控制与决策
Control and Decision

改进YOLOv2的端到端自然场景中文字符检测

刘杰， 朱旋， 宋密密

引用本文:
刘杰, 朱旋, 宋密密. 改进YOLOv2的端到端自然场景中文字符检测[J]. 控制与决策, 2021, 36(10): 2483−2489.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2020.0270

改进YOLOv2的端到端自然场景中文字符检测

刘 杰†, 朱 旋, 宋密密

(哈尔滨理工大学 测控技术与通信工程学院，哈尔滨 150080)

摘 要: 针对自然场景中文字符检测率低、小字符检测困难以及字符检测类别多样等问题,提出一种基于YOLOv2的改进方法,并将其应用于自然场景中文字符检测中. 首先利用*k*-means++聚类算法对字符目标候选框(anchor)的数量和宽高比维度进行聚类分析,提出多层特征融合策略,对原网络中第4个最大池化层前所输出的特征图经过3×3和1×1大小的卷积核进行卷积操作,并执行4倍的下采样得到局部特征;然后对第5个最大池化层前所输出的特征图经过3×3和1×1大小的卷积核进行卷积操作,并执行2倍的下采样得到局部特征,将局部特征与全局特征融合,同时增加高层卷积中的重复卷积层,将高层卷积中连续且重复的3×3×1 024大小的卷积层数由3增加为5;最后使用Chinese text in the wild (CTW)数据集对YOLOv2和改进的YOLOv2算法进行对比实验,结果表明,改进后的YOLOv2算法在中文字符检测中平均准确率均值为78.3 %,较原YOLOv2算法提升了7.3 %,且明显高于其他自然场景中的文字符检测方法.

关键词: 计算机视觉；深度学习；自然场景；中文字符检测；YOLOv2

中图分类号: TP273 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2020.0270 **开放科学(资源服务)标识码(OSID):**

引用格式: 刘杰,朱旋,宋密密. 改进YOLOv2的端到端自然场景中文字符检测[J]. 控制与决策, 2021, 36(10): 2483-2489.

End-to-end Chinese character detection in natural scene based on improved YOLOv2

LIU Jie†, ZHU Xuan, SONG Mi-mi

(School of Measurement and Control Technologe and Communication Engineering，Harbin University of Science and Technology，Harbin 150080，China)

Abstract: This paper proposes an improved method based on YOLOv2 to solve the problems of low Chinese character detection rate, difficulty in small character detection and various character detection categories in natural scenes, and applies it to Chinese character detection in natural scenes. Firstly, *k*-means++ clustering algorithm is used to cluster the number and aspect ratio of character target candidate boxes (anchors). Then the multi-layer feature fusion strategy is proposed, the feature map output before the fourth maxpooling pooling layer in the original network is convolved with 3×3 and 1×1 convolution kernels and 4 times downsampling is performed to obtain local features, and the feature map output before the fifth maxpooling pooling layer in the original network is convolved with 3×3 and 1×1 convolution kernels and 2 times downsampling is performed to obtain local features. At the same time, repeat convolution layers in high-level convolution are added, and the number of continuous and repeated 3×3×1 024 convolution layers in high-level convolution is increased from 3 to 5. Finally, the Chinese text in the wild (CTW) data set is used to compare the YOLOv2 algorithm with the improved one. The experimental results show that the improved YOLOv2 algorithm has a mean average precision (mean average precision, mAP) of 78.3% in Chinese character detection, which is 7.3% higher than mAP value of the original YOLOv2 algorithm, and is significantly higher than the one of other Chinese character detection methods in natural scenes.

Keywords: computer vision；deep learning；natural scenes；Chinese character detection；YOLOv2

0 引 言

文字作为人类文明的标志、信息交流的载体,广泛地存在于自然场景图像中,相比图像中其他自然场景内容,场景中文字的逻辑性和概括性更强,可以准确地提供高层语义信息,有助于场景内容的分析与理解. 人工智能和深度学习技术的飞速发展为研究端

收稿日期: 2020-03-12；**修回日期:** 2020-05-21.
基金项目: 国家自然科学基金项目(51607049)；黑龙江省自然科学基金项目(LH2019E067).
责任编委: 陈家伟.
†通讯作者. E-mail: liujie@hrbust.edu.cn.

到端的自然场景中文字符检测提供了新的途径[1-5].

对于自然场景下的文本检测问题,国内外学者已开展了相关研究,并取得了一定的成果[6-9]. 基于传统方法的自然场景文本检测其核心思想是对人工提取的特征进行检测,如颜色阶调、区域等. 文献[10]提出了基于最大稳定极值区域的检测算法,但该方法无法检测到位于黑色背景中的白色文字区域. 文献[11]提出了基于Adaboost的检测算法,但该方法在处理低对比度图像时鲁棒性较差. 文献[12]利用FAST角点提取,但该方法受光照变化和拍摄视角影响较大. 以上传统的自然场景文本检测方法需要人工设计特征,因而容易提取大量较差特征,导致召回率和检测准确率损失.

随着深度学习技术的不断发展,卷积神经网络(convolutional neural network, CNN)在图像分类、目标检测等领域取得了一系列成就. 文献[13]提出了对文字进行分层的检测策略,该方法首先采用CNN提取特征,然后采用随机森林算法对文字候选区域进行精细分类,但该方法后续处理通常比较复杂,无法达到实时检测的效果. 为了进一步提高目标检测的速度,文献[14]提出了一种端到端的目标检测算法YOLO(you only look once),将目标检测问题转化为回归问题,进而将目标和背景进行更好地区分,但输入图像的尺寸大小是固定的,导致在训练过程中无法适应不同物体的形状. 文献[15]通过加入方向信息使得SSD(single shot detector)检测器可以应对任意方向排列的文本检测问题,但其对于间隔较大的文本检测效果不理想. 文献[16]提出的YOLOv2算法使用darknet-19网络作为特征提取网络,大大简化了网络结构,同时提高了目标检测的准确率,但是其对自然场景下的字符检测效果较差,容易出现误将背景识别为字符的情况.

针对以上问题,为了提高网络对自然场景中文字符的检测准确率,本文以基于端到端的单阶段目标检测算法YOLOv2为基础并进行改进,从而实现对自然场景下中文字符的检测.

1 YOLOv2算法

YOLOv2算法首先将图像划分为$S\times S$个网格单元,并在每个网格单元中预测B个边界框,其中每个边界框预测5个参数,分别包括t_x、t_y、t_w、t_h以及置信度confidence .

预测边界框的位置通过框的中心坐标(b_x, b_y)和框的宽(b_w)、高(b_h)4个变量表示. 它们分别通过实际预测值t_x、t_y、t_w、t_h进行计算.

如图1所示,p_w、p_h为anchor的宽和高,b_w、b_h为预测边界框的宽和高. 其中:$\sigma(t_x)$、$\sigma(t_y)$为预测边界框中心相对于当前所在网格单元左上角的偏移,(c_x, c_y)为当前网格单元相对于图像左上角的距离. 预测边界框的真实位置的计算公式为

$$b_x = \sigma(t_x) + c_x, \tag{1}$$

$$b_y = \sigma(t_y) + c_y, \tag{2}$$

$$b_w = p_w \mathrm{e}^{t_w}, \tag{3}$$

$$b_h = p_h \mathrm{e}^{t_h}. \tag{4}$$

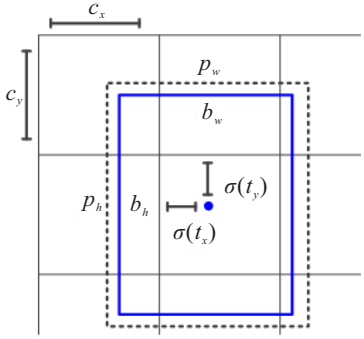

图1 边界框位置预测

confidence的值代表检测目标的准确性. 该参数的大小由每个网格单元中包含检测目标的概率和输出边界框的准确率共同决定,其中输出边界框的准确率定义为预测边界框与真实边界框的交并比(intersection over union, IOU),其计算公式为

$$\text{confidence} = \Pr(\text{object}) \times \text{IOU}_{\text{pred}}^{\text{truth}}. \tag{5}$$

其中:$\Pr(\text{object})$为网格单元中含有某类目标的概率,$\text{IOU}_{\text{pred}}^{\text{truth}}$为预测边界框与真实边界框的交集与并集之比. 剔除低于置信度阈值的大多数预测边界框,对高于置信度阈值的预测边界框采用非最大抑制操作得到最终的边界框.

YOLOv2采用了Darknet-19网络结构,共包含19个卷积层和5个最大池化层.

在经过3×3大小的卷积和2×2大小的最大池化层之后,特征图维度降低两倍,同时特征图的通道数增加两倍,并且在3×3大小的卷积核之间使用1×1大小的卷积核压缩特征图的通道数以降低模型计算量和参数,网络使用了全局平均池化(global average pooling)作为预测.

2 YOLOv2算法改进

本文主要研究自然场景中文字符检测问题,YOLOv2初始网络定义的anchor参数与网络的层级结构不适用于本文的研究对象. 因此,首先采用k-means++聚类算法对数据集中的字符目标预选框进

行重新聚类分析,然后针对字符目标检测的特殊性修改网络的层级结构.

2.1　k-means++算法的聚类分析

YOLOv2算法中引入了anchor的思想,anchor是一组宽高固定的初始候选框,对初始anchor参数的选择会直接影响网络对字符目标的检测精度. 为使得所选择的anchor更适合于自然场景下中文字符目标的检测,采用k-means++聚类算法代替k-means聚类算法对CTW数据集中的中文字符边界框的大小进行重新聚类分析. 上述两种算法均为典型的聚类算法,但由于k-means算法在初始簇心的选取上采取随机抽取k个样本的方式,这种随机不确定性会对最终聚类结果带来较大误差. 而k-means++算法按照概率抽取k个簇心,对于数据集中的每一个点x,计算它与最近聚类中心的距离$D(x)$,$D(x)$值越大,被选取作为聚类中心的概率越大,通过这样的方式使其可以选取较优的聚类中心. 因此,采用k-means++算法取代k-means算法,解决了k-means算法对初始簇心比较敏感的问题,通过k-means++算法的聚类结果选取适合中文字符检测的候选框的数量和大小.

在实现聚类算法时,采用中文字符标签真实样本框与先验框的平均交并比(avg IOU)代替传统的欧几里得距离作为目标函数,使得误差与先验框的尺寸无关. 目标函数d计算为

$$d(\text{box},\text{centriod}) = 1 - \text{IOU}(\text{box},\text{centriod}). \quad (6)$$

其中: box为样本标签的聚类框,centriod为聚类中心框.

选取$k = 1 \sim 10$,分别使用k-means++聚类算法和k-means聚类算法对数据集中的样本进行聚类分析,得到k与avg IOU之间的关系如图2所示. 可以看出,随着k值的增大,两种聚类方法的目标函数变化均逐渐增大. 但是,在目标函数上升的过程中明显可以看出,k-means++聚类算法的曲线更加平稳,在一定程度上降低了聚类分析带来的误差. 图2中曲线变化的拐点为最佳的anchor数量. 当k值大于6时,曲线开始变得平稳,所以anchor数量最终设置为6,代表使用6个大小不同的anchor进行字符目标的定位,这样既可以加快损失函数的收敛,降低损失,又可以消除候选框过多或过少带来的误差.

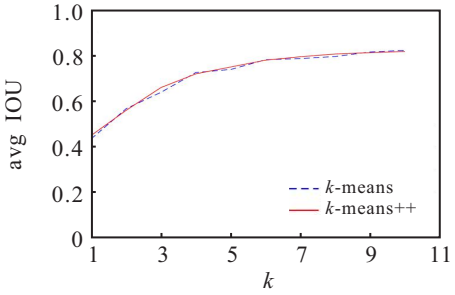

图 2　avg IOU曲线

不同k值对应的先验框宽高如表1所示. 由表1结果可见,当k值超过6时,会出现大小较为相近的聚类结果,造成冗余,因此取$k = 6$时的聚类结果作为改进后的参数. 该参数与YOLOv2算法原始5个参数(0.74, 0.87)、(2.42, 2.66)、(4.31, 7.04)、(10.25, 4.59)、(12.69, 11.87)相比,聚类结果更加集中,更适合于自然场景下中文字符的检测,同时降低了网络误将背景识别为目标的概率. 因此,使用k-means++聚类算法产生的聚类参数代替原始参数进行训练和测试.

表 1　不同k值对应的先验框宽高

$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$
(1.47, 1.95)	(1.35, 1.79)	(1.28, 1.70)	(1.23, 1.65)	**(1.16, 1.59)**	(1.13, 1.56)	(1.10, 1.51)
(3.07, 3.87)	(2.46, 3.17)	(2.19, 2.85)	(2.04, 2.95)	**(1.22, 1.95)**	(1.73, 3.87)	(1.52, 3.33)
(7.04, 8.11)	(4.48, 5.42)	(3.60, 4.47)	(4.14, 5.99)	(1.93, 2.27)	**(1.84, 2.18)**	(1.80, 2.06)
–	(9.15, 10.28)	(5.97, 6.98)	(6.82, 10.75)	(3.06, 3.45)	**(1.89, 3.16)**	**(2.66, 3.02)**
–	–	(11.25, 12.39)	(9.52, 6.53)	(4.33, 5.44)	(3.79, 4.92)	**(2.71, 3.38)**
–	–	–	(13.21, 14.45)	(7.01, 7.95)	(5.76, 6.50)	(4.05, 4.30)
–	–	–	–	(12.53, 13.76)	(8.50, 9.74)	(5.48, 6.63)
–	–	–	–	–	(14.64, 15.78)	(8.53, 9.46)
–	–	–	–	–	–	(14.44, 15.69)

2.2　改进的YOLOv2网络结构

2.2.1　多层特征融合

对于自然场景中文字符检测,每个字符之间的差异涉及字形、大小、颜色、遮挡、复杂背景等,因此为了充分利用卷积过程中产生的特征图信息,采用多层特征融合策略. 改进YOLOv2算法网络结构如图3所示. 首先在卷积过程(a)部分,使84×84×128维的输出量经过3×3和1×1大小的卷积核得到84×84×32维的特征图,并在此基础上执行4倍的下采样,得到特征1;在卷积过程(b)部分,使42×42×256维的输出量经过3×3和1×1大小的卷积核得到42×42×64维的特征图,并在此基础上执行2倍的下采样,得到特征2;最后将特征1、特征2与全局特征(特征3)进行融合, 增强网络对局部特征的提取, 进而使模型可以

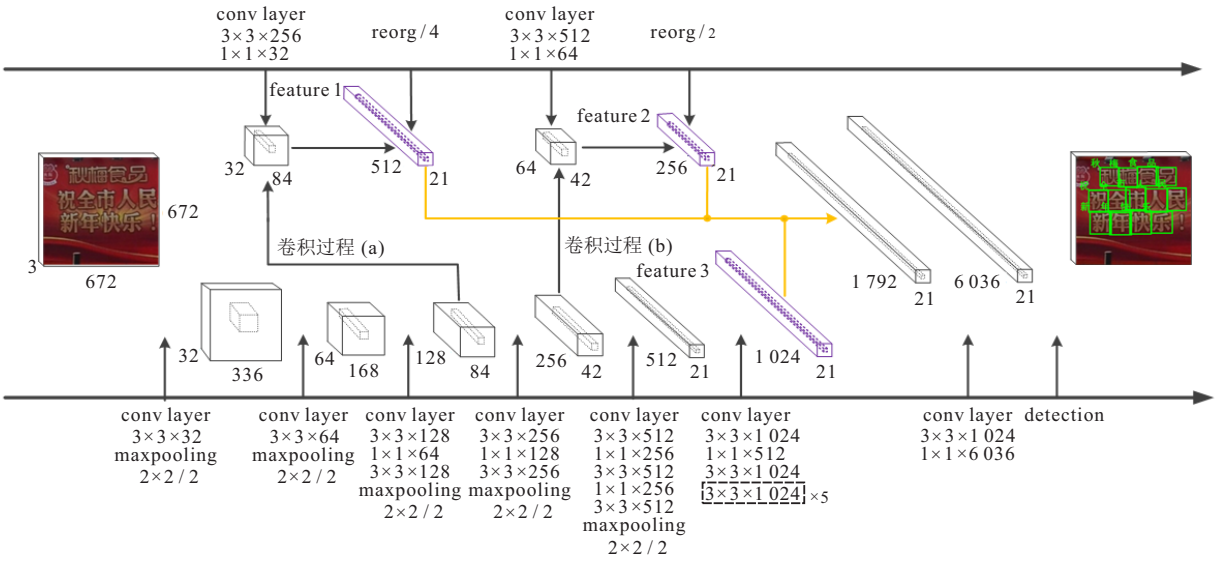

图 3 改进 YOLOv2 算法网络结构

更好地区分字符目标之间的差异以及字符目标和复杂背景之间的差异,并提高模型对小字符目标的检测能力.

2.2.2 增加高层中重复的卷积层

通常而言,YOLOv2 网络是针对通用对象检测而设计的模型,此类网络检测到的类别数量可能较少. 对于 YOLOv2 网络,在最后一个池化层后的卷积层中存在 3 个连续且重复的 3×3×1 024 大小的卷积层. 一般来说,高层卷积中重复卷积的操作可以处理类型相差较大的类别. 对于自然场景下中文字符的检测类别有 1 001 类,并且每个中文字符类型之间的差异较大,这意味着高层中的重复卷积可以提高系统的识别性能. 因此,本文改进的算法在高层中增加了重复的卷积层,将连续且重复的 3×3×1 024 大小的卷积层数增加为 5,以提高检测类别的数量.

3 实验分析

3.1 数 据 集

实验采用 CTW 数据集[17],分别对 SSD 算法、YOLOv2 算法、改进网络结构的 YOLOv2 算法以及本文提出的增加 k-means++ 聚类算法优化 anchor 参数并改进网络结构的 YOLOv2 算法进行对比实验. CTW 数据集中包含 33 285 张街景图像,其中用作检测部分的图像包含 29 016 张. 对于每张图像,所有的中文字符目标都被标注,每个字符目标都有与其对应的检测框及字符类别标注. 将数据集以 8:1 的比例划分为训练集和测试集,其中训练集包含 25 887 张图像,测试集包含 3 129 张图像. 数据集中共有 1 001 类对象,包括前 1 000 类频繁观察到的字符类别以及“其他”类别. 在训练集和测试集中包含不同大小的字符目标,大小由其边界框的长边度量,小字符为边界框长边小于 16 像素的字符,中字符为边界框长边大于 16 像素且小于 32 像素的字符,大字符为边界框长边大于 32 像素的字符.

3.2 网络训练

实验采用 CTW 数据集进行训练. 在训练过程中,将动量 (momentum) 设置为 0.9,衰减系数 (decay) 设置为 0.000 5,初始学习率 (learning rate) 设置为 0.000 1,学习率调整策略为 steps,最大迭代次数为 45 000 次,学习率在训练迭代次数为 100、25 000、35 000 次时,分别再乘以 10、0.1、0.1,使损失函数进一步收敛.

图 4 为改进的 YOLOv2 网络在训练过程时平均损失随迭代次数变化的收敛曲线. 由图 4 可见,训练初期的损失函数值约为 45,损失下降较快,随着迭代次数增加,损失下降越来越慢,迭代 25 000 次后,损失基本趋于平稳. 图 5 为预测边界框与真实边界框的 avg IOU 随迭代次数变化的曲线. 由图 5 可见,在训练初期 avg IOU 较低,约为 0.2,随着迭代次数的增加,avg IOU 逐渐增大,迭代到 25 000 次以后,avg IOU 逐渐接近 1,最终稳定在 0.8 左右,由此参数的收敛情况分析可知,改进的 YOLOv2 网络的训练结果比较理想.

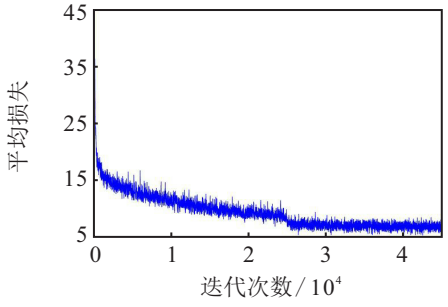

图 4 平均损失变化曲线

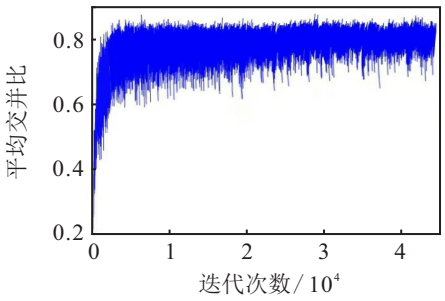

图 5　平均交并比变化曲线

3.3　实验结果分析

首先选出一些样本分别输入到 SSD、YOLOv2、improved YOLOv2 以及增加 k-means++ 聚类算法优化 anchor 参数并改进网络结构的 YOLOv2 (improved YOLOv2-k) 检测模型中进行检测, 将检测结果进行对比. 图 6(a) 是选出的 4 张测试样本, 从左至右依次为样本 1、样本 2、样本 3、样本 4. 图 6(b) ～图 6(e) 分别为测试样本在 SSD、YOLOv2、improved YOLOv2 以及 improved YOLOv2-k 模型下对应的检测结果, 其中绿框代表检测出并正确识别的中文字符, 黄框代表检测或识别错误的中文字符. 样本 1 中, SSD、YOLOv2、improved VOLOv2 均将英文字符误认为是中文字符, 而 improved YOLOv2-k 的检测结果较理想, 表明本文在对候选框的选择上更加适合检测中文字符. 样本 2

(a)　原始样本

(b)　SSD 检测结果

(c)　YOLOv2 检测结果

(d)　improved YOLOv2 检测结果

(e)　improved YOLOv2-k 检测结果

图 6　部分样本检测结果

中,由于自然场景中的光线较暗,导致SSD、YOLOv2、improved VOLOv2均出现了漏检的情况,而improved YOLOv2-k的检测结果较理想,表明本文的改进提高了复杂背景下字符检测的类别数. 样本3中,SSD、YOLOv2、improved VOLOv2对小字符目标的检测效果不理想,而improved YOLOv2-k的检测效果较理想,表明本文的改进提高了小目标的识别能力. 样本4中,由于背景与字符目标颜色相近,字符目标识别难度增大,导致YOLOv2和improved YOLOv2出现了将背景误检为目标的情况,而improved YOLOv2-k的检测效果比较理想,表明本文增加k-means++聚类算法优化anchor参数并改进网络结构后的YOLOv2检测模型,在一定程度上提高了网络对复杂场景的适应能力.

采用CTW数据集,对SSD、YOLOv2、improved YOLOv2以及improved YOLOv2-k模型进行测试,计算其对字符目标的召回率(recall, R)和检测的准确率(precision, P). 目标召回率和检测的准确率可分别表示为

$$R=\frac{X_{\mathrm{TP}}}{X_{\mathrm{TP}}+X_{\mathrm{FN}}},\tag{7}$$

$$P=\frac{X_{\mathrm{TP}}}{X_{\mathrm{TP}}+X_{\mathrm{FP}}}.\tag{8}$$

其中: X_{TP}为正确检测出来的目标数, X_{FN}为没有被检测出来的目标数, X_{FP}为被错误检出的目标数.

图7为测试集上各模型的P-R(precision-recall)曲线. 可以看出,当4个模型的准确率为66.9%、69.3%、74.1%、76.9%时,其对应的召回率分别为66.8%、69.0%、72.6%、74.8%. 可见,本文improved YOLOv2-k模型的检测精度明显高于其他模型.

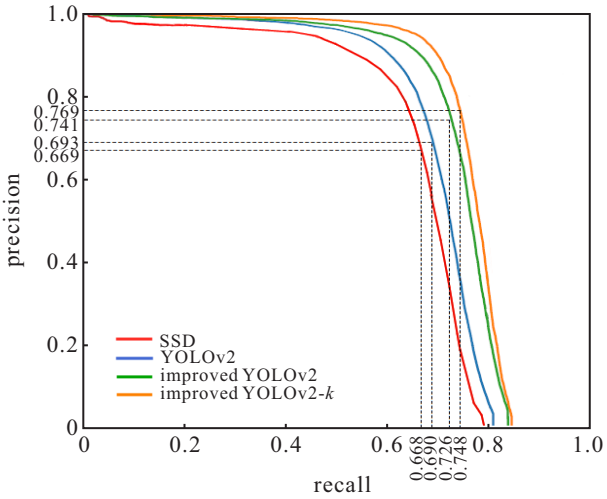

图7 P-R曲线

利用CTW中的测试集分别对SSD、YOLOv2和improved YOLOv2以及improved YOLOv2-k模型进行测试,小、中、大及其所有字符的平均准确率和平均准确率均值如表2所示,改进网络结构的improved YOLOv2模型相比未改进的YOLOv2模型,平均准确率(average precision, AP)提高了4.8%,mAP提高了4.2%;使用增加k-means++聚类算法优化anchor参数并改进网络结构的improved YOLOv2-k模型相比只改进了网络结构的improved YOLOv2模型,AP提高了2.8%,mAP提高了3.1%,整体上提高了自然场景下中文字符的检测精度.

表2 测试结果

检测模型	测试样本数量	平均检测时间/s	平均准确率(AP)/%				平均准确率均值(mAP)/%
			小字符	中字符	大字符	所有字符	
SSD	3 129	0.063	54.6	71.3	74.8	66.9	69.3
YOLOv2		0.069	56.8	74.7	76.5	69.3	71.0
improved YOLOv2		0.072	61.8	79.5	81.0	74.1	75.2
improved YOLOv2-k		0.073	64.5	82.0	84.1	76.9	78.3

4 结 论

本文提出了一种基于YOLOv2算法的改进方法,首先使用k-means++聚类算法代替k-means聚类算法对中文字符检测的anchor参数进行优化,解决了k-means算法对初始簇心较为敏感的问题,并选择6个大小不同的anchor得到适合于中文字符检测的聚类中心. 同时提出多层特征融合策略并增加高层中的重复卷积层,解决了网络对小字符漏检概率高以及检测类别多的问题. 将深度学习方法与自然场景文字检测问题相结合,实现了对自然场景中文字符的端到端检测. 实验结果表明,增加k-means++聚类算法优化anchor参数并改进网络结构的YOLOv2检测模型,在对自然场景中文字符检测中,平均准确率均值为78.3%,能够保证较高的检测准确率. 但是,改进后的YOLOv2算法仍然存在漏检的情况,下一步将会收集更多的中文自然场景图像,以进一步研究如何提高自然场景中文字符检测准确率和速度.

参考文献(References)

[1] 李健伟, 曲长文, 彭书娟. 基于级联CNN的SAR图像舰船目标检测算法[J]. 控制与决策, 2019, 34(10):

2191-2197.
(Li J W, Qu C W, Peng S J. A ship detection method based on cascade CNN in SAR images [J]. Control and Decision, 2019, 34(10): 2191-2197.)

[2] 罗俊海, 杨阳. 基于数据融合的目标检测方法综述[J]. 控制与决策, 2020, 35(1): 1-15.
(Luo J H, Yang Y. An overview of target detection methods based on data fusion[J]. Control and Decision, 2020, 35(1): 1-15.)

[3] 郭戈, 王兴凯, 徐慧朴. 基于声呐图像的水下目标检测、识别与跟踪研究综述[J]. 控制与决策, 2018, 33(5): 906-922.
(Guo G, Wang X K, Xu H P. Review on underwater target detection, recognition and tracking based on sonar image[J]. Control and Decision, 2018, 33(5): 906-922.)

[4] 张慧, 王坤峰, 王飞跃. 深度学习在目标视觉检测中的应用进展与展望[J]. 自动化学报, 2017, 43(8): 1289-1305.
(Zhang H, Wang K F, Wang F Y. Advances and perspectives on applications of deep learning in visual object detection[J]. Acta Automatica Sinica, 2017, 43(8): 1289-1305.)

[5] 鞠默然, 罗海波, 王仲博. 改进的YOLOv3算法及其小目标检测中的应用[J]. 光学学报, 2019, 39(7): 0715004.
(Ju M R, Luo H B, Wang Z B. Improved YOLOv3 algorithm and its application in small target detection[J]. Acta Optica Sinica, 2019, 39(7): 0715004.)

[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.
(Wang R M, Sang N, Ding D, et al. Text detection in natural scence image: A survey[J]. Acta Automatica Sinica, 2018, 44(12): 2113-2141.)

[7] Dai Y C, Huang Z, Gao Y T, et al. Fused text segmentation networks for multi-oriented scene text detection[J]. 2017, arXiv: 1709.03272.

[8] Jiang Y Y, Zhu X Y, Wang X B, et al. R2CNN: Rotational region CNN for orientation robust scene text detection[J]. 2017, arXiv: 1706.09579.

[9] Yin F, Wu Y C, Zhang X Y, et al. Scene text recognition with sliding convolutional character models[J]. 2017, arXiv: 1709.01727.

[10] Neumann L, Matas J. A method for text localization and recognition in real-world images[C]. Asian Conference on Computer Vision, Queenstown: Springer, 2010: 770-783.

[11] Lee J J, Lee P H, Lee S W, et al. AdaBoost for text detection in natural scene[C]. IEEE International Conference on Document Analysis and Recognition. Beijing: IEEE, 2011: 429-434.

[12] Buta M, Neumann L, Matas J, et al. FASText: Efficient unconstrained scene text detector [C]. IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1206-1214.

[13] Xu H L, Su F. A robust hierarchical detection method for scene text based on convolutional neural networks[C]. IEEE International Conference on Multimedia and Expo. Turin: IEEE, 2015: 1-67.

[14] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779-788.

[15] Shi B G, Bai X, Belongie S. Detecting oriented text in natural images by linking segments[C]. IEEE Conference on Computer Vision and Pattern Recognition. Hawaii: IEEE, 2017: 3482-3490.

[16] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger[C]. Computer Vision and Pattern Recognition. Hawaii: IEEE, 2017: 6517-6525.

[17] Yuan T L, Zhu Z, Xu K, et al. Chinese text in the wild[J]. 2018, arXiv: 1803.00085.

作者简介

刘杰 (1980−), 女, 副教授, 博士, 从事光纤光栅传感及解调技术、数字图像处理、计算机视觉等研究, E-mail: liujie@hrbust.edu.cn;

朱旋 (1996−), 女, 硕士生, 从事人工智能、深度学习等研究, E-mail: 953583040@qq.com;

宋密密 (1996−), 女, 硕士生, 从事人工智能、深度学习的研究, E-mail: 1064130028@qq.com.

(责任编辑: 郑晓蕾)