



基于可变形卷积的改进YOLO目标检测算法

黄凤琪, 陈明, 冯国富

(上海海洋大学 信息学院, 上海 201306)

摘要: 针对YOLO目标检测算法存在边界框定位不准确及对小目标检测精度低的问题, 提出一种改进的YOLO目标检测算法dcn-YOLO。使用k-means++算法聚类出更符合数据集尺寸的锚盒, 以降低初始点对聚类结果的影响并加快网络训练收敛速度。构建残差可变形卷积模块res-dcn, 分别采用将其嵌入YOLO第一特征提取头模块中和替换3个YOLO特征提取头模块的方式, 构建两种改进的dcn-YOLO算法, 使网络可以自适应地学习特征点的感受野, 从而对不同尺寸和形状的目标提取更有效的特征, 提高检测精度。在VOC数据集上的实验结果表明, 该算法能有效提高目标检测精度, mAP达到82.6%, 相比YOLO、SSD、Faster R-CNN, 分别高出了2.1、5.2、9.4个百分点。

关键词: YOLO算法; 目标检测; 感受野; 可变形卷积; k-means++算法

开放科学(资源服务)标志码(OSID):



中文引用格式: 黄凤琪, 陈明, 冯国富. 基于可变形卷积的改进YOLO目标检测算法[J]. 计算机工程, 2021, 47(10): 269-275, 282.

英文引用格式: HUANG F Q, CHEN M, FENG G F. Improved YOLO object detection algorithm based on deformable convolution[J]. Computer Engineering, 2021, 47(10): 269-275, 282.

Improved YOLO Object Detection Algorithm Based on Deformable Convolution

HUANG Fengqi, CHEN Ming, FENG Guofu

(Institute of Information Technology, Shanghai Ocean University, Shanghai 201306, China)

[Abstract] The YOLO algorithm for object detection is limited by the inaccurate positioning of the boundary box and the low detection accuracy for small objects. To address the problem, an improved YOLO algorithm, dcn-YOLO, is proposed based on deformable convolution for object detection. The algorithm employs the K-means++ to cluster anchor boxes that are more in line with the size of data set, so as to reduce the impact of initial points on clustering results and speed up the convergence of network training. Then, a residual deformable convolution module, res-dcn, is constructed. Two improved dcn-YOLO algorithms are derived by embedding res-dcn in the first YOLO feature extraction head module or replacing three YOLO feature extraction head modules with res-dcn, so the network can adaptively learn the receptive field of feature points and extract more effective features for objects of different sizes and shapes, increasing the detection accuracy. Experimental results on VOC data sets show that the propose algorithm can effectively improve the object detection accuracy. Its mAP reaches 82.6%, which is 2.1 percentage points higher than that of YOLO, 5.2 percentage points higher than that of SSD and 9.4 percentage points higher than that of Faster R-CNN.

[Key words] YOLO algorithm; object detection; receptive field; deformable convolution; k-means++ algorithm

DOI: 10.19678/j.issn.1000-3428.0059096

0 概述

目标检测需判断图像中目标的类别及不同目标的边界框位置, 是计算机视觉里的主要任务, 被广泛应用于视频监控^[1-2]、行人检测^[3-4]等领域。传统的目标检测方法依赖人工选择的特征进行计算, 如HOG^[5]、SIFT^[6]、Haar^[7]等, 但计算复杂, 运算速度慢,

且对目标形变和背景复杂变化的图像检测鲁棒性较差。而神经网络依赖其强大的特征提取和特征表示功能, 被广泛应用于目标检测任务。

目前主流的目标检测算法有以YOLO^[8-10]系列算法、SSD^[11-12]系列算法为代表的单阶段算法和以R-CNN^[13-14]系列算法为代表的两阶段算法。单阶段算法将目标检测过程看作回归问题来处理, 使用一个统

基金项目: 国家重点研发计划(2018YFD0701003); 上海市科技创新行动计划(6391902902)。

作者简介: 黄凤琪(1995—), 男, 硕士研究生, 主研方向为计算机视觉; 陈明, 教授、博士; 冯国富, 副教授、博士。

收稿日期: 2020-07-29 **修回日期:** 2020-09-27 **E-mail:** 2607546441@qq.com

一的深度神经网络进行特征提取和目标分类及边界框回归,实现了端到端的推理,具有较快的检测速度,但检测精度没有两阶段算法高;而两阶段目标检测算法先使用区域建议网络(Region Proposal Network, RPN)提取出感兴趣区域,即含有目标的区域,再用深度神经网络对感兴趣区域进行分类及边界框回归,具有更高的检测精度,但其检测速度太慢,达不到实时的要求。

目前对目标检测算法的改进方法主要包括:使用更复杂更深层的骨干网络,以提取更有效的特征;融合多尺度特征,使浅层特征传递到深层,从而提高定位精度;加入新型网络模块,以增强网络的特征提取能力等。FU等^[12]提出在SSD检测框架的基础上,采用更深的ResNet-101网络来进行特征提取,并且采用了反卷积层,引入额外的大量语义信息,改进了SSD算法对于小物体的检测能力,但使用更深的基础网络和反卷积层,增加了大量的计算,使得检测速度大幅降低,无法满足实时的要求。JIANG等^[15]通过在SSD网络中引入空洞卷积,扩大特征点的感受野,优化网络的特征提取能力,从而提高了网络的检测精度,但增加了模型的复杂度,导致检测速度的降低。XU等^[16]提出在YOLO算法中引入注意力机制和改进的特征融合方法,以提高网络的定位精度,形成了Attention-YOLO算法,但检测速度较慢。

以上研究从不同角度提升了目标检测算法的性能。但单纯地使用反卷积或空洞卷积等操作,扩大了特征点的感受野,可能增加背景区域对特征点信

息的影响,导致目标定位不准或把背景误检为目标。对不同尺寸和形状的目标自适应地提取感受野,可以有效地避免这个问题。DAI等^[17-18]提出了可变形卷积(Deformable Convolution),通过在常规卷积操作上增加平行网络预测常规卷积采样点的偏移,使得每个采样点都具有一定的偏移量,可以学习到自适应的感受野,从而提高网络对不同尺寸和形状的物体进行特征提取的能力。

本文在YOLO算法的基础上,提出基于可变形卷积的改进dcn-YOLO算法。通过构建残差可变形卷积模块res-dcn,并将其嵌入YOLO特征提取头模块和替换YOLO特征提取层,形成两种改进的dcn-YOLO模型,使网络能够自适应地学习不同尺寸和形状目标的特征,提高目标检测的精度。

1 YOLO算法

YOLO是一个端到端的深度神经网络模型,其网络结构如图1所示,主要由Darknet53主干网络、3个FPN^[20]多尺度特征提取头加上3个YOLO多尺度检测层组成,其中所有卷积模块都由卷积层、Batch Normalization(BN)层和leaky激活层组成。Residual为残差模块,将上一模块的输入累加到当前层,以实现浅层特征向深层的传递,提高定位精度。整个特征提取网络由常规卷积层组成,使用 3×3 的卷积核进行特征提取,采用 1×1 的卷积核进行通道数压缩,动用步长为2的卷积层来减小特征图分辨率,替代池化操作。

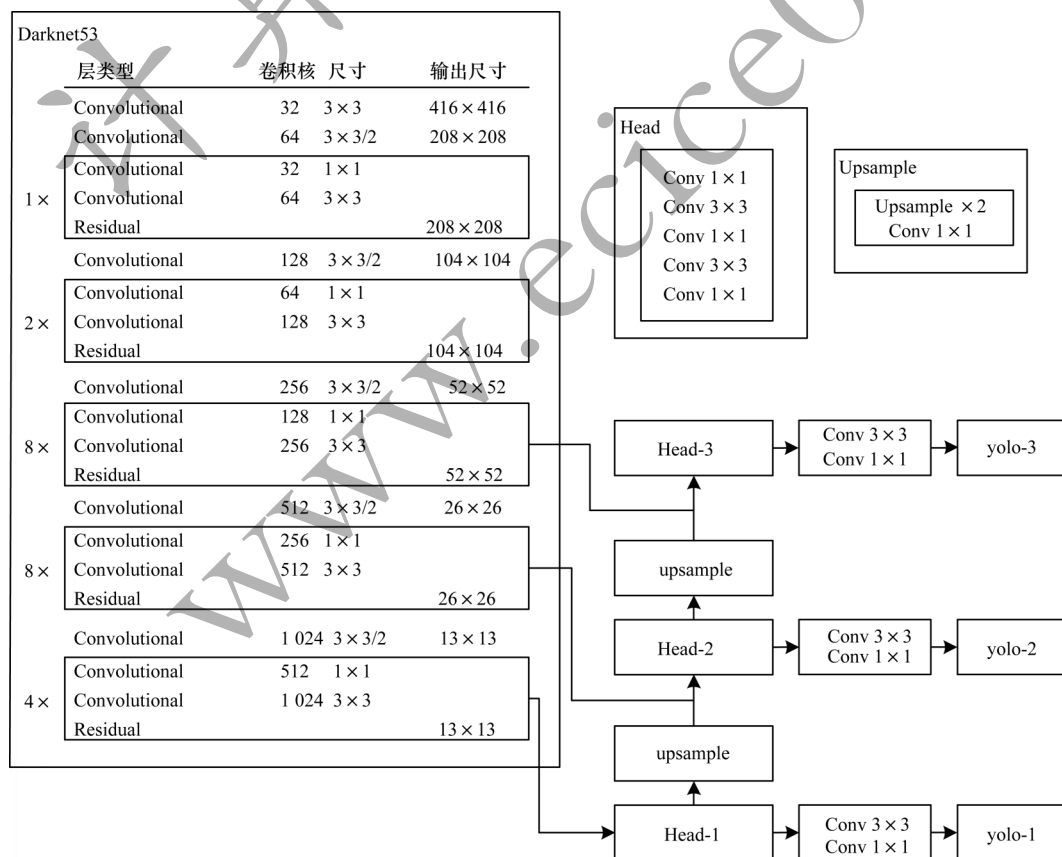


图1 YOLO网络结构

Fig.1 Network structure of YOLO

YOLO的核心思想是利用整个图像作为网络的输入,将其分成 $S \times S$ 个网格,如果某个目标的中心落在这个网格中,则该网格就负责检测这个目标,如图2所示。每个网格检测 B 个目标,每个目标需要预测边界框坐标 (x, y, w, h) ,其中 (x, y) 为目标边界框中心坐标, (w, h) 为边界框的宽和高;一个置信度得分(该位置有目标的置信度)和 C 个类别得分(该目标属于某类的概率),所以YOLO检测层输出张量大小是 $S \times S \times B \times (4 + 1 + C)$ 。以输入图像尺寸为416像素 \times 416像素和VOC数据集为例,3个多尺度YOLO检测层尺寸分为 $13 \times 13 \times 75, 26 \times 26 \times 75, 52 \times 52 \times 75$,分别负责对大、中、小型目标进行检测,提高了YOLO的检测性能。

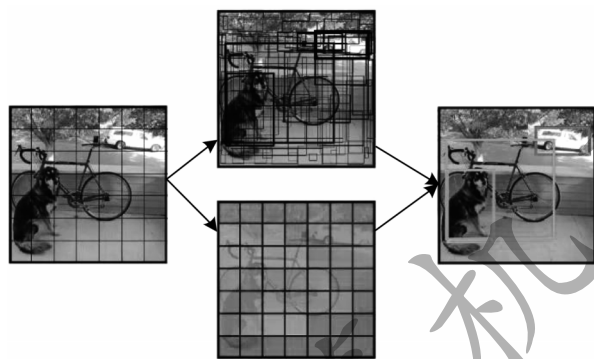


图2 YOLO的检测过程

Fig.2 Detection process of YOLO

虽然YOLO使用了FPN结构进行多尺度检测,有

效提高了对目标的检测能力,但同一特征图上的特征点具有相同的感受野,对形变较大的目标提取的特征容易受到背景区域的干扰,导致检测边界框定位不准,同时背景干扰也会导致小目标检测精度不高。

2 改进的dcn-YOLO模型

卷积网络对大尺寸多形变目标的建模存在固有的缺陷,这是因为卷积网络只对输入特征图的固定位置进行采样^[21]。例如,在同一层特征图中,所有的特征点的感受野是一样的,但由于不同位置可能对应着不同尺度或变形的物体,因此能够对尺度或者感受野大小进行自适应学习精确定位的目标。

可变形卷积可以提高模型对形变目标的建模能力,使用平行卷积层学习offset偏移,使得卷积核在输入特征图上的采样点发生偏移,可集中于人们感兴趣的区域或者目标。即对卷积核中每个采样点的位置都增加了偏移量,可以实现在当前位置附近随意采样而不局限于常规卷积的规则采样点。如图3所示为常规卷积的采样点和可变形卷积采样点的对比。图4所示为可变形卷积的计算流程,使用卷积层对输入特征图计算偏移量,偏移量和输出特征图有相同的分辨率,输出通道数为 $3N$ (N 为卷积核采样点个数),其中 $2N$ 为预测的 x, y 方向偏移量,因为不同采样点对特征有不同的贡献,所以还要预测 N 个采样点的权重。

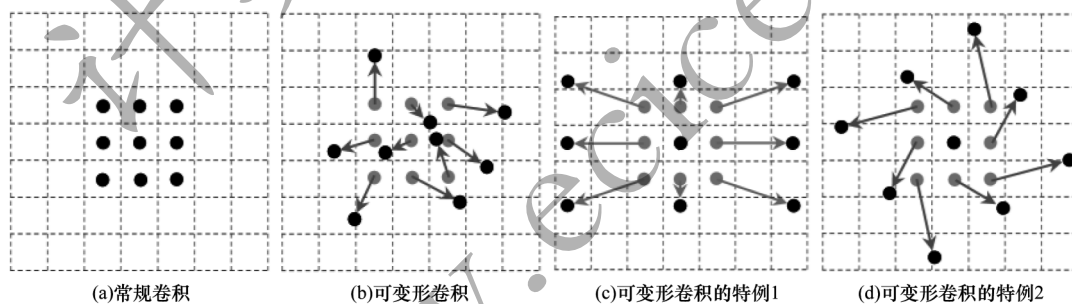
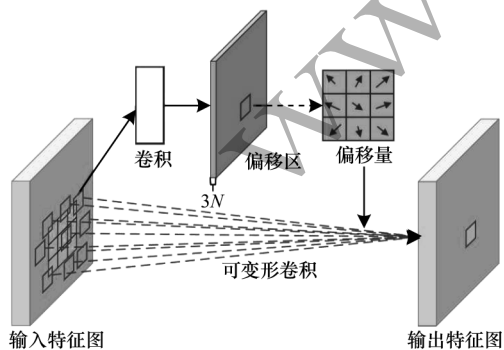
图3 常规3 \times 3卷积和可变形卷积采样点对比Fig.3 Comparison of sampling points between conventional 3 \times 3 convolution and deformable convolution

图4 可变形卷积的计算流程

Fig.4 Computational procedure of deformable convolution

常规的卷积操作主要分为2步:1)在输入的特征图上使用规则网格 R 进行采样;2)使用卷积核 w 对采样点进行加权运算。 R 定义了感受野的大小和扩张,如式(1)所示定义了一个大小为 3×3 、扩张率为1的卷积核。

$$R = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\} \quad (1)$$

对于在输出的特征图上的每个位置 p_0 ,通过式(2)计算输出值 $y(p_0)$:

$$y(p_0) = \sum_{p_n \in R} w(p_n) \times x(p_0 + p_n) \quad (2)$$

其中: p_n 是对 R 中所列位置的枚举。

在可变形卷积的操作中,通过对规则网格 R 增加一个偏移量 $\{\Delta p_n | n = 1, 2, \dots, N\}, N = |R|$ 进行扩张,

同时对每个采样点预测一个权重 Δm_n 。那么同样的位置 p_0 的值 $y(p_0)$ 变为如式(3)所示:

$$y(p_0) = \sum_{p_n \in R} w(p_n) \times x(p_0 + p_n + \Delta p_n) \times \Delta m_n \quad (3)$$

由于偏移量 Δp_n 通常是小数,因此需要通过双线性插值法计算 x 的值,如式(4)~式(6)所示:

$$x(p) = \sum_q G(q, p) \times x(q) \quad (4)$$

$$G(q, p) = g(q_x, p_x) \times g(q_y, p_y) \quad (5)$$

$$g(a, b) = \max(0, 1 - |a - b|) \quad (6)$$

图5所示为常规卷积和可变形卷积的特征点感受野之间的对比。2个特征点(分别在大羊和小羊身上)代表不同尺度和形状的目标。对比可知,通过两层卷积运算,常规卷积的特征点都具有固定尺寸的感受野,而可变形卷积可自适应的学习感受野的采样位置更符合物体本身的形状和尺寸,更有利于特征提取。

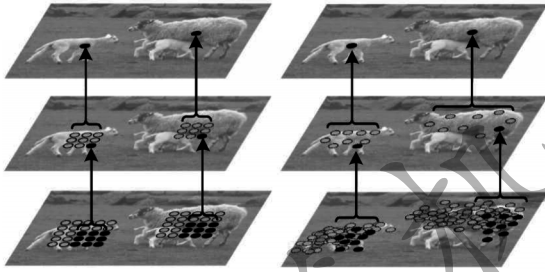


图5 常规卷积和可变形卷积的特征点感受野对比

Fig.5 Comparison of receptive fields between conventional convolution and deformable convolution

本文基于可变形卷积构建 res-dcn 模块,在常规的卷积网络中,首先使用 3×3 的卷积进行特征提取,然后使用 1×1 的卷积进行特征压缩。对于可变形卷积来说,使用 1×1 的可变形卷积对采样点计算偏移容易导致采样不稳定,因此,本文的 res-dcn 模块连续使用 3 个 3×3 的可变形卷积进行特征提取。最后再将输出和输入按通道数进行拼接,以保留多层次的语义信息。如图6所示,Deformable conv是由1个可变形卷积、1个BN层和1个leaky激活层组成。其中,concat为按通道数拼接。

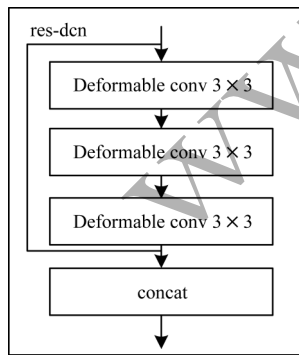


图6 res-dcn 模块结构

Fig.6 Module structure of res-dcn

本文设计了2种改进的 dcn-YOLO 模型,第1种是在第1个YOLO特征提取头部嵌入 res-dcn 模块,第

2种是将YOLO网络中3个特征提取头的卷积模块替换为 res-dcn,以实现对各个尺度都能更好的自适应提取特征。如图7所示,该设计参考了YOLO-spp^[22]的网络结构。YOLO-spp将YOLO网络中的Head-1特征提取头替换为YOLO-spp Head-1,使用 $1 \times 1, 5 \times 5, 9 \times 9, 13 \times 13$ 的最大池化层,提取不同尺寸感受野的特征,再将特征按通道数进行拼接以融合多尺度的特征,提高了YOLO的检测性能。由此可知,YOLO网络中Head特征提取头模块具有丰富的语义信息。

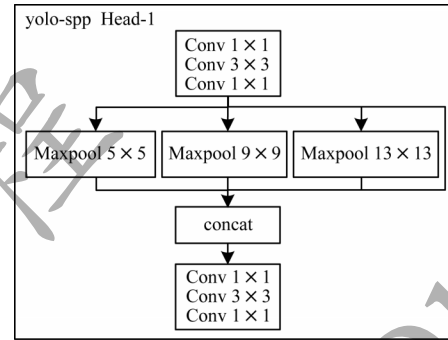


图7 YOLO-spp Head-1 模块结构

Fig.7 Module structure of YOLO-spp Head-1

本文先将 dcn-YOLO1 模型在 Head-1 的第2卷积层后使用 1×1 的常规卷积压缩特征通道数到512以减少计算量;然后将 res-dcn 模块嵌入该卷积层后进行特征提取,3个可变形卷积通道数均为512,网络中其他结构和YOLO保持一致;而 dcn-YOLO2 则将 res-dcn 模块替换YOLO中所有Head特征提取头的第2、3、4卷积层,以实现对大、中、小目标分别自适应提取特征。每个 res-dcn 模块中可变形卷积的通道数分别为512、256、128,网络中其他结构和YOLO保持一致,如图8和图9所示。

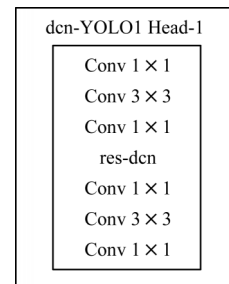


图8 dcn-YOLO1 Head-1 模块结构

Fig.8 Module structure of dcn-YOLO1 Head-1

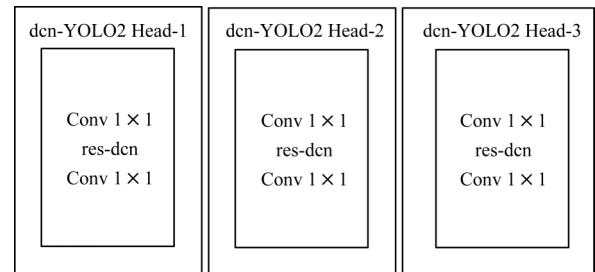


图9 dcn-YOLO2 Head 模块结构

Fig.9 Module structure of dcn-YOLO2 Head

3 实验结果与分析

本文所用到的实验平台是 Intel 酷睿 i79700K CPU, 64 GB 内存, NVIDIA RTX 2080 Ti 11 GB GPU, 软件环境是 ubuntu 18.04, pytorch 1.3, cuda 10.0, cudnn 7.5。本文实验使用 pascal voc 2007 和 voc 2012 公开数据集, 共有 20 类目标; 使用 voc2007trainval 和 voc2012trainval 作为训练集, 共有 16 551 张图像; 使用 voc2007test 作为测试集, 共有 4 952 张图像; 使用平均精度 (mAP) 和检测速度 (FPS) 来检验模型的有效性。

3.1 锚盒计算

YOLO 使用锚盒作为先验框辅助预测目标边界, 尺寸合适的锚盒可以使目标检测精度提高。为了获得平均交并比 (avg-iou) 更大的锚盒, 本文使用 k-means++ 算法进行锚盒计算, 使用 (1-iou) 作为聚类距离。算法流程如下:

算法 1 k-means++ 聚类锚盒算法

输入 训练集所有目标的宽高集合 S , 聚类中心数 K

输出 K 组锚盒

步骤 1 从 S 中随机取一个值作为初始聚类中心 C_i

步骤 2 计算 S 中所有样本与已有聚类中心的最小 iou 距离 $d(x)$, 概率计算公式如式 (7) 所示, 并选取下一个聚类中心 C_i 。

$$p = \frac{d(x)^2}{\sum_{x \in S} d(x)^2} \quad (7)$$

步骤 3 重复步骤 2 直到找到 K 个聚类中心。

步骤 4 对数据集中的每个样本 x_i , 计算其到 K 个聚类中心的 iou 距离, 并将其划分为距离最小的聚类中心所对应的类别。

步骤 5 根据划分结果重新计算 K 个聚类中心, 计算式如式 (8) 所示:

$$C_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \quad (8)$$

表 2 各 YOLO 模型纵向对比实验

Table 2 Longitudinal contrast experiment of YOLO models

网络模型	新锚盒	res-dcn	mAP/%	检测速度/(frame·s ⁻¹)	训练耗时/h
YOLO(1)	×	×	80.56	100	36
YOLO(2)	√	×	80.71	100	36
YOLO-spp	×	×	81.96	99	36
all-dcn-YOLO	√	√	79.36	41	115
dcn-YOLO1	√	√	82.23	93	44
dcn-YOLO2	√	√	82.62	79	64

从表 2 可以看出, 使用 k-means++ 聚类的锚盒获得了更高的 avg-iou, mAP 有略微提升。dcn-YOLO1 和 dcn-YOLO2 的 mAP 相比原始 YOLO 算法分别提高了 1.67、2.06 个百分点, 这是因为本文添加的 res-

步骤 6 重复步骤 4 和步骤 5, 直到聚类中心位置不再变化, 输出最终的聚类中心。

经过聚类, 得到了 9 组锚盒 $\{(28, 55), (65, 80), (66, 179), (145, 114), (116, 273), (187, 196), (334, 200), (215, 333), (362, 360)\}$, avg-iou 值为 68.26%, 比 YOLO 默认的锚盒提高了 4.10%。

3.2 网络训练

为了进行充分的对比实验, 分别对 YOLO 模型、计算新锚盒后的 YOLO 模型、YOLO-spp 模型和本文提出的 dcn-YOLO1、dcn-YOLO2 模型进行训练, 并将 YOLO 网络中所有常规卷积层简单替换为可变形卷积, 形成 all-dcn-YOLO 模型, 以对比可变形卷积的特征提取能力。设置训练迭代 50 200 轮次, 批量大小为 8, 动量为 0.9, 衰减系数为 0.000 5, 学习率更新策略见表 1。使用随机水平翻转、裁剪、调整曝光饱和度和数据增强操作。每迭代 10 轮次后, 随机调整输入图像尺寸 (从 320 像素×320 像素到 608 像素×608 像素, 以 32 的倍数变化)。

表 1 学习率更新策略

Table 1 Learning rate update strategy

迭代轮次	学习率
0	0.000 05
400	0.000 10
700	0.000 20
900	0.000 50
1 000	0.001 00
40 000	0.000 10
45 000	0.000 01

3.3 纵向对比实验

训练完成后, 使用 voc2007test 数据集进行对比测试, 将输入的图像尺寸缩放为 416 像素×416 像素, 对比结果见表 2。

dcn 模块可以对不同尺寸和形状的目标自适应提取有效特征。虽然 dcn-YOLO1 和 dcn-YOLO2 的 FPS 有一定降低, 但仍然能达到实时的效果。此外也可以看出简单地将所有常规卷积替换为可变形卷积,

并不能提高模型的检测效果。相反,由于 dcn 需要使用一个额外的卷积操作预测采样点的偏移量,在前向计算和反向传播的过程中,与标准卷积相比都增加了计算量,因此训练时间和检测时间大大增加,而检测精度却没有提升。这也证明了本文提出的 res-dcn 模型有效性,即在损失少量计算速度的前提下,有效地提高了检测精度。

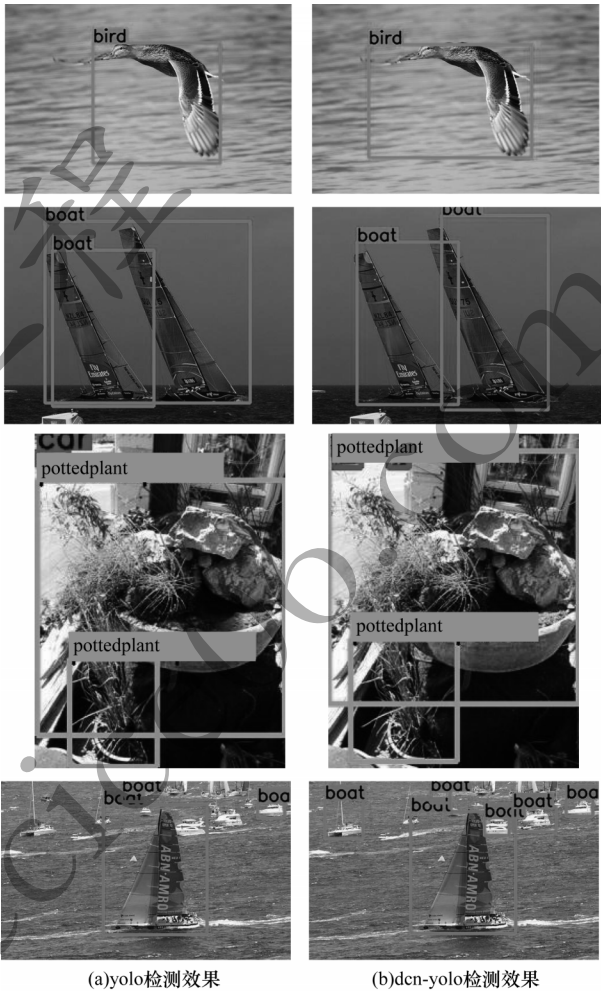
表 3 列出了各个模型对不同类别的 mAP 值,其中加粗部分为各对比模型中获得较高 mAP 的项。从表 3 可以看出,本文提出的模型与原 YOLO 模型相比,18 种类别的目标检测精度均有一定提高,尤其是 bird 和 pottedplant 等小目标分别比 YOLO 高出了 4.59、3.70 个百分点。同时,对 boat 和 cow 等尺寸和形状变化较大的目标,精度分别提高了 4.16、3.16 个百分点。虽然有些类别对 YOLO-spp 有更高的精度,但提出的模型只略低一点。总体来看,本文提出的算法仍然达到了最高的 mAP。

表 3 各模型对不同类别的 AP 对比
Table 3 Comparison of AP of different models for different categories %

类别	YOLO	YOLO-spp	dcn-YOLO1	dcn-YOLO2
bird	75.71	77.88	79.62	80.30
horse	90.03	90.02	88.63	91.59
boat	72.29	72.65	75.09	76.45
diningtable	75.93	78.46	78.03	78.17
sheep	82.17	82.73	85.11	85.27
bus	88.15	89.12	86.82	88.48
sofa	79.25	80.76	84.12	81.80
person	83.72	84.35	84.23	84.80
aeroplane	87.59	88.59	89.00	89.32
dog	87.60	89.70	89.68	88.81
car	88.79	89.97	89.69	89.93
motorbike	87.86	89.66	90.38	89.37
chair	63.15	63.29	64.37	65.24
bicycle	88.93	88.77	90.14	88.38
pottedplant	56.39	56.08	56.38	60.09
train	86.98	90.61	87.68	89.35
cat	90.25	91.76	89.97	90.18
bottle	67.13	68.98	70.26	68.42
cow	83.62	87.30	85.98	86.78
tvmonitor	75.71	78.62	79.40	79.62
mAP	80.56	81.96	82.23	82.62

图 10 所示为 YOLO 和 dcn-YOLO 的检测效果,在第 1 行中,dcn-YOLO 预测出了 bird 更精确的边界框。由于翅膀所占的像素较少,特征不明显,YOLO 提取了更多的背景干扰特征,导致定位精度不高。而 dcn-YOLO 可以自适应地提取到鸟翅膀的特征,所以定位效果更好。在第 2 行中,YOLO 将 2 个靠近

的船只检测为了 1 只船,而 dcn-YOLO 很好地区分出了 2 只船;在第 3 行中,dcn-YOLO 准确地检测出了左上角远处模糊的车辆,也得益于其良好的特征提取能力;第 4 行中由于小目标特征不明显,YOLO 漏检了很多的小船只,而与之相比,dcn-YOLO 检测出了更多的目标,取得了更好的检测效果。



(a)yolo检测效果 (b)dcn-yolo检测效果

图 10 YOLO 和 dcn-YOLO 的检测效果对比

Fig.10 Comparison of detection effect between YOLO and dcn-YOLO

3.4 横向对比实验

将提出的模型和目前性能优异的目标检测算法进行了对比,结果见表 4。可以看到 dcn-YOLO 与 YOLO-spp 和 Attention-YOLO 相比,mAP 分别高出了 0.64、0.70 个百分点,而且检测速度远远超过了 Attention-YOLO;与单阶段目标检测算法 SSD 和 DSSD 相比,dcn-YOLO 算法的 mAP 分别高出了 5.20、4.00 个百分点,且检测速度远远超过 DSSD,达到实时效果;与两阶段检测算法 FasterR-CNN 和 R-FCN 相比,dcn-YOLO 算法的 mAP 分别高出了 9.40、2.10 个百分点。

表4 dcn-YOLO与其他目标检测算法性能对比

Table 4 Performance comparison between dcn-YOLO and other object detection algorithms

算法	输入图片尺寸	mAP/%	检测速度/(frame·s ⁻¹)
YOLO	416×416	80.60	100
YOLO-spp	416×416	81.96	99
Attention-YOLO	416×416	81.90	25
dcn-YOLO	416×416	82.60	79
SSD	300×300	77.40	85
DSSD	321×321	78.60	16
FasterR-CNN	600×1 000	73.20	10
R-FCN	600×1 000	80.50	12

4 结束语

为提高YOLO算法的检测精度,本文基于可变形卷积构建res-dcn模块,使用k-means++聚类出符合数据集尺寸的锚盒,将res-dcn模块分别嵌入和替换YOLO特征提取头,最终得到改进的dcn-YOLO模型。在VOC数据集上的实验结果表明,该模型相比原YOLO算法以及其他主流的目标检测算法,都获得了更好的效果,证明了本文提出模型的有效性。但由于可变形卷积需要更多的计算量,导致模型的运算速度有一定下降,下一步将构建更好的模型,以保证在不降低网络计算速度的前提下,达到更高的检测精度。

参考文献

- [1] 李军伟,周小龙,产思贤,等. 基于自适应卷积神经网络特征选择的视频目标跟踪方法[J]. 计算机辅助设计与图形学学报,2018,30(2):273-281.
LI J W, ZHOU X L, CHAN S X, et al. A novel video target tracking method based on adaptive convolutional neural network feature[J]. Journal of Computer-Aided Design & Computer Graphics, 2018, 30(2): 273-281. (in Chinese)
- [2] 桑军,郭沛,项志立,等. Faster-RCNN的车型识别[J]. 重庆大学学报,2017,40(7):32-36.
SANG J, GUO P, XIANG Z L, et al. Vehicle detection based on faster-RCNN[J]. Journal of Chongqing University, 2017, 40(7): 32-36. (in Chinese)
- [3] 芮挺,费建超,周游,等. 基于深度卷积神经网络的行人检测[J]. 计算机工程与应用,2016,52(13):162-166.
RUI T, FEI J C, ZHOU Y, et al. Pedestrian detection based on deep convolutional neural network[J]. Computer Engineering and Applications, 2016, 52(13): 162-166. (in Chinese)
- [4] 谢林江,季桂树,彭清,等. 改进的卷积神经网络在行人检测中的应用[J]. 计算机科学与探索,2018,12(5):708-718.
XIE L J, JI G S, PENG Q, et al. Application of preprocessing convolutional neural network in pedestrian detection[J]. Journal of Frontiers of Computer Science and Technology, 2018, 12(5): 708-718. (in Chinese)
- [5] NAVNEET D, BILL T. Histograms of oriented gradients for human detection [C]//Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2005: 4-8.
- [6] DAVID G. Object recognition from local scale-invariant features [C]//Proceedings of International Conference on Computer Vision. New York, USA: ACM Press, 1999: 1150-1157.
- [7] PAPAGEORGIOU C P, OREN M, POGGIO T. A general framework for object detection [C]//Proceedings of the 6th Computer Vision, Washington D. C., USA: IEEE Press, 1998: 555-562.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 779-788.
- [9] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 6517-6525.
- [10] REDMON J, FARHADI A. YOLOv3: an incremental improvement [EB/OL]. [2020-07-28]. <https://arxiv.org/abs/1804.02767>.
- [11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 21-37.
- [12] FU C Y, LIU W, RANGA A, et al. DSSD: deconvolutional single shot detector [EB/OL]. [2020-07-28]. <https://arxiv.org/abs/1701.06659>.
- [13] 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.
- [14] DAI J F, LI Y, HE K M, et al. R-FCN: object detection via region-based fully convolutional networks [C]//Proceedings of 2016 Conference Advances in Neural Information Processing Systems. Barcelona, Spain: MIT Press, 2016: 379-387.

(下转第282页)

(上接第 275 页)

- [15] 姜竣,翟东海. 基于空洞卷积与特征增强的单阶段目标检测算法[J/OL]. 计算机工程;1-10[2020-09-27]. <https://doi.org/10.19678/j.issn.1000-3428.0058315>.
JIANG J, ZHAI D H. Single-stage object detection algorithm based on atrous convolution and feature enhancement[EB/OL]. Computer Engineering;1-10[2020-09-27]. <https://doi.org/10.19678/j.issn.1000-3428.0058315>. (in Chinese)
- [16] 徐诚极,王晓峰,杨亚东. Attention-YOLO:引入注意力机制的YOLO检测算法[J]. 计算机工程与应用,2019,55(6):13-23.
XU C J, WANG X F, YANG Y D. Attention-YOLO: YOLO detection algorithm that introduces attention mechanism[J]. Computer Engineering and Applications, 2019,55(6):13-23. (in Chinese)
- [17] DAI J F, QI H Z, XIONG Y W, et al. Deformable convolutional networks[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017:764-773.
- [18] ZHU X Z, HU H, LIN S, et al. Deformable convnets v2: more deformable, better results[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2019: 9308-9316.
- [19] MARK E, LUC VAN G, CHRISTOPHER K W, et al. The pascal visual object classes challenge[J]. International Journal of Computer Vision. 2010,88(2):303-338.
- [20] LIN T, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[EB/OL]. [2020-07-28]. <https://arxiv.org/abs/1612.03144>.
- [21] REDMON J, FARHADI A. YOLO: real-time object detection[EB/OL]. [2020-07-28]. <https://pjreddie.com/darknet/yolo/>.
- [22] 管皓,薛向阳,安志勇. 一种利用在线卷积网络的视频目标跟踪方法[J]. 小型微型计算机系统,2017,38(4):872-875.
GUAN H, XUE X Y, AN Z Y. Video object tracking via online convolutional network[J]. Journal of Chinese Computer Systems, 2017,38(4):872-875. (in Chinese)

编辑 赖玉玲