI G, LIN T Y, LE Q V. Nas-fpn: Learning scalable feature pyramid architecture for object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019; 7036-7045.
- [17] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020; 10781-10790.
- [18] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018; 7794-7803.
- [19] CHENG D, MENG G, CHENG G, et al. ScNet: Structured edge network for sea-land segmentation [J]. IEEE Geoscience and Remote Sensing Letters, 2016, 14(2): 247-251.
- [20] WOO S, PARK J, LEE J Y, et al. Cham: Convolutional block attention module [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018; 3-19.
- [21] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019; 3146-3154.
- [22] LI H, XIONG P, AN J, et al. Pyramid attention network for semantic segmentation [J]. ArXiv Preprint,2018, ArXiv:1805.10180.
- [23] YU C, WANG J, PENG C, et al. Learning a discriminative feature network for semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018; 1857-1866.
- [24] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.
- [25] CAO J, CHEN Q, GUO J, et al. Attention-guided context feature pyramid network for object detection [J]. ArXiv Preprint,2020, ArXiv:2005.11475.
- [26] CHEN K, WANG J, PANG J, et al. MMDetection: Open mmlab detection toolbox and benchmark [J]. ArXiv Preprint, 2019, ArXiv:1906.07155.

作者简介

史晨晨,硕士研究生,主要研究方向为机器学习、深度学习和目标检测等。

E-mail:706665038@qq.com

张长伦,博士,副教授,主要研究方向为机器学习、深度学习等。

E-mail:zclun@bucea.edu.cn