Journal of Computer Applications
计算机应用, 2020, 40(3): 854 - 858
ISSN 1001-9081
CODEN JYIIDU
2020-03-10
http://www.joca.cn

文章编号:1001-9081(2020)03-0854-05
DOI:10.11772/j.issn.1001-9081.2019071262

基于RetinaNet改进的车辆信息检测

刘　革,郑叶龙*,赵美蓉
(天津大学 精密测试技术及仪器国家重点实验室,天津 300072)
(*通信作者电子邮箱zhengyelongby@tju.edu.cn)

摘　要:移动端计算力不足和存储有限导致车辆信息检测模型精度不高、速度较慢。针对这一问题,提出一种基于RetinaNet改进的车辆信息检测算法。首先,开发新的车辆信息检测框架,将特征金字塔网络(FPN)模块的深层特征信息融合进浅层特征层,以MobileNet V3为基础特征提取网络;其次,引入目标检测任务的直接评价指标GIoU指导定位任务;最后,使用维度聚类算法找出Anchor的较好尺寸并匹配到相对应的特征层。与原始RetinaNet目标检测算法的对比实验表明,所提算法在车辆信息检测数据集上的精度有10.2个百分点的提升。以MobileNet V3为基础网络时平均准确率均值(mAP)可达97.2%且在ARM v7设备上单帧前向推断用时可达100 ms。实验结果表明,所提方法能够有效提高移动端车辆信息检测算法性能。

关键词:卷积神经网络;目标检测;维度聚类;特征融合;GIoU
中图分类号:TP391.4　　**文献标志码**:A

Vehicle information detection based on improved RetinaNet

LIU Ge, ZHENG Yelong*, ZHAO Meirong
(*State Key Laboratory of Precision Testing Technology and Instruments*, *Tianjin University*, *Tianjin* 300072, *China*)

Abstract: The lack of computational power and limited storage of the mobile terminals lead to the low accuracy and slow speed of vehicle information detection models. Therefore, an improved vehicle information detection algorithm based on RetinaNet was proposed to solve this problem. Firstly, a new vehicle information detection framework was developed, and the deep feature information of the FPN (Feature Pyramid Network) module was merged into the shallow feature layer, and MobileNet V3 was used as the basic feature extraction network. Secondly, the direct evaluation index of target detection task——GIoU (Generalized Intersection over Union) was introduced to guide the positioning task. Finally, the dimension clustering algorithm was used to find the better size of Anchors and match them to the corresponding feature layers. Compared with the original RetinaNet target detection algorithm, the proposed algorithm has the accuracy improved by 10.2 percentage points on the vehicle information detection dataset. When using MobileNet V3 as the basic network, the mAP (mean Average Precision) can reach 97.2% and the forward inference time of single frame can reach 100 ms on ARM v7 devices. The experimental results show that the proposed method can effectively improve the performance of mobile vehicle information detection algorithms.

Key words: Convolutional Neural Network (CNN); target detection; dimension clustering; feature fusion; Generalized Intersection over Union (GIoU)

0　引言

随着智能信息时代的到来,作为车辆信息智能审核及其他智能交通系统的关键部分,车辆外框检测[1-2]、车牌检测[3-4]有着广泛的应用。

深度学习模型具有强大的特征表达能力,结合较大数据量的保证和计算力的进步,它正逐步成为机器视觉领域的热点研究方向。目标检测是指从一张未知的图像中分离出背景和感兴趣的目标并确定这一目标的类别和位置。当前基于卷积神经网络的目标检测方法主要分两类:一类是基于Anchor-free思想的目标检测框架;另一类是基于Anchor思想的目标检测框架。基于Anchor-free思想的目标检测框架通常使用关键点检测的方法和密集预测的方法来完成检测任务,如CornerNet[5]、ExtremeNet[6]都是通过预测的关键点及关键点之间的几何信息来确定目标的位置;如CenterNet[7]通过密集预测方法直接得到检测目标的位置。这种基于Anchor-free思想模型结构设计较为简单,在移动端性能一般不如基于Anchor思想的目标检测框架。基于Anchor思想的目标检测框架分为两小类:一类是基于区域候选框的二阶段目标检测框架,这种检测框架先在图像上生成若干可能包含目标的候选区域,然后通过卷积神经网络(Convolutional Neural Network, CNN)分别对这些候选区域提取特征,最后通过卷积神经网络进行目标位置的回归与类别的识别,如Fast RCNN(Region Convolutional Neural Network)[8]、Faster RCNN[9]、R-FCN(Region-based Fully Convolutional Network)[10]等;另一类为一

收稿日期:2019-07-19;**修回日期**:2019-09-27;**录用日期**:2019-09-27。
基金项目:国家自然科学基金资助项目(51805367);天津市自然科学基金资助项目(18JCQNJC04800, 18JCZDJC31800)。
作者简介:刘革(1994—),男,湖北襄阳人,硕士研究生,主要研究方向:计算机视觉;　郑叶龙(1987—),男,浙江温州人,讲师,博士,主要研究方向:机器视觉、力学测量;　赵美蓉(1967—),女,天津人,教授,博士,主要研究方向:机器视觉、力学测量。

www.joca.cn

阶段目标检测框架,这种目标检测框架直接从图片获得预测结果,将整个目标检测任务整合成一个端到端的任务,而且只处理一次图像即可得到目标的类别与位置信息,如YOLO(You Only Look Once)[11-12]、SSD(Single Shot multibox Detector)[13]、DSSD(Deconvolutional Single Shot Detector)[14]、FSSD(Feature fusion Single Shot multibox Detector)[15]等。二阶段检测算法相比较于一阶段检测算法具有更高的识别准确率和定位精度,而由于二阶段检测算法生成的中间候选区域存在很多不必要的重叠导致检测时间增加,所以检测速度不如一阶段检测算法。一阶段检测算法一般采取密集采样候选区域的机制,这种机制通常会产生大量的负样本使正负样本不平衡,导致训练过程受到影响并降低最终的检测准确率。文献[16]提出的RetinaNet通过Focal Loss成功地解决了这一问题,使一阶段目标检测算法精度第一次高于当时精度最高的二阶段目标检测算法。

所以,本文采用RetinaNet为车辆信息检测框架,将车辆信息分为三类,包括含车头的车外框、含车尾的车外框及车牌。算法的主要改进如下:

1)为了提高模型的前向推断速度,本文以文献[17]提出的MobileNet V3作为基础特征提取网络。

2)对于车牌这一小目标的检测精度不高问题,本文将RetinaNet基础特征提取网络的C5层经过上采样融合到FPN(Feature Pyramid Network)的P3层,并通过1×1的卷积对融合之后的特征层进行特征压缩,使浅层特征层有更好的车牌高级语义信息进而提高车牌检测精度。

3)本文通过引入文献[18]提出的GIoU(Generalized Intersection over Union)这一优化目标加快模型拟合预测框与真实框的位置关系,并提高模型的预测精度。

4)由于原始RetinaNet是针对常规的目标检测数据集设置的Anchor(检测过程中的候选框)尺寸,针对本文车辆信息检测这一数据集,本文使用文献[19]提出的K-means++聚类算法得到Anchor的较好尺寸并匹配到对应的特征图,进而提高检测准确度并提高召回率。

通过实验表明,本文提出的算法相较于其他目标检测算法精度有明显的提升。

1　基于RetinaNet改进的车辆信息检测

1.1　整体结构

本文算法的整体结构如图1所示,图中的C3、C4、C5、P3、P4、P5、P6、P7特征层和原始RetinaNet网络相对应。左边部分首先对图像进行预处理,将处理后的结果作为输入送入网络,然后经过轻量级网络提取初步特征,由于需要满足移动端实时检测需求,本文使用MobileNet V3作为基础特征提取网络;中间部分使用基于文献[20]提出的FPN(Feature Pyramid Network, FPN)算法改进的结构,对基础特征提取网络的低层细节特征信息与高层语义特征信息进行融合,以丰富每一特征层的感受野,使FPN输出的每一特征层都能预测不同尺度的目标。在原始RetinaNet的FPN模块中,深层特征层经过多次卷积及上采样操作之后,深层特征层的语义信息很难传达到浅层特征层,导致浅层用于检测小目标的特征层缺乏一定的深层语义信息。在本文的应用场景中,针对车牌这一小目标检测精度较低的问题,本文将基础特征提取网络的C5特征层经过四倍上采样与FPN的P3特征层堆叠(如图1加粗箭头表示),并通过1×1的卷积层对堆叠后的特征通道进行压缩融合恢复到原来的特征层通道数,以提高浅层特征层的高层语义信息;右边部分为分类和定位模块基于FPN提取的特征对原始图像中的目标进行分类和位置回归。分类子网络的输出通道数为9,即为每个最终输出特征层的Anchor数量;定位子网络的输出通道数为9×4,即为每个最终输出特征层的Anchor数量与Anchor位置预测值数量的乘积。其中所有的分类子网络和所有的定位子网络分别共享不同的权重。

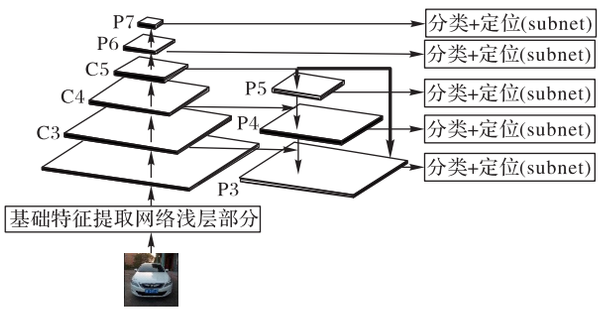


图1　改进的RetinaNet整体结构

Fig. 1　Overall structure of improved RetinaNet

1.2　目标定位的损失函数改进

当前很多目标检测算法的定位代价损失函数都使用预测框与真实框角点坐标距离的1-范数Loss或者2-范数Loss,然而在最终评测模型性能的时候是使用预测框与真实框的交并比(Intersection over Union, IoU)来判断是否正确检测到目标。由图2可以看出,模型对这两者的优化不是等价的,当预测框与真实框的2-范数Loss相等的时候两者的IoU却不相等。图中右上角在圆心的矩形框为真实框,右上角在圆上的矩形框为预测框,可以看出预测框与真实框的2-范数Loss都是相等的,但是第三个的IoU明显高于前两个,所以拟合效果也优于前两者。

但是若直接采取IoU当作评价模型性能好坏的标准,当预测框与真实框没有交集时,不管两者之间的距离有多远相对应的IoU都为零,这就导致这种情况出现时模型没有能力度量预测框与真实框的距离。基于此,本文提出了一种新的度量预测框与真实框拟合程度的方法,即GIoU。对于预测框A与真实框B,可以算出两者的最小凸集(包含A、B的最小包围框)C。通过最小凸集,可以计算GIoU。IoU和GIoU的计算公式如下:

$$IoU = \frac{A \cap B}{A \cup B} \tag{1}$$

$$GIoU = IoU - \left| C \backslash (A \cup B) \right| / |C| \tag{2}$$

式(2)中的减数部分分子为最小凸集中不包含A或者B的部分。GIoU与IoU相似,GIoU作为一个距离度量指标,其损失函数L_{GIoU}计算如式(3)所示:

$$L_{\mathrm{GIoU}} = 1 - GIoU \tag{3}$$

由GIoU计算公式可以得知,GIoU总是小于等于IoU,由IoU计算公式(1)可得$0 \leqslant \mathrm{IoU} \leqslant 1$,所以由式(3)可得$-1 \leqslant \mathrm{GIoU} \leqslant 1$,即只有当预测框与真实框完全重合时GIoU = 1。而且,当预测框与真实框IoU为零时GIoU转换为式(4):

$$GIoU = -1 + (A \cup B)/C \tag{4}$$

此时预测框与真实框相离越远GIoU越接近-1,而$A \cup B$不变,要对GIoU进行优化必须使C逐渐变小,即预测框与真

www.joca.cn

实框越来越接近。

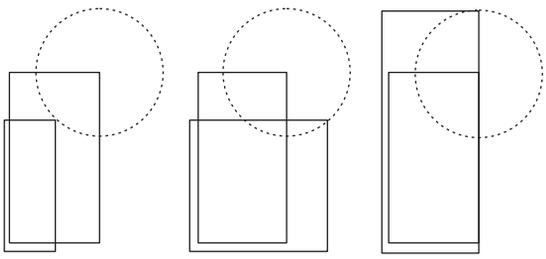


图2 相同2-范数loss不同IoU

Fig. 2 Different IoUs for same 2-norm loss

1.3 Anchor选取方法

Anchor机制起初是由Faster-RCNN提出，后来被广泛用于一阶段及二阶段目标检测框架，如SSD、YOLO、RetinaNet等。Anchor主要用来解决目标检测任务中尺度及宽高比例变化范围过大的问题，模型训练的时候真实标签被编码为相对于Anchor的偏移量，模型测试时根据模型预测相对于Anchor的偏移量来解码检测结果。通过Anchor机制，将尺度和宽高比例的空间划分为相应的几个子空间，降低了目标检测任务的难度，使模型更容易学习。针对本文的应用场景，由于RetinaNet使用的是一般物体检测数据集，所以原始RetinaNet选定的Anchor尺度和宽高比在本文的检测任务中并不适用。基于此，本文使用*K*-means++聚类算法对车辆信息数据集的真实标注框进行聚类操作，真实标注框长宽映射到模型输入大小size下的聚类结果如图3所示。

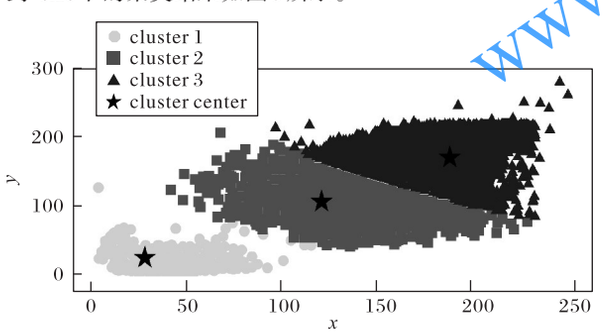


图3 真实boxes长宽聚类值

Fig. 3 Real boxes lenghth-width cluster values

由图3可以看到，有3个聚类簇，车牌的长宽对应坐标聚集在[31,19]附近，两种类型车外框的长宽对应坐标分别聚集在[124,130]、[159,171]附近。

所以，本文以[19×19,38×38,76×76,152×152,304×304]作为5个对应特征层的Anchor尺寸，以[0.6,1,1.8]作为Anchor的长宽比。使用对真实框维度聚类的方法可以使Anchor尺寸更加接近真实值，使模型更容易拟合目标的真实位置，从而降低模型的训练难度。

2 实验

2.1 实验环境说明和基本参数设置

本文实验在计算机配置为Ubuntu 16.04、CUDA（Compute Unified Device Architecture）8.0、cuDNN（CUDA Deep Neural Network library）5.1、NVIDIA Tesla P40的系统下进行，使用的深度学习框架为Pytorch 0.4.1，开发语言为Python 2.7。

本文采用的车辆信息数据集为自己采集，由Labelme标注软件标注，并转换成PASCAL VOC（Visual Object Classes）[21]的数据集格式，总样本量大小为29 340张，车辆类型、车牌颜色、拍摄车辆的角度、车辆在图片中的尺寸比例、光照、背景环境及图像分辨率在数据集中都存在很大的变化，对于检测算法有较高的要求。通常情况下，单个检测目标可以通过1 000个测试样本来评估模型的检测性能。综合考虑本文改进方法在测试集上的有效评估和模型通过训练样本得到充分训练两方面因素，本实验基于车辆信息检测模型的任务复杂度和PASCAL VOC设置数据集的方式将总样本按照大约8:2的比例随机划分为训练集和测试集。训练阶段首先对图片先进行数据增强预处理以提高模型的泛化能力，包括随机缩放、随机水平翻转、随机padding、随机裁剪、随机旋转、随机色度扭曲、随机添加噪声、随机对比度变换、像素值标准化，并将增强后的图片resize到224×224送入模型。在测试阶段，首先对输入的图片进行像素值标准化及进一步resize到224×224的操作，然后送入模型进行预测。同时在训练阶段，为了防止Anchor与真实框没有IoU大于0.5的情况，本实验设置当出现这种情况时将与真实框IoU最高的Anchor作为对应类别的正样本参与训练。

基于迁移学习的方法，本文使用在ImageNet数据集预训练的MobileNet V3作为基础特征提取网络，优化器选择小批量随机梯度下降（mini-batch Stochastic Gradient Descent，mini-batch SGD），动量为0.9，权重衰减为1E-4，batch size设置为512，初始学习率设置为0.001，每经过30个epoch学习率下降10倍，共训练100个epoch。

2.2 实验分析

本实验针对车牌、包含车头部分的车外框、包含车尾部分的车外框这三类检测目标，根据以下公式计算查准率（Precision）与查全率（Recall）。

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

其中：准确率代表模型预测出的样本中真样本的比例；查全率为模型预测出的正样本占所有正样本的比例。其中：*TP*（True Positive）为被模型预测为真的正样本；*FP*（False Positive）为模型预测为真的负样本；*FN*（False Negative）为模型预测为非真的正样本。按照PASCAL VOC计算平均准确率（Average Precision，AP）的方法，由每个类别的查准率和查全率可以计算出每个类别的AP。

实验一以MobileNet V3为基础网络，以RetinaNet为原始检测框架，以交并比阈值为0.7时每个类别的AP和三个类别的平均准确率均值（mean AP，mAP）作为不同改进方法作用于原始RetinaNet的评价标准。实验结果如表1所示，当在MobileNet V3的基础上不使用维度聚类操作而直接使用RetinaNet原始的Anchor尺寸及长宽比时，模型性能较低，尤其车牌部分AP值较低；在原始模型的基础上把位置回归的代价函数改进为GIoU-Loss后，车牌的AP值有10个百分点的提高；在原始模型的基础上加入维度聚类操作后，车辆外宽的AP值稳定的情况下车牌的AP值有接近27个百分点的提高；

www.joca.cn

在原始模型FPN的基础上嵌入特征融合分支后,车牌的AP值有10.3个百分点的提高,同时mAP有3.4个百分点的提高;当同时引入维度聚类和改进的FPN特征融合的方法后,mAP相对于只加入维度聚类有1.2个百分点的提高,同时mAP相对于只加入改进的FPN特征融合有6.5个百分点的提高;当同时引入维度聚类和GIoU-Loss的方法后,mAP相对于只引入GIoU-Loss有6.8个百分点的提高,同时mAP相对于只加入维度聚类有1.3个百分点的提高;当同时引入GIoU-Loss和改进的FPN特征融合方式的方法后,mAP相对于只引入GIoU-Loss有0.6个百分点的提高,同时mAP相对于只加入改进的FPN特征融合方式有0.4个百分点的提高;当同时引入维度聚类、改进的特征融合方式、GIoU-Loss的方法后,车牌的AP值相对于原始RetinaNet的方法有30.3个百分点的提高,同时mAP有10.2个百分点的提高。

表1　不同模块对模型的性能影响对比

Tab. 1　Comparison of effects of different modules on performance of model

维度聚类	特征融合	GIoU-Loss	AP			mAP
			车头	车尾	车牌	
无	无	无	0.980	0.993	0.638	0.870
有	无	无	0.983	0.990	0.900	0.957
无	有	无	0.981	0.991	0.741	0.904
无	无	有	0.980	0.993	0.733	0.902
无	有	有	0.981	0.992	0.752	0.908
有	无	有	0.990	0.991	0.931	0.970
有	有	无	0.991	0.993	0.923	0.969
有	有	有	0.982	0.992	0.941	0.972

改进后算法与原算法预测实验效果对比如图4所示。由图4第一行两张图可见,改进后的车牌检测效果相较于未改进的算法定位更加准确;由图4的第二行两张图可见,原始算法中有车灯被误检为车牌,改进后的车牌检测效果准确率和召回率都更高。



(a) 原始算法预测　　(b) 改进后算法预测

图4　改进后算法与原算法效果比较

Fig. 4　Comparison between improved algorithm and original algorithm

实验二为本文提出的算法与其他目标检测算法在本文的车辆信息数据集上的性能对比,以IoU阈值为0.5进行性能评估,实验结果如表2所示。由实验结果可知,本文所提出的算法准确率高于其他算法。

表2　本文算法与其他方法的性能对比

Tab. 2　Performance comparison among proposed method and other methods

Model	Backbone	mAP
Faster-RCNN	ResNet50[22]	0.976
SSD	Vgg-16[23]	0.968
RetinaNet	ResNet50	0.988
本文算法	ResNet50	0.999

3　结语

针对RetinaNet原始算法直接运用在本文的车辆信息检测数据集上性能不高的问题,本文使用维度聚类算法初始化Anchor使模型更容易收敛,在原有模型的基础上融合深层的特征信息到浅层特征层以提高浅层特征层的车牌高级语义信息,同时引入更加接近检测任务的评价标准GIoU-Loss作为代价函数以提高模型的定位能力。实验结果表明,改进后的算法对于车牌这一小目标的检测准确率及检测的平均准确率都有很大的提高。经测试,本文提出的以MobileNet V3为基础特征提取网络的车辆信息检测算法经移动端深度学习框架优化后,在ARM v7设备上单帧前向推断用时可达100 ms,可以满足工业实时检测要求。实验结果证明了本文提出算法的可行性和实用性,当然本文提出的方法还有很多地方需要改进,比如本文使用未剪枝的MobileNet V3作为基础特征提取网络,一些冗余通道的存在增加了计算量且降低了推断速度,在本文的模型基础上进行模型压缩仍然需要更进一步的研究。

参考文献(References)

[1] 徐子豪,黄伟泉,王胤. 基于深度学习的监控视频中多类别车辆检测[J]. 计算机应用, 2019, 39(3):700-705. (XU Z H, HUANG W G, WANG Y. Multi-class vehicle detection in surveillance video based on deep learning[J]. Journal of Computer Applications, 2019, 39(3):700-705.)

[2] 王卫东,程丹. 监控场景下的实时车辆检测方法[J]. 电子测量与仪器学报, 2018, 32(7):83-88. (WANG W D, CHENG D. Real-time vehicle detection method for video surveillance[J]. Journal of Electronic Measurement and Instrument, 2018, 32(7): 83-88.)

[3] LIU Y, TIAN B, CHEN S, et al. A survey of vision-based vehicle detection and tracking techniques in ITS[C]// Proceedings of 2013 IEEE International Conference on Vehicular Electronics and Safety. Piscataway: IEEE, 2013: 72-77.

[4] KIM H, LEE Y, YIM B, et al. On-road object detection using deep neural network[C]// Proceedings of the 2016 IEEE International Conference on Consumer Electronics-Asia. Piscataway: IEEE, 2016: 1-4.

[5] LAW H, DENG J. CornerNet: detecting objects as paired keypoints[C]// Proceedings of the 2018 European Conference on Computer Vision, LNCS 11218. Cham: Springer, 2018: 765-781.

[6] ZHOU X, ZHUO J, KRAHENBUHL P. Bottom-up object detection by grouping extreme and center points[C]// Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 850-859.

[7] DUAN K, BAI S, XIE L, et al. CenterNet: keypoint triplets for object detection [EB/OL]. [2019-04-19]. https://arxiv. org/pdf/1904. 08189. pdf.

[8] GIRSHICK R. Fast R-CNN[C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE,

www.joca.cn

2015: 1440-1448.

[9] 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 and Machine Intelligence, 2015, 39(6):1137-1149.

[10] DAI J, LI Y, HE K, et al. R-FCN: object detection via region-based fully convolutional networks[C]// Advances in Neural Information Processing Systems 29. Barcelona: NIPS, 2016: 379-387.

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

[12] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 6517-6525.

[13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9905. Cham: Springer, 2016: 21-37.

[14] FU C, LIU W, RANGA A, et al. DSSD: deconvolutional single shot detector [EB/OL]. [2019-01-17]. https://arxiv. org/pdf/1701. 06659. pdf.

[15] LI Z, ZHOU F. FSSD: feature fusion single shot multibox detector [EB/OL]. [2018-12-17]. https://arxiv. org/pdf/1712. 00960. pdf.

[16] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 2999-3007.

[17] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetV3 [EB/OL]. [2019-05-19]. https://arxiv. org/pdf/1905. 02244. pdf.

[18] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]// Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 658-666.

[19] ARTHUR D, VASSILVITSKII S. *K*-means++: the advantages of careful seeding[C]// Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms. New York: ACM, 2007: 1027-1035.

[20] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 936-944.

[21] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal Visual Object Classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

[22] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.

[23] SAINATH T N, KINGSBURY B, SAON G, et al. Deep convolutional neural networks for large-scale speech tasks[J]. Neural Networks, 2015, 64: 39-48.

This work is partially supported by the National Natural Science Foundation of China (51805367), the Tianjin Natural Science Foundation (18JCQNJC04800, 18JCZDJC31800).

LIU Ge, born in 1994, M. S. candidate. His research interests include computer vision.

ZHENG Yelong, born in 1987, Ph. D., lecturer. His research interests include machine vision, mechanical measurement.

ZHAO Meirong, born in 1967, Ph. D., professor. Her research interests include machine vision, mechanical measurement.

www.joca.cn