



DOI: 10.12086/oe.2021.200062

基于深度学习的门机抓斗检测方法

张文明^{1,2}, 刘向阳^{2*}, 李海滨², 李雅倩²¹燕山大学电气工程学院, 河北 秦皇岛 066004;²燕山大学工业计算机控制工程河北省重点实验室, 河北 秦皇岛 066004

摘要: 在港口门机抓斗装卸干散货的作业过程中, 人眼观察无法精确判断抓斗所在位置, 会带来工作效率低下及安全性等问题。为解决该问题首次提出了一种基于深度学习的门机抓斗检测方法。利用改进的深度卷积神经网络 YOLOv3-tiny 对抓斗数据集进行训练及测试, 进而学习其内部特征表示。实验结果表明, 基于深度学习的门机抓斗检测方法可实现门机抓斗检测速度每秒 45 帧, 召回率高达 95.78%, 在很好满足检测实时性与准确性的同时, 提高了工业现场作业的安全性及效率。

关键词: 抓斗检测; 深度学习; YOLOv3-tiny; 空间金字塔池化; 反转残差组; 空洞卷积

中图分类号: TP391.41; U653

文献标志码: A

张文明, 刘向阳, 李海滨, 等. 基于深度学习的门机抓斗检测方法[J]. 光电工程, 2021, 48(1): 200062

Zhang W M, Liu X Y, Li H B, et al. The detection method for grab of portal crane based on deep learning[J]. *Opto-Electron Eng*, 2021, 48(1): 200062

The detection method for grab of portal crane based on deep learning

Zhang Wenming^{1,2}, Liu Xiangyang^{2*}, Li Haibin², Li Yaqian²¹School of Electrical Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China;²Key Laboratory of Industrial Computer Control Engineering of Hebei Province, Yanshan University, Qinhuangdao, Hebei 066004, China

Abstract: In order to solve the problems of low work efficiency and safety caused by the inability of human eyes to accurately determine the position of the grab during the loading and unloading of dry bulk cargo by portal crane, a method of grab detection based on deep learning is proposed for the first time. The improved deep convolution neural network (YOLOv3-tiny) is used to train and test on the data set of grab, and then to learn its internal feature representation. The experimental results show that the detection method based on deep learning can achieve a detection speed of 45 frames per second and a recall rate of 95.78%. It can meet the real-time and accuracy of detection, and improve the safety and efficiency of work in the industrial field.

Keywords: grab detection; deep learning; YOLOv3-tiny; SPP; inverted residual group; dilated convolution

收稿日期: 2020-02-25; 收到修改稿日期: 2020-05-09

基金项目: 河北省自然科学基金资助项目(F2019203195)

作者简介: 张文明(1979-), 男, 博士, 副教授, 主要从事计算机视觉, 模式识别的研究。E-mail: 327897150@qq.com

通信作者: 刘向阳(1996-), 男, 硕士研究生, 主要从事计算机视觉、目标检测方面的研究。E-mail: 2041203253@qq.com

版权所有©2021 中国科学院光电技术研究所

1 引言

近几年,随着港口行业的蓬勃发展,港口吞吐量不断加大,港口码头对装卸干散货的需求也越来越大,有关资料显示,2018年中国港口完成货物吞吐量就已经达到143.51亿吨^[1]。随着科学技术的发展,特别是5G时代的来临,各行各业都开始向着信息智能化方向发展,而利用智能化来加强港口散货装卸设备的检测进而提高作业效率将是港口散货装卸行业未来的发展趋势^[2]。抓斗是港口主要干散货装卸工具,所以对于抓斗的操作控制是提高门座式起重机和卸船机生产效率和自动化程度的关键^[3]。目前,对于门机抓斗检测采用的方法主要是人为操控,即驾驶员坐在门机驾驶室内,通过肉眼观察抓斗是否到达要抓取干散货或者释放干散货的合适位置,由人来判断何时下放或升起抓斗上的钢丝绳。因此存在以下问题:第一,由于人眼距离货物比较远,在释放抓斗时钢丝绳容易过放,即支持绳有较多松绳,一次作业循环浪费几秒钟,多次循环作业加起来就浪费了大量时间,产生大量的无用功。第二,由于司机长期作业会导致人眼疲劳从而导致误判,也会出现过放问题,这对企业的发展是不利的,因为除了耗时耗力又会增加公司的输入成本。第三,操作的安全性取决于司机的专业技能熟练程度,取物的高度完全靠门机司机肉眼观察,存在一定的局限性和较大的误差。那么如何精确检测出抓斗的位置,使得抓斗装卸货物更高效安全,便成了港口行业急需解决的问题。

近几年,随着科技的发展,深度学习越来越火热,尤其在目标检测领域取得了快速发展^[4];许多学者也开始将深度学习技术应用到各种实际应用场景中,比如将深度卷积神经网络应用到车辆目标检测^[5]和遥感图像中的飞机目标检测^[6]等任务中,较传统方法都带来了巨大的性能提升。抓斗检测也是一类目标检测问题,本文便采用计算机视觉中目标检测技术替代传统肉眼去观察的方式,利用深度学习自动实现对抓斗位置的实时检测。所谓目标检测,就是区分图像或者视频中的目标与其他不感兴趣的部分目标,包括检测它的位置及类别,主要的方法分为单阶段和双阶段两种方式。双阶段中传统的方法是基于滑动窗口加分类模型,但是这样会耗费大量的计算。卷积网络的出现带来了基于区域的特征提取方法,即RCNN^[7],但是训练一个RCNN模型非常昂贵,而且步骤较多:根据选择性搜索,对每张图片提取2000个单独区域;再用

CNN提取每个区域的特征。假设我们有 M 张图片,那么CNN特征就是 $M \times 2000$ 。这样会使RCNN的速度变慢,通常每个新图片需要40s~50s的时间进行预测,达不到实时检测抓斗的目的。在Fast RCNN^[8]中,将图片输入到CNN中(也就是通过一次卷积提取)会相应地生成传统特征映射。利用这些映射,就能提取出感兴趣区域。之后,使用一个感兴趣区域池化层将所有提出的区域重新修正到合适的尺寸,以输入到完全连接的网络中,这使得处理一张图片大约需要2s。尽管比RCNN要快,但是处理速度仍然很慢,达不到实时效果。Faster RCNN^[9]是Fast RCNN的优化版本,虽然直接用整张图作为输入,但是整体还是采用了RCNN那种候选区域加分类的思想,只不过是将提取候选区域的步骤放在CNN中实现了,由于是双阶段,先找区域再分类,尽管精度较高,但是速度不容乐观。单阶段目标检测模型YOLOv1的出现解决了上述的速度问题,由于它是利用整张图作为网络的输入,直接在输出层回归边界框(bounding box)的位置和其所属的类别,所以速度相对较快^[10]。但是由于YOLOv1中所划分的每个格子可以预测2个bounding box,最终只选择交并比(intersection-over-union, IOU)最高的bounding box作为物体检测输出,即每个格子最多只预测出一个物体。那么当物体占画面比例较小,一个格子出现多个物体时,却只能检测出其中一个,说明其定位不准,容易出现漏检情况。YOLOv3^[11]的出现打破了上述尴尬的境地。YOLOv3-tiny是YOLOv3的迷你版,速度更快,但检测精度不如YOLOv3,由于工程对于实时性的需要,所以本文选择YOLOv3的tiny版本进行改进,首先在原有网络基础上引入空间金字塔池化^[12]进而提取多方位的抓斗特征;其次引入深度可分离卷积^[13]并拓展网络宽度形成反转残差组加入到网络结构中,在加深网络、提高检测性能的同时维持了较少的参数量,减少了计算消耗;最后通过在网络低层加入两个空洞卷积层和高层进行特征融合,保证在不失分辨率的情况下扩大融合后特征图的感受野。用改进后的网络与原网络在抓斗数据集上进行对比实验,实验结果表明,改进后的网络模型较原始模型在性能上有显著提升,改进后的检测模型mAP达到96.25%,较原始模型mAP值提高了17.68%,召回率高达95.78%,较原始召回率提高了17.22%,且误检率为0,检测速度达到每秒45帧,很好地满足了门机抓斗检测的实时性与准确性,在一定程度上解决了因人眼观察无法精确判断抓斗所在位置所带来的工作效率

低下及安全性的问题。

2 YOLOv3-tiny 模型

2.1 YOLOv3-tiny 模型结构

YOLOv3-tiny 是一种端到端的卷积神经网络，由卷积层、池化层、上采样层和两个 YOLO 检测层组成，其网络结构如图 1 所示。相比于 YOLOv3 少了一个 YOLO 检测层和类似残差的连接，同时减少了大量的 1×1 和 3×3 的卷积核，虽然这样会使检测精度不如 YOLOv3，但是会大幅度降低要训练的参数量和浮点运算，所以更便于在移动端部署。同时速度明显高于其它模型，在 COCO 数据集上达到了 220 f/s，是 YOLOv3 的 6 倍之多，使得实时性获得了良好的保障。

2.2 YOLOv3-tiny 检测原理

YOLOv3-tiny 将整张图片作为输入，然后通过卷积神经网络进行特征提取，得到 $s \times s$ 大小的特征图，之后划分原图为 $s \times s$ 个网格。如果真实边界框中某个目标的中心坐标落在哪个单元格中，那么就由该单元格来预测此目标。在每个网格中预测 3 个边界框，以对多类目标进行检测。输出的每个边界框包含 (x, y, w, h, c_{con}) 五个基本参数，其中 x, y 代表边界框的中心离其所在网格单元格边界的偏移。 w, h 代表边界框真实宽高相对于整幅图像的比例。 c_{con} 为边界框的置信度，它代表了所预测的边界框中含有目标物的可能性和这个预测框预测的有多准两重信息，其计算公式为

$$c_{con} = P_o \times IOU \quad (1)$$

其中： P_o 是预测为要检测的目标概率值，IoU 为预测框和真实物体边界框的交并比。当得到每个预测框的类别置信度后，设置阈值，滤掉得分低的边界框，对保留的边界框进行非极大值抑制^[7]处理，进而得到最终的检测结果。类别置信度公式为

$$P(\text{class}_i | \text{Object}) \times P_o \times IOU \quad (2)$$

YOLOv3-tiny 借鉴 FPN^[14]的思想，融合多尺度信息预测，将高层输出的特征图进行上采样与低层的特征图进行融合，实现增强语义信息的同时，加强了对位置信息的检测。

3 改进方法

由于 YOLOv3-tiny 相比 YOLOv3 少了大量的 1×1 和 3×3 卷积核，这样虽降低了大量的计算量，提高了检测速度，但是网络深度大大降低，导致检测精度下降较多，容易出现定位不准，识别错误，发生误检以及漏检现象。本文根据原 YOLOv3-tiny 网络加以改进，通过在原有网络基础上加入四个反转残差组来增加网络的深度和拓展网络宽度并引入空间金字塔池化模块提取多角度特征，同时在低层与高层特征图融合间加入两个空洞卷积层，保证在不失分辨率的情况下下来扩大低层特征图的感受野，从而形成改进后的深度卷积神经网络。

3.1 空间金字塔池化模块

在 First YOLO 检测层前加入一个空间金字塔池化模块，加入的 SPP 模块如图 2 所示。SPP 模块由三个

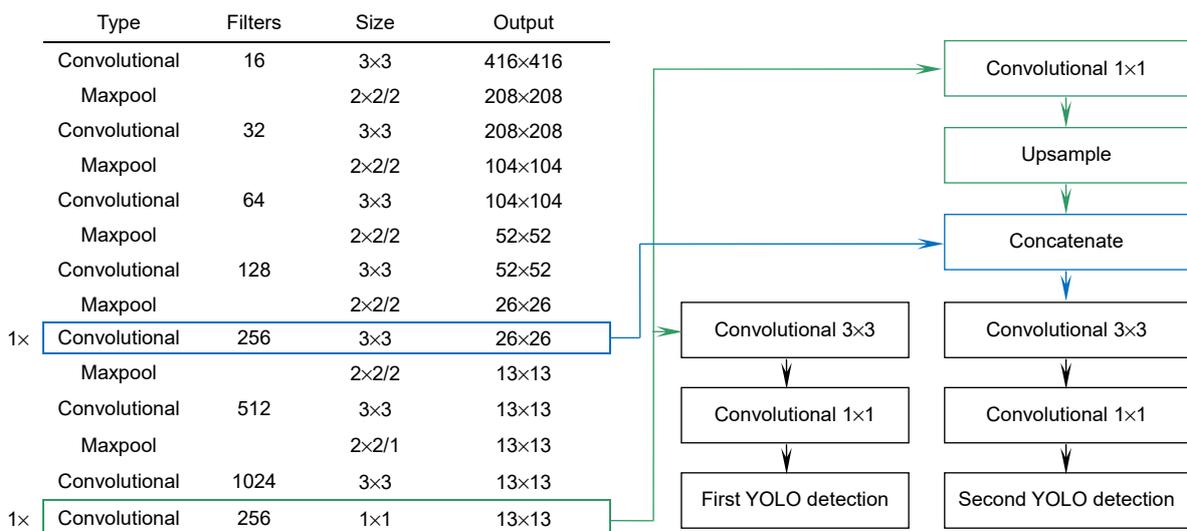


图 1 YOLOv3-tiny 网络结构
Fig. 1 Network structure of YOLOv3-tiny

平行的最大池化层、一个卷积层和 route 层组合形成。其中池化核尺寸大小分别为 5×5、9×9、13×13，步长均为 1。输入的特征图通过 1×1 卷积层以及不同大小的最大池化核作用，提取了不同角度的特征，形成了带有不同感受野的特征图，然后通过 route 层将前面获得的特征图进行通道维度上的拼接，进而提取到更多有用的多尺度信息。池化操作使得数据的维度降低，将较低层次的特征组合为较高层次的特征，通过融合提高了语义信息，可以使得模型检测的准确率提高。

3.2 反转残差组

算上最大池化层，YOLOv3-tiny 一共 24 层，网络深度较浅，但如果仅通过加深网络层数进一步提高抓斗检测的效果，伴随着网络深度的增加，势必会带来计算量的加大，进而影响到检测的实时性能，所以本文在改进的网络结构中引入深度可分离卷积^[13]，在降低计算量的同时拓展网络宽度，构成如图 3 所示的反转残差组，其中 Dwise 为深度卷积，Pwise 为逐点卷积。

当输入特征图经过深度卷积后，由于输出特征图的维度取决于输入特征图的通道数，所以提取的特征受限于输入通道数。为了提取更多的特征，反转残差组中输入的特征图首先经过 3 个分支中的 1×1 普通卷积核进行升维度，然后利用 3×3 大小的深度卷积将图像区域与通道分开考虑，对于不同的输入通道采取不同的卷积核分别进行卷积，进而收集每个通道的空间特征，在此之后分别应用三个分支中的逐点卷积，即采用 1×1 的标准卷积核将深度卷积输出的特征图进行组合，采集每个点的信息，同时将其维度降到与输入特征图相同的维度，然后汇总三个分支采集到的信息

进行元素求和，进而产生新的特征，这种网络宽度的拓展提取了输入特征图更多的信息，最后将如图 3 所示的反转残差组中的 Input 与三个分支产生的新特征构成残差组的连接来提高网络传播的性能，解决因改进后网络层数的加深可能引起的网络退化问题^[15]。

反转残差组中的每一平行层都有 BN^[16]模块和激活函数，BN 加速网络的收敛，而激活函数提高网络的表达性。为了避免网络反向传播^[17]时出现神经元死亡的现象，前两层的激活函数采用 LReLU^[18]，计算式为

$$f(x) = \begin{cases} x, & x \geq 0 \\ ax, & x < 0 \end{cases} \in R \quad (3)$$

反转残差组中的第三个平行层即逐点卷积后均采用 Linear 线性激活，因为经逐点卷积降维后，再用 ReLU 函数激活会破坏特征^[19]。

为了提高门机抓斗检测的性能，在改进的 YOLOv3-tiny 网络中加入 4 个图 3 所示的反转残差组，在实现加深网络深度的同时提高了检测的准确度。

3.3 特征融合

随着网络深度的增加，语义信息越来越强，虽然利于分类，但是所检测物体的边缘由于模糊程度加大不能得到准确回归，因此能见度下降，不利于抓斗定位。所以在改进的 YOLOv3-tiny 网络结构采用如图 4 所示的特征融合方式。Layer 9 即第九层中的特征图位于低层，含有较多的位置信息，通过添加两个 3×3 的空洞卷积层，在维持分辨率不变，增添较少参数的同时扩大其感受野^[20]，再将 Layer 62 中 2 倍上采样后的特征图与其进行维度上的连接，如图 4 所示。融合后的特征图既含有较高的语义信息，又提取了较多的位置信息，有利于抓斗的定位。

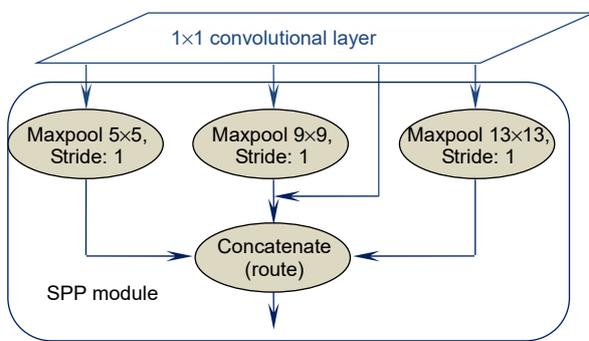


图 2 SPP 模块结构
Fig. 2 Architecture of SPP module

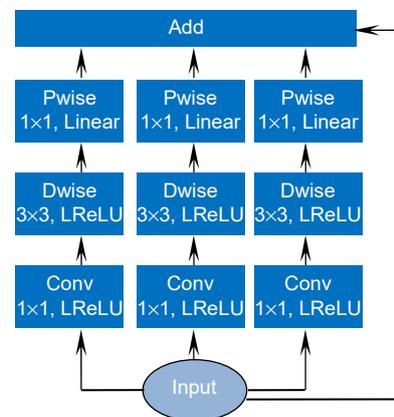


图 3 反转残差组
Fig. 3 Inverted residual group

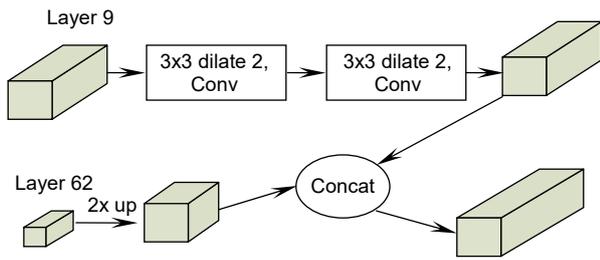


图 4 特征融合
Fig. 4 Feature fusion

3.4 抓斗检测总体思路

改进后的整体网络结构图如图 5 所示，利用改进后的 YOLOv3-tiny 网络模型，识别抓斗这个类别并实时跟踪抓斗的位置，输出相应的坐标信息。作业过程中，随着门机大臂的旋转、移动，相机和抓斗是一起移动的，这样从相机获得的二维坐标信息 X 、 Y 与测距传感器所测得的钢丝绳长度形成的深度信息 Z 一起构成三维坐标信息，但此时信息并不唯一，如图 6 所示。此时 A 、 B 、 C 三点在正上方摄像头与传感器所获得的绳长信息组合下所形成的三维信息是相同的，再加上大臂垂直面上的旋转角度 1 、 2 、 3 便唯一确定了抓斗所在的空间位置。当抓斗以某一种状态(比如 1)所对应的角度在同一水平面移动时，此时的位置信息

是不唯一的，需要角度传感器测量水平旋转角度。通过水平面与垂直面的大臂旋转角度信息和抓斗所在的 X 、 Y 、 Z 信息共同确定了抓斗在空间中的唯一位置信息，综合相机的内参，如焦距、感光单元大小以及分辨率等参数，计算出抓斗的空间位置坐标，进而实现对港口抓斗的实时检测。

4 实验结果及对比分析

4.1 摄像头选型及安装

由于港口独特的作业环境，选择适合现场抓斗检测的摄像头显得尤为重要。为了实时跟踪抓斗，检测其实时位置，所以将其安装室外，如图 7 中红色圈附近，由相机自带的机械结构固定在门机臂上，镜头垂直向下放置。作业过程中相机不能变焦，否则无法建立统一的坐标系，失去参考价值。

由于作业环境的复杂性，港口摄像头的选型需要满足以下几点要求：

- 1) 抗震性能。作业时门机臂会有震动，可能导致相机拍摄画面不清晰。
- 2) 防水性能。由于安装在室外，不得不考虑天气情况。
- 3) 工作温度范围。一般 $-30\text{ }^{\circ}\text{C}$ 到 $60\text{ }^{\circ}\text{C}$ 即可。

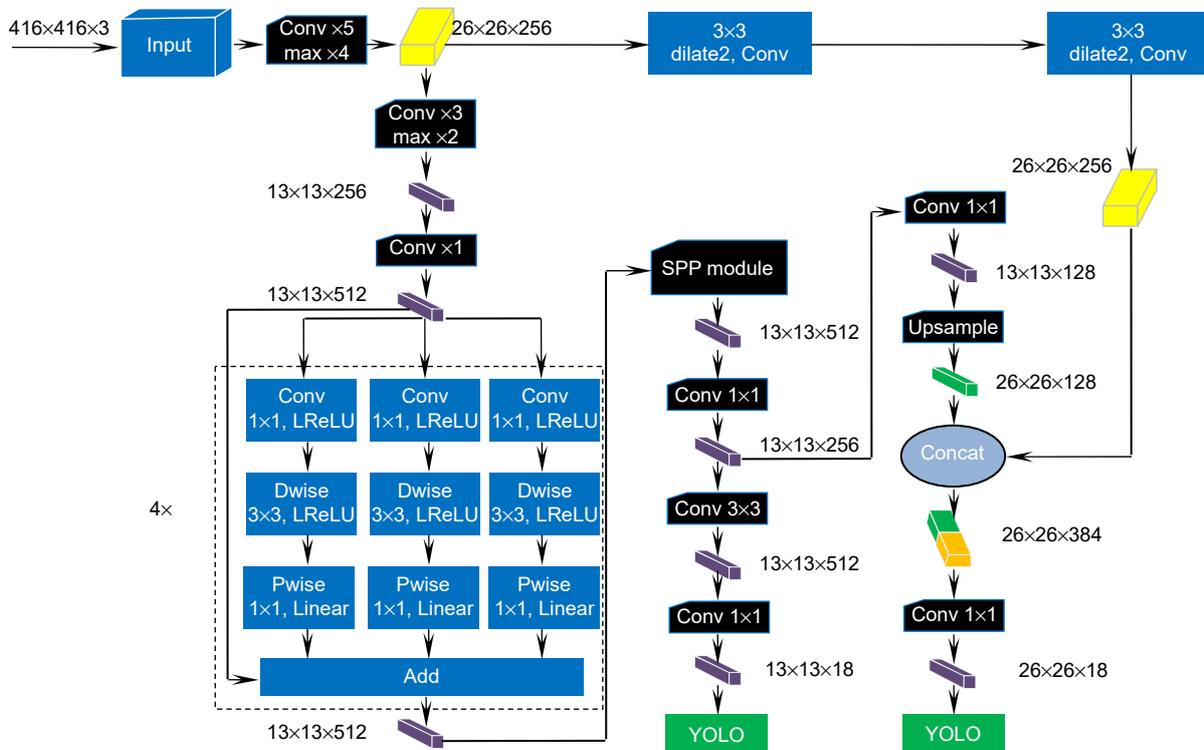


图 5 改进 YOLOv3-tiny 算法的网络结构
Fig. 5 The network structure of the improved YOLOv3-tiny algorithm

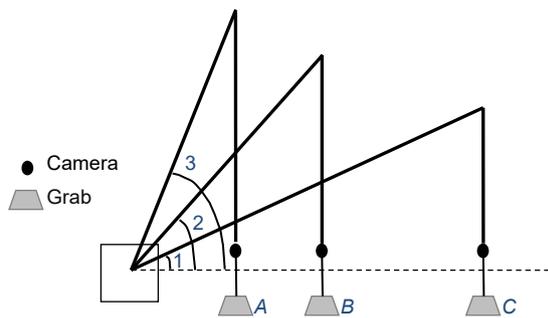


图 6 抓斗位置检测

Fig. 6 Position detection of grab

4) 抗电磁干扰。码头电气设备多, 功率大, 强大的电磁干扰经常会影响摄像头成像。

5) 对光照强度的敏感程度要小。

根据以上要求并综合实际情况选择了型号为 DS-2CD3T56WD-I8 的网络摄像机。

4.2 数据集制作

本实验中将数据集制作成 VOC2007 格式。首先通过安装在港口的摄像头采集几段抓斗作业的视频, 然后通过软件 Free Video to JPG Converter 截取其中部分帧作为要训练和测试的图片, 再通过标注软件 Labellmg 对所有图片进行标注, 如图 8 所示, 既要给出类别信息又要框住其位置, 从而形成 xml 格式的标注文件。用 python 程序将本次实验的图片分为训练集 3230 张和测试集 900 张并以 txt 格式进行保存, 最后再利用 darknet 源程序中的脚本文件将之前的 xml 标注文件进行归一化处理并转化为 txt 格式, 将其作为

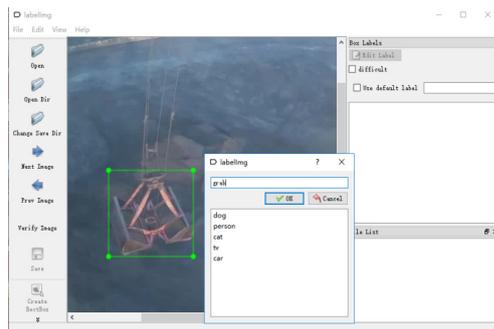


图 8 抓斗的标注

Fig. 8 The annotation of grab

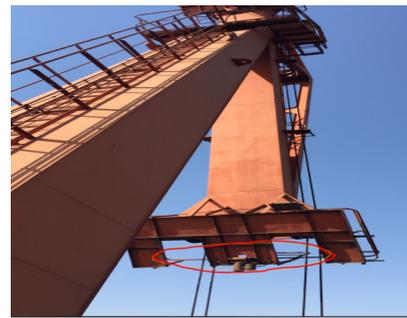


图 7 门机臂

Fig. 7 Crane arm

抓斗数据集标签, 至此, 完成抓斗数据集的制作。

4.3 实验平台及网络的训练

4.3.1 实验平台

本实验使用 Intel(R) Core(TM) i9-9900X CPU @ 3.50 GHz×20 处理器, 在 Ubuntu16.04 操作系统下进行, 为了提高运算速度, 缩短训练时长, 使用 NVIDIA GeForce RTX 2080 显卡一张, CUDA 10.0.130、CUDNN 7.5.0 对其进行运算加速。在本实验中使用的深度学习框架为 darknet, 结合 C、C++、python 编程语言以及 Opencv 和 gcc 库完成程序的编写及应用。

4.3.2 网络的训练

下载 YOLO 官网上 YOLOv3-tiny 的权重, 截取前 15 层作为改进后网络的预训练权重, 然后基于改进后的网络使用训练数据集中的图片进行训练、微调, 进而得到对于抓斗检测效果达到最佳的权重参数。实验中梯度优化采用带有动量的小批量随机梯度下降算法, 部分训练参数的设置如表 1 所示。

迭代次数小于 1000 时, 学习率从 0 开始随着迭代次数的增加逐渐变大, 直到迭代次数达到 1000 时, 学习率增加到 0.001。为了得到更好的收敛效果, 之后采用按需调整学习率策略, 本实验中设置当迭代次数达到 40000 次和 45000 次时, 学习率分别降至为 0.0001 和 0.00001。表 1 中的旋转角度、饱和度、曝光量以及色调为数据增强参数, 用来生成更多训练样本; 抖动因子作为一种数据增强手段, 通过随机调整宽高比的范围增加噪声来抑制过拟合。输入图片馈送入改进的网络后, 采取随机多尺度训练的方式对抓斗图片进行

表 1 网络参数表

Table 1 The table of network parameter

参数名	学习率	迭代次数	批量大小	动量	权重衰减	旋转角度	饱和度	曝光量	色调	抖动因子
参数值	0.001	50000	60	0.9	0.0005	3	1.5	1.5	0.1	0.3

训练。同时为了减小显存压力，将每一批次中的抓斗图片平均分成两个小批次分别进行前向传播。

在训练过程中，保存了算法各项指标的训练日志，并根据日志信息绘制相应曲线。平均损失以及平均交并比随着迭代次数的增加其变化曲线分别如图 9、图 10 所示。

从图 9 中可以看到，迭代次数在 20000 之前的这段时间，损失下降较快；随着迭代次数的增加损失下降趋于平缓；当达到 50000 次迭代时，损失值降到 0.007，达到理想状态。从图 10 中可以看出，平均交并比从 0.15 附近随着迭代次数的增加快速提高，说明模型的准确率在随之提高；当迭代至 40000 次后，平均交并比在 0.95 附近浮动。

4.4 抓斗检测模型对比实验及结果分析

为了验证改进后的网络是否有效提高了对门机抓斗检测的性能。本文首先进行了四个模型的对比实验。将原 YOLOv3-tiny 网络模型作为实验一；3.1 节中加入空间金字塔池化模块后的网络模型命名为 YOLOv3-tiny-SPP，将其作为实验二；在实验二网络模型的基础上加入 3.2 节中 4 个反转残差组后的网络模

型，命名为 YOLOv3-tiny-group，将其作为实验三；本文改进的 YOLOv3-tiny 模型命名为 YOLOv3-tiny-new，将其作为实验四。四个实验均在相同的抓斗训练集上训练，相同的测试集上测试。测试指标 mAP 值的计算按照 Pascal VOC2007 的计算标准，即 IOU 阈值设置为 0.5 时的测试结果，当抓斗的预测框与真实框的交并比大于等于 50% 时，视为匹配正确，检测出来的抓斗是正确的，即 T_p ，如图 11 所示(其中蓝色框为真实框，绿色框为匹配正确的预测框)。否则检测出来的抓斗是错误的，即 F_p ，如图 12 所示(其中蓝色框为真实框，红色框为匹配错误的预测框)。

四次实验对抓斗测试集检测的结果如图 13 所示。

根据测试结果计算抓斗检测的精确率(precision, P)以及召回率(recall, R)，计算公式如下：

$$P = T_p / (T_p + F_p) \quad (4)$$

$$R = T_p / (T_p + F_N) \quad (5)$$

其中 F_N 为没有被检测出来的抓斗数量。

四个实验的测试统计结果如表 2 所示。

从表 2 的实验结果可知，神经网络模型 YOLOv3-tiny-SPP 对抓斗检测的精确率、召回率以及 mAP 值都要高于 YOLOv3-tiny，但测试集的测试时间

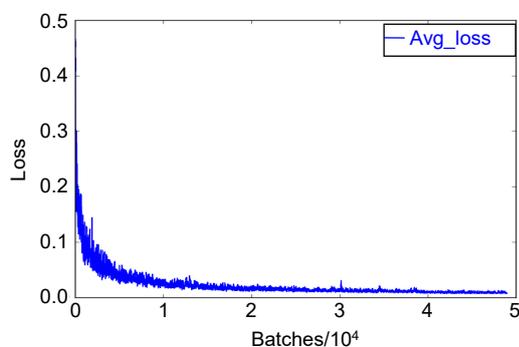


图 9 平均损失变化曲线

Fig. 9 The changing curve of average loss

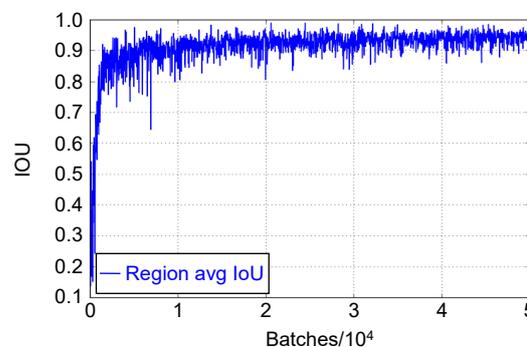


图 10 平均交并比变化曲线

Fig. 10 The changing curve of average IOU



图 11 正确的抓斗

Fig. 11 The correct grab



图 12 错误的抓斗

Fig. 12 The false grab

滞后了 0.779 s, 说明在 YOLO 检测层前加入的 SPP 模块确实提高了对抓斗识别的准确度, 但同时也带来了计算上的消耗。从表 2 中还可以看到神经网络模型 YOLOv3-tiny-group 对抓斗检测的精确率比 YOLOv3-tiny-SPP 模型提高了 1.51%, 召回率提高了 11.11%, mAP 值提高了 9.68%, 测试时间滞后 0.971 s, 说明改进的网络中通过加入 4 个反转残差组加深网络深度、拓展网络宽度的方法提高了对门机抓斗检测的准确度, 降低了漏检情况的发生。卷积神经网络模型 YOLOv3-tiny-new 相比于模型 YOLOv3-tiny-group 只多了如图 4 所示的特征融合方式, 但从表 2 的实验结果可以看到, 精确率、召回率和 mAP 值均高于网络模型 YOLOv3-tiny-group 的测试结果, 测试时间也只相差 6ms, 说明在低层与高层之间通过添加两个 3×3 的空洞卷积层维持分辨率不变、增添较少参数数量的同时扩大融合后特征图的感受野的方式是有效的, 提高了网络模型的检测性能。

由于门机抓斗作业的复杂性, 比如说光照、灰尘等因素会影响到抓斗的检测效果, 所以出现适当的漏检是可以接受的, 但是对于误判的情况要求较为严格, 即误检现象的出现, 因为要通过获取抓斗检测的实时位置信息进而完成抓斗的开闭、移动等操作。对比实验一、二、三、四的测试结果, 发现在抓斗测试集上改进后的网络 YOLOv3-tiny-new 表现最佳, 漏检数量

达到最低且不存在误检现象, 如图 13 所示。

从表 2 的实验结果中还可以发现, 虽然随着网络模型复杂度的增加, 识别抓斗的平均准确率在不断提高, 但是各模型在测试集上的测试时间也在不断延长, 说明本文改进后的网络在提高抓斗检测准确率的同时牺牲了一定的检测速度。

为了进一步验证本文改进后的网络检测性能的好坏, 以 mAP 以及每秒识别帧数(FPS)作为检测效果评价指标, 进行了如表 3 所示的抓斗检测性能对比实验。

从实验结果可以看出, 双阶段网络模型 Faster RCNN 的 mAP 值最高, 达到了 97.31%, 本文基于 YOLOv3-tiny 改进的网络模型 YOLOv3-tiny-new 的 mAP 值为 96.25%, 虽然检测准确率不及 Faster RCNN, 但检测速率比其快 225 倍, 因此, 对于门机抓斗这项实时检测任务而言, 本文改进后的深度学习网络模型表现更好。相比于原始模型 YOLOv3-tiny, 本文改进后的网络模型检测速率稍有下降, 但是检测准确率得到了大幅度提高, 从表 3 中可以看出, YOLOv3 网络模型在 mAP 以及识别帧率两项指标上都不及本文改进后的网络。由此可见, 本文改进后的网络模型兼顾了门机抓斗检测准确率以及检测速率。调用本文改进后的网络模型 YOLOv3-tiny-new 训练结束时生成的权重文件对港口抓斗作业视频进行测试, 检测效果如图 14 所示, 改进后的网络模型可以准确识别门机抓斗并

表 2 实验结果对比表

Table 2 Comparison table of experimental results

Model	Precision/%	Recall/%	Time/s	mAP/%
YOLOv3-tiny	93.64	78.56	5.203	78.57
YOLOv3-tiny-SPP	97.91	83.33	5.982	86.03
YOLOv3-tiny-group	99.42	94.44	6.953	95.71
YOLOv3-tiny-new	100	95.78	6.959	96.25

表 3 性能对比表

Table 3 Comparison table of performance

Model	FPS	mAP/%
YOLOv3-tiny	66	78.57
YOLOv3	23	94.91
YOLOv3-tiny-new	45	96.25
Faster RCNN	0.2	97.31

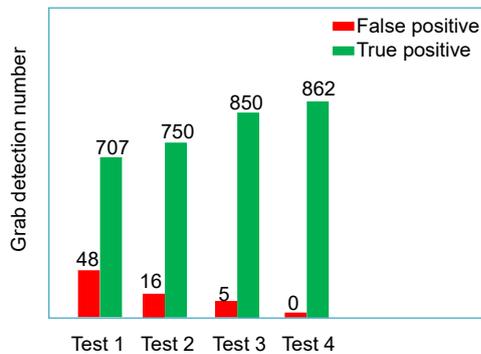


图 13 检测结果
Fig. 13 Detection results



图 14 测试抓斗视频
Fig. 14 The test of grab video

实时检测其位置,提高了作业的安全性,同时在取料的过程中可以根据抓斗所在位置提前释放抓斗上方的钢丝绳,进而在一定程度上避免抓斗过放现象的发生,在卸料时同样可以根据检测到的抓斗实时位置信息实现甩斗卸料,提高了门机抓斗作业效率。

5 结束语

本文首次提出了一种基于深度学习的门机抓斗检测方法,针对 YOLOv3-tiny 网络模型检测速度快但准确率低的特点,在其检测层前加入空间金字塔池化模块提高语义信息,同时在网络中引入深度可分离卷积并进行横向及深度扩展,形成 4 个反转残差组加入到网络中,带来较少计算消耗的同时提高检测准确率,并在改进后网络的第 9 层加入两个 3×3 空洞卷积层,在增添较少参数的同时扩大其感受野,再与第 62 层 2 倍上采样后的特征图进行特征融合,进一步提高了改进后网络的整体检测性能,使得改进后的网络检测性能达到最佳。本文改进后的网络模型 YOLOv3-tiny-new 相比于原 YOLOv3-tiny 网络模型,门机抓斗检测的精确度提高了 6.36%,召回率提高了 17.22%,mAP 值提高了 17.68%,同时检测速度达到 45 f/s,在速度与准确率之间得到了较好的平衡,更好地满足了工业现场对抓斗检测的实时性与准确性的要求,解决了在港口门机抓斗装卸干散货的作业过程中,人眼观察无法精确判断抓斗所在位置所带来的工作效率低下及安全性的问题。在未来的研究工作中,我们将进一步结合港口现场的作业环境,对多种复杂场景进行研究,将深度学习与门机作业更好地结合。

参考文献

- [1] Chen Y M. Current situation and future trend of Chinese ports[J]. *China Water Transp*, 2019(6): 7.
陈英明. 中国港口现状及未来走势[J]. 中国水运, 2019(6): 7.
- [2] Xing X J. Some thoughts on bulk cargo handling industry of domestic port[J]. *Hoist Con Mach*, 2019(10): 1.
邢小健. 对国内港口散货装卸行业的一些思考[J]. 起重运输机械, 2019(10): 1.
- [3] Ji B S. A study on the design of grab bucket control program[J]. *J Nantong Vocational Tech Shipping Coll*, 2012, 11(4): 76–79.
季本山. 现代门机抓斗控制程序的设计[J]. 南通航运职业技术学院学报, 2012, 11(4): 76–79.
- [4] Yao Z Y. Research on the application of object detection technology based on deep learning algorithm[D]. Beijing: Beijing University of Posts and Telecommunications, 2019.
姚筑宇. 基于深度学习的目标检测研究与应用[D]. 北京: 北京邮电大学, 2019.
- [5] Manana M, Tu C L, Owolawi P A. A survey on vehicle detection based on convolution neural networks[C]//*Proceedings of the 3rd IEEE International Conference on Computer and Communications*, 2017: 1751–1755.
- [6] Dai W C, Jin L X, Li G N, et al. Real-time airplane detection algorithm in remote-sensing images based on improved YOLOv3[J]. *Opto-Electron Eng*, 2018, 45(12): 180350.
戴伟聪, 金龙旭, 李国宁, 等. 遥感图像中飞机的改进 YOLOv3 实时检测算法[J]. 光电工程, 2018, 45(12): 180350.
- [7] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 580–587.
- [8] Girshick R. Fast R-CNN[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2015: 1440–1448.
- [9] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//*Proceedings of the 28th International Conference on Neural Information Processing Systems*, 2015: 91–99.
- [10] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 779–788.
- [11] Redmon J, Farhadi A. YOLOv3: an incremental improvement[Z]. arXiv:1804.02767, 2018.
- [12] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Trans Pattern Anal Mach Intell*, 2014, 37(9): 1904–1916.
- [13] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[Z]. arXiv:1704.04861, 2017.
- [14] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 936–944.
- [15] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770–778.
- [16] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[Z]. arXiv:1502.03167, 2015.
- [17] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition[J]. *Neural Comput*, 1989, 1(4): 541–551.
- [18] Maas A L, Hannun A Y, Ng A Y. Rectifier nonlinearities improve neural network acoustic models[C]//*Proceedings of the 30th International Conference on Machine Learning*, 2013.
- [19] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 4510–4520.
- [20] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[Z]. arXiv:1511.07122, 2015.

The detection method for grab of portal crane based on deep learning

Zhang Wenming^{1,2}, Liu Xiangyang^{2*}, Li Haibin², Li Yaqian²

¹School of Electrical Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China;

²Key Laboratory of Industrial Computer Control Engineering of Hebei Province, Yanshan University, Qinhuangdao, Hebei 066004, China



The correct grab

Overview: In recent years, with the vigorous development of the port industry, the port throughput is increasing, and the demand for loading and unloading dry bulk cargo is also increasing. At present, the method adopted is mainly man-made operation. The driver sits in the cab of the gantry crane, and observes whether the grab reaches the proper position to grab or release the dry bulk by naked eyes, and judges when to lower or raise the steel wire rope on the grab. Then there will be the following problems: first, because the human eyes are far away from the goods, the wire rope is easy to be over released when the driver releases the grab. A few seconds are wasted in one operation cycle, and a lot of time is wasted and a lot of idle work is produced in multiple operation cycles. Second, the driver's long-term operation will lead to eyestrain, which will lead to misjudgment and over the release. It is not conducive to the development of the enterprise, because, in addition to time-consuming and labor-consuming, it will increase the input cost of the company. So how to accurately detect the position of grab and make it more efficient to load and unload cargo has become an urgent problem for the port industry. In order to solve the problems of low work efficiency and safety caused by the inability of human eyes to accurately determine the position of the grab during the loading and unloading of dry bulk cargo by portal crane, a method of grab detection based on deep learning is proposed for the first time. The improved deep convolution neural network (YOLOv3-tiny) is used to train and test on the data set of grab, and then to learn its internal feature representation. The experimental results show that the detection method based on deep learning can achieve a detection speed of 45 frames per second, a recall rate of 95.78%, and a false detection rate of 0. Although the accuracy of detection is lower than Faster RCNN, the detection speed is 225 times faster than Faster RCNN. Compared with the original model YOLOv3-tiny, the detection speed of the improved network model in this paper is slightly reduced, but the detection accuracy has been greatly improved. Through the contrast test, we can see that the YOLOv3 network model is not as good as the improved network in the two indicators of mAP and FPS. Therefore, for the real-time detection task of gantry crane grab, the improved model in this paper performs better. It can meet the real-time and accuracy of detection, and improve the safety and efficiency of work in the industrial field.

Zhang W M, Liu X Y, Li H B, *et al.* The detection method for grab of portal crane based on deep learning[J]. *Opto-Electron Eng*, 2021, **48**(1): 200062; DOI: 10.12086/oe.2021.200062

Foundation item: Natural Science Foundation of Hebei Province (F2019203195)

* E-mail: 2041203253@qq.com