

基于单目视觉的水面目标识别与测距方法研究

苏萍¹, 朱晓辉^{1,2}

(1. 南通大学信息科学技术学院, 江苏南通 226019;

2. 南通先进通信技术研究院, 江苏南通 226019)

摘要:无人船在自主航行过程中,如何探测障碍物种类与距离是值得重点研究的问题。现阶段的无人船主要采用激光雷达、超声波等设备探测水面障碍物,但此类方法并不能识别出障碍物的具体类型。提出采用基于单目视觉的目标检测与测距算法,不仅能够检测出无人船前方是否有障碍物,还能够对水面障碍物进行识别和距离检测。首先通过训练好的Mask R-CNN模型对水面障碍物进行检测识别,随后利用改进的基于小孔成像原理的单目测距方法对识别到的障碍物测量距离。为验证该目标识别与测距方法的有效性,选取了不同形状的船只、河岸、荷叶及荷花等图像进行验证实验。实验结果表明,课题研究的目标识别与测距方法能较为准确地对水面障碍物进行分类及像素级别的实例分割,同时测量障碍物与无人船之间的距离。

关键词:单目视觉;无人船;障碍物识别;Mask R-CNN;单目测距

中图分类号:TP391.41

文献标识码:A

文章编号:1673-629X(2021)02-0080-05

doi:10.3969/j.issn.1673-629X.2021.02.015

Research on Water Surface Target Recognition and Ranging Based on Monocular Vision

SU Ping¹, ZHU Xiao-hui^{1,2}

(1. School of Information Science and Technology, Nantong University, Nantong 226019, China;

2. Nantong Research Institute for Advanced Communication Technologies, Nantong 226019, China)

Abstract: In the process of autonomous navigation, how to detect the type and distance of obstacles is worth studying. At present, unmanned ships mainly use lidar, ultrasonic and other equipment to detect obstacles on water surface, but such methods cannot identify the specific types of obstacles. The target detection and ranging algorithm based on monocular vision is proposed, which can not only detect whether there are obstacles in front of the unmanned ship, but also identify the obstacles on the water surface and detect the distance. Firstly, the trained Mask R-CNN model is used to detect and identify the obstacles on the water surface, and then the improved monocular ranging method based on keyhole imaging principle is used to measure the distance of the identified obstacles. In order to verify the effectiveness of the target recognition and ranging method, the images of water surface obstacles such as ships, riverbanks and lotus with different shapes are selected for verification experiments. The experiment shows that the target recognition and ranging method studied can classify the water surface obstacles and segment the examples at the pixel level more accurately, and measure the distance between the obstacles and the unmanned ship at the same time.

Key words: monocular vision; unmanned ship; obstacle recognition; Mask R-CNN; monocular distance detection

0 引言

21世纪以来,随着无人机、无人车技术发展成熟,无人船的应用也越来越广泛。与其他无人设备一样,无人船能够自主规划和自主航行,通过各类设备及传感器完成一系列环境感知、目标检测、自动避障等任务,主要应用于水质监测、环境监控及安全监控等^[1]方

面。在无人船自主航行过程中,对障碍物进行检测及避障是其亟需解决的一个问题,现有的微型无人船主要采用超声波、激光雷达及毫米波雷达等设备进行避障。此类设备虽然能较为精确地探测微型无人船与前方障碍物的距离,但其难以识别障碍物具体种类同时难以检测水草、垃圾等水面漂浮物,造成无人船因躲避

收稿日期:2020-01-07

修回日期:2020-05-10

基金项目:西交利物浦大学重点项目建设专项(KSF-P-02);苏州市自然科学基金(SYG201837);南通市自然科学基金(JC2018075);南通大学-智能信息技术联合研究中心开放课题(KFKT2017A06)。

作者简介:苏萍(1995-),女,硕士研究生,研究方向为计算机图像处理;朱晓辉,博士,副教授,研究方向为计算机软件与理论。

不及时而出现被水草或垃圾等缠绕的现象,影响无人船自主航行能力。因此,该文提出了基于计算机视觉的方法实现障碍物识别与测距,利用 Mask R-CNN 深度学习网络对水面障碍物进行识别定位及实例分割,采用基于小孔成像原理的单目测距算法,实现水面障碍物到无人船的距离测量。

传统的目标检测算法主要通过多尺度的窗口在目标图像上滑动产生候选区域,对候选区域内的图像计算其手工设计的特征,采用依据此特征训练好的分类器(例如 SVM、Adaboost 等)对所有的候选区域进行分类,取概率大于预设阈值的候选区域经 NMS(非极大值抑制)后的结果作为最终输出。常见的手工设计特征包括梯度方向特征图(HOG)^[2]、可变性组件模型(DPM)^[2]、尺度不变特征变换(SIFT)^[3]等。此类基于先验知识手工设计的特征,虽然在特定场景下拥有较高精度的目标检测结果,但同时由于其对先验知识具有依赖性,因此它们在自适应和泛化能力上表现不足,同时由于现有的分类器主要为二分类,因此在面对多分类问题时需要针对每个类别训练一个分类器,实用性不强。随着深度学习的发展,基于神经网络的目标检测方法能够将训练好的模型应用到不同的场景,并且能够自适应地提取目标的不同层级特征。与传统的目标检测方法相比,检测精度和泛化性都得到了有效提高。根据分类回归和区域提取是否分开将基于深度学习的目标检测模型分为两类:(1)基于分类回归的目标检测模型。此类模型先根据特征图划定选框,对目标进行分类,经典的方法有:SSD^[4]、YOLO^[5]、YOLOv3^[6]。这些方法利用回归思想提取目标的边界回归框,大大提升了检测速度,但是检测精度较差^[7];(2)基于候选区域的目标检测模型。此类模型对特征图进行边框提取,将其输出与特征图一同输入到感兴趣区域(region of interest)池化层,实现目标分类与定位,此类目标检测方法是近年来研究目标检测的热门方法。R-CNN^[8]是第一个将深度学习引入到目标检测领域并能够实现自适应检测的方法,随后很多研究人员对它进行改进,如:SPP-Net^[9]网络,它将金字塔池化层(spatial pyramid pooling)引入到 R-CNN 中,减少了输入图像的尺寸对网络的影响,同时提升了检测的精度;Fast R-CNN^[10]在 SPP-Net 的空间金字塔池化层基础上使用单尺度池化,在保证检测精度的同时进一步提高检测速度;Faster R-CNN^[11]将区域建议网络(region proposal networks)引入到 Fast R-CNN 的提取候选区域过程,实现端到端的训练,提升了网络训练的速度和提取区域的准确度,是目前应用较广的检测框架^[12];Mask R-CNN^[13]在 Faster R-CNN 的基础上增加一个预测掩码分支,将 Faster R-CNN 中的 RoI 池

化层改进为 RoIAlign,采用双线性插值法来降低边界回归框的位置误差,进一步提升目标检测识别的准确度。由文献[13]的研究表明,Mask R-CNN 识别精度比 Faster R-CNN 更高,因此该文采用 Mask R-CNN 算法对水面障碍物进行检测识别。

基于机器视觉的测距主要有双目视觉测距、结构光学视觉测距^[14]以及单目视觉测距^[15]。结构光学视觉测距易受光源影响,因此其应用场景有较多限制;双目视觉测距的研究重点主要为特征匹配^[16],其测距的速度和精度受特征点匹配的精度影响较大。相较于上述两种算法,单目视觉测距则具有运算速度快、结构简单等优点,因此其具有广泛的应用前景。现有基于单目视觉的测距算法主要包含基于小孔成像原理测距、基于序列图像测距以及基于单帧静态图像测距。基于序列图像测距的算法通过计算两帧之间目标与相机距离的差值,而不是直接计算当前帧目标与相机的距离,因此其计算误差既取决于初始计算的目标与相机的距离,同时取决于两帧之间的差值,测量结果与实际距离误差较大;基于单帧静态图像测距虽然实时性很高,但在远距离测距时,其精度较低;而基于小孔成像原理的单目测距模型具有结构简单、测距精度较高的优点,因此该文采用基于小孔成像原理的单目测距模型。

1 Mask R-CNN

Mask R-CNN 是在 Faster R-CNN 的基础上改进得到的,主要由 Faster R-CNN、RoIAlign、FCN 三个模块组成^[17],其网络结构如图 1^[13]所示。

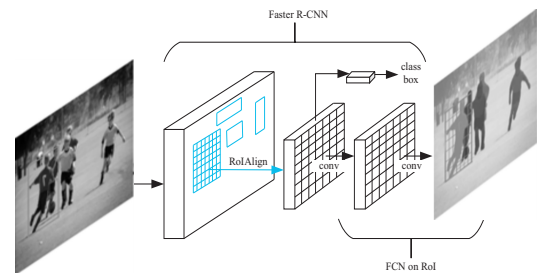


图 1 Mask R-CNN 网络结构

1.1 RoIAlign

在 Faster R-CNN 的检测框架中,使用 RoIPooling 根据预选框的坐标在特征图上对应的位置池化为固定大小的特征图,然后进行分类和边框回归。预选框位置是由模型回归得到的,一般是浮点数,且池化后的特征图尺寸要求是固定的,所以 RoIPooling 要进行两次量化过程,一是预选框位置量化成整数点的坐标值;二是将量化后的边界区域平均分割成 $n \times n$ 个单元(bin)并对每个单元的边界进行量化。由于两次量化后,得到的预选框位置与最初回归得到预选框的位置存在一定偏差,因此影响了检测框架的分割精度。

为解决 RoIPooling 检测框架分割精度不高的问题,Mask R-CNN 提出了新的检测框架方法—RoIAlign。RoIAlign 取消了 RoIPooling 中的预选框边界量化操作,使得预选框与原图目标区域的位置更好地对齐。其具体操作为:遍历每一个候选区域,保持浮点数的预选框边界不变,并将每一个候选区域 $n \times n$ 分割成个单元,再将每个单元平均分成 m 份,并采用双线性插值法计算每份中心点位置的像素值,取这些像素值中最大的像素值作为每个单元的像素值,最终由这些最大像素值组成 $n \times n$ 大小的特征图。

RoIPooling 与 RoIAlign 在反向传播过程也存在一定的差异,RoIPooling 的反向传播公式如式(1)^[17]所示,而 RoIAlign 的反向传播公式如式(2)^[17]所示。

$$\frac{\partial L}{\partial x_i} = \sum_r \sum_j [i = i^*(r, j)] \frac{\partial L}{\partial y_{rj}} \quad (1)$$

其中, x_i 表示池化前特征图上的像素点, y_{rj} 表示池化后的第 r 个候选区域的第 j 个点, $i^*(r, j)$ 表示 y_{rj} 像素值的来源。

$$\frac{\partial L}{\partial x_i} = \sum_r \sum_j [d(i, i^*(r, j)) < 1] (1 - \Delta h) (1 - \Delta w) \frac{\partial L}{\partial y_{rj}} \quad (2)$$

其中, $i^*(r, j)$ 表示一个浮点数的坐标位置, $d(i, i^*(r, j))$ 表示两点之间的距离, Δh 和 Δw 表示 x_i 与 $i^*(r, j)$ 横纵坐标的差值,然后作为双线性内插的系数乘在原始的梯度上。

1.2 FCN 网络结构

全卷积网络 (FCN) 的英文全称是 fully convolutional networks,它对图像进行像素级分类,以解决语义级图像分割 (semantic segmentation) 问题。在进行图片分类时,FCN 可以输入任意尺寸的图像,使用反卷积层对最后一个卷积层的特征图进行上采样,不断扩大特征图的尺寸,直至它恢复到与输入图像相同的尺寸,继而对每个像素都生成一个预测,同时保留输入图像的空间信息,最终逐个像素计算 softmax 分类损失,对上采样的特征图中每一个像素值进行分类,从而提高图像分类准确率^[18]。

Mask R-CNN 在 Faster R-CNN 为每个候选对象的感兴趣区域提供类标签和边框偏移量两个输出的基础上,新增了一个分割 mask 的分支。mask 分支是一个小的 FCN,应用在每一个感兴趣区域,它在每一个感兴趣区域上预测出一个大小为 $m \times m$ 的 mask,使得 mask 分支中的每层都能够明确地保持 $m \times m$ 的空间布局,避免将其用缺少空间维度的向量表示。

2 测距

摄像头采集图片时,是将三维物体投影到摄像头

的二维平面上。宽度为 W 的物体,经过针孔平面投影到摄像头成像平面之后的影像宽度为 W_1 ,针孔平面与成像平面的距离即摄像头焦距为 F ,物体平面与针孔平面的距离设为 D ,那么利用三角形相似原理,根据式(3)就可以得到物体到摄像头的距离。

$$D = \frac{W \times F}{W_1} \quad (3)$$

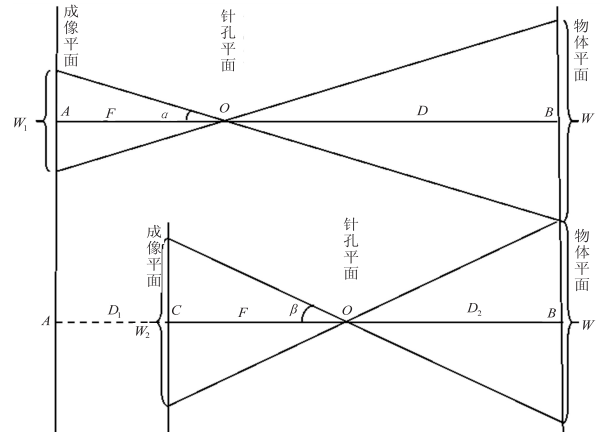


图2 摄像头成像原理

测距实验原理如图2所示。摄像头在 A 点给目标物体拍摄一张图片,假设物体宽度为 W ,目标物体在摄像头成像平面上的投影宽度为 W_1 ,摄像头的焦距为 F ,目标物体与摄像头镜头的距离为 D ,根据相似三角形原理,有式(4):

$$\frac{W_1}{F} = \frac{W}{D} \quad (4)$$

由式(4)可以推算出物体宽度的计算公式,即式(5):

$$W = \frac{W_1 \times D}{F} \quad (5)$$

摄像头从 A 点向前移动到 C 点,移动距离为 D_1 ,目标物体在原地不动,目标物体的宽度依然是 W ,摄像头的焦距 F 不变。目标物体在摄像头成像平面上的投影宽度为 W_2 ,目标物体与摄像头镜头的距离设为 D_2 ,根据三角形相似原理,有式(6):

$$\frac{W_2}{F} = \frac{W}{D_2} \quad (6)$$

其中, $D_2 = D - D_1$ 。

由式(6)同样能够得出目标物体的宽度计算公式,即式(7):

$$W = \frac{W_2 \times D_2}{F} \quad (7)$$

因为目标物体的宽度 W 是不变的,根据式(5)和式(7),可以计算出目标物体与摄像头镜头的距离 D ,如式(8):

$$D = \frac{W_2 \times D_1}{W_2 - W_1} \quad (8)$$

3 实验结果与分析

3.1 实验平台

文中的实验平台为 Ubuntu16.04,实验环境为:CPU,Intel(R) Core(TM) i5-7400,3.00 GHz×4;GPU,NVIDIA GeForce GTX 1080;软件环境 Python 2.7、Labelme、TensorFlow-GPU 1.1.0、Keras、OpenCV 等。

3.2 水面障碍物种类识别

水面障碍物识别共分为五个步骤,分别为:采集数据、标注标签、转换数据、训练数据、识别目标。其过程如图 3 所示。

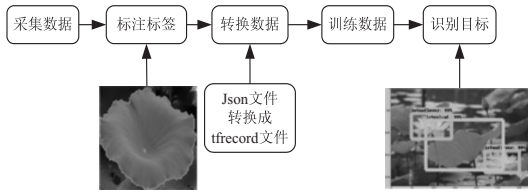


图 3 水面障碍物种类识别流程

原有的 Mask R-CNN 模型是由 COCO 数据集训练得到的,COCO 数据集中一共有 80 种常见的物种类别,但这 80 个类别中并不包含文中需要测试识别的水面障碍物种类,如果直接调用此模型会出现误检、漏检的情况,因此重新采集各类水草和船只等样本来训练新的 Mask R-CNN 模型。该实验以水面植物、船只等障碍物为实验数据集,从网络上下下载带有水面障碍物的图片共有 630 张,图片的尺寸大小没有限制,样本示例如图 4 所示。使用开源软件 Labelme 对样本进行标注,每个样本图片中至少包含一个目标物体,标注的样本示例如图 5 所示。以连续且密集的点连线将目标样本的轮廓边缘标注出来,并生成 Json 数据格式的文件,数据文件主要包括路径名、样本文件名、标注类别名、连接点坐标、标注样本长宽等数据。



图 4 水面障碍物样本示例



图 5 水面障碍物样本标注示例

从数据集中随机取 556 张图片作为训练样本,剩余 74 张图片作为测试样本,对训练样本训练 20 万次后得到 Mask R-CNN 模型。在原有的 74 张测试样本中添加额外 16 张包含水面障碍物和非水面障碍物的样本图片,基于 Mask R-CNN 模型对 90 张样本进行检测识别。在 90 个测试样本中,能够检测识别到的目标有 174 个,其中几种典型的水面障碍物识别结果如图 6 所示,不同类型的障碍物用不同颜色的矩形框框出并标出障碍物种类;统计测试样本的识别准确率和测试目标的识别准确率,结果如表 1 所示,测试样本的准确率达到 88% 以上,测试目标的识别准确率达到 97% 以上。

表 1 目标识别结果

指标	Mask R-CNN
测试样本/张	90
检测到的目标/个	174
平均识别时间/(s/张)	2.17
测试样本识别准确率/%	88.42
检测到的目标识别准确率/%	97.70

对于识别到的 174 个目标进行结果分析,如表 2 所示,其中,水面障碍物的类别被识别准确的个数为 166,水面障碍物和非水面障碍物被错误识别的个数为 8,水面障碍物识别的召回率为 99.40%,精确率为 95.95%。召回率表示的是 Mask R-CNN 模型检测到目标物体的能力,精准率表示的是 Mask R-CNN 模型能够将目标物体识别准确度。从数据上可以看出经过训练的 Mask R-CNN 模型能够将很好地检测并识别水面障碍物。

表 2 识别目标分析

类别	合计
True Positive (TP)	166
False Negative (FN)	1
False Positive (FP)	7
True Negative (TN)	0
Precision	95.95%
Recall	99.40%

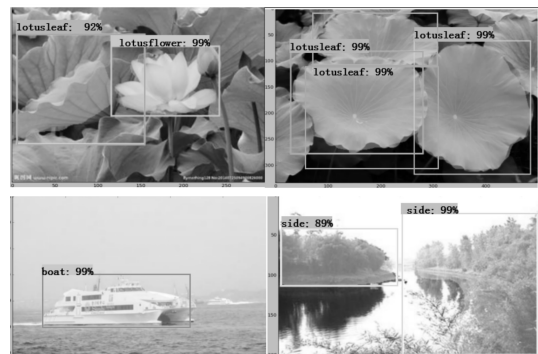


图 6 Mask R-CNN 识别示例

3.3 测 距

航行过程中,无人船以固定的速度和方向匀速向前行驶,截取摄像头拍摄的第一张图片后识别图片中尺寸最大的障碍物的类型并计算其宽度 W_1 ;一段时间后截取摄像头拍摄的第二张图片中框出障碍物的矩形框宽度 W_2 ,根据式(8)可以计算出摄像头与前方障碍物的距离,具体实验结果如表 3 所示。摄像机与障碍物的实际距离为手动测量,存在一定的误差。

表 3 测距结果

实际距离/cm	测量距离/cm	误差/%
64	65.89	2.91
121.5	126.69	3.94
224	223.90	0.04
326.5	326.62	0.04
429	431.64	0.62

根据表中实验数据,目标物体与摄像头之间的实际距离与测量距离的误差都是小于 5%,在允许误差范围之内。该测距实验虽然是基于小孔成像原理利用相似三角形计算摄像头与目标物体的距离,但是与以往不同的是,测距过程中并不需要知道目标物体的真实大小就能够计算出摄像头与障碍物之间的距离,计算简单方便,应用前景广泛。

4 结 束 语

该文利用深度学习和单目视觉的目标识别与测距方法来检测无人船前方水面障碍物的种类和距离。实验结果表明,Mask R-CNN 网络模型能够有效地对形状各异的水面障碍物进行识别,同时基于小孔成像原理的单目测距模型,与其他研究学者研究的此类模型不同的是测距计算不需要知道目标物体的真实大小,只需要知道无人船前进的距离就能够较为快速准确地测量出障碍物距离。因此,该文研究的基于单目视觉的水面障碍物识别与测距方法,能够有效地帮助无人船提高自主巡航能力。

由于 Mask R-CNN 深度学习网络模型较为庞大,且文中在训练模型时的训练样本规模较小,可能导致障碍物识别速度和精度仍然有提高的空间;测距模型的测量目标单一,只计算距离信息没有其他信息,例如方位坐标信息。所以,今后的研究方向是进一步提高 Mask R-CNN 识别效率,在测距模型中新增测量目标物体的方向坐标等信息。

参考文献:

[1] 李家良. 水面无人艇发展与应用[J]. 火力与指挥控制, 2012,37(6):203-207.
 [2] DALA N, TRIGGS B. Histograms of oriented gradients for human detection[C]//IEEE conference on computer vision

and pattern recognition. San Diego, CA, USA: IEEE, 2005: 866-893.

- [3] FELZENSZWALB P, MCALLESTER D, RAMANAN D. A discriminatively trained, multiscale, deformable part model [C]//IEEE conference on computer vision and pattern recognition. Anchorage, AK, USA: IEEE, 2008:1-8.
 [4] LOWE D G. Distinctive image features from scaleinvariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2):91-110.
 [5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//European conference on computer vision. Cham: Springer, 2016:21-37.
 [6] 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, NV, USA: IEEE, 2016:779-788.
 [7] 王振华. 基于深度学习的野外巡线系统图像目标检测研究 [D]. 北京:中国地质大学(北京), 2018.
 [8] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]//IEEE international conference on computer vision. Venice, Italy: IEEE, 2017:2999-3007.
 [9] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//IEEE conference on computer vision and pattern recognition. Columbus, OH, USA: IEEE, 2014: 580-587.
 [10] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
 [11] GIRSHICK R. Fast R-CNN [C]//IEEE international conference on computer vision. Santiago, Chile: IEEE, 2015: 1440-1448.
 [12] REN S Q, HE K M, 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, 2017, 39(6): 1137-1149.
 [13] FENG X Y, MEI W, HU D S. Aerial target detection based on improved faster R-CNN [J]. Acta Optica Sinica, 2018, 38(6):0615004.
 [14] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN [C]//IEEE international conference on computer vision (ICCV). Venice, Italy: IEEE, 2017:2980-2988.
 [15] 李冠楠. 直齿圆柱齿轮渐开线齿廓的结构光视觉测量技术 [D]. 长春:吉林大学, 2014.
 [16] 薛冰. 基于 Mask R-CNN 的电力设备锈迹检测 [J]. 计算机系统应用, 2019, 28(5): 248-251.
 [17] 韩延祥, 张志胜, 戴敏. 用于目标测距的单目视觉测量方法 [J]. 光学精密工程, 2011, 19(5): 1110-1117.
 [18] 叶海加, 陈罡, 邢渊. 双目 CCD 结构光三维测量系统中的立体匹配 [J]. 光学精密工程, 2004(1): 71-75.