文章编号:2096-1472(2023)08-0035-05
DOI:10.19644/j.cnki.issn2096-1472.2023.008.008

基于改进卷积神经网络的图像数字识别方法研究

王耀宗,张易诚,康宇哲,沈 炜

(浙江理工大学计算机科学与技术学院,浙江 杭州 310018)

✉ 781593022@qq.com; 1138263774@qq.com; kangyuzhe2018@gmail.com; latitude@126.com

摘 要:针对试卷分数的统计问题,采用一种带有特殊分值框的试卷,并提出了一种基于改进卷积神经网络的识别统计方法。首先基于 YOLO 目标检测算法对分值框进行定位,并引入膨胀卷积模块丰富感受野、调整边框损失函数、提高收敛速度,然后基于 ResNet 卷积神经网络对分数进行识别,并融合注意力机制提高特征提取能力。实验结果表明,经改进的模型对 1 000 份试卷中题目分数的识别准确率为 99.2%,可以准确、高效地识别试卷图像中的分数。

关键词:目标检测;损失函数;ResNet;注意力机制;试卷分数识别

中图分类号:TP391 **文献标志码**:A

Research on Image Digital Recognition Method Based on Improved Convolutional Neural Network

WANG Yaozong, ZHANG Yicheng, KANG Yuzhe, SHEN Wei

(*School of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou* 310018, *China*)

✉ 781593022@qq.com; 1138263774@qq.com; kangyuzhe2018@gmail.com; latitude@126.com

Abstract: Aiming at the statistical problem of test paper scores, this paper proposes a recognition statistics method based on improved convolutional neural network by using a test paper with a special score box. Firstly, the YOLO object detection algorithm is used to locate the score boxes, and the dilated convolutional module is introduced to enrich the receptive field and adjust the border loss function of the frame to improve the convergence rate. Then, the score is recognized based on ResNet convolutional neural network, and the attention mechanism is integrated to improve the feature extraction ability. The experimental results show that the improved model has a recognition accuracy of 99.2% for question scores in 1 000 test papers, and can accurately and efficiently recognize scores in test paper images.

Key words: object detection; loss function; ResNet; attention mechanism; score recognition of test papers

0 引言(Introduction)

纸质试卷作为教学考试的重要载体,被广泛应用于各种考试中。一直以来,试卷分数的统计主要有两种方式:一种是人工阅卷,教师通过口算或计算器对试卷分数进行汇总;另一种是使用特定的答题卡,通过光标阅读器对答题卡进行扫描,进而统计分数。前者不仅需要消耗大量的人力和时间,而且效率低、错误率高,后者需要制作特定的答题卡且光标阅读器价格昂贵,无法大规模应用[1]。

随着卷积神经网络的发展,数字识别技术在教学考试中得到了广泛应用。周铁军等[2]基于 Keras 构建了可以识别分数为 70 分以内的卷积神经网络,但该网络的识别准确率不高,只有 94%。仝梦园等[3]提出了一种融合贝叶斯分类器的分数识别算法,该算法虽然有着较高的识别率,但是无法准确地定位题目的分数,存在漏检风险。

针对上述问题,本文采用一种带有特殊分值框的试卷,并在此基础上提出了一种基于改进卷积神经网络的试卷分数识

收稿日期:2023-02-27

别方法。该方法分为两个部分，第一部分基于 YOLO 目标检测算法对该分值框进行定位，并引入膨胀卷积模块丰富感受野、优化调整边框损失函数、提高收敛速度[4]。第二部分基于 ResNet 卷积神经网络对分数进行识别，并融合注意力机制提高特征提取能力[5]。实验结果表明，该方法可以准确高效地识别试卷图像中的分数，大幅度降低人工阅卷的工作量。

1 模型构建(Model building)

1.1 YOLOv5

1.1.1 YOLOv5 模型结构

相较于 R-CNN(区域卷积神经网络)、Fast R-CNN(快速区域卷积神经网络)等二阶段(Two Stage)算法，作为一阶段(One Stage)算法的 YOLO 有着更快的检测速度，并且随着 YOLO 版本的不断迭代，其准确率也有了很大的提高[6-7]。本文基于 YOLOv5 目标检测算法实现对分数框的定位，该算法结构由输入端、Backbone(骨干)、Neck(瓶颈)和 Head(头部)四个部分组成，如图 1 所示。

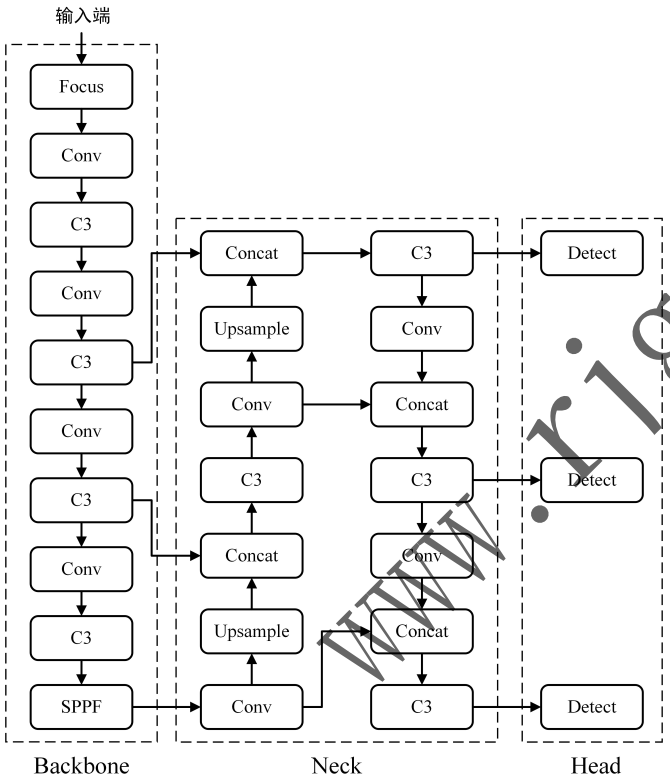


图 1 YOLOv5 结构图

Fig. 1 The structure of YOLOv5

输入端使用 Mosaic 数据增强方式随机选取数据集中的四张原始图片进行随机裁剪、缩放和排布，拼接为一张图片作为输入，这样不仅可以大幅度丰富数据集，还能丰富目标物体的尺度，提高整个网络的鲁棒性。

Backbone 中使用了 Focus 结构和 C3 结构。Focus 使用切片操作把一个大尺寸特征图拆分成多个小尺寸特征图，然后进行 Concat(拼接)操作。C3 是基于 CSP 结构将原本的残差结构分为两个部分，一部分仅经过一个基本的卷积运算模块，另一部分使用了多个残差块堆叠，最后将两个部分进行 Concat 操作，这样不仅可以实现更丰富的梯度组合，还能在增强网络特征提取能力的同时，大幅度减少参数量[8]。

Neck 中使用了特征金字塔网络(Feature Pyramid Networks，FPN)和路径聚合网络(Path Aggregation Network，PAN)相结合的结构[9-10]。FPN 自顶向下通过下采样将高阶的强语义特征和低层的位置信息融合，而 PAN 自底向上通过上采样将低层的强定位特征传递到高层，最终实现高层特征与底层特征相互融合和相互补充，丰富了模型的分类和定位能力。

Head 中使用三个 Detect 检测器分别对不同尺寸的特征图进行检测，确保了网络对大目标和小目标都有不错的检测效果。

1.1.2 YOLOv5 融合膨胀卷积模块

膨胀卷积(Dilated Convolution)通过在标准的卷积核中注入空洞增加模型的感受野[11]。相比常规的卷积操作，膨胀卷积通过膨胀率控制卷积核的膨胀程度。感受野的计算公式如下：

$$RF_i = RF_{i-1} + (k-1) \times s \tag{1}$$

其中，RF_i 是当前感受野，RF_{i-1} 是上一层感受野，k 是卷积核的尺寸，s 是步长。

本文使用膨胀率为 3 和 5、卷积核大小为 3×3 的膨胀卷积对特征图进行卷积操作，然后将两个输出特征图拼接构成膨胀卷积模块(Dilated Convolution Module，DCM)，如图 2 所示。目标检测为了保证感受野需要依靠下采样，但是下采样会导致图像的位置信息丢失，因此本文在 YOLOv5 中 SPPF(Spatial Pyramid Pooling Fast)层后添加一个该模块的分支，并将输出特征图送入第三个 Detect 检测器，以此丰富不同感受野的上下文信息。

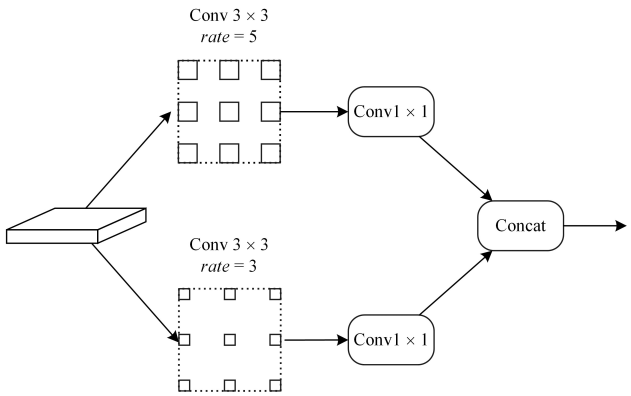


图 2 膨胀卷积模块

Fig. 2 Dilated convolution module

1.1.3 改进边界回归损失函数

YOLOv5 使用 CIoU 作为衡量边框损失的函数，如公式(2)和公式(3)所示，b 和 b^{gt} 分别代表预测框和真实框，IoU 为预测框和真实框的交并比，$\rho^2(b,b^{gt})$ 表示预测框与真实框中心点的欧式距离，c 表示预测框与真实框的最小外接矩形的对角线距离。

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b,b^{gt})}{c^2} + \frac{v^2}{1 - IoU + v} \tag{2}$$

$$\upsilon=\frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}}-\arctan\frac{w}{h}\right)^2 \tag{3}$$

ZHANG 等[12]在 CIoU 的基础上提出了 EIoU 损失函数。EIoU 不仅考虑了 CIoU 已经涉及的重叠面积和中心点距离的影响,还考虑了边长的影响,表达式如下:

$$\begin{aligned}L_{EIoU}&=L_{IoU}+L_{dis}+L_{asp}\\&=1-IoU+\frac{\rho^2(b,b^{gt})}{c^2}+\frac{\rho^2(w,w^{gt})}{c_w^2}+\frac{\rho^2(h,h^{gt})}{c_h^2}\end{aligned} \tag{4}$$

c_w 和 c_h 是包含预测框和真实框的最小闭包区域的宽度和高度,w 和 w^{gt} 表示预测框与真实框的宽度,$\rho^2(w,w^{gt})$ 表示预测框与真实框宽度的欧式距离,h 和 h^{gt} 表示预测框与真实框的高度,$\rho^2(h,h^{gt})$ 表示预测框与真实框高度的欧式距离。

GEVORGYAN[13]提出了 SIoU 损失函数,SIoU 考虑了角度损失、距离损失、形状损失和 IoU 损失。角度损失的示意图(图 3)和表达式如下:

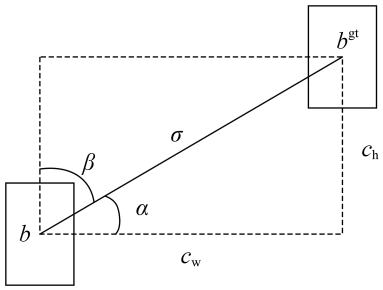


图 3　角度损失图

Fig. 3 Diagram of angle cost

$$\Lambda=1-2\sin^2\left(\arcsin x-\frac{\pi}{4}\right) \tag{5}$$

$$x=\frac{c_h}{\sigma}=\sin\alpha \tag{6}$$

$$\sigma=\sqrt{(b_{c_x}^{gt}-b_{c_x})^2+(b_{c_y}^{gt}-b_{c_y})^2} \tag{7}$$

$$c_h=\max(b_{c_y}^{gt},b_{c_y})-\min(b_{c_y}^{gt},b_{c_y}) \tag{8}$$

Λ 为角度损失,(b_{c_x},b_{c_y}) 代表预测框中心点坐标,$(b_{c_x}^{gt},b_{c_y}^{gt})$ 是真实框中心点坐标,σ 为预测框和真实框中心点的距离,c_h 为预测框和真实框中心点的高度差。形状损失的示意图(图 4)和表达式如下:

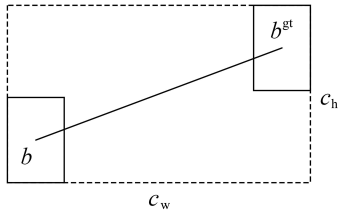


图 4　形状损失图

Fig. 4 Diagram of shape cost

$$\Omega=\sum_{t=w,h}(1-e^{\omega_t})^{\theta} \tag{9}$$

$$w_w=\frac{|w-w^{gt}|}{\max(w,w^{gt})},w_h=\frac{|h-h^{gt}|}{\max(h,h^{gt})} \tag{10}$$

其中,Ω 为形状损失,θ 为超参数,w 和 h 为预测框的宽和高,w^{gt} 和 h^{gt} 为真实框的宽和高。距离损失的表达式如下。

$$\Delta=\sum_{t=x,y}(1-e^{-\gamma\rho_t}) \tag{11}$$

$$\rho_x=\left(\frac{b_{c_x}^{gt}-b_{c_x}}{c_w}\right)^2,\rho_y=\left(\frac{b_{c_y}^{gt}-b_{c_y}}{c_h}\right)^2,\gamma=2-\Lambda \tag{12}$$

其中,Δ 为距离损失,c_w 和 c_h 为预测框和真实框的最小外接矩形的宽和高。最终 SIoU 损失函数表达式如下:

$$L_{SIoU}=1-IoU(b,b^{gt})+\frac{\Omega+\Delta}{2} \tag{13}$$

本文分别使用 EIoU 和 SIoU 替换原始的 CIoU,并在相同数据集中进行实验验证。

1.2 ResNet

1.2.1 ResNet 模型结构

相较于 LeNet 和 VGG 等相对浅层的网络,ResNet 使用了残差块结构,如图 5 所示,该结构将特征矩阵 $\boldsymbol{X}$ 进行卷积得到 $F(\boldsymbol{X})$,然后通过 shortcut(短接)分支传递特征矩阵 $\boldsymbol{X}$,再将 $\boldsymbol{X}$ 和 $F(\boldsymbol{X})$ 相加后进行激活[14-15]。ResNet 网络解决了网络过深时的梯度爆炸和梯度消失的问题,使得网络层数可达百层。

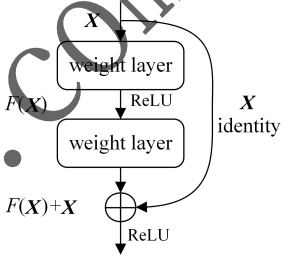
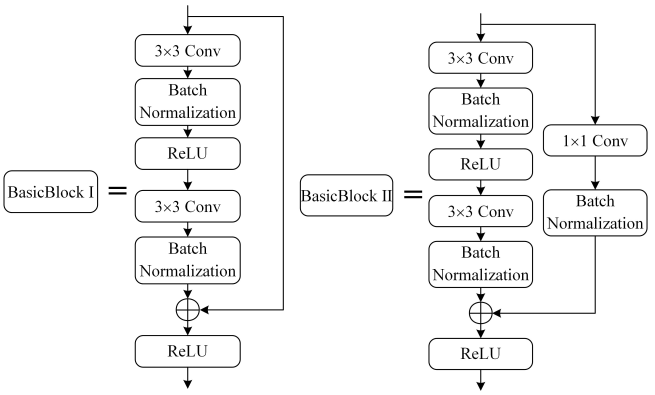


图 5　残差块结构图

Fig. 5 Residual block structure diagram

本文基于 ResNet18 卷积神经网络实现分数识别,该网络主要由基本残差块组成。图 6(a)为基本残差块Ⅰ,它包含一条直线路径和一条跳跃路径。直线路径由两个 3×3 的卷积层和两个批量归一化(Batch Normalization,BN)层组成,中间使用 ReLU 函数激活。跳跃路径则直接将输入特征图与直线路径的输出相加。最后将相加结果通过 ReLU 函数激活后输出。图 6(b)是基本残差块Ⅱ,它和普通残差块结构相似,区别是跳跃路径增加了一个 1×1 的卷积层和一个 BN 层,用于对输入特征图进行下采样。



(a)基本残差块Ⅰ　　(b)基本残差块Ⅱ

图 6　基本残差块结构图

Fig. 6 The structure diagram of basic residual block

ResNet18 结构如图 7 所示，输入图像先进行 7×7 的卷积和最大池化操作，然后经过 8 个基本残差块进行特征提取，最后经过平均池化和全连接输出预测概率。

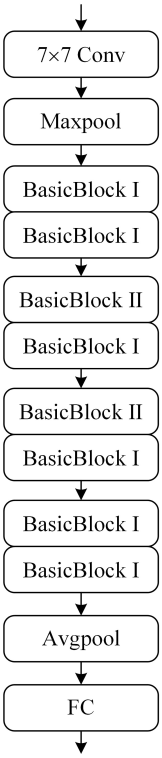


图 7 ResNet 结构图

Fig. 7 The structure diagram of ResNet

1.2.2 ResNet 融合注意力机制

当我们看到一张图像时，大脑会优先注意到图像中的重点信息，因而可能会忽视其他部分。注意力机制就是模仿人类大脑处理图像信息产生的，其可以使网络自动关注到一些重点区域。

卷积注意力模块（Convolutional Block Attention Module，CBAM）是一种轻量的注意力模块，结构如图 8 所示，其包括通道注意力模块（Channel Attention Module，CAM）和空间注意力模块（Spatial Attention Module，SAM）[16]。本文将 CBAM 融合到 ResNet 网络中，在网络的第二层和倒数第二层加入 CBAM，以此提高该网络在通道和空间上的特征提取能力。

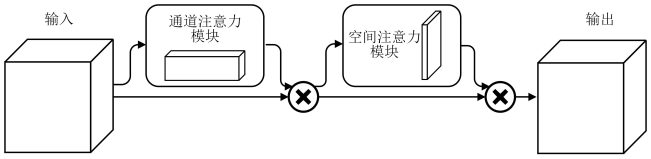


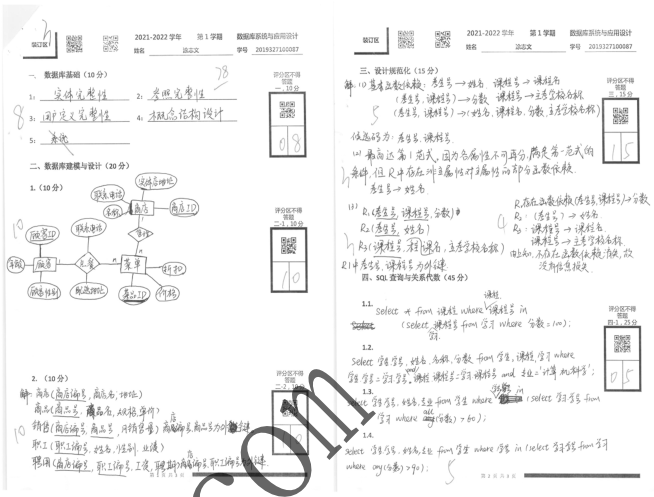
图 8 CBAM 结构图

Fig. 8 The structure diagram of CBAM

2 实验（Experiment）

2.1 数据集

本文所采用的试卷扫描图像示例如图 9 所示，试卷包含一种特殊的分值框，该分值框由两个部分组成，上半部分是可识别出题号信息的二维码，下半部分可供老师填写分数。除此之外，可通过试卷左上角的二维码识别学生的学号信息，这样可将学生的学号、题号和分数信息关联起来。本文共使用 1 000 张以 24 位、150×150 dpi 格式保存的浙江理工大学若干课程试卷扫描图像作为数据集，并按照 1∶1 的比例将其随机分为训练集和验证集。



（a）试卷第 1 页 （b）试卷第 2 页

图 9 试卷扫描图像

Fig. 9 Scanned image of examination paper

2.2 实验环境

本实验所使用的 CPU 为 Intel（R）Core（TM）i5-10400 CPU @ 2.90 GHz，操作系统为 Ubuntu 20.04，GPU 为 NVIDIA GeForce GTX1060，软件环境为 Pytorch1.12 和 Python3.7 等。

2.3 分值框识别结果

为了选择性能更好的 IoU 损失函数，在 YOLOv5 网络中基于相同数据集分别使用 CIoU、EIoU 和 SIoU 对模型进行训练，结果如表 1 和图 10 所示，其中 $mAP@0.5:0.95$ 表示 IoU 从 0.5 一直取值到 0.95，每隔 0.05 计算一次 mAP 的值。从图 10 中可以看出，相较于 CIoU，EIoU 和 SIoU 的收敛速度更快且 EIoU 和 SIoU 在 $mAP@0.5:0.95$ 指标上分别比 CIoU 高了 1.1% 和 1.0%，因此最终选择 EIoU 作为边界框的损失函数。

表 1 不同 IoU 的实验数据

Tab.1 Experimental result data of different IoU

边框损失函数	$mAP@0.5:0.95$/%
CIoU	96.9
EIoU	98.0
SIoU	97.9

为了验证膨胀卷积模块（DCM）的有效性，研究人员又使用了相同数据集分别对原始 YOLOv5 和融合膨胀卷积模块的 YOLOv5 进行实验，并采用 $mAP@0.5:0.95$ 作为评价指标，结果如表 2 所示，YOLOv5-EIoU-DCM 组合相较于 YOLOv5-CIoU 组合和 YOLOv5-CIoU-DCM 组合分别有 1.7% 和 1.1% 的提升，证明了该模块的有效性，因此使用该组合对试卷图像

中的分值框进行识别,试卷的识别效果如图 11 所示。

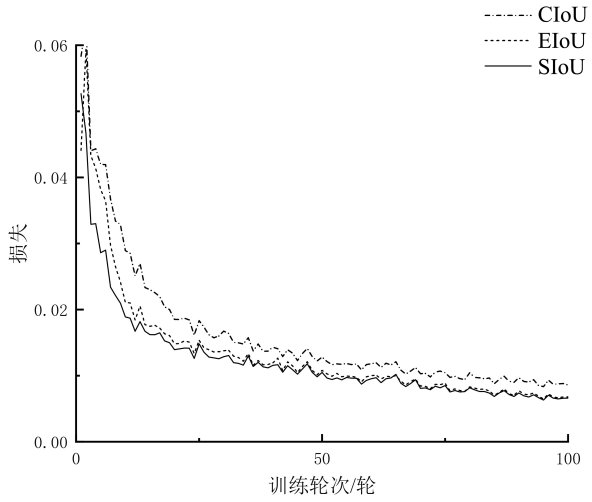


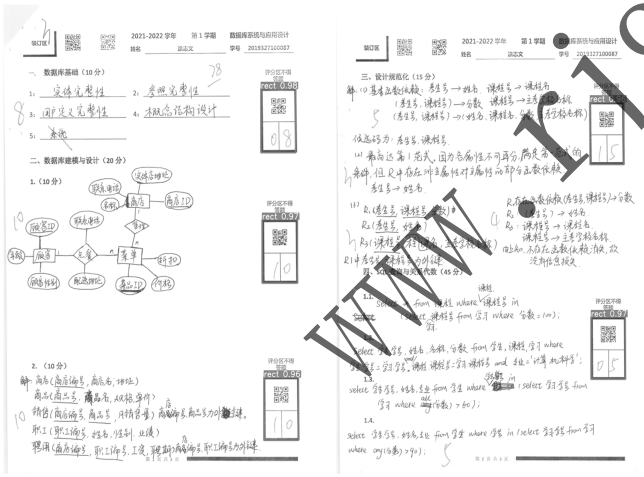
图 10　不同 IoU 方法的损失曲线

Fig. 10 Loss curve of different IoU methods

表2　YOLOv5 实验数据

Tab.2 YOLOv5 experimental result data

网络	*mAP*@0.5∶0.95/%
YOLOv5-CIoU	96.9
YOLOv5-CIoU-DCM	97.5
YOLOv5-EIoU-DCM	98.6



(a)试卷第 1 页　　(b)试卷第 2 页

图 11　识别效果图

Fig. 11 Recognition renderings

2.4　分数识别结果

本文使用 ResNet 卷积神经网络对试卷中分数进行识别。图像共有 11 个分类,即 0～9 共计 10 个分类和空白分类。因为手写数字总是会有一些倾斜角度,并且扫描的试卷难免会有一些噪声,所以需要对图像进行预处理,将图像在(−30°,30°)内随机旋转一定角度,并添加了 10%的椒盐噪声。实验结果如表 3 所示,相较于原始的 ResNet 网络,融合了 CBAM 的 ResNet 网络虽然 *FPS*(帧率)降低了 6.6%,但是准确率提高了 0.4%。

表3　ResNet 实验数据

Tab.3 ResNet experimental result data

网络	*FPS*	准确率/%
ResNet	70.1	98.9
ResNet-CBAM	65.5	99.3

2.5　实验结果与分析

为了验证改进后网络的识别效果,研究人员分别使用 4 种不同的网络组合对试卷扫描图像进行识别。结果如表 4 所示,相较于原始的 YOLOv5-CIoU 与 ResNet 的组合,YOLOv5-EIoU-DCM 与 ResNet-CBAM 的组合虽然 *FPS* 降低了约 7.7%,但准确率提高了 2.2%,证明了改进模型的有效性。

表4　实验结果数据

Tab.4 Experimental result data

网络	*FPS*	准确率/%
YOLOv5-CIoU+ResNet	5.2	97.0
YOLOv5-EIoU-DCM+ResNet	4.9	97.2
YOLOv5+ResNet-CBAM	5.0	98.5
YOLOv5-EIoU-DCM+ResNet-CBAM	4.8	99.2

3　结论(Conclusion)

针对试卷分数的统计问题,本文采用一种带有特殊分值框的试卷,并提出一种基于改进卷积神经网络的分数识别方法。该方法首先基于 YOLOv5 目标检测算法对试卷中的分值框进行定位,然后基于 ResNet 卷积神经网络对分数进行识别。研究人员在 YOLOv5 中引入了膨胀卷积模块并优化调整了边框损失函数,相较于原始的 YOLOv5,YOLOv5-EIoU-DCM 组合在 *mAP* 指标上有 1.7%的提升。此外,本文研究将 CBAM 融入 ResNet 中,相较于原始 ResNet,融合了 CBAM 的 ResNet 准确率有 0.4%的提高。最终结果表明,改进模型对题目分数的识别率更加优秀,识别准确率为 99.2%,可以准确且高效地识别试卷图像中的分数。

参考文献(References)

[1] 程淑红,尚果超. 基于视觉的答题卡自动判分系统设计[J]. 计量学报,2018,39(6):804-810.

[2] 周铁军,王外忠,蔡玉婷,等. 试卷手写分值识别方法研究[J]. 信息技术与信息化,2021(9):23-25.

[3] 仝梦园,金守峰,陈阳,等. 改进卷积神经网络的手写试卷分数识别方法[J]. 西安工程大学学报,2020,34(4):80-85.

[4] REDMON J,DIVVALA S,GIRSHICK R,et al. You only look once:unified,real-time object detection[C]//Institute of Electrical and Electronics Engineers. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Computer Society,2016:779-788.

(下转第 57 页)