

文章编号 1004-924X(2025)19-3093-13

基于 SDFSN-HiFuse 网络的减速器工件分类

于智龙¹, 张雪寒^{1*}, 齐丽华¹, 杨佳欣¹, 于广滨², 李忠刚²

(1. 哈尔滨理工大学 自动化学院, 黑龙江 哈尔滨 150080;

2. 哈尔滨工业大学 机电学院, 黑龙江 哈尔滨 150080)

摘要: 减速器相似工件的准确分类对于其精密装配至关重要。现有视觉分类方法在面对高度相似的工件时存在特征判别性不足、抗复杂背景干扰能力弱等问题, 性能表现不佳, 在精密装配中容易引入误差。针对减速器工件类内差异大、类间差异小的特点, 提出一种基于 HiFuse 的空域双焦协同网络 (Spatial Dual-Focus Synergy Network, SDFSN) 减速器工件分类方法。设计多分支空间自适应的膨胀率选择机制, 使模型对形变区域自动选择最合适的感受野。构思双阶段几何-局部协同注意力机制, 对每个膨胀分支的输出特征施加逐级精细的注意力引导, 动态调整特征权重, 有效增强模型对重要区域的判别能力, 实现由粗到细的特征提取。引入可变形几何图, 实现与几何拓扑适配的图结构, 突破传统固定网格限制, 在可变形卷积后引入曲率门控机制, 继承几何形变的适应性特征, 显著提升对复杂曲面区域的响应能力与表达精度。实验结果表明, SDFSN-HiFuse 在自制数据集上的准确率比基线提高 3.57%, 精确度提高 2.99%, 而且满足工件分类的实时性要求, FPS 达到 300.39 frame/ms。

关键词: 减速器工件分类; 深度学习; 注意力机制; 多尺度膨胀卷积

中图分类号: TP18; TH132.46 文献标识码: A

doi: 10.37188/OPE.20253319.3093 CSTR: 32169.14.OPE.20253319.3093

Reducer workpiece classification based on SDFSN-HiFuse network

YU Zhilong¹, ZHANG Xuehan^{1*}, QI Lihua¹, YANG Jiaxin¹, YU Guangbin², LI Zhonggang²

(1. College of Automation, Harbin University of Science and Technology, Harbin 150080, China;

2. College of Mechanical and Electrical Engineering, Harbin Institute of Technology,

Harbin 150080, China)

* Corresponding author, E-mail: Zxh990802@163.com

Abstract: Accurate classification of visually similar reducer parts is essential for precise assembly. Existing visual classification methods struggle with highly similar parts due to limited discriminative features and low robustness to complex background interference, which can introduce errors in assembly. To address these challenges, a HiFuse-based Spatial Dual-Focus Synergy Network (SDFSN-HiFuse) is proposed for classification of reducer workpieces, targeting scenarios with large intra-class variance and small inter-class variance. A multi-branch spatially adaptive dilation-rate selection mechanism is introduced to enable automatic determination of appropriate receptive fields for deformed regions of workpieces. A two-stage geo-

收稿日期: 2025-04-13; 修订日期: 2025-06-18.

基金项目: 黑龙江省重点研发计划资助项目 (No. 2023ZX01A03)

metric - local collaborative attention mechanism provides stepwise fine-grained guidance to features from each dilation branch, dynamically reweighting features and enhancing discrimination of salient regions via a coarse-to-fine refinement process. A deformable geometric graph is employed to model geometric topology flexibly, overcoming the constraints of traditional fixed grids. Following deformable convolution, a curvature gating mechanism preserves adaptive geometric deformation features, substantially improving responsiveness and representation accuracy on complex curved surfaces. On a custom dataset, SDFSN-HiFuse achieves a 3.57% absolute improvement in accuracy and a 2.99% increase in precision over the baseline, while meeting real-time requirements with a processing rate of 300.39 frame/s.

Key words: reducer workpiece classification; deep learning; attention mechanism; multi-scale dilated convolution

1 引 言

减速器工件是机械产品制造中的基础和关键组成部分,科学构建其多维分类体系直接影响生产线的传动效率与结构创新。分类体系不仅是传动装配设计优化的重要依据,其准确性也直接影响工件间的配合精度与生产成本控制。

随着数字孪生技术在智能制造领域的深度应用,如工具状态监测^[1]、曲面零件旋压技术^[2]等,工业机器人在减速器装配时利用数字孪生辅助装配,这对减速器工件分类方法的准确性提出了更高要求,尤其在面对背景复杂、相似度高以及姿态各异问题时。实际装配场景中,以某减速器装配企业为例,包含多类结构高度相似的齿轮轴、螺栓等工件,外观差异极为微小,如不同型号的螺栓在长度和螺纹结构上仅有约0.3 mm的差异,螺纹密度和头部轮廓相近。现有的视觉分类方法在面对这些高度相似的工件时,会因特征判别性不足、抗复杂背景干扰力弱等问题,错分率显著提升^[3]。

基于机器视觉的分类方法在实际生产中表现出巨大潜力,但是随机森林^[4]、支持向量机^[5]等传统方法却面临三大挑战,即特征设计依赖经验、语义信息会丢失以及计算效率低。为了应对这些挑战,深度学习的引入是一种变革性的发展,传统的机器学习方法已经无法适应大数据时代,迫切需要一种可以完成分类任务的学习模型。

深度学习在工件分类中展现出巨大潜力,这归功于卷积神经网络(Convolutional neural networks, CNN)强大的特征提取能力^[6]。从LeNet^[7]首次证明CNN在图像任务中的有效性,

到现在CNN已从深度^[8]、宽度^[9]、效率^[10]和结构自动化^[11]四个方面取得显著进步。MOHAN等^[12]针对工件提出优化的卷积神经网络架构,有效提高了多类别工件图像的识别准确率,并在复杂背景下展现出良好的泛化能力。传统CNN结构虽然对中等规模数据表现良好,但对小差异、高相似度工件识别能力有限。LIU等^[13]通过改进YOLOv5网络结构,实现机械零部件的快速分类,验证了视觉识别系统在工业场景下的端到端集成应用潜力。尽管集成了分类功能,但识别准确率仍依赖YOLOv5模型对小目标的检测性能,误检率偏高。王林毅等^[14]构建了一个密集堆叠数据集,训练了目标检测模型,实现了对多类别控制阀零件的检测任务,但未关注高相似度工件间的分类问题。RAJAGOUNOER等^[15]采用几何梯度特征与机器视觉系统对工件进行识别,实现了工业级在线检测系统的集成,但其方法依赖几何特征和边缘梯度,对形状接近或表面纹理接近的工件区分能力较差。洪存存等^[16]采用改进的双边滤波算法处理样本图像,构建最小二乘支持向量机自动识别模型,准确率为93.3%,但其双边滤波的高斯核的平滑作用可能会削弱微小缺陷的梯度特征。梁礼明等^[17]采用渐进和跨层的方式充分融合目标语义信息,嵌入双层路由注意力模块,并引用SIoU损失函数,更好地平衡了检测精度和速度的问题。盛彬等^[18]提出轻量化的LAI YOLOv5模型,引入注意力机制与特征融合,识别精度为98.6%,但未深入研究工件类别之间的细微结构差异。余永维等^[19]采用Inception模块结合SSD检测框架,兼顾识别准确性与实时性能,设计针对复杂工业环境的多尺度特征提取方法,增强模型对小目标检测

能力,但只关注目标识别,未深入分析高相似类别之间的分类精度。

综上所述,当前工件视觉分类研究虽已取得一定成果,但仍存在一定不足。多数模型对工件结构微差识别能力有限,缺乏在高相似度工件数据集上的有效验证。因此,本文聚焦于高相似度工件的图像分类任务,提升模型对局部结构与姿态各异的判别能力。针对工件分类算法,选择 HiFuse^[20]作为基础网络框架,它是一个三分支架构,能够将全局与局部特征进行自适应融合。本文将 HiFuse 网络应用在工件分类领域并进行分类。针对工件中存在图像背景复杂、形状相似、姿态各异问题,设计 SDFSN-HiFuse 网络结构。首先,设计多分支空间自适应膨胀卷积(Multiscale Convolution of Adaptive Expansion Rate, AERC)选择机制,突破传统膨胀卷积^[21]的固定模式。对每个膨胀分支的输出施加注意力,动态调整特征权重,使模型对形变区域自动选择最合适的感受野。然后,设计双阶段几何-局部协同注意力机制(Geometry-Local Cooperative Attention, GLCA),在每个分支上辅助自适应膨胀卷积,起到权重调整作用。层级注意力设计实现从空间粗选到特征精修的递进优化,使模型更专注于工件关键特性,显著提升分类精度。最后,引入可变形几何图动态生成与工件几何拓扑适配的图结构,突破传统固定网格限制,在可变形卷积后引

入曲率门控机制,继承几何形变适应性,增强曲面区域响应。自制工件数据集,实验结果表明, SDFSN-HiFuse 在工件数据集上实现良好的性能。

2 SDFSN-HiFuse 工件分类模型

HiFuse 模型是一种针对医学图像分类设计的深度学习模型,通过分层多尺度特征融合技术,结合 CNN 与 Transformer 的优势,显著提高分类精度与效率。其设计依然能够有效处理工件图像中复杂的结构差异,因此,基于 HiFuse-Small 网络提出 SDFSN-HiFuse 网络,用在工件的分类上。

2.1 SDFSN 模块

图 1 为 SDFSN-HiFuse 网络的整体架构。其中,局部分支内 5 个分支并行提取特征,根据生成的注意力,量化各膨胀率分支在当前位置的重要性,根据权重图对多分支输出进行加权,克服固定卷积核对不同区域特征敏感度不足问题。

通过引入膨胀卷积和 LCA 模块^[22]来设计 SDFSN,加强显著特征的学习,从而提高工件分类的准确性。SDFSN 模块如图 2 所示。整体结构主要包括深度可分离卷积、并行膨胀卷积和 GLCA,深度可分离卷积进行轻量化。

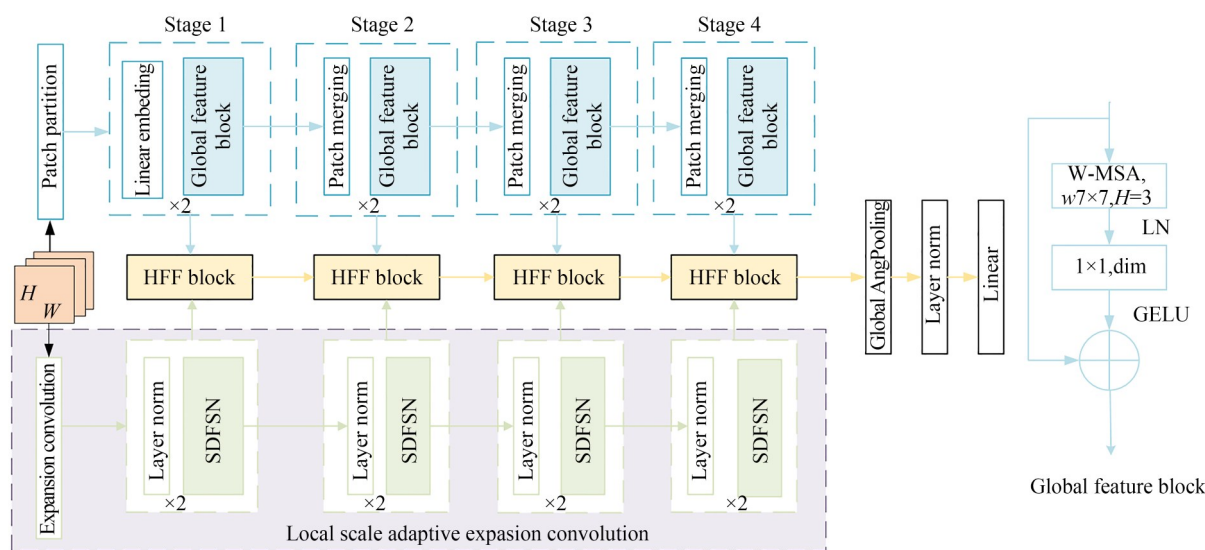


图 1 SDFSN-HiFuse 整体架构

Fig. 1 Overall architecture of SDFSN HiFuse

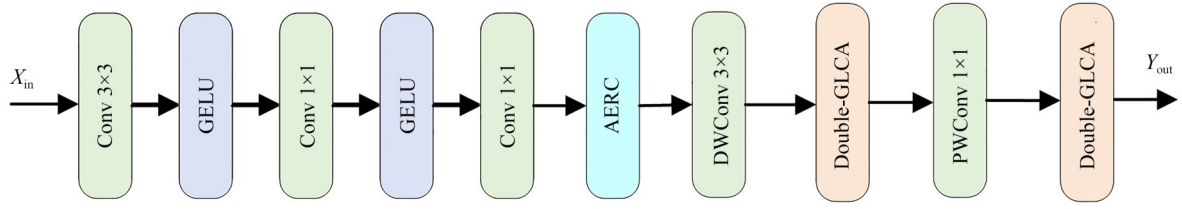


图2 SDFS模块

Fig. 2 SDFS module

2.2 AERC 模块

为应对姿态各异、形状相似问题,通过引入膨胀卷积,让其根据当前视角下特征的尺度,动态调整感受野大小,设计 AERC 模块,加强显著特征的学习,从而提高工件分类的准确性。AERC 模块包括 5 个并行分支,膨胀率分别为 1, 2, 3, 5, 7, 大膨胀率分支专注于形变区域,小膨胀率分支聚焦细节。每个位置可独立选择最优膨胀率。GLCA 对融合后的特征进行几何形变补偿,解决膨胀卷积的网格效应问题。然后通过 1×1 卷积实现跨通道信息融合。AERC 模块的整体结构如图 3 所示。模块将输入特征映射 $X_{in} \in \mathbb{R}^{c \times g \times h \times w}$ 分为 g 组,可以表示为 $X_{in} = [X_1, X_2, X_3, \dots, X_g]$,每组特征都通过两个平行的子网络来捕获跨渠道信息,并进一步生成空间注意信息 $M_i \in \mathbb{R}^{h \times w}$, $i = 1, 2, 3, \dots, g$ 。最后,

根据权重预测器生成空间注意力,量化各膨胀率分支在当前位置重要性,根据权重图对多分支输出进行加权。lexNet^[23]提出特征分组已被广泛应用于网络架构,该结构表明不同的特征组具有各种类型的特征。给定输入张量 $X_{in} \in \mathbb{R}^{c \times h \times w}$, X_{in} 沿通道划分为 g 组。每组负责学习不同特征,以减少参数的数量而不减少通道维度,从而最大程度减少信息丢失, g 可以通过相应的空间注意力来增强组。最后,通过汇总组之间的特征获得丰富的特征信息。每个组的跨空间学习,进行以下操作。使用两个平行分支以不同的尺度捕获空间信息。 1×1 标准卷积旨在调整通道,生成像素级空间注意力图。 3×3 卷积旨在进行局部特征提取,为注意力提供多尺度特征基础,称为 X^i 。随后,每个分支应用全局平均池化,以编码全局空间信息并建模长期依赖性。每个分支产生的

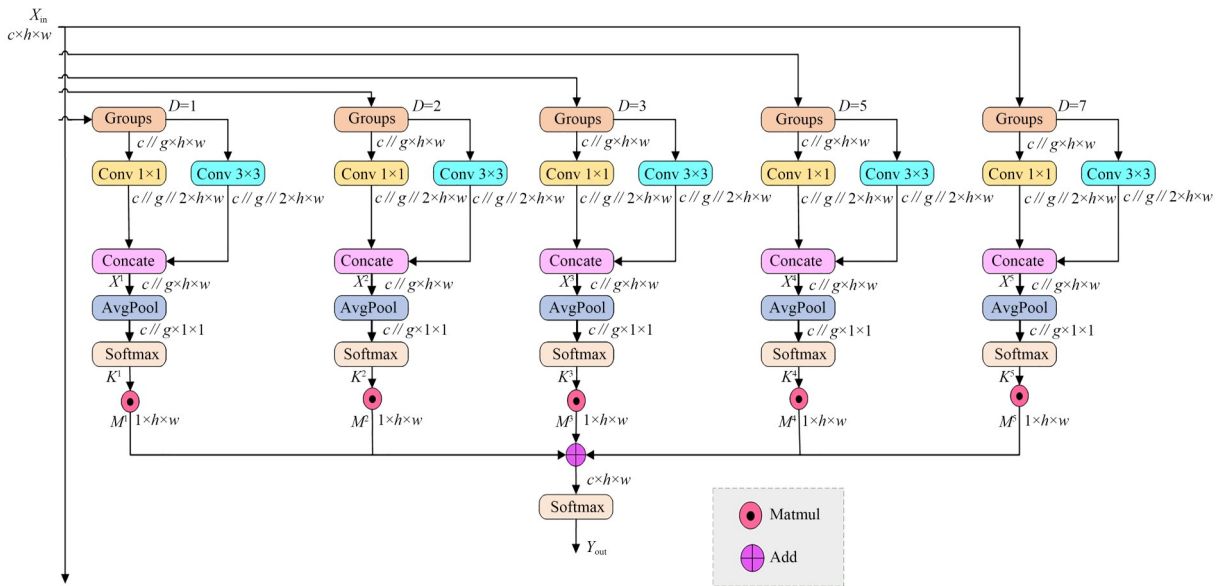


图3 AERC模块

Fig. 3 AERC module

长度为 $c//g$ 的向量分别表示为 K^i , 如下所示:

$$\bar{K}_1^i = \frac{1}{H \times W} \sum_i^H \sum_j^W X_1^i(i, j). \quad (1)$$

执行 Softmax 函数 K^i , 以获取通道组中不同通道的权重。每一条分支中的向量 K^i 为:

$$K_1^i = \text{Softmax}(\bar{K}_1^i). \quad (2)$$

然后执行矩阵乘法, 生成空间注意力图 M^i , $M^i \in \mathbb{R}^{H \times W}$, 注意图中, 每个像素代表当前空间尺度和另一个空间尺度之间的跨通道相互作用信息, 从而更好地表达当前位置对局部上下文的依赖性。注意力 M^i 如下:

$$M_1^i = X_1^i \odot K_1^i, \quad (3)$$

其中 $\odot$ 是矩阵相乘符号。

值得注意的是, 每个组中的空间注意力仅从相应组的数据中计算出来, 因此不同组之间没有信息关联。然后, 通过动态空间权重加权融合来汇总多分支特征, 表达式如下:

$$Y_{\text{out}} = \sum (Weight_d(x, y) * Feature_d(x, y)) \quad \forall d \in dilations, \quad (4)$$

其中: AERC 模块的自适应机制通过权重预测器实现形变区域的动态响应, 权重预测器赋予大膨胀率分支 ($r=5/7$) 更高权重, 大感受野可跨越形变区域捕捉上下文结构。对于细节丰富处, 权重预测器激活小膨胀率分支 ($r=1/2$)。过渡区域使用中膨胀率分支 ($r=3$), 该过程由式 (3) 的注意力机制实现位置的自适应选择。

2.3 Double-GLCA 模块

为抑制背景干扰、解决形状相似、姿态各异问题, 设计一个 Double-GLCA 模块, 如图 4 所示。Double-GLCA 具有双路结构, 一条支路处理局部特征, 另一支路处理几何形变。第一阶段几何注意力通过可变形卷积初步区分工件和背景, 抑制与工件整体几何结构无关的背景噪声。可变形卷积通过空间偏移量动态调整采样点位置, 使卷积核覆盖区域始终适配工件的投影几何形变, 从而提取到工件结构特征。第二阶段应用局部注意力在第一阶段定位的区域内, 进一步聚焦与工件自身的判别性局部细节, 同时进一步弱化残留的背景信息。局部注意力放大微小差异, 解决形状相似问题。可变形几何图动态生成与工件几何拓扑适配的图结构, 突破传统固定网格限制。对边缘或者齿轮区域分配更多采样点, 对平坦区域减少采样密度, 使其动态适配不同视角下工件的投影形变。在可变形卷积后引入曲率门控机制 G_{curv} , 继承几何形变适应性, 增强曲面区域响应。

可变形几何卷积表达式如下:

$$F_{\text{deform}} = \sum_{K=1}^K W_K \cdot F_{\text{base}}(P + P_K + \Delta P_K), \quad (5)$$

其中: P_K 为固定卷积核位置, ΔP_K 为学习到的偏移量, $F_{\text{deform}} \in \mathbb{R}^{h \times w \times c//2}$ 为形变特征。

基于可变形卷积输出的形变适应特征图, 计算逐像素高斯曲率近似值, 然后将曲率信息与原

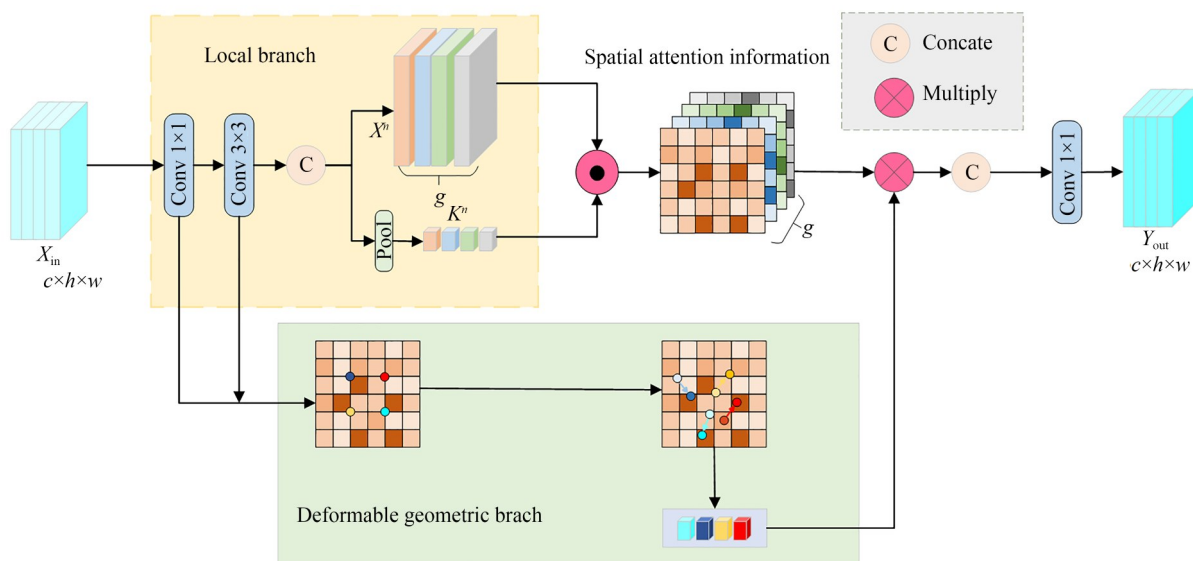


图 4 Double-GLCA 模块

Fig. 4 Double-GLCA module

始特征拼接,特征融合后生成曲率特征增强,使模型对减速器工件的曲面形态敏感。

曲率门控调制表达式如下:

$$G_{\text{curv}} = \sigma(FC(\text{AvgPool}(F_{\text{base}}))), \quad (6)$$

$$F_{\text{gate}} = G_{\text{curv}} \odot F_{\text{deform}}, \quad (7)$$

其中: $G_{\text{curv}} \in \mathbb{R}^{c//2}$ 为门控权重, F_{base} 为输入, σ 为 Sigmoid 激活函数, F_{gate} 为调制。

每组特征分别通过自适应池化操作映射到矢量 K^i 。然后,将这些向量进行相乘操作,从而产生注意力图 M^i 。基于输入特征动态生成融合权重矩阵,自适应调节几何局部贡献比,实现特征互补。

空间自适应融合表达式如下:

$$F_{\text{fused}} = W_{\text{route}}[:, 0] \odot F_{\text{gate}} + W_{\text{route}}[:, 1] \odot F_{\text{local}}, \quad (8)$$

$$W_{\text{route}} = \text{Soft max}(\text{Conv3} \times 3(F_{\text{base}})), \quad (9)$$

$$F_{\text{local}} = \text{Sigmoid}(\text{DWConv}(F_{\text{base}})) \odot F_{\text{base}}, \quad (10)$$

其中: W_{route} 为权重, F_{local} 为支路输出。

在逐通道卷积(提取初级空间特征)与点卷积(融合通道信息)后,分别嵌入两阶段 GLCA

注意力机制:第一阶段对逐通道卷积输出的底层特征图进行粗筛选,聚焦工件显著性区域;第二阶段对点卷积生成的高级特征表示进行细调整,增强判别性细节的权重。这种层级注意力设计实现了从空间粗选到特征精修的递进优化,使模型更专注于工件关键特性,显著提升分类精度。

3 实验及结果分析

3.1 工件数据集制作

以减速器为研究对象,采用实验室的自采图片制作数据集。图像采集设备为 Intel RealSense D435 深度相机,在数据采集过程中,拍摄不同光照、背景和角度下的工件图像,最终构建数据集 1 888 张,共分为四大类,九小类,分别包括:螺钉两类(304-round-screw, 400 张)、齿轮轴两类(gear shaft, 400 张)、齿轮轴承三类(reducer, 678 张)、缸套两类(steel jacket, 410 张),如图 5 所示。然后,按照 0.7:0.15:0.15 的比例划分训练集、验证集和测试集。



图 5 工件数据集

Fig. 5 Workpiece dataset

3.2 实验环境和超参数配置

实验环境的配置如下:CPU 为 Intel Core i9-13900K, GPU 为 NVIDIA RTX 4090, 操作系统为 Ubuntu 22.04.6。软件版本和深度学习框架分别为 python3.8.20, Cuda11.5, Torch2.4.0。

在实验时没有使用预训练权重,输入图像大小设置为 224×224 , 批量大小为 32。使用 AdamW 优化器,初始学习率为 $4e-4$, 动量为 0.937, 权重衰减系数为 0.05, 总训练轮数为 500。

3.3 评价指标

在检测过程中,采用以下指标来全面评估提出的分类模型性能和效率:将准确率(Accuracy, A)和精确度(Precision, P)用作主要性能指标来衡量分类的准确性。Acc 是分类正确的样本数占总样本数的比例,反映模型正确分类工件样本的能力。Pre 是指在所有被预测为正的样本中,实际为正样本的概率。

A 和 P 的计算如下:

$$A = \frac{TP + TN}{TP + FP + TN + FN}, \quad (11)$$

$$P = \frac{TP}{TP + FP}, \quad (12)$$

其中: TP 是正确预测为正样本的数量; FP 是错误预测为正样本的数量; FN 是正样本的数量错误地预测为负样本; TN 是正确预测为负样本的数量。

FLOPs 代表每秒浮点操作,表示处理单个图像所需的浮点操作数量,反映模型对计算资源的需求及其在硬件上部署的适用性。

Params 表示模型参数,使用模型中可训练参数的总计数来评估其复杂性。Params 越低,模型权重越小。

推理速度 FPS 表示模型在一次推理过程中所需时间,以 ms 为单位进行衡量。推理速度越快,模型在实际应用中的表现越好。其公式如下:

$$FPS = \frac{1000}{t}, \quad (13)$$

其中 t 表示推理一张图片需要的时间。

3.4 消融实验

为验证 AERC 模块和 Double-GLCA 模块的有效性,使用 HiFuse 模型作为基线进行工件数据集的消融研究,结果列于表 1 和图 6。

表 1 自制数据集的消融研究

Tab. 1 Research on ablation of workpiece dataset

Baseline	AERC	GLCA	Double-GLCA	$A/\%$	$P/\%$	Params/M	FLOPs/G	FPS/ms
HiFuse				94.64	95.22	123.28	18.14	537.44
HiFuse	✓			95.54	96.34	142.69	21.26	463.65
HiFuse		✓		96.41	96.12	142.7	21.27	441.09
HiFuse			✓	97.32	97.30	153.01	22.96	495.50
HiFuse	✓	✓		97.61	97.56	153.10	22.95	387.93
HiFuse	✓		✓	98.21	98.21	168.10	25.00	300.39

基线 HiFuse 模型在工件数据集上达到 94.64% 的准确率, Params 为 123.28 M, FLOPs 为 18.14 G, 其推理速度最高, 达到 537.44 帧每毫秒。通过集成 AERC 模块, 性能提高到 95.54%, 同时模型 Params 和 FLOPs 适度增加, 这证明所提出的 AERC 模块在增强模型特征表示方面的有效性。

为验证 Double-GLCA 模块的有效性, 设计仅集成 GLCA 与 Double-GLCA。仅集成 GLCA 时, 该模型 Acc 为 96.41%, 仅集成 Double-GL-

CA 模块时, 该模型准确率提升 2.68%, 模型 Params 和 FLOPs 分别为 153.01 M 和 22.96 G。这表明 Double-GLCA 模块可以有效地改善性能。将 AERC 和 Double-GLCA 模块结合在一起, 模型准确率达到 98.21%, 推理速度为 300.39 帧每毫秒, 虽然较基线下降了, 但仍适用本文应用场景。

与其他两个模型相比, 由于动态形变建模以及多尺度扩展需要并行分支处理不同感受野, 参数和计算量略有增加, 但准确率相较于基线

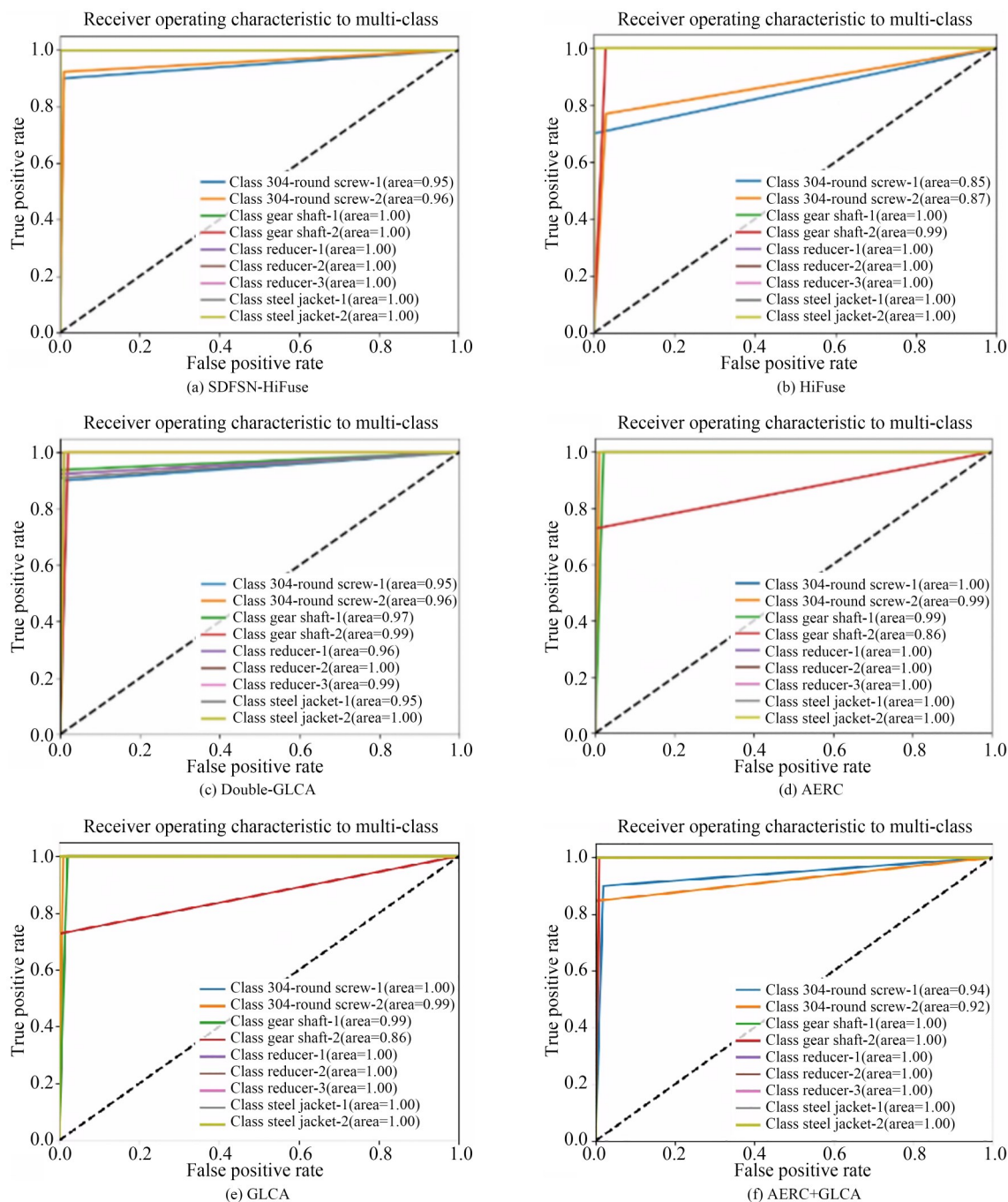


图6 ROC曲线

Fig. 6 ROC curves

HiFuse 模型提升 3.57%。消融研究验证所提出的 AERC 和 Double-GLCA 模块的有效性。图 6 为 ROC 曲线对比,从图中可以看出,SDFS-HiFuse 显示出较高的真阳性率与较低的假阳性率。HiFuse 方法在工件分类中存在 0.85~0.87 的 AUC 波动,其优势源于 SDFS-HiFuse 网络有效捕捉螺纹间隙,齿轮等微观几何

特征。

3.5 不同膨胀率对比实验

为测试自适应机制的有效性,用固定膨胀率、随机生成膨胀率以及不同组合膨胀率进行对比实验。不同膨胀率的实验结果对比如表 2 所示。当使用固定膨胀率 2 时,A 为 89.41%,模型参数为 125.50 M,FLOPs 为 18.50 G。由此看

出,使用固定膨胀率模型性能低于基线。使用随机生成膨胀率 A 为 96.43%,Params 为 134.18 M,FLOPs 为 19.89 G。使用膨胀率为 1,3,5 时,准确率为 96.67%,参数略有降低,但 A 比 SDFS-N-HiFuse 低 1.54%。而推理速度分别提高

41%,35%,3%,这是因为参数量增加,从而导致模型变大。从上述数据看出,在 SDFS-N-HiFuse 中 AERC 模块的有效性。Params 略有增加,这是因为多了膨胀率分支,自适应融合也会增加额外参数。

表 2 不同膨胀率的对比实验
Tab. 2 Comparative study of different expansion rates

膨胀率	$A/\%$	$P/\%$	Params/M	FLOPs/G	FPS/ms
2	89.41	90.05	125.50	18.50	510.58
随机生成	96.43	96.67	134.18	19.89	463.13
1,3,5	96.67	96.64	160.85	24.50	310.05

3.6 不同分支数对比实验

为测试不同分支数量的影响,分别采用分支 3 和分支 5 进行对比实验,结果如表 3 所示。当使用 3 分支时, A 为 96.64%,模型参数为 15.30 M,FLOPs 为 22.51 G;使用 5 分支时, A 为 89.41%,模型参数增加 11.84%,FLOPs 增加 11.06%,推理速度降低 38%,这是由于膨胀率分支会增加额外参数。

表 3 不同分支数的对比实验
Tab. 3 Comparative study of different numbers of branches

分支数	$A/\%$	$P/\%$	Params/M	FLOPs/G	FPS/ms
3	96.64	96.71	150.30	22.51	485.10
5	98.21	98.21	168.10	25.00	300.39

3.7 对比实验

为验证所提出的 SDFS-N-HiFuse 网络的优越性,将与图像分类的 6 种主流算法:ResNet18, MobileNetV2^[24], ShuffleNet^[25] 和 EfficientNetV2^[26],ConvNext^[27],Swin Transformer^[28] 进行比较,结果如图 7 所示。

在针对 6 种典型卷积神经网络模型的对比实验中,各模型在分类性能与计算效率方面呈现出显著的差异性。如表 4 所示,SDFS-N-HiFuse 模型以 98.21% 的准确率和 98.21% 的精确度达到最优分类性能,其双指标一致性表明该模型在特征表达与类别区分能力上具有显著优势。然而,该模型的 Params(168.10M)与 FLOPs(25.00G)

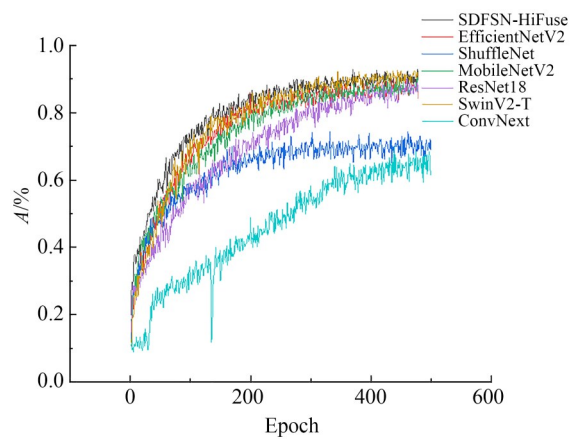


图 7 分类系列算法准确率曲线对比
Fig. 7 Comparison of accuracy curves for classification series algorithms

略高于轻量化设计代表 MobileNetV2 (Params142.57 M,FLOPs 20.68 G),显示其性能提升伴随一定的计算开销。传统模型 ResNet18 虽在分类性能 (A 为 96.43%, P 为 96.57%) 上与 MobileNetV2 相近,但其 Params (172.21 M) 与 FLOPs(26.21 G) 均高于后者,印证深度残差网络在浅层结构中存在冗余参数的问题。值得注意的是,EfficientNetV2 通过复合缩放策略以 125.10 M Params 和 17.65 G FLOPs 成为计算效率最优的模型,推理速度最快 (600.59 frame/ms),尤其适用于资源受限场景。相比之下,ShuffleNet 在 Params (170.65 M) 与 FLOPs(26.11 G) 上接近 ResNet18 的情况下,分类性能 (A 为 81.25%, P 为 82.10%) 显著落后,暗示其通道混洗机制可能在本实验任务中存在

表 4 不同分类算法的比较

Tab. 4 Comparison of different classification algorithms

模型	$A/\%$	$P/\%$	Params/M	FLOPs/G	FPS/ms
ResNet18	96.43	96.57	172.21	26.21	351.65
MobileNetV2	96.42	96.64	142.57	23.65	600.59
ShuffleNet	81.25	82.10	170.65	26.11	366.25
EfficientNetV2	97.32	97.37	165.23	24.85	500.65
ConvNext	82.14	82.61	170.55	25.11	300.52
SwinV2-T	98.15	98.26	210.30	30.67	250.53
SDFSN-HiFuse	98.21	98.21	168.10	25.00	300.39

适应性不足的局限。SwinV2-T虽然在分类精度上具有显著优势,与SDFSN-HiFuse持平,但由于Params较高,所以其推理速度SDFSN-HiFuse慢。ConvNext作为纯卷积网络现代化的优秀代表,其 A (82.14%)和推理速度(300.52 ms)在本实验中均落后于其余模型,这可能由于参数配置一些原因导致。由此可见,所提出的SDFSN-HiFuse模型在测试集上表现最优,在参数和计算量基本持平的情况下,准确率最高。

为更直观看到SDFSN通过引入膨胀率来自适应扩大感受野,从而捕捉更广泛的上下文信息,可视化如图8所示。颜色从蓝色到红色渐变(彩图见期刊电子版),表示从低响应到高响应,坐标轴上的数值范围是0到50,代表空间维度,图8(a)大部分区域呈现高响应,这说明模型在这些区域有较强的激活,有效证明SDFSN捕获关键特征的能力。图8(b)的高响应通常集中在局部区域,比如边缘。

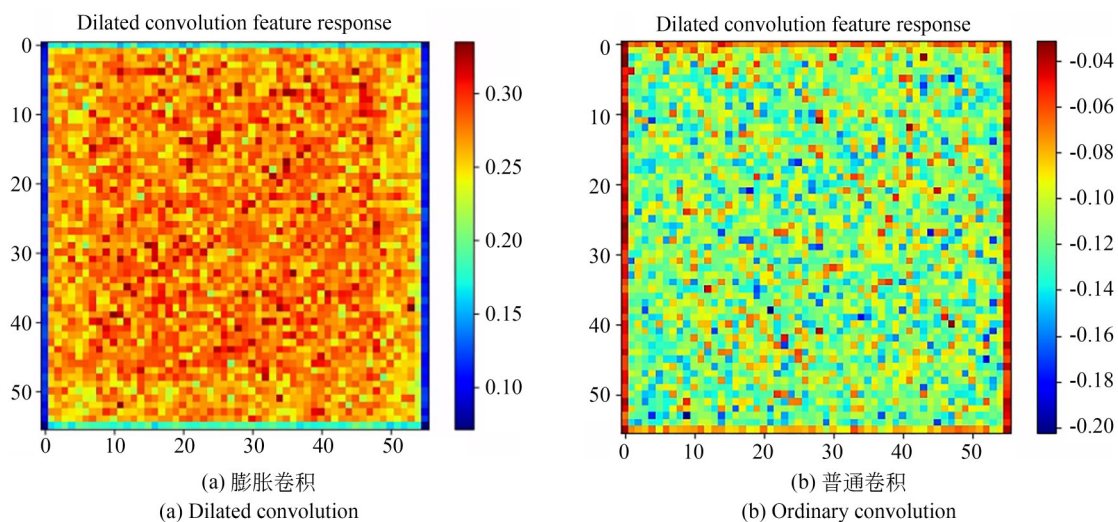


图8 卷积对比

Fig. 8 Convolutional comparison

3.8 检测结果分析

在自制数据集上的可视化结果如图9所示,9(a)~9(i)分别为九种工件在SDFSN-HiFuse上的可视

化结果。可视化结果显示,SDFSN-HiFuse模型在九类工件数据集上表现出良好的分类能力,并且对于形状特征明显的类别,模型置信度可以达到99%。

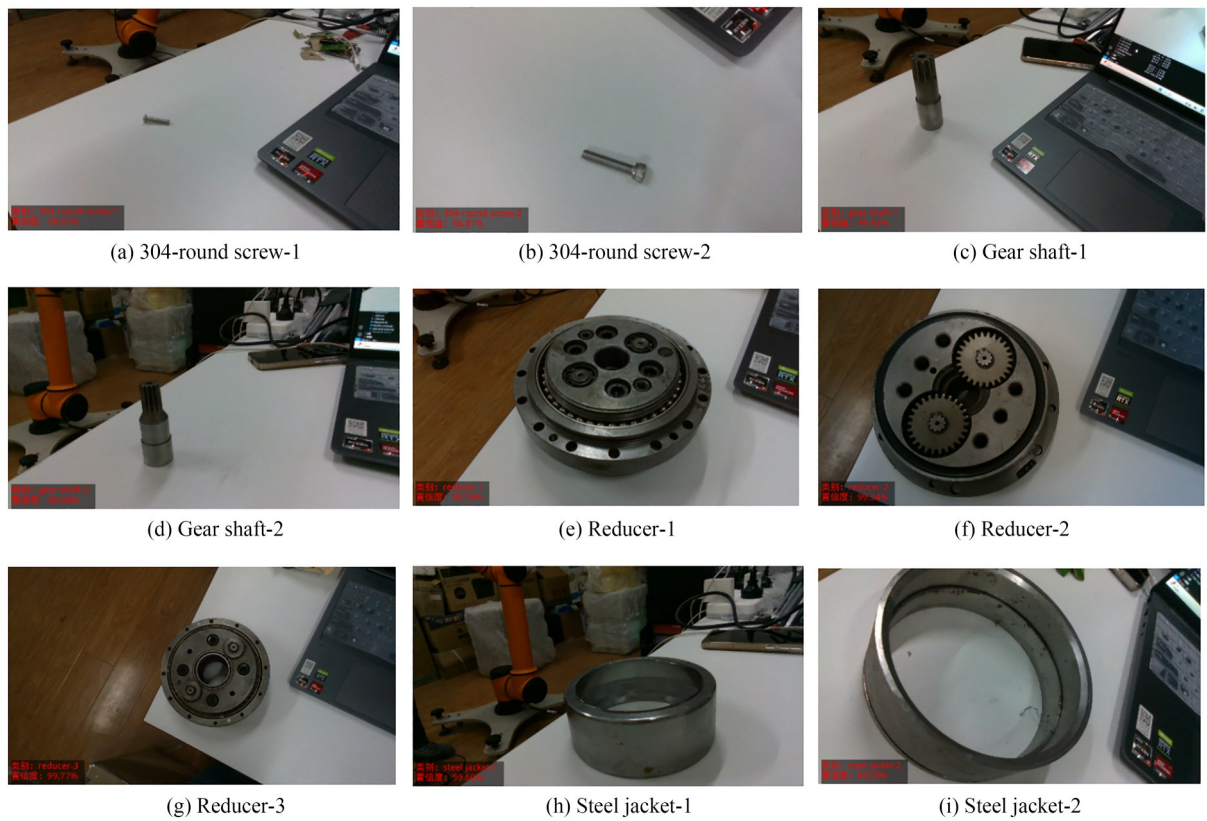


图 9 可视化结果
Fig. 9 Visualization results

4 结 论

本文提出了一种 SDFSN-HiFuse 工件分类算法,以改善因工件相似差异小导致误装问题。通过设计多分支空间自适应的膨胀率选择机制,引入轻量级 GLCA-Guided 对每个膨胀分支的输出施加几何-局部注意力,动态调整特征权重,使模型对形变区域自动选择最合适的感受野。设计了一个双阶段几何-局部协同注意力机制实现先粗略再精细细化特征的方法,增强了重要区域的权重判别性。引入可变形几何图动态生成与工件几何拓扑适配的图结构,突破传统固定网格限制,在可变形卷积后引入曲率门控机制,继承几何形变适应性,增强曲面区域响应,并在自制工件数据集上进行实验。实验结果表明,SDFSN-HiFuse 在工件数据集上的准确

率为 98.21%,精确度为 98.21%,同时保持低计算复杂性,从而验证了其出色的检测性能,有助于智能制造与工业自动化分类在制造业中的发展。

在今后的研究中,将更关注于面对几何特征高度相似、尺寸与几何公差要求严格的工件,以提升算法的精度。

作者贡献声明:

- 于智龙:方法提出及设计;
- 张雪寒:仿真实验及论文撰写;
- 齐丽华:论文审核;
- 杨佳欣:实验数据整理;
- 于广滨:实验方法指导;
- 李忠刚:实验数据分析。

参考文献:

- [1] ZHANG J, LI C C, DENG C J, et al. Toward digital twins for intelligence manufacturing: Self-adaptive control in assisted equipment through multi-sensor fusion smart tool real-time machine condition monitoring[J]. *Journal of Manufacturing Systems*, 2025, 82: 301-318.
- [2] GAO P F, LI X S, YAN X G, et al. Digital twin-driven intelligent spinning technique for curved surface parts[J]. *Journal of Industrial Information Integration*, 2025, 45: 100848.
- [3] REAL C, TORRES Y. Effect of changes in the sequence of assembly operations on error rates: a case study from the car manufacturing industry[J]. *IEEE Access*, 2024, 12: 34644-34655.
- [4] BREIMAN L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5-32.
- [5] CORTES C, VAPNIK V. Support-vector networks [J]. *Machine Learning*, 1995, 20 (3) : 273-297.
- [6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [7] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. *Proceedings of the IEEE*, 1998, 86 (11) : 2278-2324.
- [8] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 770-778.
- [9] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions [C]. 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 7-12, 2015, Boston, MA, USA. IEEE, 2015: 1-9.
- [10] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications[J/OL]. *arXiv*, preprint arXiv: 1704.04861, 2017. <https://arxiv.org/abs/1704.04861>.
- [11] ZOPH B, VASUDEVAN V, SHLENS J, et al. Learning transferable architectures for scalable image recognition[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 8697-8710.
- [12] MOHAN C, SABER T, NALLATHAMBI P J. Recognition of industrial spare parts using an optimized convolutional neural network model[J]. *Information*, 2024, 15(12): 793.
- [13] LIU Y X, DUAN S H, SHEN Z P, et al. Grasp and inspection of mechanical parts based on visual image recognition technology[J]. *Journal of Theory and Practice of Engineering Science*, 2023, 3 (12): 22-28.
- [14] 王琳毅, 白静, 李彦梅, 等. 面向工业目标检测的密集控制阀零件数据集[J]. *光学精密工程*, 2024, 32(8): 1241-1251.
WANG L Y, BAI J, LI Y M, et al. Dense control valve parts dataset for industrial object detection [J]. *Opt. Precision Eng.*, 2024, 32 (8) : 1241-1251. (in Chinese)
- [15] RAJAGOUNDER R. A machine vision system for inspecting mechanical parts[J]. *Machine Graphics and Vision*, 2025, 34(1): 75-86.
- [16] 洪存存, 王雪琨, 于文文, 等. 基于改进双边滤波的皮革缺陷检测[J]. *激光技术*, 2021, 45(3): 373-377.
HONG C C, WANG X K, YU W W, et al. Leather defect detection based on improved bilateral filtering [J]. *Laser Technology*, 2021, 45 (3) : 373-377. (in Chinese)
- [17] 梁礼明, 龙鹏威, 卢宝贺, 等. 改进 GBS-YOLOv7t 的钢材表面缺陷检测[J]. *光电工程*, 2024, 51(5): 240044.
LIANG L M, LONG P W, LU B H, et al. Improvement of GBS-YOLOv7t for steel surface defect detection [J]. *Opto-Electronic Engineering*, 2024, 51(5): 240044. (in Chinese)
- [18] 盛彬, 王晨升. 基于深度学习的装配零部件识别与定位方法[J]. *制造技术与机床*, 2024(6): 44-51.
SHENG B, WANG CH SH. An assembly part identification and localization method based on deep learning [J]. *Manufacturing Technology & Machine Tool*, 2024(6): 44-51. (in Chinese)
- [19] 余永维, 韩鑫, 杜柳青. 基于 Inception-SSD 算法的零件识别[J]. *光学精密工程*, 2020, 28(8): 1799-1809.
YU Y W, HAN X, DU L Q. Target part recognition

- tion based inception-SSD algorithm[J]. *Opt. Precision Eng.*, 2020, 28(8): 1799-1809. (in Chinese)
- [20] HUO X Z, SUN G, TIAN S W, *et al.* HiFuse: Hierarchical multi-scale feature fusion network for medical image classification[J]. *Biomedical Signal Processing and Control*, 2024, 87: 105534.
- [21] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[J/OL]. *arXiv preprint arXiv*: 1511.07122, 2015. <https://arxiv.org/abs/1511.07122>.
- [22] LUONG M T, PHAM H, MANNING C D. Effective approaches to attention-based neural machine translation[J/OL]. *arXiv preprint arXiv*: 1508.04025, 2015. <https://arxiv.org/abs/1508.04025>.
- [23] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [24] SANDLER M, HOWARD A, ZHU M L, *et al.* MobileNetV2: inverted residuals and linear bottlenecks[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 4510-4520.
- [25] ZHANG X Y, ZHOU X Y, LIN M X, *et al.* ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 6848-6856.
- [26] TAN M, LE Q. Efficientnetv2: smaller models and faster training[C]. *International conference on machine learning*. ICML, 2021: 10096-10106.
- [27] WOO S, DEBNATH S, HU R H, *et al.* ConvNeXt V2: co-designing and scaling ConvNets with masked autoencoders[C]. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 17-24, 2023, Vancouver, BC, Canada. IEEE, 2023: 16133-16142.
- [28] LIU Z, LIN Y T, CAO Y, *et al.* Swin transformer: hierarchical vision transformer using shifted windows[C]. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 10-17, 2021, Montreal, QC, Canada. IEEE, 2021: 9992-10002.

作者简介:



于智龙(1979—),男,黑龙江哈尔滨人,博士,副教授,硕士生导师,2002年于深圳大学获得学士学位,2007年于哈尔滨理工大学获得硕士学位,2013年于东北林业大学获得博士学位,主要研究方向为输配电系统的状态检测、人工智能、机器学习。E-mail: zlyu@hrbust.edu.cn

通讯作者:



张雪寒(1999—),女,黑龙江哈尔滨人,硕士研究生,2021年于哈尔滨理工大学获得学士学位,主要从事工件分类、图像处理的研究。E-mail: Zxh990802@163.com