基于 YOLO 改进残差网络结构的车辆检测方法

胡臣辰，陈贤富

（中国科学技术大学 微电子学院，安徽 合肥 230027）

摘 要： 针对车辆检测任务，设计更高效、精确的网络模型是行业研究的热点，深层网络模型具有比浅层网络模型更好的特征提取能力，但构建深层网络模型时将导致梯度消失、模型过拟合等问题，应用残差网络结构可以缓解此类问题。基于 YOLO 算法，改进残差网络结构，加深网络层数，设计了一种含有 68 个卷积层的卷积神经网络模型，同时对输入图像进行预处理，保证目标在图像上不变形失真，最后在自定义的车辆数据集上对模型进行训练与测试，并将实验结果与 YOLOV3 模型进行对比，实验表明，本文设计的模型检测精准度（AP）达 90.63%，较 YOLOV3 提高了 4.6%。

关键词： 目标检测；YOLO；残差网络；深度学习

中图分类号： TP393　　**文献标识码：** A　　**DOI：** 10.19358/j.issn.2096-5133.2020.09.011

引用格式： 胡臣辰，陈贤富. 基于 YOLO 改进残差网络结构的车辆检测方法[J].信息技术与网络安全，2020，39（9）：56-60.

Vehicle detection method based on improved residual network structure of based on YOLO

Hu Chenchen，Chen Xianfu

（Institute of Microelectronics，University of Science and Technology of China，Hefei 230027，China）

Abstract： For vehicle detection task，the design of more efficient and accurate network model is a hot research. Deep network model has better feature extraction capabilities than shallow network model，but will cause the gradient to disappear and the model to overfit and other problems. Application of residual network structure can alleviate such problems. Based on the YOLO algorithm，this paper improves the residual network structure and deepens the number of network layers. A convolutional neural network model with 68 convolutional layers is designed. At the same time，the input image is preprocessed to ensure that the target is not deformed or distorted on the image. Finally，the model is trained and tested on a custom vehicle data set，and the experimental results are compared with the YOLOV3 model. The experiment shows that the model detection accuracy（AP）designed in this paper reaches 90.63%，which is 4.6% higher than YOLOV3.

Key words： object detection；YOLO；residual network；deep learning

0 引言

车辆是目标检测任务中的重要对象之一，在自动驾驶、目标追踪等领域有着十分重要的应用。以梯度方向直方图（Histogram of Oriented Gradient，HOG）和支持向量机（Support Vector Machine，SVM）结合的传统目标检测算法先计算候选框内图像梯度的方向信息统计值，再通过正负样本训练 SVM，使用传统方法受限于候选框提取效率、HOG 特征尺度鲁棒性，在实时性以及遮挡目标检测等诸多方面有着明显缺陷[1]。近年来，基于深度学习的目标检测方法以强特征提取能力、高检测率取得了惊人的成果。近年来深度学习网络在计算机视觉上因 AlexNet 在 2012 年的 ImageNet 大赛中大放异彩而进入飞速发展。2014 年 VGGNet 在追求深层网络的性能时，发现增加网络的深度会提高性能，但是与此同时带来的梯度消失问题不可避免。2015 年 ResNet 网络较好地解决了这个问题，深层残差网络可以减少模型收敛时间、改善寻优过程，但应用尺度大的卷积核的同时增加了网络模型的参数量与计算量，降低了模型的训练与检测速度[2]。

计算机视觉中的目标检测任务关注图像中特定目标的位置信息，现有方法分为 two-stage 和 one-stage 两类。two-stage 方法先产生包含目标的候选框，再通过卷积神经网络对目标进行分类，常见的方法有 RCNN、Fast-RCNN、Faster-RCNN。one-stage 方法直接使用一个卷积网络对给定输入图像给出检测结果，以 YOLO 为代表的 one-stage 目标检测方法在检测时，将候选框的生成与目标的分类回归合并成一步，基于 YOLO 的检测算法大大提高了检测速度，但检测精度仍有待提高[3]。本文选择在基于 YOLO 方法的基础上改进主干网络的残差网络结构，设计了一种新的网络模型，经实验表明提高了检测准确率。

1 相关工作

1.1 残差网络结构

2015 年何凯明等人在论文中正式提出了 ResNet[4]，其针对常规网络在进行加深深度时出现的梯度消失、浅层网络无法提高网络的识别能力等问题取得了重大突破。残差结构的表达式为：

$$y=F(x,\{W_i\})+W_s x \tag{1}$$

若不使用残差结构，当输入为 x 时，经网络后输出为 $F(x,\{W_i\})$，而引入残差结构后，需要学习的特征为 $F(x,\{W_i\})-W_s x$，使用残差结构的网络对于微小变化更显突出。

图 1 所示的残差网络结构中，先使用 1×1 的卷积核缩小通道数，最后通过 1×1 的卷积核恢复通道，使用 1×1 与 3×3 组合结构可以减少计算量与参数。

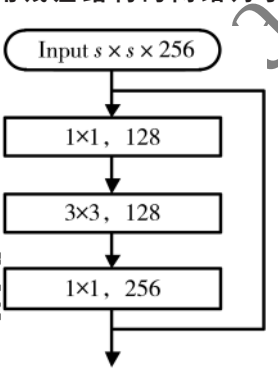

图 1 三层残差学习网络结构

1.2 YOLO 系列算法

2016 年 REDMOND J 提出 YOLO(You Only Look Once)算法[5]，将特征图分为 $S\times S$ 的网格，每个网格负责预测中心点落在网格中的目标，YOLO 将生成候选框与分类预测合并成一步，加快了检测速度。

2018 年 REDMOD J 发布了 YOLO 的第三个版本—YOLOV3[6]，YOLOV3 算法采用了不含全连接层的 darknet53 网络作为主干网络，darknet53 借鉴了 ResNet 结构，复用残差结构(Res-block)，加强主干网络提取特征的能力。

图 2 所示的 Res-block 是组成 darknet53 的重要部分，它由一个下采样的卷积块(Con-block)与若干残差单元(Res-unit)组成，其中 Con-block 由卷积层、归一化层、激活函数层构建，Res-unit 由包含两个 1×1 与 3×3 的卷积核的 Con-block 组成。

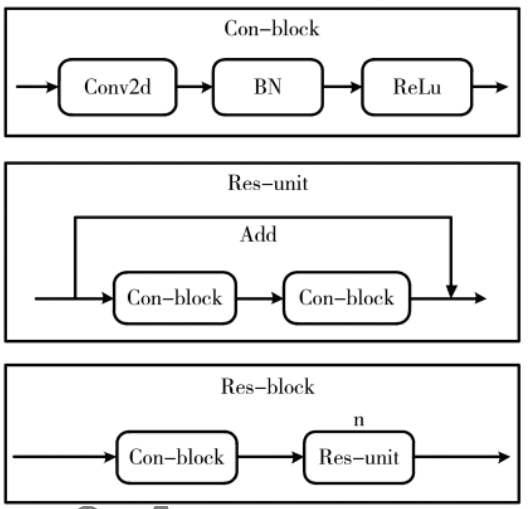

图 2 组成 darknet53 网络的基本组成单元

YOLOV3 在预测时，使用特征金字塔网络(Feature Pyramid Networks，FPN)选取主干网络上的三个尺度进行预测，有效增强不同大小的目标检测效果，同时将高层特征上采样堆叠到低层特征上，加强低层特征的语义信息。图 3 所示的 FPN 结构有三个尺度的输出，格式均为 $S\times S\times(3\times C)$，其中 S 为特征图的网格大小，3 代表每个网格负责预测 3 个候选框，5 代表每个候选框的位置信息(相对网格左上角偏移值及候选框的宽高)及包含物体的置信度，C 表示候选框所要预测目标的置信度。本文检测单一的车辆目标，C 的数值为 1。

YOLOV4 算法[7]的主干网络上采用了 CSPdarknet53，它借鉴 CSPNet 网络结构[8]，重复使用含有三个卷积块及 n 个 Res-unit 单元组合而成的网络结构(Csp-block)，在激活函数的选择上，使用 Mish 函数替换 LeakyReLu 函数。

图 4 所示的卷积块使用的激活函数为 Mish 函数：

$$\text{Mish}=x*\tanh(\ln(1+e^x)) \tag{2}$$

YOLOV4 选用 Mish 激活函数针对 ReLu 函数在负值时直接截断，梯度下降不够平滑进行了优化。

2 算法改进

2.1 模型改进

YOLOV4 在 YOLOV3 的结构基础上使用诸多了

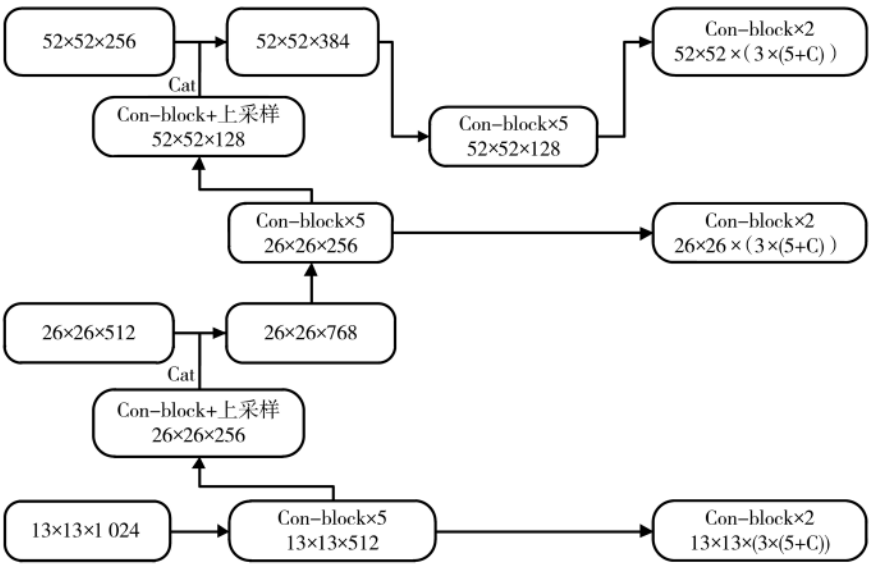

图 3 YOLOV3 采用的 FPN 网络结构

技巧，在 MSCOCO 数据集上 mAP-50 达到了 65.7%，相对于 YOLOV3 提高了大约 10%。但是所用技巧对特定场合下的目标检测效果并不是均有益。本文结合 CSPdarknet53 网络结构、darknet53 网络结构，重新设计了一种新残差单元结构 NRes-block，将输入单元分为两个部分，第一部分进行一次激活函数为 Mish 的卷积单元操作，再复用若干残差单元；第二部分只进行一次激活函数为 Mish 的卷积单元操作，最后堆叠这两部分，具体结构如图 5 所示。

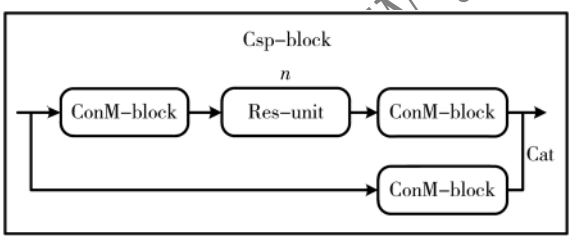

图 4 CSPdarknet53 网络的残差网络结构

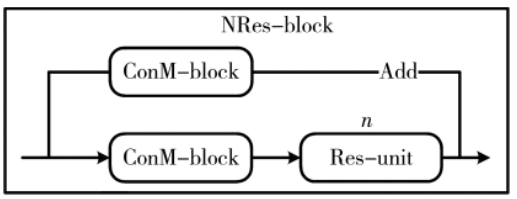

图 5 改进的残差网络结构

新设计的模型复用 NRes-block，以此构建主干网络，网络输出采用 FPN 结构，整个网络模型如图 6 所示。

图 6 所示的改进结构输入为尺寸 416×416 的三通道图像，复用卷积块和 NRes-block 网络结构后，从 52×52，26×26，13×13 三个尺度进行输出预测，13×13 尺度下的特征图经 5 次卷积块后上采样与 52×52 尺度下特征图进行堆叠，堆叠后的特征图经上采样与 52×52 尺度下的特征图进行堆叠。最终 13×13 尺度下的特征图负责预测大目标，26×26 尺度下的特征图负责预测中等目标，52×52 尺度下的特征图负责预测小目标。

2.2 数据预处理

在 YOLO 系列算法中，训练与检测的输入图像需要经过变换固定到 416×416 的大小，但是任意进行图像变换会造成目标的变形，对训练与检测有着一定影响。本文预先获取图像的宽高 w、h，使用 $\max(w,h)$ 确定变换后的一边，之后再对图像另一边进行变换，具体公式如下：

$$\mathrm{Rel}=\mathrm{Max}(w,h) \tag{3}$$

$$\mathrm{Res}=\mathrm{Min}(w,h)*\frac{416}{\mathrm{Max}(w,h)} \tag{4}$$

其中 Rel、Res 表示图像变换之后的边长，最后再对图像进行填充，使得满足训练与检测的尺度(416×416)。

图 7 所示原图像尺寸大小为 1 020×680，从图中可见预处理后的图像(右下)与直接使用 OpenCV 的 resize 方法处理后的图像(左下)对比效果。

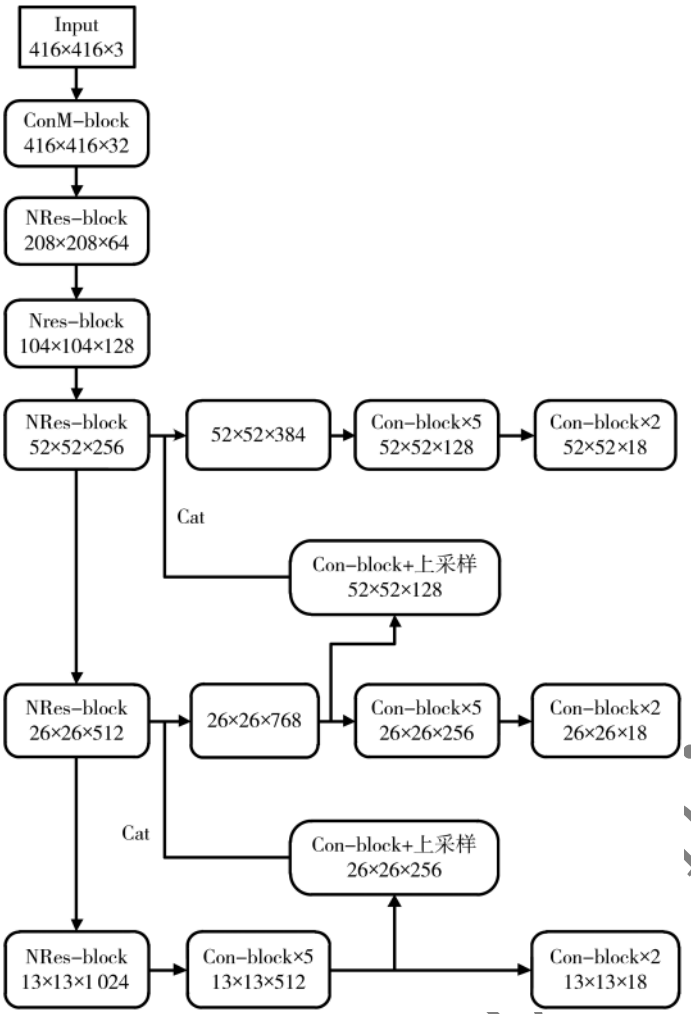

图 6 改进后网络结构

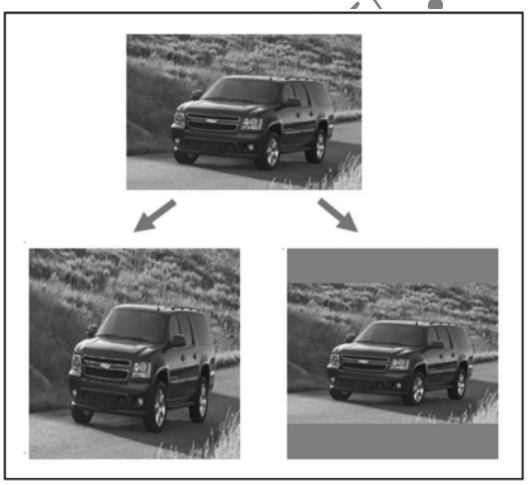

图 7 有无经预处理的图像对比

3 实验分析

3.1 模型训练

自制训练集包括 1 200 张含有车辆目标的图像，训练集与验证集划分比例为 9:1，图像目标标绘工具为 labelImg。

本文使用的硬件平台：CPU 采用 Intel E5-2678v3，GPU 采用 NVIDAI GTX1080Ti，操作系统为 Ubuntu16.04，训练框架采用 Pytorch1.2。

Batchsize 设定为 16，初始学习率设定为 $1e^{-3}$，经 1 000 次迭代后，学习率设定为 $1e^{-4}$，继续训练 400 次结束训练。训练 Loss 为 13×13、26×26、52×52 三个尺度下坐标、边框、置信度及类别损失值的总和，每经过 50 次 epoch 训练后记录 Loss 值，通过对比 YOLOV3 模型，结果如图 8 所示。

对比发现，改进残差网络结构的模型训练 Loss 下降较 YOLOV3 模型快，经 600 代训练后，Loss 逐渐稳定，且比 YOLOV3 低。

3.2 模型测试结果

在目标检测模型的评估中，精准率 P(precision)、召回率 R(recall)通常是衡量模型的重要指标，具体计算公式如下：

$$P=\frac{TP}{TP+FP} \tag{5}$$

$$R=\frac{TP}{TP+FN} \tag{6}$$

其中 TP(True Positives)表示样本被检测为正样本并且检测正确，FP 表示样本被检测为正样本但是检测错误，FN 表示样本被检测为负样本但是检测检测错误。

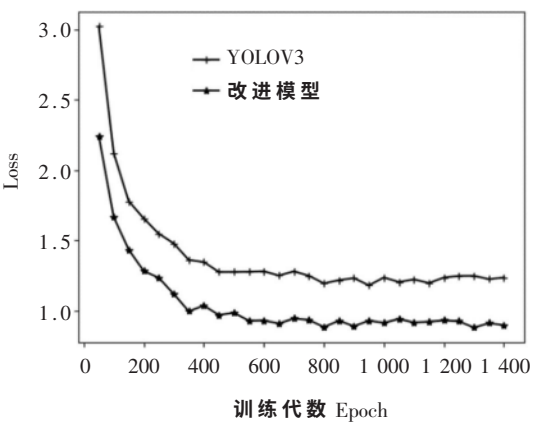

图 8 训练过程中 Loss 下降对比

精准率与召回率一般具有：精准率高、召回率低；召回率高、精准率低的特点。本文使用结合了精准率随召回率变化生成的 PR 函数与坐标轴围成的

面积——平均精度 mAP(mean Average Precision)来衡量模型性能。本文检测单一目标,mAP 即为 AP,计算公式如下:

$$AP=\int_0^1 P(R)\mathrm{d}R \tag{7}$$

将本文改进的模型与 YOLOV3 模型在自定义数据集上进行实验对比,本文设计的网络模型 mAP 为 90.63%,YOLOV3 的 mAP 为 86.47%,较 YOLOV3 提高 4.6%。

对于部分图像的检测结果如图 9 所示。

图 9 检测结果

4 结束语

本文基于 YOLO 系列模型,通过研究 ResNet、darknet53、CSPdarknet53 的网络结构,对模型结构中的残差模块进行了重新的设计,并对网络输入时的图像进行了预处理操作,最后在自行收集标绘的数据集上进行训练、测试。实验结果表明,本文设计的检测网络较 YOLOV3 的性能有着一定提升。本文设计的模型针对的是单一目标,检测目标所处背景环境非复杂场景,后续将进行复杂场景下的车辆目标识别。

参考文献

[1] LEE S, SON H, CHOI J C, et al. HOG feature extractor circuit for real-time human and vehicle detection[C]. Tencon IEEE Region 10 Conference. IEEE, 2013.

[2] 刘敦强,沈峘,夏瀚笙,等.一种基于深度残差网络的车型识别方法[J].计算机技术与发展,2018,28(5):42-46.

[3] 张富凯,杨峰,李策.基于改进 YOLOv3 的快速车辆检测方法[J].计算机工程与应用,2019,55(2):12-20.

[4] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[5] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[J]. arXiv: 1506.02640, 2015.

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

[7] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[J]. arXiv: 2004.10934, 2020.

[8] WANG C Y, LIAO H Y M, YEH I H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[J]. arXiv: 1911.11929, 2019.

(收稿日期:2020-06-25)

作者简介:

胡臣辰(1996-),男,硕士研究生,主要研究方向:智能信息处理。

陈贤富(1963-),男,副教授,主要研究方向:复杂系统与认知计算。

版权声明