文章编号：2095-2163(2022)01-0164-04 中图分类号：TP391.4 文献标志码：A

基于 YOLOv4 目标检测算法的轻量化网络设计

胡 亮，何小海，卿粼波，吴小强
（四川大学 电子信息学院，成都 610065）

摘 要：目标检测是计算机视觉领域的一个重要研究方向，在交通监控、人机交互等方面都有着广泛的应用。目前，基于深度学习的 YOLOv4 检测网络与传统目标检测相比，其检测精度虽然有所提高，但存在网络参数量大、对计算机硬件要求较高等问题。针对于此，本文对 YOLOv4 网络进行了改进，即采用 MobileNetv2 与 YOLOv4 的主干特征提取网络相结合，并利用深度可分离卷积模块，对 YOLOv4 的 PANet 和 SPP 模块中的传统卷积进行了优化，在公开数据集 VOC07+12 上进行训练，并将训练后的模型在 VOC07test 数据集上进行分析、检测。实验结果表明，改进后的 YOLOv4 卷积神经网络相比于 YOLOv4 神经网络参数量降低了 83.6%，*FPS* 提升了 5.8，*mAP*@0.5 下降了 8.5%，降低了网络对计算机硬件的要求，实现了网络模型的轻量化。

关键词：目标检测；YOLOv4；MobileNet；轻量化神经网络

Design of lightweight network based on YOLOv4 object detection algorithm

HU Liang, HE Xiaohai, QING Linbo, WU Xiaoqiang
(College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China)

【Abstract】 Object detection is an important research direction in the field of computer vision. It has a wide range of applications in traffic monitoring and human-computer interaction. At present, the YOLOv4 network based on deep learning has improved detection accuracy compared with traditional object detection, but it has a lot of parameters, which has a large amount of requirement of computer hardware. In order to reduce the number of parameters and the size of network model, this paper modify the YOLOv4 network, that is, using MobileNetv2 network combine with the feature extraction network of YOLOv4, and using the block of depthwise separable convolution optimizes traditional convolution in PANet and SPP model. Finally, the proposed model is trained on the VOC07+12 datasets and tested on the VOC07test datasets. Experiment results show that the parameters of the proposed convolutional neural network based on YOLOv4 are reduced by 83.6% compared with YOLOv4 neural network, and FPS is increased by 5.8 and the mAP@0.5 is decreased by 8.5%, which reduce the need for computer hardware and achieve the lightweight of the network model.

【Key words】 object detection; YOLOv4; MobileNet; lightweight neural network

0 引 言

在计算机视觉领域中，目标检测一直以来都是其中最重要的部分之一，其对应的算法也在不断推陈出新，目标检测技术已被广泛地应用于医疗、公共、军事等领域。当前，目标检测领域已经从传统的目标检测算法转向基于深度学习的目标检测算法。基于深度学习的目标检测算法根据其有无候选框生成，分为一阶段目标检测算法和二阶段目标检测算法。一阶段目标算法主要包括 SDD[1]、YOLO[2-3]等，其不需要候选框，将定位和分类合成一步完成，这样做虽然增加了学习的难度，但是提升了算法的速度，同时也减少了占用空间。二阶段目标检测算法主要包括 Rcnn[4]、Fast-Rcnn[5]、Faster-Rcnn[6]等，由于比一阶段目标检测算法多了前景与背景的分类和检测，所以步骤相对复杂，占用的空间相对较多。

YOLO 系列算法作为一阶段目标检测算法的主要代表之一，被广泛应用于各类目标识别场景。YOLOv4[7]算法作为 YOLO 系列的代表作，虽然在精度和速度上被推向了一个新的高度，但随着网络复杂度和层数的不断增加，其参数量不断增加、模型大小不断增大，加大了对计算机硬件的需求，无法在诸如嵌入式平台等低功耗平台上实现应用，因此需对模型进行轻量化处理[8]。针对 YOLOv4 参数量较多模型较大的问题，人们提出了不同的解决方法。例如，在特征提取层 CSPDarkNet53 中按照一定比例减少 3 个输出层中的残差结构；利用 Cross-Stage Lightweight (CSL)[9] 模块替换特征提取层

作者简介：胡 亮(1996-)，男，硕士研究生，主要研究方向：多媒体通信与信息系统；何小海(1964-)，男，博士，教授，主要研究方向：图像处理与信息系统、机器视觉与智能系统；卿粼波(1982-)，男，博士，副教授，主要研究方向：信号与信号系统、图像处理、图像通信；吴小强(1969-)，男，学士，高级工程师，主要研究方向：计算机应用、图像处理。
通讯作者：何小海 Email：hxh@scu.edu.cn
收稿日期：2021-09-28

CSPDarkNet53 中的普通卷积模块，从而得到新的 YOLO－CSL 模型；利用 GhostNet[10] 模块代替 YOLOv4 算法中传统的下采样操作，从而解决 YOLOv4 中下采样操作成本较高的问题。

本文在 YOLOv4 网络的基础上进行了轻量化处理。利用 MobileNet 轻量化网络的思想，将其融入到 YOLOv4 的特征提取层 CSPDarkNet53 和特征金字塔 PANet 以及 SPP 模块中，在适当地牺牲掉较少的精准度的情况下，大幅度地降低了网络的参数量及其模型占用空间。

1　网络结构设计

1.1　YOLOv4 网络

YOLOv4 可看作 YOLOv3 的加强版，其总体框架仍基于 YOLOv3。对于主干特征提取网络 backbone，YOLOv4 选择更优的 CSPDarknet53 网络，替换了原来的 Darknet53 网络；对于目标检测器的 neck，即 YOLOv3 中的特征金字塔，将其从原来的 FPN 修改为 PANet[11]，并新增了 SPP 结构，用于提升感受野，以较小计算量的增加换取较大准确率的提升；对获取到的特征进行预测 head 部分，仍沿用 YOLOv3 中的 YOLO 模块。最终网络结构如图 1 所示。

除在以上模块中的大改动之外，为了让其更适合在单个 GPU 上进行训练，YOLOv4 对此做出了专门的创新。例如，引用新的数据增强方法：Mosaic 和 Self-Adversarial Training[12]；使用遗传算法来选择最优的超参数；改进一些现有的算法，使其更适合高效的训练与检测。如：改进 SAM、PAN 以及 Cross mini-Batch Normalization(CmBN)。

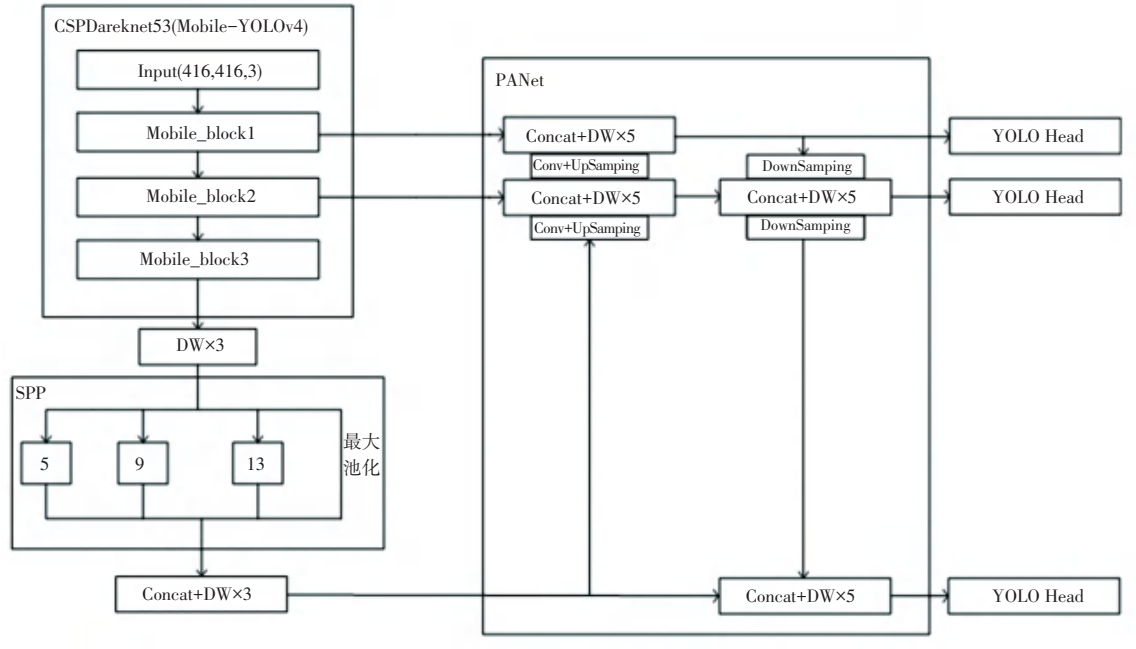

图 1　YOLOv4 结构图

Fig. 1　Structure of YOLOv4

1.2　MobileNet 网络

MobileNet 是基于深度可分离卷积构造的网络，其核心思想是将标准卷积拆分为两个部分完成，即深度卷积和逐点卷积。对于标准卷积，其所有卷积核作用到所有的输入通道上进行卷积操作，实现一步完成。而对于 MobileNet 而言，其将先进行深度卷积操作，针对输入的每个通道采用不同的卷积核，即一个卷积核对应一个输入通道，再采用逐点卷积，利用 1×1 的卷积操作来结合所有深度卷积得到输出。

假设　输入尺寸为 $D_F \times D_F \times M$，标准卷积核的尺寸为 $D_K \times D_K \times M \times N$。若采用标准卷积核进行计算，步长为 1 且 padding，则计算量为 $D_K \times D_K \times M \times N \times D_F \times D_F$。若采用深度可分离卷积进行计算，总计算量为 $D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F$。将深度可分离卷积的总计算量和标准卷积的总计算量相比可得：

$$\frac{D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F}{D_K \times D_K \times M \times N \times D_F \times D_F} = \frac{1}{N} + \frac{1}{D_K^2} \tag{1}$$

通常情况下 N 取值较大，假设采用 3×3 的卷积核，通过式(1)可计算得到采用深度可分离卷积相比标准卷积可降低大约 9 倍的计算量。

MobileNetv2[13] 总体思想仍基于深度可分离卷积，核心由 Bottleneck Residual block 模块组成，其结构如图 2 所示。

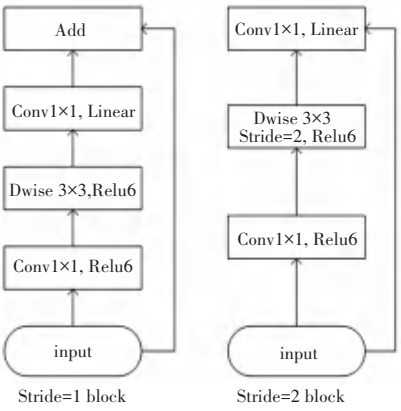

图 2 Bottleneck Residual block

Fig. 2 Bottleneck Residual block

Bottleneck Residual block 模块引入了残差结构,增强了梯度的传播;去掉了最后输出时的 ReLU,保留了特征多样性,增强了网络的表达能力;先对输入进行升维,有助于提取到整体的足够多的信息。MobileNetv2 的整体网络结构参数设置见表 1。

表 1 MobileNetv2 参数设置

Tab. 1 MobileNetv2 parameter settings

Input	Operator	t	c	n	s
$224^2\times3$	conv2d	–	32	1	2
$112^2\times32$	bottleneck	1	16	1	1
$112^2\times16$	bottleneck	6	24	2	2
$56^2\times24$	bottleneck	6	32	3	2
$28^2\times32$	bottleneck	6	64	4	2
$14^2\times64$	bottleneck	6	96	3	1
$14^2\times96$	bottleneck	6	160	3	2
$7^2\times160$	bottleneck	6	320	1	1
$7^2\times320$	conv2d 1×1	–	1 280	1	1
$7^2\times1$ 280	avgpool 7×7	–	–	1	–
1×1×1 280	conv2d 1×1	–	k	–	

1.3 整体网络结构

本文在 YOLOv4 网络基础上,对网络结构进行调整。由表 1 中数据可以看出,当输入图像尺寸为 416×416×3 时,第 3、5、7 层 bottleneck 的输出分别为 52× 52 × 32、26 × 26 × 96、13 × 13 × 320。所以,以 MobileNetv2 的第 4、6、8 层 bottleneck 为界,将其划分为 3 个模块,并将这 3 个模块代替 YOLOv4 主干特征提取网络中的 3 个有效特征层,从而实现 MobileNetv2 和 YOLOv4 的结合,替换后的模型为 Mobile-YOLOv4。

将修改后的 Mobile-YOLOv4 网络与原 YOLOv4 网络作比较,其参数量从 64,040,001 下降到了 39,062,013,参数量仅下降了 39%,参数量和网络大小下降程度并不理想。通过观察图 1 可以发现,除主干特征提取层,YOLOv4 特征金字塔 PANet 中存在着大量的标准卷积以及上采样和下采样的操作,这些卷积操作中包含的参数量极大,其存在的可压缩空间也较大。可利用深度可分离卷积的思想,将其中 3×3 卷积块中的标准卷积替换成深度可分离卷积,从而实现参数量的下降。在 SPP 模块的输入和输出端有两个 3×3 的卷积块,可利用上述相同的方法,替换其中的标准卷积,进一步压缩模型的参数量和大小。最终轻量化后的网络结构如图 3 所示。其中 Moble_block1、Moble_block2、Moble_block3 分别对应表 1MobileNetv2 结构中的 1~4 层、5~6 层、7~8 层,DW 为深度可分离卷积模块。

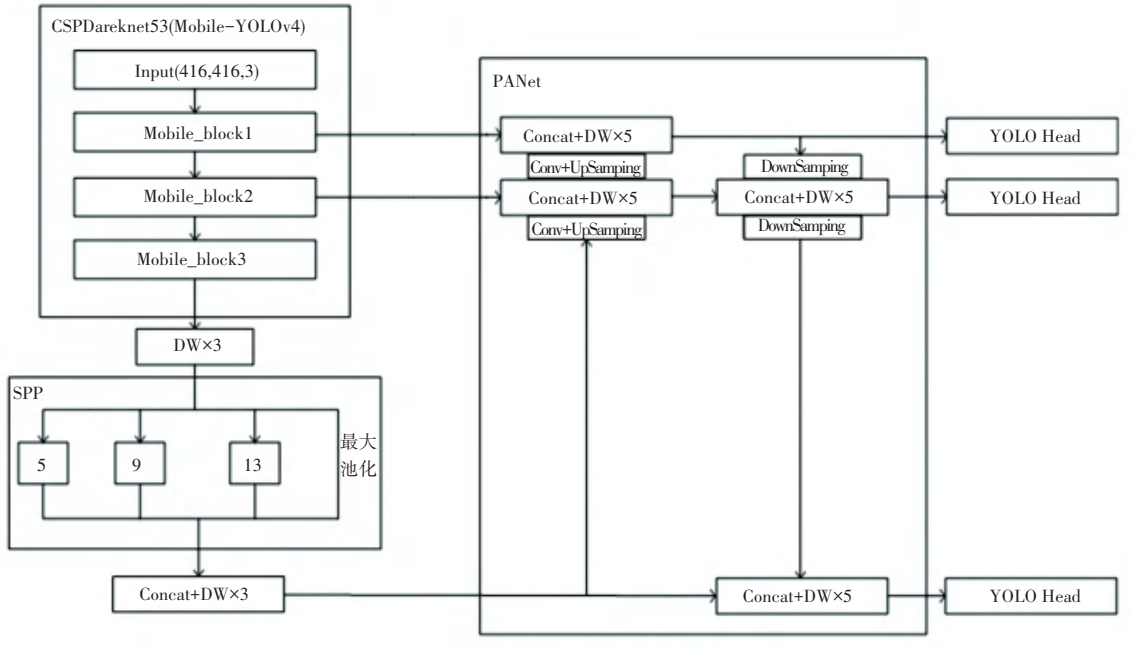

图 3 轻量化后结构图

Fig. 3 Structure of lightweight YOLOv4

2　训练及测试结果

2.1　网络训练

本文在服务器上采用 GPU 模式进行网络训练，其中训练平台配置为：Intel（R） Core（TM） i7-8700 3.2 GHz 处理器；显卡为显存 12 GB 的 NVIDIA 2080Ti SLI；Ubuntu 18.04 64 位操作系统；深度学习框架为 Pytorch。

数据集使用 VOC07+12 数据集，该数据集包含 20 个类，共有 16 551 张图片，每张图片在传入网络时，resize 大小为 608×608。按照 9∶1 的比例将数据集分为训练集和验证集。如图 4 所示，当训练约 15 个 epoch 后，网络收敛趋于平稳。

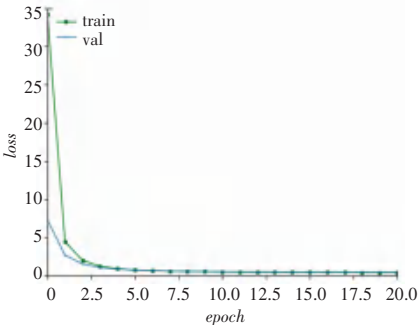

图 4　Model loss

Fig. 4　Model loss

2.2　测试结果

本文使用 VOC2007test 数据集进行测试，将修改后的模型与原 YOLOv4 模型以及 YOLO-CSL 进行对比。结果见表 2。

表 2　测试结果

Tab. 2　Test results

	parameters	*mAP*@0.5	*FPS*
YOLOv4	64,040,001	88.9	22.1
YOLO-CSL	33,594,755	85.3	24.9
Mobile-YOLOv4	39,062,013	85.5	24.2
Proposed	10,478,049	80.4	27.9

其中，Mobile-YOLOv4 为仅替换主干特征提取网络 backbone 对应的模型，Proposed 为本文所修改优化后的模型。由表 2 可看出，在仅替换 backbone 的情况下，参数量和模型大小下降并不明显，无法达到理想效果，在对 PANet 和 SPP 模块进行优化过后，参数量相较于原 YOLOv4 网络降低了 83.6%，下降程度明显，同时 *FPS* 也提升了 5.8，*mAP*@0.5 仅下降了 8.5%，其准确率足够应用于某些特定场景。同时，参数量模型大小的降低以及 *FPS* 的提升，可降低对计算机硬件的要求。

3　结束语

本文首先将 MobileNetv2 和 YOLOv4 的网络结构对比后进行融合，得到全新的 Mobile-YOLOv4 网络，但是其实际压缩效果并不理想，参数量下降不明显。通过观察 YOLOv4 的特征提取金字塔模块和 SPP 模块，得知其中参数量较大，所以利用 MobileNet 中深度可分离卷积的思想，将其中的标准卷积模块替换成深度可分离卷积模块，从而更进一步使模型得到压缩。最终本文利用上述方法改进后的网络，在牺牲可接受的准确度前提下，使得网络参数量和大小降低了 83.6%，提升了 FPS，降低了模型训练的时间及预测时间，且降低了对于计算机硬件的要求。

参考文献

[1] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of European Conference on Computer Vision. Springer, 2016: 21-37.

[2] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only look once: unified, real time object detection[C]//Computer Vision and Pattern Recognition. 2017: 6517-6525.

[3] REDMON J, FARHADI A. YOLOv3: An Incremental Improvement[J]. arXiv Preprint arXiv: 1804.02767,2018

[4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]//IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2014: 580-587.

[5] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 1440-1448.

[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] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection[C]//IEEE conference on Computer Vision and Pattern Recognition. 2020. arXiv: 2004.10934-v1[cs.CV]

[8] 王伟，何姣，石强. 复杂背景下目标识别算法分析与改进[J]. 智能计算机与应用，2020，10(4)：253-257.

[9] ZHANG Y M, LEE C C, HSIEH J W, and Fan, K.-C., CSL-YOLO: A New Lightweight Object Detection System for Edge Computing [C]//IEEE conference on Computer Vision and Pattern Recognition, 2021,arXiv:2017.04829.

[10] HAN K, WANG Y, TIAN Q, et al. GhostNet: More features from cheap operations [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580-1589.

[11] LIU S, QI L, QIN H, et al.Path aggregation network for instance segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018 :8759-8768.

[12] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-attention generative adversarial networks [C] //International conference on machine learning. PMLR,2019:7354-7363.

[13] Mark Sandler,Andrew Howard,Menglong Zhu, Andrey Zhmoginov. MobileNetV2: Inverted Residuals and Linear Bottlenecks[EB/OL].arXiv-eprints,2018.https://arxiv.org/abs/180 1.04381.