

doi:10.6041/j.issn.1000-1298.2020.05.003

基于 IFSSD 卷积神经网络的柚子采摘目标检测模型

肖德琴¹ 蔡家豪¹ 林思聪¹ 杨秋妹¹ 谢晓君² 郭婉怡¹

(1. 华南农业大学数学与信息学院, 广州 510642; 2. 广州中医药大学医学信息工程学院, 广州 510006)

摘要: 为了解决柚子采摘时传统水果检测模型对于小目标柚子漏检和将叶子误检为膨大期柚子的问题,设计了一种改进的特征融合单镜头检测器(InceptionV3-feature fusion single shot-multibox detector, IFSSD)。该检测器以特征融合单发多盒探测器(Feature fusion single shot-multibox detector, FSSD)为基础检测器,以改进的 InceptionV3 网络作为骨干网络代替超深度卷积神经网络(Very deep convolutional networks 16, VGG16),从而提高了计算效率,同时使用 Focal Loss 损失函数代替 Multibox Loss 损失函数,进而改善了由于正负样本不平衡导致的检测器误检情况。对测试数据集进行检测,结果表明,该模型的检测准确率为 93.7% (IoU 大于 0.5),在单个 NVIDIA RTX 2060 GPU 上每幅图像检测时间为 29 s。本文模型可以实现树上柚子的自动检测。

关键词: 柚子; 采摘; 实时检测; 深度学习; 特征融合; 改进的特征融合单镜头检测器

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 1000-1298(2020)05-0028-08

OSID:



Grapefruit Detection Model Based on IFSSD Convolution Network

XIAO Deqin¹ CAI Jiahao¹ LIN Sicong¹ YANG Qiumei¹ XIE Xiaojun² GUO Wanyi¹

(1. College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China

2. School of Medical Information Engineering, Guangzhou University of Chinese Medicine, Guangzhou 510006, China)

Abstract: The detection, identification and precise positioning of fruit under natural conditions based on machine vision is very important for the development of intelligent picking robots. In order to solve the problem that the traditional fruit detection model for the small target grapefruit missed detection and the leaf was falsely detected as the expansion period grapefruit, an improved feature fusion single multibox detector (InceptionV3-feature fusion single shot-multibox detector, IFSSD) was designed. The feature fusion single multibox detector (feature fusion single shot-multibox detector, FSSD) was used as a base detector and optimized in two ways. On the one hand, the improved InceptionV3 network was used instead of very deep convolutional networks 16 (VGG16) as the backbone network to improve computational efficiency. On the other hand, the Focal Loss function was used instead of the Multibox Loss function, which improved the mischeck ingress of the detector due to the imbalance of positive and negative samples. Finally, the test data set was verified, and the results showed that the model achieved an average accuracy of 93.7% (mAP) (IoU was greater than 0.5). The time of one image was 29 s. The model proposed can automatically detect the grapefruit in the grapefruit tree and locate it accurately in real time, which effectively promoted the development of intelligent picking robot.

Key words: grapefruit; picking; real-time detection; deep learning; feature fusion; IFSSD

0 引言

我国是柚子种植面积最大的国家,柚子产量仅

次于美国,位于世界第二位^[1]。自 2000 年以来,我国柚类产品得到国际市场的认可,柚类出口量急剧增长,具有巨大的发展前景。在柚子的生产过程中,

收稿日期: 2019-07-16 修回日期: 2019-12-04

基金项目: 国家重点研发计划项目(2017YFD0701601)、广东省重点领域研发计划项目(2019B020217003、2019B020214005)和广东省科技计划项目(2015A020224034)

作者简介: 肖德琴(1970—),女,教授,博士生导师,主要从事物联网和农业图像视频处理研究,E-mail: deqinx@scau.edu.cn

通信作者: 杨秋妹(1983—),女,讲师,博士,主要从事农业图像视频处理研究,E-mail: yqmbegonia@163.com

采摘是最耗时费力的过程,后续的存储和加工过程直接取决于采摘效果,所以保证柚子采摘质量成为关键环节。目前,国内的水果采摘主要是由人工完成,工作强度大,且效率低下^[2]。有研究表明,用于采摘过程的费用占生产总费用的 50% ~ 70%。随着水果的大面积种植,传统的人工采摘方式已经不能满足生产需求,研究水果采摘机械化意义重大^[3]。

在自然环境下,快速、准确识别与定位柚子是实现柚子自动化采摘的重点研究方向之一,因此,研究者提出一系列水果检测识别方法^[4-6]。深度学习在农业领域取得了显著的进展^[7-9]。在传统方法中,柚子的颜色、形状以及纹理特征因受自然环境的影响而产生变化,所以很难对柚子进行有效的描述,从而对算法的泛化性产生了极大影响。基于深度学习的方法可以通过学习训练数据的特征来提供准确的结果。深度学习方法已被用于农业研究的各个领域,如 FERENTINOS^[10]使用卷积神经网络对健康和患病植物的简单叶片图像进行植物疾病检测和诊断。彭红星等^[11]使用 Res101 结合 SSD 构建了深度水果检测模型,对自然条件下的多种水果进行检测识别。然而小目标柚子在图像视野中占比较少,边缘特征不明显,甚至缺失,由于分辨率和信息有限,使得传统基于深度学习的柚子检测算法效果并不理想^[12]。同时,常见的目标检测算法是根据重叠度 (Intersection over Union, IoU) 阈值,从先验框中挑选 IoU 小于 0.7 和 IoU 大于等于 0.7 的先验框分别作为正负样本,进行分类网络的训练。然而包含叶子背景的柚子负样本远远大于正样本,导致样本不平衡,使柚子检测模型存在很严重的偏向性。

为此,本文提出一种基于 IFSSD 卷积网络的柚子检测模型,对膨大期、成熟期两个阶段的柚子进行检测试验。

1 材料与方法

1.1 试验数据的获取

试验中的柚子图像于 2018 年 4 月 1 日—10 月 10 日在广东省梅州市的顺兴果园获得。果园中共种植柚子树 1 500 余棵,柚子树随机分布。选用海康威视 DS-2CD3T56WD-I3 型摄像头进行拍摄,将 4 个摄像头固定在离地面 3 m 的支架上,摄像头分别向正北、正东、正西、正南 4 个方向。支架之间的距离为 60 m,共 9 个支架。试验设备和现场图如图 1 所示。拍摄图像的尺寸为 1 920 像素 × 1 080 像素, JPG 格式。于每天 09:00—17:00 每隔 1 h 拍摄一次图像,其中包含了晴天正午、晴天傍晚以及阴天

正午等特殊光照条件,共获得了 5 000 幅图像。其中图像中存在枝叶遮挡和果实互相遮挡的图像数量分别为 3 563 幅和 2 469 幅。



图 1 试验设备和现场图

Fig. 1 Photos of experimental equipment and site

1.2 柚子检测数据集的构建

在 5 000 幅柚子树图像中利用随机抽取的方式选择 4 000 幅图像,使用 Labelme 软件进行标记作为训练集,其他图像作为测试集。

图 2 为训练集的样本预处理过程,通过对图像分别进行翻转、平移和裁剪来扩充数据集。



图 2 图像预处理

Fig. 2 Image preprocessing

在柚子训练集中,柚子分为膨大期果实和成熟期果实,如图 3 所示,膨大期果实和成熟期果实在颜



图 3 膨大期和成熟期柚子果实

Fig. 3 Grapefruit fruit during expansion and maturity

色上有明显的区分。

1.3 基于深度学习的柚子检测算法

1.3.1 InceptionV3 网络

InceptionV3 是 Google Inception 系列网络中的一种,该网络在 InceptionV2 的基础上提出了一种新的结构,网中网(Network in network, NIN)结构^[13]。InceptionV3 是 Inception 模块按照 NIN 结构构建的一种新型网络。InceptionV3 的 Inception 模块如图 4 所示。

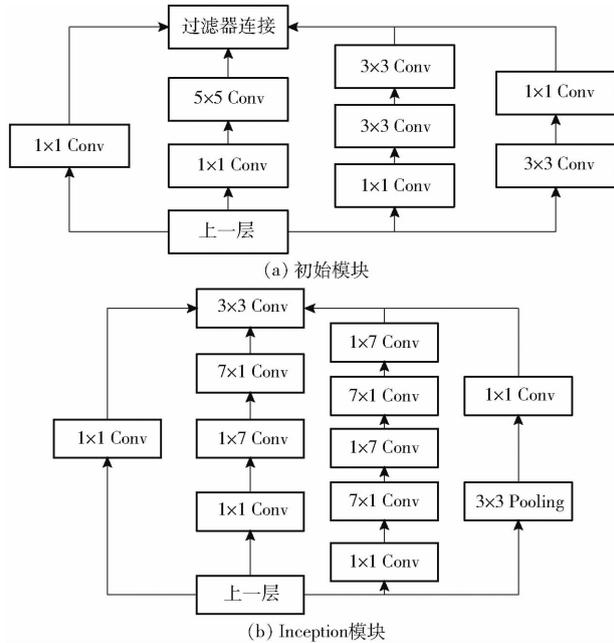


图 4 InceptionV3 的 Inception 模块结构

Fig. 4 Structure of InceptionV3 Inception module

如图 4 所示, InceptionV3 模块通过提取不同尺度的特征来减少模型的参数数量^[14],同时多尺度特征可以提高模型的识别能力。研究证明,卷积核为 1×1 的卷积层可以通过少量的计算实现特征的转换、提高网络的识别能力以及改变卷积模块输出的通道数^[15]。InceptionV3 的第 5~9 Inception 模块使用更大的卷积核来获取更加抽象的特征。针对于卷积核为 7×7 且容易产生大量参数的卷积层, InceptionV3 使用 1×7 卷积层和 7×1 卷积层,从而提高算法的效率并降低过拟合的风险。试验证明,这种不对称卷积结构可以处理越来越丰富的空间特征并增加特征多样性。

1.3.2 SSD 与 FSSD

InceptionV3 网络模型主要是对柚子的成熟情况进行分类,存在一个局限条件就是每幅图像中只能有一个柚子,柚子占据图像的面积要尽量大。为了解决这个问题,许多目标检测方法被提出,例如使用颜色和纹理特征来检测水果。由于柚子等水果是簇状结果,因此在柚子检测过程当中容易发生重叠

和误检的情况,使得检测精度大大降低。

目前有许多用于目标检测的深度学习方法被广泛应用于农业领域,如借助区域提议网络实时目标检测算法(Towards real-time object detection with region proposal networks, Faster-RCNN)^[16]、通过基于区域的全卷积网络目标检测算法(Object detection via region-based fully convolutional networks, R-FCN)^[17]、单镜头多盒探测器算法(Single shot multibox detector, SSD)^[18]、YOLO 算法(You only look once, YOLO)^[19]以及 RetinaNet^[20]等。此外,柚子检测对于算法的实时性能和小物体检测有着更高的要求,所以需要在保证精度符合要求的情况下尽量选择效率更高并且对小物体检测更有效的算法。

SSD 是一种在图像中使用同一个深度神经网络进行检测和识别检测目标的算法。SSD 生成一系列尺寸不同的候选框,通过计算标注框与候选框的偏置值来匹配两者,一般每个标注框都会匹配多个候选框。SSD 将 IoU 大于 0.5 的候选框认为是正样本,其他设置为负样本^[21]。

SSD 使用 VGG16 作为骨干网络^[22],通过在 Conv4_3、FC7、Conv7_2、Conv6_2、Conv8_2 和 pool6 中提取不同尺寸的特征图来形成多尺度检测,SSD 的网络结构图如图 5a 所示。其中特征图尺寸为 5 像素 \times 5 像素、3 像素 \times 3 像素和 1 像素 \times 1 像素,低级特征图有利于小目标的检测,特征图尺寸为 38 像素 \times 38 像素、19 像素 \times 19 像素和 10 像素 \times 10 像素的高级特征图则有利于大目标检测。SSD 使用通道数为 1024 的 3×3 卷积层和 1×1 卷积层代替全连接层和丢失层,从而达到减少模型参数、提高计算效率以及有效防止过度拟合的目的。SSD 的损失函数为

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \quad (1)$$

式中 $L(x, c, l, g)$ ——总体损失值

x ——卷积特征值 c ——真实类别

l ——预测框的位置坐标值

g ——真实框的位置坐标值

L_{loc} ——预测框和真实框间的平滑损失值

N ——候选框的数量

L_{conf} ——多类置信度下的 softmax 损失值

α —— L_{loc} 的权重

Multibox Loss 由两部分组成,前者使用了 softmax 进行分类损失计算,后者通过局部损失来预测位置。

SSD 为柚子检测任务提供了坚实的基础,其主要优点是检测速度快。然而,试验表明,SSD 存在以

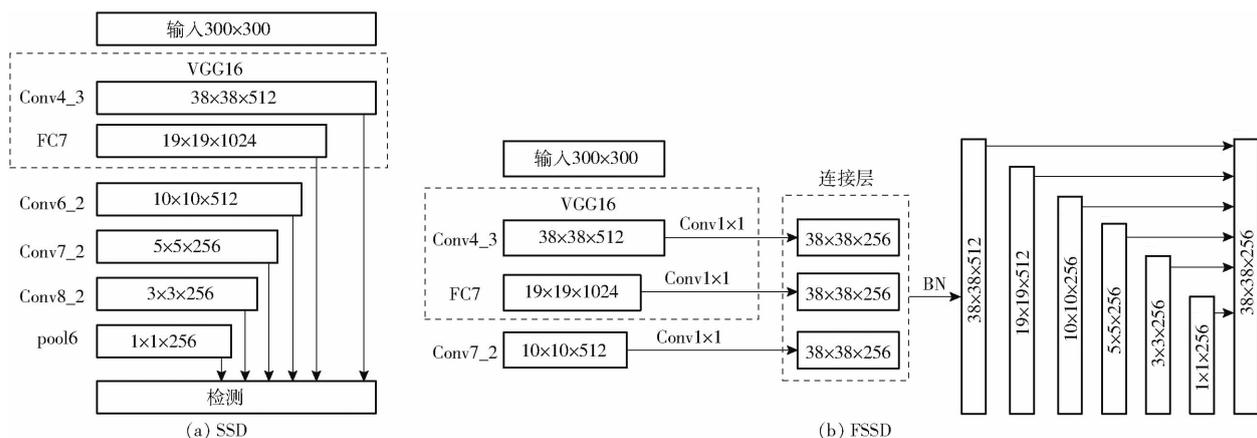


图 5 SSD 和 FSSD 模型的结构

Fig. 5 Model structure of SSD and FSSD

下问题:较小的目标难以识别;一些背景区域被识别为目标对象;重复识别目标。为了解决这些问题,SSD 模型被修改为 FSSD 模型且不增加模型参数的数量,并降低检测速度。

FSSD 模型的提出者认为 SSD 模型中用于预测的每个层之间的关系是相互独立的^[23]。FSSD 模型是将这些层与不同比例的特征图进行融合^[24],以便它们相互通信并提高准确性。但是,这种方法不适合在 FSSD 模型中融合小于 10×10 的特征映射,因为几乎没有要合并的信息。合并 3 个较大层的特征图以生成 $38 \text{ 像素} \times 38 \text{ 像素}$ 的特征图,并生成特征金字塔网络 (Feature pyramid networks, FPN)^[25]。最后,FSSD 模型从 FPN 中提取特征。

FSSD 模型的结构如图 5b 所示。FSSD 模型以 VGG16 作为主要骨干网络。Conv7_2 的特征映射变为 10×10 。FC7 和 Conv7_2 使用双线性插值将特征图尺寸调整为 $38 \text{ 像素} \times 38 \text{ 像素}$,然后将它们与 Conv4_3 连接,在该步骤中,合并层的通道数为 768 ($256 + 256 + 256$)。然后 BN 作用于合并层,该层的通道数减小到 512。另外使用 5 个卷积层来减小特征图的尺寸。最后,获得 6 种不同尺寸特征图 ($38 \text{ 像素} \times 38 \text{ 像素}$ 、 $19 \text{ 像素} \times 19 \text{ 像素}$ 、 $10 \text{ 像素} \times 10 \text{ 像素}$ 、 $5 \text{ 像素} \times 5 \text{ 像素}$ 、 $3 \text{ 像素} \times 3 \text{ 像素}$ 和 $1 \text{ 像素} \times 1 \text{ 像素}$)

并将它们用于多尺度预测。这种方法可以将浅层细节特征与高级语义特征相结合,可以比 SSD 模型更好地识别小物体^[26],并降低误检率。但 FSSD 模型的速度比 SSD 模型的速度略有下降。

1.3.3 IFSSD 模型建立

在本文中,只需要判断检测目标是否是柚子,并不需要进行多类别分类。柚子的特征相对简单,可以简化模型,所以本文选择 InceptionV3 作为 IFSSD 的骨干网络。IFSSD 模型结构如图 6 所示。

(1) InceptionV3 深度网络优化

为了减少参数数量和提高了对小目标的检测效果,本文对 InceptionV3 进行了调整,以实现检测器基于不同感受野的多尺度检测。本文对 mixed7 层进行了删除。IFSSD 模型在 mixed6 层之后添加 1×1 卷积,步长为 2,通道数为 768,然后进行批量归一化。新添加的图层 mixed6_s 的特征图尺寸为 $18 \text{ 像素} \times 18 \text{ 像素}$ 。mixed7 层的特征图尺寸从原来的 $35 \text{ 像素} \times 35 \text{ 像素}$ 变成 $18 \text{ 像素} \times 18 \text{ 像素}$ 。在 mixed7 之后加入卷积核为 1×1 、步长为 2、通道数为 768 的卷积层。mixed7_s 的特征图尺寸为 $9 \text{ 像素} \times 9 \text{ 像素}$ 。图 6 中改进后的网络结构用黄色方框表示。

IFSSD 模型与 FSSD 模型类似,都对较大的特征图进行合并。在 IFSSD 模型的 InceptionV3 中,分别

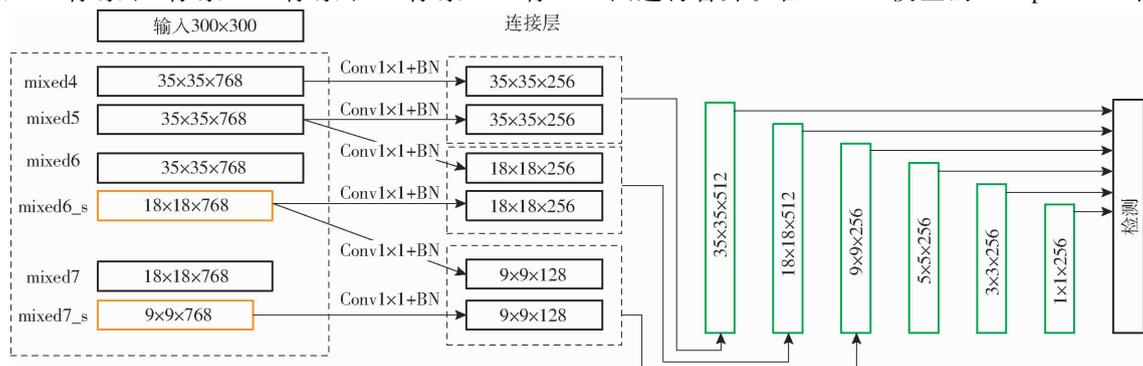


图 6 基于 IFSSD 的柚子检测模型结构

Fig. 6 IFFSD-based grapefruit detection model structure

选择特征图尺寸各不相同的4层模块,分别是mixed4、mixed5、mixed6_s、mixed7_s。在第1个连接层中,卷积核为 1×1 、步长1、通道数为256的卷积层被应用于mixed4和mixed5。这样做的目的是在不改变特征图尺寸的前提下减小模型参数,从而减少模型中的参数数量并增加增强模型的非线性识别能力。卷积后生成的特征图尺寸为35像素 \times 35像素,然后将mixed4和mixed5卷积生成的特征图合并到一起。在第2个连接层中,在mixed5模块之后添加一层卷积核为 1×1 、步长为2、通道数为256的卷积层,其生成的特征图尺寸为18像素 \times 18像素。在mixed6_s之后添加一个卷积核为 1×1 、步长为1、通道数为256的卷积层。其生成的特征图尺寸为18像素 \times 18像素。然后,对两个特征图进行合并。在第3个连接层中,一个卷积核为 1×1 、步长为2、通道数为128的卷积层被添加到mixed6_s之后,生成了 $9 \times 9 \times 128$ 的特征图。最后卷积核为 1×1 、步长为1、通道数为128的卷积层被添加到mixed7_s之后,其生成了 $9 \times 9 \times 128$ 的特征图,并将两个特征图进行合并。

在IFSSD模型中, $35 \times 35 \times 512$ 、 $18 \times 18 \times 512$ 和 $9 \times 9 \times 256$ 的连接层用于生成FPN。在 $9 \times 9 \times 256$ 的连接层之后添加一个卷积核为 3×3 、步长为2、通道数为256的SAME卷积层,生成特征图尺寸为5像素 \times 5像素。其后,添加一个卷积核为 3×3 、步长为 1×1 、通道数为256以及使用VALID的卷积层,生成特征图尺寸为3像素 \times 3像素。最后,在 $3 \times 3 \times 256$ 层之后,添加一个卷积核为 3×3 、步长为1、通道数为256的VALID卷积层,生成的特征图尺寸为1像素 \times 1像素。图中绿色方框所示的6层不同尺度的预测框用于获得柚子的位置和置信度。

可以看出,IFSSD模型使用双线性插值从VGG16中3个不同层获得相同的尺寸,并将它们融合在一起以获得输出层并生成FPN。IFSSD模型使用 1×1 卷积来修改InceptionV3中的mixed4、mixed5、mixed6_s、mixed7_s层的通道数和图像尺寸,并且成对地融合上述修改后的卷积层以获得3个不同层,然后基于这3个层生成FPN。

(2) 损失函数优化

一阶段算法(one-stage)和二阶段算法(two-stage)的表现不同主要是由大量前景背景类别不平衡导致^[27]。二阶段算法中,在候选框阶段,通过得分和非极大值抑制(Non-maximum suppression, NMS)筛选过滤掉了大量的负样本,然后在分类回归阶段又固定了正负样本比例,或者通过难例挖掘(OHEM)在线挖掘困难样本使得前景和背景相对平

衡。而一阶段算法需要产生约100 kB的候选位置。虽然有重复采样,但是训练仍然被大量负样本所主导。本文使用的IFSSD模型属于one-stage模型,使用Focal Loss代替原有的损失函数。

Focal Loss主要是为了解决one-stage目标检测中正负样本比例严重失衡的问题^[28]。该损失函数降低了大量简单负样本在训练中所占的权重,也可理解为一种困难样本挖掘。Focal Loss是在交叉熵损失函数基础上进行的修改。二分类交叉熵损失值计算式为

$$L = \begin{cases} -\lg y' & (y=1) \\ -\lg(1-y') & (y=0) \end{cases} \quad (2)$$

式中 L ——二分类交叉熵的损失值

y ——真实分类 y' ——预测分类

y' 是激活函数的输出,取值在0~1之间。可见普通的交叉熵对于正样本而言,输出概率越大损失越小;对于负样本而言,输出概率越小则损失越小。此时的损失函数在大量简单样本的迭代过程中比较缓慢且可能无法优化至最优。

Focal Loss在原有基础上加了一个常数因子 γ ,以减少易分类样本的损失,从而更关注于困难的、错分的样本。此外,加入平衡因子 a_i ,用来平衡正负样本本身的比例不均,具体公式如下

$$F_L(p_i) = -a_i(1-p_i)^\gamma \lg p_i \quad (3)$$

式中 p_i ——不同类别的概率

a_i ——正样本和负样本的比例,前景类别使用 a_i 时,对应的背景类别使用 $1-a_i$

2 试验与结果分析

2.1 试验方法与参数设置

柚子数据集中包含一定量果实膨大期和成熟期的图像。训练集和验证集以4:1的比例进行随机划分。其中训练集包含4000幅图像,使用Labelme标注软件对每幅图像中成熟期和果实膨大期的柚子进行标注,最后按照PASCAL VOC数据格式^[29]构建训练集,另外1000幅未被标注图像作为测试集,用来评价模型在未知柚子图像数据集上的泛化性能。

为了评估IFSSD模型对柚子的检测识别性能,首先通过预处理的4000幅RGB图像创建IFSSD模型。剩余的1000幅测试图像用于IFSSD模型的测试。测试环境为IntelXeon E3-1245 v3 CPU、主频3.40 GHz、内存32 GB, GPU为GeForce RTX 2060、操作系统为Ubuntu 16.04, Cuda版本为9.0, Keras深度学习框架, OpenCV 4.0、Python3.7编程语言。

IFSSD模型使用自适应矩估计(Adaptive moment estimation)作为优化算法,同时以TensorFlow作为张量操作库的Keras深度学习框架

在 GeForce RTX2060 显卡上进行加速运算。

本文模型设置不同参数,通过交叉验证选取最优参数设置,初始学习率设置为 0.001,经过多次迭代后微调为 10^{-4} ,偏差 1 设置为 0.9,偏差 2 设置为 0.999,归一化的最小值边界设置为 0.000 1,权重衰变为 5×10^{-4} ,图像批量为 16。

本文模型对 300 像素 \times 300 像素图像的检测速率为 29 f/s,训练过程中损失值变化曲线如图 7 所示,随着迭代次数增加,IFSSD 模型的损失值不断下降,经过 100 次迭代后,下降趋势趋向于平缓,并在 200 次迭代后趋于稳定。

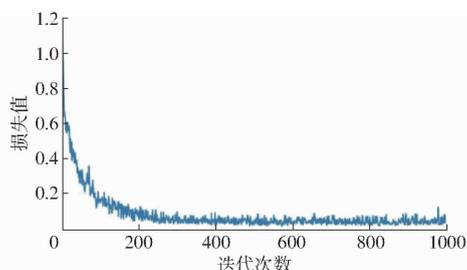


图 7 IFSSD 模型训练过程中损失值的变化曲线

Fig.7 Change of loss value during IFSSD training

2.2 检测结果分析

本文使用 IFSSD、SSD、FSSD 模型分别对数据集进行检测。表 1 为检测结果,其中检测出来的目标定为正样本,未检测出来的目标定为负样本。IoU 阈值为 0.5,试验结果有 4 种,分别为 IoU 值小于等于 0.5 (FP)、IoU 值大于 0.5 (TP)、未检测出真实目标 FN、TN。本文不统计 TN 类样本。准确率计算公式为

$$P = \frac{T_p}{T_p + F_p} \times 100\% \quad (4)$$

式中 T_p ——检测出的正样本数量

F_p ——检测出的负样本数量

表 1 不同检测模型的检测结果

Tab.1 Detection results of different detection models

模型	准确率/%	模型规模/ MB	每幅图像检测 时间/s
SSD	84.3	28.2	38
FSSD	91.5	28.2	41
IFSSD	93.7	19.7	29

在检测精度方面,IFSSD 模型的准确率为 93.7%,分别比 SSD 模型和 FSSD 模型提高了 9.4 个百分点和 2.2 个百分点。在图像检测速率方面,IFSSD 模型每幅图像检测时间为 29 s,低于 SSD 模型的 38 s 和 FSSD 模型的 41 s。

如表 2 所示,IFSSD 模型的准确率比梯度直方图 (Histogram of oriented gradient, HOG) + 支持向量

机 (Support vector machine, SVM) 和变形零件模型 (Deformable parts model, DPM) + SVM 提高了 14.1 个百分点和 16.4 个百分点,这是由于传统目标检测算法只适用于特征明显、背景简单的图像,而在实际应用中,背景复杂多变,且待检测的目标复杂多变,很难通过一般的抽象特征完成对目标的检测,而深度学习可以提取同一目标的丰富特征,完成目标的检测。然而在模型规模上 IFSSD 模型要大于 HOG + SVM 和 DPM + SVM,这是由于 IFSSD 模型的参数较多。从效率方面看,IFSSD 模型的效率与传统方法较接近。

表 2 IFSSD 与传统检测模型性能的比较

Tab.2 Comparison of IFSSD and traditional grapefruit detection models

模型	准确率/ %	模型规模/ MB	每幅图像 检测时间/s
IFSSD	93.7	19.7	29
HOG + SVM	79.6	3.0	27
DPM + SVM	77.3	2.5	31
Faster R-CNN	85.9	52.0	30
YOLOv2 - DarkNet - 19	84.3	23.5	21

相对于 Faster R-CNN 和 YOLOv2 - DarkNet - 19,由于 IFSSD 模型加入了 FPN 结构和 Inception 骨干网络,使得准确率分别提高了 7.8 个百分点和 9.4 个百分点;从检测时间看,IFSSD 模型与 Faster R-CNN 模型较接近。

本文分别从目标大小、物候期、重叠情况、光照情况以及有无枝叶遮挡等方面分别对 SSD、FSSD、IFSSD 模型进行比较。从表 3 可以看出,IFSSD 模型的总体性能优于 FSSD、SSD 模型。从目标大小方面看,加入了 FPN 模块的 FSSD、IFSSD 模型在小目标检测方面有了一定的提升,这是由于 FPN 能够从更

表 3 不同图像类型下不同模型的检测准确率

Tab.3 Detection accuracy of different models under different image types

图像类型	SSD	FSSD	IFSSD
小目标	79.63	85.15	86.68
中目标	85.64	86.24	89.53
大目标	92.12	92.14	94.16
膨胀期	85.42	86.12	90.63
成熟期	89.63	91.18	92.78
相互重叠	77.14	79.36	83.14
无重叠	90.88	91.24	93.26
强光	86.11	86.26	88.66
弱光	85.31	87.22	89.67
正常光	92.11	92.65	94.68
有枝叶遮挡	88.12	88.21	89.74
无枝叶遮挡	90.41	93.12	96.88

小的特征跨度进行特征融合,来进行多尺度预测。从不同物候期方面进行分析,IFSSD模型对膨胀期的识别提升较为明显,相对于SSD、FSSD模型分别提高了5.21个百分点和4.51个百分点,这是由于膨胀期的柚子颜色为绿色,容易因正负样本不平衡而导致错检和漏检,Focal Loss的引入缓解了这个问题。但是在重叠、光照和有枝叶遮挡的情况下,IFSSD模型检测效果没有明显的提高,仍然需要进一步优化。由于摄像机型号不同,不同的焦距和距离导致柚子图像的尺寸不同,本文采用将摄像头固定并通过调整焦距来获得小目标、中目标和大目标3种类型的图像。图8是使用以上3种模型对部分图像处理的效果。从图中可以看出,对于中型和大型目标,3种模型都能检测到大部分的目标。对于大型目标,3种模型的准确率均达到92%以上。但是对于极小的目标,SSD模型效果要比IFSSD模型和FSSD模型差,SSD模型容易漏检大部分小目标柚子。这是由于SSD模型虽然从不同层次的特征进行预测,ConvNets提取的特征随着层次的增加语义越来越强,但是SSD模型却把它们当成一样的层次去预测,不能充分地利用局部细节特征和全局语义特征。FSSD模型采用FPN方法把细节特征(定位)和全局语义(识别)结合起来,即把浅层的细节特征和高层的语义特征结合起来,因此IFSSD、FSSD模型在小目标检测方面得到了进一步的提升。

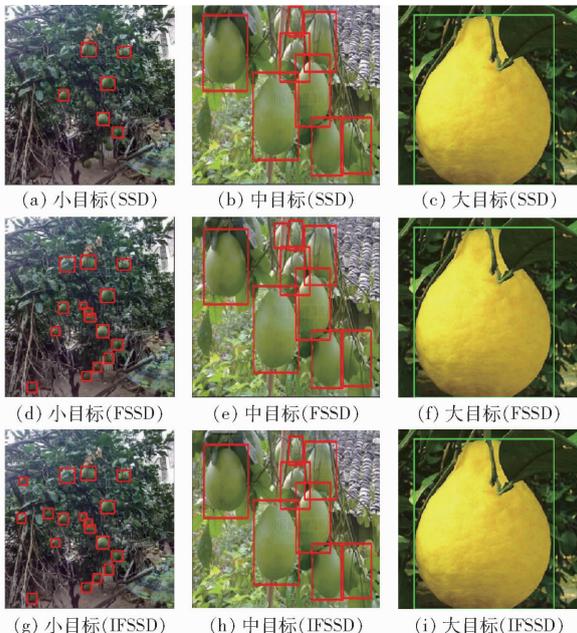


图8 3种模型对图像的处理效果

Fig. 8 Processing effect of three models on images

进一步通过试验来分析IFSSD模型在数据不平衡下对柚子检测效果的改进。

如图9所示,对于成熟期的柚子而言,IFSSD、FSSD、SSD模型都具有较精确的检测效果。这是由

于成熟期的柚子是黄色的,相对于背景和果实膨大期的柚子有着明显的颜色区别。但是SSD模型和FSSD模型会将某些叶片检测为果实膨大期的柚子,这是因为叶子与果实膨大期的柚子的颜色相似,同时在数据不平衡的情况下,以叶子为主的背景样本明显要多于果实膨大期的柚子样本,这会造成模型对于果实膨大期柚子的识别效果较差,同时在光线等自然条件的影响下,容易将成簇的叶子错认为是处于膨大期的果实。

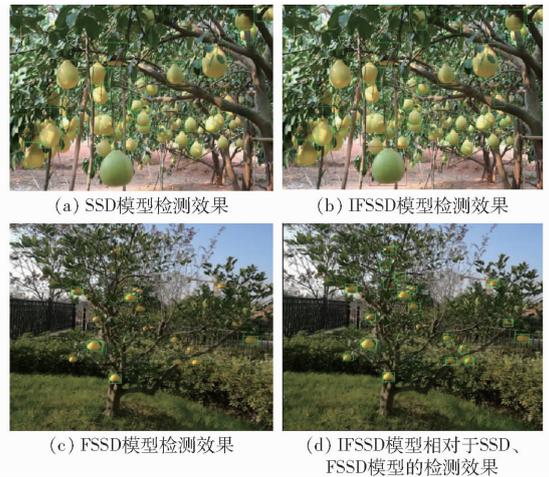


图9 柚子识别检测结果示例

Fig. 9 Examples of grapefruit identification results

图10为SSD、FSSD模型将绿色叶子错误检测为膨大期柚子的情况,而IFSSD模型能够极大地避免此类情况的发生,这表明Focal Loss损失函数的引入能够改善数据不平衡带来的影响,从而提高柚子的检测精度。

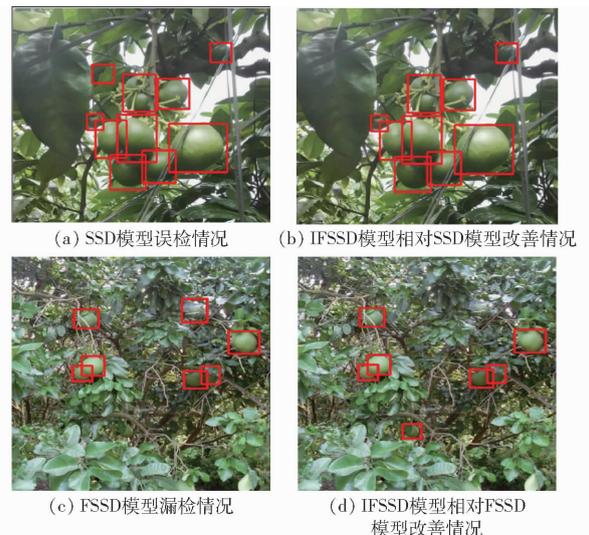


图10 膨大期柚子漏检示例

Fig. 10 Examples of expanded leaves misidentified as grapefruit

3 结束语

本文设计了一种改进的特征融合单镜头检测器

(IFSSD),提出了基于IFSSD卷积神经网络的柚子检测模型,并对所提模型进行检测试验。结果表明,本文模型对柚子检测的准确率达到93.7%,每幅图像的检测时间为29 s。

参 考 文 献

- [1] 马超侠,柯佑鹏. 世界柚子生产及贸易的发展趋势分析[J]. 热带农业科技, 2008, 31(4):32-36.
MA Chaoxia, KE Youpeng. Analysis on the development trend of world grapefruit production and trade [J]. Tropical Agricultural Science and Technology, 2008, 31(4): 32-36. (in Chinese)
- [2] 崔玉洁,张祖立,白晓虎. 采摘机器人的研究进展与现状分析[J]. 农机化研究, 2007,29(2):4-7.
CUI Yujie, ZHANG Zuli, BAI Xiaohu. Research progress and present situation of picking robots[J]. Journal of Agricultural Mechanization Research, 2007,29(2): 4-7. (in Chinese)
- [3] 王丹丹,宋怀波,何东健. 苹果采摘机器人视觉系统研究进展[J]. 农业工程学报, 2017,33(10):67-77.
WANG Dandan, SONG Huaibo, HE Dongjian. Research progress of vision system for apple picking robot[J]. Transactions of the CSAE, 2017,33(10): 67-77. (in Chinese)
- [4] 郭艾侠,邹湘军,朱梦思,等. 基于探索性分析的荔枝果及结果母枝颜色特征分析与识别[J]. 农业工程学报, 2013, 29(4):191-198.
GUO Aixia, ZOU Xiangjun, ZHU Mengsi, et al. Analysis and recognition of color characteristics of litchi fruits and fruiting branches based on exploratory analysis[J]. Transactions of the CSAE, 2013, 29(4):191-198. (in Chinese)
- [5] ZHAO C. Immature green citrus detection based on colour feature and sum of absolute transformed difference (SATD) using colour images in the citrus grove[M]. Elsevier Science Publishers B. V, 2016.
- [6] LU Q, CAI J R, LIU B, et al. Identification of fruit and branch in natural scenes for citrus harvesting robot using machine vision and support vector machine[J]. International Journal of Agricultural & Biological Engineering, 2014, 7(2):115-121.
- [7] 郭祥云,台海江. 深度学习在大田种植中的应用及展望[J]. 中国农业大学学报, 2019, 24(1):125-135.
GUO Xiangyun, TAI Haijiang. Application and prospect of deep learning in field planting[J]. Journal of China Agricultural University, 2019, 24(1): 125-135. (in Chinese)
- [8] 岳学军,凌康杰,洪添胜,等. 基于高光谱图像的龙眼叶片叶绿素含量分布模型[J/OL]. 农业机械学报, 2018,49(8): 18-25.
YUE Xuejun, LING Kangjie, HONG Tiansheng, et al. Chlorophyll content distribution model of Longan leaf based on hyperspectral image[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2018,49(8): 18-25. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20180802&flag=1. DOI:10.6041/j.issn.1000-1298.2018.08.002. (in Chinese)
- [9] 毕松,高峰,陈俊文,等. 基于深度卷积神经网络的柑橘目标识别方法[J/OL]. 农业机械学报, 2019,50(5):181-186.
BI Song, GAO Feng, CHEN Junwen, et al. Citrus target recognition method based on deep convolutional neural network [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019,50(5): 181-186. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20190521&flag=1. DOI:10.6041/j.issn.1000-1298.2019.05.021. (in Chinese)
- [10] FERENTINOS K P. Deep learning models for plant disease detection and diagnosis [J]. Computers & Electronics in Agriculture, 2018, 145:311-318.
- [11] 彭红星,黄博,邵园园,等. 自然环境下多类水果采摘目标识别的通用改进SSD模型[J]. 农业工程学报, 2018, 34(1):155-162.
PENG Hongxing, HUANG Bo, SHAO Yuanyuan, et al. General improved SSD model for multi-category fruit picking target recognition in natural environment[J]. Transactions of the CSAE, 2018, 34(1): 155-162. (in Chinese)
- [12] CHANG J R, CHEN Y S. Batch-normalized maxout network in network[J]. Computer Science. arxiv:1511.02583v1.
- [13] SZEGEDY C, IOFFE S, VANHOUCKE V. Inception-v4, inception-resnet and the impact of residual connections on learning [C]//AAAI Conference on Artificial Intelligence, 2016.
- [14] 林宇,陈晓勇. 基于inception v3模型的道路交通标志识别研究[J]. 江西科学, 2018, 36(5):147-150.
LIN Yu, CHEN Xiaoyong. Study on the identification of road traffic signs based on the inception v3 model [J]. Jiangxi Science, 2018, 36(5):147-150. (in Chinese)
- [15] 潘旭冉,杨帆,潘国峰. 采用改进全卷积网络的“高分一号”影像居民地提取[J]. 电讯技术, 2018, 58(2):5-11.
PAN Xuran, YANG Fan, PAN Guofeng. Resident site extraction of “High Score One” image using improved full convolution network [J]. Telecommunication Engineering, 2018, 58(2): 5-11. (in Chinese)
- [16] 伍伟明. 基于Faster R-CNN的目标检测算法的研究[D]. 广州:华南理工大学, 2018.
WU Weiming. Research based on Faster R-CNN's target detection algorithm [D]. Guangzhou: South China University of Technology, 2018. (in Chinese)
- [17] 王丹丹,何东健. 基于R-FCN深度卷积神经网络的机器人疏果前苹果目标的识别[J]. 农业工程学报, 2019, 35(3):164-171.
WANG Dandan, HE Dongjian. Identification of apple targets before the robot dredging the rout of the R-FCN deep convolutional neural network [J]. Transactions of the CSAE, 2019, 35(3):164-171. (in Chinese)
- [18] 何东健,刘建敏,熊虹婷,等. 基于改进YOLOv3模型的挤奶奶牛个体识别方法[J/OL]. 农业机械学报, 2020,51(4):250-260.
HE Dongjian, LIU Jianmin, XIONG Hongting, et al. Individual identification of dairy cows based on improved YOLOv3 [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020,51(4): 250-260. http://www.j-csam.org/jcsam/ch/reader.view_abstract.aspx?file_no=20200429&flag=1. DOI:10.6041/j.issn.1000-1298.2020.04.029. (in Chinese)
- [19] 薛月菊,黄宁,涂淑琴,等. 未成熟芒果的改进YOLOv2识别方法[J]. 农业工程学报, 2018, 34(7):173-179.
XUE Yueju, HUANG Ning, TU Shuqin, et al. Improved YOLOv2 identification method for immature mangoes [J]. Transactions of the CASE, 2018, 34(7):173-179. (in Chinese)
- [20] KHOA P, MUHAMAD K M A, ATSUO Y. Segmentation-driven hierarchical RetinaNet for detecting protozoa in micrograph [J]. International Journal of Semantic Computing, 2019, 13(3):393-413.

- [24] 李永磊, 宋建农, 董向前, 等. 基于土槽试验台的旋转耕作部件试验装置设计[J]. 农业工程学报, 2012, 28(17): 38-43.
LI Yonglei, SONG Jiannong, DONG Xiangqian, et al. Design of test device for rotary tiller components based on soil bin [J]. Transactions of the CSAE, 2012, 28(17): 38-43. (in Chinese)
- [25] 颜华, 吴俭敏, 林金天. 环形土槽微耕机试验平台设计[J]. 农业机械学报, 2010, 41(增刊): 68-72.
YAN Hua, WU Jianmin, LIN Jintian. Design of micro-cultivator testing platform with annular soil bin [J]. Transactions of the Chinese Society for Agricultural Machinery, 2010, 41(Supp.): 68-72. (in Chinese)
- [26] 祝英豪, 张居敏, 曾荣, 等. 人字型水旱两用旋埋刀辊设计与试验[J/OL]. 农业机械学报, 2019, 50(4): 49-57, 273.
ZHU Yinghao, ZHANG Jumin, ZENG Rong, et al. Design and experiment of herringbone type rotary blade roller for burying stubble in paddy field and dry land [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(4): 49-57, 273. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190406&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2019.04.006. (in Chinese)
- [27] 张居敏, 夏俊芳, 贺小伟, 等. 旋耕埋草机立刀设计功能实现程度的分析[J]. 华中农业大学学报, 2014, 33(3): 124-127.
ZHANG Jumin, XIA Junfang, HE Xiaowei, et al. Analysis on design functions realization degree of vertical blade for stubble burying rotary tiller [J]. Journal of Huazhong Agricultural University, 2014, 33(3): 124-127. (in Chinese)
- [28] 张居敏, 周勇, 夏俊芳, 等. 旋耕埋草机螺旋横刀的数学建模与参数分析[J]. 农业工程学报, 2013, 29(1): 18-25.
ZHANG Jumin, ZHOU Yong, XIA Junfang, et al. Mathematical modeling and analysis of helical blade for stubble burying rotary tiller [J]. Transactions of the CSAE, 2013, 29(1): 18-25. (in Chinese)
- [29] 苗中华, 陆鸣超, 胡晓东, 等. 基于虚拟仪器技术的采棉机智能监控系统开发与应用[J]. 农业工程学报, 2014, 30(23): 35-42.
MIAO Zhonghua, LU Mingchao, HU Xiaodong, et al. Development and application of intelligent monitoring and controlling system of cotton-picking machine based on virtual instrument technology [J]. Transactions of the CSAE, 2014, 30(23): 35-42. (in Chinese)
- [30] 乔晓东, 王晓燕, 颜华, 等. 基于虚拟仪器的后悬挂农具田间测试系统[J/OL]. 农业机械学报, 2013, 44(10): 98-103.
QIAO Xiaodong, WANG Xiaoyan, YAN Hua, et al. Field test system of rear suspension tools based on virtual instrument [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(10): 98-103. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20131017&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2013.10.017. (in Chinese)
- [31] 杨俊峰. 基于虚拟仪器的动态扭矩实验测试系统的研究[D]. 秦皇岛: 燕山大学, 2007.
- [32] 汲文峰, 贾红雷, 韩伟峰, 等. 通用刀片功率消耗影响因素分析与田间试验[J]. 农业机械学报, 2010, 41(2): 35-41.
JI Wenfeng, JIA Honglei, HAN Weifeng, et al. Analysis of influencing factors on power consumption and field test of universal blade [J]. Transactions of the Chinese Society for Agricultural Machinery, 2010, 41(2): 35-41. (in Chinese)
- [33] 张秀梅, 张居敏, 夏俊芳, 等. 水旱两用秸秆还田耕整机关键部件设计与试验[J]. 农业工程学报, 2015, 31(11): 10-16.
ZHANG Xiumei, ZHANG Jumin, XIA Junfang, et al. Design and experiment on critical component of cultivator for straw returning in paddy field and dry land [J]. Transactions of the CSAE, 2015, 31(11): 10-16. (in Chinese)

(上接第 35 页)

- [21] TIAN Xiaojie, CHENG Yaoyu, CHANG Guoli. Detect and recognize hidden cracks in silicon chips based on deep learning SSD algorithm[J]. Machine Tool & Hydraulics, 2019, 47(1): 36-40.
- [22] 童基均, 常晓龙, 赵英杰, 等. 基于深度学习的运动目标实时识别与定位[J]. 计算机系统应用, 2018, 27(8): 32-38.
TONG Jijun, CHANG Xiaolong, ZHAO Yingjie, et al. Real-time recognition and localization of moving targets based on deep learning[J]. Journal of Computer Systems, 2018, 27(8): 32-38. (in Chinese)
- [23] LI Z, PENG C, YU G, et al. Detnet: a backbone network for object detection[J]. arXiv preprint arXiv:1804.06215, 2018.
- [24] 董镛刚. 特征金字塔网络在图像检测中的应用[J]. 科学技术创新, 2018(10): 85-86.
DONG Leigang. Application of feature pyramid network in image detection [J]. Science and Technology Innovation, 2018(10): 85-86. (in Chinese)
- [25] 刘晓楠, 王正平, 贺涛涛, 等. 基于深度学习的小目标检测研究综述[J]. 战术导弹技术, 2019(1): 100-107.
LIU Xiaonan, WANG Zhengping, HE Yuntao, et al. Review of research on small target detection based on deep learning [J]. Tactical Missile Technology, 2019(1): 100-107. (in Chinese)
- [26] SUN Haoze, CHANG Tianqing, WANG Quandong, et al. Image detection method for tank and armored targets based on hierarchical multi-scale convolution feature extraction[J]. Acta Armamentarii, 2017, 38(9): 1681-1691.
- [27] 李艳霞, 柴毅, 胡友强. 不平衡数据分类方法综述[J]. 控制与决策, 2019, 34(4): 673-688.
LI Yanxia, CHAI Yi, HU Youqiang. Overview of imbalanced data classification methods[J]. Control and Decision, 2019, 34(4): 673-688. (in Chinese)
- [28] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017(99): 2999-3007.
- [29] SHETTY S. Application of convolutional neural network for image classification on Pascal VOC Challenge 2012 dataset[J]. Computer Vision and Pattern Recognition, 2016, 34(1): 155-162.