

DOI: 10.19650/j.cnki.cjsi.J2006768

基于注意力机制的曲轴瓦盖上料机器人 视觉定位和检测方法*

朱江¹, 杜瑞^{1,3}, 李建奇^{1,2,3}, 蔡慕尧¹, 许海霞¹

(1. 湘潭大学自动化与电子信息学院 湘潭 411105; 2. 湖南文理学院计算机与电气工程学院 常德 415000;
3. 分布式电推进飞行器控制技术湖南省重点实验室 常德 415000)

摘要:为了解决曲轴瓦盖人工上料效率低下、易出错的难题,研究了基于注意力机制的曲轴瓦盖上料机器人视觉定位和检测方法,实现自动上料。针对图像特征不明显,在 Faster R-CNN 的特征提取网络引入注意力机制,将曲轴瓦盖图像不同位置的权重映射到特征通道,使深度学习模型能够更多地关注曲轴瓦盖的边缘和中心语义信息。为进一步提高定位精度,本文还改进了候选框生成方法和损失函数。实验结果表明,与传统机器学习方法及经典深度学习目标检测模型相比,检测速度达 0.419 s,定位精度最优(IOU 和 GIOU 分别为 0.941 3 和 0.940 9)。该方法还具有良好的鲁棒性。现场测试表明,该方法引导上料机器人抓取和放置曲轴瓦盖组的成功率达 95.14%,提升了发动机装配生产线的效率。

关键词:视觉引导机器人;注意力机制;目标检测;深度学习

中图分类号: TP391.4 TH89 **文献标识码:** A **国家标准学科分类代码:** 510.80

Visual location and detection method of crankshaft bearing cap feeding robot based on attention mechanism

Zhu Jiang¹, Du Rui^{1,3}, Li Jianqi^{1,2,3}, Cai Muyao¹, Xu Haixia¹

(1. College of Automation and Electronics Information, Xiangtan University, Xiangtan 411105, China;

2. School of Computing and Electrical Engineering, Hunan University of Arts and Science, Changde 415000, China;

3. Hunan Province Key Laboratory for Control Technology of Distributed Electric Propulsion Aircraft, Changde 415000, China)

Abstract: To solve the problem of low efficiency and error prone of manual feeding of the crankshaft bearing caps (CBCs), the visual location and detection method of CBC feeding robot based on attention mechanism is studied to realize automatic feeding. Aiming at the unapparent image features, the attention mechanism is introduced into the feature extraction network of Faster R-CNN to map the weights of the CBC image at different positions to the feature channel, so that the deep learning model can pay more attention to the edge and center semantic information of the CBC. To further improve the location accuracy, this paper also improves the candidate box generation method and loss function. Experiment results show that compared with those of traditional machine learning methods and classic deep learning target detection models, the detection speed of this method reaches 0.419s, the location accuracies are the best (IOU and GIOU are 0.941 3 and 0.940 9, respectively). In addition, the proposed method possesses good robustness. On site test shows that the success rate for the method guide the feeding robot to grasp and place the CBCs reaches 95.14%, which improves the efficiency of the engine assembly line.

Keywords: vision guided robot; attention mechanism; object detection; deep learning

收稿日期:2020-08-10 Received Date:2020-08-10

* 基金项目:湖南省自然科学基金(2020JJ6063)、湖南省教育厅科学研究项目(18A360,19B393)、湖南省高等学校省特色学科(湘教通[2018]469)资助



0 引言

曲轴瓦盖(crankshaft bearing cap, CBC)是汽车发动机的一种关键零部件,用于紧固曲轴和发动机气缸。一台汽车发动机需要在不同的位置装配多个曲轴瓦盖。例如,一台四缸发动机需要配备5个曲轴瓦盖。由于发动机的特殊结构,安装于发动机上不同位置的曲轴瓦盖在尺寸和结构方面略有不同。为了提高装配的效率,在零部件准备阶段,根据组装一台发动机所需的曲轴瓦盖类型,将若干个不同型号的曲轴瓦盖组合在一起,分组放置在料框中。值得注意的是,每一组内的曲轴瓦盖需按照各自在发动机上装配的位置关系,从左至右顺序摆放。在装配阶段,工人从料框取出一组曲轴瓦盖放置在曲轴瓦盖压装机的的工作台上,人工操作压装机按顺序把单个曲轴瓦盖安装到发动机的对应位置。为了确保每种曲轴瓦盖都能安装在发动机的正确位置,装配工人从料框中取料时,必须对当前一组曲轴瓦盖从3个方面检查:1)整组瓦盖是否被保护用的薄膜全部或部分覆盖(即遮挡),如有遮挡,需手动揭开覆盖的薄膜。2)整组曲轴瓦盖的放置方向,对于正向放置的整组曲轴瓦盖,其方向保持不变;否则,将整组曲轴瓦盖的方向旋转 180° 。3)该组曲轴瓦盖中的各个曲轴瓦盖是否按照装配顺序依次摆放。装配时,一旦未能按正确的方向或顺序装配曲轴瓦盖,会引起发动机故障,甚至将导致发动机报废。

随着汽车生产的规模化,传统人工上料方式由于劳动强度大、效率低和出错率高,无法满足汽车制造的需要。配置视觉的机器人作为非常有前景和价值的自动检测和装配技术,已成功应用于一些行业的产品质量控制^[1-2]或产品零件组装^[3]。因此,采用视觉引导机器人代替装配工人,自动从料框中抓取一组曲轴瓦盖,放置到曲轴瓦盖压装机的的工作台上指定位置成为一种可行的方案。显然,在图像中精确定位整组曲轴瓦盖在料框中的位置,准确检测整组曲轴瓦盖是否被遮挡,整组曲轴瓦盖放置的方向和组内曲轴瓦盖的顺序异常是机器人上料的关键环节。高赢等^[4-5]采用模板匹配的方法定位曲轴瓦盖的位置,并提出了一种结合工件的先验信息和工件上几何基元位置的检测算法识别单个工件摆放错误和整体倒置。然而,该方法对遮挡的曲轴瓦盖的定位和检测效果不佳。

类似的工作可参考金属表面图像中的故障检测,例如铁路接触电网、铁路油罐车、铁路轨道、铁路列车车轮等。模板匹配作为一种简单有效的检测方法,被用于铁路接触电网中电线的悬链线组件检测^[6-7]。然而,悬链线组件容易变形,其形态各式各样。由于它们的形状不同,可能会导致非唯一的特征。Xiang等^[8]提

出了一种形状匹配的改进方法来定位铁路油罐车的开口,解决了特征不唯一的问题。但是,其匹配结果很大程度上取决于所选的匹配模板。近年来,越来越多的研究人员致力于基于机器学习的视觉检测方法。Harr^[9],定向直方图^[10]和局部二进制模式^[11]已被广泛用于提取对象的特征,然后在特征基础上训练支持向量机(support vector machine, SVM)等各种分类器实现目标的检测定位。Stella等^[12]提出了一种使用小波变换、主成份分析和神经网络分类器来检测铁路轨道中缺少紧固件的方法。Mazzeo等^[13]应用AdaBoost算法检测铁轨上的紧固件。为了检查货运列车缺少的转向架块,Liu等^[14]提出结合提取梯度编码共生矩阵特征和SVM进行故障区域实时定位的方法。但是,此类方法需要人工设计提取特征,算法的鲁棒性不强,特别是容易受光照不均等环境因素的影响,且分类器的性能也会影响检测的性能。而且,对于曲轴瓦盖的上料应用,这些方法不能同时得到定位与检测结果,需要在定位的基础上进一步设计算法检测曲轴瓦盖的放置方向和摆放顺序异常。近年来,深度卷积神经网络由于其出色的性能而受到关注。其中,Faster R-CNN^[15],YOLO^[16]和SSD^[17]是深度学习目标检测的代表性算法。但是,这些现有的深度学习缺陷检测算法侧重点在于目标的分类,对于目标的具体位置只需要给出粗略的候选框即可。而上料机器人需要获取曲轴瓦盖在图像坐标系下的精确位置才能准确、可靠地抓取。

针对曲轴瓦盖自动装配的视觉引导机器人上料,面临的难点总结如下:

1) 高精度定位

发动机和曲轴瓦盖作为精密的零部件,对装配的精度有极高的要求。机器人需要准确地确定曲轴瓦盖组在当前图像中的位置,在经过坐标转换后,得到世界坐标系下曲轴瓦盖在料框中位置,其执行机构才能可靠抓取曲轴瓦盖组到指定位置。

2) 实时定位和检测

作为工业应用,最终的目标是提高生产的效率。因此,要求上料机器人能够尽可能快地确定曲轴瓦盖在图像中的坐标,检测整组曲轴瓦盖是否被遮挡、整组曲轴瓦盖放置的方向和组内曲轴瓦盖的顺序异常。

3) 遮挡的目标

料框的表面预先覆盖了一层薄膜用于保护曲轴瓦盖。料框运送到装配线时,由于工人操作失误或疏忽,薄膜可能未完全被掀开至料框之外,在料框边缘附近的个别曲轴瓦盖组可能全部或部分被薄膜覆盖。因此,需要检测曲轴瓦盖组是否被遮挡和定位其位置,启用气枪喷气吹开薄膜。图像中被遮挡的曲轴瓦盖区域与背景的分度不高,给定位和识别带来了困难。

4) 小样本和样本不平衡

生产过程中产生的组内顺序错误的样本相比其他类别少,存在类别不平衡问题。同时由于样本采集困难和手动标注耗时,整体曲轴瓦盖属于小样本数据。

针对上述问题,本文提出一种基于注意力机制的曲轴瓦盖上料机器人视觉定位和检测深度学习方法,将注意力机制引入 Faster R-CNN 中,构建了新的两阶段检测模型,并改进了候选框的生成方式和损失函数,实现高精度、实时的定位与检测。训练阶段,首先构建曲轴瓦盖图像数据集。然后,将数据集图像用于训练提出的深度模型。在应用阶段,将获取的曲轴瓦盖图像输入训练好的模型推理,输出图像中曲轴瓦盖的定位坐标、曲轴瓦盖组放置的方向和顺序,以及是否有遮挡,实现引导机器人完成曲轴瓦盖的自动上料。

1 瓦盖图像定位和检测算法

本文的定位和检测的对象如图 1 所示。本文以更快的区域卷积神经网络 (faster region-convolutional neural network, Faster R-CNN) 作为基本结构,提出一种基于注意力机制的两阶段检测器。如图 2 所示,该模型包括特征提取网络、候选区域生成网络 (region proposal network, RPN) 和由 Fast R-CNN 构成的分类回归网络三部分。其中,注意力网络位于特征提取网络中,用于将曲轴瓦盖图像的不同位置的权重映射到特征通道,使模型能更多地关注曲轴瓦盖的边缘和中心语义信息。

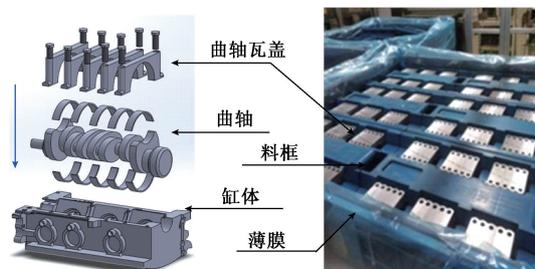


图 1 汽车发动机曲轴瓦盖

Fig. 1 Automobile engine crankshaft bearing caps

1.1 基于注意力机制的特征提取网络

1) 基于注意力机制的特征提取网络

在深度学习目标检测模型中,图像经过卷积操作后的输出通常被看作是特征映射。与处理前的图像相比,卷积网络输出的特征图有着更好的分类和定位性能。因此,选取合适的特征提取网络对于深度模型的性能至关重要。典型的 Faster R-CNN 目标检测模型使用 VGG16 作为特征提取网络。为了能够使提取的特征能更好地表征曲轴瓦盖组图像,本文在特征提取网络中加入了注意

力机制。

本文模型输入为 $1 \times 1\,000 \times 600$ 的单通道图像。在第 1 阶段,特征提取网络由 13 个卷积层和 6 个注意层组成,VGG16 用作基本的特征提取器,提取曲轴瓦盖图像的特征。在第 2 阶段,由 RPN 生成候选区域和感兴趣区域合并层。然后,将 RoI Pooling 层连接到两个并行的完全连接层,分别对应一个 4 个神经元的回归结构(即生成曲轴瓦盖组在图像中的四个坐标)和一个 4 个神经元的分类结构(即曲轴瓦盖组的四种状态——整组正向、整组反向、组内顺序异常和被遮挡)。表 1 给出了本文提出的特征提取网络的结构及其相关参数。

表 1 特征提取模型参数

Table 1 Feature extraction model parameters

网络层	输入大小	核大小	输出大小
CONV1_1	$1 \times 1\,000 \times 600$	3×3	$64 \times 1\,000 \times 600$
CONV1_2	$64 \times 1\,000 \times 600$	3×3	$64 \times 1\,000 \times 600$
CONV2_1	$64 \times 500 \times 300$	3×3	$128 \times 500 \times 300$
CONV2_2	$128 \times 500 \times 300$	3×3	$128 \times 500 \times 300$
CONV3_1	$128 \times 500 \times 300$	3×3	$256 \times 250 \times 150$
CONV3_2	$256 \times 250 \times 150$	3×3	$256 \times 250 \times 150$
CONV3_3	$256 \times 250 \times 150$	3×3	$256 \times 250 \times 150$
CONV4_1	$256 \times 125 \times 75$	3×3	$512 \times 125 \times 75$
CONV4_2	$512 \times 125 \times 75$	3×3	$512 \times 125 \times 75$
CONV4_3	$512 \times 125 \times 75$	3×3	$512 \times 125 \times 75$
GP1	$512 \times 125 \times 75$	2×2	$512 \times 62 \times 37$
FC4	$1\,174\,528 \times 1 \times 1$	1×1	$4\,096 \times 1 \times 1$
ATT1	$4\,096 \times 1 \times 1$	1×1	$512 \times 1 \times 1$
CONV5_1	$512 \times 62 \times 37$	3×3	$512 \times 62 \times 37$
CONV5_2	$512 \times 62 \times 37$	3×3	$512 \times 62 \times 37$
CONV5_3	$512 \times 62 \times 37$	3×3	$512 \times 62 \times 37$
GP2	$512 \times 62 \times 37$	2×2	$512 \times 31 \times 18$
FC5	$285\,696 \times 1 \times 1$	1×1	$4\,096 \times 1 \times 1$
ATT2	$4\,096 \times 1 \times 1$	1×1	$512 \times 1 \times 1$
FC6	$1\,174\,528 \times 1 \times 1$	1×1	$4\,096 \times 1 \times 1$
FC7	$4\,096 \times 1 \times 1$	1×1	$4\,096 \times 1 \times 1$

不同于标准的 Faster R-CNN 模型,为了使提取的特征图能更准确地生成候选区域,本文使用注意力机制对高维特征通道 CONV4_3 层和 CONV5_3 层的多尺度特征进行加权,图 2 中的实线框即为注意力网络在本文提出的网络结构中的位置。即在标准特征提取网络 VGG16 的基础上,增加了两组新的全局池化层和完全连接层,生成长度均为 512 的注意力权重向量。通过此向量对高维

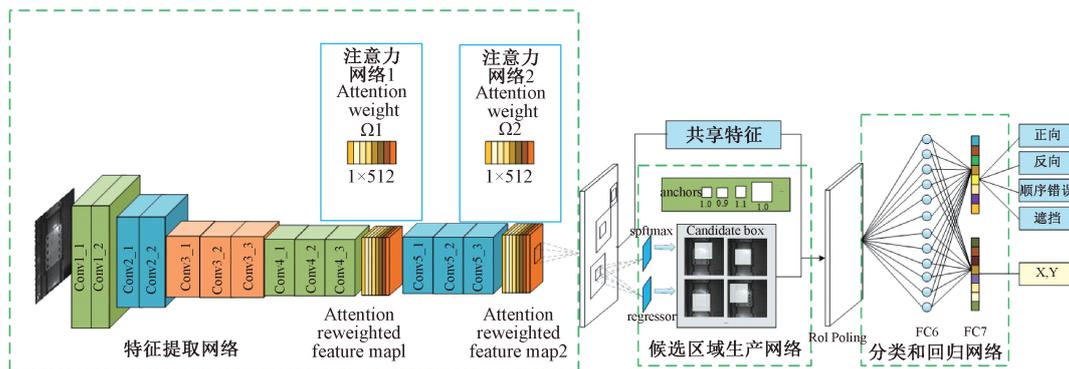


图 2 基于注意力机制的改进 Faster R-CNN

Fig. 2 The improved Faster R-CNN based on attention mechanism

特征通道(CONV4_3 层或 CONV5_3 层)的特征图加权,将其结果输入到各自相邻的下一层。

在特征提取网络中加入注意力网络后,特征图中的不同特征通道能够分配到不同的权重。因此,映射到原始图像中相对重要的位置(如曲轴瓦盖的边缘和油墨标记线条)的特征通道具有较高的权重。与直接在原始图像上加权相比,本文方法需要的先验信息较少。通过在 CONV4_3 和 CONV5_3 上对两个不同尺度的特征图进行加权使得模型更注重细节,原理如下:

$$f_{att}(n) = \Omega_n^T f_{chn} \quad (1)$$

式中: $f_{att}(n)$ 为注意力网络产生的高维特征图; n 表示从第 1 个输入开始的卷积层数,本文取值 4 或 5; f_{chn} 表示由 CONV4_3 或者 CONV5_3 输出的特征图; Ω_n^T 是相应层的注意力权重。权重与原始图像权重之间的映射关系为:

$$\Omega = F(G^T) \quad (2)$$

式中: F 是从原始图像权重到注意力网络中关注通道的映射; G^T 是原始图像的加权建议。为了获得通道权重,训练阶段直接将注意力网络添加到标准 VGG16 特征提取网络中训练。本文提出的注意力网络如图 3 所示,通过全局池化和完全连接,生成 512 个权重项,对应于 CONV4_3 和 CONV5_3 层的特征通道数。GP 代表全局池化层,FC 代表完全连接层。

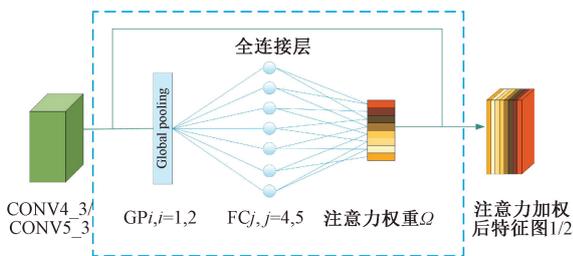


图 3 注意力网络结构

Fig. 3 Attention network structure

2) 候选区域生成网络

RPN 将原始图像上生成的候选框映射到 Attention net2 输出的注意力加权特征图上生成候选框。RPN 是分类和回归的并行结构,回归结构生成具有 4 个坐标的候选框;分类结构确定候选框中对象是背景还是目标。若是背景,丢弃候选框。否则,保留候选框并映射到先前的高维特征图。

在生成候选框过程中,为了能准确地检测到具有不同形状和大小的目标,标准的 Faster R-CNN 模型设置宽高比为 0.5、1.0 和 2.0 的 3 种候选框。由于单组曲轴瓦盖的长宽比接近 1,因此,本文将候选框的长宽比改为 0.9、1.0 和 1.1,从而能够获得更准确的定位结果。同时,本文根据候选框大小筛选,去除与曲轴瓦盖组偏差大的候选框,有效地减小了冗余计算。候选框生成示意图如图 4 所示。

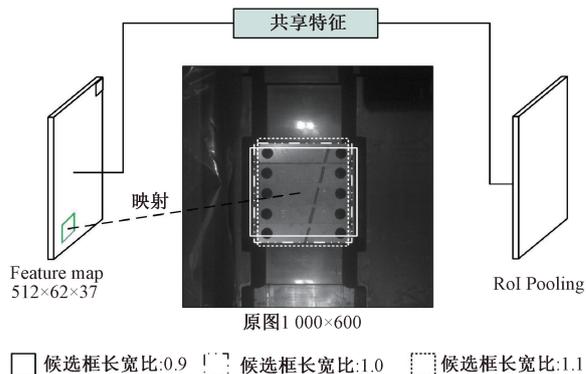


图 4 候选框生成示意图

Fig. 4 Candidate box generation schematic diagram

第 2 阶段为检测阶段,RPN 与 Fast R-CNN 共享特征,通过 Fast R-CNN 在此高维特征图上对产生的候选框微调,输出最后的检测和定位结果。

回归网络获得曲轴瓦盖组在图像中的坐标,而分类网络将曲轴瓦盖组分为曲轴瓦盖组正向、曲轴瓦盖组反向、曲轴瓦盖组内顺序异常和曲轴瓦盖组被遮挡 4 种类



型。标准的 Faster R-CNN 网络中,一旦候选框交并比大于阈值,则将候选框视为目标。由于实际应用中,每幅图像中只有一组完整的曲轴瓦盖,因此本文的网络仅保留交并比最高的候选框作为最终检测结果。本网络输出的矩形框的4个坐标为曲轴瓦盖组的左上角和右下角的水平垂直坐标。

1.2 损失函数

本文的损失函数分为 RPN 的损失函数和 Fast R-CNN 的损失函数,分别与候选框的生成性能和获取的最终定位结果相关。它们的损失函数具有相同的结构,均由分类损失和回归损失组成,表示为:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (3)$$

式中: p_i, p_i^* 分别表示被预测为目标概率和被预测为非目标的概率; t_i, t_i^* 分别表示真实值和预测值的矩形框。

式(3)的第1项为分类损失。RPN 主要任务之一是区分候选框内区域是否为待检测的目标。本文正类别为曲轴瓦盖,负类别为图像的背景。在训练阶段,RPN 网络的分类损失参数 N_{cls} 取 256,即采用 256 个矩形框来生成候选框。Fast R-CNN 的主要任务是输出候选框内区域的最终类别。训练时,将计算网络预测的不同类别之间的损失,该部分分类损失参数 N_{cls} 为 128。

式(3)的第2项为回归损失,用于拟合预测框和真实框之间的差距。 N_{reg} 为正负样本的总数, λ 设置为 1。传统的两级检测器中,回归损失函数通常使用 Smooth L1 函数^[18]。然而,以该损失函数训练时可能会存在局部最优解,无法得到最优解^[16]。为此,本文将表示真实矩形框和预测矩形框的广义交并比(generalized intersection over union, GIOU)^[19]作为回归损失函数:

$$L_{reg}(t_i, t_i^*) = 1 - GIOU(t_i, t_i^*) \quad (4)$$

与交并比(intersection over union, IOU)^[20]相比,GIOU 通过反映预测值和真实值之间的距离更精确地表示他们之间的差距,定义如下:

$$GIOU = IOU - \frac{A^c - U}{A^c} \quad (5)$$

式中:IOU 为交并比,表示目标检测中真实框和预测框交集和并集之比:

$$IOU = \frac{REC_1 \cap REC_2}{REC_1 \cup REC_2} \quad (6)$$

U 为两个矩形框的交集, A^c 代表最小闭包区域:

$$A^c = (x_2^c - x_1^c) \times (y_2^c - y_1^c) \quad (7)$$

其中,

$$x_1^c = \min(\hat{x}_1^p, \hat{x}_1^g), x_2^c = \max(\hat{x}_2^p, \hat{x}_2^g) \quad (8)$$

$$y_1^c = \min(\hat{y}_1^p, \hat{y}_1^g), y_2^c = \max(\hat{y}_2^p, \hat{y}_2^g)$$

\hat{x}_1^p 表示预测矩形框的最小横坐标, x_1^g 表示真实值矩形框的最小横坐标,以此类推,可以分别获得矩形框预测值和真实值的最大横坐标、最小纵坐标和最大纵坐标。

综上所述,本文的损失函数由 RPN 分类损失、RPN 回归损失、Fast R-CNN 分类损失和 Fast R-CNN 回归损失求和得到,且各个损失的权重相同。对于本文的应用,RPN 生成候选框的过程既要调整曲轴瓦盖目标的候选框定位,还要对该候选框目标进行曲轴瓦盖组或背景的二分类。故本文对 RPN 和 Fast R-CNN 的分类和回归损失采用同样的权重。

2 实验与分析

为了验证本文提出的深度学习模型性能,本文将该模型与传统的机器学习方法、当前主流的目标检测深度学习模型 Faster R-CNN、SSD、YOLO 和 YOLO v3 从定位精度、运行速度和鲁棒性等方面对比。执行深度学习模型的配置如下:Tensorflow 1.8.0, Ubuntu 16.04, Intel® Xeon® Gold 6130 CPU @ 2.10 GHz, Tesla T4 GPU 和 192 GB 内存。CUDA 和 CUDNN 的版本分别为 9.0 和 7.1.2。传统机器学习方法在 OpenCV 3.0 下实现。

2.1 数据集构建

本文构建的数据集共有 906 张曲轴瓦盖图像,分为 4 类:216 张整组曲轴瓦盖顺序正常且正向图像;290 张整组曲轴瓦盖顺序正常且反向的图像;100 张曲轴瓦盖组内顺序异常的图像;300 张曲轴瓦盖被部分遮挡的图像。其中,组内顺序异常且反向和组内顺序异常且正向图像各有 40 张和 60 张。4 种典型样本如图 5 所示。

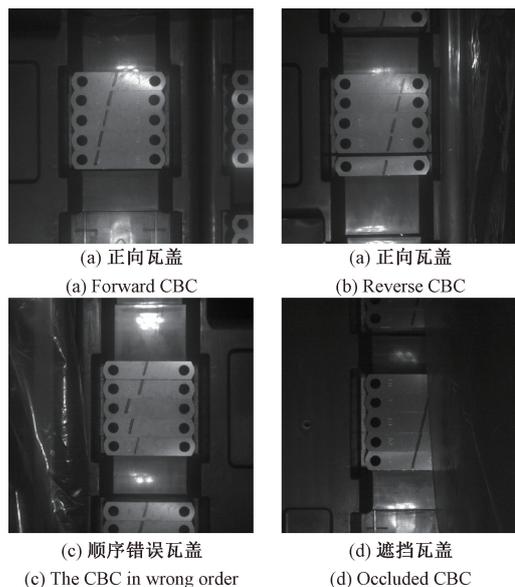


图5 4种瓦盖类型

Fig. 5 Four types of the CBCs

首先,对每一幅图像标记整组曲轴瓦盖的方向、被遮挡情况和组内顺序,整组曲轴瓦盖在图像中的最小水平坐标、最大水平坐标、最小垂直坐标和最大垂直坐标等信息。由于样本的数量太少,训练时本文先使用 Imagenet 数据集的预训练模型对特征提取器微调。该方法克服了深度学习需要大量样本的问题,模型能以较短的训练时间获得较好的效果。

2.2 放置方向和顺序异常的检测

本文模型对 4 类典型样本的推理结果如图 6 所示。

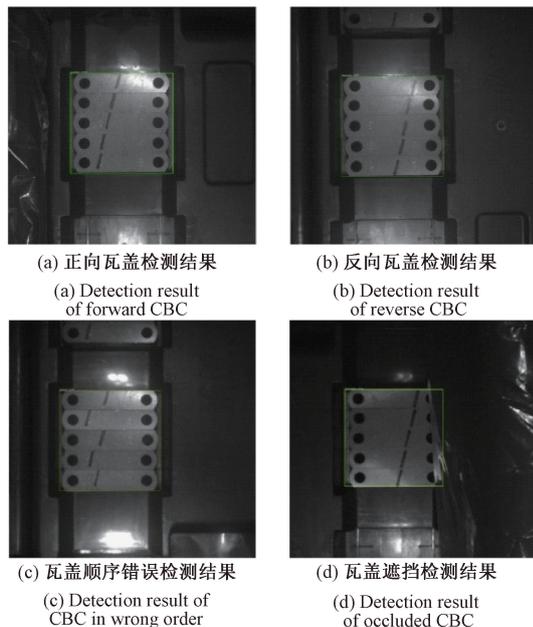


图 6 检测的视觉结果

Fig. 6 Visual results of the detection

本文采用精度和召回率评估模型的检测性能,检测的定量结果如表 2 所示,所有的方法的性能相当,几乎都可以准确地检测曲轴瓦盖在图像中的放置方向、组内顺序和遮挡情况。多次实验下,每个模型中只有极个别图像出现分类错误。文献[4-5]对正放和倒置的曲轴瓦盖的识别率均为 100%,曲轴瓦盖组内顺序异常且正向和曲轴瓦盖组内顺序异常且反向的识别率分别为 99.77% 和 99.54%。但是,该方法仅仅考虑了前 3 类样本情况,对于第 4 类样本一曲轴瓦盖被部分遮挡并未考虑。因此,对于本文的包含了遮挡的测试集,该方法将曲轴瓦盖被部分遮挡误检为其它类型,导致准确率和召回率急剧下降。

表 2 检测结果指标

Table 2 Detection result index

方法	正向准确率	正向召回率	反向准确率	反向召回率
本文	0.990 0	0.990 0	0.990 0	0.990 0
Faster R-CNN	0.990 0	0.990 0	0.990 0	0.990 0
文献[4-5]	0.666 7	0.600 0	0.581 3	0.683 5

2.3 定位的精度

本文采用 IOU 和 GIoU 定量评估本文的定位性能。本文不仅将提出的模型与经典的目标检测深度学习模型比对,而且为体现注意力机制对模型的定位性能影响,进一步对比了无注意力机制、只在 CONV5_3 层中使用注意力机制(称为单尺度注意力)、在 CONV4_3 和 CONV5_3 层中都使用注意力机制(称为多尺度注意)的定位精度,定位结果如表 3 所示。

表 3 无数据扩充定位精度

Table 3 Location accuracy without data augmentation

方法	IOU	GIoU
无注意力机制	0.929 2	0.928 4
单尺度注意力	0.938 0	0.937 6
多尺度注意力	0.938 9	0.938 4
Faster R-CNN	0.920 5	0.920 1
SSD	0.913 6	0.912 9
YOLO	0.891 4	0.891 1
YOLOv3	0.925 6	0.925 2
Harr +ADABOOST	0.824 8	0.823 9
Hog + SVM	0.871 3	0.870 9

结果表明,本文根据曲轴瓦盖图像的特点改进了损失函数和生成候选框的方法,即使没有增加注意力机制的网络的定位精度也比标准 Faster R-CNN 的定位精度更优。与单阶段检测的 SSD、YOLO 和 YOLOv3 网络相比,本文采用两级检测来生成候选框,因此具有较高的准确性。此外,具有注意力机制的网络的定位精度优于没有注意力机制的网络;多尺度注意力网络比单尺度注意力网络的定位结果更好。本文方法还与典型的“特征提取+分类器”式机器学习方法 Harr+ADABOOST 和 Hog+SVM 做了对比。分别采用 Harr 和 Hog 提取图像的特征,然后分别训练分类器。结果表明,本文的定位精度也优于这两种方法。

2.4 数据扩充对定位精度的影响

单组曲轴瓦盖组内顺序错误的图像只有 100 张,与其他类别相比存在类别不平衡。为此,本文通过对其旋转 180°,将其扩充到 200 张。然后,先按 0.2 的比例随机划分出测试集,剩下的图像中按照 7:3 的比例随机划分训练集和验证集。扩充之后的训练集、验证集和测试集分别为 562、242 和 202,其分布如表 4 所示。

利用扩充后的数据集,重新训练了各个模型,并将测试集送入训练的模型,定位结果如表 5 所示。与表 3 相比,扩充单组曲轴瓦盖组内顺序错误的样本,本文的定位精度有一定的提升。其他模型定位精度也相应提升,但



各深度学习模型之间的结果精度大小关系保持不变,本文提出的模型依然最佳。样本的扩充,对 Harr + ADABOOST 和 Hog+SVM 的定位精度未见提升。

表4 数据集分布情况
Table 4 Dataset distribution

类别	训练集	验证集	测试集
整组曲轴瓦盖正向	117	49	50
整组曲轴瓦盖反向	155	80	55
曲轴瓦盖组内顺序异常	122	41	37
被遮挡瓦盖	168	72	60
总计	562	242	202

表5 数据扩充后定位精度
Table 5 Location accuracy after data augmentation

方法	IOU	GIOU	是否提升
无注意力机制	0.936 9	0.936 4	↑
单尺度注意力	0.940 1	0.939 6	↑
多尺度注意力	0.941 3	0.940 9	↑
Faster R-CNN	0.923 8	0.923 0	↑
SSD	0.918 5	0.917 7	↑
YOLO	0.897 7	0.895 3	↑
YOLOv3	0.930 7	0.930 5	↑
Harr + ADABOOST	0.824 8	0.823 9	-
Hog + SVM	0.871 3	0.870 9	-

2.5 算法的实时性

表6所示为本文算法与经典目标检测深度学习模型的训练时间和测试时间。传统机器学习方法的运行时间是本文方法消耗时间的3~4倍。尽管在所有的深度学习方法中,本文的总训练时间和单次测试时间都最长,但是最大的差别在于训练模型的耗时。本文提出的模型为两阶段目标检测模型,与单阶段检测模型(如YOLO)相比,消耗时间长是两阶段目标检测模型的固有问题。需要说明的是,尽管本文方法单次的测试时间在深度学习方法中最长,大约是单阶段检测模型中耗时最小的YOLO检测时间的1.5倍,仅比同是两阶段检测模型的Faster RCNN多0.01s,且在0.5s之内。相对机器人执行机构的移动、抓取等动作耗时范围在几秒至十几秒来说,处于可接受范围之内,且拥有最高的定位精度。

2.6 算法的鲁棒性

实际的工业应用中,由于现场环境的变化或设备长时间运行后老化,获取的曲轴瓦盖图像中可能会出现噪声的干扰、光源老化后光照不均或照明的亮度变化等新情况。尽管采集到的曲轴瓦盖图像包含了上述噪声情况,在深度模型训练时增加了监督。但是,为了应对实际

表6 训练和测试时间
Table 6 Training and testing time

方法	总训练时间/s	单张测试时间/s
本文方法	13 822	0.419
Faster R-CNN	12 915	0.408
SSD	4 968	0.277
YOLO	3 694	0.225
YOLOv3	5 177	0.253
Harr+ADABOOST	22 360	1.415
Hog+SVM	12	1.546

生产中可能产生新的无监督噪声,需要验证本文模型的鲁棒性,本文从之前的数据集中任意抽取了51张图像随机引入两组强弱不同但均远比当前现场条件恶劣的噪声。其中强噪声组为椒盐比例为0.1的椒盐噪声和均值为0、方差为0.02的高斯噪声。弱噪声组为椒盐比例为0.05的椒盐噪声和均值为0、方差为0.01的高斯噪声。在此基础上或改变光照的强度,或同时加入噪声和改变光照强度,模拟实际工业生产中可能发生的情况。

得到的新图像样本,送入到之前训练好的模型中重新评估定位性能。与之前的实验相同,仍然用图像的IOU和GIOU作为衡量定位精度的依据。经10次实验平均得到原样本图像、加噪的图像和加大亮度的图像的定位结果如图7和表7所示。图7中,第1行从左至右为椒盐噪声样本(样本一)、高斯噪声样本(样本二)、椒盐+高斯噪声样本(样本三)、光照变化样本(样本四)和光照+椒盐+高斯噪声样本(复杂样本);第2行从左至右为上述样本图像对应的定位结果。

表7 抗干扰实验
Table 7 Anti-interference experiment

场景	IOU	GIOU	GIOU下降百分比/%
原始图片	0.941 3	0.940 9	-
椒盐噪声(弱)	0.924 6	0.924 4	1.65
椒盐噪声(强)	0.918 5	0.918 2	2.27
高斯噪声(弱)	0.936 3	0.936 1	0.48
高斯噪声(强)	0.932 1	0.931 9	0.90
椒盐噪声(弱)+高斯噪声(弱)	0.924 4	0.924 4	1.65
椒盐噪声(强)+高斯噪声(强)	0.916 9	0.916 6	2.43
光照变化	0.918 2	0.917 8	2.31
光照变化+椒盐噪声(弱)+高斯噪声(弱)	0.903 4	0.903 3	3.76
光照变化+椒盐噪声(强)+高斯噪声(强)	0.897 3	0.896 8	4.41

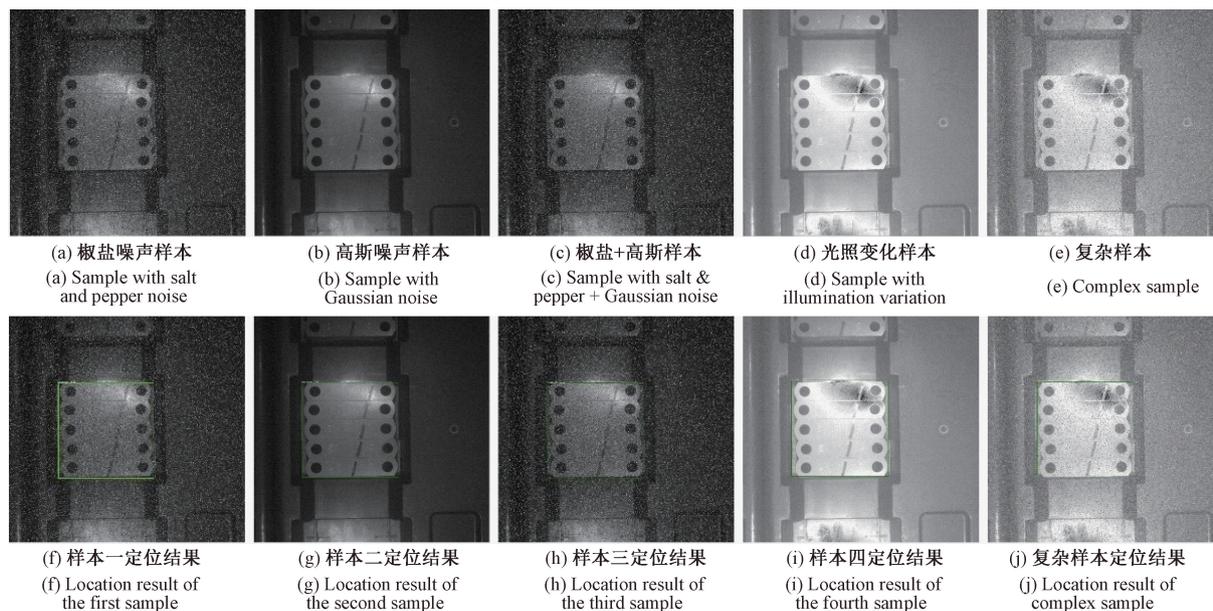


图 7 抗干扰实验视觉结果

Fig. 7 Visual results of anti-interference experiment

由图 7 和表 7 可知,虽然相对于原始图像,引入的噪声和光照的变化会引起定位精度一定的下降,但是仍然在可接受的范围之内。上述实验中,直接将加噪图像和改变光照的图像输入之前训练好的模型中,并没有将这些图像作为训练样本重新训练模型,在此情况下,定位的精度(GIOU)变化最多的也仅仅下降 4.41%,说明本文模型具有较好的环境适应性。

接下来,本文方法与其他经典目标检测深度学习模型比较鲁棒性。对于引入光照变化+椒盐噪声+高斯噪声等干扰的数据集,在同等条件下本文方法与经典目标检测深度学习模型经 10 次实验,得到该噪声环境下平均定位精度如表 8 所示。本文提出的方法在噪声条件下仍然拥有最高的定位精度,而且 FasterR-CNN、SSD 和 YOLOv3 在此情况下的定位精度相对于原始图像中的定位精度,其下降程度也大于本文提出的方法。此外,YOLO 甚至一次都未能成功定位目标。

表 8 与其他模型的鲁棒性对比

Table 8 Comparison of robustness with other models

方法	原 IOU	原 GIOU	噪声 IOU	噪声 GIOU	GIOU 下降百分比/%
本文方法	0.941 3	0.940 9	0.897 3	0.896 8	4.41
Faster R-CNN	0.923 8	0.923 0	0.874 2	0.874 0	4.90
SSD	0.918 5	0.917 7	0.857 3	0.856 9	6.08
YOLO	0.897 7	0.895 3	0	0	89.53
YOLOv3	0.930 7	0.930 5	0.881 4	0.881 2	4.93

3 基于视觉引导的曲轴瓦盖上料机器人

曲轴瓦盖上料机器人以 M-20iA 发那科六轴机器人为原型,在其手臂上定制加装了 LED 光源、分辨率为 1 598×1 550 的康耐视工业相机和末端执行机构。末端执行机构采用气缸夹爪。控制器对获取的料框中单组曲轴瓦盖的图像分析处理,根据定位和检测结果,引导机器人移动、夹取、搬运和放置曲轴瓦盖组等动作。曲轴瓦盖上料机器人和作业现场如图 8 所示,对曲轴瓦盖自动上料,其工作流程如下。

1) 在背景中寻找曲轴瓦盖

光电传感器周期性检测生产线上的空置情况,如果检测到存在料框,上料机器人的各个关节相互配合移动,

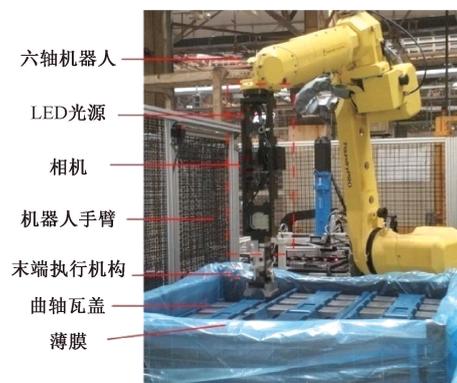


图 8 抓取作业中的上料机器人

Fig. 8 Feeding robot in grasping operation

将机器人上臂移动至料框上方且方向朝下,使得相机基本水平面向料框,从而控制光源打光、拍照。

2) 曲轴瓦盖定位、放置方向和组内顺序异常检测

如图9所示,将采集的图像送入训练好的深度学习模型进行推理,得到单组曲轴瓦盖的定位框在图像中的坐标、放置方向和组内的顺序。

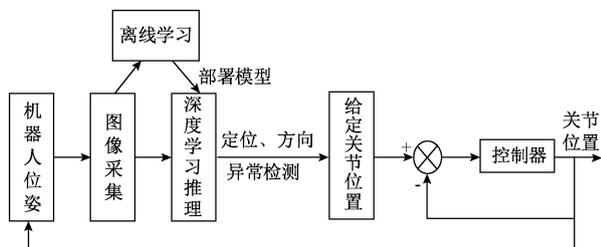


图9 上料机器人工作流程

Fig. 9 Feeding robot workflow

3) 曲轴瓦盖的抓取与搬运

对于组内顺序正确的曲轴瓦盖,由预先采用张正友方法^[21]对相机标定的结果,将该图像坐标系下坐标转换到世界坐标系的坐标。如有遮挡,控制气枪移

动到对应位置吹起薄膜,将覆盖在曲轴瓦盖上的薄膜吹至料框之外,解除薄膜对曲轴瓦盖的遮挡。若没有遮挡,则跳过上一步。然后,利用逆运动学计算出上料机器人所有关节的位置,从而改变机器人的位置和姿态,控制末端执行机构抓取曲轴瓦盖组,移动至曲轴瓦盖压装机的的工作台,根据整组曲轴瓦盖的放置方向,是否旋转机器人手臂,完成曲轴瓦盖组的放置。

为了测试本文方法的准确性和引导机器人的有效性,对其引导性能进行了对比测试。实验将设计的定位方法与机器人示教定位进行比较,将机器人根据定位的坐标执行“抓取-移动-放置”这一系列动作的成功率作为衡量引导性能的指标。对生产线上一个装满曲轴瓦盖的料框(共计144组曲轴瓦盖,分3层叠放)进行了工件定位抓取实验。设置的实验组别如下:A组一曲轴瓦盖组的三维位置完全由示教获得;B组一曲轴瓦盖组的二维平面坐标由文献[4-5]方法获得,深度信息采用预先确定的固定值;C组一曲轴瓦盖组的二维平面坐标由本文所述方法获得,深度信息采用预先确定的固定值。

实验中,除了记录上料机器人末端执行机构抓取曲轴瓦盖的成功次数以外,还特别关注了抓取的曲轴瓦盖放置在曲轴瓦盖压装机工作台上的插销上的次数。考虑到末端执行机构手爪张开的角度设计中已留有一定的余量,张开的角度实际比曲轴瓦盖的宽度大,因此,对抓取来说定位允许存在一定的偏差。但是,定位的偏差对曲轴瓦盖是否能准确放置到工作台上的插销上的影响较大。

所以,即使能够成功的抓取,也不一定能准确地将曲轴瓦盖组放置到指定的位置。显然,一次成功的引导以放置的曲轴瓦盖组能够成功装配在发动机上对应位置为准,故本文的成功率定义如下:

$$\text{成功率} = \frac{\text{放置成功次数}}{\text{试验次数}} \quad (9)$$

如表9所示,由于曲轴瓦盖组在料框中的平面位置变动几率较大,示教上料的成功率较低。文献[4-5]的方法在无遮挡时成功率高于99%,但是在包含遮挡时,其成功率仅为88.8%。本文提出的方法,在即使存在遮挡的情况下,引导上料机器人抓取和放置曲轴瓦盖的成功率有显著提高,可达95.14%,提升了生产效率。

表9 引导性能对比

Table 9 Guide performance comparison

测试项目	实验次数	抓取成功次数	放置成功次数	成功率/%
A组	144	107	86	59.72
B组	144	130	128	88.80
C组	144	139	137	95.14

4 结 论

本文研究了基于注意力机制的曲轴瓦盖上料机器人视觉定位和检测方法。在Faster R-CNN模型中引入注意力机制,使检测模型更好地学习目标的特征。在此基础上,在损失函数中引入了GIOU和改进了搜索候选框的大小,进一步提高了定位的精度。即使在被遮挡的情况下,也能快速、准确地定位。同时,本文提出的模型对噪声、光照变化等也有较好的鲁棒性,能较好地适应环境的变化,便于工业现场的部署应用。

实际应用中,由于料框中用于承重的衬垫变形,导致某些曲轴瓦盖组的高度发生变化,这时再采用二维视觉定位难以满足要求。在后续的研究中,将在现有方法的基础上,研究基于3D视觉的引导方法,实时确定曲轴瓦盖在空间中的高度和姿态,从而进一步提高上料的成功率。

参考文献

- [1] 张辉,易俊飞,王耀南,等. 医药质量检测关键技术及其应用综述[J]. 仪器仪表学报, 2020, 41(3): 1-17. ZHANG H, YI J F, WANG Y N, et al. Review on key technologies and applications of pharmaceutical quality testing [J]. Chinese Journal of Scientific Instrument, 2020, 41(3): 1-17.
- [2] 项荣,徐哈升. 用于刹车片外观质量检验的照明系统质量检验[J]. 仪器仪表学报, 2018, 39(11): 248-257.



- XIANG R, XU H SH. Quality inspection of lighting system used in surface quality inspection of brake pads[J]. Chinese Journal of Scientific Instrument, 2018,39(11):248-257.
- [3] 张辉,金侠挺. 基于曲率滤波和反向 P-M 电动车充电孔检测方法[J]. 仪器仪表学报, 2016, 37(7): 1626-1638.
- ZHANG H, JIN X T. Detection method for electric vehicle charging hole based on curvature filter and inverse P-M diffusion [J]. Chinese Journal of Scientific Instrument, 2016,37(7):1626-1638.
- [4] 高赢,张莹,闫璠,等. 发动机瓦盖上料机器人研究[J]. 计算机工程与应用,2016,52(20):257-262.
- GAO Y, ZHANG Y, YAN F, et al. Study on feeding robot of engine tile [J]. Computer Engineering and Applications,2016,52(20):257-262.
- [5] 高赢. 发动机瓦盖上料和检测系统研究[D]. 湘潭:湘潭大学,2016.
- GAO Y. Study on engine tile feeding and detection system[D]. Xiangtan: Xiangtan University, 2016.
- [6] YE H, LIU Z G, HAN Z W, et al. Fracture detection of ear pieces of catenary support devices of high-speed railway based on SIFT feature matching[J]. Journal of the China Railway Society, 2014, 36(2):31-36.
- [7] YANG H, LIU Z, HAN Y, et al. Defective condition detection of insulators in electrified railway based on feature matching of speeded-up robust features [J]. Power System Technology, 2013, 37(8):2297-2302.
- [8] XIANG X, YAN Z, ZHANG L, et al. A system and vision localization method for the opening of railway oil tank wagon based on shape matching[J]. IEEE Sensors Journal, 2016, 16(9):3008-3016.
- [9] PAPAGEORGIOU C P, OREN M, POGGIO T. General framework for object detection [C]. Sixth International Conference on Computer Vision, Bombay, India, 1998: 555-562.
- [10] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]. IEEE Computer Society Conference on Computer Vision & Pattern Recognition, San Diego, CA, USA, 2005: 886-893.
- [11] WANG X, HAN T X, YAN S. An HOG-LBP human detector with partial occlusion handling[C]. IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 2009: 32-39.
- [12] STELLA E, MAZZEO P, NITTI M, et al. Visual recognition of missing fastening elements for railroad maintenance [C]. In Proc. IEEE Int. Conf. Intell. Transp. Syst., Singapore, 2002: 94-99.
- [13] MAZZEO P L, NITTI M, STELLA E, et al. Visual recognition of fastening bolts for railroad maintenance[J]. Pattern Recognit. Lett, 2004, 25(6): 669-677.
- [14] LIU L, ZHOU F, HE Y. Automated visual inspection system for bogie block key under complex freight train environment[J]. IEEE Transactions on Instrumentation and Measurement, 2014, 65(1):2-14.
- [15] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.
- [16] 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,2016:779-788.
- [17] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision, 2016:21-37.
- [18] GIRSHICK R. Fast R-CNN [C]. IEEE International Conference on Computer Vision, Santiago, 2015: 1440-1448.
- [19] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA,2019:658-666.
- [20] YU J, JIANG Y, WANG Z, et al. UnitBox: An advanced object detection network [C]. ACM, 2016: 516-520.
- [21] ZHANG Z. Flexible camera calibration by viewing a plane from unknown orientations [C]. Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 1999: 666-673.

作者简介



朱江,2001年于湘潭大学获得学士学位,2005年于湖南大学获得硕士学位,2011年于湖南大学获得博士学位,现为湘潭大学自动化与电子信息学院副教授,主要研究方向为智能信息处理、模式识别、嵌入式并行计算。

E-mail:zhu_jiang@xtu.edu.cn

Zhu Jiang received his B. S. degrees from Xiangtan University in 2001, and received M. S. and Ph. D. degrees both from



Hunan University in 2005 and 2011, respectively. He is currently an associate professor in School of Automation and Electronics Information, Xiangtan University. His current research interest includes intelligent information processing, pattern recognition, and embedded parallel computing.



杜瑞, 2020 年获得湘潭大学信息与通信工程硕士学位。主要研究方向包括智能信息处理和模式识别。

E-mail: 905155893@qq.com

Du Rui received M. S. degree in information and communication engineering from Xiangtan University in 2020. His current research interest includes intelligent information processing and pattern recognition.



李建奇 (通信作者), 2002 年于中南大学获得学士学位, 2007 年于湖南大学获得硕士学位, 2013 年于中南大学获得博士学位, 现为湖南文理学院计算机与电气工程学院教授, 主要研究方向为先进传感与检测技术、飞行器智能控制技术、机器人视觉与图像处理等方面研究。

E-mail: jianqi_li@126.com

Li Jianqi (Corresponding author) received his B. S. degree from Central South University in 2002, M. S. degree from Hunan University in 2007 and Ph. D. from Central South University in 2013. He is now a professor in School of computing and

Electrical Engineering, Hunan University of Arts and Science. His main research interest includes advanced sensing and detection technique, aircraft intelligent control technique, robot vision and image processing.



蔡慕尧, 2020 年获得湘潭大学电子与通信工程硕士学位。主要的研究方向为智能信息处理和模式识别。

E-mail: 554904186@qq.com

Cai Muyao received M. S. degree in electronic and communication engineering from Xiangtan University in 2020. His current research interest includes intelligent information processing and pattern recognition.



许海霞, 2001 年于湘潭大学获得学士学位, 2005 年于湘潭大学获得硕士学位, 2011 年于湖南大学获得博士学位, 现为湘潭大学自动化与电子信息学院副教授, 主要研究方向为模式识别。

E-mail: xhxia2002@126.com

Xu Haixia received her B. S. and M. S. degrees both from Xiangtan University in 2001 and 2005, respectively, received Ph. D. degree from Hunan University in 2011. She is currently an associate professor in School of Automation and Electronics Information, Xiangtan University. Her current research interest is pattern recognition.