

杨洋, 雷钧, 王伟鸣, 等. 融合 Swin Transformer 的变速箱中壳检测方法研究[J]. 智能计算机与应用, 2025, 15(7): 168–174.
DOI: 10.20169/j. issn. 2095–2163. 250725

融合 Swin Transformer 的变速箱中壳检测方法研究

杨 洋, 雷 钧, 王伟鸣, 徐洪胜

(湖北汽车工业学院 电气与信息工程学院, 湖北 十堰市 442000)

摘 要: 针对汽车变速箱中壳生产线中存在壳体表面孔系定位效率低, 检测样本不均衡等问题, 本文提出了一种融合 Swin Transformer 的变速箱中壳检测方法。首先, Swin Transformer 作为骨干网络, 并结合多尺度特征融合结构, 使输出特征能够同时捕捉浅层和深层语义的信息, 从而增强检测方法对特征的提取能力; 其次, 针对 Swin Transformer 训练需大量数据问题, 采用 Affine Transformation 的方法来提升模型的鲁棒性; 最后, 通过引入 Huber Loss 和 Focal Loss 优化损失函数, 以平衡检测样本不均衡的问题。实验结果表明, 融合 Swin Transformer 的检测方法相较于现阶段主流的目标检测算法, 在 5%~95% IoU 的平均精度 AP_{5-95} 、50% IoU 的平均精度 AP_{50} 和 75% IoU 的平均精度 AP_{75} 指标上提高了 0.4%~4.5%。同时通过消融实验, 进一步验证了该方法在变速箱中壳检测的有效性。

关键词: 目标检测; Swin Transformer; 多尺度特征融合; 变速箱中壳

中图分类号: TP391

文献标志码: A

文章编号: 2095–2163(2025)07–0168–07

Research on the detection method of transmission case based on the fusion of Swin Transformer

YANG Yang, LEI Jun, WANG Weiming, XU Hongsheng

(School of Electrical and Information Engineering, Hubei University of Automotive Technology, Shiyan 442000, Hubei, China)

Abstract: In response to the problems of low efficiency in locating surface holes and uneven detection samples in the production line of automotive gearbox shells, this paper proposes a gearbox shell detection method that integrates Swin Transformer. Firstly, Swin Transformer is selected as the backbone network, combined with a multi-scale feature fusion structure, to enable the output features to capture both shallow and deep semantic information simultaneously, thereby enhancing the detection method's ability to extract features; Secondly, in response to the large amount of data required for Swin Transformer training, the Affine Transformation method is adopted to improve the robustness of the model; Finally, the loss function is optimized by introducing Huber Loss and Focal Loss to balance the problem of imbalanced detection samples. The experimental results show that the detection method integrating Swin Transformer has improved by 0.4%~4.5% compared to mainstream object detection algorithms in terms of average accuracy of 5%~95% IoU AP_{5-95} , average accuracy of 50% IoU AP_{50} , and average accuracy of 75% IoU AP_{75} indicators. At the same time, the effectiveness of this method in gearbox casing detection was further verified through ablation experiment.

Key words: object detection; Swin Transformer; multi scale feature fusion; transmission casing

0 引言

随着科学技术的不断发展, 汽车工业突飞猛进, 成为了中国的支柱产业^[1]。汽车变速箱作为车辆传动系统的核心组件, 其功能和性能影响汽车的安全与使用寿命^[2]。变速箱中壳孔系作为变速箱的

一部分, 包含了各种孔洞和连接接口, 承载着变速箱内部传动部件的固定和传动功能^[3–4]。在变速箱中壳加工生产线中, 数控机床的上下料由机器人完成, 为了保证中壳加工精度, 要求机械手爪抓取中壳位置不变, 因此必须准确识别变速箱中壳的定位孔和抓取孔等孔系, 便于机械手抓取。工业领域目前主

基金项目: 湖北省教育科学规划重点课题(2022GA049)。

作者简介: 杨 洋(1999—), 男, 硕士研究生, 主要研究方向: 深度学习, 机器视觉; 王伟鸣(1997—), 男, 硕士研究生, 主要研究方向: 深度学习, 机器视觉。

通信作者: 雷 钧(1975—), 男, 硕士, 教授, 主要研究方向: 工业互联网与智能制造。Email: huatlei@163.com。

收稿日期: 2023–11–28

要采用数字图像处理结合机器学习的方法进行识别,但这些方法存在一系列问题,包括复杂的特征工程、对遮挡和复杂背景敏感、不适应尺度变化、模型泛化能力有限等^[5]。

随着机器视觉和深度学习技术的迅猛发展,目标检测技术为自动化识别提供了强大的工具,相较传统机器视觉,能够实现更为精准的检测结果,并具备更强的鲁棒性。Girshick 等^[6]提出区域卷积神经网络(Region-based Convolutional Neural Network, R-CNN)模型,该模型引入了候选区域的概念,首次将深度学习与目标检测结合,但该模型计算资源消耗高、速度慢、难以实时应用;Ren 等^[7]提出了 Faster R-CNN (Faster Region-based Convolutional Neural Network),通过引入了区域提议网络(Region Proposal Network, RPN)替代 R-CNN 的选择性搜索(Selective Search),使得候选框的生成更加高效,但该模型需要训练 RPN 和目标检测两个子网络,复杂度较高。为了提高训练速度,Redmon 等^[8]提出了 YOLOv3 (You Only Look Once version 3),通过牺牲一些精度以换取更快的训练速度和推理速度。以上基于 CNN 的目标检测算法只能捕获局部信息,Carion 等^[9]提出 DETR (Detection Transformer); Liu 等^[10]提出 Swin Transformer,将自然语言处理(Natural Language Processing, NLP)领域 Transformer 模型应用到视觉任务中,增强了图像对全局的感知能力。在实际应用中, Swin Transformer 已被广泛应用于多个领域。在缺陷检测中, Gao 等^[11]改进 Swin Transformer 结构,有效提升了工业材料和零部件的检测精度; Zhang 等^[12]采用可迁移的 Swin Transformer 进行领域适配,成功解决了不同轮胎类型间的缺陷检测问题。此外, Swin Transformer 还被应用于机械故障诊断,刘子昌等^[13]结合小波时频图与 Swin Transformer 提出一种柴油机故障诊断方法,显著提高了故障识别准确率。

在工业领域,基于深度学习的目标检测算法主要应用于工业自动化、质检以及安全监控等方面。Tang 等^[14]提出一种基于改进 CenterNet 的工业物体检测方法,使用 ResNet-50 作为 Backbone,提高检测速度,并在特征处理网络中添加上采样层以提高检测精度;单明陶等^[15]提出改进的 Faster R-CNN 算法解决了传统目标检测算法对刹车零件缺陷识别率不高的问题;徐健等^[16]在 SSD (Single Shot Multibox Detector) 网络中融入感受野模块,提高网络的特征提取能力,解决了柔性智能制造过程中对密集工件

漏检率高的问题;代晓林等^[17]提出基于 Swin Transformer 骨干网络的检测方法,使用窗口划分代替卷积,避免造成数据丢失和冗余计算,提高了钢板表面缺陷识别的精度和速度。

以上研究为变速箱中壳目标检测的研究提供了参考。针对复杂工况下 CNN 网络无法充分提取图像的全局信息,本文提出了一种融合 Swin Transformer 变速箱中壳的检测方法。该方法在 Faster R-CNN 网络模型的基础上进行改进,将骨干网络替换为 Swin Transformer 与多尺度特征融合,以增强模型对特征的提取能力;使用 Affine Transformation 技术提升模型的鲁棒性,同时引入了 Huber Loss 和 Focal Loss 优化损失函数,提高模型的识别精度。

1 融合 Swin Transformer 的检测算法

1.1 网络整体架构

融合 Swin Transformer 网络的整体架构如图 1 所示,主要包括骨干网络(Backbone)、区域生成网络(RPN)、感兴趣区域(RoI)以及目标分类(Classification)和边界框回归(Box Regression)。

首先,骨干网络部分将 Faster R-CNN 的 ResNet-50 替换为 Swin Transformer 模型,对输入图像进行特征提取,得到不同层次的特征图,通过特征金字塔网络 (Feature Pyramid Network, FPN) 自上而下和自下而上的连接方式,将不同层级的特征图融合,生成一组具有多尺度信息的特征金字塔;其次,将多尺度信息输入到一个小型的卷积神经网络 RPN,通过滑动窗口在图像上提取各种大小和宽高比的锚框,对这些锚框进行分类和回归,从而得到候选区域;在 RoI 池化中,将候选区域映射回特征图上,并将每个候选区域划分为固定大小的网格,通过对每个网格内的特征值进行池化操作,将不同大小的候选区域转换为固定大小的特征图;最后,通过全连接层 (Fully Connected, FC) 对每个 RoI 池化后的特征图进行目标分类和边界框回归,从而得到每个候选区域的最终预测结果。

1.2 Swin Transformer 模型

Swin Transformer 模型由 4 个网络层构成,每个网络层都包括一个线性嵌入层和一个 Swin Transformer 块。线性嵌入层负责将输入数据进行高效编码, Swin Transformer 块用于执行复杂的特征提取和建模操作。Swin Transformer 模型结构如图 2 所示。

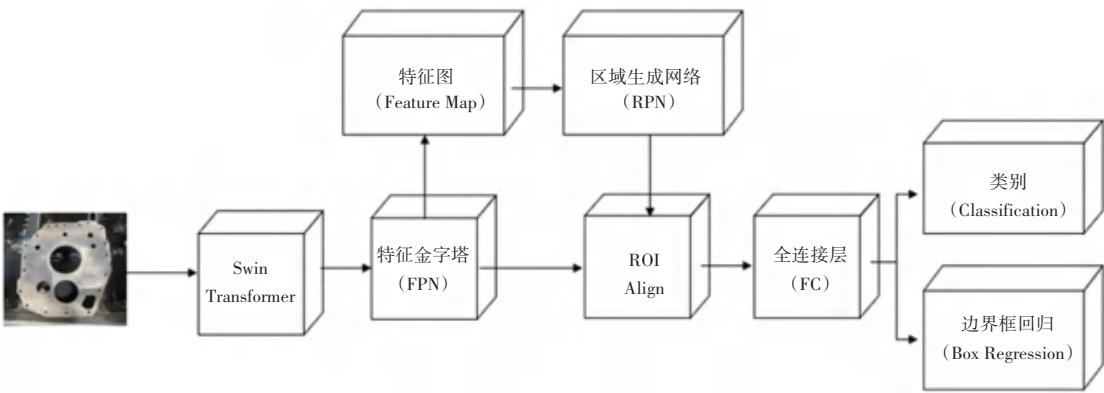


图 1 融合 Swin Transformer 网络的整体架构

Fig. 1 Integrated Swin Transformer Network Architecture

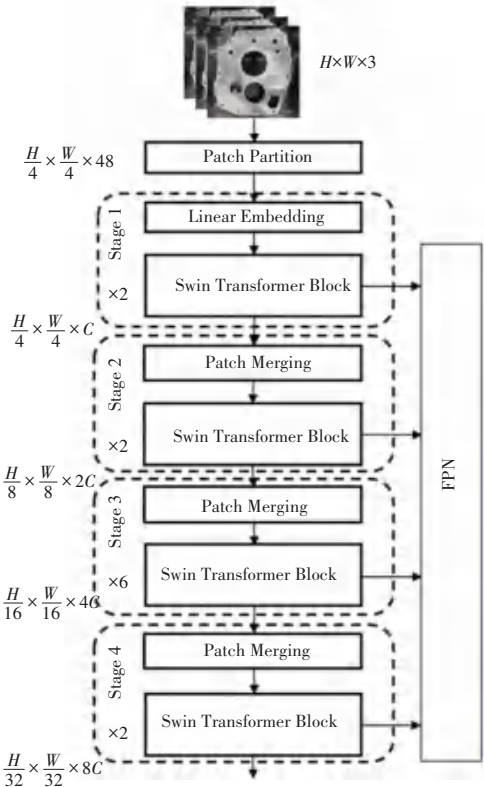


图 2 Swin Transformer 模型结构

Fig. 2 Swin Transformer model structure

Swin Transformer 处理图像的流程:首先,输入图像大小为 $H \times W$ 且为 RGB 三通道,通过图块分割 (Patch Partition) 将输入的图像分割为一系列大小为 4×4 的非重叠图块,并将这些图块在 RGB 三通道方向上展平成一维向量,展平后大小为 $4 \times 4 \times 3$,最终得到的图像大小为 $H/4 \times W/4 \times 48$;其次,将分割的图块输入到线性嵌入层 (Linear Embedding) 中,每个小块的向量通过映射到更高维度的向量空间中,以增加模型的表征能力,这种映射过程利用线性变换矩阵来完成,即通过矩阵乘法将输入向量映射到更高维度的空间;最后,线性变换后的特征图输入到

Swin Transformer 块中进行注意力计算,并通过 4 个相同阶段的处理输出 4 个不同的特征图,输入到 FPN 模块中。

1.3 Swin Transformer 块

Swin Transformer 块是 Swin Transformer 模型的基本构建块,通过自注意力和前馈网络,对输入特征图进行转换和特征提取。具体包括:局部自注意力 (Local Self-Attention) 在每个图像块内部,引入一个局部的自注意力机制,用于对图像块内的信息进行交互和整合;窗口多头自注意力 (Windows Multi-head Self-Attention, W-MSA) 和移位窗口多头自注意力 (Shifted Windows Multi-head Self-Attention, SW-MSA) 将图像块分成不同的局部区域,并在这些局部区域内进行自注意力的计算;层归一化 (Layer Normalization, LN) 用于对每个图像块的特征进行归一化操作;多层感知器 (Multilayer Perceptron, MLP) 用于对输入的特征进行处理和转换。Swin Transformer 块的输出特征变换过程共分为 4 个阶段。首先,通过 W-MSA 对上一层的特征图进行处理,并使用层归一化进行标准化,通过残差连接生成当前层的特征图,计算公式如下:

$$\hat{z}^l = W - \text{MSA}(\text{LN}(z^{l-1})) + z^{l-1} \quad (1)$$

其中, z^{l-1} 是上一层的输出特征图, $\hat{z}^l$ 是经过窗口化多头自注意力计算后得到的特征图。

其次,进行多层感知机 (MLP) 处理,并通过层归一化进一步优化其特征表示,通过残差连接得到新的输出特征图,计算公式如下:

$$z^l = \text{MLP}(\text{LN}(\hat{z}^l)) + \hat{z}^l \quad (2)$$

其中, z^l 是当前层的输出特征图。

在下一阶段,引入了 SW-MSA 机制,进行层归一化处理,应用 SW-MSA 操作并通过残差连接得到

特征图, 计算如下:

$$\hat{z}^{l+1} = \text{SW} - \text{MSA}(\text{LN}(z^l)) + z^l \quad (3)$$

其中, $\hat{z}^{l+1}$ 是经过 SW-MSA 模块处理后的特征图。

最后, 进行 MLP 处理, 进一步增强特征表达能力, 并通过残差连接得到最终的输出特征图, 计算如下:

$$z^{l+1} = \text{MLP}(\text{LN}(\hat{z}^{l+1})) + \hat{z}^{l+1} \quad (4)$$

其中, z^{l+1} 是当前层经过 MLP 处理后的输出特征图。

Swin Transformer 块结构如图 3 所示。

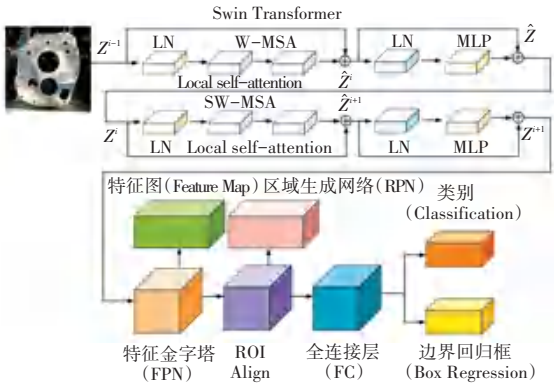


图 3 Swin Transformer 块结构

Fig. 3 Swin Transformer block structure

1.3.1 窗口多头自注意力

窗口多头自注意力 (W-MSA) 将图像划分为一系列固定大小的非重叠窗口, 然后在每个窗口内进行自注意力计算, 这样只需要处理每个窗口内的局部关系, 大大减少了计算和内存开销。假设每个窗口包含 $M \times M$ 个块 (patch), 一个全局的 MSA 模块和一个基于窗口的模块的计算复杂度在 $H \times W$ 块的图像中计算公式如下:

$$\Omega(\text{MSA}) = 4hwc^2 + 2(hw)^2C \quad (5)$$

$$\Omega(\text{W-MSA}) = 4hwc^2 + 2M^2hwC \quad (6)$$

其中, C 为特征维度。

在计算复杂度时 MSA 的 $H \times W$ 需要计算二次函数, 而 W-MSA 是线性函数, 因此 W-MSA 缩小了计算资源。

1.3.2 滑动窗口多头自注意力

尽管窗口多头自注意力在处理图像时减少了计算和内存开销, 但存在局部信息传递限制问题。因此引入滑动窗口多头自注意力 (SW-MSA), 在窗口之间加入了一定的重叠区域, 使得相邻窗口可以共享一些信息, 这样边缘窗口将从相邻的窗口获取额

外的信息, 并更好地捕捉了全局信息。通过滑动窗口, 模型可以在局部信息和全局信息之间取得更好的平衡。

在 SW-MSA 中, 每个窗口的中心会稍微移动一些距离, 而不是固定在原始位置上, 这样平移操作增强了信息的交互, 有助于模型在全局范围内更好地捕捉长距离依赖关系。SW-MSA 模块如图 4 所示。



图 4 SW-MSA 模块

Fig. 4 SW-MSA module

在滑动窗口的过程中, 窗口数由 L 层的 4 个增加到 $L+1$ 层的 9 个。为了减少窗口数量增加带来的计算量, 引入循环移位操作来处理滑动窗口。通过旋转输入图片, 可以创建一个新的待处理窗口, 其中包含不相邻窗口的信息; 通过添加掩码操作, 将自注意力的计算限制在子窗口中, 从而避免了窗口数量的增加, 有效减少了模型的计算量。循环移位的规则如图 5 所示。

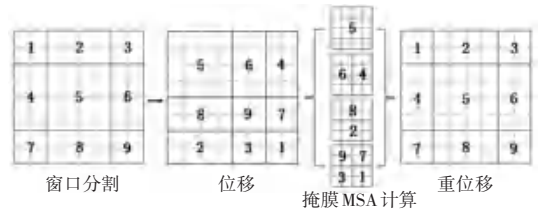


图 5 窗口循环移位规则

Fig. 5 Window loop shift rule

窗口循环移位中掩膜 MSA 计算公式 (7):

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}} + \mathbf{B}\right)\mathbf{V} \quad (7)$$

其中, $\mathbf{Q}$ 表示查询矩阵 Query; $\mathbf{K}$ 表示键矩阵 Key; $\mathbf{V}$ 表示值矩阵 Value; d 表示 $\mathbf{Q}$ 、 $\mathbf{K}$ 的维度; $\mathbf{B}$ 表示相对位置偏差。

1.4 数据增强

由于 Swin Transformer 模型训练需要大量的样本来学习, 因此采用数据增强来扩充有限的数据集。通过对原始数据进行变换和扩展, 增加模型的多样性, 有助于提高泛化性能, 降低过拟合的风险, 使得模型能够从相对较少的原始数据中学到更多有用的信息。

数据增强采用仿射变换 (Affine Transformation), 通过对图像中的目标对象进行形变、旋转和缩放来

改变其形状、方向和尺寸,减轻目标检测中的视角和尺度变化对算法性能的影响,使模型更容易捕捉目标的不同表现形式,从而提高了检测准确性和稳定性^[18]。Affine Transformation 能够改善模型对复杂场景中目标变换的适应性,增强目标检测系统的鲁棒性和可靠性。数据增强后数据样本如图 6 所示。

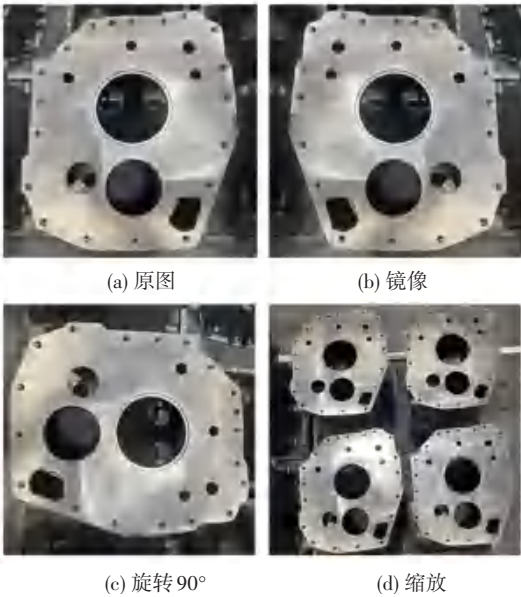


图 6 数据增强后数据样本

Fig. 6 Data samples after data augmentation

1.5 损失函数优化

融合 Swin Transformer 检测算法中,损失函数主要包括回归损失和分类损失两部分。回归损失函数采用 Huber Loss 损失函数,是均方误差 (Mean Squared Error, MSE) 和平均绝对误差 (Mean Absolute Error, MAE) 的一种折中,具有平滑性,能够减少离群点损失对函数的影响,Huber Loss 损失函数公式如下:

$$L_{\delta}(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - \hat{y})^2, & |y - \hat{y}| \leq \delta \\ \delta |y - \hat{y}| - \frac{1}{2}\delta^2, & |y - \hat{y}| > \delta \end{cases} \quad (8)$$

其中, y 表示真正的目标值; $\hat{y}$ 表示预测值; δ 是一个超参数,控制平方损失和线性损失之间切换的阈值。

如果 $|y - \hat{y}| \leq \delta$, 那么损失函数采用平方误差形式,属于均方误差;如果 $|y - \hat{y}| > \delta$, 损失函数采用线性误差形式,属于平均绝对误差。Huber Loss 在接近真实值时对平方误差进行惩罚,而在远离真实值时对绝对误差进行惩罚,从而使其对异常

值更具鲁棒性。

由于 RPN 的候选框中,正样本的数量明显不足,导致训练样本之间出现了类别分布的不均衡情况,进而影响了模型的泛化能力。为了平衡检测样本不均衡的问题,在分类损失中引入了 Focal Loss 损失函数,公式如下:

$$FL_{cls} = \begin{cases} -\alpha(1 - p_t)^{\gamma} \ln(p_t), & y = 1 \\ -(1 - \alpha)p_t^{\gamma} \ln(1 - p_t), & y = 0 \end{cases} \quad (9)$$

其中, p_t 表示模型对于样本属于正类的概率预测值,范围在 $[0, 1]$ 之间; $\alpha \in [0, 1]$ 表示正负样本权重参数; γ 是一个可调参数,用于控制对易分类样本和困难样本的关注程度。

2 实验及分析

2.1 实验环境

实验使用 Windows10, 64 位操作系统, Pytorch 框架进行训练,显卡为 GETFORCE RTX3090, 处理器为 i9-12900K, CUDA 环境。具体实验环境见表 1。

表 1 实验环境

Table 1 Experimental environment	
名称	参数
操作系统	Windows 10
内存	24 G
CPU	GETFORCE RTX3090
CUDA	11.6
开发软件	Pycharm
框架	Pytorch

2.2 数据集

汽车变速箱中壳数据集来源于某汽车动力有限公司,并使用 Labelme 标注软件进行数据集的标注,共含有 5 个类别,标注后生成的 .json 格式的文件使用脚本转换成微软通用目标语义数据集 (MS COCO) 标准格式。数据集共 848 张图片,经过数据增强处理后,扩充至 2 544 张图片。将中壳数据集按照 7 : 3 的比例随机划分成训练集和测试集,训练集图片 1 780 张,测试集图片 764 张。

2.3 评价指标

在目标检测中,算法的性能通常采用平均精度 (Average Precision, AP) 值来评估, AP_{5-95} 、 AP_{50} 、 AP_{75} 分别表示 AP 在 IoU 阈值为 5% - 95%、50% 和 75% 时的平均精度。AP 综合考虑了准确率和召回率两个指标,计算公式如下:

$$AP_j = \int_0^1 P(R) dR \tag{10}$$

其中, P 、 R 分别表示准确率和召回率, j 表示置信度阈值。

准确率和召回率计算公式:

$$\begin{aligned} P &= \frac{TP}{TP + FP} \\ R &= \frac{TP}{TP + FN} \end{aligned} \tag{11}$$

其中, TP 表示模型正确预测为正例的样本数量; FP 表示模型错误预测为正例的负例样本数量; FN 表示模型错误预测为负例的正例样本数量。

表 2 对比实验结果							
Table 2 Comparison of experimental results							
算法	Backbone	Epoch	AP/%	AP ₅₀ /%	AP ₇₅ /%	Param/ M	FPS
Fsster R-CNN	ResNet-50	300	51.6	56.6	54.3	48.9	15.5
YOLOv3	Darknet-53	300	53.2	57.7	55.6	24.3	42.7
SSD	VGG16	300	52.8	57.1	54.9	40.2	39.3
DETR	ResNet-50	300	55.7	59.1	57.3	53.6	19.8
本文方法	Swin-T	50	56.1	60.2	57.9	61.4	22.4

通过表 2 可以看出,融合 Swin Transformer 检测算法在 AP_{5-95} 、 AP_{50} 和 AP_{75} 指标相较于 Fsster R-CNN、YOLOv3、SSD、DETR 算法, AP_{5-95} 指标提高了 0.4%~4.5%, AP_{50} 指标提高了 1.1%~3.6%, AP_{75} 指标提高了 0.6%~3.6%,本文方法相较于其他算法精度最高,满足变速箱中壳孔系检测精度的需求;从 Epoch、Param 和 FPS 参数可以看出,由于 SSD 和 YOLO 目标检测算法属于一阶段算法,因此相较于改进算法在检测速度和参数量方面有一定的优势,但是检测精度偏低;本文方法相比其他算法 Epoch 较少,在两阶段检测算法中速度最快,实现了检测速度和性能的平衡。

为了验证具体的训练效果,将变速箱中壳检测效果可视化,结果如图 7 所示。对于变速箱表面不同类别的孔系,预测框均能准确完成分类和识别任务。

2.5 消融实验

为了验证数据增强和损失函数优化对整体检测效果的影响,在相同的实验环境下设置消融实验。策略一和策略二分别表示数据增强和损失函数优化,消融实验一采用策略一数据增强,消融实验二采用策略二损失函数优化,实验结果见表 3。

其他衡量算法优劣的性能指标还包括参数量 (Param) 和帧率 (FPS)。参数量用于衡量计算网络的参数大小;帧率用于衡量算法执行的速度。本文实验通过 AP、Param 和 FPS 指标的对比,综合衡量改进算法的性能。

2.4 对比实验

为验证融合 Swin Transformer 检测算法的有效性,实验在变速箱中壳数据集上进行训练,训练完成后得到对应的模型及权重文件。在同样的实验环境和数据集下,选择 4 个现阶段主流的目标检测算法 Faster R-CNN、YOLOv3、SSD^[19]、DETR 进行对比实验,对比实验结果见表 2。

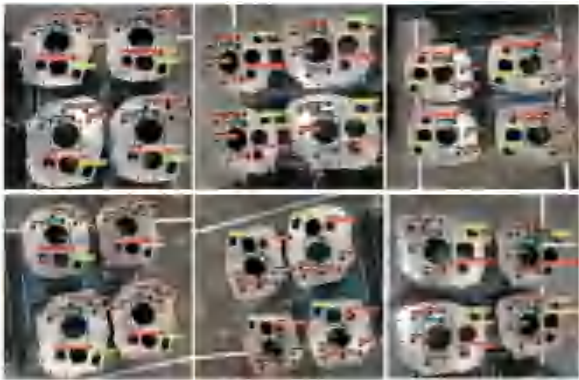


图 7 变速箱中壳检测效果可视化
Fig. 7 Visualization of gearbox casing detection effect

表 3 消融实验			
Table 3 Ablation Experiment			
算法	策略一	策略二	AP ₅₋₉₅ /%
原算法	×	×	55.6
消融实验一	√	×	56.0
消融实验二	×	√	55.7
本文方法	√	√	56.1

注:×表示不采用策略一或二;√表示采用。

由表 3 可以看出,原算法在仅加入数据增强时 AP 值提升了 0.4%;在仅优化损失函数时 AP 值提升了 0.1%;在同时加入数据增强和优化损失函数

时 AP 值提升了 0.5%,说明改进算法有效提升了变速箱中壳孔系的检测效果和准确性。为了直观地显示表 3 中消融实验的结果,绘制了平均精度和损失函数曲线对比图,如图 8、9 所示。

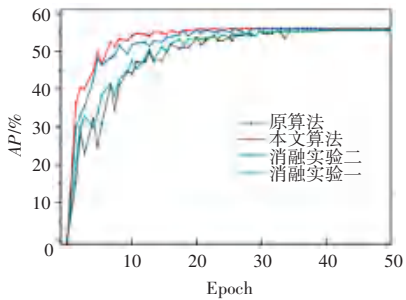


图 8 平均精度曲线对比

Fig. 8 Comparison of Average Accuracy Curves

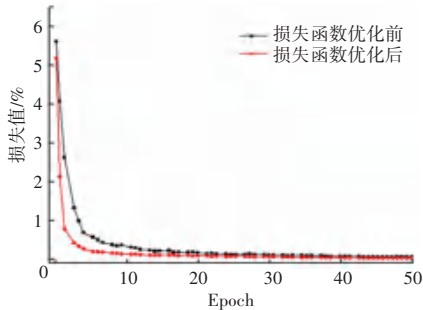


图 9 损失函数曲线对比

Fig. 9 Comparison of Loss Function Curves

3 结束语

本文针对汽车变速箱中壳生产线存在的壳体表面孔系定位效率低和检测样本不均衡等问题,提出了一种融合 Swin Transformer 的变速箱中壳检测方法。使用 Swin Transformer 作为骨干网络,并结合多尺度特征融合结构,捕捉浅层和深层语义信息,提升了特征提取能力。为了解决 Swin Transformer 训练所需大量数据问题,采用 Affine Transformation 技术来提升模型泛化能力;引入 Huber Loss 和 Focal Loss 优化损失函数,解决检测样本不均衡问题。对比实验表明,本文方法在 AP_{5-95} 、 AP_{50} 和 AP_{75} 指标上提高了 0.4%~4.5%;消融实验中, AP_{5-95} 值提升了 0.5%,验证了融合 Swin Transformer 检测方法在汽车变速箱中壳检测方面的有效性。

由于 Swin Transformer 网络的参数量较大,影响了检测速度,后续需要进一步优化模型或尝试采用轻量级的骨干网络来提升检测速度。

参考文献

[1] 佚名. 开启汽车产业中国式现代化新征程[J]. 汽车纵横,

2022,141(12):14.
[2] 卢昌云. 汽车发动机和变速箱自动化装配技术分析[J]. 时代汽车,2021,359(11):164-165.
[3] 周倩,张纪立,王俊有,等. 基于 MAGMASoft 的铝合金变速箱壳体压铸工艺优化[J]. 机械设计与制造,2023,392(10):201-204.
[4] 朱剑峰,林逸,陈潇凯,等. 汽车变速箱壳体结构拓扑优化设计[J]. 吉林大学学报(工学版),2013,43(3):584-589.
[5] 孙雄峰. 机器人视觉系统中物体检测技术的研究与应用[D]. 沈阳:中国科学院大学(中国科学院沈阳计算技术研究所),2019.
[6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Richfeature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ:IEEE,2014:580-587.
[7] 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, 2016,39(6):1137-1149.
[8] REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
[9] CARION N, MASSA F, SYNNAEVE G, et al. End-to-End Object Detection with Transformers [C]// European Conference on Computer Vision. Cham: Springer,2020: 213-229.
[10] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ:IEEE,2021: 10012-10022.
[11] GAO L, ZHANG J, YANG C, et al. Cas-Vswin transformer: A variant swin transformer for surface-defect detection [J]. Computers in Industry,2022, 140: 103689.
[12] ZHANG Y, WANG Y, JIANG Z, et al. Domain adaptation via transferable swin transformer for tire defect detection [J]. Engineering Applications of Artificial Intelligence, 2023, 122: 106109.
[13] 刘子昌,白永生,李思雨,等. 基于小波时频图与 Swin Transformer 的柴油机故障诊断方法[J]. 系统工程与电子技术, 2023,45(9):2986-2998.
[14] TANG C, WU Z, WANG S, et al. Industrial object detection method based on improved CenterNet [C]// Proceedings of the 2021 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI). Piscataway, NJ: IEEE, 2021: 121-125.
[15] 单明陶,高玮玮,杨亦乐,等. 基于改进 Faster R-CNN 的刹车零件缺陷检测算法[J]. 国外电子测量技术,2022,41(4):22-28.
[16] 徐健,陆珍,刘秀平,等. 基于改进 SSD 的密集工件检测[J]. 组合机床与自动化加工技术,2022,585(11):70-74.
[17] 代晓林,刘梦玫,生群,等. 基于改进 Swin Transformer 的钢板表面缺陷检测方法[J]. 装备制造技术,2022,328(4):88-91.
[18] XUE X, ZHANG K, TAN K C, et al. Affine transformation-enhanced multifactorial optimization for heterogeneous problems [J]. IEEE Transactions on Cybernetics, 2020, 52(7): 6217-6231.
[19] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[J]. CORR, abs/1512.02325, 2015.