

基于反残差结构的轻量级多目标检测网络

刘万军, 高明月, 曲海成*, 刘腊梅

辽宁工程技术大学软件学院, 辽宁 葫芦岛 125105

摘要 针对 YOLO 系列的目标检测方法参数多、计算量大、生成检测模型规模大等导致对运行硬件平台计算资源要求高的问题, 提出一种基于反残差结构的轻量级多目标检测网络 (IR-YOLO)。首先, 利用深度可分离卷积减少模型参数和计算量; 其次, 基于深度可分离卷积构造反残差模块, 提取高维特征; 最后, 根据反残差结构特点, 利用线性激活函数减少通道组合过程激活函数的信息损失。IR-YOLO 算法较 YOLOv3-Tiny 算法模型尺寸减少 47.7%。实验结果表明 IR-YOLO 算法在不影响检测精度的前提下, 可有效减少网络计算量和存储量。

关键词 图像处理; 目标检测; 反残差结构; 深度可分离卷积

中图分类号 TP391

文献标识码 A

doi: 10.3788/LOP56.221003

Light-Weight Multi-Object Detection Network Based on Inverted Residual Structure

Liu Wanjun, Gao Mingyue, Qu Haicheng*, Liu Lamei

College of Software, Liaoning Technical University, Huludao, Liaoning 125105, China

Abstract To solve high computational resource requirement for running hardware platform of the series of the YOLO object detection method due to the huge parameters, the large amount of calculation, and the large scale of detection model, this paper developed a light-weight object detection network based on inverted residual structure (IR-YOLO). First, it used depth separable convolution to reduce detection model parameters and computational quantities. Secondly, it constructed inverted residual block based on depth separable convolution to extract high-dimensional feature. Finally, according to the characteristic of inverted residual structure, it used a linear activation function to reduce the information loss during the process of channels combination. The experimental results show that the IR-YOLO detection model is reduced by 47.7% compared to the YOLOv3-Tiny detection model, it validated that the IR-YOLO algorithm can effectively compress the model while maintaining detection accuracy.

Key words image processing; object detection; inverted residual structure; depth separable convolution

OCIS codes 100.2000; 100.4996; 100.2960

1 引言

目标检测是计算机视觉领域研究的热点问题之一, 在医学、交通、军事等各个领域有广泛的应用前景^[1]。目前, 目标检测方法主要分为传统目标检测方法和基于深度学习的目标检测方法^[2]。传统的目标检测方法^[3]主要步骤为: 1) 选择候选区域; 2) 提取候选区域的特征; 3) 应用分类器检测候选区域是否

含有目标。但传统的目标检测方法由于应用手工特征, 导致对图像中多尺度目标检测时稳健性差, 利用滑动窗口选择候选区域时时间复杂度高。基于深度学习的目标检测无需人工设计特征, 且具有较高的检测精度, 逐渐成为目标检测的主流算法, 主要可分为两种: 一是基于区域的目标检测方法; 二是基于回归的目标检测方法。基于区域的方法由区域提取和目标分类两步构成, 其典型代表有 Girshick 等^[4]提出

收稿日期: 2019-03-27; 修回日期: 2019-04-27; 录用日期: 2019-05-17

基金项目: 国家自然科学基金青年基金项目(41701479)、辽宁省自然科学基金(20180550529)、第六批生产技术问题创新研究基金(20160092T)

* E-mail: quhaicheng@lntu.edu.cn

的 R-CNN 目标检测框架。R-CNN 框架首先提取候选框,调整候选框为固定尺寸;其次应用卷积神经网络(CNN)对候选区域特征进行提取;最后应用支持向量机(SVM)分类及回归算法得到预测框。为优化提取候选框过程和提高检测精度,R-CNN 框架的变体 SPP-Net^[5]、Fast-R-CNN^[6]、Faster-R-CNN^[7]、R-FCN^[8]等逐一出现,并在各个领域得到广泛应用^[9-11]。虽然基于区域的目标检测方法能产生较高的精度,但是检测速度慢。为解决检测速度的问题,基于回归的目标检测方法应运而生。基于回归的方法,不产生候选框,将目标位置和类别问题转换为回归问题,典型代表为 YOLO 和 SSD 算法。2016—2018 年,Redmon 等^[12-14]提出 YOLO 框架,将目标分类和定位转换为回归问题。2016 年 Liu 等^[15]提出的 SSD 框架沿用了 YOLO 基于回归的思想,结合 Faster-R-CNN 的 anchor box(候选框),提出 prior box(预设框),加入特征金字塔,有效平衡检测速度与精度,但是 SSD 算法对训练样本依赖严重。2018 年,Redmon 等^[14]提出的 YOLO 第 3 个版本借鉴了残差网络,引入 anchor box,使用多尺度预测,具有较强的实时检测能力,因此本文引入 YOLO 端到端的思想。

目标检测可以应用于安防、汽车辅助驾驶等方面,但上述目标检测方法参数多、计算量大、生成模型大等导致数据难以移植到计算能力及存储空间有限的嵌入式终端应用。目前,优化 CNN 的方法主要有 3 类:一是采用手工方式设计高效的网络结构,典型代表有 SqueezeNet^[16]、MobileNet^[17-18];二是通过微调等方法细化结构,典型代表有剪枝^[19-20]、量化^[21]、低秩分解^[22]等;三是使用迁移知识,通过训练好的大模型得到小模型,典型代表有知识蒸馏^[23]。其中手工设计结构深度可分离卷积,比常规卷积效率更高,同时仍能保持高精度。因此本文应用深度可分离卷积,旨在解决深度学习方法对计算资源的依赖问题。

综上所述,本文针对基于深度学习的目标检测模型对运行平台计算力大,难以移植到嵌入式平台的问题,提出轻量级多目标检测网络(IR-YOLO)。该网络引入反残差块,利用多尺度特征图检测不同尺寸目标,在不影响检测精度的同时,减少了模型计算量和参数存储量。

2 YOLOv3-Tiny 目标检测原理

基于回归的目标检测算法是高效完成检测任务

的最常用的方法,将其目标分类和定位转换为回归问题,达到了实时检测的目的^[24]。YOLOv3-Tiny 框架是在 YOLOv3 基础上应用嵌入式平台的轻量级目标检测网络,虽然 YOLOv3-Tiny 在检测精度上较 YOLOv3 降低,但是能实现模型尺寸压缩,即 YOLOv3-Tiny 可将 YOLOv3 特征检测网络 darknet-53 缩减为 7 层常规卷积和 6 层 max pooling(最大池化)下采样层。

首先将输入图片调整成 10 种不同尺寸,将调整后的图片按 13×13 、 26×26 划分网格(grid cell);然后在非重叠的 $S \times S$ 个 grid cell 上回归得到目标的位置和类别。对于每个 grid cell,模型输出 B 个边界框(bounding box)和相应的置信度分数及目标对应类别的条件概率,然后利用非极大值抑制过滤冗余的边界框,置信度得分公式为

$$S_{\text{score}} = P_{r,\text{object}} \times I_{\text{truth,pred}} \quad (1)$$

式中: $P_{r,\text{object}}$ 为有无目标,如果不含目标,则 $P_{r,\text{object}} = 0$;如果含有目标,则 $P_{r,\text{object}} = 1$; $I_{\text{truth,pred}}$ 为边界框与真实框之间的交并比(IOUS)。

YOLOv3-Tiny 网络损失函数由边界框中心坐标误差项、边界框的宽与高误差项、预测置信度误差项和预测类别误差项组成,损失函数计算公式为

$$\begin{aligned} l_{\text{loss}} = & \lambda_{\text{coord}} \sum_{i=0}^{s^2} \sum_{j=0}^B 1_{\text{obj},ij} [(x_i - \hat{x}_i)^2 + \\ & (y_i - \hat{y}_i)^2] + \lambda_{\text{coord}} \sum_{i=0}^{s^2} \sum_{j=0}^B 1_{\text{obj},ij} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + \\ & (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] + \sum_{i=0}^{s^2} \sum_{j=0}^B 1_{\text{obj},ij} (C_i - \hat{C}_i)^2 + \\ & \lambda_{\text{noobj}} \sum_{i=0}^{s^2} \sum_{j=0}^B 1_{\text{obj},ij} (C_i - \hat{C}_i)^2 + \\ & \sum_{i=0}^{s^2} 1_{\text{obj},ij} \sum_{c \in c_{\text{classes}}} [p_i(c) - \hat{p}_i(c)]^2 \quad (2) \end{aligned}$$

式中:等式右第一项为中心坐标损失,第二项为边界框的宽高损失,第三项为置信度损失,第四项为类别预测损失。其中: λ_{coord} 为定位误差经验权重,为使回归边界框定位准确,设定 $\lambda_{\text{coord}} = 5$; s 为输入图像划分的网格数; B 为每个 grid cell 预测的边界框个数; $1_{\text{obj},ij}$ 为第 i 个 grid cell 中第 j 个边界框中含有目标,若不含目标则不对损失函数进行反向传播; x_i 、 y_i 和 \hat{x}_i 、 \hat{y}_i 分别为真实框和预测框的中心坐标; w_i 、 h_i 和 \hat{w}_i 、 \hat{h}_i 分别为真实框和预测框的宽和高; C_i 、 \hat{C}_i 分别为 grid cell 中目标的真实框和预测框的置信度得分; λ_{noobj} 为误差经验权重,为平衡有无目标样本,设定 $\lambda_{\text{noobj}} = 0.5$; c_{classes} 为总目标类别; $p_i(c)$ 、 $\hat{p}_i(c)$ 分别

为目标属于类别 c 的真实和预测的条件概率。

3 IR-YOLO 网络

YOLOv3-Tiny 框架可以实现端到端的目标检测,然而随着网络层数的增长,参数和计算量也随之成倍增长,从而导致模型计算资源开销增大。因此为减少网络模型参数存储量和计算量,利用深度可分离卷积构造反残差块代替常规卷积层,设计轻量级多目标检测网络。IR-YOLO 网络由多个反残差块堆叠来进行特征提取,其实现分为两个部分,一是

构造反残差块,二是设计网络结构。

3.1 反残差块的构造

为解决常规卷积参数存储量和计算量随网络层数加深而成倍增长,导致模型尺寸增大,难以在计算资源受限的硬件平台应用的问题,应用深度可分离卷积代替常规卷积,构造反残差块,将“空间跨通道”特征学习过程转变为空间特征学习和通道组合两部分:一是深度卷积在每个输入通道上独立地进行空间卷积;二是点卷积将深度卷积输出结果映射到新的通道空间。深度可分离卷积解耦过程如图 1 所示。

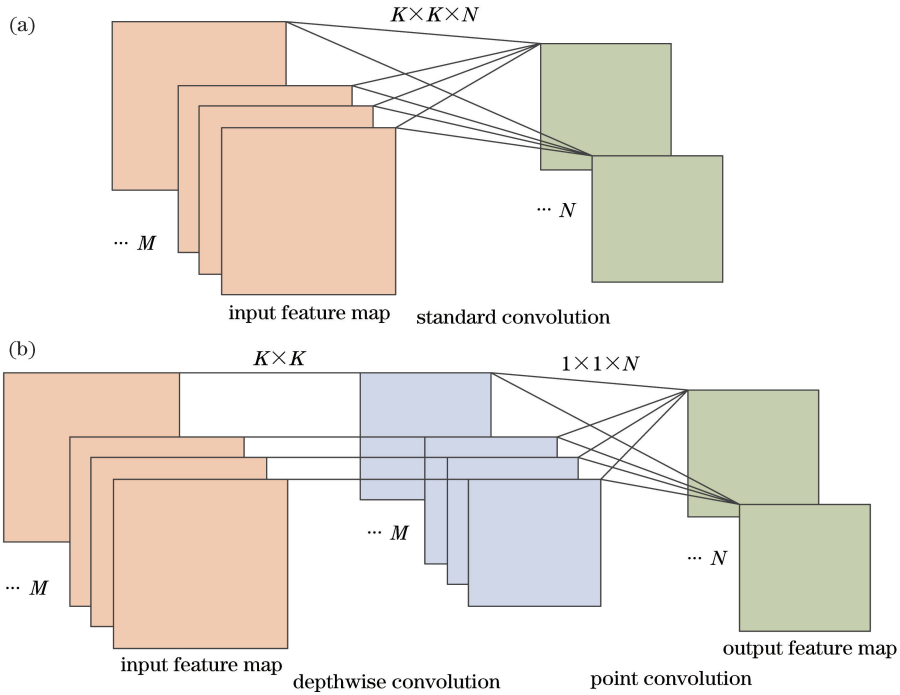


图 1 深度可分离卷积解耦过程图。(a)标准卷积;(b)深度可分离卷积

Fig. 1 Decoupling process of the depth separable convolution. (a) Standard convolution; (b) depth separable convolution

图 1 中 $K \times K \times N$ 为常规卷积核大小, $K \times K$ 为深度卷积核大小, $1 \times 1 \times N$ 为点卷积核大小, M 为卷积计算中的输入特征图通道数, N 为输出特征图通道数。

深度可分离卷积能够有效减少参数,常规卷积计算可表示为

$$G_N = \sum_M K_{M,N} * F_M \quad (3)$$

深度卷积计算可表示为

$$\hat{G}_M = \sum \hat{K}_{1,M} * F_M \quad (4)$$

式中, K 和 \hat{K} 为过滤器, F 为输入尺寸, G 和 \hat{G} 为输出尺寸。由 (3) 式得到常规卷积计算量为 $F \times F \times M \times N \times K \times K$, 由 (4) 式得到深度卷积计算量为 $K \times K \times M \times F \times F$, 则深度可分离卷积与常规卷积计算量之比为

$$\frac{K \times K \times M \times F \times F + M \times N \times F \times F}{K \times K \times M \times N \times F \times F} = \frac{1}{N} + \frac{1}{K^2} \quad (5)$$

综上所述,深度可分离卷积有效降低了网络的计算复杂度。虽然深度可分离卷积有效降低了模型参数存储量,但会导致网络层数加深。残差结构能够解决随网络层数加深梯度消失的问题,常规残差模块先通过 1×1 卷积层压缩特征图通道,然后在低通道应用 3×3 卷积层提取特征,最后再通过 1×1 的卷积层扩张特征图通道。但是残差结构在压缩特征图时存在损害特征表达的问题,因此在特征提取过程中利用反残差模块,先通过 1×1 卷积层扩张通道,然后在高维特征图应用 3×3 的深度卷积层提取特征,最后再通过 1×1 的点卷积层将深度卷积结果

映射到新的通道空间。残差结构与反残差结构如图 2 所示。

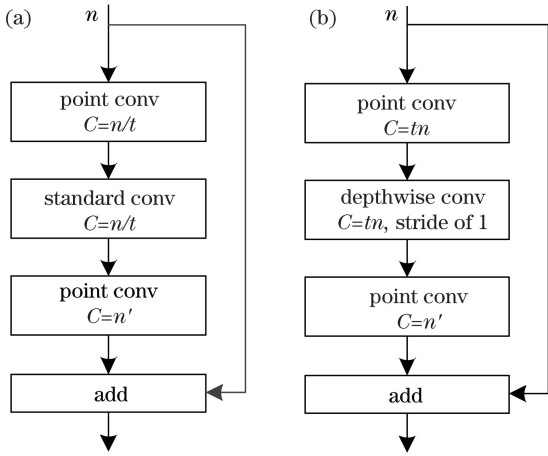


图 2 残差结构与反残差结构。(a)残差结构;

(b)步长为 1 的反残差结构

Fig. 2 Residual block and inverted residual block.

(a) Residual block; (b) inverted residual block when stride is 1

图 2 中 point conv 为点卷积操作, depthwise conv 为深度卷积操作, n 为反残差结构的输入通道数, t 为扩充或压缩通道的倍数, C 为通道数, n' 为经过扩充(或压缩)处理之后的通道数。反残差块参数计算见表 1。

表 1 中: h 、 w 分别为特征图的高和宽, k 为特征

type	output
convolution	$416 \times 416 \times 16$
max pooling	$208 \times 208 \times 16$
inverted residual block	$208 \times 208 \times 32$
max pooling	$104 \times 104 \times 32$
inverted residual block	$104 \times 104 \times 64$
max pooling	$52 \times 52 \times 64$
inverted residual block	$52 \times 52 \times 128$
max pooling	$26 \times 26 \times 128$
inverted residual block	$26 \times 26 \times 256$
max pooling	$13 \times 13 \times 256$
inverted residual block	$13 \times 13 \times 512$
max pooling	$13 \times 13 \times 512$
inverted residual block	$13 \times 13 \times 1024$

表 1 反残差块参数

Table 1 Parameters of inverted residual block

Input	Operation	Output
$h \times w \times k$	1×1 point conv, ReLU	$h \times w \times 2k$
$h \times w \times 2k$	$3 \times 3/s$ depth conv, ReLU	$\frac{h}{s} \times \frac{w}{s} \times 2k$
$\frac{h}{s} \times \frac{w}{s} \times 2k$	1×1 point conv, linear	$\frac{h}{s} \times \frac{w}{s} \times 2k$

图的通道数; s 为步长; operation 为反残差块的计算过程。由表 1 可知, 反残差结构中扩充通道的点卷积和深度卷积均应用非线性激活函数(ReLU)。

由于组合通道数的点卷积层使用 ReLU 会导致负值变为 0, 从而失去部分信息, 因此使用线性激活函数可解决组合通道过程中信息损失问题。

3.2 IR-YOLO 网络架构

为解决目标检测网络参数存储量大的问题, 提出 IR-YOLO 轻量级多目标检测网络, 网络结构如图 3 所示, 其中虚线框为 IR-YOLO 网络特征提取部分。首先, 特征提取网络由 6 个下采样层和 6 个反残差块构成, 通过反残差块, 扩张特征图通道提取高纬度特征, 再进行通道降维, 得到特征图。同时基于深度可分离卷积的反残差块能够有效降低模型计算复杂度。其次, 采用多尺度检测目标, 应用浅层特征图 26×26 检测较小目标, 应用 13×13 检测较大目标。

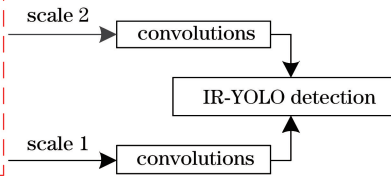


图 3 IR-YOLO 网络框架

Fig. 3 IR-YOLO network architecture

4 实验结果及分析

4.1 实验环境

从计算力消耗、模型大小、检测速度和检测精度 4 个方面验证 IR-YOLO 网络性能。本文实验环境为: Cuda 8.0 加速; 硬件配置为 Intel(R) Core (TM) i7-8750H CPU 2.20 GHz、GeForce GTX

1070 Mobile 显卡; 操作系统为 Ubuntu 18.04.1 LTS。实验采用 VOC 2007 + 2012 数据集, 其中: 训练集和验证集共 21503 张图片, 标注 64046 个目标; 测试集 7013 张图片, 共 12032 个目标。VOC 数据集标记 20 个类别, 数据集中各类数据不平衡, 其中 Person 类别目标较多。各个类别目标占比见表 2。

表2 VOC数据集
Table 2 VOC dataset

Category	Train set	Test set
Aeroplane	1171	285
Bicycle	1064	337
Bird	1605	459
Boat	1140	263
Bottle	1764	469
Bus	822	213
Car	3267	1201
Cat	1593	358
Chair	3152	756
Cow	847	244
Dining table	824	206
Dog	2025	489
Horse	1072	348
Motor bike	1052	325
Person	13256	4528
Potted plant	1487	480
Sheep	1070	242
Sofa	814	239
Train	925	282
TV monitor	1108	308
Total	40058	12032

4.2 网络的训练

为验证 IR-YOLO 框架的有效性,选取 VOC 训练集分别训练 YOLOv3-Tiny 模型和 IR-YOLO 模型。实验中超参数设置如表 3 所示。

基于经验及实验验证设定超参数,根据实验平台硬件资源,设计 Batch(批次)为 64;基于训练卷积神经网络经验值,设计初始学习率为 0.001,然后根据迭代次数更新学习率;参考 YOLO 系列算法设计

表4 浮点运算次数对比

Table 4 Comparison on number of floating point operations

Input	Output	Number of floating point operations in standard conv / 10^9	Number of floating point operations in inverted residual block / 10^9		
			Expand point conv	Depth conv	Squeeze point conv
$208 \times 208 \times 16$	$208 \times 208 \times 32$	0.399	0.044	0.025	0.089
$104 \times 104 \times 32$	$104 \times 104 \times 64$	0.399	0.044	0.012	0.089
$52 \times 52 \times 64$	$52 \times 52 \times 128$	0.399	0.044	0.006	0.089
$26 \times 26 \times 128$	$26 \times 26 \times 256$	0.399	0.044	0.003	0.089
$13 \times 13 \times 256$	$13 \times 13 \times 512$	0.399	0.044	0.002	0.089
$13 \times 13 \times 512$	$13 \times 13 \times 1024$	1.595	0.177	0.003	0.354

从表 4 可知,应用深度卷积和点卷积分解标准卷积时,模型计算力消耗明显减少;同时输入特征图越大,计算力消耗减少效果也越明显。在 104×104 的特征图提取过程中,深度卷积需要 0.012×10^9 次

表3 超参数

Table 3 Hyper parameters

Parameters name	Value
Batch	64
Momentum	0.9
Weight decay	0.0005
Learning rate	0.001

权重衰减参数为 0.0005,动量为 0.9。

YOLOv3-Tiny 模型和 IR-YOLO 模型的训练损失如图 4 所示。

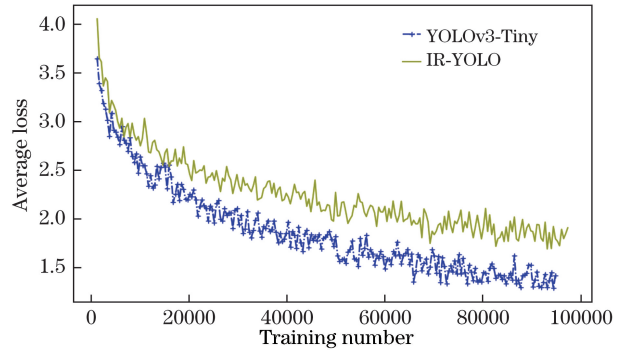


图4 训练损失图

Fig. 4 Train loss curves

由图 4 及损失日志文件可知,模型在训练 65000 次时损失趋于平稳,本文应用 65000 次至 95000 次的权重进行实验。

4.3 实验结果及分析

为进一步验证 IR-YOLO 模型能有效减少参数和计算量,将 IR-YOLO 与 YOLO 中轻量级检测模型 YOLOv3-Tiny 进行比较,分析常规卷积及基于深度可分离卷积的反残差结构需要的浮点运算次数,对比结果如表 4 所示。

浮点运算,在 13×13 特征图提取过程中,深度卷积只需 0.002×10^9 次浮点运算。进一步,IR-YOLO 的模型大小为 18.2 MB,YOLOv3-Tiny 的模型大小为 34.8 MB,由此验证 IR-YOLO 有效减小了模型大小,其模型尺寸缩减了 47.7%。

为验证模型检测速度,分别在图形处理单元(GPU)和中央处理单元(CPU)环境下进行验证实验,两种模型对比结果如表5所示。

表5 IR-YOLO模型与YOLOv3-Tiny模型检测速度对比

Table 5 Comparison detection speed of IR-YOLO model and YOLOv3-Tiny model

Model	CPU speed / (frame · s ⁻¹)	GPU speed / (frame · s ⁻¹)
YOLOv3-Tiny	1.2	31.3
IR-YOLO	1.7	31.2

从表5中可知,IR-YOLO模型在GPU上的检测速度可以达到实时检测效果;在CPU上的检测速度较YOLOv3-Tiny提高了0.5 frame · s⁻¹;由于实验平台GPU与CPU型号限制,若将模型部署到算力更高的GPU与CPU上,检测速度还会有所提升。

根据图4可知,在迭代65000次时,模型损失值趋于稳定,可以近似得到网络特征输出,并判断网络的检测精度。为验证模型的检测精度,本文选取网络收敛后的权重进行对比,应用平均精度均值(mAP, m_{AP})作为评价指标。

检测准确率表示为

$$P_c = \frac{T_c}{T_c + F_c}, \quad (6)$$

式中, T_c 为正确检测类别 c 的个数, F_c 为错误识别类别 c 的个数。实验阈值为0.5,IOU大于阈值则为正确识别。

召回率表示为

$$R_c = \frac{T_c}{T_c + N_c}, \quad (7)$$

式中, N_c 为漏检类别 c 的个数。单个类别检测平均精确率 A_c 为P-R曲线求积分,其表示为

$$A_c = \int_0^1 P(R) dR, \quad (8)$$

式中, P 为类别 c 的准确率, R 为类别 c 的召回率。mAP为数据集中 N 个类别平均精确率之和求均值,表示为

$$m_{AP} = \frac{1}{N_{classes}} \sum A_c, \quad (9)$$

式中, $N_{classes}$ 为数据集所含类别个数。

两种模型对应权重的检测精度如表6所示。

由表6可知,YOLOv3-Tiny模型在训练75000次时,检测精度最佳;IR-YOLO模型在训练95000次时,检测精度最佳。

为进一步验证IR-YOLO模型的检测精确度,

选取YOLOv3-Tiny模型75000次权重和IR-YOLO模型95000次权重进行实验验证,在VOC 2007测试集上对比IR-YOLO与YOLOv3-Tiny的检测精度,结果如表7所示。

表6 不同训练次数mAP对比

Table 6 Comparison mAP of different training numbers

Training number	YOLOv3-Tiny mAP / %	IR-YOLO mAP / %
65000	45.15	43.33
75000	45.60	44.37
85000	45.17	45.23
90000	42.75	44.20
95000	42.76	46.07

表7 IR-YOLO模型与YOLOv3-Tiny模型在VOC数据集上检测结果对比

Table 7 Comparison of detection results of IR-YOLO and YOLOv3-Tiny on VOC dataset %

Category	YOLOv3-Tiny	IR-YOLO
Aeroplane	54.78	56.38
Bicycle	60.79	57.86
Bird	27.24	28.19
Boat	27.9	28.92
Bottle	14.8	17.58
Bus	56.98	58.48
Car	63.8	64.05
Cat	50.39	53.57
Chair	25.77	23.25
Cow	46.43	45.48
Dining table	39.66	45.48
Dog	46.09	45.68
Horse	66.62	62.45
Motor bike	64.09	62.85
Person	59.23	59.4
Potted plant	18.22	17.22
Sheep	47.57	44.68
Sofa	39.39	43.11
Train	54.02	58.25
TV monitor	50.34	48.62
mAP	45.60	46.07

两种模型在VOC数据集的检测精度直方图如图5所示。

YOLOv3-Tiny在VOC 2007+2012数据集上mAP为45.60%,IR-YOLO算法在VOC 2007+2012数据集上mAP为46.07%,两种模型的检测精度基本一致。

YOLOv3-Tiny模型和IR-YOLO模型实验效果对比如图6所示。

从图6(b)、(c)中可以直观地看到,YOLOv3-

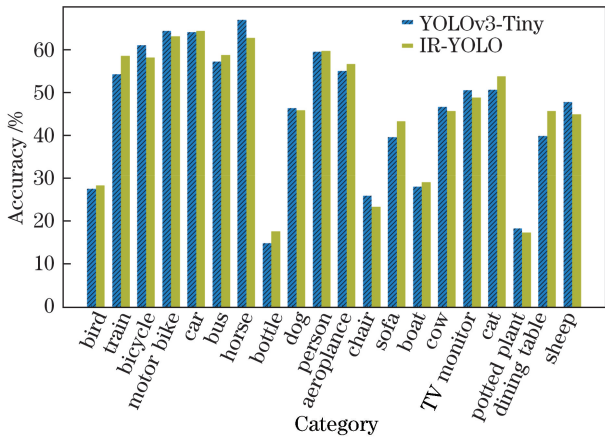


图 5 检测精度直方图

Fig. 5 Class detection accuracy histogram

Tiny 模型自行车漏检, 而 IR-YOLO 模型可以检测出两辆自行车; 从图 6(e)、(f) 中可以直观地看到, YOLOv3-Tiny 模型没有检测出图片中的公交车, 而 IR-YOLO 模型可以检测出公交车。由图 6

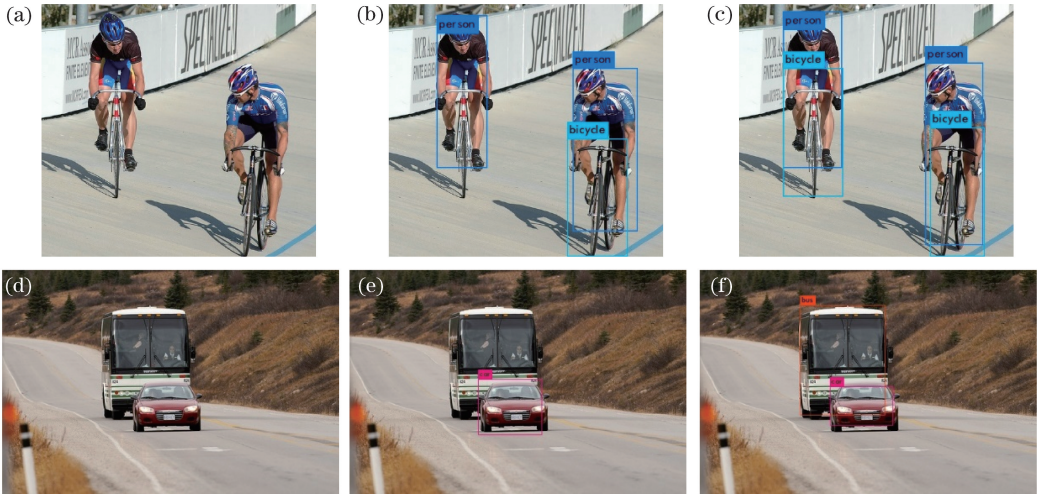


图 6 实验效果对比。(a)(d)原始输入图像;(b)(e) YOLOv3-Tiny 模型检测结果;(c)(f) IR-YOLO 模型检测结果

Fig. 6 Comparison of detection results. (a)(d) Original input images ; (b)(e) detection results with YOLOv3-Tiny Model; (c)(f) detection results with IR-YOLO Model

参 考 文 献

[1] Wang Y Q, Ma L, Tian Y. State-of-the-art of ship detection and recognition in optical remotely sensed imagery[J]. Acta Automatica Sinica, 2011, 37(9): 1029-1039.
王彦情, 马雷, 田原. 光学遥感图像舰船目标检测与识别综述 [J]. 自动化学报, 2011, 37(9): 1029-1039.

[2] Zhang H, Wang K F, Wang F Y. Advances and perspectives on applications of deep learning in visual object detection [J]. Acta Automatica Sinica, 2017, 43(8): 1289-1305.

可知, IR-YOLO 模型在模型规模较小时仍可保证检测精度, 有效解决了目标检测网络对运行硬件平台计算资源要求高的问题。

5 结 论

针对深度学习目标检测方法计算复杂度高、计算资源开销大、生成检测模型规模大, 难以在硬件资源受限平台应用的问题, 提出一种轻量级的多目标检测网络 IR-YOLO。通过深度可分离卷积代替常规卷积, 解决了模型参数存储量和计算量大的问题。通过反残差结构先升维、提取特征、再降维的特征提取过程, 及组合通道过程应用线性激活函数, 降低信息损失, 解决了由于减小网络模型尺寸导致的精度下降问题。实验结果表明 IR-YOLO 模型在不影响检测精度的前提下, 模型尺寸减小 47.7%。下一步工作将进一步研究网络结构对提升模型训练速度及目标检测精度的影响。

张慧, 王坤峰, 王飞跃. 深度学习在目标视觉检测中的应用进展与展望 [J]. 自动化学报, 2017, 43(8): 1289-1305.

[3] Yin H P, Chen B, Chai Y, et al. Vision-based object detection and tracking: a review [J]. Acta Automatica Sinica, 2016, 42(10): 1466-1489.
尹宏鹏, 陈波, 柴毅, 等. 基于视觉的目标检测与跟踪综述 [J]. 自动化学报, 2016, 42(10): 1466-1489.

[4] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York:

- IEEE, 2014: 580-587.
- [5] He K M, Zhang X Y, Ren S Q, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [6] Girshick R. Fast R-CNN [C] // 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 1440-1448.
- [7] Ren S Q, He K M, Girshick R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [8] Dai J F, Li Y, He K M, *et al.* R-FCN: object detection via region-based fully convolutional networks [C] // Advances in Neural Information Processing Systems 29 (NIPS 2016), December 5-10, 2016, Centre Convencions Internacional Barcelona, Barcelona SPAIN. Canada: NIPS, 2016: 379-387.
- [9] Cao Y J, Xu G M, Shi G C. Low altitude armored target detection based on rotation invariant faster R-CNN[J]. Laser & Optoelectronics Progress, 2018, 55(10): 101501.
曹宇剑, 徐国明, 史国川. 基于旋转不变 Faster R-CNN 的低空装甲目标检测[J]. 激光与光电子学进展, 2018, 55(10): 101501.
- [10] Feng X Y, Mei W, Hu D S. Aerial target detection based on improved faster R-CNN[J]. Acta Optica Sinica, 2018, 38(6): 0615004.
冯小雨, 梅卫, 胡大帅. 基于改进 Faster R-CNN 的空中目标检测 [J]. 光学学报, 2018, 38(6): 0615004.
- [11] Zhang C, Chen Y. Object detection based on hard examples mining using residual network[J]. Laser & Optoelectronics Progress, 2018, 55(10): 101003.
张超, 陈莹. 残差网络下基于困难样本挖掘的目标检测[J]. 激光与光电子学进展, 2018, 55(10): 101003.
- [12] Redmon J, Divvala S, Girshick R, *et al.* You only look once: unified, real-time object detection[C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 779-788.
- [13] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 6517-6525.
- [14] Redmon J, Farhadi A. YOLOv3: an incremental improvement[J/OL]. (2018-04-08) [2019-03-01]. <https://arxiv.org/abs/1804.02767>.
- [15] Liu W, Anguelov D, Erhan D, *et al.* SSD: single shot multibox detector[M] // Leibe B, Matas J, Sebe N, *et al.* European conference on computer vision-ECCV 2016. Lecture Notes in Computer Science. Cham: Springer, 2016, 9905: 21-37.
- [16] Iandola F N, Han S, Moskewicz M W, *et al.* SqueezeNet: AlexNet-level accuracy with $50 \times$ fewer parameters and < 0.5 MB model size [J/OL]. (2016-11-04) [2019-03-01]. <https://arxiv.org/abs/1602.07360>.
- [17] Howard A G, Zhu M L, Chen B, *et al.* MobileNets: efficient convolutional neural networks for mobile vision applications [J/OL]. (2017-04-17) [2019-03-01]. <https://arxiv.org/abs/1704.04861>.
- [18] Sandler M, Howard A, Zhu M L, *et al.* MobileNetV2: inverted residuals and linear bottlenecks [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 4510-4520.
- [19] Yoon J, Hwang S J. Combined group and exclusive sparsity for deep neural networks[C] // Proceedings of the 34th International Conference on Machine Learning, August 6-11, 2017, Sydney, NSW, Australia. New York: ACM, 2017, 70: 3958-3966.
- [20] Liu Z, Li J G, Shen Z Q, *et al.* Learning efficient convolutional networks through network slimming[C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2736-2744.
- [21] Zhou S C, Wu Y X, Ni Z K, *et al.* DoRefa-Net: training low bitwidth convolutional neural networks with low bitwidth gradients [J/OL]. (2018-02-02) [2019-03-01]. <https://arxiv.org/abs/1606.06160>.
- [22] Sainath T N, Kingsbury B, Sindhwani V, *et al.* Low-rank matrix factorization for Deep Neural Network training with high-dimensional output targets[C] // 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, May 26-31, 2013, Vancouver, BC, Canada. New York: IEEE, 2013: 6655-6659.

- [23] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network[J/OL]. (2015-03-09) [2019-03-01]. <https://arxiv.org/abs/1503.02531>.
- [24] Wu S, Xu Y, Zhao D N. Survey of object detection based on deep convolutional network [J]. Pattern Recognition and Artificial Intelligence, 2018, 31(4): 335-346.
- 吴帅, 徐勇, 赵东宁. 基于深度卷积网络的目标检测综述[J]. 模式识别与人工智能, 2018, 31(4): 335-346.