

基于细化多尺度深度特征的目标检测网络

李雅倩, 盖成远, 肖存军, 吴超, 刘佳甲

(燕山大学工业计算机控制工程河北省重点实验室, 河北秦皇岛 066004)

摘 要: 现有深度卷积神经网络中感受野尺度单一, 无法适应目标的尺度变化和边界形变, 故此本文提出了一种提取并融合多尺度特征的目标检测网络. 该网络通过减少池化并在网络底层加入空间加信道压缩激励模块来突出可利用的细节信息, 生成高质量的特征图; 此外, 在深层网络中加入可变多尺度特征融合模块, 该模块具有多种尺度的感受野并可根据物体边界预测采样位置, 最后通过融合多尺度特征使网络具有更强的特征表达能力并且对不同尺度实例及其边界信息更具鲁棒性. 实验证明, 本文结构实现了比原有结构更高的平均精度, 与目前主流目标检测算法相比也具有一定优势.

关键词: 目标检测; 特征金字塔网络; 可变形卷积; 信道空间压缩激励; 多尺度特征融合

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 0372-2112 (2020)12-2360-07

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2020.12.011

Object Detection Networks Based on Refined Multi-scale Depth Feature

LI Ya-qian, GAI Cheng-yuan, XIAO Cun-jun, WU Chao, LIU Jia-jia

(1. Key Lab of Industrial Computer Control Engineering of Hebei Province, Yanshan University, Qinhuangdao, Hebei 066004, China)

Abstract: In the existing deep convolution neural network, the scale of receptive field is single, which could not adapt to the scale change and boundary deformation of the target. Therefore, a target detection network based on multi-scale feature extraction and feature fusion is proposed in this paper. The proposed network reduces pooling and adds space as well as channel compression excitation module at the bottom of the network to highlight the details and generate high-quality feature map. Besides, a variable multi-scale feature fusion module is added to the deep network, which has a multi-scale receptive field and can predict the position according to object boundary. Finally, the multi-scale feature fusion is used to enable the network of stronger ability of feature expression and is more robust to different scale and flexible boundary of instances. Experimental results show that the proposed structure achieves higher average accuracy than the original structure, and also has certain advantages compared with the state-of-the-art algorithms.

Key words: object detection; feature pyramid network; deformable convolution; channel spatial squeeze excitation; multi-scale feature fusion

1 引言

目标检测作为计算机视觉领域最重要最具挑战性的基础任务之一, 在自动驾驶、安防监控等领域发挥着重要作用. 近年来, 卷积神经网络(CNN)广泛应用于计算机视觉领域, CNN 可以对图像进行连续的滤波操作, 高层语义信息由原始数据输入并逐层抽象, 所以基于深度 CNN 框架的目标检测算法相对于传统目标检测算法取得了明显的性能提升.

目标检测的难点主要在于小目标的检测, 通常的

思路是利用多尺度特征. 2015 年 Liu 等人提出了一种利用多层次特征图的目标检测 SSD^[1] 算法, 利用多种分辨率特征图同时进行检测任务, 一定程度上提升了对小目标检测的性能. 2017 年 Tsung-Yi Lin 等人提出的特征金字塔网络 FPN^[2], 将网络不同层级特征图进行自上而下的连接, 综合高层语义信息与低层级位置信息, 成为现在目标检测领域普遍使用的一种结构.

FPN 算法是一种加强特征提取网络特征表达的方法, 该方法利用 CNN 模型来高效提取和融合图片不同维度特征, 可在单一图片尺度下生成多维度特征的表

达,相比于传统图像金字塔大大降低了计算开销.但 FPN 算法仅考虑到不同维度之间的特征融合,忽略了对同维度视觉信息在不同感受野下进行处理的重要性,并且特征提取网络固定形状的卷积过程无法适应实例的尺度变化和边界形变.

基于上述工作和问题,本文在 FPN 算法基础上进行改进,提出一种对目标尺度和边界形变具有强鲁棒性的可变多尺度特征感知网络(Deformable Multi-scale Feature Perception Network, DMFPN).本算法采用信道加空间压缩激励模块强化细节特征,并使用可变多尺度特征融合模块(Deformable Multi-scale Feature Fusion Module, DMFFM)提升网络有效宽度,最后将深层网络提取的多尺度特征以特征金字塔的方式进行回传,充分结合高层多尺度语义特征和低层级细节特征,进而提升算法的检测性能.

2 基本原理

FPN 算法流程如图 1 所示,该算法使用残差网络^[3]进行特征提取并采用特征金字塔结构将高层网络强语义信息与底层网络的精确位置信息相结合,形成更加丰富的特征表示.但由于残差网络中池化操作较多,网络深层输出的特征图分辨率过低,融合操作无法将小物体语义信息向下回传,导致算法对小物体无法得到较高的分类得分.并且目标检测任务的准确度很大程度上取决于网络生成特征图的精细程度,所以本文的主要动机是形成一系列具有高细节特征的特征图和更具鲁棒性的特征描述.

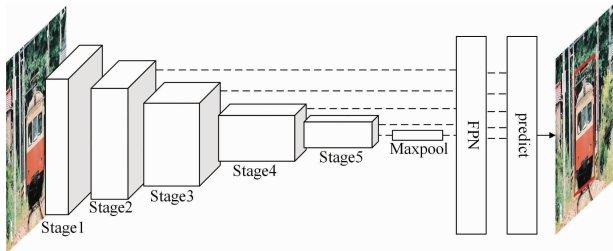


图1 基于残差网络的FPN目标检测算法结构

图 2 给出了本文算法的网络结构,相对于 FPN 特征提取网络,具体改进内容如下:

(1) 在 ResNet50 基础上,删除 Stage4 与 Stage5 之间的池化层,并复制一个 Stage5 来替代原本 Stage5 后的最大池化层,作为 Stage6,生成具有更高语义信息的特征图.池化的移除可以保证算法在较高分辨率特征图上进行检测,为了弥补移除池化层导致的感受野减小,同时更好的对目标边界进行采样,Stage5 和 Stage6 均使用膨胀率为 2 的可变形卷积^[4].

(2) 在特征提取网络存在池化操作的 Stage1、

Stage2、Stage3 阶段使用信道加空间压缩激励模块,该模块可使网络自行学习到特征图中信息的重要程度,强化细节和前景信息,弱化背景信息和噪声,有助于细节信息向网络后端流动.

(3) 为了提升对图片中的尺度差异、长宽比例、视角变化和非刚性形变的建模能力,在提取高级语义信息的深层网络中,使用由不同感受野的可变形卷积组成的可变多尺度特征融合模块(DMFFM),对不同尺度卷积核提取的特征进行融合,形成对尺度和形变更具鲁棒性的特征表征.

通过上述改进,本文算法网络结构输出的最小分辨率特征图相较于 FPN 算法,尺度由原图的 1/16 提升至 1/8,众多研究表明,特征图分辨率与小目标物体的检测性能成正相关^[5].结合信道加空间压缩激励模块对底层特征图的处理,使本文算法输出的特征图可以更好的保留小目标信息及边缘位置信息;其次,本文算法的可变多尺度特征融合模块可以提取并融合更加丰富的多尺度信息并充分发挥了可变形卷积对物体边界强大的建模能力,使本文网络更适合对多尺度及复杂边界物体的检测.

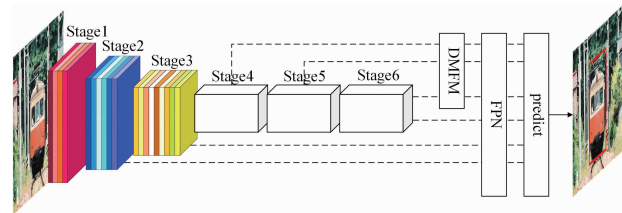


图2 可变多尺度特征感知网络(DMFPN)结构

2.1 信道加空间压缩激励

特征图中不同通道间和不同空间位置包含不同的特征响应,某些通道和空间位置存在噪声并对当前检测任务存在干扰.针对这些问题,文本使用文献[6]提出的信道压缩激励模块并在其基础上添加空间压缩激励操作,对特征图不同信道及空间位置重新进行权重赋值,减弱多个池化对细节特征的影响.

2.1.1 信道压缩激励

信道压缩激励通过显式的建模信道间的相互关系,来自适应的校准信道上的特征响应,区分各个信道所包含的不同特征的重要程度.信道压缩激励模块结构如图 3 所示,与文献[6]不同的是,该方法仅使用全局平均池化进行特征图的压缩,本文使用全局平均池化和最大池化两种池化方式将输入特征图进行压缩,生成两个 1 维特征图 $S_1, S_2 \in \mathbf{R}^{1 \times 1 \times c}$,该一维特征图某种意义上具有输入图片的全局特征.将特征图 S_1, S_2 输入到由两个全连接层 FC_1, FC_2 和一个非线性激活函数 ReLU 组成的多层感知机 MLP,其作用及超参数设置可

参考文献[6]. 图3中‘ \oplus ’表示将MLP输出的两个1维特征图按通道进行加和操作,得到最终的压缩特征图 $M \in \mathbf{R}^{1 \times 1 \times C}$, M 的值代表网络学习到的各个通道的相关程度,最后将 M 使用 Logistic 函数进行归一化,得到 $M^* = (m_1^*, m_2^*, \dots, m_c^*)$, 图3中‘ \otimes ’表示将 M^* 特征图中元素按通道与输入特征图 F 相乘,完成对原始特征图的重校正,整个过程可由式(1)、(2)表示为:

$$M^* = \text{Logistic} \left\{ \text{MLP} \left[\frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W f_x(i, j) \right] + \text{MLP} \left[\max_{i \in H, j \in W} f_x(i, j) \right] \right\} \quad (1)$$

$$F^* = F \otimes M^* \quad (2)$$

式中 $F \in \mathbf{R}^{H \times W \times C}$ 为输入特征图,其中 H, W, C 分别为特征图的高、宽和通道数; $f_x(i, j)$ 为输入特征图 F 的第 x 通道中坐标为 (i, j) 点像素值; F^* 为模块产生的新特征图.

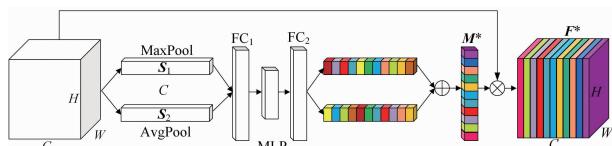


图3 信道压缩激励结构图

2.1.2 空间压缩激励

与信道压缩激励不同,空间压缩激励注重区分图片不同空间位置的重要程度,其结构如图4所示.该模块以 F^* 作为输入,首先将 F^* 在通道维度做最大池化和平均池化得到特征图 $P_1, P_2 \in \mathbf{R}^{H \times W \times 1}$, 图4中‘ \ominus ’表示通道级联操作,级联后得到的特征描述为 $P_3 \in \mathbf{R}^{H \times W \times 2}$, P_3 中包含两种池化操作后得到的所有空间信息,之后对 P_3 使用卷积尺寸为 $7 \times 7 \times 2$ 的卷积层对其需要加强或抑制的位置信息进行编码和融合,编码得到的特征描述作为 $P_4 \in \mathbf{R}^{H \times W \times 1}$, 图4中‘ \otimes ’表示将 P_4 中各个位置像素值按通道对输入特征图 F^* 对应像素位置进行乘法操作,完成对不同空间位置信息的重校正,空间压缩激励可用式(3)、(4)表示.

$$P_4 = \text{Logistic} \{ f^{7 \times 7} [\text{AvgPool}(F^*); \text{MaxPool}(F^*)] \} \quad (3)$$

$$P^* = P_4 \otimes F^* \quad (4)$$

式中: $f^{7 \times 7}$ 为 7×7 大小卷积层; AvgPool 和 MaxPool 分别为以通道维度对特征图进行平均池化和最大池化操作; P^* 为空间压缩激励模块输出的特征图.

2.2 可变多尺度特征融合模块(DMFFM)

不同尺度卷积核的感受野范围不同,得到的特征信息也不同,可变多尺度特征融合模块可以提取并融合不同尺度特征,从而获得更具鲁棒性的特征表征;此外,该模块使用的可变形卷积可以对卷积的采样位置进行动态调整,提升特征对物体尺度和非刚性形变的

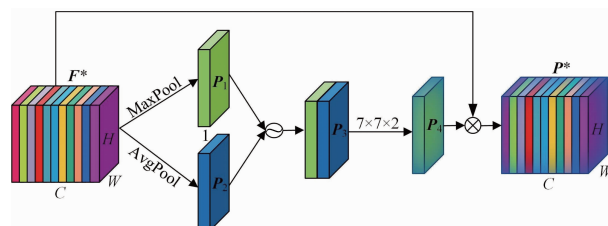
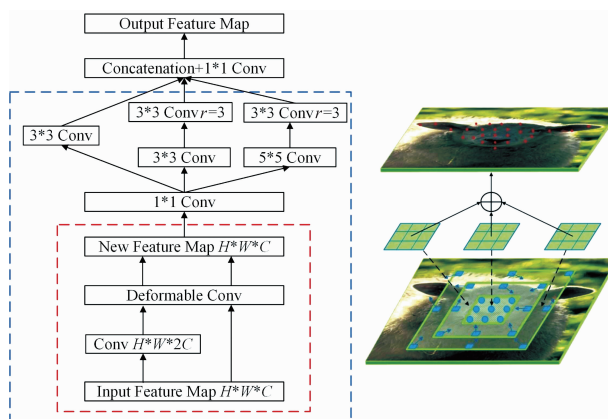


图4 空间压缩激励结构图

表征能力.

可变多尺度特征融合模块的结构和功能如图5所示,该模块结构类似 Inception^[7] 网络结构,为了在特征提取网络同一阶段收集多尺度空间信息,模块设计了感受野范围分别为 $3 \times 3, 9 \times 9, 15 \times 15$ 的三个并行支路,图中 r 代表膨胀率. 模块由偏移采样模块和特征融合模块两部分组成,其中偏移采样模块结构如图5(a)蓝色虚线区域所示,偏移采样模块分为偏移预测和求取偏移采样点响应两个步骤. 图5(b)表示多尺度特征融合模块三个支路对应于原图的感受野范围,蓝色点表示原有采样位置,蓝色箭头表示偏移预测步骤预测的偏移量,红色点表示最终采样位置.

可变多尺度特征融合模块具体运算步骤如下:首先,对于输入到该模块的特征图使用偏移采样模块对其像素偏移量进行训练和学习,利用预测偏移量可以对原有特征图每个像素位置的像素值进行重新计算并替换生成新特征图;然后,对于偏移采样模块输出的新特征图,使用 1×1 卷积核对其进行降维,减少模块后续计算复杂度,并使用三种不同感受野大小卷积核分别对新特征图进行采样和抽象,从而得到多尺度深度特征;最后,特征融合模块负责对之前卷积采样得到的多尺度深度特征进行融合并输出. 下面,将对各个功能模块分别进行介绍.



(a) 结构图 (b) 功能图

图5 可变多尺度特征融合模块结构及功能图

2.2.1 偏移采样模块

偏移采样模块可弥补传统卷积规则格点采样策略

的不足,提升网络对非刚性形变的建模能力,具体计算过程如下:

(1) 偏移预测:偏移预测操作如图 5(a) 红色虚线区域所示,该操作对输入特征图 $I \in \mathbf{R}^{H \times W \times C}$ (Input Feature Map) 使用 $2C$ 个卷积核构成的卷积层进行处理,得到偏移特征图 $O \in \mathbf{R}^{H \times W \times 2C}$ (Offset Feature Map),加入采样点偏移量后的卷积操作由式(5)表示为:

$$y(p_n) = \sum_{p_n \in \mathbf{R}} w(p_n) \cdot x(p_n + \Delta p_n) \quad (5)$$

式中: p_n 为 I 上某个坐标位置, $x(p_n)$ 代表 p_n 位置像素值; $w(p_n)$ 为 p_n 位置对应的滤波器权重; Δp_n 为预测得到的偏移量, I 的单个维度对应 O 相同像素位置两个维度,分别代表 x 、 y 轴方向偏移量, Δp_n 由 x 、 y 组成; $p_n + \Delta p_n$ 代表新采样点坐标值。

(2) 偏移采样点响应:由于偏移预测得到的 x 、 y 偏移量可能是任何数值,新采样点可能无法对应输入特征图的像素位置,为此,我们使用双线性插值的方式计算偏移采样点响应值.设偏移采样点响应值为 $x(p_n^*) = x(p_n + \Delta p_n)$, $x(q)$ 为特征图 q 点像素值,则 $x(p_n^*)$ 如式(6)所示:

$$x(p_n^*) = \sum_q \text{Bilinear}(p, q) \cdot x(q) \quad (6)$$

式中: Bilinear 为双线性插值公式。

设新采样点 $p_n + \Delta p_n$ 的坐标值是 (a, b) , 该坐标点可能为浮点数,首先将 (a, b) 转换为四个整数 $\text{floor}(a) = x_1$ 、 $\text{ceil}(a) = x_0$ 、 $\text{floor}(b) = y_1$ 、 $\text{ceil}(b) = y_0$, 将四个整数值进行组合就可以得到四组整数点像素值 $\text{PX}_1 = \text{Pixel}(x_0, y_0)$ 、 $\text{PX}_2 = \text{Pixel}(x_0, y_1)$ 、 $\text{PX}_3 = \text{Pixel}(x_1, y_0)$ 、 $\text{PX}_4 = \text{Pixel}(x_1, y_1)$, 其中 floor 为向上取整操作, ceil 为向下取整操作, Pixel 为像素值, 根据双线性插值公式, 式(6)可由式(7)表示:

$$\begin{aligned} x(p_n^*) = & \frac{(x_1 - a)(y_1 - b)}{(x_1 - x_0)(y_1 - y_0)} \cdot \text{PX}_1 \\ & + \frac{(a - x_0)(y_1 - b)}{(x_1 - x_0)(y_1 - y_0)} \cdot \text{PX}_2 \\ & + \frac{(x_1 - a)(b - y_0)}{(x_1 - x_0)(y_1 - y_0)} \cdot \text{PX}_3 \\ & + \frac{(a - x_0)(b - y_0)}{(x_1 - x_0)(y_1 - y_0)} \cdot \text{PX}_4 \end{aligned} \quad (7)$$

新得到的偏移点像素值 $x(p_n^*)$ 将代替原有输入特征图 I 上对应像素值 $x(p_n)$, 输出新特征图, 此新特征图作为接下来卷积操作的输入, 完成对新采样点的卷积采样操作。

2.2.2 特征融合模块

特征融合模块主要对偏移采样模块中提取出的多尺度特征通过卷积的方式进行融合. 由于卷积操作可以对不同通道所包含的多尺度信息进行交互来实现跨

通道的信息组合, 可最终形成对尺度和形变具有鲁棒性的特征描述. 如图 5(a) 所示, 其中为了减少模块参数量, 新特征图 (New Feature Map) 首先使用 1×1 大小卷积层将通道数降低为原来的 $1/8$, 通道串联后再将其还原为原始通道数. 通道升降维的过程可以视作对通道间信息的线性组合变化, 该操作配合卷积后接的非线性激活函数可以大大提升模型的非线性特性, 非线性的提升可以使网络形成更加抽象的特征表达并且增加了模型的容错能力。

其具体实现步骤为: 首先, 采集三个支路输出特征图并使用 Concatenate 函数进行通道维度的数组拼接, 拼接生成的特征图通道数是输入通道数的 $3/8$; 然后, 在 Stage4 ~ Stage5 分别使用 1024、256 和 256 个尺寸为 $1 \times 1 \times 384$ 的卷积核对特征图维度进行升维, 以得到期望的输出通道数. 最后, 特征图经 ReLU 处理后输出, 参加构建特征金字塔结构。

3 实验结果与分析

实验在 Ubuntu16.4 系统下进行, 深度学习框架为 Pytorch-0.3.1, 显卡型号为 NVIDIA GTX1080Ti, GPU 通用并行计算架构为 CUDA8.0 和 cuDNN5.0. 数据集为 PASCAL VOC 数据集, 其中包含 33101 张训练图片和 4952 张测试图片, 目标类别为 20 类, 实验选取均值平均精度 (mAP) 作为客观评价指标。

3.1 算法实现细节

特征提取网络使用 ImageNet^[8] 预训练模型初始化, 新添加的网络层采用文献[9]提出的初始化方法进行权重填充, 损失函数同文献[10]. 训练集将训练 15 轮次 (epoch), 训练初始学习率为 0.001, 之后每 5 轮次下降为原来的 0.1, 动量因子设置为 0.9, 每批次样本数量设置为 2. 热力图可视化方法使用文献[11]提出的 Grad-CAM 方法。

3.2 实验结果对比

3.2.1 客观评价指标对比

表 1 为 DMFPN 算法与基于三种不同网络结构的 FPN 算法的客观评价指标对比. DMFPN 算法网络层数为 96 层, 其中包含信道加空间压缩激励层 39 层, 包含神经元总数为 25752 个, 平均精度达到 81.98%, FPN 算法在特征提取网络分别 ResNet50、ResNet101、DetNet59 的均值平均精度分别为 78.42%、80.25% 和 80.41%. PASCAL VOC 数据集中 15 类目标本文算法均高于其他三种网络结构。

表 2 中对可变多尺度特征融合模块 (DMFFM) 进行消融实验并分别与 Cascade RCNN^[12]、R-FCN^[13]、DSSD^[14]、YOLOv3^[15] 等算法进行了精度及速度对比, 在此实验中为了验证 DMFFM 模块有效性, 特征提取网络使用 Res-

Net101, 实验中仅使用 DMFFM 模块, 验证算法为 FPN 算法. 实验结果显示, 本文算法速度上与主流双阶段算法持平, 精度上均高于其他目标检测算法.

表 1 DMFPN 算法与基于三种特征提取网络的 FPN 算法精度对比

Class	FPN (ResNet50)	FPN (ResNet101)	FPN (DetNet59)	DMFPN
Aero	86.51	85.82	86.87	87.22
Bike	87.59	87.10	80.91	85.62
Bird	78.92	81.92	79.40	83.19
Boat	70.58	74.13	72.57	75.28
Bottle	55.20	58.94	68.23	67.88
Bus	86.70	87.55	88.07	87.91
Car	87.14	88.56	88.68	88.69
Cat	87.96	89.21	88.72	89.77
Chair	58.58	65.89	64.37	67.05
Cow	84.72	86.34	86.40	88.31
Table	73.44	76.74	72.11	77.32
Dog	84.76	80.07	88.17	89.41
Horse	89.20	88.06	88.74	89.45
M-bike	86.10	85.12	84.09	87.00
Person	78.53	83.16	84.12	85.53
Plant	51.77	52.38	55.62	55.15
Sheep	78.62	81.91	82.36	84.71
Sofa	74.46	79.15	78.57	78.82
Train	86.78	87.35	88.12	88.82
TV	80.74	78.66	81.94	82.39
mAP	78.42	80.25	80.41	81.98

表 2 DMFFM 模块有效性分析及与各类目标检测算法客观评价指标对比

算法类型	算法	网络结构	mAP	FPS
双阶段 算法	Faster-RCNN	ResNet101	76.62	7
	Cascade RCNN	ResNet101	79.67	10
	Cascade RCNN + FPN	ResNet101	80.68	9
	FPN	DetNet59	80.41	8
	FPN	ResNet101	80.25	7
	R-FCN	ResNet101	79.51	7
	R-FCN	ResNet50	78.63	9
	FPN + DMFFM	ResNet101	81.12	7
单阶段 算法	SSD513	VGG16	79.83	19
	SSD513	ResNet101	80.62	9
	SSD321	VGG16	77.24	32
	SSD321	ResNet101	77.31	23
	DSSD513	ResNet101	80.73	5
	YOLOv3	DarkNet53	79.42	37

表 3 将本文网络结构所包含参数量、训练时间和最佳轮次与骨干网络分别为 ResNet50、ResNet101、DetNet59 的 FPN 算法进行对比, 其中训练时间是在相同硬件条件下训练数据集 15 个轮次所需时间; 网络模型大小为算法中所包含的参数总量, 其中存储着网络中所有神经元所对应的参数值; 最佳轮次为测试结果最好的模型轮次数, 轮次数越低表明算法训练拟合速度越快. 本文算法训练所需时长为 23 小时, 网络生成模型大小为 625.2 MB, 测试最佳轮次为 8, 通过对比可以发现, 本文模型参数和训练所需时间相较于其他三种网络结构有所增加, 但本文算法训练拟合速度更快.

表 3 四种网络结构的模型参数量大小、训练时长和最佳轮次对比

算法	FPN (ResNet50)	FPN (ResNet101)	FPN (DetNet59)	DMFPN
模型大小	247.0 MB	416.8 MB	598.8 MB	625.2 MB
训练时长	14 h	20 h	22 h	23 h
最佳轮次	14 epoch	12 epoch	11 epoch	8 epoch

3.2.2 主观视觉效果对比

本文 DMFPN 算法和 FPN 算法部分可视化结果如图 6 所示, 可视化图片选择场景中实例尺度变化较大的图片, 左侧为 FPN 算法可视化结果, 右侧为本文算法的可视化结果, 密集且存在差异区域使用绿色虚线框标记并放大. 对比图 6 (a) 和图 6 (b) 发现, 图片中存在一个尺度较小的鸟类目标, 本文算法成功检测, FPN 算法对此目标发生漏检. 图 6 (c)、(d) 图片中实例存在重叠且存在大量小目标, 本文算法对图片中实例无一漏检, FPN 算法对汽车、摩托车、人各漏检一例. 图 6 (e)、(f) 图中两个船类目标存在重叠且尺度差异较大, 两只船本文算法都检测到, FPN 算法漏检一个船类目标. 图 6 (g)、(h) 图片中有两个人类目标背景阴暗且存在大面积重叠, 本文与 FPN 算法对此目标均发生漏检, 但右侧人类目标实例本文算法成功检测, FPN 算法对此发生漏检. 图 6 (i)、(j) 图片中车辆目标互相存在遮挡, 本文算法在目标存在大面积遮挡情况下依然可以很好的检测到目标, FPN 算法漏检两个车辆目标.

图 7 展示了卷积响应热力图, 图 7 (a) 为 FPN 算法卷积响应热力图; 图 7 (b) 为 DMFPN 算法卷积响应热力图. 由于信道加空间压缩激励模块和可变多尺度特征融合模块的共同作用, 可以发现本文网络的卷积响应可以更好的覆盖物体边界, 无论是响应强度还是响应范围都要比原始 FPN 算法更好. 观察发现, 图 7 (a) 所采用的传统卷积其采样方式为规则格点采样, 使热力图响应也呈现矩形格点式分布并且卷积响应较弱; 图 7 (b) 为本文网络结构的热力图响应, 由于可以使采样位置根据物体形状发生偏移, 采样点可以更好地覆

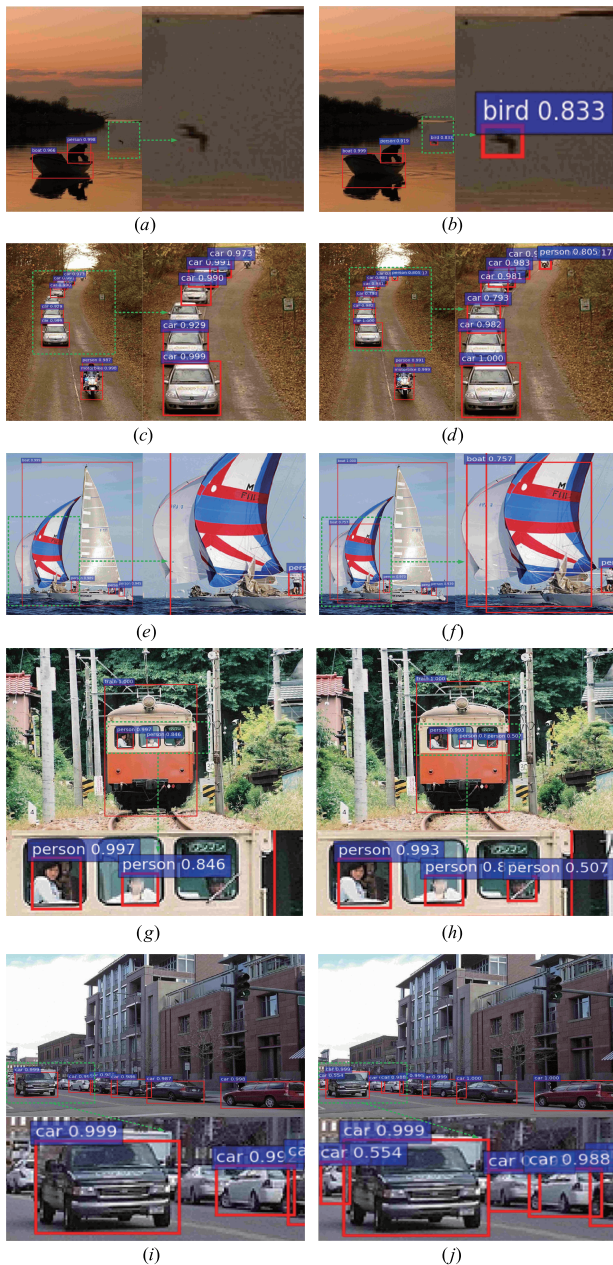


图6 FPN算法(左)与DMFPN算法(右)可视化效果对比

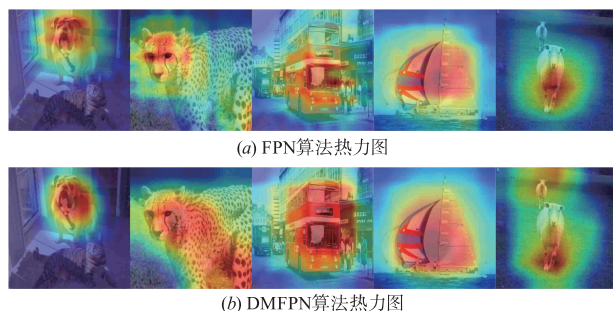


图7 卷积响应热力图对比

盖物体范围,并且信道加空间压缩激励模块可以增强

前景位置的卷积响应强度.

4 结论

本文针对目标检测任务中存在的不足,基于 FPN 网络提出了 DMFPN 网络,该网络结合信道加空间压缩激励模块和可变多尺度特征融合模块,可加强细节特征并形成更加丰富的特征表示,对小尺度目标具有很强的鲁棒性,也能适应目标轮廓的多变性.为验证算法有效性,在 PASCAL VOC 数据集进行实验及可视化并与其他算法进行对比.实验结果表明,相比于其他算法,本文算法对目标检测任务具有更好的检测效果.

参考文献

- [1] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multi-box detector [A]. European Conference on Computer Vision (ECCV) [C]. Amsterdam, Netherlands: Springer, 2016. 21-37.
- [2] Lin T Y, Dollar P, Girshick R, et al. Feature pyramid networks for object detection [A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Honolulu, USA: IEEE Computer Society Press, 2017. 963-944.
- [3] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Boston, USA: IEEE Computer Society Press, 2015. 770-778.
- [4] Dai J F, Qi H Z, Xiong Y W, et al. Deformable Convolutional Networks [EB/OL]. http://openaccess.thecvf.com/content_ICCV_2017/papers/Dai_Deformable_Convolutional_Networks_ICCV_2017_paper.pdf. 2018-10-6.
- [5] 刘颖,刘红燕,范九伦,等.基于深度学习的小目标检测研究与应用综述[J].电子学报,2020,48(03):590-601.
Liu Y, Liu H Y, Fan J L, et al. A survey of research and application of small object detection based on deep learning [J]. Acta Electronica Sinica, 2020, 48(03): 590-601. (in Chinese)
- [6] Hu J, Shen L, Albanie S, et al. Squeeze-and-Excitation Networks [EB/OL]. <https://arxiv.org/abs/1709.01507>. 2018-7-11.
- [7] Christina S, Vincent V, Sergey I, et al. Rethinking the inception architecture for computer vision [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Las Vegas, USA: IEEE Computer Society Press, 2016. 2818-2826.
- [8] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge [J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [9] He K M, Zhang X Y, Ren S Q, et al. Delving deep into rec-

- tifiers: Surpassing human-level performance on imagenet classification [A]. IEEE International Conference on Computer Vision (ICCV) [C]. Santiago, Chile: IEEE Computer Society Press, 2015. 596 – 604.
- [10] Girshick R. Fast R-CNN [A]. IEEE International Conference on Computer Vision (ICCV) [C]. Santiago, Chile: IEEE Computer Society Press, 2015. 1440 – 1448.
- [11] Ramprasaath R, Michael C, Abhishek D, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization [EB/OL]. <https://arxiv.org/abs/1610.02391>. 2019-1-12.
- [12] Cai Z W, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection [A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Salt Lake City, USA: IEEE Computer Society Press, 2018. 2603 – 2611.
- [13] Dai J F, Li Y, He K M, et al. R-FCN: Object Detection via Region-based Fully Convolutional Networks [EB/OL]. <https://arxiv.org/abs/1605.06409>. 2018-7-29.
- [14] Fu C Y, Liu W, Ranga A, et al. DSSD: Deconvolutional Single Shot Detector [EB/OL]. <https://arxiv.org/abs/1701.06659>. 2018-6-15.
- [15] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement [EB/OL]. <https://pjreddie.com/media/files/papers/YOLOv3.pdf>. 2019-4-11.

作者简介



李雅倩 (通信作者) 女, 1982 年 4 月出生, 湖南岳阳人. 2010 年毕业于法国鲁昂应用科学学院信息学专业, 现为燕山大学电气工程学院副教授, 研究方向为: 模式识别.
E-mail: yaqianli@126.com



盖成远 男, 1993 年 11 月出生, 黑龙江大兴安岭人. 燕山大学硕士研究生, 主要研究方向为: 深度学习与模式识别.
E-mail: gaichengyuan@126.com