

DOI:10.19651/j.cnki.emt.2212127

基于 ViT 燃气表外观零件识别与定位方法研究^{*}

高泽铭¹ 刘桂雄¹ 陈国宇² 黄 坚^{1,3}(1. 华南理工大学机械与汽车工程学院 广州 510640; 2. 广州能源检测研究院 广州 511447;
3. 广州计量检测技术研究院 广州 510663)

摘要: 关键零件完整性是燃气表重要检定要求之一,经典图像特征匹配方法实现其完整性检测,存在通用性、泛化能力较低问题。本文提出一种改进 Faster R-CNN 多视角燃气表关键零件识别定位方法,该方法首先采用 Vision Transformer(ViT)替代 Faster R-CNN 卷积神经网络,其自注意力机制促进学习图像块特征之间相关性,强化表征能力;其次研究 ViT 优化结构参数,在 Transformer 层数 $L=14$ 、自注意头数 $m=12$ 下,模型可达到相对较优准确率。实验表明,最优模型 mAP 达 86.71%,较 ResNet50 提高 2.48%,与 ResNet101 检测准确率相当,能有效降低模型复杂性,检测效率提高 5.8%;燃气表关键零件单次检测耗时 1.13 s,可满足燃气表外观关键零件检测的准确性、实时性要求。

关键词: 燃气表;深度学习;目标检测;Vision Transformer;自注意力

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 460

Research on recognition and localization inspection of appearance for parts of gas meter based on ViT

Gao Zeming¹ Liu Guixiong¹ Chen Guoyu² Huang Jian^{1,3}(1. School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510640, China;
2. Guangzhou Institute of Energy Testing, Guangzhou 511447, China;
3. Guangzhou Institute of Measurement and Testing Technology, Guangzhou 510663, China)

Abstract: The completeness of key parts is an important verification requirement for gas meters. Although the traditional image feature matching method is used to realize the automation of part detection, its universality is poor. This paper proposes an improved method for Faster R-CNN to identify and locate key parts of gas meters from multiple perspectives. First, Faster R-CNN utilizes Vision Transformer (ViT) to replace the convolutional neural networks, whose self-attention mechanism can help to learn the correlation between image block features and strengthen the representation ability. And then the ViT structure with 14 Transformer layers and 12 self-attention heads is optimized to achieve optimal accuracy. Experimental results show that the mAP of the optimal model is 86.71%, 2.48% higher than that of ResNet50. It is equivalent to the detection accuracy of ResNet101, whose detection efficiency is increased by 5.8%, and effectively reduces the complexity of the model. It takes 1.13 s to accomplish the single detection of key parts of gas meter. The method balances the accuracy and real-time ability for key parts detection of gas meter.

Keywords: gas meter; deep learning; object detection; Vision Transformer; self-attention

0 引 言

燃气表是重要民用计量仪表,国家标准 GB/T 6968-2019、JJG 577-2012 等明确了燃气表外观关键零部件完整、齐全方面检定要求,检定项目依靠人工目视检测,会存在工作量大、容易出错的问题。目前燃气表自动化检定研究主

要集中示值误差、压力损失等读数识别^[1],外观关键零部件自动识别定位方面已开始有学者开展研究^[2],具体包括图像特征匹配识别与定位、基于深度学习的识别与定位等方法。

图像特征匹配识别与定位方法通过多尺度滑动窗口在图像人为设计特征提取器,通过所提取的多尺度特征经图

收稿日期:2022-11-18

^{*} 基金项目:广东省市场监督管理局科技项目(2022CJ04)、广东省市场监督管理局科技项目(XMBH20220614019)资助

像描述与零件模板比较,完成实现分类与定位,但经典图像特征匹配识别与定位方法,存在泛化能力较低问题,难以实现特征复杂零部件稳定识别。如中北大学电子测试技术国家重点实验室的任楷飞等^[3]采用灰度值金字塔算法进行模式匹配,对识别零件进行矩形框标识并输出像素坐标;为实现自适应模板匹配,基于图像灰度值梯度变化提取边缘信息解决燃气表齿轮尺寸变化与位置不确定性问题^[4];本团队前期基于均值滤波、Otsu 自适应分割、连通域分析等机器视觉算法,可实现燃气表划痕、孔洞、表皮剥落等表观缺陷检测,每个面检测速度不超过 0.42 s^[5]。

基于深度学习的识别与定位方法则通过卷积神经网络(convolutional neural network, CNN)能自动学习零件各种特征,从而减少繁琐图像处理环节,可以实现泛化能力好、通用性佳的零件识别、定位,具体分为一阶段、两阶段的检测方法^[6-7]。其中,一阶段检测方法将识别与定位一体化,如改进目标检测模型优化器,采用 AdaBelief 优化算法代替原有随机梯度下降算法,将原有预测边界框改进为凸包、最小外接矩形框,在零件数据集上测试准确率 mAP 达 80.06%^[8];河北工业大学的张明路等^[9]结合深度可分离卷积逆残差结构改进目标检测模型的主干网络,在检测网络前添加注意力机制,提高零件检测效率,准确率 mAP 达到 66.35%;武汉工程大学的冯晨光等^[10]针对装配零件类别及安装位置检测,研究装配缺陷目标检测模型的特征提取网络结构,通过压缩冗余特征层,提高模型推理速度。两阶段检测方法将识别与定位分段,检测精度优于一阶段检测方法。如通过在线硬样本挖掘、减少感兴趣区域数量方法优化快速区域卷积神经网络(faster region-based convolutional neural networks, Faster R-CNN)零部件检测框架^[11];华北理工大学的王一等^[12]Faster R-CNN 目标检测模型引入 ResNet101 特征提取网络、新增先验锚框、采用软非极大抑制方法,提高在零件识别定位准确性。

本文则选用 Faster R-CNN 检测模型研究多视角下燃气表外观关键零件识别与定位方法,通过引入视觉注意力模型(vision transformer, ViT)替换目标检测模型特征提取网络,进而改善模型表征性能^[13];结合 ViT 关键参数优化,提高模型准确率,从而实现燃气表外观关键零件检测的准确性、实时性要求。

1 燃气表外观关键零件检测需求分析

燃气表外观关键零件为计数器、机械封印、螺钉、气体流向标识、密封盖等,通过识别图像各类零件并框选其位置可统计零件数量、检出缺失零件位置,快速判断燃气表是否合格。单一视角下关键零件特征不明显甚至无法识别与定位全部零件,为采集燃气表多视角图像,在检定工序固定工业相机以及燃气表位置后,可旋转燃气表使工业相机采集其正视、后视、左视、右视以及俯视等多视角图像,图 1 为燃气表多视角图像及零件分布示意图。不同视角下零件特征

各异,可选取如图中所示图像零件显著特征作为检测模型学习标签,降低检测模型学习难度,提高模型识别准确率以及定位精度。燃气表大小、形状各异,检测模型应兼顾不同尺度零件特征,减少小零件漏检几率。

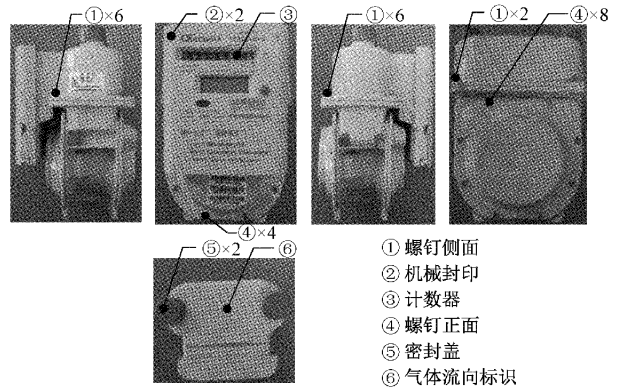


图 1 燃气表多视角图像及零件分布示意图

以实时性、准确性为评价指标,基于 Faster R-CNN 目标检测模型,设计面向燃气表关键零件多尺度快速检测系统,实现燃气表计数器、机械封印、螺钉、气体流向标识、密封盖等共 32 个零件的识别与定位,每个燃气表关键零件检测耗时不超过 1.5 s。

2 基于 Faster R-CNN+ViT 的燃气表外观关键零件视觉检测技术

图 2 为基于 Faster R-CNN+ViT 的燃气表关键零件视觉检测技术总流程框图,主要包括离线构建与优化零件检测模型、在线实时关键零件检定两部分。

离线构建与优化零件检测模型流程中,最优化模型检测合格图像,依据表 1 生成关联向量 (H_n, S_n, C_n, P_n) ,作为合格检定模板库。

在线实时关键零件检定流程中,检测图像输出关联向量 (H_i, S_i, C_i, P_i) ,由视角类型与零件类别搜索模板库,当 P_n 与 P_i 误差在 5%判为同一零件。

若检出所有零件则燃气表零件齐全,否则依据匹配关系输出缺失零件及其位置。

2.1 基于 Faster R-CNN 的燃气表特征检测技术

图 3 为燃气表特征检测的 Faster R-CNN 模型结构图, Faster R-CNN 是两阶段目标检测模型,第 1 阶段特征提取低层网络学习输入图像底层特征并获得较准确位置信息,高层网络学习输入图像较丰富语义特征,输出一系列特征图,并经特征金字塔网络(feature pyramid networks, FPN)融合低层与高层特征图输出多尺度特征图,提高对小目标物体检测精度。第 2 阶段区域推荐网络(region proposal network, RPN)检测图像各个区域是否存在关键零件并初步生成边界回归框,并在不同特征图上生成候选区域,候选区域经感兴趣区域池化(RoI pooling)、全连接层对可能存在关键零件的区域分类,并精调边界回归框。令边界框对

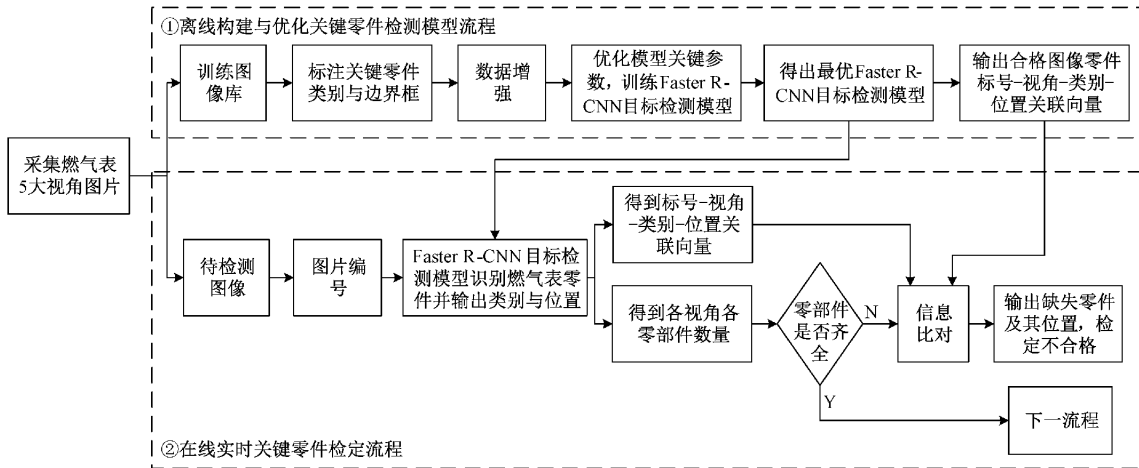


图 2 基于 Faster R-CNN+ViT 的燃气表关键零件视觉检测技术总流程

表 1 图像零件排序-视角-类别-位置关联表

类型	范围	说明
零件排序 H_n	$H_n \in [1, 32]$	32 个检测零件
视角类型 S_n	$S_n \in [1, 5]$	5 个视角
零件类别 C_n	$C_n \in [1, 6]$	6 种零件
边界框中心 P_n	(u_n, v_n)	零件中心

角位置、中心位置分别为 R_{bbox-n} 、 P_n ，边界框左上角、右下角、中心点坐标分别为 $(u_{bbox-n}^{\min}, v_{bbox-n}^{\min})$ 、 $(u_{bbox-n}^{\max}, v_{bbox-n}^{\max})$ 、 (u_n, v_n) ，则预测边界框公式为：

$$R_{bbox-n} = (u_{bbox-n}^{\min}, v_{bbox-n}^{\min}, u_{bbox-n}^{\max}, v_{bbox-n}^{\max}) \quad (1)$$

$$u_n = \frac{1}{2}(u_{bbox-n}^{\min} + u_{bbox-n}^{\max}) \quad (2)$$

$$v_n = \frac{1}{2}(v_{bbox-n}^{\min} + v_{bbox-n}^{\max}) \quad (3)$$

$$P_n = (u_n, v_n) \quad (4)$$

从输入图像到输出结果的端到端流程提高检测效率，满足实时性要求。同时两次粗、精分类以及两次边界回归可实现较高精度的特征检测。令 Faster R-CNN 模型总损失函数，RPN 网络分类损失、RPN 网络回归损失，输出分类损失、输出回归损失分别为 L_{loss} 、 L_{rpn_cls} 、 L_{rpn_reg} 、 L_{bbox_cls} 、 L_{bbox_reg} ，则有：

$$L_{loss} = L_{rpn_cls} + L_{rpn_reg} + L_{bbox_cls} + L_{bbox_reg} \quad (5)$$

进一步采用梯度下降法训练 Faster R-CNN 模型，使损失函数 L_{loss} 达到最小值，获得最佳性能的最优模型权重。

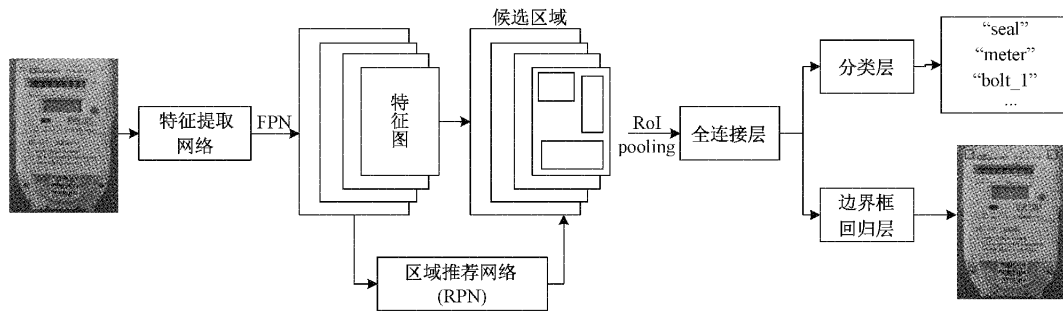


图 3 燃气表特征检测的 Faster R-CNN 模型结构

2.2 基于 ViT 的特征提取强化技术

Faster R-CNN 大多采用卷积神经网络作为主干网络，图 4 为 Faster R-CNN 图像特征提取机理示意图。每层卷积层多个卷积核通过滑窗方式提取图像或特征图上相邻区域信息输出特征图并传递给下一卷积层，单一卷积层提取局部信息，通过堆叠卷积层加大感受野捕获全局上下文信息。

ViT 将图像转化为一列向量序列输入 Transformer 提取特征。图 5 为采用 ViT 的 Faster R-CNN 图像特征提取示意图，首先将输入图像切分为 N 个图像块，再通过降维、线性变换为维度为 d 的向量序列，按顺序赋予位置编

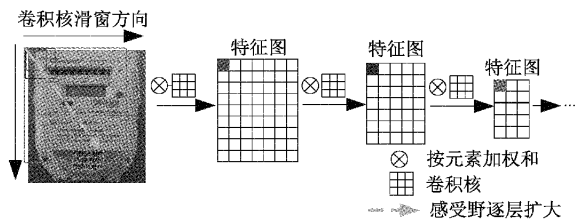


图 4 Faster R-CNN 图像特征提取机理示意图

码构成二维矩阵 $I \in \mathbf{R}^{N \times d}$ 输入到 Transformer 结构，输出维度不变的二维特征图。

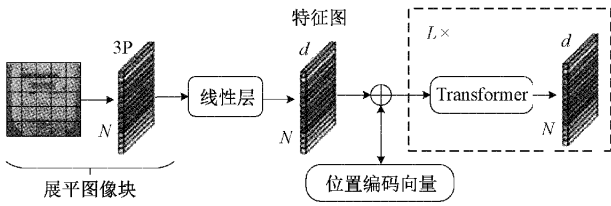


图 5 采用 ViT 的 Faster R-CNN 图像特征提取示意图

与 CNN 相比,Transformer 的多头自注意力 (multi-headed self-attention, MHSA) 结构强化对图像特征提取能力。在 MHSA 中, I 投影到不同线性空间, 可得到查询矩阵 $Q \in \mathbf{R}^{N \times d}$ 、键矩阵 $K \in \mathbf{R}^{N \times d}$ 、值矩阵 $V \in \mathbf{R}^{N \times d}$ 。为得到各特征之间多种匹配模式, 采用 m 个注意力头, 将 Q 、 K 、 V 按列分别划分 m 份, 得 $Q^i \in \mathbf{R}^{N \times k}$ 、 $K^i \in \mathbf{R}^{N \times k}$ 、 $V^i \in \mathbf{R}^{N \times k}$ ($i \in [1, m]$), 投影到不同线性子空间做 MHSA 转换, 拼接所有输出头的特征图 $O^i \in \mathbf{R}^{N \times k}$, 最后得 $O \in \mathbf{R}^{N \times d}$, 具体计算式如下:

$$k = d/m \quad (6)$$

$$O^i = \text{SoftMax} \left(\frac{Q^i (K^i)^T}{\sqrt{k}} \right) V^i \quad i \in [1, m] \quad (7)$$

$$O = \text{concat} (\{O^1, \dots, O^i, \dots, O^m\}) \quad (8)$$

式中: Q^i 、 K^i 对应行向量内积得出的自注意力权重矩阵, 反映图像块特征之间相关性, 具有较强提取长距离依赖关系能力; concat 函数使输出结果, 既能聚焦重点又关注全局信息。

2.3 基于 Faster R-CNN + ViT 的燃气表特征检测优化方法

将 Faster R-CNN 模型的特征提取网络由 CNN 替换为 ViT 来优化模型, 通过堆叠多层 Transformer 结构, 逐渐增强对全局信息整合能力, 学习更抽象的语义信息^[14]。选取 4 个不同层的特征图, 分别通过反卷积上采样 4 倍、反卷积上采样 2 倍、保持不变、最大池化下采样 2 倍, 通过 FPN 实现特征融合输出 4 种尺度特征图。

为验证 Transformer 堆叠层数 L 对检测模型准确率影响, 以指标 $mAP (IoU = 0.50 : 0.95)$ 、帧率 (frames per second, FPS) 考察目标检测模型准确率、效率, 头数 $m = 12$, 层数 L 分别设置为 8、10、12、14、16、18, 其中 ViT 原始模型 $L = 12$ 。表 2 为 Transformer 层数 L 对模型准确率、效率的影响表, 可以看出通常层数 L 增加, 模型准确率提高, 但是提升效果逐渐减慢或甚至倒退, 并造成模型参数增多、运算量变大。本文选择最优层数 $L = 14$ 。

MHSA 头数 m 直接影响检测模型准确性, m 过小则不能充分学习图像特征, 头数过大则会捕获更多不相关或匹配到无关特征。实验令层数 $L = 14$, 分别设置 $m = 1, 4, 8, 12, 16$, 其中 ViT 原始模型 $m = 12$ 。表 3 为 MHSA 头数 m 对模型准确率、效率的影响表, 可选最优 MHSA 头数 $m = 12$ 。

表 2 Transformer 层数 L 对模型准确率、效率影响

层数	FPN 融合层	$mAP/\%$	FPS
8	[2,4,6,8]	84.71	5.03
10	[4,6,8,10]	84.80	4.88
12	[4,6,8,12]	84.93	4.76
14	[4,6,10,14]	86.71	4.42
16	[4,8,12,16]	87.63	4.25
18	[4,9,14,18]	85.10	3.96

表 3 MHSA 头数 m 对模型准确率、效率影响

头数	$mAP/\%$	FPS
1	73.00	5.10
4	79.55	4.77
8	83.31	4.52
12	86.71	4.42
16	82.81	4.30

3 实验与分析

本文实验训练集、测试集由 3 种燃气表 5 个视角的图像构造, 构造并训练本文改进 Faster R-CNN + ViT 模型, 测试改进燃气表外观关键零件视觉检测模型准确性、实时性。工业机器人受拍摄距离、角度偏差等影响会采集到燃气表多尺度、形状各异外观零件图像, 为拟合多种实际场景, 采用随机缩放、随机裁剪、随机水平翻转等数据增强方法扩充数据集。

实验装置采用 Nvidia Tesla V100 GPU 深度学习加速卡, 表 4 为 Faster R-CNN + ViT 模型主要超参数设置表, 迭代训练 300 轮, 其他超参数采用推荐值^[15]。

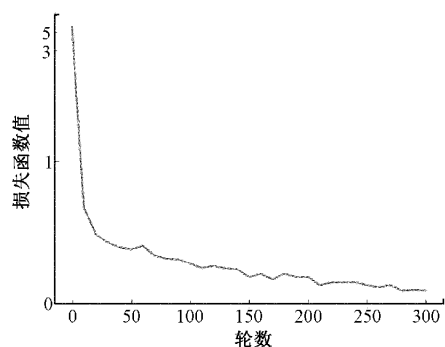
表 4 Faster R-CNN + ViT 主要超参数设置表

超参数类型	阈值	作用
rpn_nms	0.6	减少冗余锚框
bbox_nms	0.4	
bbox 类别置信度	0.8	提高检测准确率

其中 rpn_nms、bbox_nms 是判别不同锚框是否同属某零件的阈值, 当两个锚框交并比大于该阈值时判别检测到同一零件, 丢弃置信度低的锚框。bbox 类别置信度是判别某一类别的最低分数。本文燃气表外观零件不存在紧密相邻情况, 可调低 rpn_nms、bbox_nms 阈值; 本文学习标签特征显著, 类间差异大, 可提高 bbox 类别置信度。

图 6 为 Faster R-CNN + ViT 模型训练过程损失 L_{loss} 曲线, 横坐标为训练轮数, 纵坐标为训练过程损失函数值 L_{loss} 。随着模型权重不断优化, L_{loss} 逐渐收敛, 训练得出最优检测模型。

选取不同特征提取网络 ResNet50、ResNet101、ViT ($L = 14, m = 12$), 采用准确率 mAP 比较锚框位置准确性、查全率 R 比较零件检测齐全性, 在 85 张验证图片测试效

图 6 Faster R-CNN+ViT 模型训练过程 L_{loss} 曲线

果,其中:

$$R = \frac{n_r}{N} \quad (9)$$

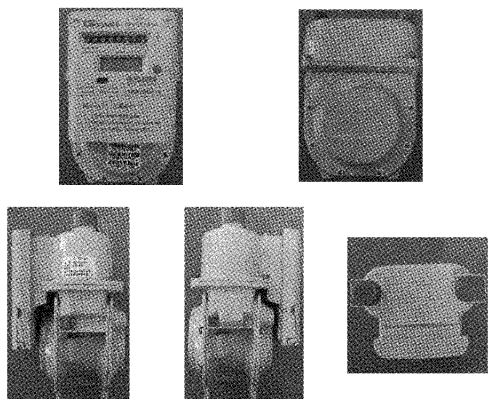
式中: N 为验证集中待识别图像数, n_r 为实际完全检出图像数。

表 5 列出 ResNet50、ResNet101、ViT 检测效果比较,其中 Faster R-CNN+ViT 最优模型 mAP 较 ResNet50 模型提高 2.48%,可检出所有外观零件,表明 ViT 自注意力机制能有效提取长距离依赖关系,提高特征表征性能,优胜于一般堆叠残差结构。Faster R-CNN+ViT 的检测效果与 ResNet101 相当,但减少冗余残差结构,模型复杂性降低。

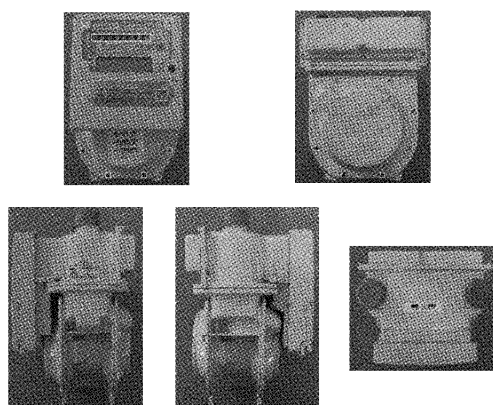
表 5 ResNet50、ResNet101、ViT 检测效果 %

模型	mAP	R
ResNet50	84.23	97.65
ResNet101	87.12	100
ViT	86.71	100

选取市面上知名品牌如金卡智能集团股份有限公司、广州金燃智能系统有限公司等燃气表,实验 Faster R-CNN+ViT 检测性能。结果表明,Faster R-CNN+ViT 可实现燃气表 5 个视角的多种外观关键零件识别定位,单次检测耗时 1.13 s,相比 ResNet101 检测耗时 1.2 s 提高 5.8%,可满足燃气表外观关键零件检测的准确性、实时性要求。图 7 为多种燃气表外观关键零件识别与定位效果图。



(a) 金卡 JGD-N2.5-1 燃气表



(b) 金燃 JRGD-N2.5 燃气表

图 7 几种燃气表外观关键零件识别与定位效果

4 结 论

本文面向燃气表检定中外观零件完整性检测需求,研究一种基于 Faster R-CNN+ViT 燃气表外观关键零件识别定位方法,可满足燃气表外观关键零件检测准确性、实时性要求。创新点为:

1) 提出一种改进 Faster R-CNN 多视角燃气表关键零件识别定位方法,采用 ViT 替代 Faster R-CNN 的 CNN 主干网络,可学习提取图像块特征,具有强特征表征能力;研究 ViT 的 Transformer 层数、自注意力头数等结构参数,在 Transformer 层数 $L=14$ 、自注意力头数 $m=12$ 时模型可达到最优准确率 $mAP=86.71\%$,较 ResNet50 提高 2.48%,与 ResNet101 检测准确率相当,检测效率提高 5.8%。

2) 在多种燃气表 5 个视角的多种外观关键零件识别定位试验表明,单次检测耗时 1.13 s,可满足燃气表外观关键零件检测的准确性、实时性要求。通过识别定位合格燃气表外观关键零件,生成合格检定模板库;待检燃气表与模板匹配检定合格性,可快速定位缺失零件,检测准确性高,适用于不同燃气表检测场景。

ViT 通过提取图像块长距离依赖关系获取全局视野,具有全局表征能力,可结合 CNN 的归纳偏置特点捕获局部感受野,提高空间感知能力,进一步提高检测模型准确性。

参考文献

- [1] 卢其伦,张圆明,马健,等. 燃气表示值误差检定时数据自动采集存在的问题及改进措施[J]. 中国测试, 2021,47(S2):11-14.
- [2] 刘桂雄,何彬媛,吴俊芳,等. 面向视觉检测的深度学习图像分类网络及在零部件质量检测中应用[J]. 中国测试, 2019,45(7):1-10.
- [3] 任楷飞,孟令军,顾泽凌. 基于灰度值金字塔算法的零件识别系统设计[J]. 中国测试, 2018,44(7): 83-87.

- [4] YE Q, DU C, LIN H, et al. A novel gas meter verification method via adaptive template matching and pulse activation decision [C]. 2021 33rd Chinese Control and Decision Conference (CCDC), 2021: 7170-7175.
- [5] 卢其伦, 闫奕樸, 张圆明, 等. 基于机器视觉的燃气表表观缺陷自动检测方法[J]. 计量与测试技术, 2022, 49(7):54-56.
- [6] 程旭, 宋晨, 史金钢, 等. 基于深度学习的通用目标检测研究综述[J]. 电子学报, 2021, 49(7):1428-1438.
- [7] 王鑫鹏, 王晓强, 林浩, 等. 深度学习典型目标检测算法的改进综述[J]. 计算机工程与应用, 2022, 58(6): 42-57.
- [8] 杨琳, 陈赛旋, 崔国华, 等. 基于改进 YOLOv4 算法的零件识别与定位[J]. 组合机床与自动化加工技术, 2021(10):28-32.
- [9] 张明路, 郭策, 吕晓玲, 等. 改进的轻量化 YOLOv4 用于电子元器件检测[J]. 电子测量与仪器学报, 2021, 35(10):17-23.
- [10] 冯晨光, 魏巍, 陈灯, 等. 基于 SlimYOLO 的控制箱零件检测方法[J]. 电子测量技术, 2022, 45(17): 120-126.
- [11] SONG Q, GUO Y, JIANG J, et al. High-speed railway fastener detection and localization method based on convolutional neural network [J]. ArXiv Preprint, 2019, ArXiv:1907.01141.
- [12] 王一, 马钰东, 董光林. 基于改进 Faster RCNN 的零件识别方法研究[J]. 应用光学, 2022, 43(1):67-73.
- [13] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. ArXiv Preprint, 2020, ArXiv:2010.11929.
- [14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]. Advances in Neural Information Processing Systems, 2017:5998-6008.
- [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 and Machine Intelligence, 2017, 39(6): 1137-1149.

作者简介

高泽铭, 硕士, 主要研究方向为现代传感技术与机器视觉。

E-mail: megao_zm@mail.scut.edu.cn

刘桂雄, 教授, 主要研究方向为先进传感与仪器研究。

E-mail: megxliu@scut.edu.cn

陈国宇, 高级工程师, 主要研究方向为智慧检测与能源计量。

E-mail: mchenguoyu@163.com

黄坚, 博士后, 主要研究方向为人工智能机器视觉技术及系统。

E-mail: mehuangjian@mail.scut.edu.cn