基于组卷积特征融合的 One-Stage 目标检测模型

鲍先富,强赞霞,李丹阳,杨 瑞
(中原工学院,河南 郑州 450007)

摘 要:由于移动终端计算能力和内存大小的限制,在移动设备上进行实时目标检测具有非常大的挑战性。为了更好地在无人驾驶汽车等移动设备上进行目标检测,该文以 YOLOv3 单阶段目标检测模型为基础,对部署在移动设备上的目标检测模型进行优化,以提高模型的检测精度(MAP)并降低计算复杂度。具体改进措施为:基于 DarkNet-53 为主干网络引入组卷积和通道洗牌技术;基于 M. G. Hluchyj 等学者提出的网络设计指导原则,对主干网络的残差单元和下采样单元进行修改优化;为减轻 YOLOv3 模型对于密集目标的漏选和标签重写问题,引入特征混合金字塔模型。通过在 Pascal VOC2007 和 VOC2012 数据集上进行实验对比,优化模型的整体精度较 YOLOv3 提高 8.17%,模型参数量降低 1.21 M,在与 YOLOv4 的参数量大体相等的情况下达到了 YOLOv4 的检测精度。

关键词:卷积神经网络;目标检测;残差网络;特征融合金字塔;通道洗牌;组卷积
中图分类号:TP391 **文献标识码**:A **文章编号**:1673-629X(2021)11-0086-09
doi:10.3969/j.issn.1673-629X.2021.11.015

One-Stage Target Detection Model Based on Group Convolution Feature Fusion

BAO Xian-fu,QIANG Zan-xia,LI Dan-yang,YANG Rui
(Zhongyuan University of Technology,Zhengzhou 450007,China)

Abstract:Due to the limitations of mobile terminal computing power and memory size,real-time target detection on mobile devices is quite challenging. In order to better perform target detection on mobile devices such as driverless cars,based on YOLOv3 single-stage target detection model,the target detection model deployed on the mobile device is optimized to improve the detection accuracy (MAP) and reduce the computational complexity. The specific improvement measures are as follows:the introduction of volume and channel shuffling technology based on the DarkNet-53 backbone network; based on the network design guidelines proposed by scholars such as MG Hluchyj,the residual unit and down-sampling unit of the backbone network are modified and optimized; the feature mixture pyramid model is introduced to reduce the YOLOv3 model's omission of dense targets and label rewriting. Through experimental comparison on the Pascal VOC2007 and VOC2012 data sets,the overall accuracy of the optimized model is 8.17% higher than that of YOLOv3,and the model parameter is reduced by 1.21 M. The detection accuracy of YOLOv4 is reached when the parameter quantity of YOLOv4 is roughly equal.

Key words:convolutional neural network;target detection;residual network;feature fusion pyramid;channel shuffle;group convolution

0 引 言

随着社会车辆的快速增多,道路交通变得愈加复杂。为了提高道路安全,避免人为驾驶失误造成不必要的交通事故,越来越多的研究学者开始对无人驾驶领域进行研究,其中包括目标识别在内的计算机视觉任务。在车辆行驶过程中,针对汽车、行人等关键目标的识别与检测任务对车辆安全行驶和避障有着举足轻重的作用,得益于近些年来深度学习的快速发展和硬件算力的巨大飞跃,基于深度学习的目标检测算法取得的检测效果受到各方学者的青睐。该文针对无人驾驶领域已有的关于车辆和行人的检测和识别任务进行优化,针对车辆密集场所提高模型检测精度,检测算法以基于 YOLOv3[1] 目标检测模型对目标检测框架进行优化,其中的改进主要为:(1)对 Darknet-53[2] 主干网络中的残差模块和基于步长为 2 的卷积下采样方法进行改进;(2)为提升原 YOLOv3 目标检测网络对于不

收稿日期:2020-09-23 **修回日期**:2021-01-25
基金项目:河南省科技计划项目(182102210126)
作者简介:鲍先富(1996-),男,硕士,研究方向为数字图像处理、深度学习和目标检测;强赞霞,博士,副教授,研究方向为图像及信号处理、目标识别与检测、模式识别与智能系统等。

同尺度目标的检测能力,该模型在原模型的基础上加入自适应空间特征融合模块(ASFF),在提升对不同目标尺度包容能力的同时提升网络的检测能力,降低对关键目标的漏检率;(3)实验在 PASCAL VOC2007 和 PASCAL VOC2012 数据集[3]上进行对比测试,取得了比原始 YOLOv3 检测框架精度高 12% 的检测效果,且对改进后主干网络的推理速度没有显著影响。

1　相关工作

传统的目标检测算法多是基于设计手工特征,通过观察待检目标进行人为设计特征学习方式,其算法检测过程主要分为区域选择、特征提取和分类器分类三个步骤。区域选择一般通过滑动窗口的方式对图像区域进行遍历,其中滑动窗口的大小及长宽与检测模型的检测精度与速度密切相关。传统特征提取算法主要包含 SIFT[4]、HOG[5]等,这种传统算法对于检测目标的多样性、光照强度、背景的复杂性具有较差的鲁棒性。

自 2013 年以来,深度学习得到迅速发展和广泛研究,其中基于深度学习的目标检测算法分为:区域法和回归法。基于区域提议方面的目标算法如 Faster RCNN[6],由于实时检测效果不够理想,无法用于无人驾驶和检测等领域。在基于回归的单阶段目标检测算法中,张海涛等学者[7]基于 SSD 算法[8]引入注意力机制和扩大感受野的方式,增强高层特征图所包含的高级特征信息,实现检测效果的提升,但是其总体精度仍然较低;Redmon 等学者[9]在 YOLOv2 的基础上结合 ResNet、特征金字塔等思想提出 YOLOv3 算法,该算法在实时性和检测精度方面得到广泛提升,但是在小目标和目标密集环境中存在漏选和目标重写现象;顾恭等学者[10]在 YOLOv3 的基础上通过增加主干网络输出特征图数量,增加对不同尺寸目标的检测能力;Bochkovskiy A 等[11]在综合许多已有学者研究成果的基础上,通过组合不同的优化技巧对 YOLOv3 进行优化,使其精度达到新的高度。

综合当今研究的优势与不足,该文以 YOLOv3 目标检测网络为基础作进一步改进,选择 YOLOv3 目标检测网络基于如下原因:(1)YOLOv3 为单阶段目标检测网络,在实际检测应用中能够达到实时检测效果;(2)为了单独分析优化方法的效果,避免受 YOLOv4 中的多优化方式影响,单独分析文中优化方法的优劣,所以不使用最新的 YOLOv4 作为基准网络比较;(3)YOLOv3 目标检测模型对于密集、多尺度目标存在漏选特点、密集目标检测回召率低等问题,使用该优化手段验证对 YOLOv3 的改进程度和效果。

2　模型介绍

2.1　YOLOv3 模型

YOLOv3[1]是 Redmom 等学者于 2018 年提出的单阶段目标检测网络,该算法主干网络结构为 Darknet-53[2],如图 1 所示。该算法结合残差网络、特征金字塔多尺度检测等一系列优秀的网络设计思想,能够达到较好的检测效果和几乎实时的检测速度。主干网络 Darknet-53 引入了 ResNet[12]网络模型中的残差单元并进行重新组合,主干网络中的残差单元将传递的特征图依次进行卷积核为 3 × 3、步长为 2 的卷积操作和下采样处理,再依次进行卷积核为 1 × 1、步长为 1,卷积核为 3 × 3、步长为 2 的卷积处理,之后再与输入特征相加,由此组成残差单元。主干网络 Darknet-53 通过残差单元堆叠、卷积和下采样处理得到不同尺度的特征图,并通过上采样和卷积处理结合不同尺度的特征图,形成特征金字塔结构,从而实现不同尺度目标的

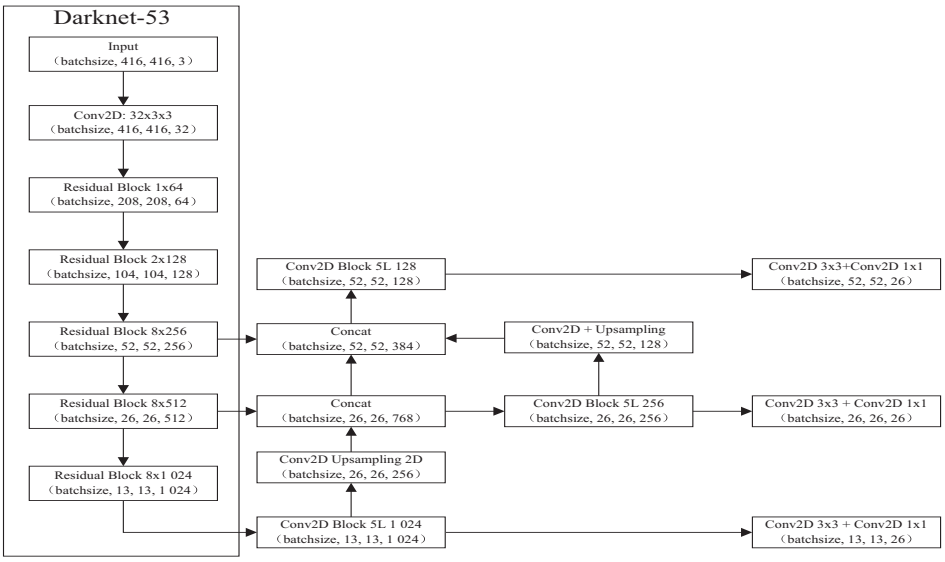


图 1　Darknet-53 结构

检测,有效避免由于网络深度过高而造成的梯度消失问题。

2.2 分组卷积与通道洗牌

基于 YOLOv3 目标检测框架和最新的目标检测算法思想,该文以主干网络 DarkNet-53[2] 为基础对网络结构进行修改,重新优化其中的残差单元并引入组卷积和通道洗牌技术。组卷积技术是由 Cohen T 等学者[13] 提出,主要是对输入的特征图进行通道分组,然后对每组特征图分别进行卷积操作,如图 2 左为传统卷积技术,右为分组卷积技术。在分组卷积中,若输入特征图的大小为 $C \times H \times W$,输出的特征图数量为 N,如果要分为 G 个分组,则每组输入特征图输入通道为 C/G,每组输出特征图的数量为 N/G,每个卷积核尺寸为 $C/G \times K \times K$。假设卷积核的总数仍为 N,每组卷积核为 N/G,由于卷积核只与同组的输入特征图进行卷积操作,每个卷积组的总参数量为 $C/G \times N \times K \times K$,所以由计算对比可知,分组卷积比传统卷积在参数量上减少为原来的 $1/G$,其组操作如图 2 右图所示。分组 1 的输出特征图数量为 2,使用 2 个卷积核,每个卷积核的输入通道数为 4,分组中每个卷积核计算所用通道数与输入特征图通道数相同,卷积核只和同组的输入特征图做卷积操作,而不与其他组的输入特征图做卷积操作。

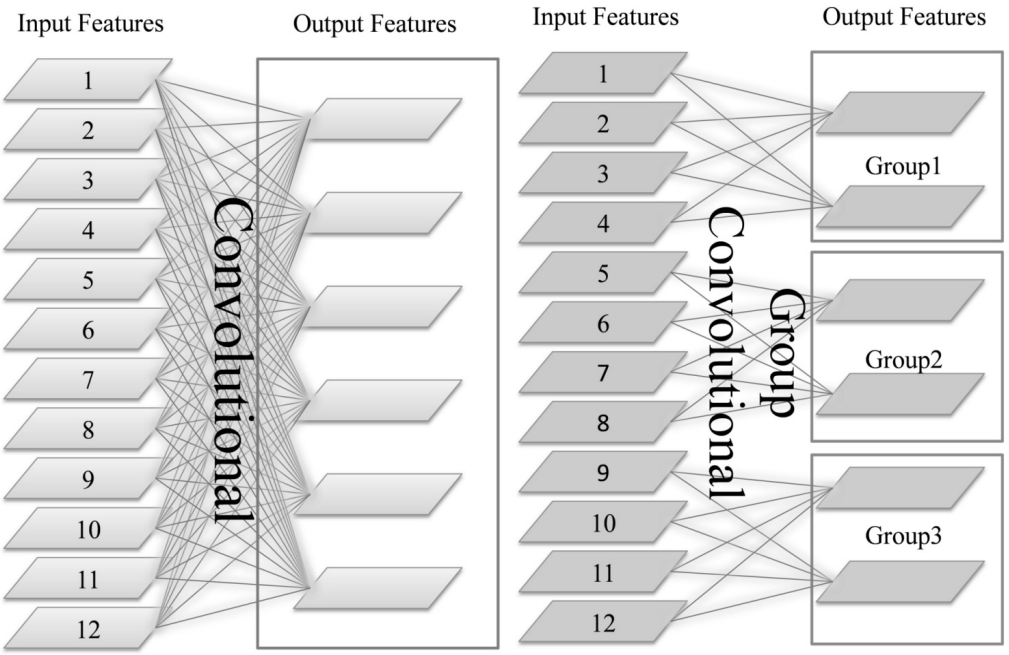


图 2 普通卷积与分组卷积示意图

YOLOv3 网络模型采用传统卷积和深度可分离卷积(depth separable convolution,DSC)进行特征提取和特征筛选,其中可分离卷积对网络结构存在性能瓶颈[14],如果直接在通道组内进行逐点卷积(point wise convolution,PWC),会导致各个通道内的信息不能进行相互流通交流。为了解决瓶颈问题,文中对主干网络结构引入通道洗牌技术(channel shuffle,CS)[15]。通道洗牌技术[16]是一种组内卷积和整组卷积的折中解决方案,通过组合 3 × 3 和 1 × 1 卷积的方式进行深度可分离卷积。假设输入的特征图大小为 $h \times w \times c_1$,输出的特征图为 $h \times w \times c_2$,此处进行 1 × 1 逐点卷积的浮点运算量为:

$$F = h \cdot w \cdot c_1 + h \cdot w \cdot c_2 \tag{1}$$

由公式(1)知:当 $c_1 \times c_2$ 远大于 9 时,可以发现其可分离卷积的计算量增长主要在 1 × 1 逐点卷积上,引入分组卷积后,在组内进行 1 × 1 逐点卷积,对于分成 g 组的分组卷积的计算量为(FLOPs):

$$F = 9 \cdot h \cdot w + \frac{h \cdot w \cdot c_1 \cdot c_2}{g} \tag{2}$$

对比公式(1)和公式(2)可以发现,通道内分组后再进行卷积可以有效降低逐点卷积的计算量,同时为了解决深度可分离卷积的各特征图通道之间信息沟通不畅的问题,检测模型引进了通道洗牌技术。如果分组的特征图尺寸为 $w \times h \times c_1$,其中 $c_1 = g \times n$,g 表示分组卷积过程中的分组数,进行通道洗牌的操作如下:(1)将特征图展开成 $g \times n \times w \times h$ 的四维矩阵,此处将 $w \times h$ 用 s 表示;(2)将 $g \times h \times s$ 的矩阵分别对 g 轴和 n 轴进行转置操作后,把得到的转置结果矩阵进行平铺,最后使用卷积核为 1 × 1 的组卷积操作,如图 3 所示,先将得到的特征图通道数目分为 9 个相同的通道数,并将得到的 9 个通道集合顺序打散,将其与对应卷积核进行卷积操作后,将得到的特征图恢复到开始之前

的张量结构。

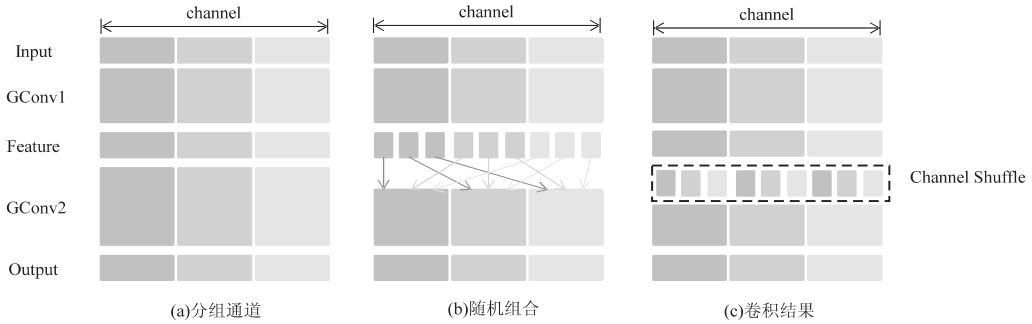
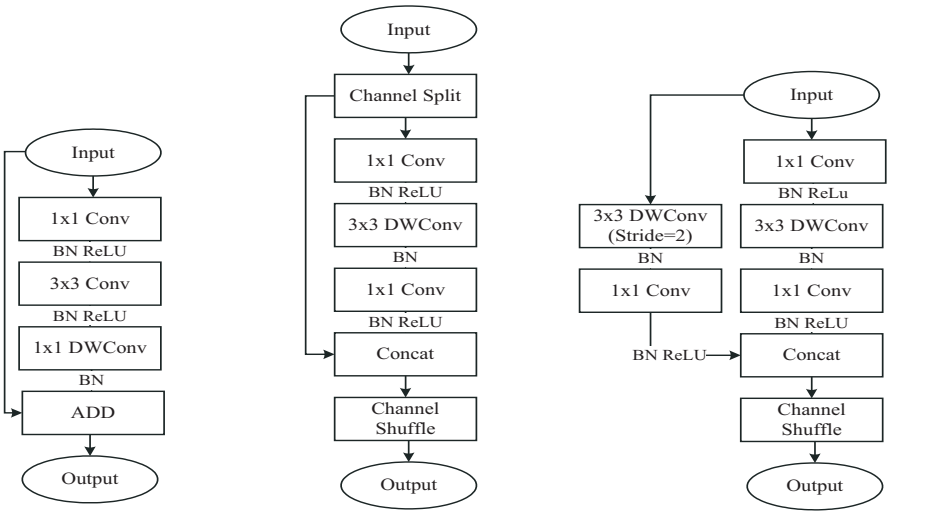


图 3　分组卷积和通道洗牌

2.3　残差单元改进

深度卷积神经网络如 ResNet[14] 和 DenseNet[15] 等类型的复杂网络模型推理速度较慢,不能满足实时检测需求,为了更好地在移动设备上运行,模型设计需要考虑模型的参数规模和移动设备的内存大小。Ma N 等学者结合 Shufflenet 和 Mobilenet 设计思路,提出了关于轻量级网络的设计观点[16],其中轻量级的神经网络应当符合如下设计准则:(1)使用轻量级网络模型中的深度可分割卷积(depthwise separable convolutions,DSV),在输入通道和输出通道采用相同通道大小的情况下可以最小化内存访问量;(2)过量使用组卷积会增加模型的内存访问量;(3)对于 Inception 类网络的“多路”结构,会导致网络结构的碎片化并降低网络模型并行度;(4)网络模型中的元素级操作虽然有较大的时间开销,但具有很大的作用,能提升特征的可代表性。根据这四条轻量级网络设计原则,文中对主干网络 DarkNet-53 中的残差单元进行修改,具体改进结构如图 4 所示。



(a)ResNet 残差单元　　(b)改进的残差单元结构　　(c)改进的下采样方式

图 4　改进结构

原 YOLOv3 模型中所采用的主干网络 DarkNet-53 的残差单元如图 4(a)所示,在该结构单元采用逐元素操作相加方式(Add)对两分支信息进行整合,这样的元素级操作会增加主干网络模型计算量[16]。如图 4(b)所示,改进后的残差结构将其更改为通道连接操作(Concat),同时为增加各通道的信息交流,在结构中加入通道洗牌操作,先将输入的网络特征图分为 c' 和 $c-c'$,为了符合 Ma N 提出的网络设计规则[16],一般 $c'=c/2$。其中图 4(c)左网络分支作为输入特征图的同等映射,对输出特征图进行复制,右分支对输入特征图连续进行 3 次卷积操作,令其输入输出通道数相等,对左右两分支进行通道连接操作(Concat),并进行通道洗牌以保证残差结构和特征图内各通道的信息交流,舍去原网络模型中使用步长为 2 的卷积下采样方式,改为图 4(c)所示的下采样方式,以此避免原特征图中的信息丢失,并对其进行特征筛选和深加工。

2.4　自适应空间特征融合

为了充分利用高层特征的语义信息和底层特征的细粒度特征,文中结合基于特征金字塔(feature pyramid networks for object detection,FPN)思想改进而来的自适应空间特征融合金字塔(adaptive spatial feature fusion pyramid,ASFF)[17]。ASFF 是一种特征混合的方法,可以在空间上学习其他尺寸特征图的特征信息,并保留有用的特征信息。对于待融合的特征

图后的特征图信息,网络使用卷积操作将其他尺度大小的特征图进行融合,此时通过使用上采样和 1 × 1 卷积进行通道变换,将尺寸需要调到相同的大小,然后进行加权,通过训练学习找到最好的参数组合。在每一个特征空间位置上,不同的特征会被自适应融合,如果有矛盾信息,通过训练可通过小权重参数将其过滤掉。ASFF 具备很多的优点,如实现成本低,几乎不增加模型推理时间,对一般的主干网络模型也具备一定的泛化能力,适用于类似 YOLOv3 等一系列具有特征金字塔结构的 One-Stage 目标检测器。针对 YOLOv3 中提取的 3 个不同尺寸的特征图,文中通过将三个不同尺度和权重的特征图进行结合,将特征金字塔结构进行修改,有效提升了对不同尺寸目标的检测精度,在一定程度上解决了模型的漏检问题。

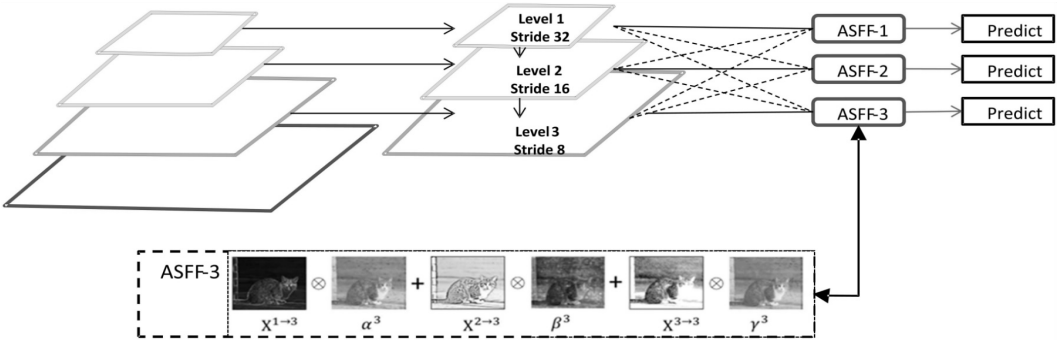


图 5 混合特征金字塔模型

针对来自不同层的特征进行融合,每一层有它对应的权重系数。为了确保在融合时不同层输出的特征和通道数是相同的,当特征图尺寸不相同的时候可以通过上采样或者下采样来进行调整,其中权重系数是由预定义尺寸的特征图经过 1 × 1 卷积得到的,其各权重矩阵中权重因子的累加和为 1,权重系数在[0,1]之间,特征融合的公式如(3)所示:

$$y_{ij}^{l} = \alpha_{ij}^{l} \cdot x_{ij}^{1\to l} + \beta_{ij}^{l} \cdot x_{ij}^{2\to l} + \gamma_{ij}^{l} \cdot x_{ij}^{3\to l} \tag{3}$$

如图 5 所示, $x_{ij}^{1\to l}$, $x_{ij}^{2\to l}$, $x_{ij}^{3\to l}$ 分别代表来自三个不同层的所提取的特征图,三个不同尺寸的特征图分别与对应权重系数 α_{ij}^{l}, β_{ij}^{l}, γ_{ij}^{l} 进行相乘,得到三个特征图的融合特征 y_{ij}^{l},即图 5 所示 ASFF-1、2、3。在训练过程中,为了保证每个层的主要信息的有效性,将每个层的初始权重预设为 0.5。

2.5 模型概述

基于 YOLOv3 进行改进,文中在主干网络基础上对网络残差单元和下采样方式进行优化,并将提取到的特征图结合 ASFF 模型进行混合特征提取,残差单元改进与组卷积、通道洗牌相结合是针对主干网络的改进,ASFF 是针对特征检测层的特征提取优化,整体目标检测框架是基于 YOLOv3 进行优化而来,如图 6 所示。其中 Stage2 模块由图 4(b)所示的改进残差单元组成,且下采样方法如图 4(c)模块所示,Stage2 部分是由改进的残差单元重复 4 次得到的,Stage3 由改进残差单元重复 8 次得到,同理 Stage4 也是由改进后的残差单元重复 8 次组合得到。由此,改进后的主干网络(restruct network,RN)结构如图 6 所示。

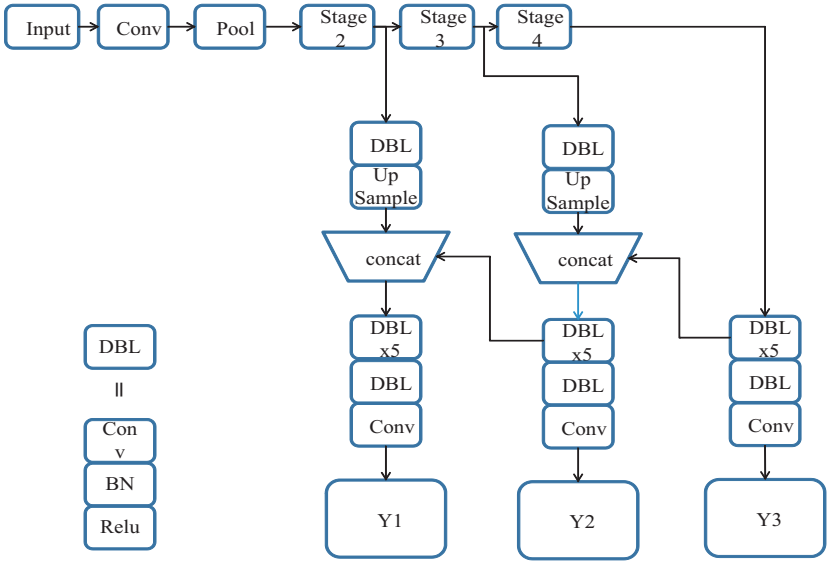


图 6 改进后的主干网络模型

设置目标检测的损失函数是为了让候选框坐标、置信度、分类损失三者之间达到平衡,如果简单地将各个损失相加,会存在以下问题:(1)不同维度的分类损失同等重要,简单将其相加会将二者视为同等重要,这种做法不够合理;(2)大目标物体的定位损失偏大,小目标物体的定位损失偏小,直接进行损失相加,会导致网络发散无法收敛。为缓解这些问题,将各类损失进行加系数和变形的方式进行改写,损失函数如公式(4)所示:

$$\begin{aligned}\mathrm{Loss}=&\lambda_{\mathrm{coord}}\sum_{i=0}^{N\times N}\sum_{j=0}^{K}I_{ij}^{\mathrm{obj}}\left[(t_x-t_x^{'})^2+(t_y-t_y^{'})^2\right]+\lambda_{\mathrm{coord}}\sum_{i=0}^{N\times N}\sum_{j=0}^{K}I_{ij}^{\mathrm{obj}}\left[(t_W-t_w^{'})^2+(t_h-t_h^{'})^2\right]+\\&\left\{-\sum_{i=0}^{N\times N}\sum_{j=0}^{K}I_{ij}^{\mathrm{obj}}\left[C_X^{'}\log(C_i)+(1-C_X^{'})\log(1-C_i)\right]-\right.\\&\left.\lambda_{\mathrm{noobj}}\sum_{i=0}^{N\times N}\sum_{j=0}^{K}I_{ij}^{\mathrm{obj}}\left[C_X^{'}\log(C_i)+(1-C_X^{'})\log(1-C_i)\right]\right\}+\\&\left\{-\sum_{i=0}^{N\times N}I_{ij}^{\mathrm{obj}}\sum_{C\in\mathrm{Class}}\left[p_i^{'}(c)\log(p_i(c))+(1-p_i^{'})\log(1-p_i(c))\right]\right\}\end{aligned}\tag{4}$$

其中,λ_{coord},λ_{noobj} 用于控制模型的分类损失和几何损失的均衡问题;i,j 分别用于标记第几个单元格和某单元格第几个候选框;$N\times N$ 表示单元格的总数;t_x,t_y,t_w,t_h 和 $t_x^{'}$,$t_y^{'}$,$t_w^{'}$,$t_h^{'}$ 分别为标签和侯选窗窗口的左上角坐标的宽度和高度;C 为预测框的置信度,$p(c)$ 为物体类别概率。文中继续使用 YOLOv3 模型中的损失计算思想:每个标记的候选框值(ground truth,GT)只对应一个预测框(predict bounding boxes,PBB),没有分配到的候选框的预测框只产生置信度损失,不产生几何损失和分类损失;I_{ij}^{obj} 表示第 i 个单元格点的第 j 个候选框,若存在预测框,将其返回 1,否则返回 0,而 I_{ij}^{noobj} 则相反,在第 i 个单元格中的第 j 个预测框,若分配有对应的标记候选框,则返回 1,否则返回 0,这样可有效避免模型因候选框较多而反向传递错误的梯度信息。

在训练开始阶段,实验对模型进行如下设计:(1)对没有目标的候选窗口的置信度损失赋予更小的损失权重,并记为 λ_{noobj},在数据集 PASCAL VOC 中取 0.5;(2)为了使模型更加重视有目标的单元格,并记为 λ_{coord} 将这些损失钱赋予更大的权重,在 PASCALVOC2007 和 2012 数据集上训练时设置为 5;(3)侯选窗的置信度和类别损失初始设置为 1,对不同大小侯选框进行预测,较大的侯选窗置信度预测存在偏大问题,为避免小候选窗预测值偏低的情形,将候选窗高宽取平方根代替原本的宽高值。

3　实　验

主干网络是以 DarketNet-53 为基础进行改进得到,整个检测模型的优化效果是通过 VOC2007 和 VOC2012 数据集进行验证和比较。在实验部分,通过设置相同的初始化变量和超参,使用不同的方法对整个模型结构进行测试,通过不同的实验结果对模型优化程度进行说明,并将改进后的目标检测架构运用于车辆和行人的检测过程,进行密集目标检测测试。实验环境:在 Windows10 系统环境下,使用 16 GB RTX2080ti 显卡进行测试,深度学习框架采用 Tensorflow-GPU1.4、CUDA10.2。

3.1　数据集

此次实验使用的数据集是作为基准数据集之一的 Pascal VOC2012 和 Pascal VOC2007,该数据集在目标检测、图像分割网络对比实验与模型效果评估中得到广泛应用。Pascal VOC 数据集主要是针对视觉任务中监督学习提供标签数据,共有二十个类别数据,主要分为四个大类别,如人、常见动物、交通车辆、室内家具用品等。VOC 数据集主要由 Annotation、ImageSets、JPEGImages、SegmentationClass 文件夹组成。Annotation 文件夹是 XML 文件,是对 JPEGImages 文件中每个图片的标注信息,一张图片对应一个 XML 文件;ImageSets 文件存放的是 txt 文件,这些文件将图片切分为各种集合;JPEGImages 文件夹存放该数据集所有的图片;SegmentationClass 文件夹适用于语义分割任务。文中主要使用 VOC 数据集进行目标检测,通过将改进后的网络模型与原模型效果进行对比,同时为了加大数据集容量,将 VOC2007 和 VOC2012 的数据集结合进行综合训练,并对训练后的网络模型进行评估。

3.2　实验结果

文中分别对修改后的主干网络(RestructNet)和混合特征金字塔模型(ASFF)进行测试,通过与 YOLOv3 的实验结果进行对比,分别在平均精度(MAP)、总体损失(Total Loss)和每秒传输帧数(FPS)指标上验证不同优化方法的检测效果和处理速度。实验结果如表 1 所示,修改后的主干网络和自适应特征混合模型组合的方法对于模型检测精度有明显优化效果。由实验结果可知,对原残差单元的修改和主干网络的重构对检测精度有明显的提升效果,在精度方面提升 4.36%,其中混合空间特征金字塔模型的使用在精度方面提升 3.0%,综合精度提升 8.31%,在不影响检测

速度的情况下,实现检测模型精度方面的优化。

表 1 实验结果对比

Model	Val mAP/%	FPS(2080ti)	Total Loss	Dataset
YOLOv3	51.97	41	15.49	Train(VOC2007+2012)、Val(VOC2007)
YOLOv3(Restruct)	55.61	38	13.28	Train(VOC2007+2012)、Val(VOC2007)
YOLOv3+ASFF	53.97	35	14.23	Train(VOC2007+2012)、Val(VOC2007)
YOLOv3(Restruct)+ASFF	60.28	45	10.21	Train(VOC2007+2012)、Val(VOC2007)

为了与主流 One-Stage 目标检测模型进行对比,文中在相同实验环境和训练参数下,与最新 YOLOv4 和其他 YOLO 系列目标检测算法进行对比,分别比较推理速度(FPS)、在 VOC2007+2012 数据集上的测试精度(MAP)、网络参数数量(Parameter)等指标,进一步说明该网络模型的优化效果。通过实验可知,结合残差网络的更改和混合特征融合金字塔优化后的网络结构,大体可以达到和 YOLOv4 的精度,且模型推理速度及网络参数数量较 YOLOv4 减少 1.21 MB,整体精度较 YOLOv3 提高 8.17%。

表 2 与目前流行的 One-Stage 算法对比

Model	Backbone	InputShape	Val mAP/%	Parameter/M	FPS(2080ti)/(f/s)
YOLOv4	CSPDarknet	512	61.12	46.34	67
YOLOv4	CSPDarknet	416	61.26	46.34	65
YOLOv4	CSPDarknet	320	58.19	46.34	61
YOLOv3	MobileNetv2	608	55.76	37.1	53
YOLOv3	MobileNetv2	512	55.35	37.1	65
YOLOv3	MobileNetv2	416	54.97	37.1	72
YOLOv3	CSPDarknet	320	52.11	48.64	58
YOLOv3	CSPDarknet	416	52.21	48.64	51
YOLOv3	CSPDarknet	608	53.62	48.64	48
Restruct (Ours)	RestructNet	416	60.28	45.13	43
Restruct (Ours)	RestructNet	512	60.31	45.13	41
Restruct (Ours)	RestructNet	608	61.86	45.13	38

3.3 实验结果分析

在 YOLOv3 模型中,将特征图直接输入主干网络结构的残差单元,用分支结构将卷积处理的特征图与原特征图进行相加,这样虽避免梯度爆炸和梯度消失问题,但新生成的特征图包含了许多不必要的背景信息。通过使用图 4(b)所示的残差单元进行改进,同时对两个网络分支进行信息处理,由实验结果可知,检测模型精度提升了 8.07%。

由 Hurtik P 提出[18],原 YOLOv3 网络模型对于密集目标存在漏选和标签重写问题[1],为了解决该问题,实验基于数据驱动的金字塔特征融合方式,该方法通过学习在空间上过滤冲突信息以抑制梯度反传时的不一致性,以此增加待检目标尺度的容纳性,同时降低推理开销。通过使用统计的日志文件绘制训练损失图,如表 2 所示,使用 ASFF 的模型在测试集上的验证损失明显低于原 YOLOv3 模型的训练损失,且比原模型具有更快的收敛速度,借助 ASFF 和组卷积优化残差单元组合的方式,在 VOC 数据集上实现了 60.28% 的平均精度以及 43 FPS 的运算速度。为了更好地进行对比,训练均未采用预训练的主干网络权重,因为改进的网络结构没有预训练权重可供参考。

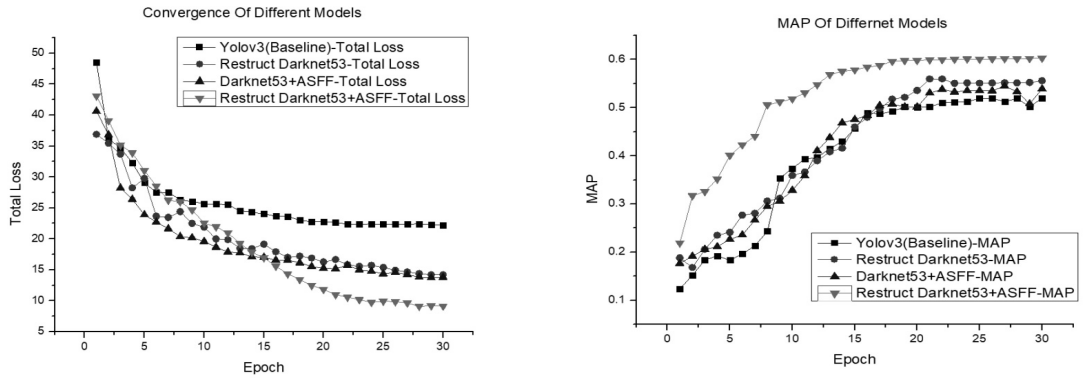


图 7 模型训练收敛及精度提升变化

通过对残差单元、下采样方式进行修改,然后和 ASFF 进行组合得到新的检测模型,通过实验进行综合测试,新的检测模型可以在 VOC2007 和 VOC2012 得到 60.28% 的检测精度。由图 7 对比所示,为了发挥 One-Satge 目标检测模型实时检测的优势,在不影响改进模型精度的条件下,主干网络(beckbone network,BN)和检测网络(neck network,NN)对不必要的卷积层进行删除,可明显看出新模型较原 YOLOv3 模型,关于平均精度的提升和训练的快速收敛情况,在每张图(416 × 416)的检测时间仍然可达 43 ms 帧率,与原 YOLOv3 模型相比,仍然可以达到实时性的检测效果。

3.4　模型运用

由实验证明,通过引入组卷积和通道洗牌技术对残差单元进行修改,和自适应空间特征混合(ASFF)组合的方法,可以取得明显的优化效果。为了验证在实际环境中改进后的目标检测模型的检测效果,文中将数据集 COCO2017 内的关于无人驾驶相关的检测类别进行分离,对分离的数据集使用 K-means 方法产生锚框并进行矛框大小设置。实验对分离后的数据集使用 K-means 方法进行锚框大小选取,通过聚类分出各个锚框的类别,然后分别对各个类别的锚框宽高取均值,得到目标的候选锚框大小分别为(10,13),(16,30),(33,23),(30,61),(62,45),(59,119),(116,90),(156,198),(373,326)。改进后的模型在分离的数据集上进行训练得到的检测效果如图 8、图 9 所示,训练后的模型在密集车流下进行测试,改进后的目标检测模型可以取得理想的检测效果且能够达到实时的检测速度,原 YOLOv3 模型存在的密集目标漏选和标签重写现象[18]也得到改善。

图 8　YOLOv3 测试效果

图 9　改进模型的测试效果

4　结束语

文中以 YOLOv3 为基础进行改进,得到一种单阶段实时目标检测模型,旨在针对无人驾驶、安全监控等领域进行目标检测和识别。首先,引入组卷积和通道洗牌技术,并对原 Darknet-53 网络的残差结构进行优化改写,为了更多地保留特征图的有效信息,使用了全新的下采样方式对特征图进行尺寸缩减;其次,为了克服原 YOLOv3 检测模型对密集目标存在的漏选和标签重写问题,使用自适应特征混合金字塔对输出的特征图进行空间特征混合处理,加强不同尺寸的检测特征图之间的信息交流,以此加强对密集目标的检测能力;最后,使用 PASCAL VOC2007 和 VOC2012 进行测试,改进后的目标检测模型相较于 YOLOv3 提升了 8.17%,取得了和 YOLOv4 大体相同的精度,并且可以达到实时的检测速度。通过实验进行测试,该模型可以有效地运用于交通监测和交通目标识别应用中,具有很强的应用性。

参考文献:

[1] REDMON J, FARHADI A. YOLOv3:an incremental improvement[J]. arXiv:1804.02767,2018.

[2] MOORE D,RID T. Cryptopolitik and the Darknet[J]. Survival,2016,58(1):7-38.

[3] EVERINGHAM M, ESLAMI S M A, VAN GOOL L, et al. The pascal visual object classes challenge: a retrospective [J]. International Journal of Computer Vision, 2015, 111 (1): 98-136.
[4] NG P C, HENIKOFF S. SIFT: predicting amino acid changes that affect protein function [J]. Nucleic Acids Research, 2003, 31(13): 3812-3814.
[5] WANG X. An HOG-LBP human detector with partial occlusion handling[C]//Proc. IEEE international conference on computer vision. Kyoto, Japan: IEEE, 2009.
[6] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
[7] 张海涛, 张 梦. 引入通道注意力机制的 SSD 目标检测算法[J]. 计算机工程, 2020, 46(8): 264-270.
[8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//Computer vision - ECCV 2016. Amsterdam, The Netherlands: Springer, 2016: 21-37.
[9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [s. l.]: IEEE, 2016: 779-788.
[10] 顾 恭, 徐旭东. 改进 YOLOv3 的车辆实时检测与信息识别技术[J]. 计算机工程与应用, 2020, 56(22): 173-184.
[11] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[J]. arXiv: 2004.10934, 2020.
[12] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[J]. arXiv: 1602.07261, 2016.
[13] COHEN T, WELLING M. Group equivariant convolutional networks [C]//International conference on machine learning. Anaheim, California, USA: PMLR, 2016: 2990-2999.
[14] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: an extremely efficient convolutional neural network for mobile devices [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City, UT, USA: IEEE, 2018: 6848-6856.
[15] IANDOLA F, MOSKEWICZ M, KARAYEV S, et al. Densenet: implementing efficient convnet descriptor pyramids [J]. arXiv: 1404.1869, 2014.
[16] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: practical guidelines for efficient CNN architecture design [C]// Proceedings of the European conference on computer vision (ECCV). Munich, Germany: [s. n.], 2018: 116-131.
[17] LIU S, HUANG D, WANG Y. Learning spatial fusion for single-shot object detection[J]. arXiv: 1911.09516, 2019.
[18] HURTIK P, MOLEK V, HULA J, et al. Poly-YOLO: higher speed, more precise detection and instance segmentation for YOLOv3[J]. arXiv: 2005.13243, 2020.

(上接第 85 页)

[9] SHI Guiming, SUO Jidong. Remote sensing image edge-detection based on improved Canny operator [C]//2016 8th IEEE international conference on communication software and networks (ICCSN). Beijing: IEEE, 2016.
[10] JIN J, FU L. Edge detection based on Canny-oscillation algorithm[C]//2nd international conference on automatic control and information engineering (ICACIE). Hong Kong, China: [s. n.], 2017.
[11] DENG C X, WANG G B, YANG X R. Image edge detection algorithm based on improved Canny operator[C]//International conference on intelligent systems design & applications. Rio de Janeiro, Brazi: IEEE, 2007.
[12] COPE R K, ROCKETT P I. Efficacy of Gaussian smoothing in Canny edge detector [J]. Electronics Letters, 2000, 36 (19): 1615-1617.
[13] SMITH T G, MARKS W B, LANGE G D, et al. Edge detection in images using Marr-Hildreth filtering techniques[J]. Journal of Neuroence Methods, 1988, 26(1): 75-81.
[14] SHI T, KONG J Y, WANG X D, et al. Improved Sobel algorithm for defect detection of rail surfaces with enhanced efficiency and accuracy[J]. Journal of Central South University, 2016, 23(11): 2867-2875.
[15] 李忠海, 宋智钦, 王崇瑶. 非整数步长的分数阶微分 Sobel 算子的应用[J]. 计算机工程与应用, 2018, 54(17): 192-197.
[16] WANG Y, CHEN Y, WANG M. A new vehicle license plate correction method based on sobel operator and priori knowledge[C]//2016 6th international conference on machinery, materials, environment, biotechnology and computer (MMEBC 2016). Tianjin, China: [s. n.], 2016.
[17] 李绍丽, 苑玮琦, 李德健. 基于并查集和约束集合的雪糕棒表面污染检测[J]. 计算机应用研究, 2018, 35(8): 2527-2531.