

基于双检测头的轻量级目标检测算法^①



郭昕刚, 纪超群, 翟 双, 程 超

(长春工业大学 计算机科学与工程学院, 长春 130012)

通信作者: 程 超, E-mail: chengchao@ccut.edu.cn

摘 要: 为解决轻量级目标检测算法中由于分类损失较大导致算法精确度低的问题, 提出一种对目标的位置与分类使用双检测头的检测方法. 算法中用卷积头对位置进行检测, 用全连接头对分类进行检测; 分类检测时特征图经过卷积层后融合位置回归分支的特征图, 再使用全连接层对特征图进行处理; 并提出分组全连接的方式进一步减少全连接层的计算量. 在 VOC 数据集上对算法进行训练, 结果表明, 改进后模型的分类损失有了明显的下降, 有效地提升了轻量级目标检测算法的检测精确度, 算法在 VOC 测试集上达到 70.08% 的精确度.

关键词: 深度学习; 目标检测; 轻量级; 双检测头; 特征融合

引用格式: 郭昕刚, 纪超群, 翟双, 程超. 基于双检测头的轻量级目标检测算法. 计算机系统应用, 2022, 31(11): 254-260. <http://www.c-s-a.org.cn/1003-3254/8768.html>

Lightweight Target Detection Algorithm Based on Double Detection Head

GUO Xin-Gang, JI Chao-Qun, ZHAI Shuang, CHENG Chao

(School of Computer Science and Engineering, Changchun University of Technology, Changchun 130012, China)

Abstract: In order to solve the problem of low accuracy caused by large classification loss in the lightweight target detection algorithm, a method of detecting the location and classification of the target with double detection heads is proposed. In the algorithm, the convolution head is used to detect the position, and the full connector is used to detect the classification. In the classification detection, after the feature map passes through the convolution layer, the feature map of the fused position regression branch is processed through the full connection layer. A grouping full connection method is proposed to further reduce the amount of calculation in the full connection layer. The algorithm is trained in VOC datasets. The results show that the classification loss of the improved model is significantly reduced, and the detection accuracy of the lightweight target detection algorithm is effectively improved. The accuracy of the algorithm on the VOC test set has reached 70.08% mAP.

Key words: deep learning; target detection; lightweight; double detection head; feature fusion

目标检测是计算机视觉任务中最为基础的任务也是较为关键的任务, 目标检测任务统一了图像中目标的类别与位置的信息, 为计算机视觉中的场景解析提供了详细的基本信息. 目前表现较好的目标检测算法大多数都是基于深度学习的方法^[1], 包括早期的经典 CNN 算法^[2], 以及两阶段算法中最具代表性的算法 Faster R-CNN^[3]; 目前应用最为广泛的一阶段检测算法

YOLO 系列 (v1^[4]、v3^[5]) 和 SSD 算法^[6]. 最初此类算法大多使用特征金字塔结构^[7]、残差网络^[8] 等不同的技术手段提高模型的精确度以使模型能够在实际场景下应用. 在精确度方面有了较大的提升, 但模型不断的变大, 计算参量越来越多, 导致一些模型只能在计算资源较大的服务器上运行, 对计算与存储资源受限的移动终端设备并不友好. 轻量级目标检测算法的出现为移

① 基金项目: 吉林省教育厅科研规划项目 (JKKH20210739KJ)

收稿时间: 2022-02-11; 修改时间: 2022-03-14; 采用时间: 2022-03-21; csa 在线出版时间: 2022-08-12

动终端设备的算法应用提供了一种可能。

轻量级目标检测算法使用较为简练的特征提取网络,且特征提取网络一般会使用特殊的手段来简化计算,从而达到减少计算参量的目的。如 MobileNet 网络^[9]使用的深度可分离卷积; ShuffleNet 网络^[10]使用的分组卷积; Tiny YOLO 网络^[11]使用轻量化的 Darknet 特征提取网络。这些算法都在一定程度上减少了计算参量从而简化了模型的计算参数。然而在使用 MobileNetV3 网络与 SSD 算法结合的实验中发现,在训练过程中算法对位置的回归较好,但对类别的回归就会出现较高的分类损失,使网络整体的损失较高,从而导致模型检测精确度较低的问题。

在目标检测算法中需要检测目标的类别与位置,其中位置信息只需用目标的中心点坐标以及宽高就可以表示,但分类信息需要较多的目标特征才能对目标有较好的描述。目标检测算法使用轻量级的特征提取网络会导致算法对图片的特征提取不充分,而且轻量级的检测算法中为了简化运算大多数会使用卷积代替分类网络中的全连接操作,但 Wu 等人^[12]证明了全连接层对分类有较好的效果,使用卷积层会影响算法对分类检测的处理效果。

本文以降低轻量级目标检测算法中分类损失从而提升算法的检测精确度为目标,提出了一种轻量化双检测头目标检测算法:1) 构建 MobileNetV3-SSD 轻量级目标检测算法,使用 MobileNetV3 替换原 SSD 算法中的 VGG16 网络,并用深度可分离卷积替代常规卷积进一步减小网络计算参数;2) 对特征图使用双检测头进行处理,位置回归部分使用卷积头处理特征图,分类回归部分使用分组全连接层处理特征图,并将位置回归部分的特征图与分类回归部分进行特征融合,进而增强网络对分类回归的处理能力从而提高算法的检测精确度。

1 相关理论

1.1 通道注意力机制

在已有的研究中已经证明注意力机制对各种计算机视觉任务都有帮助^[13],其中通道注意力机制能够较好地整合不同通道间的信息,使网络能够更加侧重信息量较大的通道,从而提升网络的性能。在 MobileNetV3 网络中使用轻量化的 Squeeze-and-Excitation (SE) 模块^[14]作为通道注意力机制,并取得了较好的结果,Coordinate

Attention^[15]中提出在 SE 模块中用平均池化粗暴对特征图进行压缩,虽然能够对通道间的信息量有较好的评判,但其忽略了通道上的位置信息。SE 模块结构如图 1 所示。

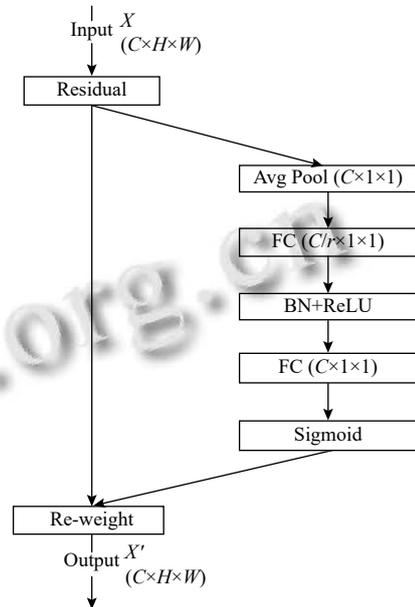


图 1 SE 模块结构图

SE 模块使用平均池化对输入的特征图进行特征压缩,在使用两层全连接层对通道信息进行交互,最后用非线性激活函数对各个通道的信息量进行量化。Coordinate Attention 通过对特征图从 H, W 两个方向进行平均池化,在一定程度上保留了对位置信息的分析。Coordinate Attention 的结构如图 2 所示。

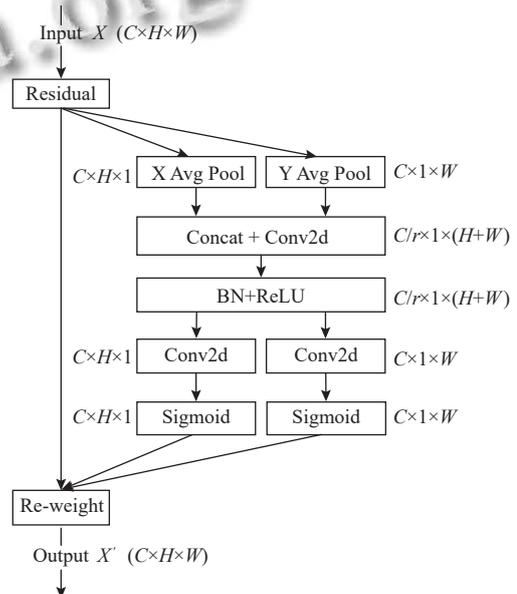


图 2 Coordinate Attention 结构图

Coordinate Attention 模块对特征图的 H 与 W 两个方向分别进行平均池化,生成两组针对 H 和 W 不同方向的特征向量,然后对拼接后的两组向量使用卷积层代替全连接层进行信息交互,拆分拼接的张量分别进行一次卷积操作与一次非线性变换,最后将得到的两组不同方向上的权重与原特征图进行通道权重相乘。

本文采用 Coordinate Attention 模块代替 SE 模块重新分配通道的权重,使网络在关注整体分类信息的基础上增加了对位置信息的参数学习,从而可以更好地利用有限的网络结构对图像进行更加细致的特征提取,充分利用网络的每一层,进而减少网络参量。

1.2 SSD 目标检测算法

SSD 目标检测算法是最早使用多张特征图进行目标检测的算法,相比较于 YOLO 算法虽然在速度上有所减少但精确度以及对小目标的处理能力都有较大的提升.算法的结构如图 3 所示。

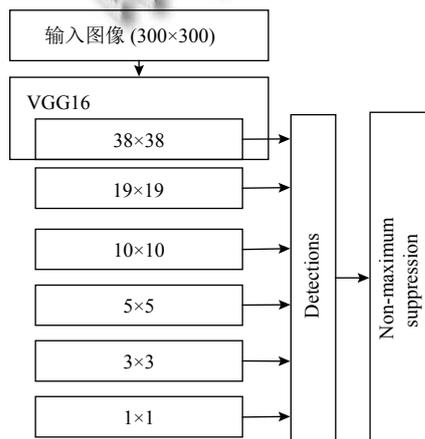


图 3 SSD 算法结构图

SSD 算法采用 VGG16 作为特征提取网络,使用 VGG16 的 Conv4_3 层输出的 38×38 大小的特征图作为基准特征图,在其后增加卷积网络得到 5 种大小不同的特征图. SSD 算法利用 Faster R-CNN 的锚框机制使用多级特征图在大尺寸特征图上检测较小目标,在小尺寸的特征图上检测较大目标.这种多级特征图检测的方式有效地改善了网络对不同大小目标检测的准确性。

2 本文算法

本文通过对 MobileNetV3-SSD 网络进行调整,使用双检测头分别对位置信息以及分类信息进行检测.位置检测时使用卷积的方式对特征图处理;分类检测

时使用全连接层对融合位置检测部分后的特征图进行处理,实现在不增加网络结构的前提下对网络结构的复用,并针对全连接的计算量太大的问题借鉴分组卷积的思想提出分组全连接的方式以达到减少计算量的目的。

2.1 分组全连接

全连接层在网络中将上一层的所有神经元与本层的神经元都进行连接,可以看作是对前边网络提取到特征的综合.虽然全连接层能将所有特征进行综合但其需要对所有的神经元计算导致全连接层的计算量太大.在现有的目标检测网络中大部分使用卷积层来代替全连接操作从而减少网络的计算量。

Zhang 等人^[16]证明了全连接层对分类网络的重要性,Wu 等人^[12]提出的双检测头算法也证明了全连接层在图像分类方面的优势,本文同样使用全连接层对用于分类的特征图进行处理,并针对其计算量大的问题根据分组卷积的思想提出分组全连接进行优化处理.全连接层与分组全连接层示意图如图 4 所示。

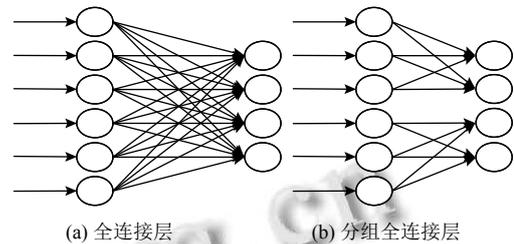


图 4 全连接层与分组全连接层示意图

图 4(a) 为全连接层示意图,网络的每个输出神经元与每个输入神经元之间都会进行连接.图 4(b) 是分组全连接层,借鉴分组卷积的思想对全连接层的计算做分组处理,对输入的神经元个数进行平均分组,分组个数为 N ,同时对输出的神经元个数进行相同的分组处理,将输入神经元的每一小组与其相对应的输出神经元的每一小组对应组成 N 个较小单位的全连接.假设全连接层的输入神经元个数为: in_size , 输出神经元的个数为 out_size , 分组的个数为 G , 全连接层的计算参数量为 F , 分组全连接层的计算参数量为 F_g . 则有全连接层参数量与分组全连接参数量的计算公式如式 (1)、式 (2) 所示:

$$F = in_size \times out_size \quad (1)$$

$$F_g = (in_size/G) \times (out_size/G) \times G \quad (2)$$

为对比全连接层与分组全连接层计算参数量,用 F 与 F_g 做除法进行比较,计算公式如式(3)所示:

$$F_g = \frac{in_size \times out_size}{(in_size/G) \times (out_size/G) \times G} = G \quad (3)$$

通过式(3)可以看出相比较于全连接分组全连接的计算量少了几组的个数倍.虽然分组全连接层能够有效地减少计算参数,但是直观分析其结构会发现这种分组会导致神经元之间的信息交互受到一定的影响,这有违背于全连接的宗旨.经过实验验证发现使用两层分组全连接与一层全连接的结构会保证所有特征完成信息交互,从而解决了分组带来的影响.分组全连接层与全连接层结合的网络示意图如图5所示.

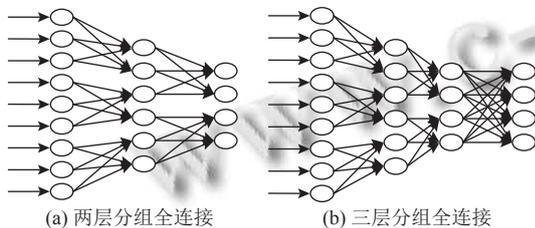


图5 分组连接示意图

从图5中可以看出使用两层分组全连接,前一层的组数大于后一层的组数时,能够使组间的神经元完成信息交互,在两层分组全连接层后使用一个全连接层就能够保证所有的神经元都有一个信息的交互,在两层分组全连接后神经元的个数会缩减为网络识别的类别数目,所以最后一层的全连接层并不会带来较大的计算量.本文后续实验部分使用 MobileNetV3 网络对分组全连接层与卷积层进行对比分析,证明在对图像分类上分组全连接的表现好于卷积层.

2.2 特征图处理模块

本文采用对特征提取网络 MobileNetV3 输出的 6 张特征图使用双检测头的处理方式.位置检测时特征提取网络输出的特征图经过 4 个卷积层后会直接送入到位置回归的检测头中进行位置回归计算.对于类别回归分支,使用 3 个卷积层对特征图处理后在与位置回归分支的特征图进行特征融合处理,用两层分组全连接与一层全连接层对特征图进行处理后送入到分类回归检测器中进行类别回归.其网络结构如图6所示.

模块中位置回归部分使用卷积层处理后直接送入到检测器中进行 loss 计算;分类回归部分使用位置回归的特征图进行特征融合处理,能够有效地增加特征

图上的特征信息,提高了网络的利用率,使用分组全连接层与全连接层处理特征图能为后期的分类处理提供更加丰富的语义信息,进而提高了网络对分类处理的能力.

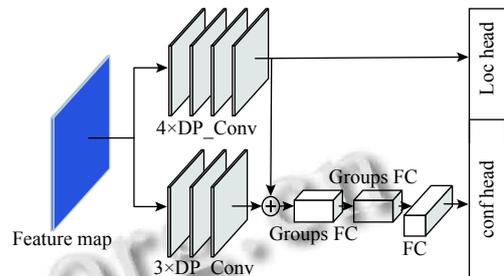


图6 特征提取模块示意图

2.3 网络结构

本文选用 MobileNetV3-SSD 作为本次实验的算法框架.实验中将轻量级特征提取网络 MobileNetV3 的 SE 模块替换为 Coordinate Attention 模块,以便在网络传播过程中保留更多的位置信息进而弥补后期对位置回归处理过程中的简单操作.使用特征提取网络 MobileNetV3 第 16 层与第 20 层输出 19×19 、 10×10 两种尺寸的特征图,对 10×10 大小的特征图依次进行 4 次步长为 2 的卷积操作生成 5×5 、 3×3 、 2×2 与 1×1 大小的 4 种特征图,最终得到 6 种不同大小的特征图.将 6 层特征图中的 2×2 、 5×5 、 10×10 大小的 3 层特征图依次进行上采样处理后进行特征融合.算法框架图如图7所示.

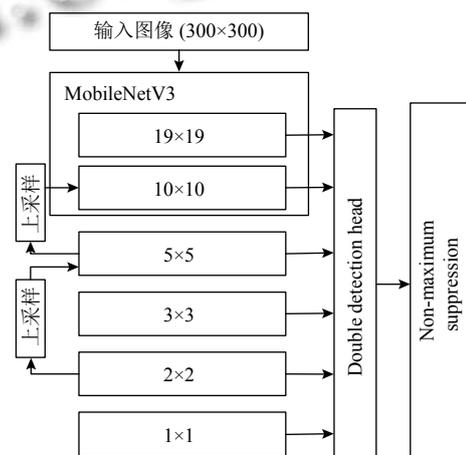


图7 本文算法框架图

MobileNetV3 网络输出的 6 张特征图都经过双检测器分别进行位置与分类的回归,对位置回归使用卷

积层处理后将6层处理的结果拼接成一个特征向量送入到检测器使用 smooth L1 函数进行位置 loss 计算; 对类别回归分支, 特征提取网络输出的特征图在经过卷积层后融合位置检测部分的特征图, 再用两层分组全连接与一层全连接对特征图上每个点的不同预测框进行类别预测。

在改进的算法中对网络结构进行调整, 针对不同的检测任务使用与之对应的检测方式. 对特征提取网络输出的特征图在位置检测分支使用卷积头检测; 在分类检测部分使用全连接头进行检测, 并复用位置检测的特征图, 以增强算法对分类检测的处理能力。

2.4 损失函数

在算法中两个检测头 (位置检测与分类检测) 一起进行回归预测, 算法的整体损失为两部分损失之和, 总损失计算如式 (4) 所示:

$$L(x, c, l, g) = \frac{1}{N} (L_{\text{conf}}(x, c) + \alpha L_{\text{loc}}(x, l, g)) \quad (4)$$

其中, 全连接层检测头的类别回归损失, 使用交叉熵损失函数得到. 卷积层检测头的位置回归损失, 使用 smooth L1 损失函数得到。

3 实验分析

3.1 实验数据及环境

本次实验在验证分组全连接部分使用的数据集为计算机视觉和机器学习文献中标准的基准数据集 CIFAR-10, CIFAR-10 数据集包含 10 个类别, 每个类别有 6000 张图像, 每张图片的尺寸为 32×32 大小, 数据集中一共有 50 000 张训练图片和 10 000 张测试图片. 目标检测网络使用的数据集为公开的目标检测领域较为基准的 PASCAL VOC 数据集, 测试数据集使用 PASCAL VOC2007_test. 使用 PASCAL VOC2007 与 PASCAL VOC2012 作为训练数据集, 两个版本的数据集都包含 20 个类别, 两个数据集共有 16 551 张训练图片。

实验平台为 Intel Core i7 处理器与两张 Quadro RTX5000 显卡, 总显存为 32 GB; 软件平台为 Ubuntu 18.04 操作系统与 CUDA 10.0、cudnn 7.6.5、PyTorch 深度学习框架。

3.2 分组全连接验证

为了验证分组全连接的有效性使用 MobileNetV3 网络作为分类网络, 对 CIFAR-10 数据集进行分类训练, 在实验过程中只改变网络的最后部分, 即对特征图

最后处理使用卷积层与分组全连接两种形式做比较. 对两种网络在 CIFAR-10 数据集上进行 100 个 epoch 的训练, 训练结果如图 8 所示。

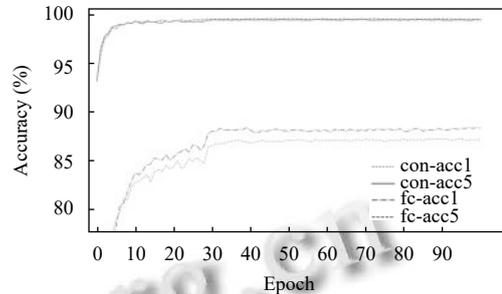


图8 卷积头与分组全连接头精确度对比

图 8 中 con-acc1 与 con-acc5 表示全连接层的 top1、top5 精度; fc-acc1 与 fc-acc5 表示分组全连接层的 top1、top5 精度. 由图 8 可得, 在 top5 精度下两种方法的分类精确度较为接近; 在 top1 精度下两种方法在前 10 个 epoch 中卷积头的表现比较好, 在之后的 epoch 中分组全连接层的精确度逐渐超过卷积层的精确度. 证明分组卷积能够在不损失精度的前提下有效的减少计算参数量。

3.3 目标检测网络训练

使在 PASCAL VOC2007 与 PASCAL VOC2012 两个数据集下对网络进行训练. 在实验中使用 MobileNetV3-SSD 算法对上述实验的正确性做出了验证, 实验证明通过对网络结构的调整以及使用双检测头等手段能够有效地解决算法在训练过程中对分类回归不友好的问题. 在数据集上对 MobileNetV3-SSD 算法以及改进后的算法进行 800 个 epoch 的训练。

为了便于比较算法中位置回归与分类回归的变化情况以及对整体算法的影响情况, 对训练过程中位置回归与分类回归的 loss 变化情况进行可视化, 其训练结果对比如表 1 所示。

表1 训练结果对比表

Epoch	MobileNet-SSD		本文算法	
	位置损失	分类损失	位置损失	分类损失
100	1.066914797	3.111261606	1.065126181	2.886003184
200	1.025483727	2.538415424	1.033130605	2.418995404
300	1.060274959	2.627688408	1.059114933	2.317253661
400	0.848206937	2.090683546	0.863407135	1.918196213
500	0.863017678	2.102664709	0.848778903	1.516336465
600	0.979661703	2.431370502	0.977824986	1.430396795
700	0.809709013	1.977009654	0.796233833	1.499994962
800	0.811149274	1.987000942	0.808602752	1.389887293

表 1 是单检测头算法与本文的双检测头算法在训练中位置回归与分类回归的损失变化情况. 在训练过程中前 400 个 epoch 无论是位置损失还是分类损失下降都较快, 且位置损失在 400 个 epoch 后逐渐趋于平稳, 而分类损失还在不断的缓慢下降. 由表 1 可以看出本文算法的分类损失相比较于原算法能够有较好的回归, 且与位置回归的差别有了明显的缩小. 本文提出的双检测头的思想通过降低分类检测部分的损失有效减小算法整体训练过程中的损失从而提高算法的精确度. 算法的整体训练对比结果如图 9 所示.

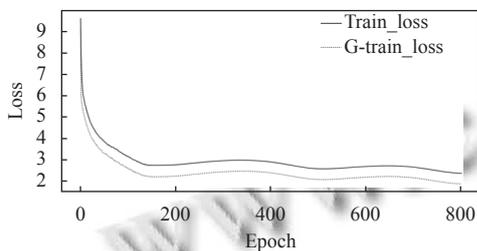


图 9 算法整体训练对比

图 9 中实线为 MobileNetV3-SSD 算法的训练 loss 变化情况, 虚线为本文算法的训练 loss 变化情况. 从图 9 可以看出本文算法在训练过程中的 loss 对比原有算法有了明显下, 算法的收敛效果较好, 模型的检测精确度有明显提高.

3.4 实验结果分析

在 VOC2007_test 测试数据集对下对本文算法进行测试, 检测精度结果如图 10 所示.

```
Average Precision Per-class:
aeroplane: 0.7149119399119398
bicycle: 0.7863999291847394
bird: 0.7221005346548918
boat: 0.6315080648151344
bottle: 0.27738539908149443
bus: 0.8041025641025641
car: 0.6736507936507936
cat: 0.8498930303343233
chair: 0.6060983780541408
cow: 0.7641152525343207
diningtable: 0.7797521089856206
dog: 0.8343020467015482
horse: 0.8841461207052067
motorbike: 0.491794841183597
person: 0.6928614612398607
pottedplant: 0.24449247509596023
sheep: 0.4989430376143175
sofa: 0.780958755560911
train: 0.8738972496107941
tvmonitor: 0.6819206597885217
FPS:43.850845

Average Precision Across All Classes:0.7008572281368257
```

图 10 本文算法评估结果

本文算法在 VOC2007_test 测试集上 20 个类的 mAP (平均检测精度) 为 70.08%, 算法的检测速度为 43.85 fps, 在达到了实时性要求的前提下网络的检测精度有了较大提高.

为了更好地验证算法的可行性, 在相同实验条件

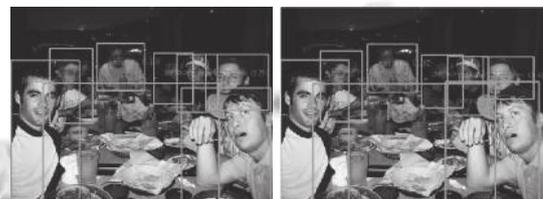
下, 通过对比不同轻量级算法之间的评估结果, 其比较结果如表 2 所示. 算法使用双检测头在只增加少量计算参数的情况下能够有效地提升算法的检测精确度, 与同等结构下使用单检测头的算法相比, 平均检测精确度提高了 2.48%.

由表 2 可得, 本文算法与原有的 VGG16-SSD 算法相比较在模型精确度仅下降 7.17% 的基础上得到的模型大小缩小了近 10 倍. 且本文算法与同类算法 Tiny YOLOv3 相比模型大小与精确度上都有很大的提高. 所以本文提出的轻量级双检测头目标检测算法具有实际的应用价值.

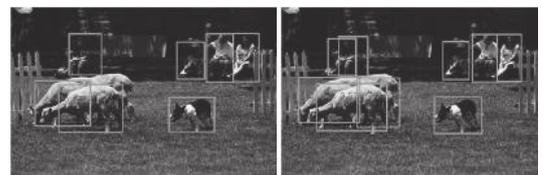
表 2 算法比较结果

算法	模型尺寸 (M)	mAP (VOC) (%)	FPS
Tiny YOLOv3	33.4	58.4	34
VGG16-SSD	105	77.26	35
MobileNetV3-SSD	21.3	65.2	38
MobileNetV3-SSD-lite	9.9	63.6	43
Fpn-MobileNetV3-SSD	10.2	67.6	42
本文	11.6	70.08	43

图 11 列出了本文算法与原算法的检测结果对比图, 可以看出算法对各个类别的检测精确度均有差别, 本文算法在类别准确度与检测物体个数均优于 MobileNet-SSD 算法.



(a) 本文算法检测结果 1 (b) 原算法检测结果 1



(c) 本文算法检测结果 2 (d) 原算法检测结果 2



(e) 本文算法检测结果 3 (f) 原算法检测结果 3

图 11 算法检测结果对比图

4 结论与展望

文中提出一种基于 MobileNet-SSD 网络改进的算法,使用 Coordinate Attention 模块代替 MobileNetV3-SSD 网络中的 SE 模块,对 SSD 算法位置回归与分类回归分别使用卷积层与全连接层进行处理,并将分类回归的特征图与位置回归部分的特征图进行特征融合,同时提出分组全连接的方法减少全连接层的计算量,进一步优化算法的性能.使用 PASCAL VOC2007 与 PASCAL VOC2012 数据集对本文算法训练,相比较于单检测头本文提出的算法分类损失能够有效地减小,从而提高了算法的检测精确度,算法在 VOC2007_test 上取得了 70.08% 的精确度,对比原有的 MobileNetV3-SSD 网络,精确度提升了 2.48% 的精确度,进一步验证了该方法的有效性.

参考文献

- 1 Minar MR, Naher J. Recent advances in deep learning: An overview. arXiv: 1807.08169, 2018.
- 2 Sharif Razavian A, Azizpour H, Sullivan J, *et al.* CNN features off-the-shelf: An astounding baseline for recognition. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Columbus: IEEE, 2014. 512–519.
- 3 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]
- 4 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6517–6525.
- 5 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv: 1804.02767, 2018.
- 6 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot MultiBox detector. Proceedings of the 14th European Conference on European Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37.
- 7 吴之昊,熊卫华,任嘉锋,等.基于轻量级 SSD 的电力设备锈蚀目标检测.计算机系统应用,2020,29(2): 262–267. [doi: 10.15888/j.cnki.csa.007280]
- 8 He KM, Zhang XY, Ren SQ, *et al.* Identity mappings in deep residual networks. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 630–645.
- 9 Howard AG, Zhu ML, Chen B, *et al.* MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv: 1704.04861, 2017.
- 10 Zhang XY, Zhou XY, Lin MX, *et al.* ShuffleNet: An extremely efficient convolutional neural network for mobile devices. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 6848–6856.
- 11 Ma J, Chen L, Gao ZY. Hardware implementation and optimization of tiny-YOLO network. Proceedings of the 14th International Forum on Digital TV and Wireless Multimedia Communications. Shanghai: Springer, 2017. 224–234.
- 12 Wu Y, Chen YP, Yuan L, *et al.* Rethinking classification and localization for object detection. Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 10186–10195.
- 13 Tsotsos JK. A Computational Perspective on Visual Attention. Cambridge: MIT Press, 2011.
- 14 Hu J, Shen L, Sun G. Squeeze-and-Excitation networks. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141.
- 15 Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design. Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13708–13717.
- 16 Zhang CL, Luo JH, Wei XS, *et al.* In defense of fully connected layers in visual representation transfer. Proceedings of the 18th Pacific-Rim Conference on Multimedia Information Processing. Harbin: Springer, 2017. 807–817.

(校对责编:牛欣悦)