文章编号：2095-2163(2022)12-0164-05 中图分类号：TP391 文献标志码：A

基于改进 Faster R-CNN 的算式检测与定位

王 巍，周庆华
（长沙理工大学 物理与电子科学学院，长沙 410004）

摘 要：算术题批改是小学数学老师的一项重要任务。为了提高批改效率，可使用机器视觉的方法来检测和识别。算式检测与定位的准确性会影响后续的识别与批改结果，为了提高其准确性，提出了一种基于改进 Faster R-CNN 的基础算式检测与定位的方法。通过聚类分析数据集中算式的参数，对区域建议网络（Region Proposal Network，RPN）中锚框（Anchor boxes）的尺寸和比例进行了调整，减少了训练中的冗余计算，提高了收敛速度；同时用 ROI Align 替换 ROI Pooling，避免了 2 次量化对检测精度带来的影响。实验表明，改进的 Faster R-CNN 提升了基础算式的检测定位效果。

关键词：算式检测；更快的区域卷积神经网络；聚类；感兴趣区域对准；深度学习

Detection and location of formulas based on improved Faster R-CNN

WANG Wei, ZHOU Qinghua
(School of Physics and Electronics, Changsha University of Science and Technology, Changsha 410004, China)

【Abstract】 Correcting formula exercise is an important task for primary school teachers. In order to improve the efficiency of grading, machine vision methods can be used to detect and recognize. The accuracy of formulas detection and location will affect the results of recognition and correction. The paper proposes a method of basic formulas detection and location based on improved Faster R-CNN. Through clustering analysis parameters of the formulas in the dataset, the scales and ratios of the anchors are adjusted in Region Proposal Network, which would reduce the redundant calculation in the training to improve the speed. At the same time, ROI Align is used to replace ROI Pooling to avoid the impact of twice quantization. The experiments show that the improved Faster R-CNN improves the detection and location effect of basic formulas.

【Key words】 arithmetic formula detection; faster R-CNN; clustering; ROI align; deep learning

0 引 言

小学四则混合运算题一般由加、减、乘、除、括号和数字组成，数字又包含印刷体和手写体两种类型。从广义上来讲，用机器视觉方法对这种基础算式的检测与识别，属于光学字符识别（Optical Character Recognition，OCR）。传统 OCR 技术主要分为文字区域定位、行列分割、分类器识别等几个步骤，其中文字区域定位、行列分割在本质上就是文本检测与定位。研究可知，文字区域定位是对文字颜色、亮度、边缘等信息进行聚类，进而分离出文字区域与非文字区域。一般情况下多是采用投影法[1]进行行列分割，行列分割的目的是提取出单字，主要方法是利用文字在行列间存在间隙的特征，由此来找出行列分割点。在背景单一、数据简单、容易分割的情况下，比如车牌识别[2-3]和身份证识别[4-5]等，传统 OCR 方法一般都能达到较好的效果。然而，在自然场景中的图像文本检测[6]时，往往存在背景复杂、有噪声干扰、字符之间有粘连/重叠等情况。具体来说，本文研究的算术题题卡一般是由家长自行打印，打印纸中会有背景图案干扰，手写答案也会有粘连/重叠的情况，使得传统 OCR 方法的区域定位和行列分割准确度大打折扣。在此背景下，近年来提出的深度学习[7]OCR 技术就表现出明显优势。

与传统 OCR 方法相比而言，深度学习 OCR 算法无需进行文字的单字分割，可以直接对整行文字进行识别[8]。因此，在深度学习的 OCR 技术中，使用合适的目标检测算法对文本进行检测与定位是至关重要的一个环节。Ren 等人[9]在 2017 年提出了 Faster R-CNN，被证实是一种较高效的目标检测算法。由于 Faster R-CNN 的检测对象只是 PASCAL VOC 2007 数据集上的 20 类目标，故对其它特定目标的检测效果并不理想。因此，后期又相继提出了基于 Faster R-CNN 的改进算法[10~13]。冯小雨等

基金项目：国家自然科学基金(42074198)。

作者简介：王 巍(1988-)，男，硕士研究生，主要研究方向：数字图像处理；周庆华(1977-)，男，博士，教授，博士生导师，主要研究方向：人工智能及其应用。
通讯作者：周庆华 Email：zhouqinghua@csust.edu.cn

收稿日期：2022-03-22

人[14]对 Faster R-CNN 算法进行改进，将其专门用来检测空中目标。黄继鹏等人[15]提出面向小目标的多尺度 Faster R-CNN 检测算法，提高了 Faster R-CNN 在小目标检测任务上的平均精度。王宪保等人[16]提出了一种分裂机制的改进 Faster R-CNN 算法，获得了比原始 Faster R-CNN 更好的检测效果。黄宁霞等人[17]增加基础网络的深度，采用双线性插值和 soft-NMS[18]等方法改进 Faster R-CNN，在不同场景的人行道障碍物检测中获得了不错的鲁棒性。

然而，本文处理的基础算式具有长短不一、手写数字随机、定位要求高等特点，上述改进算法不适用于算式的检测与定位。考虑到基础算式的上述特点，本文提出了一种基于聚类的快速区域卷积网络（faster region based convolutional neural network based on clustering，CF R-CNN）算法，CF R-CNN 算法主要有 2 方面的改进：

（1）用 K-means 聚类算法[19]得出更适合的 anchor 参数，使模型收敛更快。

（2）参考 Mask R-CNN[20]的处理方法，把 ROI Pooling 改为 ROI Align，避免了 ROI Pooling 中 2 次量化带来的影响，提高了检测精度。实验表明，在基础算式检测定位中，本文提出的 CF R-CNN 有更好的性能。

1　数据集

为了完成用于基础算式检测定位的 CF R-CNN 网络的训练，研究制作了一个包含每个算式定位框信息的题卡数据集。考虑到教师和家长制作题卡并进行识别的常见场景，每页题卡采用每行 2 个算式、每页 25 行、A4 纸打印的形式。研究中以随机的方式生成了 500 页的口算题，数据集的一个样本页如图 1 所示，每个算式由 2~3 个整数或小数的四则混合运算组成，部分含有括号、铅笔书写答案，用手机摄像头拍照。

图 1　数据集样本示例

Fig. 1　Sample of the dataset

2　CF R-CNN 算法

CF R-CNN 的网络结构如图 2 所示。选择 Faster R-CNN 作为基础的网络框架，主要由特征提取网络、区域建议网络、ROI Align 和分类回归网络组成。对此拟做阐释分述如下。

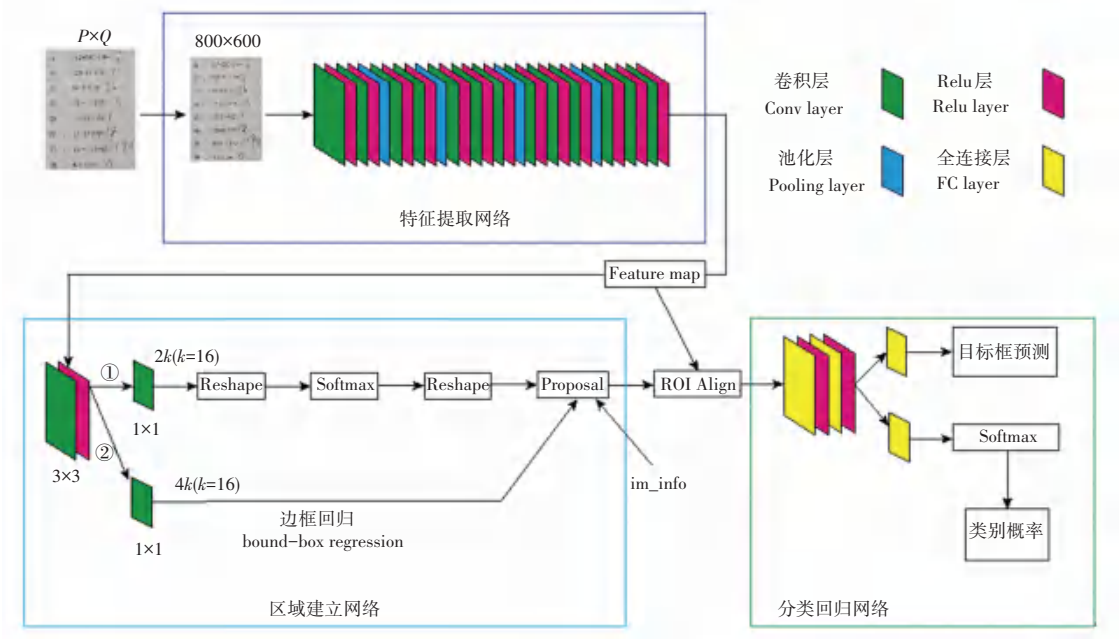


图 2　CF R-CNN 网络结构框架图

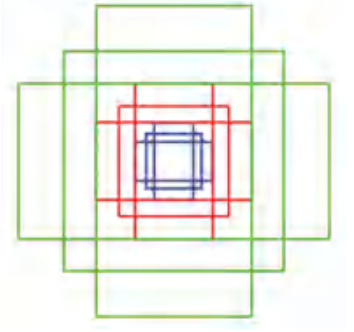
Fig. 2　Frame of CF R-CNN

2.1 特征提取网络

图片输入进来,先归一化到 800×600 像素,接下来是选取 VGG16 的前面部分层的基本结构作为特征提取网络,具体见图 2。由图 2 看到,主要由 13 个卷积层(Conv Layer)、13 个激励层、4 个池化层(Pooling Layer)组成,卷积层全部采用 3×3 大小的卷积核,步长为 1(*stride* = 1),填充 1 圈 0(*pad* = 1);所有激励层采用修正线性单元函数(the rectified linear unit, *Relu*);所有池化层采用最大池化(Max Pooling),2×2 大小的池化核,步长为 2(*stride* = 2),不填充(*pad* = 0)。经过一次池化层尺寸会变为原来的一半,因此最终得到的特征图(feature map)尺寸变为原图的 1/16。

2.2 改进的区域建议网络

为了减少参数数量,节省训练时间,改进的区域建议网络(Region Proposal Network,RPN)与分类回归网络共享特征提取网络,在特征图上做 3×3 的滑动窗口,把每个滑动窗口的中心映射到原图,生成 k 种不同的锚框来得到期望的目标建议框,原始 Faster R-CNN 采用 $k=9$ 种锚框,如图 3(a)所示,这 9 个锚框由 3 种长宽比[1∶2,1∶1,2∶1],3 种尺寸[128^2,256^2,512^2]组成,锚框长宽比和尺寸均不能适用于本文的算式检测与定位。为了找到适合算式检测定位的锚框,本文采用 K-means 聚类算法,在第 3 节实验与分析得出 $k = n_w \times n_r = 16$、共 16 种锚框($n_w = 4$ 种宽度值[64,80,96,112]和 $n_r = 4$ 种高宽比例[0.20,0.25,0.30,0.36]),如图 3(b)所示。



(a) Faster R-CNN 算法采用的锚框

(b) K-means 算法采用的锚框

图 3 锚框

Fig. 3 Anchors

生成的锚框(anchor boxes)中有一些是超出图片边界的,把这部分锚框去除,利用非极大值抑制(Non-Maximum Suppression,NMS)[21]去除重叠的框。对剩下的锚框分 2 条路线处理。线路一负责判断是否为目标对象,线路二负责计算目标对象的锚框(anchor boxes)与真实框(ground truth)在原图中的偏移量,最终得到区域建议框(region proposal)。

2.3 ROI Align

原始 Faster R-CNN 在 RPN 后使用 ROI Pooling,用来接收 RPN 输出的区域建议框(proposal)和原始特征图(feature map)。该过程可详见图 4 中的虚线流程,假如原图(image)为 800×800 大小的正方形,目标建议框为 665×665 大小的正方形。第一步,特征提取下采样 16 倍,得到边长为[800/16]=50 的正方形特征图(其中,[·]表示向下取整),同时,建议框(proposals)映射到特征图(feature map),边长大小变成[665/16]=41。第二步,把尺寸池化到统一大小 7×7,进入分类回归网络的全连接层,每块的边长为 41/7=5.86,是浮点数,做取整操作,取[41/7]=5。

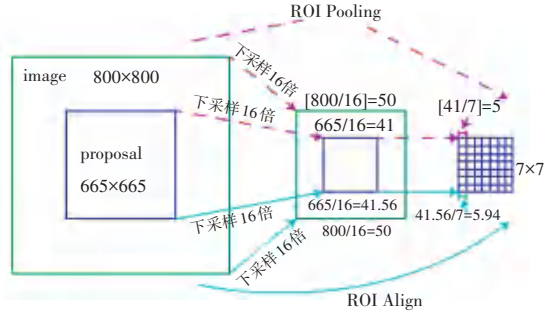


图 4 ROI Pooling 和 ROI Align

Fig. 4 ROI Pooling and ROI Align

以上 2 步都存在取整操作,此后映射回原图时会带来坐标的偏差,使得最终的检测框出现偏差,导致定位不准确,为了解决定位偏差的问题,研究采用 ROI Align 替代 ROI Pooling。ROI Align 不采用取整操作,而是直接利用浮点数来进行操作。该过可详见图 4 中的实线流程。在第一步中,特征图(feature map)中的建议框(proposals)边长为 665/16=41.56;在第二步中,池化到 7×7 大小时,每块的边长为 41.56/7=5.94。ROI Align 有效避免了取整带来的量化误差,使得映射回原图的定位框更精确。

2.4 分类回归网络

ROI Align 得到的 7×7 建议框特征图(proposal feature maps)进入分类回归网络的全连接层,后分 2 路同时进行。一路进行框坐标值的回归,得到更精确的目标检测框;另一路利用 *softmax* 函数进行计算,判断出目标的种类。

3　实验与分析

本次研究采用图 2 中的基本网络框架，通过聚类方法得出 n_w 种宽度值和 n_r 种高宽比，把单个锚点的锚框数目记为 k（$k = n_w \times n_r$），不同的 n_w，n_r 值决定了不同的 k 值。通过实验得到网络在不同 k 值下的性能，取综合性能最佳者作为 CF R - CNN 网络，并将其与原始的 Faster R - CNN 进行对比验证。实验中，把数据集分为训练验证集和测试集，分别占比 80% 和 20%，训练 50 轮：前 40 轮学习率为 0.001，后 10 轮学习率衰减为 0.000 1。

3.1　实验环境

实验均在 Ubuntu18.04 LTS 操作系统下进行，电脑的中央处理器为 Intel(R) Core(TM) i7-10870H，运行内存 16 GB，一块 NVIDIA GeForce RTX 2060 显卡，使用 CUDA 10.0 并行计算架构，cuDNN7.4 深度神经网络 GPU 加速库以及 TensorFlow1.13 深度学习框架，所用的编程语言为 Python。

3.2　k 值的确定

考虑到特征提取网络的 16 倍下采样，对 25 000 个算式的宽度值除以 16，作为新的宽度值，对其进行聚类分析，得出宽度值分布在 4～7 之间（一共 4 个整数值：4，5，6，7）。考虑到取 1～2 种宽度值，数目过少，因此，仅选取 $n_w = 3$ 和 4 两个值。取 $n_w = 3$ 时，宽度值为 4.5、5.7、6.7，选取 5、6、7 这 3 个值；取 $n_w = 4$ 时，宽度值为 4.3、5.3、6.0、7.0，选取 4、5、6、7 这 4 个值。接着对高宽比进行聚类分析，高宽比的值分布在 0.2 ～ 0.4 之间。分别按 3 ～ 6 类进行聚类，得出数据见表 1。选取表 1 中的几种 $n_w \times n_r$ 的组合做对比测试，得出训练时间及 AP 值。综合考虑选择 $k = n_w \times n_r = 4 \times 4$ 的组合作为 CF R-CNN 网络参数。

表 1　不同 k 值组合的参数及性能对比

Tab. 1　Performance comparison of different k-value combinations

$n_w \times n_r$	宽度	高度比	*time*/（h：m）	*AP*
3×3	5,6,7	0.22,0.27,0.34	2：39	0.999 9
4×4	4,5,6,7	0.20,0.25,0.30,0.36	2：34	1
4×5	4,5,6,7	0.20,0.23,0.27,0.31,0.37	2：35	0.999 8
4×6	4,5,6,7	0.19,0.22,0.25,0.28,0.32,0.38	2：35	0.999 8

3.3　性能对比分析

实验得出 Faster R-CNN 的训练时间和 *AP* 值，并与 CF R-CNN 进行对比，对比结果见表 2。图 5 为二者的损失曲线。由图 5 可以看出，CF R-CNN 的训练时间更短，收敛更迅速，准确度更高。对 CF R-CNN 与 Faster R-CNN 进行对比测试，测试结果如图 6 所示。图 6 中，图 6(a) 是部分的实验结果，图 6(b) 是基于同样的图片，运用 Faster R-CNN 得到的实验结果。

表 2　不同组合的训练时间和 *AP* 值

Tab. 2　Training time and *AP* values for different combinations

网络	*time*/（h：m）	*AP*
Faster R-CNN	2：41	0.797 4
CF R-CNN	2：34	1.000 0

从图 6 对比可以看出，CF R-CNN 的检测效果更好，把所有的基础算式都定位出来，而 Faster R-CNN 每张图都存在没有检测出来的算式（算式 35 和算式 48）。主要原因是 Faster R-CNN 的锚框与基础算式匹配度太低，而 CF R-CNN 选用了更准确的锚框尺寸和比例，使得检测效果得到提升。

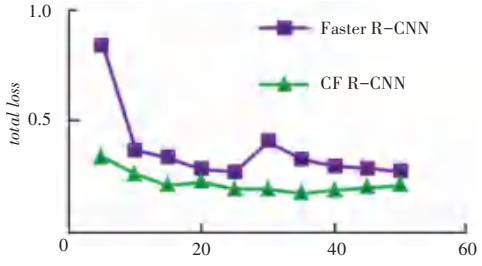
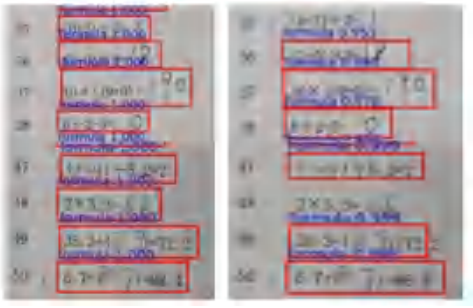


图 5　CF R-CNN 与 Faster R-CNN 损失曲线

Fig. 5　Loss curve of CF R-CNN and Faster R-CNN

在定位准确度方面，从图 6(b) 可以看出，算式 37 中手写数字 9 的上部落在了定位框外；图 6(b) 的算式 49，最后一个数字 2 全部都处于定位框之外。这些定位的不准确都会带来后续的识别偏差，从而造成最终批改系统的错误判别。分析主要原因应为 ROI Pooling 的 2 次量化造成偏差，使得定位框不够准确，而用 ROI Align 替换 ROI Pooling 后，消除了这部分偏差，这样一来定位就会更加准确。



(a) CF R-CNN 算法 (b) Faster R-CNN 算法

图 6 测试结果对比图

Fig. 6 Comparison chart of test results

4 结束语

为了提高基础算术题检测与定位的准确度,研究提出了基于改进 Faster R-CNN 的 CF R-CNN 方法。由于传统的 Faster R-CNN 中 RPN 的 3×3、共 9 种锚框,与基础算式匹配度不高,不再适用于基础算式检测定位。通过创建数据集,并对基础算式的尺寸和比例进行聚类分析,得出多种初始组合,进行实验对比分析,最终选择效果最佳的 4×4 的 16 种组合。用 ROI Align 替换 ROI Pooling,有效避免了 ROI Pooling 两次量化带来的偏差。在 500 张基础算式的图片数据集上进行训练测试,相比于传统的 Faster R-CNN,CF R-CNN 的收敛速度更快、损失更小、定位更准确,为后续的基础算式识别和自动批改提供保障。

参考文献

[1] 赵雪春,戚飞虎. 基于彩色分割的车牌自动识别技术[J]. 上海交通大学学报,1998,32(10):6-11.

[2] 周世杰,李顶根. 基于卷积神经网络的大场景下车牌识别[J]. 计算机工程与设计,2020,41(09):2592-2596.

[3] 陈黎,黄心汉,王敏,等. 基于聚类分析的车牌字符分割方法[J]. 计算机工程与应用,2002(06):221-222,256.

[4] 严国莉,黄山,李岱璋,等. 印刷体数字快速识别算法在身份证编号数字识别中的应用[J]. 计算机工程,2003,29(01):178-179.

[5] 殷瑞祥,李国华. 身份证号码的自动识别系统[J]. 华南理工大学学报(自然科学版),2002,30(02):94-96.

[6] 王润民,桑农,丁丁,等. 自然场景图像中的文本检测综述[J]. 自动化学报,2018,44(12):2113-2141.

[7] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015,521(7553): 436-444.

[8] SHI Baoguang, BAI Xiang, YAO Cong. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(11): 2298-2304.

[9] REN Shaoqing, HE Kaiming, 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.

[10] 魏相站,邵丽萍,周骅. 基于改进的 Faster RCNN 模型在车辆类型检测中的应用[J]. 智能计算机与应用,2020,10(07):97-100,103.

[11] 唐茂俊,黄海松,张松松,等. 改进的 Faster-R-CNN 在焊缝缺陷检测中的应用[J]. 组合机床与自动化加工技术,2021(12):83-86.

[12] 殷小芳,辛月兰,兰天,等. 改进 Faster R-CNN 的多通道检测算法[J]. 计算机工程与设计,2021,42(12):3453-3460.

[13] 张莹,刘子龙,万伟. 基于 Faster R-CNN 的无人机车辆目标检测[J]. 电子科技,2021,34(11):11-20.

[14] 冯小雨,梅卫,胡大帅. 基于改进 Faster R-CNN 的空中目标检测[J]. 光学学报,2018,38(06):250-258.

[15] 黄继鹏,史颖欢,高阳. 面向小目标的多尺度 Faster-R-CNN 检测算法[J]. 计算机研究与发展,2019,56(02):319-327.

[16] 王宪保,朱啸咏,姚明海. 基于改进 Faster R-CNN 的目标检测方法[J]. 高技术通讯,2021,31(05):489-499.

[17] 黄宁霞,张荣芬,刘宇红. 改进深度学习框架 Faster R-CNN 的人行道障碍物目标检测[J]. 机械设计与研究,2021,37(02):7-12.

[18] BODLA N, SINGH B, CHELLAPPA R, et al. Soft-NMS --Improving object detection with one line of code[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017:5562-5570.

[19] KRISHNA K, MURTY M N. Genetic k-means algorithm [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 1999,29(3):433 - 439.

[20] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]// IEEE International Conference on Computer Vision (ICCV). Venice, Italy:IEEE,2017:2980-2988.

[21] NEUBECK A,GOOL V L . Efficient non-maximum suppression [C]// International Conference on Pattern Recognition. Hong Kong, China:IEEE Computer Society, 2006:850-855.

(上接第 163 页)

[11] FAN Heng, LIN Liting, YANG Fan, et al. Lasot: A high-quality benchmark for large-scale single object tracking [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE 2019: 5374-5383.

[12] KRISTAN M, LEONARDIS A, MATAS J, et al. The sixth visual object tracking vot2018 challenge results[C]// Proceedings of the European Conference on Computer Vision (ECCV) Workshops. Cham:Springer,2018:3-53.

[13] MUELLER M, SMITH N, GHANEM B. A benchmark and simulator for uav tracking [C]// European Conference on Computer Vision. Cham: Springer, 2016: 445-461.