

基于多向特征金字塔的轻量级目标检测算法

白创, 王英杰, 闫昱, DJUKANOVIC Milena

引用本文:

白创, 王英杰, 闫昱, 等. 基于多向特征金字塔的轻量级目标检测算法[J]. 液晶与显示, 2021, 36(11): 1516–1524.

BAI Chuang, WANG Ying-jie, YAN Yu, et al. Lightweight object detection algorithm based on multi-directional feature pyramid[J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(11): 1516–1524.

在线阅读 View online: <https://doi.org/10.37188/CJLCD.2021-0018>

您可能感兴趣的其他文章

Articles you may be interested in

一种高精度的卷积神经网络安全帽检测方法

High precision detection method of safety helmet based on convolution neural network

液晶与显示. 2021, 36(7): 1018–1026 <https://doi.org/10.37188/CJLCD.2020-0309>

基于深度可分离卷积的交通标志识别算法

Traffic sign recognition algorithm based on depthwise separable convolutions

液晶与显示. 2019, 34(12): 1191–1201 <https://doi.org/10.3788/YJYXS20193412.1191>

基于特征融合的轻量级SSD目标检测方法

Lightweight SSD object detection method based on feature fusion

液晶与显示. 2021, 36(10): 1437–1444 <https://doi.org/10.37188/CJLCD.2021-0007>

基于YOLOv4-tiny改进的轻量级口罩检测算法

Lightweight mask detection algorithm based on improved YOLOv4-tiny

液晶与显示. 2021, 36(11): 1525–1534 <https://doi.org/10.37188/CJLCD.2021-0059>

基于注意力掩模融合的目标检测算法

Object detection algorithm based on attention mask fusion

液晶与显示. 2019, 34(8): 825–833 <https://doi.org/10.3788/YJYXS20193408.0825>

文章编号:1007-2780(2021)11-1516-09

基于多向特征金字塔的轻量级目标检测算法

白 创^{1,2}, 王英杰^{1*}, 闫 昱¹, DJUKANOVIC Milena³

- (1. 长沙理工大学 物理与电子科学学院, 湖南 长沙 410114;
2. 柔性电子材料基因工程湖南省重点实验室, 湖南 长沙 410114;
3. 黑山大学 电气工程学院, 黑山共和国 波德戈里察 81000)

摘要:针对实时目标检测算法 Tiny YOLOv3 存在的深层特征难以提取、实时性不能满足需求、边界框定位不够准确的问题,提出一种改进的轻量级检测模型 MTYOLO (MdFPN Tiny YOLOv3)。该模型构造多向特征金字塔网络 (MdFPN) 代替简单级联,充分完成了多层语义信息的提取与融合;使用深度可分离卷积代替标准卷积,有效降低网络复杂度,提升了检测的实时性能;采用 Complete IOU loss (CIOU loss) 代替 MSE 作为回归损失函数,极大地提高了边界框的回归精度。在 PASCAL VOC 和 COCO 数据集上对 MTYOLO 进行测试,结果表明,改进后模型的 mAP 最高可达到 83.7%,检测速度最快可达到 205 fps。

关 键 词: 目标检测;特征金字塔网络;深度可分离卷积;Complete IOU loss

中图分类号: TP391.41 **文献标识码:** A **doi:** 10.37188/CJLCD.2021-0018

Lightweight object detection algorithm based on multi-directional feature pyramid

BAI Chuang^{1,2}, WANG Ying-jie^{1*}, YAN Yu¹, DJUKANOVIC Milena³

- (1. College of Physics and Electronic Science, Changsha University of Technology, Changsha 410114, China;
2. Hunan Provincial Key Laboratory of Flexible Electronic Materials Genome Engineering, Changsha 4100114, China;
3. Faculty of Electrical Engineering, University of Montenegro, Podgorica 81000, Montenegro)

Abstract: Aiming at the problems that the deep features are difficult to extract, the real-time performance cannot meet the requirements, and the bounding box positioning is not accurate enough of the real-time object detection algorithm Tiny YOLOv3, an improved lightweight detection model MTYOLO

收稿日期: 2021-01-23; **修订日期:** 2021-03-28.

基金项目: 中国-黑山科技合作委员会第三届例会交流合作项目 (No. 3-7); 长沙理工大学“双一流”科学研究国际合作拓展项目 (No. 2019ic18); 柔性电子材料基因工程湖南省重点实验室开放基金 (No. 202005)
Supported by China-Montenegro 3rd Science & Technology Exchange and Cooperation Project (No. 3-7); Double First-class Scientific Research International Cooperation Expansion Project of Changsha University of Science and Technology (No. 2019ic18); Open Research Fund of Hunan Provincial Key Laboratory of Flexible Electronic Materials Genome Engineering (No. 202005)

* 通信联系人, E-mail: 1051815150@qq.com

(MdFPN Tiny YOLOv3) is proposed. The model constructs multi-directional feature pyramid network (MdFPN) instead of simple concatenation to adequately complete the extraction and fusion of multi-layer semantic information. The deep separable convolution is used instead of standard convolution to effectively reduce the complexity of the network and improve the real-time performance of detection. The complete IOU loss (CIOU loss) is used instead of MSE as the regression loss function, which greatly improves the regression accuracy of bounding box. The results which MTYOLO is tested on PASCAL VOC and COCO datasets show that the mAP of the improved model can reach 78.7% and the detection speed can reach 205 fps.

Key words: object detection; feature pyramid network; deep separable convolution; complete IOU loss

1 引 言

目标检测是机器视觉领域的研究热点之一,因其快速、准确、可识别性高等特点广泛应用于无人驾驶、行人检测和视频监控等领域。传统的目标检测方法^[1]主要是根据先验知识建立数学模型并求解该模型得到检测结果,而此种人工设计特征加浅层分类器的方案只能提取细节信息较少的低层特征,检测精度和鲁棒性受到极大限制^[2]。由于卷积神经网络能够自动学习到更深层次的特征表达,与传统方法相比,学习到特征的鲁棒性更强、精度与速度更高,因此广泛应用于目标检测领域。

目前基于卷积神经网络的检测方法主要包括两大类:基于区域的双阶段目标检测方法,如 R-CNN 系列^[3-5];基于回归的单阶段目标检测方法,如 RetinaNet^[6]、SSD^[7]、YOLO^[8-10]等。双阶段检测器使用生成建议网络提取特征后送入分类器得到类别,而单阶段检测器不需要额外的区域分类步骤,直接对特征映射图上每个位置的目标进行分类预测。双阶段检测器较单阶段检测器检测精度更高但实时性明显降低,故 YOLO 作为检测精度与速度均衡发展的单阶段检测器得到更加广泛的研究与应用。YOLO 是一种端到端的实时目标检测框架,将物体检测作为回归问题进行求解,不需要预先通过生成建议网络(RPN)得到感兴趣区域(ROI),而是直接将图片送入特征金字塔网络^[11](FPN)提取特征经过分类回归器得到输出。轻量级目标检测模型 Tiny YOLOv3 较 YOLOv3 减少了大量的特征提取卷积层与残差块,同时仅使用 FPN 简单级联,忽略了更深层次

的语义特征及梯度消失问题,损失了较多的有效信息从而检测精度明显下降;其次,网络模型尺寸虽然随着卷积层的减少而降低,但仍然存在计算量(FLOPs)冗余影响检测速度的问题;最后,在计算损失函数时未考虑坐标点之间的相关性及关键指标 IOU (Intersection over Union)的影响,导致边界框定位不够准确,回归精度明显下降。

本文针对 Tiny YOLOv3 存在的特征提取不充分、检测速度受限、边界框回归精度低的问题,通过改进金字塔特征提取网络、使用深度可分离卷积与引入 CIOU loss 函数,提出实时高精度 MTYOLO 检测模型。

2 Tiny YOLOv3 目标检测原理

Tiny YOLOv3 以回归的方式训练由 7 层类 Darknet-19 构成的主干网络,同时通过 K-means 聚类算法对样本数据集进行聚类学习,分别在 13×13 和 26×26 两个尺度上得到两组 anchor box 作为先验框进行后续的回归操作。在针对多目标问题时,使用 logistic 分类器取代 softmax 分类器,解决了同一物体属于多个类别的表征问题。

在 Tiny YOLOv3 中,将输入图像送入卷积神经网络提取特征并划分为 $S \times S$ 的网格,每个网格包含 3 个预测框,每个预测框产生 4 个用于回归的参数 (x, y, w, h) 与 $1 + n$ 个用于分类的参数 $(c, p_1, p_2, \dots, p_n)$,送入预测层产生预测框返回值参数 $(o_x, o_y, o_w, o_h, o_c, o_{p1}, o_{p2}, \dots, o_{pn})$,可以表示为:

$$o_x = \sigma(x) + c_x, \quad (1)$$

$$o_y = \sigma(y) + c_y, \quad (2)$$

$$o_w = p_w e^w, \quad (3)$$

$$o_h = p_h e^h, \quad (4)$$

$$o_c = \sigma(c) = P_r \times \text{IOU}_{\text{pred}}^{\text{gt}}, \quad (5)$$

$$o_{pi} = \sigma(p_i), \quad (6)$$

式中: o_x 、 o_y 、 o_w 、 o_h 为预测框的中心点坐标和宽高, $\sigma(\cdot)$ 代表 sigmoid 函数, $\sigma(x)$ 、 $\sigma(y)$ 为预测框中心对应网格左上角的横坐标、纵坐标, c^x 与 c^y 为当前网格相对于左上角网格偏移的网格数, p_w 与 p_h 为边界框的宽度与高度, e^w 与 e^h 为横纵方向上的缩放因子。 c 代表该边界框的置信度, 当网格中存在目标时 P_r 为 1, 否则为 0, $\text{IOU}_{\text{pred}}^{\text{gt}}$ 为预测框和真实框的交并比; b_{pi} 为 n 个类别通过分类器的概率, 当 b_{pi} 大于 0.5 时, 表示该边界框负责的目标属于此类。最后, 将多个边界框的参

数通过非极大值抑制 (Non-maximum suppression, NMS) 运算得到最终的位置信息和类别预测值。

3 改进的 MTYOLO 检测模型

改进的 MTYOLO 检测模型包括主干网络、特征融合网络、预测层与输出层。如图 1 所示, 主干网络将输入图像经过一系列卷积操作得到不同感受野的特征图, 特征融合网络将生成的各层特征图通过多向级联得到两种尺度的特征向量, 送入预测层产生大量带有位置与类别信息的预测框, 最后由 NMS 得到输出层的检测结果。

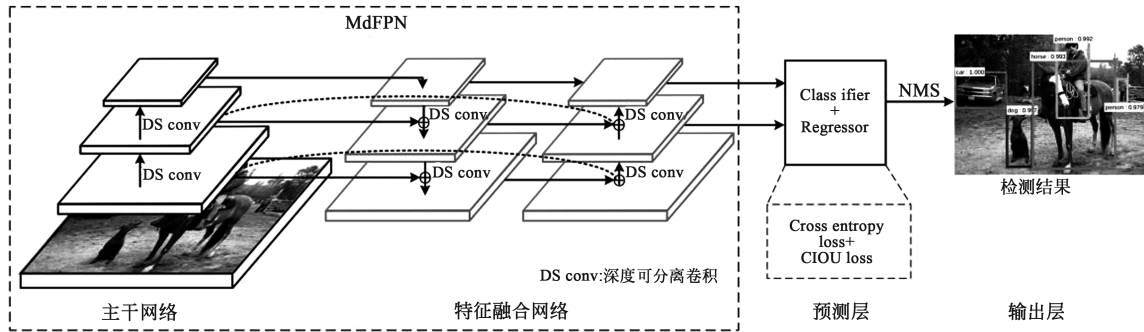


图 1 MTYOLO 检测模型

Fig.1 MTYOLO test model

3.1 MdFPN 特征提取网络

特征提取网络是影响目标检测算法性能的关键因素之一。Tiny YOLOv3 使用传统的 FPN 采用 bottom-up 跨层连接方式保留了高层特征的语义信息与低层特征的细节信息, 提升了小物体的检测效果。然而传统的 FPN 只能学习相邻尺度的特征, 固地接收单向信息流, 导致网络过多关注低层特征的优化, 累计更多的梯度从而不能充分学习高层特征。

MTYOLO 采用 MdFPN (Multi-directional feature pyramid network) 结构通过在传统的 FPN 提取网络上添加双向交叉连接及侧向级联后充分地进行多层间不同分辨率、不同语义特征的融合, 缓解了原网络在特征提取过程中对边缘信息的忽略, 极大地共享了网络权重参数, 提高了网络敏感度。此外, 由于偶数卷积核因其感受野不对称会在特征图中累计位置像素的偏移, 致使

空间信息受到严重侵蚀, 故选择奇数卷积核作为核函数。与较大的卷积核相比, 3×3 的卷积核可在保证获得相同感受野和更多特征信息的同时消耗更少的计算资源。因此, MdFPN 选择堆叠多个 3×3 卷积核代替较大的卷积核以便获得更好的非线性判决网络。算法的整体网络结构如图 2 所示。

(1) 构建了 7 层的主干网络, 使用 6 个卷积模块 (包含卷积、标准化和 Leaky ReLU 激活 3 个过程) 与最大池化提取复杂背景下的物体信息, 将输入图像 X_1 经过 32 倍降采样后得到 Y_6 , 保证在提取到 (208, 104, 52, 26, 13, 7 pixel) 各尺度特征图的情况下控制网络层数。

(2) 添加图像融合与跨层连接运算。将主干网络特征图 $Y_2 \sim Y_6$ 与跨层上采样后的 $Y_9 \sim Y_{11}$ 进行初次融合得到特征融合层 $C_1 \sim C_5$ 。

(3) 利用 3 个深度可分离卷积与 2 个 3×3 卷

积完成对初步融合图像特征的再次提取,获得更多有效的特征。此时,由于 C_4 、 C_5 特征通道数过大故使用普通卷积防止计算量骤增。

(4) 移除 Tiny YOLOv3 中单一特征融合与预测的卷积层,通过对(1)、(2)、(3)中各层特征图进行最大池化、深层卷积与特征融合操作得到不

同尺度的输出向量 $C_6 \sim C_9$,使其既包含池化聚焦的空间信息与前一层融合变换后的局部特征,又包含本层与后一层的全局特征。

(5) 将特征融合后的 Y_{12} 、 Y_{13} 分别卷积后得到 output1、output2,直接送入预测层用于多分类和回归操作,不断训练迭代模型,得到最终的检测结果。

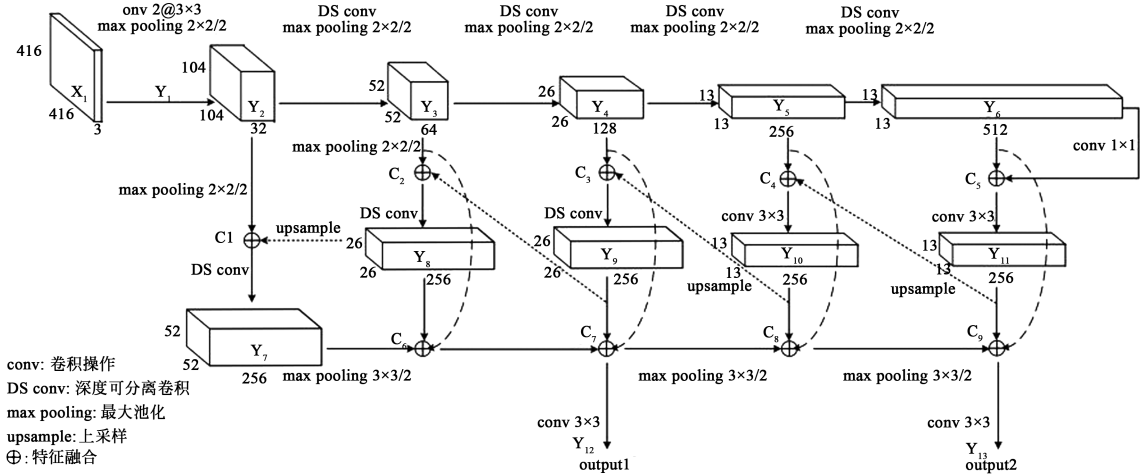


图 2 MdFPN 特征提取网络

Fig.2 MdFPN feature extraction network

表 1 Tiny YOLOv3 与模型 A 在 VOC2007/COCO 数据集上的 mAP

Tab.1 mAP of Tiny YOLOv3 and model A on VOC2007 / COCO dataset

模型	mAP/%
Tiny YOLOv3	58.4/33.1
模型 A	71.7/47.2

将仅使用 MdFPN 代替 Tiny YOLOv3 网络的简单级联,不改变其他结构的检测模型记为模型 A。由表 1 可以看出,相对于传统的 Tiny YOLOv3 网络,模型 A 在 VOC2007 和 COCO 数据集上的 mAP 分别提升了 13.3% 和 14.1%,表明 MdFPN 可有效融合多层信息,提高检测精度。

3.2 深度可分离卷积

传统的卷积计算每一步都需要关注所有通道

的计算区域从而累积较多的参数。针对深层网络导致模型计算复杂度过高与检测速度受限的问题,可采用模型裁剪^[12]与使用深度可分离卷积构建网络来减少参数占用内存设计轻量级模型,但由于模型裁剪可能会带来不可逆转的细节丢失,故选择深度可分离卷积对 MTYOLO 模型进行简化。

深度可分离卷积将标准卷积分解成两层。一层用于滤波,在输入的每个通道上独立执行空间卷积;一层用于组合,将深度卷积的通道混合后使用点卷积输出。如图 3 所示,输入图像为 $D_F \times D_F \times M$,输出图像为 $D_W \times D_H \times N$,当采用标准卷积核 $D_K \times D_K \times M \times N$ 时,此卷积层的计算量为 $D_K \times D_K \times M \times N \times D_F \times D_F$;而使用深度可分离卷积时卷积层的计算量为 $D_K \times D_K$ 深度卷积和 1×1 点卷积的计算量之和。故采用深度可分离卷积与使用标准卷积的网络计算量之比为:

$$\frac{D_K \times D_K \times M \times D_W \times D_H + M \times N \times D_W \times D_H}{D_K \times D_K \times M \times N \times D_W \times D_H} = \frac{1}{N} + \frac{1}{D_K^2} \quad (7)$$

由于实际应用中 $N \gg D_k^2$, 故使用深度可分离卷积代替标准卷积的计算量下降为原来的 $\frac{1}{D_k^2}$ 。

以 MdFPN 的第二层结构为例, 网络中输入特征图为 $104 \times 104 \times 32$, 输出特征图为 $104 \times 104 \times 64$ 。由(7)式可知, $D_k = 3$ 时, 网络参数可被压缩为原来的 $1/9$ 。因此, MdFPN 使用多个深度可分离卷积可构建紧凑的模型, 从而提升

检测速度。从理论上来说, 模型的时间复杂度可以用来衡量检测速度, 即一张图像送入模型进行一次完整的前向传播所需的浮点运算个数 (FLOPs)。由于普通卷积相比于深度可分离卷积而言, 在进行卷积运算时的数据调用的逻辑更加繁琐, 且在对每个特征图的卷积操作完成后还会对这些特征图进行累加操作, 大量的中间数据需要更多的寄存器来存储, 故在前向传播时将花费更多的时间。

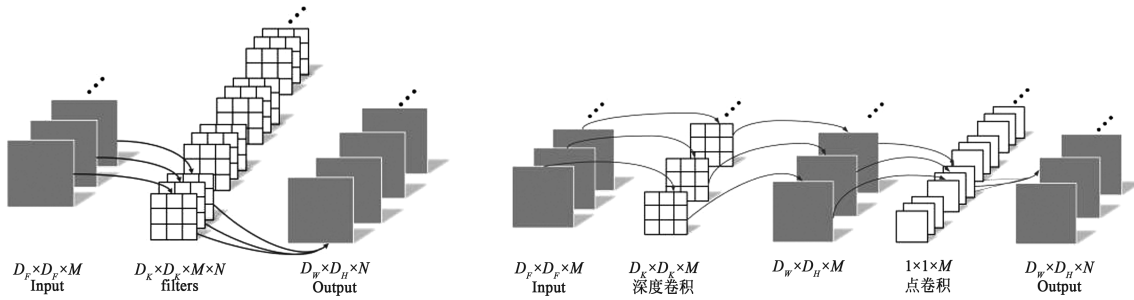


图 3 标准卷积和深度可分离卷积

Fig.3 Standard convolution and depthwise separable convolution

表 2 Tiny YOLOv3 与模型 B 的参数总量、FLOPs 及 FPS

Tab.2 Total parameters, FLOPs and FPS of Tiny YOLOv3 and Model B

模型	参数总量/M	FLOPs/billion	FPS/GPU/CPU
Tiny YOLOv3	8.71	5.52	160/24.6
模型 B	7.64	1.74	225/36.7

将仅使用深度可分离卷积代替部分标准卷积, 其他网络结构均不改变的检测模型记为模型 B。由表 2 可知, 相比传统的 Tiny YOLOv3 网络, 模型 B 的参数总量下降了 12.3%, 模型尺寸减小了 68.5%, FPS 在 GPU 和 CPU 上分别提升了 40.6% 和 49.2%。结果表明, 使用深度可分离卷积代替标准卷积极大地节省了计算资源, 提高了检测速度。

3.3 损失函数的优化

Tiny YOLOv3 仅使用 MSE 函数作为网络的回归函数, 独立地计算边界框 4 个坐标点的损失, 未考虑坐标点之间的相关性与距离损失的尺度不变性, 故存在收敛速度慢和回归不准确的问题。针对 MSE 函数存在的问题, 充分考虑边界框回归的三要素: 重叠区域、中心点距离和长宽比, MTYOLO 采用 CIOU 损失代替 MSE, 通过

引入 IOU (图 4)、添加惩罚项 av 来最小化两个边界框, 进一步明确了梯度优化的方向。MSE 函数与 CIOU 损失可描述为式 (8)、(9):

$$L_{\text{MSE}} = \sum_{m \in \{x, y, w, h\}} (o_m, o_m^{\text{gt}})^2, \quad (8)$$

$$L_{\text{CIOU}} = 1 - \text{IoU} +$$

$$\frac{d^2(o, o^{\text{gt}})}{c^2} + av = 1 - \text{IoU} + \frac{L_{\text{MSE}}}{c^2} + av, \quad (9)$$

式中, gt 代表真实框, o 与 o^{gt} 分别为预测框与真实框的中心点, d 表示欧氏距离; c 表示预测框和真实框最小外接矩形的对角线距离, 与 IOU 共同实现对预测框相交面积的约束; a 是一个正的权衡参数, 控制梯度下降方向; 而 v 约束两框的相交比例, 使模型能够向区域重叠更密集的方向优化, 进一步提高检测结果。将 a 与 v 表示如下:

$$a = \frac{v}{(1 - \text{IoU}) + v}, \quad (10)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2. \quad (11)$$

对于分类损失与置信度损失,使用二值交叉熵损失函数,则总损失可以表示为式(12):

$$L = \sum_{i=0}^{S \times S} \left[\sum_{j=0}^2 (\lambda_{coord} I_{ij}^{obj} L_{CIOU} + I_{ij}^{obj} L_{conf} + \lambda_{noobj} I_{ij}^{noobj} L_{conf}) + \sum_{c \in classes} I_{ij}^{obj} L_{cls} \right], \quad (12)$$

$$L_{conf} = -t_c \log o_c - (1 - t_c) \log(1 - o_c), \quad (13)$$

$$L_{cls} = -t_{pi} \log o_{pi} - (1 - t_{pi}) \log(1 - o_{pi}), \quad (14)$$

其中, i 代表不同尺度, I_{ij}^{obj} 与 I_{ij}^{noobj} 分别表示第 i 个网格的第 j 个边界框是否负责该目标。 λ_{coord} 与 λ_{noobj} 分别为定位误差与分类误差的权重。 t_{pi} 与 b_{pi} 分别表示真实目标预测目标的类别概率, t_c 与 b_c 分别表示真实目标预测目标的置信度。

表 3 Tiny YOLOv3 与模型 C 在 VOC2007/COCO 数据集上的 mAP

Tab. 3 mAP of Tiny YOLOv3 and Model C on VOC2007/COCO dataset

模型	mAP/%
Tiny YOLOv3	58.4/33.1
模型 C	68.7/41.3

将仅使用 CIOU 损失代替 MSE, 其他网络结构均不改变的检测模型记为模型 C。由表 3 可以看出, 相对于传统的 Tiny YOLOv3 网络, 模型 C 在 VOC 和 COCO 数据集上的 mAP 分别提升了 10.3% 和 8.2%, 表明使用 CIOU loss 可提高回归

框的检测精度进而提升 mAP。

3.4 模型传播过程分析

模型训练过程分为前向计算及反向传播过程。在两个过程交叠往复中持续对权值参数保持更新, 使损失函数结果达到最小。在 MTYOLO 前向传播的过程中第 i 个卷积层可以表示为式(15):

$$\mathbf{Y}_i = F(\mathbf{X}_i) = F(\mathbf{X}_{i-1} * \mathbf{w}_i + \mathbf{b}_i), \quad (15)$$

式中, $\mathbf{X}_i$ 、 $\mathbf{Y}_i$ 分别代表输入与输出张量, $\mathbf{w}_i$ 、 $\mathbf{b}_i$ 表示第 i 层的权重信息与偏移量, $F(\cdot)$ 是激活函数。将输入图像通过特征提取等操作得到网络最后一层的输出参数 $\mathbf{Y}_i$ 。进行后向传播时, 第 i 层的损失函数的梯度信息可以表示为式(16)。其中, $\odot$ 为逐元素乘积 (Hadamard 乘积)。

$$\delta i = \frac{\partial L}{\partial \mathbf{Y}_i} = \frac{\partial L}{\partial \mathbf{Y}_{i+1}} \cdot \frac{\partial \mathbf{Y}_{i+1}}{\partial \mathbf{Y}_i} =$$

$$\delta_{i+1} * (\mathbf{w}_{i+1})^T \odot F(\mathbf{X}_{i-1} * \mathbf{w}_i + \mathbf{b}_i), \quad (16)$$

此外, $L_{conf}(L_{cls})$ 对 $\mathbf{Y}_j$ 的梯度信息可以表示为式(17), 则参数 o_j 向着 t_j 方向梯度更新时可以得到理想的置信度与分类误差。

L_{CIOU} 对 $\mathbf{Y}_j$ 的梯度信息由式(18)可得, 主要由 IOU、 d^2 、 c 对输出坐标的偏导组成。结合图 4 分

$$\frac{\partial L_{conf(cls)}}{\partial \mathbf{Y}_j} = \frac{\partial L_{conf(cls)}}{\partial o_j} \cdot \frac{\partial o_j}{\partial \mathbf{Y}_j} = \left(\frac{-t_j}{o_j} + \frac{1-t_j}{1-o_j} \right) \sigma(\mathbf{Y}_j) (1 - \sigma(\mathbf{Y}_j)) = o_j - t_j, \quad (17)$$

$$\frac{\partial L_{CIOU}}{\partial \mathbf{Y}_j} = \frac{\partial L_{CIOU}}{\partial o_m} \frac{\partial o_m}{\partial \mathbf{Y}_j} = \left(-\frac{\partial IOU}{\partial o_m} + \frac{1}{c^2} \frac{\partial L_{MSE}}{\partial o_m} - \frac{2L_{MSE}}{c^3} \frac{\partial c}{\partial o_m} + v \frac{\partial \alpha}{\partial o_m} + a \frac{\partial v}{\partial o_m} \right) \sigma(\mathbf{Y}_j) (1 - \sigma(\mathbf{Y}_j)). \quad (18)$$

析可知, IOU 越大, $\frac{d^2(o, o^{gt})}{c^2} \rightarrow 0$ (两框中心点重合), $a \rightarrow 1$, 此时回归损失的求解主要取决于 v 。

因此, 对约束参数 v 进行求导得到式(19)、(20), 其中, 对于每一轮训练的 w^{gt} 与 h^{gt} 均为已知常数。由式(19)、(20)可知, 若预测框的长和宽 $\subset [0, 1]$ 时, $w^2 + h^2$ 的值会很小, 可能导致梯度爆炸, 因此应将其替换成 1。

$$\frac{\partial v}{\partial w} = -\frac{8}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right) \times \frac{h}{w^2 + h^2}, \quad (19)$$

$$\frac{\partial v}{\partial h} = \frac{8}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right) \times \frac{w}{w^2 + h^2}. \quad (20)$$

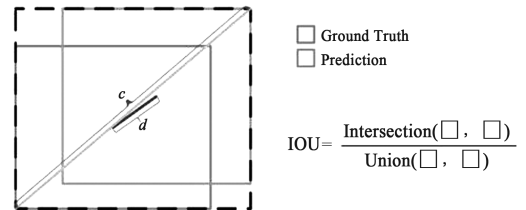


图 4 IOU、 c 及 d 的示意图。

Fig.4 Schematic diagram of IOU, c and d .

4 实验结果及分析

研究基于上述理论与消融实验的分析,在网络结构上将 Tiny YOLOv3 原有的类 Darknet-19 替换为 MdFPN 特征提取网络,使用深度可分离卷积代替部分标准卷积,并在预测层利用 CIOU loss 取代 MSE 函数,构造新的 MTYOLO 检测模型。实验环境为:Ubuntu 18.04;TensorFlow Version: 1.13; CUDA Version: 10.1; 内存: 32 G; GPU: Tesla P4。实验采用的数据集为 PASVAL VOC 2007/2012 和 COCO 2014。VOC 数据集由 20 种不同类型的物体标注的自然图像组成,包含训练验证集 train_val 共 16 551 张,测试集 test 4 952 张;COCO 数据集包含了 80 个种类的目标,包含训练验证集 train_val 共 123 287 张,测试集 test 40 775 张。

4.1 训练过程

MTYOLO 输入图片大小为 416×416 ,批处理大小为 64,最大迭代次数为 20 000,初始学习率为 0.001,动量为 0.9,权重衰减率为 0.000 5。训练过程中使用交叉熵损失(Cross entropy loss)和 CIOU loss 对网络权重参数进行反向随机梯度下降更新权重,实现网络的有效学习。当迭代次数到达 5 000 次,学习率降为 0.000 1;当迭代次数达到 10 000 次,学习率进一步降为 0.000 01。

图 5 为使用 VOC 2007+2012 数据集进行训练验证的回归误差(val_loss)与每 2 500 轮计算一次得到的分类准确率(Accuracy)得到的曲线。由图可知,经过 20 000 次迭代后,val_loss 逐步

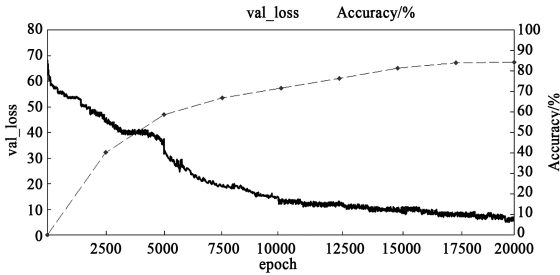


图 5 回归误差与分类正确率

Fig.5 Regression error and classification accuracy

趋于稳定,最终下降到 5.5 左右,分类正确率上升至 84.2%,与 Tiny YOLOv3 相比较,回归误差下降约 50.9%,分类正确率上升约 9.3%,表明训练得到的模型较为理想。同时,当 MTYOLO 的 mAP 达到 56.9%时,仅耗时 1 036 800 s,较 Tiny YOLOv3 减少了近 25.9%,训练速度得到了较大的提升。

4.2 测试结果

为了评估改进算法的准确率,对 MTYOLO 及马立^[13-16]等人改进的 Tiny YOLOv3 模型分别在 VOC 2007 与 COCO 数据集上进行测试,结果如表 4 所示。

由表 4 可以看出,MTYOLO 较 Tiny YOLOv3 在 VOC 和 COCO 数据集上 mAP 分别提升了 25.3%和 27.5%,模型尺寸下降了 15.2%,FLOPs 下降了 5.6%,FPS 在 GPU 与 CPU 分别提升了 25.6%和 36.2%。并且,与马立等人改进的模型及 YOLO 系列最新版本 v4、v5 相比,在损失较小精度的情况下,检测速度更理想,更适用于嵌入式平台。因此,通过使用 MdFPN 网络、深度可分离卷积和 CIOU 损失构建目标检测模型,能

表 4 七种模型在 VOC/COCO 上的 mAP 和模型尺寸、FLOPs、FPS

Tab.4 mAP on VOC / COCO and Model size, FLOPs, FPS for seven models

模型	mAP/%	模型尺寸/M	FLOPs/billion	FPS(GPU/CPU)
Tiny YOLOv3	58.4/33.1	33.4	5.52	160/24.6
Improved algorithm of Tiny YOLOv3[13]	68.1/—	35.1	5.73	—/32.7
Tinier-YOLO[14]	65.7/34.0	8.9	2.563	—/25.1
YOLO Nano[15]	69.1/—	4.0	4.57	—/26.9
YOLOv4[16]	—/62.8	244	—	96/—
YOLOv5	—/50.1	27	—	167/—
MTYOLO	83.7/60.6	28.32	5.21	201/33.5

在提升检测精度的基础上,有效地减少网络参数和计算量,保证检测速度提升到一个更高的水平。

5 结 论

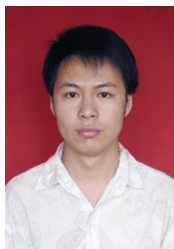
基于 Tiny YOLOv3 改进的检测模型 MTYOLO,通过构造多向特征金字塔与深度可分离卷积

的特征提取网络代替标准卷积堆叠级联的网络结构,并使用 CIOU 损失代替 MSE 作为该模型的回归损失函数,在保证检测速度较理想的情况下提高了 mAP。在 PASCAL VOC 和 COCO 数据集上与其他轻量级的目标检测算法进行对比实验,结果表明,MTYOLO 具有更高的 mAP,并且取得了较快的检测速度。

参 考 文 献:

- [1] 王春哲,安军社,姜秀杰,等.基于卷积神经网络的候选区域优化算法[J].中国光学,2019,12(6):1348-1361.
WANG C Z, AN J S, JIANG X J, *et al.* Region proposal optimization algorithm based on convolutional neural networks [J]. *Chinese Optics*, 2019, 12(6): 1348-1361. (in Chinese)
- [2] MINAR M R, NAHER J. Recent advances in deep learning: an overview [J]. *arXiv*:1807.08169, 2018.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation [C]//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus: IEEE, 2014.
- [4] GIRSHICK R. Fast R-CNN [C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago: IEEE, 2015: 1440-1448.
- [5] 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.
- [6] LIN T Y, GOYAL P, GIRSHICK R, *et al.* Focal loss for dense object detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 318-327.
- [7] LIU W, ANGUELOV D, ERHAN D, *et al.* SSD: single shot MultiBox detector [C]//*Proceedings of the 14th European Conference on Computer Vision*. Amsterdam, The Netherlands: Springer, 2016: 21-37.
- [8] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You only look once: unified, real-time object detection [C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 779-788.
- [9] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017: 6517-6525.
- [10] REDMON J, FARHADI A. YOLOv3: an incremental improvement [J]. *arXiv*:1804.0276, 2018.
- [11] LIN T Y, DOLLÁR P, GIRSHICK R, *et al.* Feature pyramid networks for object detection [C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017.
- [12] LUO J H, WU J X, LIN W Y. ThiNet: a filter level pruning method for deep neural network compression [C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017.
- [13] 马立,巩笑天,欧阳航空.Tiny YOLOv3 目标检测改进[J].光学精密工程,2020,28(4):988-995.
MA L, GONG X T, OUYANG H K. Improvement of Tiny YOLOV3 target detection [J]. *Optics and Precision Engineering*, 2020, 28(4): 988-995. (in Chinese)
- [14] FANG W, WANG L, REN P M. Tinier-YOLO: a real-time object detection method for constrained environments [J]. *IEEE Access*, 2019, 8: 1935-1944.
- [15] WONG A, FAMUORI M, SHAFIEE M J, *et al.* YOLO Nano: a highly compact you only look once convolutional neural network for object detection [J]. *arXiv*:1910.01271, 2019.
- [16] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [J]. *arXiv*:2004.10934, 2020.

作者简介:



白 创(1983—),男,博士,讲师,2015年于华中科技大学获得博士学位,主要从事机器视觉与模式识别技术、深度神经网络学习算法方面的研究。E-mail: 154317586@qq.com



王英杰(1994—),女,硕士研究生,2017年于天津师范大学获得学士学位,主要从事深度学习、机器视觉方面的研究。E-mail: 1051815150@qq.com