DOI: 10.16452/j.cnki.sdkjzk.2023.03.011 文章编号:1672-3767(2023)03-0094-09

基于卷积神经网络和混合注意力机制的书标检测算法

张 岩[1,2],赵蒙蒙[2],孙英伟[2],常艳康[2]
(1.青岛科技大学 图书馆,山东 青岛 266061; 2.青岛科技大学 机电工程学院,山东 青岛 266061)

摘 要:为实现图书馆中机器人智能排架,提出一种基于卷积神经网络和混合注意力机制的书标检测模型。将DenseNet121 引入 YOLOv4 以提高特征和梯度之间的传递效率,利用 SPDC 模块实现局部和全局特征融合,进而通过通道和空间混合注意力提高模型的特征表征能力。实验结果表明,模型的平均准确率、整体性能、参数量和模型大小均优于对比方法,且易于部署到嵌入式设备中实现在线检测,从而提高图书乱架治理的智能化水平。

关键词:卷积神经网络;混合注意力机制; 书标;目标检测;智慧图书馆
中图分类号:TP393;G250.7 **文献标志码**:A

Title label detection using convolutional neural network and hybrid attention mechanism

ZHANG Yan[1,2], ZHAO Mengmeng[2], SUN Yingwei[2], CHANG Yankang[2]
(1. Library of Qingdao University of Science and Technology, Qingdao 266061, China;
2. College of Electromechanical Engineering, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: To realize robot intelligent shelving in library, this paper proposed a title label detection model based on convolutional neural network and hybrid attention mechanism. DenseNet121 was applied to YOLOv4 to improve the transfer efficiency between features and gradients. The spatial pyramid dilated convolution(SPDC) module was used to achieve local and global feature fusion. Then the model's feature representation ability was improved through channel and spatial attention. Experimental results show that the average accuracy, overall performance, parameter amount and model size of the proposed model outperforms those of the compared methods and it is easy to deploy to embedded devices to achieve online detection, thus improving the intelligent level of book disorder management.

Key words: convolutional neural network; hybrid attention mechanism; title label; object detection; smart library

随着计算机视觉技术的快速发展,视觉检测技术受到广泛关注。为实现智能图书馆建设,研究人员基于计算机视觉技术对图书书脊分割、书标识别等问题进行了广泛研究,并提出诸多识别算法。Lee 等[1]利用互补金属氧化物半导体(complementary metal-oxide-semiconductor,CMOS)相机获取书脊图像,将颜色信息转化为直方图,通过判断检测结果的直方图是否与预存图像的直方图相匹配,来验证图书摆放是否正确。该方法检测前需图书摆放整齐且图像完整。方建军等[2]提出一种图书索书号提取分割算法,将 HSV(hue saturation value)空间分布特征与霍夫变换结合,利用边缘检测实现对书脊的分割。该方法涉及较多经验阈值,易受环境影响,鲁棒性较差。

目前,卷积神经网络由于其特征提取鲁棒性强、泛化能力好、可以实现端到端检测等优点,被广泛应用于目标检测等领域并取得较好效果。作为计算机视觉技术的基本问题之一,目标检测可为图像和视频的语义理解提供有价值的信息,被广泛应用于图像分类、人类行为分析、人脸识别和自动驾驶等领域。目标检测算

收稿日期:2022-09-29
基金项目:国家自然科学基金项目(62172248);山东省自然科学基金项目(ZR2019MEE066)
作者简介:张 岩(1980—),男,山东菏泽人,副教授,硕士生导师,主要从事计算机视觉和人工智能的研究.
E-mail:zy@qust.edu.cn

法主要分为两类:一类是首先生成区域建议,然后将每个建议分类为不同的对象类别,如 R-CNN[3]、Faster R-CNN[4]等两阶段目标检测算法;另一类是将目标检测视为回归或分类问题,采用统一的框架直接获得最终结果,如 YOLO、SSD(single shot multibox detector)[5]等单阶段目标检测算法。这些方法中,以 YOLO 系列为代表的检测器效果较好。

2015 年,YOLOv1 目标检测算法[6]被提出,该算法采用回归思想,对目标框直接进行回归和类别分类,省掉了显式生成候选区域的过程,与两级检测器 Faster R-CNN 相比,能够大幅降低检测时间、提高检测效率。但由于目标漏检现象严重,而且对边界框的定位精度低,导致算法精度不高。YOLOv2 对网络结构进行多方面改进,包括引入 Anchor 机制和多尺度训练策略,并对网络损失函数中位置回归的损失部分进行了优化,在 Pascal VOC 2007 上的检测精度达到 78.6%[7]。然而,由于缺乏底层信息的特征层,YOLOv2 对小尺寸目标检测精度较低。在 YOLOv2 的基础上,YOLOv3 的主干网络被替换成 Darknet-53,进一步提升了网络的特征提取能力。YOLOv4 算法在 YOLOv3 的基础上引入 Mosaic 数据增强、CSP 模块、Mish 激活函数、空间特征金字塔池化,进一步提升了网络的特征提取能力。YOLOv5 通过引入 Focus 结构降低 Backbone 中提取特征信息的损失,利用自适应图片缩放技巧减少网络计算量,在灵活性和速度上优于 YOLOv4,但在检测性能上稍弱于 YOLOv4[8]。

为实现图书馆中机器人智能排架,相关学者利用深度学习模型开展了广泛研究。邓三鸿等[9]利用长短期记忆(long-short term memory,LSTM)模型和字嵌入实现中文图书标签分类,通过构建二元分类器解决多标签分类问题,取得不错的分类效果,但是 LSTM 结构复杂,网络训练难度大且耗时较长。王崇等[10]将边缘算子和直线检测算法结合,实现对图书书脊的提取,并基于 Faster R-CNN 对图书书标进行定位和提取,实现了图书智能识别,但是两级检测器需要先对候选框进行提取,检测速度慢。此外,其较大的网络参数量对算力有限的移动设备的网络训练和实时检测带来挑战。苏志芳等[11]将机器视觉和无线传输技术相结合,设计出基于数字信号处理器(digital singnal processor,DSP)的嵌入式图书清点装置,提高了图书乱架清点的效率,但该方法需要人工手持数据采集器提取书脊图像的特征,调用图书馆自动化集成系统(integrated library automation system,ILAS)模块得到索取号。刘芳等[12]基于 Mask R-CNN 引入 ResNet50 并结合 Triplet Loss 实现对甲骨文拓片的自动识别,准确率可达 95%。王晓刚等[13]基于 Mask R-CNN 实现对书脊图像的分割,通过文字识别技术实现书标提取。但是,由于两阶段的检测策略和较大的参数量导致算法模型较大,网络训练和检测速度较慢,难以在移动设备上实现实时检测。

针对现有检测算法的不足,为提升检测准确率和效率、降低训练复杂度,基于 YOLOv4 网络提出一种基于卷积神经网络和混合注意力机制的书标检测算法,以实现移动设备端的高精度实时检测。通过在主干特征提取网络中引入 DenseNet121,使特征和梯度之间的传递更有效,缓解梯度消失问题,降低网络训练难度和模型大小。同时,将空间金字塔空洞卷积(spatial pyramid dilated convolution,SPDC)模块引入特征融合网络中,实现局部特征和全局特征融合,提升特征表达能力;在特征融合网络中引入混合注意力模块,增强有用特征,弱化不相关特征,提升算法检测精度。

1　相关工作

1.1　卷积神经网络

卷积神经网络(convolutional neural network,CNN)[14]具有强大的特征提取能力、较少的算法参数以及可并行计算等优点,被广泛应用于图像分类和识别任务中。CNN 一般由一个或多个卷积层、池化层以及顶部的全连接层组成,其核心部分是卷积层。卷积层在图像特征提取过程中实质上是一种线性运算,对图像局部特征进行整合和归一化映射,提取图像底层特征的细节信息和高层特征的语义信息。不同卷积核大小可以提取不同的图像特征。

卷积层内的每个神经元通过卷积核与前一层的局部区域相连接。相比传统神经网络,神经元不用与图片上每个像素相连接,极大地减少了权重的数量,减少了训练的参数。在不同卷积核和激活函数的共同作用下,输出不同的特征图。卷积过程运算式为:

$$\boldsymbol{F}_k^t=\sigma\left(\sum_{r=U_k}\boldsymbol{V}_{rk}^t f_r^{t-1}+b_k^t\right)。\tag{1}$$

式中：$\boldsymbol{F}_k^t$ 为卷积层的特征图，t 表示输出特征图的空间位置，k 表示卷积核的通道数，U_k 为卷积核的集合，$\boldsymbol{V}_{rk}^t$ 为特征图 f_r^{t-1} 所对应的权值，r 表示卷积核的空间位置，b_k^t 为特征图的偏置，$\sigma()$ 为激活函数。

1.2 注意力机制

人类视觉系统会选择性地关注感兴趣区域并获得其详细特征信息，而忽略背景区域。受此启发，注意力机制被提出并成为深度学习领域最重要的概念之一，被广泛应用于计算机视觉和自然语言处理等领域[15]。注意力机制的核心思想是找出原始数据的相关性并突出其特征，在计算能力有限的情况下，该机制类似于一种算力资源分配模式，将有限的计算资源用于处理更重要的信息。根据应用场景和对象的不同，分为空间注意力、通道注意力、分支注意力、时间注意力和混合注意力。其中，空间注意力聚焦特征图上有效的空间信息，通道注意力通过特征内部之间的关系聚焦图像中的有用信息，分支注意力和时间注意力则分别用于关注重要的分支和时间信息。

由于卷积神经网络在特征提取过程中会产生大量的冗余特征，研究模型如何重点关注有用特征、弱化对不相关特征的关注成为重点。相关研究中，挤压与激励网络的通道注意力机制(channel attention mechanism，CAM)被广泛使用。CAM 通过最大池化(MaxPool)和平均池化(AvgPool)操作将特征图 $\boldsymbol{F}$ 在空间维度上进行压缩后，输入特征图的空间维数进行逐元素求和，最终生成通道注意力特征图：

$$\begin{aligned}M_c(\boldsymbol{F})&=\sigma(\mathrm{MLP}(\mathrm{AvgPool}(\boldsymbol{F}))+\mathrm{MLP}(\mathrm{MaxPool}(\boldsymbol{F})))\\&=\sigma(\boldsymbol{W}_1(\boldsymbol{W}_0(\boldsymbol{F}_{\mathrm{avg}}^c))+\boldsymbol{W}_1(\boldsymbol{W}_0(\boldsymbol{F}_{\mathrm{max}}^c)))。\end{aligned}\tag{2}$$

式中：$M_c(\boldsymbol{F})$为最终生成通道注意力特征图，C 为通道维度或通道数，μ 为 Sigmoid 激活函数，MLP 表示两层共享神经元，MaxPool($\boldsymbol{F}$)和 AvgPool($\boldsymbol{F}$)分别表示最大池化操作和平均池化操作，$\boldsymbol{F}_{\mathrm{avg}}^c$和 $\boldsymbol{F}_{\mathrm{max}}^c$分别为通过全局平均池化和全局最大池化聚合通道信息后的特征，W_0 和 W_1 为感知机网络的共享参数。

CAM 可以增强网络的表征能力，关注重要特征，抑制不必要特征，但 CAM 只关注图像通道的重要性，而忽略了特征像素点间的注意力信息，造成空间上注意力信息的浪费。为了关注空间位置上的有用特征，空间注意力机制(space attention mechanism，SAM)将最大池化和平均池化在通道维度下得到的特征进行堆叠，经过 7×7 卷积核和 Sigmoid 激活函数得到的结果对特征空间权重归一化，最后将权重系数和输入特征相乘：

$$M_s(\boldsymbol{F})=\sigma(f^{7\times7}([\mathrm{AvgPool}(\boldsymbol{F}),\ \mathrm{MaxPool}(\boldsymbol{F})]))=\sigma(f^{7\times7}([\boldsymbol{F}_{\mathrm{avg}}^s,\boldsymbol{F}_{\mathrm{max}}^s]))。\tag{3}$$

式中：$M_s(\boldsymbol{F})$ 为最终生成空间注意力特征图，s 表示注意力机制中的空间维度或空间位置，$\boldsymbol{F}_{\mathrm{avg}}^s$ 和 $\boldsymbol{F}_{\mathrm{max}}^s$ 分别表示经全局平均池化和全局最大池化聚合空间信息后的特征。

SAM 可建立高纬度的空间特征相关性，但忽略了通道间特征的信息交互。为了汇总通道和空间的注意力信息，Woo 等[16]提出混合注意力模块(combined attention module，CBAM)，通过构建 CAM 和 SAM 串联通道和空间两个维度的注意力信息获取更可靠的特征，实现对计算资源的合理分配。

2 基于 YOLO 和混合注意力机制的书标检测算法

本研究提出的书标检测算法是将 YOLOv4 目标检测网络作为基础网络改进，网络结构由主干特征提取网络、特征融合网络和 YOLO 检测头 3 部分组成，整体结构如图 1 所示。首先，对输入图像进行预处理，然后利用主干特征提取网络 DenseNet121 模型进行特征提取，通过建立不同层之间的连接关系，减轻梯度消失问题。其次，将提取的特征图通过特征融合层 CBAM 进行 3 个不同尺度目标的特征融合，以增强网络对不同尺度目标的表征能力。同时，为提取不同尺寸的空间特征信息，提升算法对于空间布局和物体变性的鲁棒性，在特征融合网络中引入 SPDC 模块，以更好地实现对书标细线边缘等小目标特征的提取。最后，通过网络中 YOLO 检测头实现目标的定位和分类。

2.1 主干特征提取网络

YOLOv4 中的主干特征提取网络的使用基于 CSPNet 和 DarkNet53 的 CSPDarkNet53 结构，比 Dark-

Net53 具有更好的识别精度和计算效率。考虑 CSPDarkNet53 结构仍然具有大量的网络参数,不适合在计算资源有限的移动平台上部署,引入 DenseNet121 对 YOLOv4 主干特征提取网络进行改进。DenseNet121 通过建立不同层之间的连接关系充分利用特征信息,进一步减轻梯度消失问题。同时利用 Transition layer 缩减网络宽度,使网络参数量和计算量降低,有效抑制过拟合现象。

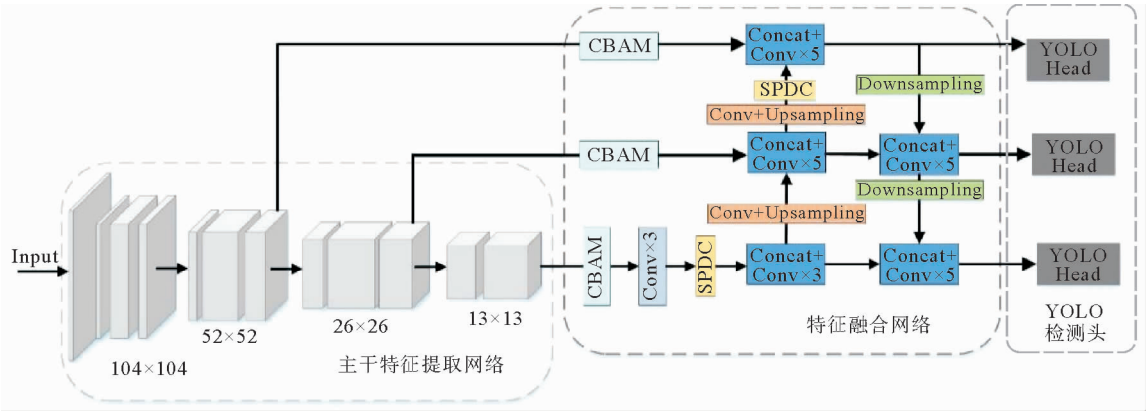


图 1　基于卷积神经网络和混合注意力机制的书标检测网络结构

Fig. 1　Network structure of the proposed detection model for title label

2.2　SPDC 模块

为检测不同尺度目标,通常在网络结构中采用具有不同感受野的空间金字塔池化(spatial pyramid pooling,SPP)模块。YOLOv4 中的 SPP 模块如图 2(a)所示,在检测任务中 SPP 被用于提取多尺度目标的特征信息,分离出具有不同感受野的最显著的上下文特征。但 SPP 模块中的池化操作易丢失信息使图片变模糊,考虑到图书书标的特点,应在网络中尽量减少池化操作,提取特征时尽量保留精细结构特征。因此,本研究将 SPP 模块中的池化操作用空洞卷积代替。图 2(b)为 SPDC 模块,将原始 SPP 中尺寸大小分别为 5×5、9×9 和 13×13 的池化核用空洞率分别为 2、4、6,且卷积核大小为 3×3 的空洞卷积代替。然后,将输入特征与经过 3 个空洞卷积运算后的 3 个不同尺度的特征通过连接操作进行聚合。

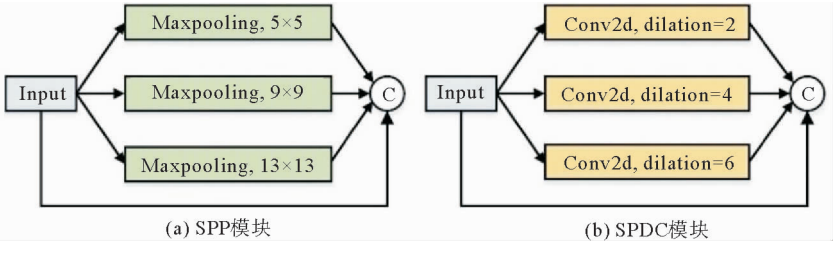


图 2　空间金字塔池化和空间金字塔空洞卷积模块结构示意图

Fig. 2　The architecture of the spatial pyramid pooling and spatial pyramid dilated convolution

2.3　混合注意力机制(CBAM)

由于神经网络在特征提取过程中会产生大量冗余特征。考虑到不相关特征对索书号区域检测的挑战,本研究将空间与通道混合注意力模块嵌入特征融合网络中,通过增大权重使网络更加关注目标区域有用特征、弱化不相关特征,以获得更精确的检测和定位。

CBAM 包含通道和空间注意模块,提供每个通道特征图中重要特征的位置信息,其结构如图 3 所示。在通道注意力模块中,特征图 $\boldsymbol{F}$ 通过最大池化和平均池化生成通道描述符,依次送入权重共享的两层神经元中,经过激活函数后,获得通道注意力特征图

$\boldsymbol{F}_1=\mu(\mathrm{MLP}(\mathrm{AvgPool}(\boldsymbol{F})+\mathrm{MaxPool}(\boldsymbol{F})))$。 (4)

在空间注意力模块中,基于通道注意力生成的特征图 $\boldsymbol{F}_1$,分别使用最大池化和平均池化生成特征描述符并按通道拼接,利用 7×7 卷积和激活函数生成空间注意力特征图

$\boldsymbol{F}_2=\mu(f^{7\times7}([\mathrm{AvgPool}(\boldsymbol{F}_1),\mathrm{MaxPool}(\boldsymbol{F}_1)]))$。 (5)

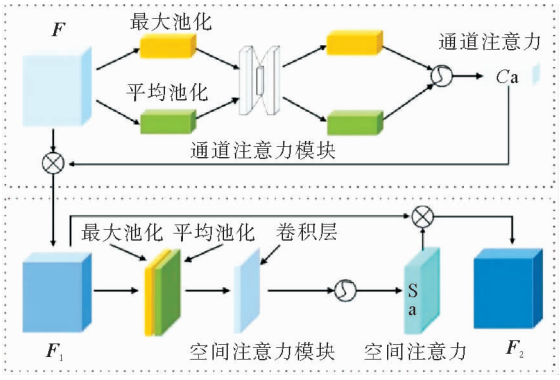


图 3 CBAM-空间与通道注意力模块

Fig. 3 CBAM-spatial and channel attention module

3 实验结果与分析

本研究所有实验均在相同的环境中进行,均使用默认参数在相同的训练和测试数据集上执行。

3.1 数据集

采用的数据集系青岛科技大学图书馆馆藏。构建数据集时,从 4 032×3 024 像素的图像中截取 608×608 像素包含图书及书标的图像作为数据集样本。采用 335 个含有索书号的图书图像和 LabelImg 工具制作相应的像素级高质量 XML 标签。为避免主观因素影响,标签由 3 名该研究方向研究生独立标记,然后通过投票确定。为获得合适的模型参数,使用交叉验证训练算法,数据集被随机划分为 70%的训练集、20%的验证集和 10%的测试集。为了提高训练速度,训练前将图片像素由 608×608 调整为 416×416。

3.2 实验设置

对于 YOLOv4 骨干网络,将迁移学习[17]在 Pascal VOC2007 公共数据集上训练得到的权重作为 DenseNet121 的初始权重。SGD 优化器用于最小化 YOLOv4 损失函数。为加速模型的收敛,前 35 个 Epoch 的初始学习率和 Batch_size 分别设置为 0.001 和 8。35 个 Epoch 之后,为了获得稳定的高质量模型,学习率和 Batch_size 分别降至 0.000 1 和 4。实验环境如表 1 所示。

表 1 实验环境

Table 1 Experimental environment

实验环境	实验配置
操作系统	Windows 10
编程语言	Python 3.7.6
深度学习框架	Pytorch 1.2.0
内存、线数	16 GB、6 核
GPU 型号	NVIDIA GeForce RTX 2060

3.3 评估指标

为准确评估模型的性能,使用目标检测常用的 4 个评估指标:精确率(P)、召回率(R)、F1 值(F_{M})和平均精度(A_{P})。计算式分别为:

$$P=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{P}}}, \tag{6}$$

$$R=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{N}}}, \tag{7}$$

$$F_{\mathrm{M}}=2\times\frac{P\times R}{P+R}, \tag{8}$$

$$A_{\mathrm{P}}=\int_0^1(P\times R)\mathrm{d}R。 \tag{9}$$

式中:T_{P}为检测到 Ground truth 的检测框数量,F_{P}为检测到同一个 Ground truth 的冗余检测框的数量,F_{N}为没有检测到 Ground truth 的数量。高精确率表示可以正确检测到更多的图书索书号,高召回率表示检测过程中遗漏的索书号更少。A_{P}是 precision-recall 曲线下的面积,被用于计算索书号的不同召回率级别(0~1)的平均精确率。F_{M}值是通过计算精确率和召回率的加权调和平均值来评估分类器的性能,其值范围在 0~1 之间。本研究主要以A_{P}作为模型性能的参考标准。

3.4 消融和对比实验

为验证所提算法的有效性,进行消融和对比实验。首先验证 DenseNet121 的有效性,其次进行 SPDC 模块不同空洞率的对比实验和不同注意力模块的消融实验,最后验证使用 SPDC 和 CBAM 对本算法性能的

影响。此外,与两阶段检测算法 Faster R-CNN、单阶段检测算法 SSD 和 RetinaNet[18] 3 种主流目标检测算法进行对比实验。

考虑到所采用数据集的规模,训练过程易产生过拟合现象,本研究将 DenseNet121 引入 YOLOv4 主干特征提取网络。2 种不同主干提取网络实验结果如表 2 所示,改进后的平均精度提升了 0.7%。这是因为神经网络每一层提取的特征都相当于对输入数据的一个非线性变换,而随着深度的增加,变换的复杂度也逐渐增加。相比于一般神经网络的分类器直接依赖于网络最后一层的特征,DenseNet 由于密集链接方式,提升了梯度的反向传播,可以综合利用浅层复杂度低的特征,因而更容易得到一个光滑的具有更好泛化性能的决策函数。

为验证 SPDC 模块中空洞率 2、4、6 组合为最优解,进行多组不同空洞率实验,如表 3 所示,当空洞率为 2、4、6 时,SPDC 模块可以将网络的平均精度提高到 97.44%,比其他组合都高。这是因为空洞卷积的空洞率过小,获得的感受野小于待提取的特征区域,使获得的局部信息过多,导致全局信息的丢失,影响识别率;若空洞率过大,获得的感受野大于待提取的特征区域,会忽略被检测的物体,导致目标成为背景,使细小特征检测效果较差。因此确定 2、4、6 作为 SPDC 模块中空洞卷积的空洞率。

表 2　不同主干提取网络性能对比

Table 2　Performance comparison of different backbones

主干网络	精确率/%	召回率/%	F1 值	平均精度/%
CSPDarkNet53	95.77	87.40	91	96.51
DenseNet121	93.02	92.07	93	97.21

表 3　SPDC 模块中不同空洞率下的性能对比

Table 3　Performance comparison of SPDC modules with different dilation rate

空洞率	精确率/%	召回率/%	F1 值	平均精度/%
[1,2,3]	89.97	95.24	92	97.03
[2,4,6]	90.84	96.02	93	97.44
[2,5,7]	90.47	95.68	93	97.25
[2,6,8]	90.42	95.63	93	97.11

为了证明 CBAM 比一般注意力模块更有效,分别与通道和空间注意力模块进行对比实验,如表 4 所示,在 YOLOv4 基础上引入 CAM,平均精度为 95.4%;在 YOLOv4 基础上引入 SAM,平均精度为 97.3%;而在 YOLOv4 引入 CBAM 后,能够获取更全面可靠的注意力信息,检测精度进一步提升为 98.7%。

表 4　不同注意力模块的消融实验

Table 4　Ablation experiment of different attention modules

算法	精确率/%	召回率/%	F1 值	平均精度/%
YOLOv4+CAM	86.14	93.50	90	95.40
YOLOv4+SAM	95.22	93.09	94	97.30
YOLOv4+CBAM	90.79	96.14	93	98.70

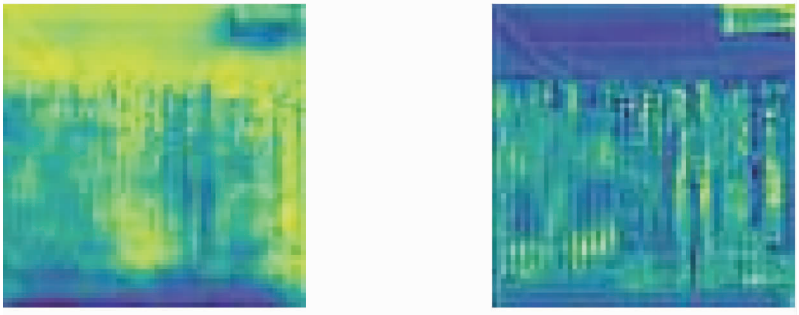
表 5 给出了本研究算法和对比算法在书标数据集上的检测结果。从表 5 中可以看出,两阶段目标检测网络 Faster R-CNN 的平均精度值最低,仅为 91.08%,单阶段目标检测网络 YOLOv4 的平均精度达 96.51%,比 Faster R-CNN 模型的平均精度提高 5.43%。在单阶段目标检测网络中,YOLOv4 的检测性能也优于 SSD 和 RetinaNet。

表 5　对比算法的实验结果

Table 5　Experimental results of algorithms in comparison

算法	主干网络	精确率/%	召回率/%	F1 值	平均精度/%
Faster R-CNN	VGG	60.79	97.36	75	91.08
SSD	VGG	97.84	82.93	90	96.30
RetinaNet	ResNet50	96.46	83.13	89	94.84
YOLOv4+SPDC	DenseNet121	90.84	96.02	93	97.44
YOLOv4+CBAM	DenseNet121	90.79	96.14	93	98.70
本算法	DenseNet121	90.81	97.22	93	98.85

如表 5 所示,在引入 DenseNet121 作为 YOLOv4 的主干网络的基础上,引入 SPDC 模块后,比主干特征提取网络为 DenseNet121 的 YOLOv4,平均精度提高了 0.23%,表明 SPDC 模块能够提升特征的表达能力,适用于检测目标差异较大的情况。同时,引入 CBMA 混合注意力机制之后,相比标准 YOLOv4 和改进后的 YOLOv4,检测结果平均精度分别提高了 2.19%和 1.51%,表明 CBAM 可以增强有用特征,弱化无效特征。图 4 给出了在同等条件下未引入 CBAM 和引入 CBAM 时模型特征图的对比结果,可以看出,引入 CBMA 后网络的特征提取能力有所增强,图书和书脊边界有用特征更加显著,而背景等无关信息得到了削弱。因此,在网络训练过程中,混合注意力模块能够显著加强对图书书脊中索书号区域特征信息的提取。



(a) 原特征图 (b) 引入CBAM后的特征图

图 4 特征可视化结果

Fig. 4 Feature visualization results

为了进一步验证本算法的性能优越性,图 5 给出对比算法的精确率-召回率曲线。通常,精确率-召回率曲线下的面积越大,检测精度越高。为获得更好的显示效果,将水平和垂直坐标范围均调整为[0.5,1.02]。使用式(9)计算书标检测结果的平均精度。当召回率小于 0.9 时,本算法的精确度保持相对稳定,而其他算法的准确度明显降低。此外,本算法的精确率-召回率曲线下的面积显著大于其他算法,表明本算法具有更高准确度。

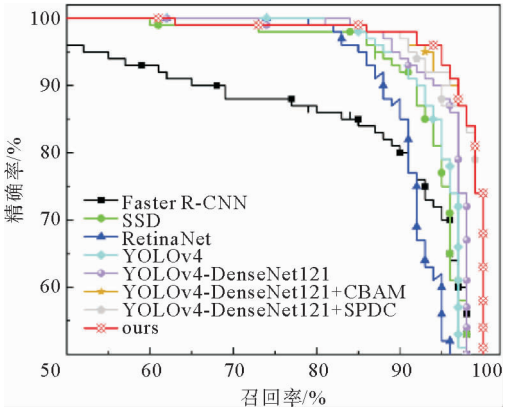


图 5 对比算法的精确率-召回率曲线

Fig. 5 Precision-recall curves of algorithms in comparison

3.5 检测效率

为检测本算法的优势,表 6 从模型大小、网络参数和检测时间方面给出了对比数据。从表 6 可以看出,Faster R-CNN 在模型大小和网络参数方面都相对较大,给算力带来了较大挑战,不适用于移动设备在线检测。相比之下,将 DenseNet121 引入 YOLOv4 中、通过 Concat 特征短路连接实现特征重用、采用较小的 growth rate,使得每层所独有的特征图较小,本算法模型大小减小到 61.7 M,参数量缩减为 17.11 M,实现了参数更小、计算高效,更适合部署于嵌入式设备。

此外,在主干网络 DenseNet121 的基础上,引入 SPDC 模块和 CBAM 混合注意力,使得模型大小和网络参数分别增加了 1.5 M 和 5.68 M,但精度提升较大,而且检测时间仅增加 1.4 ms 左右。

综上,相比标准 YOLOv4 网络以及其他主流算法,本算法能够更好地平衡检测精度和速度,且在实际乱架图书检测部署中,训练速度快,对硬件设备的性能要求低,更容易部署于移动式设备中实现实时检测。图 6 给出了本算法对图书书标检测的实际结果。

表 6　对比算法在大小、参数和时间方面的比较

Table 6　Comparison of different algorithms in terms of size, parameters and time

算法	主干网络	模型大小/M	网络参数/M	检测时间/ms
Faster R-CNN	VGG	521.0	137.08	108.29
SSD	VGG	90.5	26.15	105.03
RetinaNet	ResNet50	138.0	37.97	76.14
YOLOv4	CSPDarkNet53	244.0	64.36	70.64
YOLOv4	DenseNet121	61.7	11.43	68.50
YOLOv4+SPDC	DenseNet121	62.1	13.42	68.69
YOLOv4+CBAM	DenseNet121	62.4	16.78	69.57
本算法	DenseNet121	63.2	17.11	69.92

虽然本研究算法在书标检测精度和效率方面均取得了较好的效果,但是仍然存在 8%左右的检测失败样本,如图 7 所示,其中红色边界框表示算法检测的结果,黄色边界框表示检测出现的漏检情况。出现漏检的原因主要有:①由于时间久远,书标出现颜色变淡、字迹模糊或破损的情况,导致标签与背景对比度低,特征显著性变差;②图书较薄导致书标仅有较少部分得以展示,网络在特征提取的过程中会缺失大量细节信息。

图 6　书标检测结果示例

Fig. 6　Example of title label test results

图 7　图书索书号区域检测失败样本

Fig. 7　Sample of failed book claim number area detection

4　结论

本研究提出一种基于改进的 YOLOv4 的书标检测算法,相比改进前 YOLOv4 网络的平均精度提高了 2.34%。与其他现有主流算法相比,具有性能优势,且参数量和模型的大小均远低于对比方法,更容易部署到嵌入式设备中,能够满足基于机器视觉智能图书识别检测系统的要求,提高了图书乱架治理的智能化水平。未来考虑引入自注意力机制,对全局进行上下文建模,使网络注意到整个输入中不同部分的相关性,加深对细小、对比度低特征的理解,以减少误检和漏检的概率。此外,用于小样本学习的弱监督网络也是后续的研究方向。

参考文献:

[1] LEE D J,CHANG Y,ARCHIBALD J K,et al.Matching book-spine images for library shelf-reading process automation[C/OL]// IEEE International Conference on Automation Science and Engineering.IEEE,2008:738-743.DOI:10.1109/COASE.2008.4626503.

[2] 方建军,赵强强.图书馆在架图书的索书号图像提取与分割[J].北京联合大学学报(自然科学版),2015,29(1):87-92.
FANG Jianjun,ZHAO Qiangqiang.Extraction and segmentation of call number image for books on shelves[J].Journal of Beijing Union University,2015,29(1):87-92.

[3] GIRSHICK R,DONAHUE J,DARRELL T,et al.Rich feature hierarchies for accurate object detection and semantic segmentation[J/OL].2014 IEEE Conference on Computer Vision and Vison and Pattern Recognition(CVPR),2014:580-587.DOI:10.1109/CVPR.2014.81.

[4] REN S,HE K,GIRSHICK R,et al.Faster R-CNN:Towards real-time object detection with region proposal networks[J].IEEE Transactions on Pattern Analysis & Machine Intelligence,2017,39(6):1137-1149.

[5] LIU W,ANGUELOV D,ERHAN D,et al.SSD:Single shot multibox detector[C/OL]// Computer Vision-ECCV 2016:14th European Conference.Netherlands: Springer,Oct.11-14,2016.DOI:10.1007/978-3-319-46448-0_2.

[6] REDMON J,DIVVALA S,GIRSHICK R,et al.You only look once:Unified,real-time object detection[C/OL]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.IEEE,2016:779-788.DOI:10.1109/CVPR.2016.91.

[7] REDMON J,FARHADI A.YOLO9000:Better,faster,stronger[C/OL]// IEEE Conference on Computer Vision & Pattern Recognition.IEEE,2017:6517-6525.DOI:10.1109/CVPR.2017.690.

[8] ZHOU F,ZHAO H,NIE Z.Safety helmet detection based on YOLOv5[C/OL]// IEEE International Conference on Power Electronics,Computer Applications(ICPECA).IEEE,2021.DOI:10.1109/ICPECA51329.2021.9362711.

[9] 邓三鸿,傅余洋子,王昊.基于 LSTM 模型的中文图书多标签分类研究[J].数据分析与知识发现,2017,1(7):52-60.
DENG Sanhong,FU Yuyangzi,WANG Hao.Multi-label classification of Chinese books with LSTM model[J].Data Analysis and Knowledge Discovery,2017,1(7):52-60.

[10] 王崇.基于机器视觉的智能图书识别检测系统的设计[D].沈阳:沈阳工业大学,2020:6-24.
WANG Chong.Intelligent book recognition and detection system based on machine vision[D].Shenyang:Shenyang University of Technology,2020:6-24.

[11] 苏志芳,徐德刚,袁小一.基于机器视觉的图书乱架清点系统—以中南大学图书馆为例[J].高校图书馆工作,2020,40(6):60-65.
SU Zhifang,XU Degang,YUAN Xiaoyi.Disorganized books inventory system based on machine vision technology:Taking Central South University Library for example[J].Library Work in Colleges and Universities,2020,40(6):60-65.

[12] 刘芳,李华飙,马晋,等.基于 Mask R-CNN 的甲骨文拓片的自动检测与识别研究[J].数据分析与知识发现,2022,5(12):88-97.
LIU Fang,LI Huabiao,MA Jin,et al.Automatic detection and recognition of oracle rubbings based on Mask R-CNN[J].Data Analysis and Knowledge Discovery,2022,5(12):88-97.

[13] 王晓刚,钱思文,张继,等.基于计算机视觉和人工智能技术的图书馆图书盘点系统的探索与应用[J].图书馆杂志,2022,41(7):96-100.
WANG Xiaogang,QIAN Siwen,ZHANG Ji,et al.Exploration and application of library automatic book inventory checking system based on computer vision and artificial intelligence [J].Library Journal,2022,41(7):96-100.

[14] 季长清,高志勇,秦静,等.基于卷积神经网络的图像分类算法综述[J].计算机应用,2022,42(4):1044-1049.
JI Changqing,GAO Zhiyong,QIN Jing,et al.Review of image classification algorithms based on convolution neural network[J].Journal of Computer Applications,2022,42(4):1044-1049.

[15] 王旭强,岳顺民,张亚行,等.基于注意力机制的特征融合序列标注模型[J].山东科技大学学报(自然科学版),2020,39(5):79-88.
WANG Xuqiang,YUE Shunmin,ZHANG Yahang,et al.Attention based sequence labeling model with feature fusion[J].Journal of Shandong University of Science and Technology(Natural Science),2020,39(5):79-88.

[16] WOO S,PARK J,LEE J Y,et al.CBAM:Convolutional block attention module[C/CD]// Proceedings of the 15th European Conference on Computer Vision(ECCV 2018).Munich,Sep. 8-14,2018.

[17] 包翔,汪满容,刘桂锋.一种基于主题模型与迁移学习的文本分类方法[J].山东科技大学学报(自然科学版),2021,40(3):80-88.
BAO Xiang,WANG Manrong,LIU Guifeng.A novel text classification method based on topic model and transfer learning [J].Journal of Shandong University of Science and Technology(Natural Science),2021,40(3):80-88.

[18] LIN T Y,GOYAL P,GIRSHICK R,et al.Focal loss for dense object detection[J/OL].IEEE Transactions on Pattern Analysis & Machine Intelligence,2020.DOI:10.1109/tpami.2018.2858826.

(责任编辑:傅 游)