融合深度扩张网络和轻量化网络的目标检测模型

权 宇[1],李志欣[1],张灿龙[1],马慧芳[1,2]
(1. 广西师范大学广西多源信息挖掘与安全重点实验室,广西桂林 541004;2. 西北师范大学计算机科学与工程学院,甘肃兰州 730070)

摘 要: 目标检测作为计算机视觉的一个重要研究方向,近年来在算法性能上有了突破性进展. 为了更好的提升两阶段目标检测的精度与速度性能,提出了一种基于迁移学习方法的融合深度扩张卷积网络和轻量化网络的检测模型. 首先用扩张卷积网络替换主干网络中部分的卷积残差模块——深度扩张卷积网络 D_dNet-65;然后对预训练后的特征图进行压缩操作,并增加一个 81 类的全连接层以确保正常进行分类和回归操作——轻量化网络结构;最后,引入迁移学习方法并融合 D_dNet 和轻量化网络结构,通过迁移实现模型的进一步优化. 实验在典型的数据集 MSCOCO 以及 VOC07 上进行. 实验评估表明,本文提出的方法具有良好的有效性和可扩展性.

关键词: 图像目标检测; 迁移学习; 扩张卷积网络; 轻量化网络; 卷积神经网络
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2020)02-0390-08
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2020.02.023

Fusing Deep Dilated Convolutions Network and Light-Weight Network for Object Detection

QUAN Yu[1], LI Zhi-xin[1], ZHANG Can-long[1], MA Hui-fang[1,2]
(1. *Guangxi Key Lab of Multi-source Information Mining and Security, Guangxi Normal University, Guilin, Guangxi* 541004, *China*;
2. *College of Computer Science and Engineering, Northwest Normal University, Lanzhou, Gansu* 730070, *China*)

Abstract: Object detection is an important research direction in the field of computer vision. In recent years, object detection has made great advances in public datasets, and there are also breakthroughs in algorithmic performance. In order to improve the accuracy and speed performance of two-stage object detection, this paper proposes a detection model based on transfer learning method that fuses the deep dilated convolutions network and the light-weight network. First, the dilated convolutions network is used to replace the convolutional residual module in the backbone network, namely deep dilated convolution network (D_dNet-65). Then, by compressing the pretrained feature map and adding an 81-class fully connected layer to replace the original two layers, namely light-weight network. Finally, the transfer learning method is introduced in the pre-training to optimize the model (D_dNet and light-weight network). The experiment was carried out on a typical data set, MSCOCO and VOC07. And the experiment shows that the method proposed in this paper has good effectiveness and scalability.

Key words: image object detection; transfer learning; dilated convolution network; light-weight network; convolution neural network

1 引言

目标检测作为图像处理和计算机视觉领域中的经典课题. 其关注的是检测特定的物体目标,并同时获得该目标的类别以及位置信息. 此外,目标检测也在实例分割[1]和目标跟踪上有很大的延展空间.

基于传统图像处理和机器学习算法的目标检测方法和基于深度学习的目标检测方法[2]作为目标检测的

收稿日期:2019-04-22;修回日期:2019-10-17;责任编辑:马兰英
基金项目:国家自然科学基金(No. 61966004, No. 61663004, No. 61762078, No. 61866004);广西自然科学基金(No. 2019GXNSFDA245018, No. 2016GXNSFAA380146, No. 2017 GXNSFAA198365, No. 2018GXNSFDA281009);广西多源信息挖掘与安全重点实验室基金(No. 16-A-03-02, No. MIMS18-08)

两大类方法. 前者提出的特征提取方法与后者基于滑动窗口的候选区域(Region Proposal)选择策略相比在针对性、时间复杂度和窗口冗余上效果都不是很好. 而基于区域建议的目标检测算法从最初的 R-CNN[3]、Fast R-CNN[4]到后来的 Faster R-CNN[5]、R-FCN[6]逐步实现了端到端的目标识别与检测网络,使得计算机视觉在目标检测、实例分割以及目标跟踪方面从精确度和速度方面都有了很大空间的提升.

本文提出一种基于迁移学习融合深度扩张网络和轻量化网络的目标检测模型,如图 1 所示. 首先,针对生成候选区域过程的主干网络部分,延用目标检测经典残差网络[7](Residual Network,ResNet)结构,从降低深层网络参数数目的角度出发,融合瓶颈设计网络结构,与之前的参数数目相比降低了接近 17 倍.

此外,为获得较强的语义特征,提高目标检测和实例分割性能,在卷积操作后加入了改进的空间特征金字塔[8](Feature Pyramid Networks,FPN)处理,因而构建了深度扩张卷积网络(Deep Dilated Convolution Network,D_dNet-65);其次是基于候选区域再识别过程的头部网络部分,考虑到全连接卷积神经网络层[9](Fully Convolutional Layer, FC)的计算量过大且耗时. 以 MSCOCO[10]数据集为例,将特征图(Feature Maps)分类压缩到 10 类,并在区域建议网络(Region Proposal Network,RPN)之前采用扩大卷积操作生成 Feature Map. 因压缩了分类的数量,故而把分类和回归操作放到后面进行,并将进行单独分类回归操作的全连接层改成了一次性全连接操作,形成本文的 Light-weight Network;为了更好的融合这两个网络模型,在预训练阶段加入迁移学习方法,通过权重的迁移与变相增加数据集的方法进一步优化模型,提高精确度和检测速度.

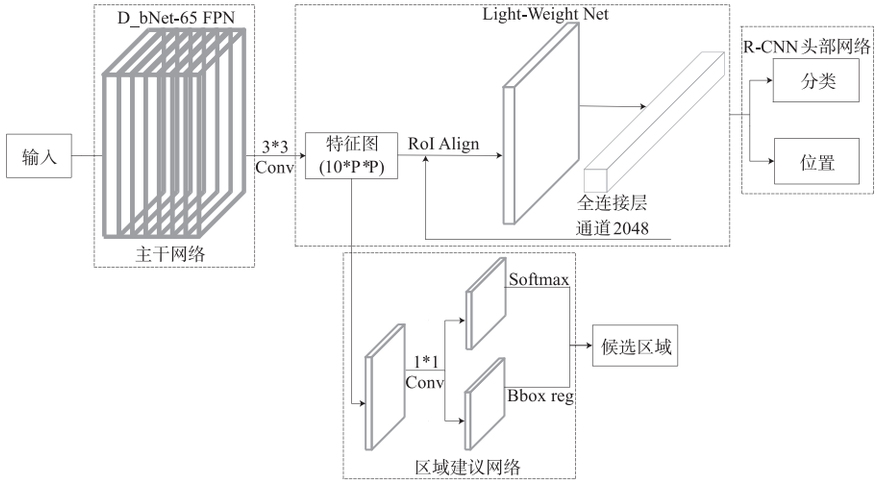

图1　基于迁移学习融合深度扩张卷积网络与轻量化网络的目标检测框架

2　相关工作

随着计算机视觉的发展,目标检测的主流方法也逐渐从传统的目标检测方法转移到基于深度学习的目标检测方法. 2013 年,Girshick 等提出 R-CNN,通过训练 AlexNet[11]对生成的 Region Proposals 进行特征提取,为目标检测发展方向提供了新的思路. 继 R-CNN 后,Girshick 等推出 Fast R-CNN,整个目标检测网络构思更精巧,流程更为紧凑,大幅提升目标检测的速度. 经过了 R-CNN 和 Fast R-CNN 的积淀,Girshick 等提出新的 Faster R-CNN. 其将特征抽取、Region Proposals 提取、边框回归以及分类整合在一个网络框架中. 2017 年,何凯明基于 Faster R-CNN 架构提出 Mask R-CNN 网络,一举完成目标实例分割(Object Instance Segmentation)和关键点检测[12],该算法的提出不仅有效的完成了目标检测,同时也实现了高质量的语义分割. 故而,本文的实现在很多方面受到该篇论文的启发.

3　融合深度扩张网络与轻量化网络目标检测模型

3.1　概述

近几年,目标检测算法在主干网络的设计上也从 VGG 网络换成 ResNet,传统的 ResNet 或者 VGG 的步长等于 32,按照特征图尺寸的减少,主干网络通常有 5 个阶段(P1-P5),RetinaNet 中存在 P6 和 P7.

由于较大物体是在比较深的特征图上进行预测,对应原图比例的感受野需求也大,但同时特征图越深,物体边缘清晰度越模糊,故对应的回归就较弱;相反,

较小物体在小分辨率特征图上便很难可见. FPN[13]及RetinaNet等网络使用分辨率较大且比较浅的层来解决这类问题,但小物体目标在“深层”网络中已经消失,所以即使把浅层与语义强的深层相加,很大部分的语义信息还是会丢失. 因此,本文在主干网络上引入了更多的阶段(如P6)加入预训练. 此外,考虑到本文是基于两阶段目标检测框架,与基于单阶段目标检测框架相比精度高,但速度慢. 故通过在主干网络中引入深度扩张卷积来提高精度,减少计算量. 此外,为了轻量化头部网络,本文在对卷积层的特征进行池化操作之后,对输出的最终的特征图进行压缩操作(以MSCOCO数据集为例从原来的$81\times p\times p$压缩到$10\times p\times p$或者$5\times p\times p$),为了保证正常完成目标检测任务,在最后增加一个81类的全连接层以确保可以正常进行分类和回归操作,详细内容见3.3节. 此外,利用迁移学习[14]可以很好解决数据尴尬的优势,在模型中引入迁移学习方法,并获取较好的效果.

3.2 深度扩张卷积网络

本文保留了ResNet-50中的stage1-4,融合DetNet[15]网络结构并额外增加stage6,将stage5-6同时加入预训练. 此外,在stage5-6中一改常规的卷积残差模块堆积的操作,加入了扩张卷积的瓶颈网络结构[16].

图2中分别给出ResNet残差模块、DetNet核心模块以及本文的D_dNet-65核心模块. 其中D_dNet-65 R-CNN核心模块是基于ResNet与DetNet的核心模块改进而来. 其结构特征依然遵循ResNet模块的结构,并在DetNet模块的每个卷积层操作后都增加了批量归一化(Batch Normalization,BN)和激活函数层(Rectified Linear Unit,ReLU). 前者加速了训练进程,减少了达到同等要求的训练次数,使训练深度网络的能力极大的加强;后者不仅使部分神经元的输出为0,也增加了网络的稀疏性,减少了参数的相互依存关系,缓解了过拟合问题的发生,并使整个过程的计算量节省很多.

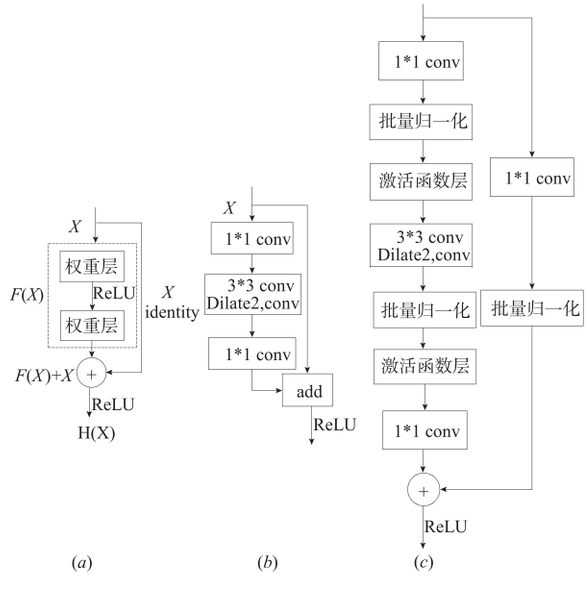

图2 ResNet残差模块、DetNet核心模块和D_dNet-65核心模块

因此,在主干网络的设计上第1阶段通过$7\times7\times64$的卷积操作、BN、ReLU和最大池化层(Max pooling),以保证输入图像经过stage1后只为原来的1/4,从而确保足够大的感受野. 考虑stage1阶段获取的Feature Map较大相应的耗时也大,故依然延用Mask R-CNN做法让stage1阶段只参与预训练阶段. 另外,stage2-4中的每一个大层都是由相同的残差模块1×1,3×3,1×1的卷积层重复组合叠加而成.

而在主干网络stage5-6层中,为保持特征图的分辨率和感受野大小,这里增加额外的stage6阶段并加入预训练模型中,如图3所示.

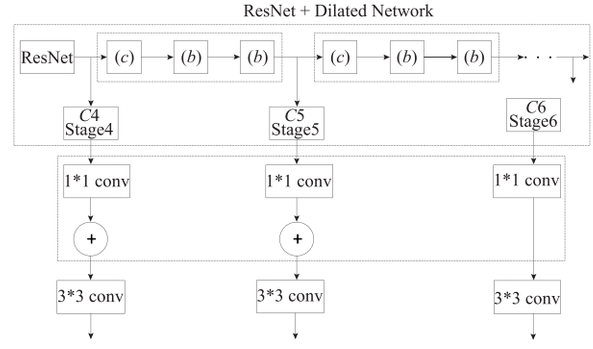

图3 D_dNet-65主干网络结构

首先,本文在stage5-6两个阶段保持特征图大小一致且都是原图尺寸的1/16,与原来的ResNet-50的stage5的特征图相比是原图尺寸的1/32,无论是在特征图的分辨率还是感受野的大小上都好很多. 其次,stage5-6两个阶段的第一层分别加入了深度扩张卷积网络,其由主路和旁路两个分支构成. 在主路上,即把1×1,3×3和1×1的3个卷积层作为基础模块;在卷积层间分别加入BN和ReLU的处理操作;由于输入先经过1×1卷积核不会改变特征图大小,接着进入3×3卷积层(padding=1),故也不会改变输入特征图的尺寸,故在主路上就可以保证特征图的尺寸不变. 在旁路上,为了保证输入的特征图可以和主路上输出的特征图相加,故在旁路上设置了$1\times1\times256$的卷积操作. 而在stage5-6两个阶段的第2层和第3层继续沿用残差模块即可. 另外通过分析发现,直接输出一个256维与先经过一个$1\times1\times64$的卷积层,再经过一个$3\times3\times64$的卷积层,最后经过一个$1\times1\times256$的卷积层,输出256维,参数量降低了1/9,因此,增加深度扩张卷积网络模块

在一定程度上对计算量和内存需求的压力有所降低.

3.3　轻量化头部网络

传统的头部网络较为“厚重”,如 Fast R-CNN 将计算量放在 ROI 操作后面;R-CNN 将计算量放在 ROI 操作前面;Faster R-CNN 则用 2 个厚重的全连接层做特征区域预测等. 这些操作都在头部网络引入了较大的计算量. 考虑到池化操作输出特征图较厚和特征图分类回归的操作是引起头部网络复杂度增加的原因,本文尝试在 Mask R-CNN 目标检测框架上对池化操作输出的特征图从原来的 $81\times p\times p$ 压缩到 $10\times p\times p$,相当于把原来的 3900 多个通道压缩到 490 个通道. 并在后面加入一个 81 类的 FC,从而顺利完成目标检测任务,如图 4 所示.

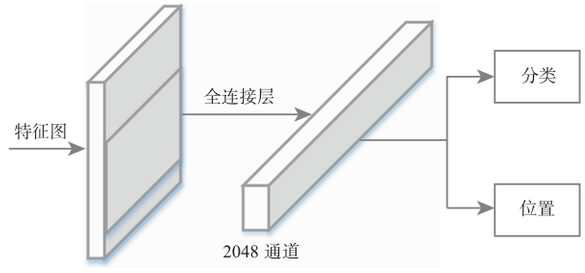

图4　轻量化网络框架图

3.4　基于迁移学习的预训练模型

迁移学习是指利用已有知识对不同但相关领域问题求解. 将迁移学习应用到深度卷积神经网络中主要体现于在特定任务上所训练得到的权重参数上,因而迁移学习的本质即为权重的迁移.

文献[17]基于迁移学习实现分类和行为识别. 文献[17]的研究证明了在卷积神经网络间迁移知识并不需要具有很强的语义相关性. 本文根据目标任务,修改输出层神经元数目,并随机初始全连接层神经元的权重,同时利用 MSCOCO 数据集预训练得到的权重参数初始化其余网络层的参数. 最后利用目标任务训练集训练整个网络,得到最终的模型. 图 5 给出本文模型中嵌入迁移学习后的具体流程步骤:

(1) 在源任务 VOC 数据集上预训练模型;

(2) 修改全连接层,更改神经元数目,并随机初始化全连接层的网络参数;同时利用步骤(1)中训练得到的权重初始化其余网络层参数;

(3) 将网络的学习率缩小 10 倍,最后利用目标任务数据集对参数初始化后的网络进行精调,得到目标模型.

本文借用 Mask R-CNN 算法模型作为源任务,目标任务为本文的 D_dNet-65 R-CNN 模型算法. 在源任务中,将 VOC07 作为训练集,对应的主干网络延用 Mask R-CNN 网络结构,总共有 5 个阶段. 其中 C1 为图像输入的第一阶段,因为 C1 对应提取的特征图较大计算耗时,所以后续处理弃用;而 S2、S3、S4、S5 则继续后续的特征提取操作;将 4 个阶段通过 FPN 以及上采样得到对应的 P2-P5,P6 为 P5 通过上采样获取的;最后对 P2-P6 进行特征融合,从而将权重输送到目标任务. 并将 MSCOCO 作为训练集,与源任务不同的是增加了一个阶段,总共有 6 个阶段. 其中 C1 依然作为图像输入的第一阶段,后续经过 S2-S6 阶段进行特征值的层层提取. 最后将提取的特征融合源任务输送过来的权重进行训练,便完成了迁移的过程.

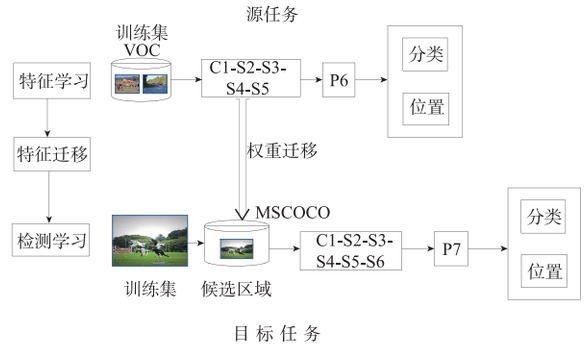

图5　迁移学习过程示意图

4　实验结果与分析

4.1　数据集和评估指标

实验评价标准选用:本文采用 ResNet-50 网络作为基准网络,并在 MSCOCO 数据集上对文中提出的方法从局部到全局进行综合性评价. 这里选用标准的 COCO 数据集指标对实验进行评估,主要包括平均准确率(Average Precision, AP)、均召回率(Average Recall, AR)、训练时间(Train Time)、测试速率(Test Rate)以及不同指标的变种标准衡量.

实验数据集说明:实验上主要依托 MSCOCO 数据集. 该数据集有 80 个对象类别,其中训练集有 80k 个图像,测试集和验证集各有 40k 个图像.

本文的验证实验会把 40k 的验证集分为 35K 和 5k 的数据集. 然后,将 80k 的训练集与 35k 的验证集结合获得 115k 的训练集,以及 5K 的小型验证集. 由于选用 GPU1080 的配置,迭代次数尽量设置的小,但不会浪费大量时间在验证上面,从而使得 Tensorboard 的更新频率更高. 文中使用 trainval 115k 进行训练,最初训练将学习率设置为 10^{-3}并进行 160k 迭代,后续训练学习率为 10^{-4}的 40k 迭代以及学习率为 10^{-5}的 40k 次迭代.

由于处理器限制,文中每个 GPU 一次只处理 1 张图像,且每张图像会设置 2000 个感兴趣区域用来做训练和 1000 个感兴趣区域用来做测试. 其次,验证次数也

多次精调,防止过大会减慢训练,过小降低验证精度.图像的输入用 1024×1024 大小,但 FPN 操作输入尺寸为 256×256 大小的特征图.其次,扩张网络模块中依然选用简单又可以避免梯度消失的 ReLU,并加入 BN.最后主干网络引入迁移学习的方法,分别迁移提前训练好的 Mask R-CNN 和 Inception 模型应用到自己的模型中,池化层的大小设置为 7,通过迁移学习进一步修正训练中的参数和 BN 的结果,提升了训练速度.

4.2 深度扩张卷积网络实验

考虑到 FPN 可以很好地在速度和准确率之间进行权衡.为了进一步验证提出的 D_dNet-65 网络在 FPN 中的有效性,将 D_dNet-65 网络与 ResNet-50 网络做实验对比.首先,文中会对 D_dNet-65 网络进行分类训练,详细结果如表 1 所示.

其中,Toperr(Top Error Rate)为错误率,描述被分类器错分比例;FLOPS(Float Operations Per Second)为每秒浮点运算次数、每秒峰值速度;AP 是指平均精确率(average precision);mAP 是多个类别 AP 的平均值;AP_{50}、AP_{75}、AP_s、AP_m、AP_l是在基于 COCO 数据集定义出来评论检测模型的指标.以一张图为例,AP_{50},AP_{75}等指的是提取检测器的 IoU 阈值大于 0.5,大于 0.75 等,对应的数值越高,精确率越低,表明提取越难;AP_s、AP_m、AP_l则是根据 COCO 数据集中物体的大小尺度进行划分,s(small)为 area < 322、m(medium)为 322 < area < 962、l(large)为 area > 962,area 是分割掩码(segmentation mask)中的像素数量.

表 1 各种主干网络在 MSCOCO 上对 FPN 的影响结果(%)

主干网络	分类		FPN 结果					
	Toperr	FLops(G)	mAP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
ResNet-50	24.1	3.8	37.9	60.0	41.2	22.9	40.6	49.2
ResNet-101	23.0	7.6	39.8	62.0	43.5	24.1	43.4	51.7
MaskR-CNN-50	**23.9**	**4.3**	**37.8**	**60.2**	**41.5**	**20.1**	**41.1**	**50.4**
MaskR-CNN-101	23.6	4.6	38.7	61.1	42.8	22.4	42.5	51.6
DetNet-59	23.5	4.8	40.2	61.7	43.7	23.9	43.2	52.0
D_dNet-65	**23.8**	**5.2**	**39.5**	**61.2**	**43.2**	**22.6**	**42.7**	**51.9**

从表中分析 D_dNet-65 的分类错误率低于 Mask R-CNN-50,从侧面也说明文中模型在准确率上有所提升.然后用 D_dNet-65 对 FPN 进行训练,并与基于 ResNet-50 的 FPN 进行比较.表 1 中 D_dNet-65 比 ResNet-50 和 Mask R-CNN 都有更好的性能(在 mAP 中最高超过 1.7 的增益).网络层数增加使得 D_dNet-65 网络中的参数高于 ResNet-50.在参数增加的同时,本文仍进一步验证 D_dNet-65 网络自身结构的有效性,见表 1.实验数据给出 ResNet-101 网络的复杂度为 7.6G,结果为 39.8 mAP.相比较 ResNet-101 发现,D_dNet-65 的错误率和复杂度都较低,并且 D_dNet-65 的准确度也更高.结果表明,D_dNet-65 比 ResNet 更合适.

鉴于本文是基于文献[18]的模型做出的改进及优化,表 2 将文中模型与目前先进模型进行实验对比.不难发现,虽本文的方法和最新提出的 DetNet 网络方法有一定差距,但是与文献[18]中方法对比提升了 1.3 – 2.2 个百分点,说明本文方法确实有效.

表 2 MSCOCO 上各主干网络对边框回归的影响结果(%)

模型	主干网络(FPN)	Bounding box AP					
		mAP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
Mask R-CNN	ResNet-50	**39.1**	**61.7**	**42.9**	**21.3**	**42.3**	**50.7**
	ResNet-101	38.2	60.3	41.7	20.1	41.1	50.2
	DetNet-59	40.7	62.5	44.1	24.6	43.9	52.2
	D_dNet-65	**40.4**	**62.1**	**43.5**	**22.7**	**42.8**	**52.0**

表 2 中发现,在 Mask R-CNN 网络模型中,若改变其主干网络结构,对边框回归的影响很大.首先,对比主干网络同为 ResNet 网络结构的实验结果,ResNet-50 在边框回归的准确率上高于 ResNet-101.这也是本文选用 ResNet-50 作为整个模型的基础网络进行改进的原因.另外,再对比将 DetNet-59 作为 Mask R-CNN 网络模型的的主干网络发现:对应的平均准确率比 ResNet-50 的主干网络高了 1.6 个百分点.分析 DetNet-59 的核心网络结构,其网络中增加了扩张卷积网络层,对提升计算速度与减少参数数量有很大的益处.最后分析用文中的 D_dNet-65 模块作为 Mask R-CNN 的主干网络,实验结果高于任何一种主干网络结构,最主要原因在于

文中的网络结构是在 ResNet-50 与 DetNet-59 的基础上提出来的,汲取了两种网络结构的优势,并且提出深度扩张卷积网络结构,减少了深度网络计算量过大、参数数量过多的劣势. 考虑到主干网络上的改进依托于 Mask R-CNN 并增加了第 6 阶段和第 7 阶段. 所以,本文为了使自身网络结构更有说服性,对表 3 中 4 个主干网络进行对比实验,即对 FPN 进行从零开始的训练. 在这里将 D_dNet-65 网络与 Mask R-CNN-50 网络进行再一次的验证实验,详情如表 3 所示.

表 3　MSCOCO 上各主干网络对 FPN 从零训练的结果(%)

主干网络	mAP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
ResNet-50	34.5	55.2	37.7	20.4	36.7	44.5
Mask R-CNN-50	**34.9**	**56.1**	**38.4**	**21.3**	**37.4**	**45.1**
DetNet-59	36.3	56.5	39.3	22.0	38.4	46.9
D_dNet-65	**35.6**	**56.3**	**38.9**	**21.7**	**38.1**	**46.4**

文中将 D_dNet-65 分别与 ResNet-50、Mask R-CNN-50 和 DetNet-59 网络进行对比实验. 分析发现,本文的方法在准确率上有很大的提升. 与 Mask R-CNN-50 的主干网络相比平均准确率提升了 1.1%. 接下来,希望进一步做出 D_dNet-65 网络关于不同尺度目标的有效性实验.

此外,分析表 1、表 2 以及表 3 的实验结果发现,DetNet-59 网络模型的实验结果始终高出本文的基于深度扩张卷积的 D_dNet-65 网络模型,考虑到本文模型的出发点在于保证准确率提升的前提下,提高目标检测速度. 本文模型在核心模块中增加了 BN 与 ReLU 层,极大程度的加快了训练速度、收敛速度以及降低了计算复杂度,在这些指标下都优于 DetNet-59 网络模型. 虽然最终的准确率与 DetNet-59 网络模型相比低了 0.3 ~ 0.7 个百分点,但也是在可接受范围内.

4.3　轻量化网络实验

考虑到利用轻量化网络压缩特征图通道会对接下来的特征图提取有所影响. 本文将原始 3969(81 × 7 × 7)的通道修改成 490(10 × 7 × 7)通道数,并在最后加入一个简单的全连接层. 将 ResNet-50 网络作为对比实验的主干网络,在 MSCOCO 小型数据集上进行一系列的对比实验来验证自身方法的有效性. 分别将轻量化网络嵌入到 R-FCN、Mask R-CNN-50 模型中,并与 4 个原始网络结构的实验结果进行对比分析,如表 4 所示.

表 4　MSCOCO 上各模型嵌入轻量化网络的训练结果(%)

模型	mAP	AP_s	AP_m	AP_l
R-FCN	33.1	18.8	36.9	48.1
Mask R-CNN-50	**37.9**	**21.1**	**40.5**	**51.2**
Light-Head R-CNN	41.5	25.2	45.3	53.1
D_dNet-65 R-CNN	**39.7**	**22.3**	**42.7**	**52.6**

嵌入轻量化网络后,Mask R-CNN-50 在 COCO mini-validation set 上的精确率降低了点,但对应的训练速度却明显提升了,并且从表中可以看出 R-FCN 的准确度相比原来提升高于 1%. 虽与目前较为先进的 Light-Head R-CNN 方法相比确实差了点,但与 Light-Head R-CNN 这种局部提升速度的方法相比,本文的方法在目标检测上还是相对稳定的,且本文的回归损失值明显比分类损失小了一些.

4.4　基于迁移学习方法训练的实验

首先,文中对 ResNet-50、ResNet-101 以及 Mask R-CNN 系列分别加入迁移学习方法进行模型训练,并用验证集对训练好的模型进行验证. 验证集是从 MSCOCO 的验证集分离出来的 5k 张小型验证集. 针对以上几个模型分类以及目标检测的测试结果如表 5 所示.

表 5　MSCOCO 上各模型加入迁移学习后的训练结果(%)

模型	网络训练精度	目标检测性能测试
ResNet-50	94.9	–
ResNet-101	94.7	–
Mask R-CNN-50	**95.1**	**39.3**
Mask R-CNN-101	95.3	39.1
D_dNet-65 R-CNN	**96.3**	**39.7**

从表 5 看出基于迁移学习得到的模型,在网络训练精度上和目标检测性能测试上效果不错,加入迁移学习的 Mask R-CNN-50 网络与文献[18]中相比提升了 1 个百分点,而 D_dNet-65 R-CNN 与改进的 Mask R-CNN-50 网络相比在性能测试上高出 0.4 个百分点. 因此,迁移学习方法除了在解决训练样本不足的问题. 同时使得目标任务具有更高的分类和目标检测准确率,能够更好的提升模型的性能.

4.5　模型的训练与测试时间结果分析

表 6 中给出了一阶段目标检测模型方法(SSD、YOLO)以及两阶段目标检测的模型方法(R-CNN、Fast R-CNN、Faster R-CNN、Mask R-CNN)与本文方法的对比实验. 实验中使用几个预先训练过的 ImageNet 基础网络模型以及文中自己的网络模型. 其中,VGG16 网络模型称为 V_{16};$ResNet_{50}$ 网络模型称为 R_{50};文中自己的网络模型称为 D_{65}. 其次,表中也给出了各种方法训练、测试的训练时间(Train Time)、训练速度(Train Speedup)、测试速率(Test Rate)、测试速度(Test Speedup)以及在 VOC07、COCO17 数据集上的平均准确率与我们自己的模型 D_dNet-65 R-CNN 进行实验对比的结果.

训练与测试速度的提升均以 R-CNN 为基数,表 6 中发现在几种两阶段目标检测方法中,当将其基础网络 V_{16} 更换成 R_{50} 后各方面性能都有所提升,Faster R-CNN 的测试速率是 R-CNN 的 600 多倍;而文中的 D_

dNet-65 R-CNN 方法与 Mask R-CNN 方法在同为 D_{65} 的网络模型情况下,测试速度提升 1.5 倍. 表 6 直观的说明了在两阶段目标检测中,文中的方法在检测速度的提升上取得了较好的结果. 与此同时,表 6 中也给出了关于一阶段目标检测的实验对比结果. 在同为 R50 的基础网络模型下,虽然文中的方法在测试速率上与 YOLO 方法仅仅相差了 0.003 个百分点,对于两阶段目标检测来说是一个很好的结果. 实验对比发现,在主干网络上构建一个轻量化网络模型可以大大的改善两阶段目标检测检测速度较慢的劣势.

表 6 多种模型基于 GPU 的训练与测试速率分析结果

测试项	R-CNN		Fast R-CNN		Faster R-CNN		Mask R-CNN		D_dNet-65 R-CNN		SSD		$YOLO_{V3}$	
	V_{16}	R_{50}	V_{16}	R_{50}	V_{16}	R_{50}	R_{50}	D_{65}	R_{50}	D_{65}	V_{16}	R_{50}	V_{16}	R_{50}
Train Time(h)	84	75	9.5	8.0	8.7	7.7	44	15	**11**	**10**	10	8.4	10.4	6.4
Train Speedup(×)	1	1.12	8.8	10.5	9.6	10.9	1.9	5.6	7.6	8.4	8.4	10	8	13
Test Rate(s/im)	47	5	0.32	0.25	0.14	0.11	0.2	0.07	**0.05**	**0.045**	0.045	0.038	**0.047**	0.029
Test Speedup(×)	1	9.4	146	188	335	427	522	671	**940**	**1044**	1044	1236	1000	1620
VOC07/COCO17	45.7	49.6	65.1	67.5	70.4	77.6	70.5	77.3	**79.6**	**80.7**	76.3	78.1	57.9	73.8

4.6 在 MSCOCO 数据集上的实验结果分析

本文通过四步进行有效性的实验验证. 第一步,针对提出的深度扩张卷积网络,通过对比实验分析 D_dNet-65 网络结构对 FPN 的影响、对边框回归的影响以及分析从头训练特征金字塔的效果. 第二步,将轻量化网络嵌入多个模型中进行对比实验. 第三步,在训练模型中加入迁移学习方法,并对多个模型进行迁移学习的嵌入实验分析. 第四步,如图 6 将本文方法与 Mask R-CNN 系列以及其他网络在 MSCOCO 数据集上进行实验对比. 从图中可知:对比正三角标记折线与倒三角标记折线,前者(本文的方法)明显优于本文依托的网络模型,进一步证明本文网络模型的有效性;其次将正三角标记折线与星花标记折线进行对比,虽然本文的结果与其相比略差一点,但基于正三角标记折线是一个基于全局网络模型的预测结果,所以略有差距是可以接受的. 总体上,本文的方法无论在性能还是有效性上都取得了较好的效果.

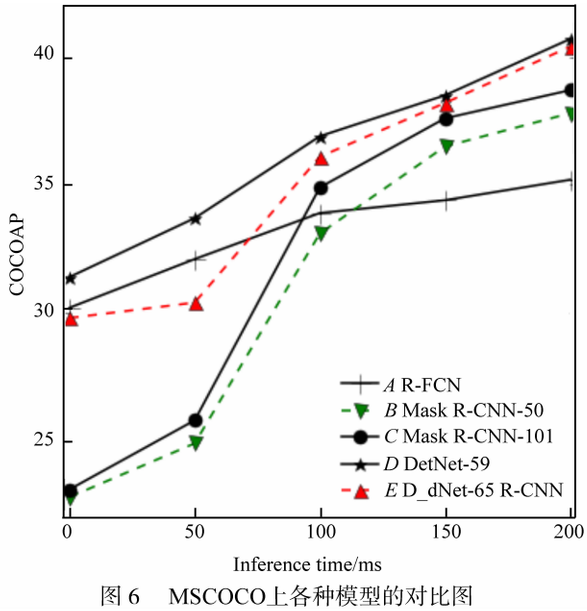

图 6 MSCOCO上各种模型的对比图

5 结束语

本文的 D_dNet-65 R-CNN 模型基于 MSCOCO 数据集从主干网络、头部网络和新增的迁移学习方法三个部分进行实验验证. 在头部网络通过压缩特征图来降低头部网络的厚重感,从而在保证精确度前提下提升网络训练的速度;最后,为了进一步优化本文的模型,提出了利用迁移学习的方法进行模型的训练,使得本文提出的模型进一步取得较好的效果. 文中通过多个实验对比分析验证,无论是在做单一训练还是综合性实验,本文的模型在精确度上都取得很有效的结果.

从研究方向考虑,将目标检测框架迁移到实例分割以及关键点检测的方向上. 除此之外,对于迁移学习的方法可以进一步深入研究,结合无监督学习与迁移学习方法,进一步提高目标检测的性能.

参考文献

[1] 许新征,丁世飞,史忠植,等. 图像分割的新理论和新方法[J]. 电子学报,2010,38(2A):76-82.
Xu X Z,Ding S F,Shi Z Z,et al. New theories and methods of image segmentation[J]. Acta Electronica Sinica,2010,38(2A):76-82. (in Chinese)

[2] Erhan D,et al. Scalable object detection using deep neural networks[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. Los Alamitos:IEEE Computer Society Press,2014. 2147-2154.

[3] Girshick R,Donahue J,Darrell T,et al. Rich feature hierarchies for accurate object detection and semantic segmentation[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. Los Alamitos:IEEE Computer Society Press,2014. 580-587.

[4] Girshick R. Fast R-CNN[A]. Proceedings of the IEEE International Conference on Computer Vision[C]. Los Alamitos:IEEE Computer Society Press,2015. 1440 – 1448.
[5] Ren S,He K,Girshick R,et al. Faster R-CNN:Towards real-time object detection with region proposal networks [A]. Advances in Neural Information Processing Systems [C]. Cambridge:MIT Press,2015. 91 – 99.
[6] Dai J,Li Y,He K,et al. R-FCN:Object detection via region-based fully convolutional networks[A]. Advances in Neural Information Processing Systems[C]. Cambridge: MIT Press,2016. 379 – 387.
[7] He K,Zhang X,Ren S,et al. Deep residual learning for image recognition[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. Los Alamitos:IEEE Computer Society Press,2016. 770 – 778.
[8] Lin T Y,et al. Feature pyramid networks for object detection[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. Los Alamitos: IEEE Computer Society Press,2017. 2117 – 2125.
[9] Long J,Shelhamer E,Darrell T. Fully convolutional networks for semantic segmentation[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. Los Alamitos: IEEE Computer Society Press, 2015. 3431 – 3440.
[10] Vinyals O,Toshev A,Bengio S,et al. Show and tell:Lessons learned from the 2015 MSCOCO image captioning challenge[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(4):652 – 663.
[11] Krizhevsky A,Sutskever I,Hinton G E. Imagenet classification with deep convolutional neural networks[A]. Advances in Neural Information Processing Systems[C]. Cambridge:MIT Press,2012. 1097 – 1105.
[12] Deng Z,Li K,Zhao Q,et al. Effective face landmark localization via single deep network[OL]. https://arxiv.org/ abs/1702. 02719,2017-02-09.
[13] Ghiasi G,Lin T Y,Le Q V. Nas-fpn:Learning scalable feature pyramid architecture for object detection[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. Los Alamitos: IEEE Computer Society Press,2019. 7036 – 7045.
[14] Pan S J,Yang Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering,2010, 22(10):1345 – 1359.
[15] Li Z,Peng C,Yu G,et al. DetNet:A backbone network for object detection[DB/OL]. https://arxiv. org/abs/1804. 06215,2018-04-17.
[16] Shrivastava A,Sukthankar R,Malik J,et al. Beyond skip connections: Top-down modulation for object detection [DB/OL]. https://arxiv. org/abs/1612. 06851, 2018-12-20.
[17] Oquab M,et al. Learning and transferring mid-level image representations using convolutional neural networks[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. Los Alamitos: IEEE Computer Society Press,2014. 1717 – 1724.
[18] He K,Gkioxari G,Dollár P,et al. Mask R-CNN[A]. Proceedings of the IEEE International Conference on Computer Vision[C]. Los Alamitos:IEEE Computer Society Press,2017. 2980 – 2988.

作者简介

权　宇　女,1992 年 9 月出生,江苏徐州人. 广西师范大学计算机科学与信息工程学院硕士研究生. 研究方向为图像理解与机器学习.
E-mail:quanyu0919@ 163. com

李志欣(通信作者)　男,1971 年 10 月出生,广西桂林人. 现为广西师范大学计算机科学与信息工程学院教授、博士生导师. 研究领域为图像理解,机器学习与跨媒体计算.
E-mail:lizx@ gxnu. edu. cn

张灿龙　男,1975 年 10 月出生,湖南娄底人. 现为广西师范大学计算机科学与信息工程学院教授、博士生导师. 研究领域为目标跟踪与模式识别.
E-mail:zcltyp@ 163. com

马慧芳　女,1981 年 7 月出生,甘肃兰州人. 博士,硕士生导师,现为西北师范大学计算机科学与工程学院教授. 研究领域为数据挖掘与机器学习.
E-mail:mahuifang@ yeah. net