

基于改进的 DSSD 的小目标检测算法研究

杨朝晨¹, 陈佳悦², 邢可¹, 刘梦尼¹, 高涛¹

(1. 长安大学 信息工程学院, 陕西 西安 710064;

2. 西北大学 信息科学与技术学院, 陕西 西安 710127)

摘要:随着深度学习的迅速发展,图像识别技术也随之日益提高,其中目标检测在辅助驾驶系统、医学领域和车流监测系统占有重要地位。大多目标检测算法对大目标较为敏感,且并未考虑特征与特征之间的相互关系及重要程度,然而小目标在图像中覆盖区域小,分辨率低,携带信息较少,导致小目标的误检或漏检率较高。针对以上问题,对小目标检测的难点进行研究,提出了一种基于改进的 DSSD(deconvolutional single shot detector)的小目标检测算法。该算法引入混合注意力机制,在通道维度上增加权重分量进行加权求和表示信息相关度,并将图片中的空间域信息做对应空间变换,提取关键信息,突出局部重点区域,有利于前景小目标的特征学习。实验结果表明,该算法在 VOC2007 测试集上的精确度达到 81.02%,比原 DSSD 算法高出 1.3%,且均优于其他对比算法,证明了算法的有效性。

关键词:深度学习;DSSD;残差网络;小目标检测;图像处理

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2022)06-0063-05

doi:10.3969/j.issn.1673-629X.2022.06.011

Small Target Detection Algorithm Based on Improved DSSD

YANG Zhao-chen¹, CHEN Jia-yue², XING Ke¹, LIU Meng-ni¹, GAO Tao¹

(1. School of Information Engineering, Chang'an University, Xi'an 710064, China;

2. School of Information Science and Technology, Northwest University, Xi'an 710127, China)

Abstract: With the rapid development of deep learning, image recognition technology is also improving, and target detection has become an important and highly relevant topic in assistant driving system, medical field and traffic flow monitoring system. Most of the target detection algorithms are sensitive to large targets and unable to take into account the relationship and importance between features. However, small targets have small coverage area, low resolution and less information, which leads to high false detection or missed detection rate of small targets. In view of the above problems, the difficulties of small target detection are studied, and a small target detection algorithm based on DSSD is proposed. In this algorithm, the hybrid attention mechanism is introduced, and the weighted component is added to the channel dimension to represent the information correlation, and the spatial domain information in the image is transformed in corresponding space to extract key information and highlight local key areas, which is conducive to feature learning of small foreground targets. The experimental results demonstrate that the proposed algorithm achieves 81.02% accuracy on the VOC2007 test set, which is 1.3% higher than that of the original DSSD algorithm, and is also superior to other comparison algorithms, which proves the effectiveness of the algorithm.

Key words: deep learning; DSSD; residual network; small target detection; image processing

0 引言

随着人工智能技术的飞速发展,目标检测(object detection)已成为计算机图像处理中一个重要研究课题,在智能交通^[1]、人脸识别^[2]、智能图像检索^[3]等领域发挥着重要作用。

传统的目标检测算法采用穷举策略使得模型产生

过多冗余窗口,导致模型检测速度缓慢,面对庞大数据集时过程繁琐,且对于复杂场景泛化性能差,无法有效处理由于不同姿势、遮挡等引起的非线性变化,在特征识别上有着难以弥补的劣势,进而难以有效提高分类水平。

近年来,深度学习凭借其优异的特征提取能力被

收稿日期:2021-07-12

修回日期:2021-11-16

基金项目:国家重点研发计划项目(2019YFE0108300);国家自然科学基金项目(62001058);陕西省重点研发计划项目(2019GY-039);长安大学中央高校基本科研业务费专项资金资助项目(300102241201)

作者简介:杨朝晨(1997-),女,硕士研究生,研究方向为数字图像处理与深度学习;高涛,教授,博士,研究方向为数字图像处理与模式识别。

广泛应用于目标检测领域。目前,基于深度学习的目标检测算法可大致分为两类:第一类为基于候选区域的两阶段算法,如以 RCNN^[4]、Fast R-CNN^[5] 和 Faster R-CNN^[6] 等为代表的一系列目标检测框架,这类算法准确度较高,但速度较慢;第二类为基于端到端模型的一阶段算法,典型代表有 SSD^[7]、YOLOv3^[8]、DSSD^[9] 等,这类方法利用回归思想,通过卷积网络预测回归边框及分类概率,无需预先提取候选网络,直接对图像中的目标进行位置预测和类别判断,较二阶段算法准确度略低,但检测速度更快,实时性更高。

然而在实际情况中,图像中的小目标分辨率低,图像模糊,携带信息较少,使得其特征表达能力弱,算法对小目标的检测不敏感,误检率或漏检率较高。针对以上问题,一些学者对普通目标检测算法做出了改进。郑浦等^[10]采用多尺度特征融合的方式,增强浅层特征表达能力对小目标进行检测。刘顺民等^[11]使用特征金字塔替代多尺度特征层,充分融合了底层和高层的信息。赵文清等^[12]通过将改进的分段比例惩罚因子 NMS 算法和连续比例惩罚因子 NMS 算法相结合以提高检测性能。马跃等^[13]设计了一种融合结构用以丰富特征图细节信息和语义信息。

上述研究均使用完整特征图作为特征输入,然而在实际分类任务中,并非所有特征对特征识别的作用程度都是相同的。因此,针对算法对小目标敏感性较低的问题,该文提出了一种改进的 DSSD 算法,引入注意力机制突出局部重点区域,在通道维度上增加权重

分量进行加权求和表示信息相关度,并将图片中的空间域信息做对应空间变换,有利于前景小目标的特征学习。

1 算法介绍

1.1 SSD 网络

SSD(single shot multibox detector)模型是一种基于回归的高精度、强实时性的目标检测算法,其主要设计思想是利用不同尺度的特征图进行目标的检测和识别。网络框架分为基础网络和附加网络两部分。基础网络使用改进的 VGG16 网络进行特征提取,附加网络新增四级级联卷积层以获得高维特征图用于目标检测。在网络模型进行训练时,预测框会以先验框为基准,通过 softmax 分类和边界回归来获得真实目标的位置,多尺度特征图共同参与目标检测有利于提升目标检测的精确度。

1.2 DSSD 模型

SSD 模型检测速度较快,对大目标检测精度较高,但对小目标不敏感,漏检或误检率较高。DSSD 模型在此基础上进行了改进,主要是采用 ResNet101 代替 VGG16 作为主干网络,其中 ResNet101 的 Conv3_x 卷积层与其后 4 层级联卷积层共同构成 DSSD 模型的特征层;特征层与 5 层反卷积层共同构成反卷积模块以扩大模型在小尺度上的高层特征信息,有效改善了小目标的检测能力;并引入预测模块提升检测精度, DSSD 模型结构如图 1 所示。

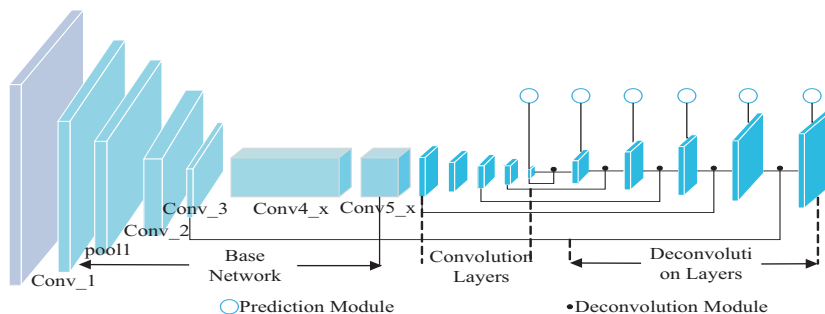


图 1 DSSD 模型结构

1.2.1 反卷积模块

为了得到更多的检测信息,DSSD 模型引入反卷积层组成反卷积模块,反卷积模块结构如图 2 所示。

DM 模块里的需要反卷积的高层特征图(图中以 A 表示)、SSD 的卷积低层特征图(图中以 B 表示)和最后得到的融合后的卷积图(图中以 C 表示)与 DSSD 的沙漏结构中的对应位置如图所示。融合后的特征图 C 送入预测模块中,作为下一级 DM 模块要反卷积的特征图再进行下一级的 DM 模块操作。至于 2 H 和 2 W 经过一系列的 padding=1, stride=1 的 3*3 卷积后仍保持 2 H*2 W 的特征图尺寸。

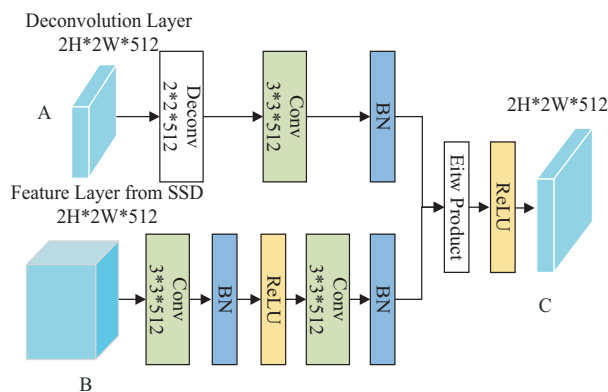


图 2 反卷积模块结构

1.2.2 预测模块

DSSD 模型在特征层和反卷积模块之间又加入了预测模块,该模块在 SSD 预测模块的基础上增加了一

个残差学习单元,进一步提高模型的精确度。预测模块结构如图 3 所示。

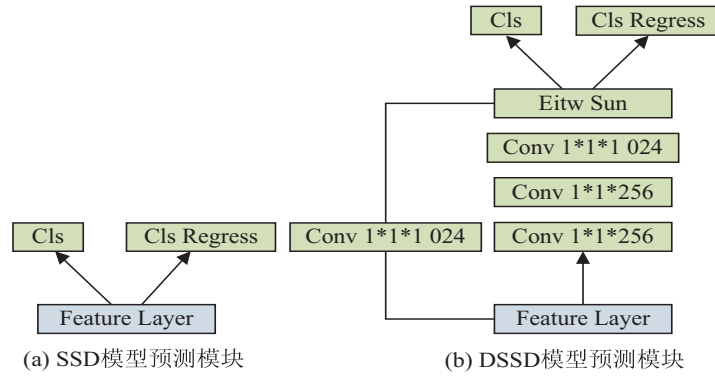


图 3 反卷积模块结构

2 CBAM+DSSD 算法

虽然 DSSD 模型的预测模块提取更深维度的特征用于分类和回归,网络整体构成不对称的“沙漏”结构,充分利用了上下文信息和层的特征,从而在小目标和密集目标的检测率上有很大的提高。但上述均使用完整特征图作为特征输入,然而在实际分类任务中,并非所有特征对特征识别的作用程度都是相同的。因此,为了突出对特征识别有效的信息并抑制作用小或无效的信息,一些研究引入了注意力机制。

2.1 CBAM 混合注意力机制

CBAM (convolutional block attention module)^[14] 是一种简单有效的混合注意力模块。通过将注意力过程分为两个独立的部分,通道注意力模块和空间注意力模块,沿着空间和通道两个维度依次推断出注意力权重,然后与原特征图相乘来对特征进行自适应调整,其网络结构如图 4 所示。

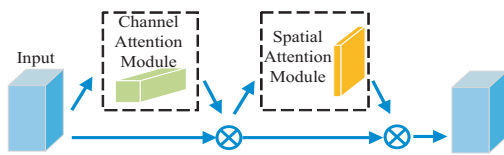


图 4 CBAM 模型结构

为了汇总空间特征,通道注意力模块采用了全局平均池化和最大池化两种方式分别利用不同的信息。输入是一个 $H \times W \times C$ 的特征 F ,先分别进行一个空间的全局平均池化和最大池化得到两个 $1 \times 1 \times C$ 的通道描述。接着将它们分别送入一个两层的神经网络,第一层神经元个数为 C/r ,激活函数为 Relu,第二层神经元个数为 C 。这个两层的神经网络是共享的。然后,再将得到的两个特征相加后经过 Sigmoid 激活函数得到权重系数 M_c 。最后,拿权重系数和原来的特征 F 相乘即可得到缩放后的新特征。

在通道注意力模块之后再引入空间注意力模块,给定一个 $H \times W \times C$ 的特征 F ,分别进行一个通道维度的平均池化和最大池化得到两个 $H \times W \times 1$ 的通道描述,并将这两个描述按照通道拼接在一起。然后经过一个 7×7 的卷积层,激活函数为 Sigmoid,得到权重系数 M_s 。最后,拿权重系数和特征 F 相乘即可得到缩放后的新特征。

2.2 模型结构

本实验采用优化后的 DSSD 模型,引入 CBAM 混合注意力机制,在保证检测速度的同时,提高了检测精度,改进的模型结构如图 5 所示。

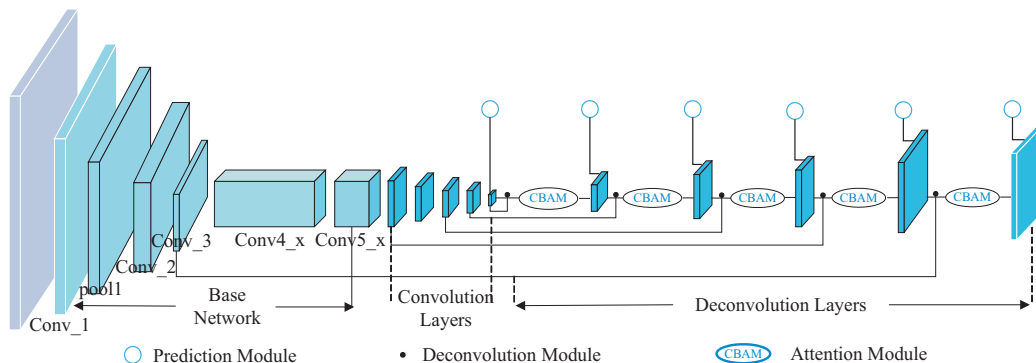


图 5 改进的 DSSD 模型结构

在图 5 中,反卷积模块中各卷积层既具有良好的浅层特征表征能力,又具有深层特征语义信息,故在反卷积模块引入 CBAM 注意力机制。由于 CBAM 模块属于轻量级模块,不会带来大量复杂参数,故对改进的 DSSD 模型检测速度基本没有影响,且一定程度上提高了网络整体检测精度和对小目标的检测精度。

3 实验设置、结果与分析

3.1 实验环境与评价指标

实验使用的深度学习框架为 pytorch,计算机系统为 Ubuntu16. 04, 显卡型号为 NVIDIA Quadro P5000。

采用平均精度 (average precision, AP) 和平均精度均值 (mean average precision, mAP) 作为模型检测性能的评价指标,其计算过程如下:按照目标检测模型预测结果对样本进行排序,针对逐个样本将置信度作为划分正负样本的阈值,计算查准率 (Precision) 和召回率 (Recall),绘制 P-R (Precision and Recall) 曲线,将 PR 曲线上召回率相同的点记为一组,共分为 n 组,AP 的计算公式如下:

$$AP = \sum_{i=1}^n \frac{P_i}{N} \quad (1)$$

其中, P_i 表示第 i 组的最大精度值, N 表示该类别在测试集中的所有实例个数。

最后计算各个类别 AP 值的均值即为 mAP。AP 和 mAP 越高,目标检测模型的性能越好。

3.2 实验数据集

实验采取 2 个通用数据集,分别为 PASCAL VOC2007 和 PASCAL VOC2012 数据集。PASCAL VOC^[15] 数据集是一个包含 Person、Bird、Cat、Cow、Dog、Horse、Sheep、Aero-plane、Bicycle、Boat、Bus、Car、Motorbike、Train、Bottle Chair、Table、Plant、Sofa、Monitor 等 20 个类别物体的目标检测数据集。该文使用 PASCAL VOC2007 和 PASCAL VOC2012 的 trainval 部分共 16 551 张图片作为训练集,使用 PASCAL VOC2007 的 test 部分作为测试集。

3.3 模型训练参数

通过一系列的调参实验后总结出了最适合该模型训练的相关网络超参数,具体如下:训练模型过程中采用 Momentum 优化算法,动量衰减项为 0.9,权重衰减因子设定为 0.000 1,步长设定为 120 000, batch size 为 16。学习率为动态的即初始学习率为 0.001,前 500 步学习率为 0.000 1,500 步后学习率为 0.001,80 000 步后学习率下降为原来的 0.1 倍,调整至 0.000 1,100 000 步后学习率调整为 0.000 01。当模型构建完成及参数确定后进行 MFF-SSD 网络模型的训练,网

络模型训练的全过程被记录在图 6 中。

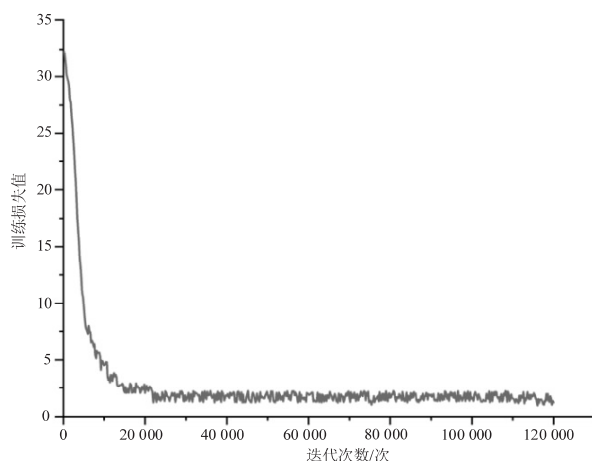


图 6 模型训练损失值

由图 6 可知:损失值在前 5 000 步迅速下降;5 000 至 21 000 左右时损失值逐渐趋于稳定状态;21 000 步后损失值趋于稳定。

3.4 实验结果与分析

利用训练好的网络对测试集进行检测,网络对 VOC2007 test 部分的检测精度如表 1 所示。可以看出所提算法对每个类别的识别精度都有一定程度的提升。

表 1 不同算法各个类别目标检测精度对比结果 %

类别	SSD	DSSD	Ours
Aero-plane	80.19	85.02	87.91
Bicycle	82.79	86.36	89.74
Bird	72.63	81.19	85.26
Boat	67.83	71.31	74.39
Bottle	50.18	51.71	55.12
Bus	82.30	85.79	88.65
Car	83.06	90.26	90.70
Cat	84.13	85.87	89.88
Chair	58.60	57.58	58.88
Cow	79.71	84.80	85.19
Dog	82.63	87.31	88.76
Horse	85.06	87.95	89.55
Motorbike	80.74	87.29	90.57
Person	76.50	82.04	83.25
Plant	50.10	54.58	55.87
Sheep	75.70	77.00	80.89
Sofa	80.43	80.13	80.23
Train	87.09	89.32	86.62
Tv	76.48	78.15	79.53

表 2 是不同模型在 VOC2007 数据集上的实验结果,对比 DSSD 和所提方法的实验结果可知,在平均检测时间略微提升的情况下,所提方法得到了比 DSSD

更优的实验精度,证明了算法的有效性。

表2 VOC2007 测试集上测试结果

方法	mAP/%	平均检测时间/s
Faster R-CNN	69.07	0.141
SSD	74.65	0.193
YOLO	72.44	0.154
DSSD	79.72	0.132
Ours	81.02	0.198

4 结束语

针对当前 DSSD 模型没有考虑小目标特点及特征图通道和空间之间的关系的问题,通过引入 CBAM 混合注意力模块,设计了一个针对小目标检测任务的改进 DSSD 模型。相比于原 DSSD 模型,设计的新目标检测模型在实验精度提升 1.3% 情况下,基本没有增加模型的参数量,检测速度无明显降低。实验的训练和推理过程仍旧是在高性能计算机上进行,下一步将考虑将检测模型进一步简化,在保证精度的同时降低检测速度。

参考文献:

- [1] FENICHE M, MAZRI T. Lane detection and tracking for intelligent vehicles: a survey [C]//2019 international conference of computer science and renewable energies (ICCSRE). Agadir, Morocco; IEEE, 2019: 1-4.
- [2] 姚乃明, 郭清沛, 乔逢春, 等. 基于生成式对抗网络的鲁棒人脸表情识别[J]. 自动化学报, 2018, 44(5): 865-877.
- [3] 李均鹏, 祝开艳, 杨 澍. 基于迁移学习的复杂场景海洋鱼类识别方法[J]. 计算机应用与软件, 2019, 36(9): 168-174.
- [4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Columbus, OH, USA; IEEE, 2014: 580-587.
- [5] GIRSHICK R. Fast R-CNN [C]//Proceedings of the IEEE international conference on computer vision. Santiago, Chile; IEEE, 2015: 1440-1448.
- [6] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]//Proceedings of the advances in neural information processing systems. Montreal, Canada; IEEE, 2015: 91-99.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//Proceedings of the European conference on computer vision. Amsterdam, The Netherlands; Springer, 2016: 21-37.
- [8] REDMON J, FARHADI A. YOLOv3: an incremental improvement [J]. arXiv: 1804. 02767, 2018.
- [9] FU C Y, LIU W, RANGA A, et al. DSSD: deconvolutional single shot detector [J]. arXiv: 1701. 06659, 2017.
- [10] 郑 浦, 白宏阳, 李 伟, 等. 复杂背景下的小目标检测算法[J]. 浙江大学学报: 工学版, 2020, 54(9): 1777-1784.
- [11] 刘顺民, 马致明, 陈炳才. 基于 SSD 算法的交通指示牌目标检测[J]. 现代电子技术, 2021, 44(13): 144-147.
- [12] 赵文清, 严 海, 邵绪强. 改进的非极大值抑制算法的目标检测[J]. 中国图象图形学报, 2018, 23(11): 1676-1685.
- [13] 马 跃, 赵志浩, 尹震宇, 等. 分段反卷积改进 SSD 的目标检测算法[J]. 小型微型计算机系统, 2021, 42(7): 1415-1420.
- [14] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]//Proceeding of the European conference on computer vision (ECCV). Munich, Germany; Springer, 2018: 3-19.
- [15] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City, UT, USA; IEEE, 2018: 7132-7141.