

基于并行附加特征提取网络的 SSD 地面小目标检测模型

李宝奇, 贺昱曜, 强 伟, 何灵蛟

(西北工业大学航海学院, 陕西西安 710072)

摘 要: 针对 SSD 原始附加特征提取网络 (Original Additional Feature Extraction Network, OAFEN) 中 stride 操作造成图像小目标信息丢失和串联结构产生的多尺度特征之间冗余度较大的问题, 提出了一种计算量小、感受野大的深度可分离空洞卷积 (Depthwise Separable Dilated Convolution, DSDC), 并利用 DSDC 设计了一个包含三个独立子网络的并行附加特征提取网络 (Parallel Additional Feature Extraction Network, PAFEN). PAFEN 上路用两个 DSDC 提取尺寸为 $19 * 19$ 和 $3 * 3$ 的特征图; 中路用一个 DSDC 提取尺寸为 $10 * 10$ 的特征图; 下路用两个 DSDC 提取尺寸为 $5 * 5$ 和 $1 * 1$ 的特征图. 实验结果表明, 在 SSD 框架内, PAFEN 在 mAP 和检测时间等方面均优于 OAFEN, 适用于地面小目标的检测任务.

关键词: 目标检测; SSD; 深度可分离卷积; 空洞卷积; 深度可分离空洞卷积; 并行附加特征提取网络

中图分类号: TP183 **文献标识码:** A **文章编号:** 0372-2112 (2020)01-0084-08

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2020.01.010

SSD with Parallel Additional Feature Extraction Network for Ground Small Target Detection

LI Bao-qi, HE Yu-yao, QIANG Wei, HE Ling-jiao

(School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China)

Abstract: Aiming at the problems of small target information loss caused by stride operation and large redundancy among multi-scale feature maps generated by serial structure in original additional feature extraction network (OAFEN) of SSD, a depthwise separable dilated convolution (DSDC) with small computation and large field of receptivity is proposed; then a parallel additional feature extraction network (PAFEN) with three independent subnetworks is designed by using five DSDCs. In upper subnetwork of PAFEN, two DSDCs are used to extract $19 * 19$ and $3 * 3$ feature maps. In intermediate subnetwork of PAFEN, one DSDC is used to extract $10 * 10$ feature maps. In lower subnetwork of PAFEN, two DSDCs are used to extract $5 * 5$ and $1 * 1$ feature maps. The experimental results show that within the framework of SSD, PAFEN is superior to OAFEN in terms of mAP and detection time, and is suitable for ground small target detection tasks.

Key words: target detection; SSD; depthwise separable convolution; dilated convolution; depthwise separable dilated convolution; parallel additional feature extraction network

1 引言

地面目标检测在无人机导航、搜索、精确打击和毁伤后评估中发挥着重要作用^[1-3]. 考虑飞行安全和飞行距离, 无人机通常会在较远距离获取地面目标图像, 由此造成的结果是图像内目标像素比较小, 这会进一步增加目标检测的难度^[4].

通过将深度学习^[5-7]模型 CNN (Convolutional Neural Networks)^[8-10]嵌入到目标检测模型之中, 目标检测精度在过去几年中不断提高, 结合 CNN 的目标检测算法可分为基于候选区域和基于回归两类. Girshick 等^[11]第一个将 CNN 用于目标检测, 并提出 R-CNN (Region-based Convolutional Neural Networks) 模型. R-CNN 使用区域建议方法从输入图像中生成 2000 个候选区域, 并

将所有的候选区域缩放到固定尺寸. 然后,使用 CNN 在这些候选区域上提取特征. 在 CNN 的最后一个特征层加入两个全连接层(SVM 和回归层). 由于 SVM 和回归层是分开训练,R-CNN 很难优化而且占用内存空间非常大. Girshick 等^[12]又提出 Fast R-CNN 模型. Fast R-CNN 首先在图像中提取感兴趣区域 (Regions of Interest, RoI); 然后对每幅图像只做一次卷积处理,在最后一个卷积层输出的特征图上对每个 RoI 进行映射,并送入 RoI 池化层把各尺寸的特征图统一到相同的大小^[13];最后利用 Softmax Loss 和 Smooth L1 Loss 对分类概率和边框回归联合训练. 联合训练省去了特征存储,提高了空间和时间利用率. 然而,提取感兴趣区域占用了整个检测过程的大部分时间. Ren 等^[14]提出了目标检测模型 Faster R-CNN. 与 Fast R-CNN 相比,Faster R-CNN 利用 RPN (Region Proposal Network) 在 CNN 最后一个连接层中自行产生建议框,因此建议框生成网络和目标检测网络共享卷积网络. 同时建议框数目从原有的约 2000 个减少为 300 个,重要的是建议框的质量也有本质的提高,但 Faster R-CNN 检测的速度依然有待提高.

基于区域建议的目标检测方法不能利用局部目标在整幅图像中的空间信息,所以一些研究者开展了无区域建议的目标检测研究,主要采用回归的思想. Redmon 等^[15]提出了一种无区域建议的目标检测模型,称为 YOLO (You Only Look Once). YOLO 通过采用空间限制,减少了对同一目标的重复检测,大大提高了效率,能够达到实时的效果. 但是 YOLO 的检测精度不如 Faster R-CNN. 针对 YOLO 存在的不足, Liu 等^[16]提出 SSD (Single Shot Detector) 模型. SSD 模型主要由四部分组成:基础网络 VGG-16 (Visual Geometry Group),附加特征提取层部分,default boxes 生成部分和卷积预测部分. SSD 通过融合六个尺度的特征来提高目标检测的精度. 原始附加特征提取网络从基础网络特征层 conv4_3 (38 * 38) 开始,然后通过标准卷积层中使用 stride 操作依次生成 fc7 (19 * 19)、conv6_2 (10 * 10)、conv7_2 (5 * 5)、conv8_2 (3 * 3) 和 conv9_2 (1 * 1) 五个尺度的特征图. stride 操作会造成图像目标信息的丢失^[17-19],尤其是图像中的小目标. 原始附加特征提取网络中六个尺度的特征图采用串联结构生成,彼此之间具有很高的冗余度,特征图之间冗余度较高不利于图像小目标的精准定位. 为了提高 SSD 对小目标的检测精度, Cao 等^[20]提出 FFSSD (Feature-Fused SSD) 模型. FFSSD 利用 Feature Fusion Module 对 VGG-16 中 conv5_3 进行 2 倍上采样操作,并与 conv4_3 融合来提高小尺寸目标的检测精度. Fu 等^[21]提出 DSSD (Deconvolutional Single Shot Detector) 模型. DSSD 基础网络为特征提取能力更

强的 Resnet-101,并利用 Deconvolution Module 扩展低维度信息的上下文信息来提高小尺度目标的检测精度. FFSSD 和 DSSD 都是通过整合更多尺度上的语义特征来提高模型对小目标的检测精度,但是这些方法,在提高对于目标物体特别时小物体识别效果的同时由于加入了额外的层,也增大了计算量,影响了检测的实时性. Zhou 等^[22]提出了 STDN (Scale-Transferable Object Detection) 模型. STDN 基础网络为 DenseNet-169,并通过 Scale-Transfer 层实现了在几乎不增加参数数量和计算量的前提下生成大尺寸的特征层来提高模型对小目标的检测精度. 由于 Scale-Transfer 层是一个转换操作,因此基本不会引入额外的参数数量和计算量. 为了提高 SSD 的检测速度,Howard 等^[23]提出了轻量化的卷积神经网络 MobileNet. MobileNet 用深度可分离卷积 (Depthwise Separable Convolution, DSC) 替换标准卷积来减少模型的参数和计算量,它在不影响目标检测精度的条件下能极大地提高 SSD 的检测速度.

除了上述针对 SSD 的改进方法外,还有一些其它提高小目标检测精度的研究工作,例如设计专用的目标检测的骨干网络^[24],优化检测模型训练过程^[25],IoU 阈值动态设计^[26]和生成高分辨率小目标特征^[27]等. 特别是文献[27], Li 等提出了一种基于 PGAN (Perceptual Generative Adversarial Networks) 的小目标检测方法. PGAN 通过训练条件生成网络使小目标的特征表示与大目标特征表示类似,并利用一个新的感知分类器来监督小目标特征的生成从而更精确地检测小目标.

针对上述改进方法不能兼顾小目标检测精度和速度的问题,本文提出了一种计算量小,覆盖范围大的卷积单元-深度可分离空洞卷积 (Depthwise Separable Dilated Convolution, DSDC),并利用 DSDC 设计了一种包含三个子网络的并行附加特征提取层网络 (Parallel Additional Feature Extraction Network, PAFEN),同时通过优化 DSDC 空洞率 (Dilation Rate) 的选取来改善 PAFEN 的性能,最后在 SSD 框架内通过结合 MobileNet (基础网络) 和 PAFEN (特征提取网络) 实现对地面小目标的快速、准确的检测.

2 基于并行附加特征提取网络的 SSD 小目标检测模型

本文通过改进附加特征提取网络来提高 SSD 对地面小目标的检测精度和速度. 在新的附加特征提取网络中,深度可分离空洞卷积作为网络基本单元(深度可分离空洞卷积摒弃了 stride 操作,它通过通道分解和卷积核空洞化来减少计算量和增大感受野);三个独立子网络结构替代原来的单网络串联结构以增加多尺度特征图的多样性,并为与 conv4 (在 MobileNet 中与 VGG-16

中 conv4_3 对应的层为 conv4) 直接相连的三个多尺度特征 fc7、conv6_2 和 conv7_2 设计 convex 空洞率策略来增加上下两个子网络的信息以改善新附加特征提取网络的性能。

2.1 深度可分离空洞卷积单元

深度可分离卷积通过将标准卷积分解成深度卷积和点卷积来大幅降低模型的参数,重要的是卷积层的特征提取能力基本不受影响. 但为了生成尺寸更小的特征图,深度可分离卷积只能通过增加卷积核尺寸或 stride 操作实现,过大尺寸的卷积核会大幅增加模型参数,而 stride 操作会造成图像小目标信息的丢失. 深度可分离空洞卷积是在深度可分离卷积的基础上,通过对深度卷积空洞化实现. 空洞化是向标准卷积中引入

了一个称作空洞率的新参数^[17],并利用扩张率控制卷积核处理数据时各值的间距,从而在计算量相当条件下实现卷积层感受野的增大. 深度可分离空洞卷积与深度可分离卷积相比具有更大的感受野,与空洞卷积相比具有更小的计算量. 深度可分离空洞卷积(DS-DC)、深度可分离卷积(DSC)和标准卷积之间的关系如图 1 所示.

对于 M 个尺寸为 $D_F * D_F$ 的输入特征图 F ,经尺寸为 $D_K * D_K$ 的卷积核操作后,输出 N 个尺寸为 $D_C * D_C$ 的特征图 G ,其中 D_F 是输入的特征图的宽度和高度, M 是输入通道数, D_C 是输出特征图的宽度和高度, N 是输出通道数. 标准卷积、深度可分离卷积和深度可分离空洞卷积的计算分解过程如图 2 所示.

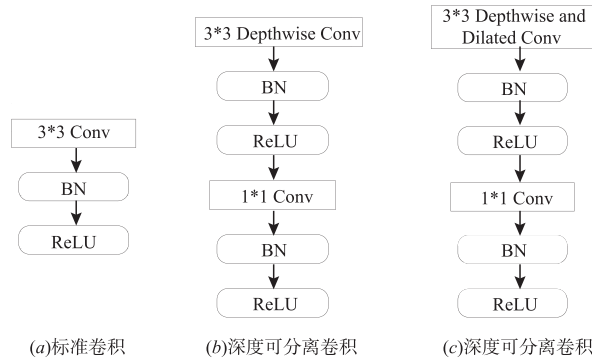


图1 标准卷积、深度可分离卷积和深度可分离空洞卷积

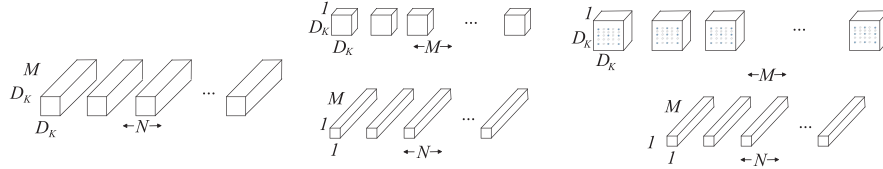


图2 标准卷积、深度可分离卷积和深度可分离空洞卷积分解

标准卷积生成特征图 G 的计算成本为:

$$D_K * D_K * M * N * D_C * D_C \quad (1)$$

深度可分离卷积生成特征图 G 的计算成本为:

$$D_K * D_K * M * D_C * D_C + M * N * D_C * D_C \quad (2)$$

深度可分离空洞卷积生成特征图 G 的计算成本为:

$$D_K * D_K * M * D_C * D_C + M * N * D_C * D_C \quad (3)$$

深度可分离卷积和深度可分离空洞卷积与标准卷积的计算成本比值均为:

$$\frac{D_K * D_K * M * D_C * D_C + M * N * D_C * D_C}{D_K * D_K * M * N * D_C * D_C} = N + D_K^2 \quad (4)$$

但在相同计算量的条件下,深度可分离空洞卷积的覆盖范围更大. 深度可分离空洞卷积实际空洞滤波

器(RDF)尺寸与空洞率之间的关系如下:

$$K_{\text{rdf}} = D_K + (D_K - 1)(R - 1) \quad (5)$$

其中 K_{rdf} 为该层 RDF 尺寸, D_K 为该层卷积核尺寸, R 为该层空洞率大小. 例如,一个卷积核尺寸为 $3 * 3$,空洞率 $R=2$ 的空洞卷积层, RDF 的实际覆盖范围为 $5 * 5$,即 $K_{\text{rdf}}=5$. 同时可以通过进一步增大空洞率 R 来扩大卷积层的感受野.

2.2 基于并行附加特征提取网络的 SSD 小目标检测模型设计

附加特征提取网络并行化设计是由 conv4 直接生成剩余的多尺度特征图,每个子网络之间保持独立,生成的多尺度特征图之间的冗余度也相对较小^[23]. 考虑 conv7_2($5 * 5$)、conv8_2($3 * 3$)和 conv9_2($1 * 1$)三个尺度的特征图尺寸比较接近,因此将剩余的五个尺度

特征图设计为三个独立的子网络结构,即 conv4 与 fc7、conv6_2 和 conv7_2 直接相连接,同时将 conv8_2 和 conv9_2 分配到上路和下路两个子网络有助于提升 fc7

和 conv7_2 特征图的质量,也可以避免与 conv7_2 特征尺度上的重复. 基于 MobileNet 和 PAFEN 的 SSD 地面小目标检测模型结构如图 3 所示,记作 MPSSD.

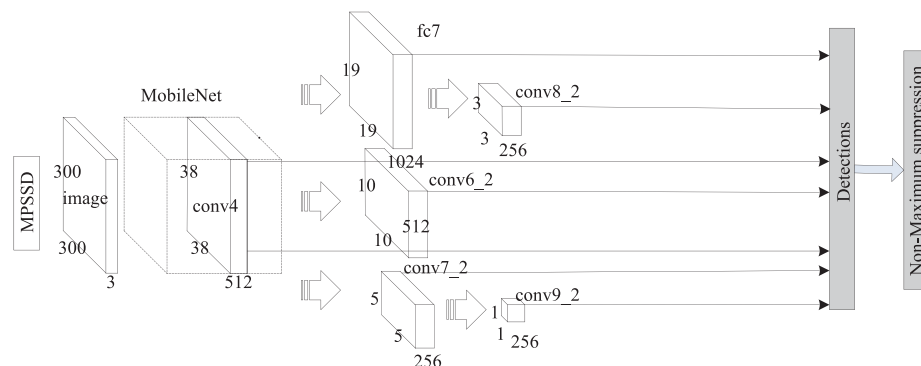


图3 MPSSD目标检测模型结构

地面小目标图像首先进入 SSD 基础网络部分,其中基础网络为 MobileNet 被截断至 conv4 的部分网络,并将 conv4 作为 PAFEN 的第一个特征层;PAFEN 由 conv4 生成三个独立的子网络:上路子网络、中路子网络和下路子网络. 上路子网络由 conv4 (38 * 38) 生成 fc7 (19 * 19) 和 conv8_2 (3 * 3) 两个尺度的特征图;中路子网络由 conv4 (38 * 38) 生成 conv6_2 (10 * 10) 一个尺度的特征图;下路子网络由 conv4 (38 * 38) 生成 conv7_2 (5 * 5) 和 conv9_2 (1 * 1) 两个尺度的特征图. default boxes 生成部分根据预先定义的 scales 和 aspect ratios 从上述六个尺度的特征层中提取数量和大小不同的 default boxes;卷积预测部分则是对 default boxes 内目标的类型和位置进行判断.

利用深度可分离空洞卷积构建 PAFEN 需要首先解决网格问题^[17-19],网格问题就是空洞卷积网络每一层实际感受野叠加操作后无法实现对输入图像的完全覆盖,因此会造成信息的丢失. 文献[17]和文献[19]针对并行和串行空洞卷积网络分别提出了卷积核尺寸固定下的空洞率选取准则. 考虑 PAFEN 并非严格意义上的串联或并联结构,同时 PAFEN 主要是为了提取图像小目标特征,因此文献[17]和文献[19]中的空洞率选取策略并不适用 PAFEN 中深度可分离空洞卷积空洞率的设计. 为了更好地提取图像小目标特征,PAFEN 中使用的空洞率数值整体相对较小. 同时,由于 PAFEN 上路和下路子网络均包含两个 DSDC,因此 fc7 和 conv7_2 使用更小的空洞率有利于提高两个子网络特征图的质量. PAFEN 中五个深度可分离空洞卷积层的参数设置如表 1 所示.

为了严格保证 PAFEN 与 OAFEN 中多尺度特征图尺寸完全一致,需要对部分 DSDC 进行填充,例如 fc7、conv6_2 和 conv9_2.

表 1 并行附加特征提取网络参数设置

Name	fc7	conv6_2	conv7_2	conv8_2	conv9_2
Kernel Size	8 * 8	7 * 7	12 * 12	5 * 5	3 * 3
Pad	1	1	0	0	1
Dilation	3	5	3	4	3

3 仿真试验

为了验证 PAFEN 的有效性以及特征层组合方式和 DSDC 空洞率选取对 PAFEN 性能的影响,实验以 mAP、平均检测时间和参数大小作为模型定量评价指标. 设计实验 1,以 MOSSD^[23] (基础网络为 MobileNet,特征提取网络为 OAFEN)为参考,比较分析 Faster-RCNN^[14]、SSD^[16]、FFSSD^[20]、DSSD^[21] 与本文地面小目标检测方法 MPSSD (基础网络为 MobileNet,特征提取网络为 PAFEN)之间的性能差异. 设计实验 2,以附加特征提取网络中不同尺度的特征图为研究对象,比较分析特征图不同的组合方式对 PAFEN 性能的影响. 设计实验 3,以深度可分离空洞卷积的空洞率为研究对象,比较分析不同的空洞率选取方式对 PAFEN 性能的影响. SSD、FFSSD、DSSD、MOSSD 和 MPSSD 由 Caffe 工具箱设计,采用 GPU (Titan X) 计算方式^[28],其中检测时间为模型检测 100 幅图像的平均前向时间,参数大小指模型(.caffemodel 文件)的实际大小. Faster-RCNN 实验模型由 TensorFlow 工具箱设计,同样采用 GPU (Titan X) 计算方式,平均检测时间为模型检测完整测试数据集的平均时间,参数大小指模型(.ckpt 文件)的实际大小.

3.1 实验数据集

为了更好地检验基于 PAFEN 的 SSD 模型 (MPSSD)对图像小目标的检测性能,我们建立了一个地面小目标数据集:SGT-DET. SGT-DET 包括四种地面

目标:军用卡车、直升机、导弹和坦克,共计 3350 幅图像,其中 2410 幅用于模型训练,940 幅图像用于模型测试,如表 2 所示。

表 2 地面小目标数据集组成

	训练(幅)	测试(幅)
军用卡车	600	235
直升机	600	235
导弹	600	235
坦克	610	235
总计	2410	940

地面小目标是一个相对的概念,其中“小”指的是图像中目标的像素比值小,SGT-DET 图像中四个地面小目标的像素比均小于 0.05。

3.2 实验 1:目标检测算法的性能比较

本实验比较分析 Faster-RCNN、SSD、FFSSD、DSSD、MOSSD 与本文地面小目标检测方法 MPSSD 在数据集 SGT-DET 上的性能差异。Faster-RCNN 的基础网络为 VGG-16,特征提取网络为 RPN;SSD 的基础网络为 VGG-16,特征提取网络为 OAFEN;FFSSD 的基础网络为 VGG-16,特征提取网络为 Feature Fusion Module;DSSD 的基础网络为 ResNet101,特征提取网络为 Deconvolution Module;MOSSD 的基础网络为 MobileNet,特征提取网络为 OAFEN;MPSSD 的基础网络为 MobileNet,特征提取网络为 PAFEN,其中 PAFEN 包含三个

独立的子网络,其中 fc7、conv6_2 和 conv7_2 三个卷积层的卷积核,空洞率和填充为[8,3,1]、[7,5,1]和[12,3,0]。分别记录检测模型在迭代 30000 次时对 SGT-DET 测试数据集的 mAP 数值、平均检测时间和参数大小。

从表 3 可以发现,Faster-RCNN 的检测精度最高为 97.6%,MPSSD 的检测时间最短为 7.18ms,MPSSD 的参数大小最少为 7.9MB。FFSSD、DSSD 和 SSD 的实验结果表明增加特征提取网络的层数(增加 contextual information)能够改善 SSD 对小目标检测的精度,但不可避免地会增加计算量和模型参数。MOSSD 和 SSD 的实验结果表明利用 MobileNet 替换 VGG-16 在减少模型参数和降低检测时间方面带来的效果是明显的,但也存在检测精度上的下降。MPSSD 和 MOSSD 的实验结果表明 PAFEN 在三种定量评价指标中明显优于 OAFEN,在地面小目标的检测任务中,PAFEN 比 OAFEN 精度更高,速度更快。综合考虑检测精度(mAP)、平均检测时间(Times)和参数大小(Parameters)三个因素,MPSSD 优于其它检测模型,更适合无人机远距离地面实时检测任务。

为了更直观的说明 MPSSD 对地面小目标的检测效果,通过 Jupyter notebook 利用训练 30000 次的 MPSSD 模型分别对 4 种地面小目标图像进行检测,检测结果如图 4 所示。从图 4 可以看出,MPSSD 模型对四种地面小目标能够实现准确检测。

表 3 目标检测模型性能比较

Models	基础网络	提取网络	mAP/%	Time/ms	Parameters/MB
Faster-RCNN	VGG-16	RPN	97.6	65.03	1092.3
SSD	VGG-16	OAFEN	96.7	20.23	105.2
FFSSD	VGG-16	Feature Fusion Module	96.9	27.43	123.2
DSSD	ResNet101	Deconvolution Module	97.2	41.19	645.2
MOSSD	MobileNet	OAFEN	93.2	7.69	14.8
MPSSD	MobileNet	PAFEN	96.5	7.18	7.9

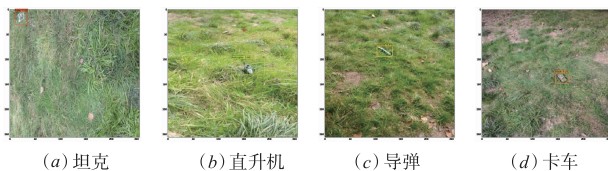


图 4 MPSSD 对地面小目标检测效果图

3.3 实验 2:子网络数量对 PAFEN 性能的影响

本实验比较不同数量子网络对 PAFEN 性能的影响。实验以包含三个子网络的 PAFEN 为参考,记 PAFEN-3;设计包含一个子网络的 PAFEN-1 和包含五个子网络的 PAFEN-5。PAFEN-1 可解释为 OAFEN 的深

度可分离空洞化,其中 fc7、conv6_2、conv7_2、conv8_2 和 conv9_2 五个卷积层的卷积核、空洞率和填充为[8,3,1]、[4,3,0]、[4,3,2]、[3,2,1]和[3,2,1]。PAFEN-5 为包含五个独立子网络的并行附加特征提取网络,即 fc7、conv6_2、conv7_2、conv8_2 和 conv9_2 直接与 conv4 相连,其中五个卷积层的卷积核、空洞率和填充为[8,3,1]、[8,4,0]、[8,5,1]、[8,5,0]和[8,6,3]。记录模型迭代 30000 次时模型对 SGT-DET 测试数据集的 mAP 数值、平均检测时间和参数大小。

从表 4 可以看出,随着 PAFEN 子网络数量的增加,模型的检测时间和参数都存在一定的增加,其中 PAFEN-3 比 PAFEN-1 时间增加 0.09ms、参数增加

0.4MB,PAFEN-5 比 PAFEN-1 时间增加 0.24ms,参数增加 1.4MB.但检测精度并没有随着子网络数量的增加而改善,PAFEN-5 的 mAP 比 PAFEN-3 低 3.3%.综合考虑检测精度、平均检测时间和参数大小三个评价指标,将 PAFEN 设计成三个子网络效果最好.

表 4 子网络数量对 PAFEN 性能的影响

	mAP/%	Time/ms	Parameters/MB
PAFEN-1	93.8	7.11	7.5
PAFEN-3	96.5	7.18	7.9
PAFEN-5	93.2	7.35	8.9

3.4 实验 3:空洞率选取对 PAFEN 性能的影响

fc7、conv6_2、conv7_2 与 conv4 直接相连,这三个卷积层对 PAFEN 的性能影响最大,因此本实验主要分析

表 5 空洞率选取对 PAFEN 性能的影响

	fc7	conv6_2	conv7_2	mAP/%	Time/ms	Parameters/MB
rise	3	5	7	93.7	7.16	7.8
decline	7	5	3	93.8	7.22	7.9
uniformity	5	5	5	94.1	7.17	7.7
convex	3	5	3	96.4	7.18	7.9

3.5 讨论

实验从 mAP、平均检测时间和参数大小三个方面比较了本文小目标检测方法 MPSSD 与经典算法(Faster-RCNN 和 SSD)和最新算法(FFSSD 和 DSSD)性能上的差异,也进一步分析了子网络的数量和空洞率的选取如何影响 PAFEN 的性能.本文地面小目标检测模型 MPSSD 的基础网络为 MobileNet,特征提取网络为 PAFEN.PAFEN 利用计算量小、感受野大的深度可分离空洞卷积单元建附加特征提取网络能消除 stride 操作造成图像中小目标信息丢失的问题.PAFEN 采用三个独立的子网络结构,使其能从 conv4 生成三组尺度独立的特征图,减少多尺度特征图之间的冗余度,提高目标检测的精度.但并非子网络的数量越多越好,例如 PAFEN-5 的检测精度低于 PAFEN-3.在 PAFEN-5 中,conv4(38*38)直接生成 conv8_2(3*3)或 conv9_2(1*1),由于两个子网络尺度之间跨度较大(38-3 或 38-1),在一定程度上会影响 conv8_2(3*3)和 conv9_2(1*1)两个尺度特征图的质量.同时,PAFEN-5 用五个独立的子网络生成五个尺度的特征图,特征图之间的独立性能为目标检测提供更丰富的选择,但过于独立的特征会影响图像目标的定位.因此,PAFEN 需要兼顾特征图的质量和特征图的多样性.PAFEN 使用 convex 空洞率策略获得了最高的检测精度,事实上,四种空洞率策略下 PAFEN 的检测精度均高于 OAFEN.在 convex 策略中,上路和下路网络初始特征层使用较小空洞率能

和比较 fc7、conv6_2、conv7_2 三个卷积层空洞率选取对 PAFEN 性能的影响.实验设计四种形式的空洞率策略:rise、decline、uniformity 和 convex,其中 convex 为 PAFEN 采用的策略.rise、decline 和 uniformity 三种策略的空洞率分别为[3,5,7]、[7,5,3]和[5,5,5].记录模型迭代 30000 次时对 SGT-DET 测试数据集的 mAP 数值、平均检测时间和参数大小.

从表 5 可以看出,四种空洞率策略下模型的检测时间和参数大小基本相同.rise、decline、uniformity 三种策略下模型的 mAP 数值也保持在同一水平,但 convex 策略下模型的 mAP 数值高于其他三种策略,比 rise 高 2.7%,比 decline 高 2.6%,比 uniformity 高 2.3%.换句话说,为包含更多层的子网络的初始层设置较低的空洞率有助于提升 PAFEN 的性能.

从 conv4 中获取更多的信息,增强 conv8_2 和 conv9_2 两个尺度特征图的质量,进而提高模型的检测精度.

Faster-RCNN 和 SSD 是经典目标检测算法的代表,Faster-RCNN 侧重于目标检测精度,而 SSD 侧重于目标检测速度,实验 1 的结果也验证了这一观点.MOSSD 用 MobileNet 替换基础网络 VGG-16 来进一步提高 SSD 的检测速度.对于本文的 SGT-DET 数据集,MOSSD 比 SSD 检测精度低 3.5%,精度下降的幅度还是比较大,这也间接表明 MobileNet 作为小目标检测模型的基础网络还存在改进的空间.FFSSD 和 DSSD 是针对小目标检测任务提出改进模型:FFSSD 通过引入 conv5_3 的信息;DSSD 通过融合特征提取网络反卷积变化后的特征.虽然 FFSSD 和 DSSD 在一定程度上提升了小目标检测的精度,但却增加了计算量.对于无人机远距离目标检测任务而言,除了满足检测精度要求外,图像的实时检测是另外一个重要的指标.MPSSD 在减少模型参数和计算量的前提下,提升了对地面小目标的检测速度.

4 结论

地面小目标检测任务具有重要的理论研究和实际应用价值.在 SSD 检测模型框架内,本文提出了一种并行附加特征提取网络 PAFEN.PAFEN 有效提升 SSD 对地面小目标检测精度和速度,并经理论分析和仿真实验证明了 PAFEN 的有效性.

对于基于 SSD 的地面小目标检测任务,改进基础

网络结构和 default box 生成策略同样能提升 SSD 模型的性能. 下一步的研究重点包括: (1) 研究适合小目标特征提取的基础网络; (2) 研究适合捕获小目标特征的 default box 生成策略.

参考文献

- [1] RAJA ABDULLAH R S A, SALAH A, ISMAIL A, et al. LTE-based passive bistatic radar system for detection of ground-moving targets[J]. *ETRI Journal*, 2016, 38(2): 302 – 313.
- [2] SUNGHO K, WOO-JIN S, SO-HYUN K. Robust ground target detection by SAR and IR sensor fusion using ada-boost-based feature selection[J]. *Sensors*, 2016, 16(7): 1117 – 1134.
- [3] XU H, YANG Z, CHEN G, et al. A ground moving target detection approach based on shadow feature with multi-channel high-resolution synthetic aperture radar[J]. *IEEE Geoscience and Remote Sensing Letters*, 2016, 13(10): 1 – 5.
- [4] BREITENSTEIN M D, REICHLIN F, LEIBE B, et al. On-line multiperson tracking-by-detection from a single, uncalibrated camera[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(9): 1820 – 1833.
- [5] 罗会兰, 童康, 孔繁胜. 基于深度学习的视频中人体动作识别进展综述[J]. *电子学报*, 2019, 47(5): 1162 – 1173.
LUO Hui-lan, TONG Kang, KONG Fan-sheng. The progress of human action recognition in videos based on deep learning: a review[J]. *Acta Electronica Sinica*, 2019, 47(5): 1162 – 1173. (in Chinese)
- [6] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. *Nature*, 2015, 521(7553): 436 – 444.
- [7] SCHMIDHUBER, JÜRGEN. Deep learning in neural networks: An overview[J]. *Neural Networks*, 2015, 61: 85 – 117.
- [8] KRIZHEVSKY A, SUTSKEVER I, HINTON G. Imagenet classification with deep convolutional neural networks[A]. *International Conference on Neural Information Processing Systems* [C]. Doha: Curran Associates Inc, 2012. 1097 – 1105.
- [9] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[A]. *Conference on Computer Vision and Pattern Recognition* [C]. Las Vegas: IEEE, 2016. 770 – 778.
- [10] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[A]. *Conference on Computer Vision and Pattern Recognition* [C]. Honolulu: IEEE, 2017. 1492 – 1500.
- [11] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[A]. *Conference on Computer Vision and Pattern Recognition* [C]. Columbus: IEEE, 2014. 580 – 587.
- [12] GIRSHICK R. Fast R-CNN[A]. *Proceedings of the IEEE International Conference on Computer Vision* [C]. Santiago: IEEE, 2015. 1440 – 1448.
- [13] HE K, ZHANG X, REN S, 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.
- [14] 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]. Canada: Curran Associates Inc, 2015. 91 – 99.
- [15] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[A]. *Conference on Computer Vision and Pattern Recognition* [C]. Las Vegas: IEEE, 2016. 779 – 788.
- [16] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[A]. *European Conference on Computer Vision* [C]. Cham: Springer, 2016. 21 – 37.
- [17] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[J]. *ArXiv Preprint*, 2015, arXiv: 1511.07122.
- [18] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, Atreus convolution, and fully connected CRFS[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 40(4): 834 – 848.
- [19] WANG P, CHEN P, YUAN Y, et al. Understanding convolution for semantic segmentation[A]. *Winter Conference on Applications of Computer Vision (WACV)* [C]. Lake Tahoe: IEEE, 2018. 1451 – 1460.
- [20] LI Z, ZHOU F. FSSD: feature fusion single shot multibox detector[J]. *ArXiv Preprint*, 2017, arXiv: 1712.00960.
- [21] FU C Y, LIU W, RANGA A, et al. DSSD: Deconvolutional single shot detector[J]. *ArXiv Preprint*, 2017, arXiv: 1701.06659.
- [22] ZHOU P, NI B, GENG C, et al. Scale-transferrable object detection[A]. *Conference on Computer Vision and Pattern Recognition* [C]. Salt Lake City: IEEE, 2018. 528 – 537.
- [23] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. *ArXiv Preprint*, 2017, arXiv: 1704.04861.
- [24] LI Z, PENG C, YU G, et al. Detnet: A backbone network for object detection[J]. *ArXiv Preprint*, 2018, arXiv: 1804.06215.

- [25] SINGH B, DAVIS L S. An analysis of scale invariance in object detection snip[A]. Conference on Computer Vision and Pattern Recognition[C]. Salt Lake City: IEEE, 2018. 3578 – 3587.
- [26] CAI Z, VASCONCELOS N. Cascade R-CNN; Delving into high quality object detection[A]. Conference on Computer Vision and Pattern Recognition[C]. Salt Lake City: IEEE, 2018. 6154 – 6162.
- [27] LI J, LIANG X, WEI Y, et al. Perceptual generative adversarial networks for small object detection[A]. Conference on Computer Vision and Pattern Recognition[C]. Honolulu: IEEE, 2017. 1222 – 1230.
- [28] 李宝奇, 贺昱曜, 何灵蛟, 王伟. 基于全卷积神经网络的非对称并行语义分割模型[J]. 电子学报, 2019, 47(5): 1058 – 1064.
- LI Bao-qi, HE Yu-yao, HE Ling-jiao, QIANG Wei. Asymmetric parallel semantic segmentation model based on full convolutional neural network[J]. Acta Electronica Sinica, 2019, 47(5): 1058 – 1064. (in Chinese)

作者简介



李宝奇(通信作者) 男, 1985年12月生, 天津宝坻人. 现于西北工业大学航海学院攻读博士学位, 研究方向为目标检测、识别和跟踪, 深度学习理论.
E-mail: bqli@mail.nwpu.edu.cn



强 伟 男, 1986年12月生, 陕西延安人. 现于西北工业大学航海学院攻读硕士学位, 研究方向为图像分类、图像语义分割及目标检测与识别等.



贺昱曜 男, 1956年生, 陕西富平人. 教授, 西北工业大学博士生导师, 主要研究方向为精确制导与仿真、智能控制与智能优化理论、图像处理理论与算法.
E-mail: heyyao@nwpu.edu.cn



何灵蛟 男, 1994年12月生, 甘肃会宁人. 现于西北工业大学航海学院攻读硕士学位, 研究方向为图像增强、图像语义分割及目标检测与识别.