

# 基于 SSD 深度网络的河道漂浮物检测技术研究

唐小敏, 舒远仲, 刘文祥, 刘金梅

(南昌航空大学 信息工程学院, 江西 南昌 330100)

**摘要:**针对无人机拍摄河流视频影像进行河流漂浮物的自动提取识别,提出了一种基于 SSD (single shot multibox detector) 深度网络的河流漂浮物识别方法,SSD 利用 COCO 数据集的预训练网络模型进行迁移训练。为得到漂浮物数据,利用传统图像处理技术对视频帧进行河流提取,河流提取准确率达 91.4%。以河流视频截取图像为样本并采用软数据增强技术对漂浮物图像进行一定扩充,利用多种特征提取网络的 SSD 和 Faster R-CNN 深度网络进行样本训练并比较结果。实验结果表明,基于 ResNet-101 的 SSD 和 Faster R-CNN 深度网络模型召回率为 61.67% 和 58.83%,  $F_1$  值为 71.29% 和 69.55%, 精度为 84.47% 和 85.05%。经实验数据对比分析,基于 ResNet-101 的 SSD 深度网络提高了河流漂浮物的精确检测。利用传统图像处理技术和深度学习方法相结合能够准确、高效地识别出河流漂浮物,为无人机边缘计算提供研究基础和参考。

**关键词:**河流漂浮物;图像处理;深度学习;SSD;目标识别

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 1673-629X(2020)09-0154-05

doi:10.3969/j.issn.1673-629X.2020.09.028

## Research on River Floating Object Detection Technology Based on SSD Deep Network

TANG Xiao-min, SHU Yuan-zhong, LIU Wen-xiang, LIU Jin-mei

(School of Information Engineering, Nanchang Hangkong University, Nanchang 330100, China)

**Abstract:** To automatically extract and identify river floats from river video by UAV, we propose a method for river float identification based on SSD (single shot multibox detector) depth network. The pre-trained network model of COCO dataset is used for migration training in SSD. In order to obtain floating object image data, river image extraction is performed on video frames using traditional image processing techniques. The river image extraction accuracy rate is 91.4%. Taking river video intercepted images as samples and using soft data enhancement technology to expand the floating object image to a certain extent, the SSD and Faster R-CNN deep network of various feature extraction networks are used for sample training and the results are compared. The experiment shows that the recall rate of ResNet-101 based SSD and Faster R-CNN deep network model are 61.67% and 58.83% respectively,  $F_1$  values are 71.29% and 69.55%, and the accuracy is 84.47% and 85.05%. Based on the comparative analysis of experimental data, the SSD depth network based on ResNet-101 is more suitable for the accurate detection of river floating objects. The combination of traditional image processing techniques and deep learning methods can identify floating objects in rivers accurately and efficiently, which provides a research basis and reference for UAV edge computing.

**Key words:** river floating; image processing; deep learning; SSD; target recognition

## 0 引言

随着 Hinton 领导的小组采用深度学习模型 AlexNet<sup>[1]</sup> 在 ImageNet 图像识别大赛中一举夺冠,人工智能技术在目标检测中得到了广泛使用,例如车辆目标检测<sup>[2]</sup>、人脸识别<sup>[3]</sup>、智能安防<sup>[4]</sup>、智能医院等。近年来,很多学者针对自然环境为河道漂浮物识别带

来的负面影响提出了自己的研究成果。

左建军等<sup>[5]</sup>采用背景减除分割出图像中漂浮物区域,再利用 BP 算法进行漂浮物特征训练并构建漂浮物分类器对分割区域进行漂浮物识别。祝典等<sup>[6]</sup>设计了一种水域漂浮物收集监测系统。许静波等<sup>[7]</sup>成功搭建了一套水面漂浮物检测及估算系统。雷李义等<sup>[8]</sup>

收稿日期:2019-11-13

修回日期:2020-03-16

基金项目:国家自然科学基金(71761028)

作者简介:唐小敏(1994-),女,硕士研究生,研究方向为软件工程、人工智能、图像处理;舒远仲,博士,教授,硕导,研究方向为计算机网络、信息安全、物联网技术、管理信息系统等。

提出了一套基于深度学习的水面漂浮物目标检测方案。朱贺等<sup>[9]</sup>结合使用灰度阈值分割和河道轮廓识别方法实现对河道区域的准确提取。利用河道轮廓形态进行河道分割可以减少河岸上垃圾对检测结果的影响,但是不能排除相似形态水体上的漂浮物对检测结果造成影响。

文中首先对河道进行分割提取,然后对已分割出的河道区域进行漂浮物检测。基于3种特征提取网络分别进行 Faster R-CNN<sup>[10]</sup> 网络模型和 SSD<sup>[11]</sup> 网络模型的训练,对比分析这三种特征提取网络的性能,选择性能最优的特征提取网络进行两种模型的对比,经过实验分析,选择 SSD 网络模型进行漂浮物检测效果最佳。

## 1 实验数据制作

实验所需数据是利用无人机对3条河流分别进行正射视频采集而得,其中河流总长5.6公里。为减小河岸垃圾对检测结果的影响,利用传统图像处理技术先进行河道检测。

对影像进行滤波处理提升河道整体灰度值并增强河道连通性,采用自适应阈值化操作将滤波图像二值化,使算法具有更好的鲁棒性;为使河道区域与非河道区域分离,先对影像进行形态学处理去除小的干扰块和加大河道的连通性,最后对图像进行轮廓提取。经过以上步骤,发现在河道旁出现水体或水域等客观条件下,河流检测不准确。为解决此问题,根据视频中相邻帧之间河道相似且河道在影像中间的特点,提出当检测区域中心线偏离图像中心线一定阈值时,取图像宽的0.1倍,使用上次河道中心点对影像进行纠正的方法最终获得河流区域图像准确率为91.4%。

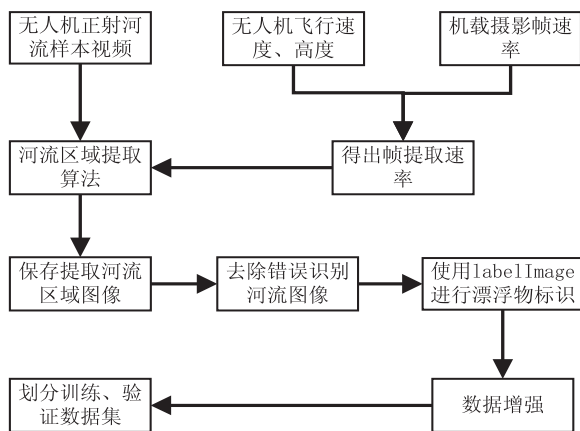


图1 河流漂浮物数据集制作技术路线图

对以上获取的河流区域数据进行漂浮物识别检测。由于没有漂浮物检测公共数据集,则需制作漂浮物数据集,其制作技术路线如图1所示。河流视频帧率为50 fps,无人机飞行速率保持在15 m/s,飞行高度

保持在55 m,对河流视频每隔90帧保存图片,共计7 800张。去除无人机起飞、降落、某些河流区域被遮挡以及错误识别等原因导致保存图片中不包含河流图像,最终可用于标注影像2 400张,利用 labelImage 对影像进行漂浮物标识并保存标注 xml 数据集,将标注完成图像按8:2进行训练集与验证集划分,数量分别为1 920和480张。

由于数据量过小,同时为了不增加数据标注时的工作量以及训练时对机器的内存消耗,对数据集使用软数据增强技术。该方法是一种在数据训练过程中对每批次的数据进行数据增强,再进行数据训练的方法。之后将对每批次训练数据进行随机翻转、随机颜色抖动和随机裁剪,使每次训练的数据不相同从而达到数据软扩增效果,该方法在一定程度上提高了模型泛化性和鲁棒性。网络训练图像大小为512\*512分辨率,并将影像R、G和B波段分别减去123.68、116.78和103.9进行归一化<sup>[12]</sup>,使模型训练更容易。

## 2 基于 SSD 的河流漂浮物检测方法

以河流影像漂浮物检测为研究目标,考虑到实时性,将使用单步检测 SSD 算法作为漂浮物检测核心算法。SSD 算法继承了 Faster RCNN 和 YOLO<sup>[13-14]</sup> 的不同优点,一方面,SSD 采用不同尺寸和长宽比的先验框(Faster RCNN 称为 Anchor),在一定程度上克服了目标检测尺寸多变问题,同时使算法训练简易化,另一方面,SSD 采用 YOLO 一步检测思想,其对特征图直接利用 CNN 进行回归预测,这与 YOLO 将特征图进行向量化再全连接后进行回归预测不同,此方法可以充分利用特征图的空间结构信息来确保预测分类准确性。

### 2.1 SSD 模型算法

SSD 深度网络由前端特征提取网络 and 后端多尺度特征检测网络组成,前端特征提取网络以 VGG16 卷积神经网络为例生成特征图,并与多尺度特征检测网络共享卷积特征。该特征提取网络包括13个3×3卷积层和5个最大池化层(Max pooling),每次卷积后均使用 ReLU 激活函数进行激活,每层组成形式均为卷积层(convolution)+激活函数(ReLU)+池化层(Max pooling)的组合方式,此沿用 VGG16 卷积网络,但该特征提取网络与 VGG16 网络略微有所差异,其将 VGG16 网络的全连接层 fc6 和 fc7 分别替换成扩张率为6的3×3空洞卷积层 conv6 和1×1的卷积层 conv7。为了不减小特征图维度,将原来池化层 pool5 的卷积核由2×2、步数2变为3×3、步数1,同时为了配合后端多尺度特征检测网络,其移除了 VGG16 的 dropout 层和 fc8 层。

后端多尺度特征检测网络共提取了 6 个网络特征图用于漂浮物的检测,第一个特征图取自 Conv4\_3 层,其大小为  $38 \times 38$ ,由于第一特征图靠前,为减小与后面检测层的差异,第一特征图后增加 L2 Normalization 层,其仅对每个像素点在通道维度上进行归一化,将归一化特征图与一个可训练缩放变量 gamma 相乘。后面 5 个特征图由于靠后无需做过多变换,其分别取自 Conv7 层、Conv8\_2 层、Conv9\_2 层、Conv10\_2 层、Conv11\_2 层,大小分别为  $19 \times 19$ 、 $10 \times 10$ 、 $5 \times 5$ 、 $3 \times 3$ 、 $1 \times 1$ 。

对于不同特征图设置先验框数目不同,先验框的设置主要体现在尺寸和长宽比上,先验框尺寸遵循线性递增规则,递增公式为:

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m - 1}(k - 1), k \in [1, m] \quad (1)$$

其中,  $m$  为特征图个数,文中为 5 (除第一特征图外,其单独处理),  $s_k$  表示先验框大小与图片大小的尺度比例,  $s_{\max}$  和  $s_{\min}$  分别表示该比例的最大值和最小值,文中取 0.9 和 0.2。第一特征图先验框尺度比例为  $s_{\min}/2$ ,对于后面 5 个特征图按照式(1)线性增加,然后将这些比例与图片大小相乘得各个尺度的特征图先验框。先验框长宽比一般为  $\alpha_i \in \{1, 2, 3, 1/2, 1/3, 1\}$ ,在实际中 Conv4\_3、Conv10\_2 和 Conv11\_2 特征图不使用长宽比为  $\{3, 1/3\}$  的先验框,因此总共得到 8 732 个先验边界框。

网络损失函数计算公式如下:

$$L(x, l, g) = \frac{1}{N} [L_{\text{conf}}(x, c) + \alpha L_{\text{loc}}(x, l, g)] \quad (2)$$

其中,  $N$  表示正样本数量,  $c$  表示类别置信度值,  $l$  表示先验框与对应边界框置信度值,  $g$  表示真实目标的位置,  $\alpha$  表示权重系数,一般设为 1。  $x_{ij}^p \in \{1, 0\}$  为指示参数,  $x_{ij}^p = 1$  表示第  $i$  个先验框与第  $j$  个真实目标匹配且类别为  $p$ 。

式(2)中位置误差  $L_{\text{loc}}$  定义如下:

$$L_{\text{loc}}(x, l, g) = \sum_{i \in \text{Pos}} \sum_{m \in \{cx, cy, w, h\}} x_{ij}^h \text{smooth}_{L1}(l_i^m - \hat{g}_j^m),$$

$$\hat{g}_j^{cx} = \frac{g_j^{cx} - d_i^{cx}}{d_i^w}, \hat{g}_j^{cy} = \frac{g_j^{cy} - d_i^{cy}}{d_i^h}, \quad (3)$$

$$\hat{g}_j^w = \log\left(\frac{g_j^w}{d_i^w}\right), \hat{g}_j^h = \log\left(\frac{g_j^h}{d_i^h}\right)$$

其中,  $d$  表示默认边界框,中心为  $(cx, cy)$ ,宽度为  $w$ ,高度为  $h$ 。

式(2)置信度误差  $L_{\text{conf}}$  定义如下:

$$L_{\text{conf}}(x, c) = - \sum_{i \in \text{Pos}} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in \text{Neg}} \log(\hat{c}_i^0)$$

$$\hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)} \quad (4)$$

## 2.2 目标检测度量标准

为评估模型对河流漂浮物识别的有效性,采用 recall (召回率)、precision (精确率) 以及引入  $F_1$  值<sup>[15]</sup> 为调和平均值对模型进行评估。假设模型正确识别漂浮物的数量为 TP (真正例),模型错将其他类识别为漂浮物的数量为 FP (假正例);模型识别非漂浮物的正确数量为 TN (真反例),模型错将漂浮物识别成其他类别数量为 FN (假反例)。

$P$  是在识别漂浮物中真正例所占的比例,则它表示为:

$$P = \frac{TP}{TP + FP} \quad (5)$$

$R$  表示正确识别出的漂浮物与存在的漂浮物数量的比值,其表示为:

$$R = \frac{TP}{TP + FN} \quad (6)$$

$F_1$  分数认为召回率和精确率同等重要,可表示为:

$$F_1 = \frac{2PR}{P + R} \quad (7)$$

## 3 结果与讨论

### 3.1 实验平台和参数设置

实验是在 Ubuntu16.04 LTS 操作系统,处理器是 Intel® CoreTM i7-6800k CPU @ 3.4 GHz×12 (64 GB 运行内存) 和显卡是 NVIDIA GeForce GTX 1080 Ti GPU (11 GB 运行显存) 的电脑设备上,深度学习框架为 Tensorflow1.8。

文中采用的是端到端识别,并使用 coco 的预训练模型进行参数迁移初始化设置。模型优化器使用 Momentum 算法,初始学习率为  $3e-4$ ,动量因子为 0.9,最大迭代步数为 100 000 步。

为加速网络迭代速度,当模型迭代达 40 000 步时,学习率调整为  $3e-5$ ,迭代到 80 000 步时,学习率进一步调整为  $3e-6$ 。训练结束后保存好训练模型,并使用测试河流视频对模型进一步测试,测试保留类别置信度大于 0.7 的河流漂浮物,最后在测试河流视频上保存输出结果。

### 3.2 训练结果分析

实验将在 Faster R-CNN 深度网络模型和 SSD 网络模型上,分别使用深度网络 ResNet-50、ResNet-101 和 Inception-v1 作为骨干网。随着训练过程中网络迭代次数的增加,网络损失都趋于稳定,随后分别比较了两种深度学习目标检测模型各自基于不同特征提取网络时整体损失的下降情况,具体对比情况如图 2 和图 3 所示。



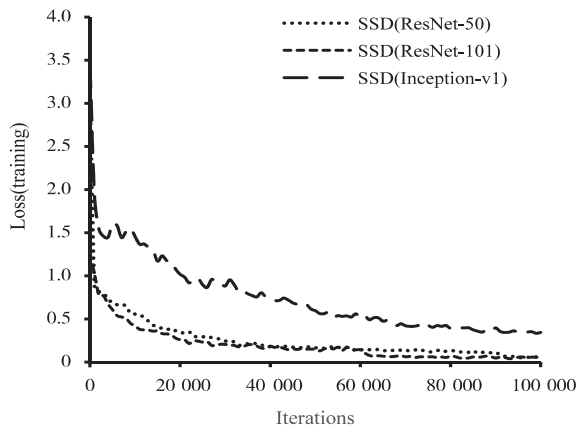


图2 SSD 模型整体损失对比

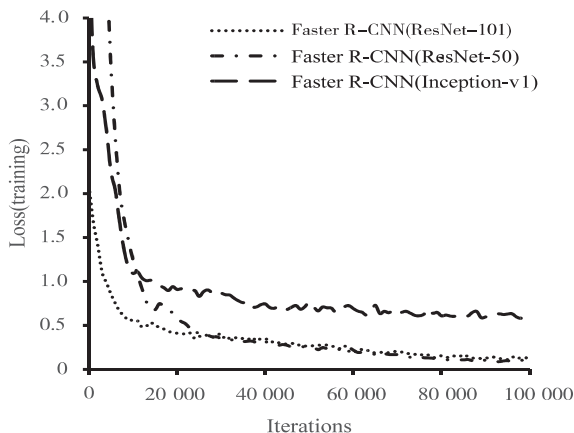


图3 Faster-RCNN 模型整体损失对比

图2显示了3种特征提取网络在SSD模型上的损失情况,显然基于ResNet-101特征提取网络的SSD模型损失下降最快且网络收敛最为稳定,图中可以明显看出随着特征提取网络深度增加,SSD模型的整体损失下降越大,模型收敛越稳定。图3显示了3种特征提取网络在Faster R-CNN模型上的损失情况,基于ResNet-101特征提取网络的Faster R-CNN收敛速度较快,明显优于其他2种模型,而基于ResNet-50和Inception-v1特征提取网络的Faster R-CNN模型都出现了不同程度的震荡,且基于Inception-v1特征提取网络的Faster R-CNN模型收敛情况相较于其他两种模型收敛程度较差。

为了选出优秀的深度网络模型与特征提取网络,根据各模型总体损失下降函数对比结果为2种模型都选择了ResNet-101作为模型特征提取网络并做了进一步对比。2种模型随迭代次数增加模型整体损失下降情况对比如图4所示。图4表明以ResNet-101为特征提取网络的SSD模型,随着迭代次数增加整体损失要优于基于ResNet-101特征提取网络的Faster R-CNN模型,虽然Faster R-CNN最后趋于稳定,但SSD模型最终损失趋于更小的数值,这表明基于ResNet-101特征提取网络的SSD模型能够更好地拟合河流漂浮物数据。

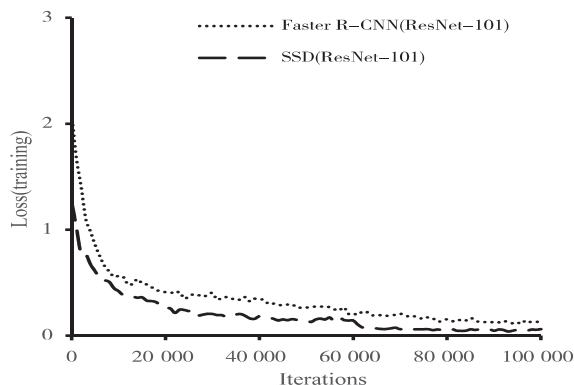


图4 SSD 与 Faster R-CNN 模型整体损失对比

根据模型评价指标,文中对以上不同特征提取网络的SSD和Faster R-CNN模型针对精确率和召回率两项指标以及 $F_1$ 值和单张图像的检测时间做出了详细对比,对比结果如表1所示。表1统计结果为训练好的不同特征提取网络的SSD模型和Faster R-CNN模型分别对河流漂浮物验证集进行漂浮物识别并计算各指标值。由表1对比可得,随着特征提取网络深度增加,基于2种模型的河流漂浮物识别精确率不断增加,模型性能不断提升。通过对比可知,基于ResNet-101特征提取网络的SSD和Faster R-CNN深度网络模型都表现出优良的性能,Faster R-CNN深度网络模型河流漂浮物识别精确率达85.05%,SSD深度网络模型河流漂浮物识别精确率达84.47%略低于前者,但SSD模型整体性能上明显优于Faster R-CNN模型,SSD模型召回率达到61.67%,高出Faster R-CNN模型2.84%,且SSD模型 $F_1$ 值达到71.29%,高于Faster R-CNN模型 $F_1$ 值69.55%,同时SSD模型在单张图像的检测时间上表现出明显优势。综上,基于ResNet-101特征提取网络的SSD深度网络模型在漂浮物识别上更具优势,符合文中对视频中河流漂浮识别的目标。

表1 不同特征提取网络的SSD和Faster R-CNN深度网络模型对比

Model	Extraction network	Precision /%	Recall /%	$F_1$ /%	Time/ms
SSD	Inception-v1	74.09	56.14	63.88	68
SSD	ResNet-50	79.57	49.58	61.10	95
SSD	ResNet-101	84.47	61.67	71.29	112
Faster R-CNN	Inception-v1	78.48	50.02	61.10	75
Faster R-CNN	ResNet-50	81.17	46.03	58.75	109
Faster R-CNN	ResNet-101	85.05	58.83	69.55	177

### 3.3 模型识别结果

根据训练结果分析,实验使用基于ResNet-101特征提取网络的SSD深度网络模型,进行漂浮物多尺度预测及位置回归,此方法有效提高了网络对不同尺度

漂浮物的识别精度,具体漂浮物识别如图 5 所示。

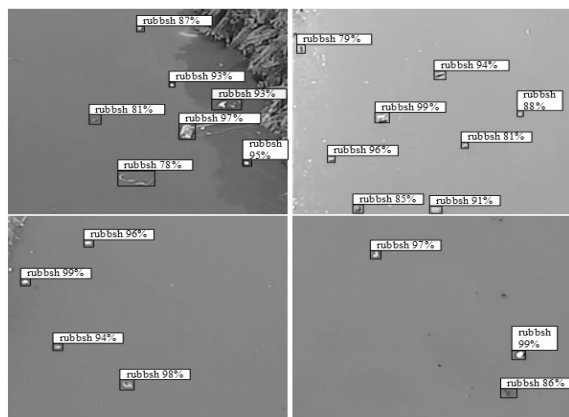


图 5 河流漂浮物识别结果

### 3.4 讨论

根据上述网络模型分析可知, Faster R-CNN 网络模型在河流漂浮物识别精度上相较于 SSD 深度网络模型提高了 0.58%, 但 Faster R-CNN 召回率减少了 2.84%,  $F_1$  值减少了 1.74%。导致以上差距的原因如下: 第一, Faster R-CNN 深度网络模型将特征提取网络提取的特征经过区域建议网络, 生成  $k$  个区域建议框后送入 RoI 子网以准确提取漂浮物的感兴趣区域, 此过程与 SSD 深度网络模型采用将整个检测网络作为单个神经网络进行漂浮物识别相比非常耗时, 所以 Faster R-CNN 模型检测速度较慢。第二, 由于无人机对河流进行拍摄时飞行高度较高, 这将导致河流漂浮物可视尺度非常小, 而 SSD 深度网络模型采用多尺度特征提取, 分辨率较大的特征图将被用于漂浮物尺度较小的目标识别, 与 Faster R-CNN 只使用单一尺度特征图提取特征进行目标识别相比, Faster R-CNN 很难将所有较小漂浮物识别出, 进而导致 Faster R-CNN 整体性能低于 SSD 深度网络模型。

## 4 结束语

提出了一种基于 SSD 深度学习网络的漂浮物检测研究方法。首先根据河道水面反光及水体颜色相近等特性, 采用 OpenCV 图像处理技术对河流区域进行分割提取。然后在此基础上进行漂浮物检测研究, 漂浮物识别算法在 Tensorflow 框架的基础上, 搭建了基于 ResNet-101 特征提取网络的深度神经网络 SSD 模型, 通过利用 COCO 数据集预训练模型进行网络迁移训练。同时介绍了从原始视频数据到漂浮物数据集的制作全过程, 并对数据集进行软数据增强, 有效缓解了数据量不足的问题。实验分析表明, 随着特征提取网络的加深, 两种深度网络模型在性能上都有不同程度的提升, 而基于 ResNet-101 特征提取网络的 SSD 模型相较于其他三种特征提取网络具有明显优势, 并能

够高效地对漂浮物进行识别, 这体现出深度学习方法在目标识别领域的强大之处。目前, 该算法已应用于无人机机载端的河流巡查中, 通过无人机实时拍摄河流视频, 已做到对实时视频进行边缘计算分析河流中漂浮物数目, 然后实时发送至服务器进行河流污染度评分。

### 参考文献:

- [1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [2] 金宇尘, 罗娜. 结合多尺度特征的改进 YOLOv2 车辆实时检测算法 [J]. 计算机工程与设计, 2019, 40(5): 1457-1463.
- [3] 王智飞. 低分辨率人脸识别算法研究 [D]. 北京: 北京交通大学, 2013.
- [4] 杨哲. 人工智能在安防领域的应用分析 [J]. 中国新通信, 2019, 21(16): 156.
- [5] 左建军, 吴有富. 水面漂浮物智能监控技术 [J]. 软件导刊, 2013, 12(4): 150-152.
- [6] 祝典, 廖敏, 刘新顺, 等. 一种新型水域漂浮物收集监测系统 [J]. 现代电子技术, 2019, 42(5): 27-31.
- [7] 许静波. 水面漂浮物监测及估算系统开发与建设 [J]. 江苏水利, 2018(8): 51-58.
- [8] 雷李义, 艾娇燕, 彭婧, 等. 基于深度学习的水面漂浮物目标检测评估 [J]. 环境与发展, 2019, 31(6): 117-120.
- [9] 朱贺, 李臣明, 张丽丽, 等. 联合灰度阈值分割及轮廓形态识别的河道提取 [J]. 电子测量与仪器学报, 2014(11): 1288-1296.
- [10] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]//Advances in neural information processing systems. Montreal: MIT Press, 2015: 91-99.
- [11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//Proceedings of the 2016 European conference on computer vision. Berlin: Springer, 2016: 21-37.
- [12] SIMONYAN K, ZISSERMAN A. Very deep convolutional net-works for large-scale image recognition [J]. Computer Science, 2014, 22(9): 1-14.
- [13] REDMON J, FARHADI A. Yolo9000: better, faster, stronger [C]//Computer vision and pattern recognition. Hawaii: IEEE, 2017: 7263-7271.
- [14] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE, 2016.
- [15] 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): 156-163.