

自动化与智能化技术

基于视觉感知的机器人工件识别方法研究

崔新霞, 卢硕晨, 孙敦凯

(中国矿业大学 机电工程学院, 江苏 徐州 221116)

摘要: **目的** 解决定制化木门尺寸规格不统一、表面纹理多样而导致的堆垛分类困难、搬运效率低下等问题。**方法** 提出采用深度学习方法进行定制式木门工件检测, 以 YOLO V3 网络为基本框架开展机器人工件识别方法研究。首先, 通过图像数据增强和预处理, 扩充定制式木门数据; 然后, 进行 YOLO V3 损失函数改进, 并根据木门特征进行定制式木门数据集锚框尺度的重新聚类; 最后, 应用空间金字塔池化层进行 YOLO V3 中特征金字塔网络改进, 并通过随机选取的测试集验证本文方法的有效性。**结果** 测试数据集的平均检测准确率均值达到 98.05%, 检测每张图片的时间为 137 ms。**结论** 研究表明, 本文方法能够满足木门生产线对准确率和实时性的要求, 可大大提高定制化木门转线及堆垛效率。

关键词: 视觉感知; 目标检测; 深度学习; 卷积神经网络; YOLO V3 网络

中图分类号: TP241 **文献标识码:** A **文章编号:** 1001-3563(2023)07-0186-10

DOI: 10.19554/j.cnki.1001-3563.2023.07.021

Robot Workpiece Recognition Method Based on Visual Perception

CUI Xin-xia, LU Shuo-chen, SUN Dun-kai

(School of Mechatronic Engineering, China University of Mining and Technology, Jiangsu Xuzhou 221116, China)

ABSTRACT: The work aims to solve the problems such as the difficulty of stacking classification and the low handling efficiency caused by the non-uniform size and specification of customized wooden doors and the diversity of surface textures. A deep learning method was proposed to detect customized wooden door workpieces, and a robot workpiece recognition method was studied based on YOLO V3 network. First, through image data enhancement and preprocessing, the customized wooden door data were expanded. Then, the YOLO V3 loss function was improved, and the anchor frame scale of the customized wooden door data set was re-clustered according to the characteristics of the wooden doors. Finally, the spatial pyramid pooling layer was applied to improve the feature pyramid network in YOLO V3, and the effectiveness of this method was verified by a randomly selected test set. The average detection accuracy of the test data set reached 98.05%, and the detection time of each image was 137 ms. The research shows that this method can meet the requirements of the wooden door production line for accuracy and real-time nature, and can greatly improve the turning line and stacking efficiency of customized wooden doors.

KEY WORDS: visual perception; object detection; deep learning; convolutional neural network; YOLO V3 network

我国木工家具制造产业规模庞大, 并在未来有相当大的发展潜力。随着社会经济发展与人民生活水平的提高, 家具产品由千篇一律逐渐向定制化、

个性化的方向发展。针对定制式木门产业化生产的需求, 将机器视觉技术与工业机器人结合, 实现柔性化生产。为了解定制式木门生产线效率差、

收稿日期: 2022-10-06

基金项目: 国家重点研发计划 (2018YFB1308303)

作者简介: 崔新霞 (1982—), 女, 博士, 讲师, 主要研究方向为机器视觉。

自动化水平低的问题, 以及为实现木门生产智能化提供了新的方案^[1]。

机器视觉主要首先通过相机获取生产线信息, 然后通过与工业计算机的通信进行工件目标检测, 最后将需要执行的指令传递给工业机器人实现堆垛动作^[2]。这种方式的出现使机器人能够完成自主决策, 提高了定制式家具智能制造生产线的灵活性。目前, 目标检测算法主要分为以下 2 类: 一类是以传统图像处理方法为核心的模板匹配方法, 一类是基于统计学的模式识别方法^[3]。其中, 定制式家具生产线小批量、多样化、个性化的要求, 模板匹配方法很难提取到合适特征, 检测精度和速度存在缺陷。随着模式识别的发展, Girshick 等^[3]在深度学习方法逐渐取得成功后, 提出使用区域提议方法配合卷积神经网络算法进行目标识别。该算法大幅提高了目标识别精度。该方法通过选择搜索生成独立候选区域, 并通过支持向量机算法进行分类和修正。Ren 等^[4]提出了一种真正实现端到端计算的算法, 有效地解决了由于选择性区域提议产生多余的候选框而造成的冗余计算。Redmon 等^[5]提出了 YOLO 系列算法, 该类算法是一种基于回归的检测方法。由于舍弃了候选框生成过程, 检测速度得到明显提升, 但影响了准确率。陈宏彩等^[6]基于 YOLOv3 算法建立深度卷积神经网络学习检测模型, 通过将神经网络结构浅层和深层特征向量连接并标准化, 提取玻璃瓶多尺度特征信息。

近几年, 随着深度学习的不断发展, 卷积神经网络在目标检测的各个领域中, 取得了突破性的进展^[7]。特别是针对每扇木门都有独特的外观, 卷积神经网络通过自适应学习图像特征, 无须复杂的人工提取工程, 实现对目标的精准检测^[8]。因此, 本文以单阶段目标检测算法 YOLO V3 为基础, 针对原始锚框尺度不适合定制式木门数据集的情况, 利用 K-means 算法实现新锚框尺度的聚类; 然后采用广义交并比损失函数代替原始损失函数实现更加准确的模型拟合; 最后采用特征金字塔网络改进网络结构, 提高定制式木门检测算法的准确率, 并与几种算法进行性能对比。

1 数据准备

1.1 木门图像数据集

定制式木门生产线往往需要加工造型多样, 尺寸不一的木门。木门工件由多种木材复合加工而成, 通常表面呈现纹理特征较弱, 存在大面积的重复纹理等特点。此外, 木门加工过程中会进行包浆工序处理, 在环境光照下容易产生影响, 这些因素都会影响检测准确率与定位精度, 而卷积神经网络通过自适应学习这些木门特征便可实现木门的精准识别。

根据木门加工厂商的生产情况, 为保证数据集中图片数据量满足模型训练要求, 优先选择了出货量大的木门样式。为更好地体现本文方法的检测性能, 主要选择了几种尺寸规格不同的定制式木门, 如图 1 所示。

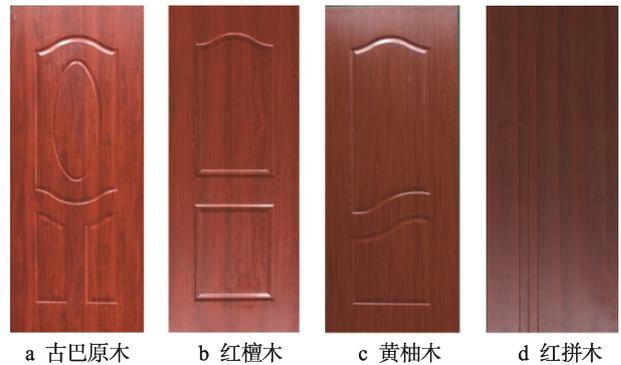


图 1 定制式木门工件实物

Fig.1 Physical drawing of customized wooden door

1.2 相机选型及布置方案

机器视觉系统作为本文重要研究内容, 对相机进行准确选型至关重要。相较于普通单目相机, 深度相机具有更强的实时场景描述能力, 能够以不同方法获得场景的深度信息, 目前已经广泛应用于各种工业领域。几种深度相机的对比如表 1 所示。

表 1 深度相机方案对比
Tab.1 Comparison of in-depth camera solutions

方案	结构光法	飞行时间法	双目视觉法
基础原理	条纹光线解调	反射时差	双目三角测量
分辨率	中	低	高
测量精度	中	高	高
抗光照能力	低	中	高
硬件成本	中	高	低
算法开发难度	中	低	高
内外参标定	需要	不需要	需要

结合木门加工要求, 选择抗光照能力强、硬件成本低和能生成高分辨率视差图的双目视觉方案。相机选择 Stereo labs 公司研发的 ZED 相机, 如图 2 所示, 相机的主要参数如表 2 所示。



图 2 ZED 相机外观

Fig.2 Appearance of ZED camera

表2 ZED 双目相机主要参数
Tab.2 Main parameters of ZED binocular camera

性能指标	参数
镜头类型	广角
最大分辨率	4 416×1 242
视场角	90°(H)×60°(V)×100°(D)
基线距离	120 mm
接口类型	USB3.0
质量	159 g
相机焦距	2.8 mm

对工业机器人的机器视觉系统来说,主要布置方案分为2种,如图3所示。第1种为眼在手型布置方案,将相机安装于机械臂上;第2种为眼在手外型布置方案,相机独立固定于相机支架,不随机械臂运动,且相机与机器人基座位置保持不变。

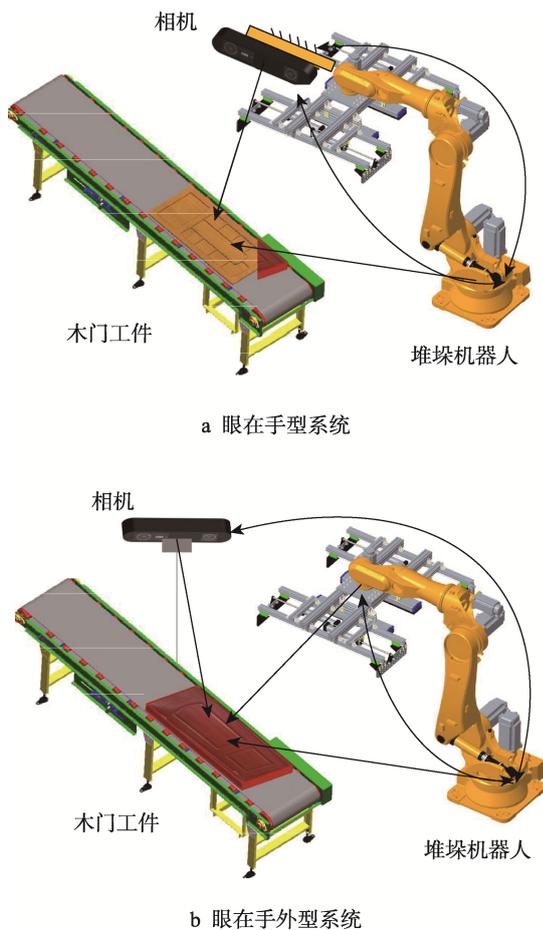


图3 相机布置方案比较
Fig.3 Comparison of camera layout

2种相机布置方案的手眼标定方式各不相同,坐标系转化的过程也不同。对于眼在手型系统,相机拍摄定制式木门工件生产场景图像并进行处理,得到定

制式木门在相机坐标系下的坐标,根据手眼标定结果得出定制式木门在机器人基坐标系的位置,由机器人完成指定抓取动作。该系统在工作过程中相机采集的图像不会受到机械臂遮挡,但是相机的固定改变了机械臂结构,并增加了机械臂负载,且双目相机有视场范围约束,当末端执行器与目标物体距离过近时,无法进行准确的深度测量。眼在手外型系统通常固定在独立于机械臂之外的相机支架上,相机镜头平面与木门工件表面及传送带表面严格水平,同样通过立体匹配算法计算定制式木门在相机坐标系位置,然后通过固定位置关系计算机器人基坐标系下工件的位置,根据基坐标系与末端执行器之间的固定关系获得末端执行器坐标系下的工件位置,完成机械臂抓取操作。由上述原理可知,眼在手外型系统视场范围大,可以获得全局的木门生产场景信息,但相机位置固定于相机支架,在实际工作过程中,工件会受到堆垛机器人运动的遮挡,导致部分时间内工件图像缺失,系统灵活性不如眼在手型系统,所以常应用于机器人基座固定的场景。根据上述需求分析,由于本文研究的智能识别与定位系统应用于定制式木门搬运机器人,不需要完全依靠于图像反馈控制,灵活性要求低,所以采用眼在手外型的双目相机布置方案。

1.3 数据预处理及数据增强

训练数据来自于定制式木门工件进行的多角度图像采集,数据分布如表3所示。深度学习模型的性能往往依赖于是否有充足的样本数据,需要为模型提供足够的木门工件特征进行学习。原始的定制式木门工件数据集样本量有限,训练过程中容易产生过拟合现象,从而对新的定制式木门图像失去检测能力。采用图像增强方法来改善数据量有限的问题并改善模型的泛化能力^[9]。图像增强方法包括翻转、旋转、缩放、裁剪和添加高斯噪声。本文通过旋转 90°、180°和 270°以及添加高斯噪声从原始定制式木门工件图像生成4个附加图像。

表3 定制式木门数据分布情况
Tab.3 Data distribution of customized wooden door

木门种类	原始样本数量	增强样本数量
古巴原木(type1)	146	730
红檀木(type2)	155	775
黄柚木(type3)	163	815
红拼木(type4)	151	755

在木门加工生产工艺流程中,为保证工件准确进入各生产工序,通常在传送带末端放置木门挡板,一方面准确调整工件位姿,从而实现准确的木门位置标注,另一方面防止工件滑落。木门挡板的主要结构如图4所示。

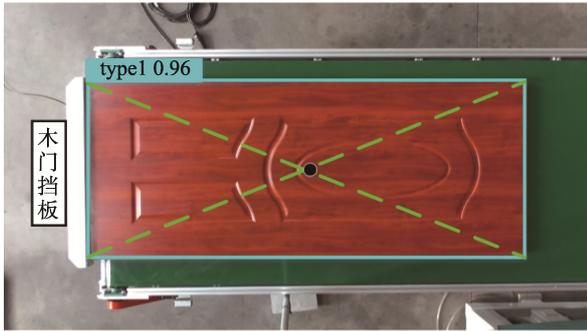


图 4 传送带中的木门位姿矫正效果
Fig.4 Effect diagram of wood door position and posture correction in conveyor belt

选择 Pascal VOC 数据格式, 利用图片标注软件 LabelImg 进行定制式木门的标注及 xml 文件的保存。标记的主要信息包括定制式木门工件类别、真实边界框的位置坐标以及边界框的高度和宽度。实验随机选取数据集中的 80% 作为训练集, 其余 20% 作为测试集, 用于评估模型的泛化性能。

2 目标识别方法

为了实现定制式木门加工生产线中准确的工件检测, 选择 YOLO V3 算法作为目标检测的基础网络框架。该算法是经过 YOLO V1、YOLO V2 改进而来的, 是目前兼顾实时性与准确性效果的单阶段目标检测方法。此外由于木门成品具有造型多样、尺寸不一, 表面纹理特征较弱, 存在大面积重复纹理, 木门表面在环境的光照下容易产生反光, 影响识别的准确率与定位精度等问题, 因此对识别算法的要求较高。YOLO V3 算法在保持识别速度的前提下, 提升了预测精度, 尤其是加强了对物体细节的识别能力, 因此选用 YOLO V3 对木门进行检测。区别于以 RCNN 网络为代表的双阶段目标检测方法, YOLO V3 方法将目标检测问题转化为逻辑回归问题, 将输入图形分割为不同的栅格, 每个栅格各自负责对应的物体, 最后用非最大值抑制法 (Non Maximum Suppression, NMS) 选出合适的预选框。检测结果包含检测框的木门种类、中心点坐标、检测框的高度和宽度, 以及反映当前边界框检测出物体的概率 $P_{\text{confidence}}^{[10]}$, 其定义如式 (1) 所示。

$$P_{\text{confidence}} = P_{\text{object}} \times I_{\text{ou}} \quad (1)$$

其中, 边界框若包含检测物体, 则 P_{object} 为 1, 否则为 0; I_{ou} 为预测边界框与标注的真实边界框之间的交集面积。

其次, 在 YOLO V3 的边界框的回归机制中, 每个边界框用 4 个参数确定, 分别为中心点坐标 (x, y) 、宽度 (d) 以及长度 (l) 。为了加快回归速度, 将边

界框参数转化为相对于栅格的相关系数, 使边界框参数变化范围回归到 0~1 内, 解决边界框训练难度大的问题, 具体回归方式如图 5 所示。

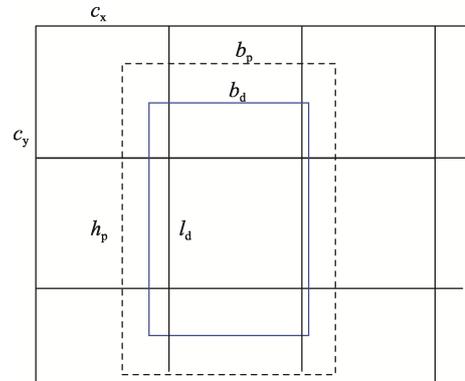


图 5 YOLO V3 位置预测原理
Fig.5 Principle of YOLO V3 position prediction

图 5 中, c_x 、 c_y 分别为边界框中心点坐标距离第 1 个栅格的 x 方向和 y 方向数目; l_d 、 b_d 分别为预测边界框对应的长和宽; t_x 、 t_y 分别为预测的边界框相对于真实边界框 x 方向和 y 方向的偏移量; $\sigma(x)$ 函数为激活函数, 用于偏移量的归一化; t_w 、 t_h 为预测边界框相对于真实边界框长和宽的尺度缩放。 b_p 、 h_p 分别为锚框的宽和高, 该初始值由聚类方法得到, 边界框的预测的参数由式 (2) 计算得出。

$$\begin{cases} b_x = \sigma(t_x) + c_x \\ b_y = \sigma(t_y) + c_y \\ b_d = l_d e^{t_w} \\ b_p = h_p e^{t_h} \end{cases} \quad (2)$$

在基本的图像特征提取方面, 输入图像的大小为 416 像素×416 像素。其原理如图 6 所示, YOLO V3 采用了称之为 Darknet-53 的网络结构 (含有 53 个卷积层) [11], 它借鉴了残差网络的做法, 层与层之间设置了快捷链路, 在保证速度的情况下, 更多地提取了图像特征。

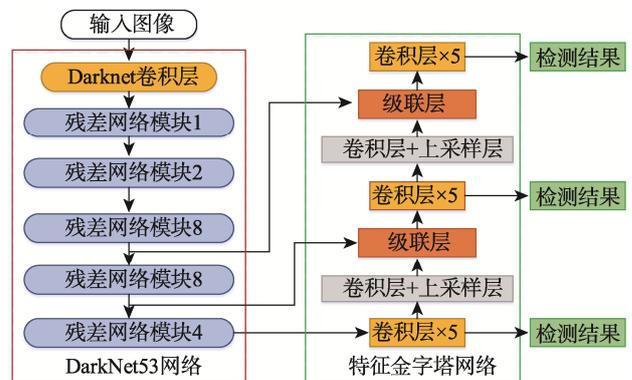


图 6 YOLO V3 基本框架
Fig.6 Basic frame of YOLO V3

YOLO V3 使用特征金字塔网络进行检测任务^[12]。来自各个图层的特征图通过构成特征金字塔网络的级联与上采样的特征合并, 通过结合 Darknet53 网络将待检测木门图片转化为 3 个不同尺度的特征图对应检测大中小 3 种不同类型的目标, 从而提高了检测各种尺度的目标的性能, 最终得到精确的分类信息和检测框的准确位置。图 6 中残差网络模块的数字代表着不同数量的残差单元, 卷积层为 DarkNet 卷积层, 包含二维卷积层、归一化层及非线性激活函数。级联层为张量拼接操作, 其作用为将浅层特征输出与上采样后的特征输出进行拼接, 最终输出的检测结果包括了 3 个不同尺度。

2.1 边界框损失函数改进

YOLO V3 的损失函数主要由分类损失、候选框的回归损失以及置信度的损失构成, 其中对于边界框回归损失, 传统的 YOLO V3 算法采用均方差损失函数, 通过直接对边界框参数 (x, y, d, l) 进行回归。这种计算方式独立了边界框坐标位置的影响和边界框长宽参数的影响, 导致边界框的回归准确率差、模型收敛效果差等。YOLO V3 虽然通过尺度因子减少了不同锚框尺度的影响, 但实际训练过程中并未取得本质改善, 因此在边界框的回归损失方面必须引入新的计算方式。随着目标检测算法得发展, I_{ou} 损失函数通常被应用在各类目标检测算法中, 目标检测方法中的交并比即预测与真实的相交与相并的面积比值, 如式 (3) 所示。

$$I_{ou} = \frac{area(B_{bdt} \cap B_{bgt})}{area(B_{bdt} \cup B_{bgt})} \quad (3)$$

针对当前函数存在的缺陷, 使用广义交并比损失函数 (Generalized Intersection Over Union, G_{iou}) 代替^[13]。 G_{iou} 具有非负性以及区分方向性能, 一定程度上克服了 I_{ou} 的缺点, 同时具备 I_{ou} 的尺度不变的优点, 主要计算过程如式 (4) 所示。

$$G_{iou} = I_{ou} - \frac{|area(C) - area(B_{bdt} \cap B_{bgt})|}{|area(C)|} \quad (4)$$

式 (4) 中引入面积单位 C , C 为包围 B_{bdt} 和 B_{bgt} 的最小面积, 面积 C 的具体几何意义如图 7 所示。虚线框的部分为 C , 点划线框为真实边界框, 黑色框为预测边界框, 阴影部分为 2 类边界框的交集。面积单位 C 的引入使 G_{iou} 的计算考虑到了非重叠区域, 能够进一步反映预测框与真实框重叠方式的细节。当预测边界框与真实边界框重叠面积较大时, G_{iou} 趋近于 1, 反之重叠区域较小。随着预测框与真实框的距离增加, G_{iou} 不断减小, 且趋近于-1, 当两者不重叠时, 取最小值-1。

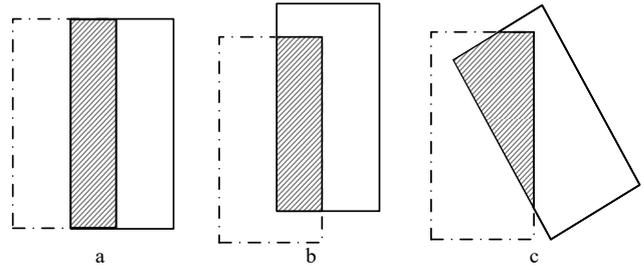


图 7 不同重叠方式下 G_{iou} 计算原理
Fig.7 G_{iou} calculation principle under different overlapping modes

2.2 K-means 聚类方法更新训练锚框尺度

在数据训练的过程中, 上述 YOLO V3 基本框架中的锚框尺度针对 COCO 数据集进行设计, 对于所需检测的定制式木门工件, 原始锚框尺寸并不能与木门工件尺寸进行对应, 从而影响网络训练效率和检测准确率。针对上述情况, 本文提出采用 K-means 聚类的方法来确定锚框方案, 其主要原理是将输入包含 N 个标注点的数据集分为 K 个聚类, 将数据集中所有点划分到距离各个聚类点最近的聚类中心中。这个过程中使用欧式距离作为分类衡量标准, 欧式距离的主要表达式如式 (5) 所示。

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (5)$$

在卷积神经网络训练过程中, 通常利用池化层来降低图像特征的维数, 保持特征、提高模型鲁棒性。常用的池化层通常通过计算图像局部区域的最大值和平均值, 所求值代表这一区域的特征^[14]。由于网络训练过程中固定了输入图像尺寸, 这种池化层仅有一个特征尺度的输出, 导致检测准确率的下降。为了消除这一现象带来的影响, 引入空间金字塔池化的方法, 该方法可以接受不同图像尺寸的输入, 一次性计算整张图像中不同大小的特征图, 然后进行任意区域的池化, 输出不同特征尺度的融合特征, 从而实现更高性能的特征提取, 原理如图 8 所示。

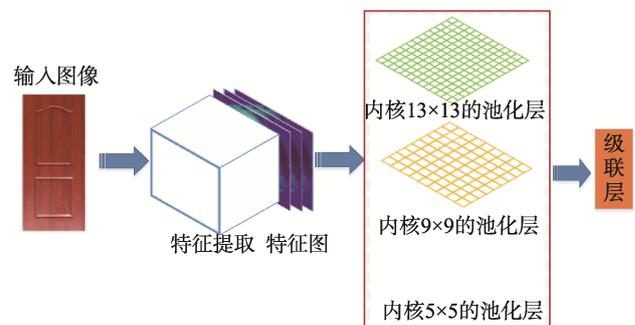


图 8 空间金字塔池化网络原理
Fig.8 Principle of spatial pyramid pooling network

空间金字塔池化方法降低了训练时用于固定输入图像尺寸的裁剪、缩放等操作带来的特征丢失的概率。图 8 中输入图像的特征映射到空间金字塔池化 (Spatial Pyramid Pooling, SPP) 网络中, 内核尺寸为 5×5 、 9×9 、 13×13 的最大池化, 将得到的特征串联输入级联层进行特征整合^[14]。由于不同尺度的特征提取被合并, 空间金字塔池化网络获得了更加丰富的特征信息, 避免重复提取。空间金字塔池化模块安排在 Darknet53 网络特征提取之后, 对 3 种不同尺度的目标识别均产生影响, 改进后的 YOLO V3 网络结构如图 9 所示, 阴影部分模块即为空间金字塔池化网络模块。

3 实验结果与分析

3.1 实验环境及性能评价指标

所有训练和测试选用来自谷歌的 TensorFlow 深度学习训练框架, 软件环境为 Windows10 操作系统, 开发工具选择 Python3.7.9, 并使用 JetBrains 公司开发的编译器 PyCharm 编译。硬件方面处理器为 Intel Core I7-6700HQ, 主频为 2.6 GHz。独立显卡为 GeForce RTX 2060, 显存为 6 GB。

由于定制式木门工件的特征复杂, 零基础开始训练很难进行有效的模型拟合, 所以本文模型训练以基于开源数据 ImageNet 的训练结果作为预训练模型^[15], 初始化模型部分参数。为使训练模型接近理想状态下的收敛, 将初始学习率、输入图像大小、学习衰减系数等模型参数分别调整如表 4 所示。模型训练应用包含动量的随机梯度下降 (Stochastic Gradient Descent with Momentum, SGDM) 算法^[16]。

为了验证本文算法的准确率与实时性是否能够满足定制式木门工件系统需求, 首先进行混淆矩阵的

定义, 其中包括真阳性 (True Positive, T_p)、假阳性 (False Positive, F_p)、真阴性 (False Negative, F_N) 和假阴性 (True Negative, F_N) 4 个指标。

对于本文的应用对象, T_p 表示定制式木门类别正样本被正确识别, F_p 表示背景或其他种类木门负样本被错误识别为正样本, F_N 表示定制式木门正样本被错误识别为背景或其他负样本, T_N 表示背景区域或其他负样本被正确识别为负样本^[17]。根据混淆矩阵, 选择精确率 (P)、召回率 (R) 以及平均精确率 (Average Precision, A_p) 等作为定制式木门工件检测算法的主要评估指标, 定义如式 (6)、式 (7) 所示。

$$P = \frac{T_p}{T_p + F_p} \tag{6}$$

$$R = \frac{T_p}{T_p + F_N} \tag{7}$$

A_{p_i} 是由 Precision-Recall 曲线与坐标轴所围成的面积, 该面积表示某类木门的检测准确率。平均精确率均值 (Mean Average Precision, M_{ap}) 表示所有定制式木门类别的平均精度均值, 与单一木门检测准确率之间转化关系如式 (8) 所示。

$$M_{ap} = \frac{1}{N} \sum_{i=1}^N A_{p_i} \tag{8}$$

3.2 K-means 锚框聚类前后结果分析

通过 K-means 聚类方法进行了定制式木门数据集的锚框尺度聚类, 经过聚类获得能够适应该数据集的新锚框为 $[(674,273), (707,287), (719,295), (1\ 308,538), (1\ 329,545), (1\ 343,553), (1\ 365,539), (1\ 386,552), (1\ 403,570)]$ 。为验证聚类锚框尺度是否适应定制式木门数据, 采用原始 YOLO V3 方法进行了不同锚框尺度的模型训练, 进行聚类前后训练结果对比, 使用 G_{iou} 损失与 M_{ap} 对比分析模型训练情况, G_{iou} 损失曲线如图 10 所示。

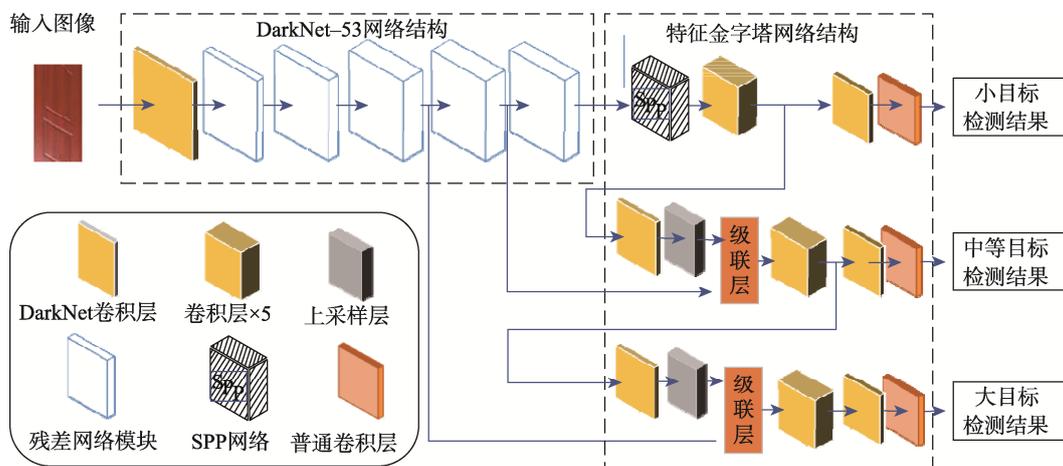


图 9 改进后的 YOLO V3 网络结构
Fig.9 Improved YOLO V3 network structure

表 4 训练超参数值设置
Tab.4 Setting of training hyperparametric values

参数名称	参数值
批处理大小(Batch size)	4
输入图像大小(Image size)	416×416
基础学习率(Base learning rate)	0.001
学习率衰减比率(Learning rate decay)	0.1
动量(Momentum)	0.9
权值衰减(Weight decay)	0.005

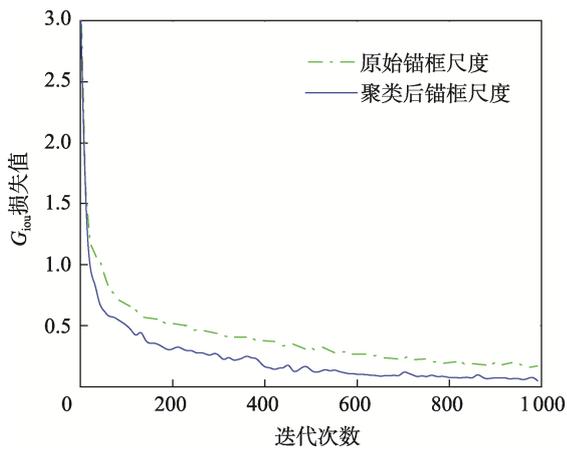


图 10 锚框聚类前后 G_{ious} 损失曲线
Fig.10 G_{ious} loss curve before and after anchor box clustering

由图 10 可见，随着模型训练的迭代训练过程，2 种锚框尺度的 G_{ious} 损失曲线均有不同程度的振荡。主要原因为木门纹理的特征复杂，图片拍摄分辨率限制等，使得数据的学习变得困难。2 种锚框尺度的训练均从 600 次开始收敛。聚类后的锚框尺度稳定时的 G_{ious} 值为 0.06，相较于原始锚框的 G_{ious} 值低了 0.11。利用上述 2 种锚框尺度的训练模型在测试集中实验，并采用平均精确率均值曲线比较定制式木门检测性能，如图 11 所示。

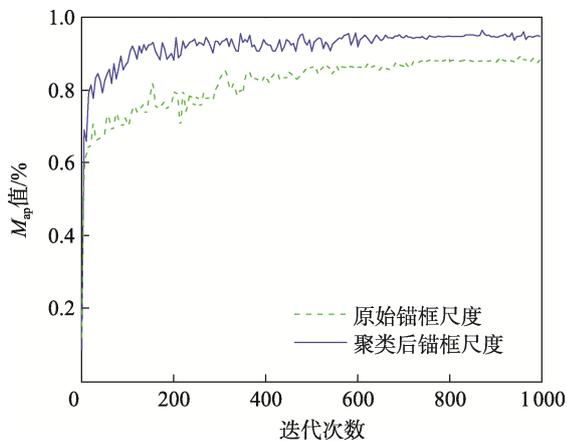


图 11 锚框聚类前后平均精确率均值曲线
Fig.11 mAP curve before and after anchor box clustering

在木门数据训练过程的开始阶段，由于模型训练初期拟合不准确，识别准确率均有不同幅度的振荡，在 600 次后趋于平稳，使用聚类后的锚框尺度训练模型的 M_{ap} 值为 93.49%，高于使用原始锚框尺度训练模型的 M_{ap} 值（89.43%）。该测试结果证明经过聚类后锚框尺度更适合定制式木门数据，检测精度更高，所以在后续实验对比中采用聚类后的锚框尺度。

3.3 空间金字塔池化网络改进前后对比

在确定了适合于检测模型的锚框尺度后，为了对比基于空间金字塔池化层改进的有效性，对比了原始 YOLO V3、YOLO V3-tiny 及改进后的 YOLO V3 的总体损失曲线，总体训练损失值越小也代表着模型更好的建立。针对定制式木门数据集进行训练的 YOLO V3-tiny、YOLO V3 以及空间金字塔池化层改进后的 YOLO V3 总体损失曲线如图 12 所示。

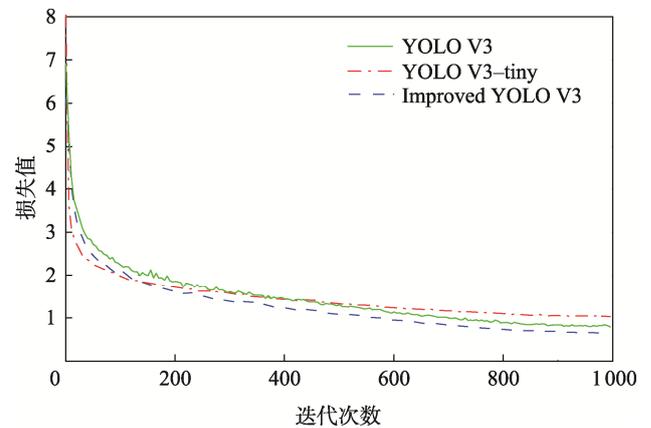


图 12 不同 YOLO V3 模型的总体损失曲线对比
Fig.12 Comparison of overall loss curves of different YOLO V3 models

由图 12 可知，YOLOV3-tiny 由于压缩了网络层数，只使用了 6 种不同锚框尺度的输出，最终总体损失稳定在 1.04 左右，与其他 2 种模型收敛效果仍有差距。接下来对比 2 种 YOLO V3 模型，改进后的 YOLO V3 模型表现出更快的收敛速度，最终损失为 0.66，相较于原始 YOLO V3 的损失（0.82）降低了 0.16，这一结果表明改进后的 YOLO V3 网络的模型收敛效果的提高。在平均检测准确率方面，改进后的 YOLO V3 算法的平均检测准确率提高到了 95.98%。

3.4 数据增强前后对比

图像增强技术将定制式木门数据集中的样本数量从 615 个增加到 3 075 个，使用上述实验确定的锚框尺度和超参数进行实验。实验过程中由于数据集中的批量大小和样本数量造成的模型训练时间大幅增加，从 23 h 增加至 112 h。以改进的 YOLO V3 训练模型在原始数据测试集和增强数据测试集进行测试，实验结果如表 5 所示。

表 5 定制式木门增强数据集检测结果对比
Tab.5 Comparison of test results of customized wooden door enhancement data set

数据集	木门种类	A_p 值/%	M_{ap} 值/%	交叉验证 M_{ap} 值/%
未增强数据集	古巴原木 (type 1)	96.54	95.98	96.74
	红檀木 (type 2)	96.77		
	黄柚木 (type 3)	93.94		
	红拼木 (type 4)	96.67		
已增强数据集	古巴原木 (type 1)	98.63	97.95	98.05
	红檀木 (type 2)	97.42		
	黄柚木 (type 3)	96.93		
	红拼木 (type 4)	98.01		

由表 5 可见, 通过增强数据集训练的模型提高了 4 种类型木门的检测准确率, 在五折交叉验证实验中总体检测准确率最终提高到了 98.05%, 增强了模型性能。在保证检测精度的基础上, 进一步比较了不同算法每张图像的平均检测时间, 实验结果如表 6 所示。

表 6 不同检测算法性能对比
Tab.6 Performance comparison of different detection algorithms

算法	M_{ap} 值/%	平均检测时间/ms
Faster RCNN	97.56	356
YOLO V3	95.93	112
YOLO V3-tiny	90.89	74
Improved YOLO V3	98.05	137

通过以上几种算法的对比, 双阶段目标检测法 Faster RCNN 虽然达到很高的检测准确率, 但检测所需时间较长, 不适合生产线实时采集的工件视频流分析, 相比之下 YOLO 系列算法的检测速度均能满足实时检测的要求。其中, YOLO V3-tiny 算法虽然计算开销低、检测速度快, 但不能满足生产线的准确率要求。综合考虑生产线的实时性要求, 本文算法的计算时间必须控制在 200 ms 以内。为保证生产线实际应用中, 检测速率达到条件的情况下取得尽量高的准确率, 选择改进后的 YOLO V3 算法, 平均检测时间达到 0.137 s/幅, 检测帧率达到 7 帧。木门数据检测结果如图 13 所示, 检测结果包括木门类型与被检测类型的置信度, 置信度的数值用来判断当前木门类型的概率。

由图 13 可见, 检测种类后的数字代表定制式木门检测置信度, 及判定为当前木门的可能性大小。图 13 中置信度均接近于 1, 代表改进后的 YOLO V3 模型准确检测了 4 种类型的木门。

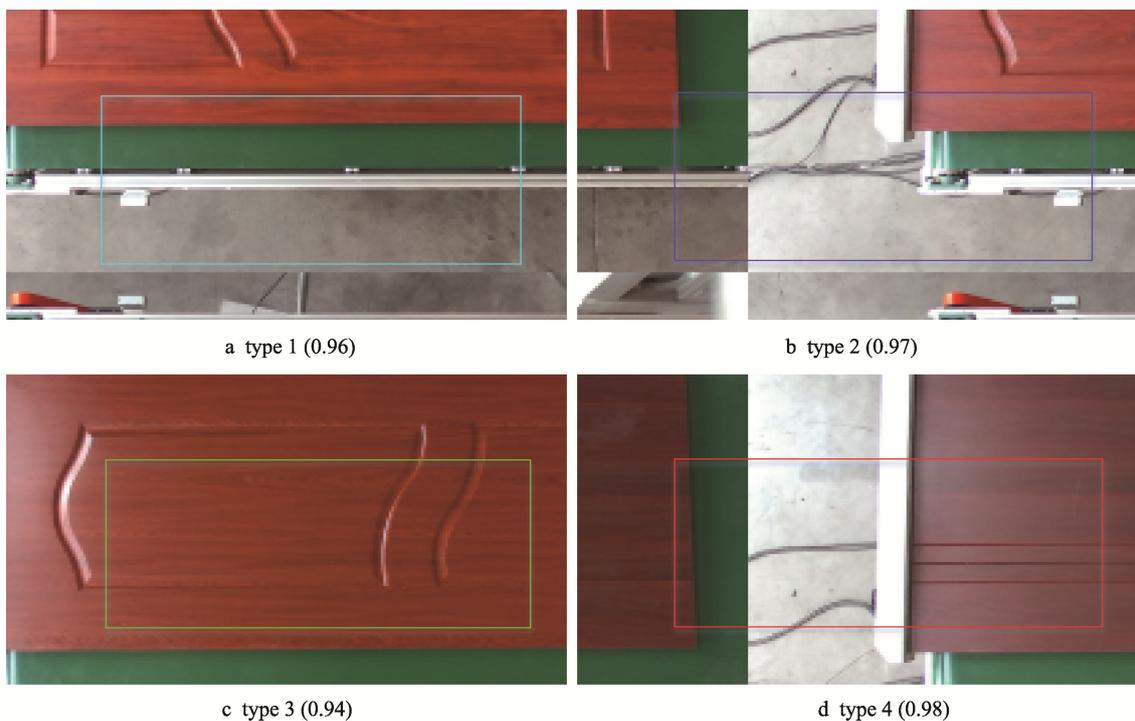


图 13 定制式木门测试集抽样检测情况
Fig.13 Sampling inspection of customized wooden door test set

4 结语

本文对定制式木门加工生产线的工件检测方法进行研究,提出一种基于卷积神经网络的工件检测方法,具体结论如下:

1) 考虑实际生产过程中出现的难点,进行机器视觉系统中相机的硬件选型及视觉系统布置方案的确定,使得对木门的识别视场范围更大,所获得的木门生产场景信息更加丰富。

2) 针对传统 YOLOV3 算法损失函数的缺陷,使用广义交并比损失函数替代原损失函数。由于原始锚框尺度不能很好地适应木门数据集的问题,使用 K -means 方法重新聚类锚框尺度并进行模型训练,进一步地使用空间金字塔池化网络改进算法结构。

3) 结果表明,改进后的算法具有更优的检测性能,空间金字塔池化网络进一步强化了定制式木门工件的特征传播,减少了训练过程中低语义信息到高语义信息的转化损失,提高了检测准确率,为辅助机器人实现自主决策、木门堆垛流程自动化提供了依据。

参考文献:

- [1] 顾聪聪,刘送永,姚遥,等.高速重载码垛机器人动力学仿真[J].包装工程,2021,42(7):186-194.
GU Cong-cong, LIU Song-yong, YAO Yao, et al. Dynamic Simulation of High Speed and Heavy Load Palletizing Robot[J]. Packaging Engineering, 2021, 42(7): 186-194.
- [2] 徐思源,储开斌,张继,等.面向小目标检测的改进 YOLOv3 算法[J].电光与控制,2022,29(8):35-39.
XU Si-yuan, CHU Kai-bin, ZHANG Ji, et al. An Improved YOLOv3 Algorithm for Small Target Detection[J]. Electronics Optics & Control, 2022, 29(8): 35-39.
- [3] Girshick R. Fast r-cnn[C]. Proceedings of the IEEE international conference on computer vision. Santiago: IEEE, 2015: 1440-1448.
- [4] REN Shao-qing, HE Kai-ming, 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.
- [5] 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: 779-788.
- [6] 陈宏彩,任亚恒,郝存明,等.基于 YOLOv3 的医药玻璃瓶缺陷检测方法[J].包装工程,2020,41(7):241-246.
CHEN Hong-cai, REN Ya-heng, HAO Cun-ming, et al. Defect Detection Method for Medical Glass Bottles Based on YOLOv3[J]. Packaging Engineering, 2020, 41(7): 241-246.
- [7] 简川霞,陈鑫,林浩,等.基于卷积神经网络的印刷套准识别方法[J].包装工程,2021,42(15):275-283.
JIAN Chuan-xia, CHEN Xin, LIN Hao, et al. Printing Registration Recognition Method Based on Convolutional Neural Network[J]. Packaging Engineering, 2021, 42(15): 275-283.
- [8] 李宜汀,谢庆生,黄海松,等.基于卷积神经网络快速区域标定的表面缺陷检测[J].计算机集成制造系统,2019,25(8):1897-1907.
LI Yi-ting, XIE Qing-sheng, HUANG Hai-song, et al. Surface Defect Detection Based on Fast Regions with Convolutional Neural Network[J]. Computer Integrated Manufacturing Systems, 2019, 25(8): 1897-1907.
- [9] 徐望明,刘召,伍世虔,等.基于虚拟多曝光融合的不均匀光照图像增强方法[J].华中科技大学学报(自然科学版),2020,48(8):79-84.
XU Wang-ming, LIU Zhao, WU Shi-qian, et al. Non-Uniform Illumination Image Enhancement Method Based on Virtual Multi-Exposure Fusion[J]. Journal of Huazhong University of Science and Technology (Nature Science Edition), 2020, 48(8): 79-84.
- [10] JIANG B, LUO R, MAO J, et al. Acquisition of Localization Confidence for Accurate Object Detection[C]// Proceedings of the European Conference on Computer Vision (ECCV), Munich: Springer, 2018: 784-799.
- [11] ZHANG L, SCHAEFFER H. Forward Stability of ResNet and Its Variants[J]. Journal of Mathematical Imaging and Vision, 2020, 62(3): 328-351.
- [12] KIM S, KOOK H, SUN J, et al. Parallel Feature Pyramid Network for Object Detection[C]// Proceedings of the European Conference on Computer Vision (ECCV), Munich: Springer, 2018: 234-250.
- [13] REZATOFI H, TSOI N, GWAK J, et al. Generalized Intersection over Union: A Metric and a Loss for Bounding Box Regression[C]// 2019 IEEE/CVF Confe-

- rence on Computer Vision and Pattern Recognition, Long Beach: IEEE, 2019: 658-666.
- [14] 陈金立, 王亚鹏, 李家强, 等. 冲击噪声下基于张量分解和 K -means 聚类的 MIMO 雷达阵列诊断[J]. 电子学报, 2021, 49(12): 2315-2322.
- CHEN Jin-li, WANG Ya-peng, LI Jia-qiang, et al. Tensor Decomposition and K -Means Clustering Based Array Diagnosis for MIMO Radar in Impulsive Noise Environment[J]. Acta Electronica Sinica, 2021, 49(12): 2315-2322.
- [15] RUSSAKOVSKY O, DENG JIA, SU HAO, et al. ImageNet Large Scale Visual Recognition Challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [16] SUTSKEVER I, MARTENS J, DAHL G, et al. On the Importance of Initialization and Momentum in Deep Learning[C]// International conference on machine learning: PMLR, 2013: 1139-1147.
- [17] KIM J, SONG J, LEE J. Recognizing and Classifying Unknown Object in BIM using 2D CNN[C]// International Conference on Computer-Aided Architectural Design Futures, Singapore: Springer, 2019: 47-57.

责任编辑: 曾钰婵