

文章编号: 1671-7449(2026)04-0423-12

J. Test Measur. Technol., 2026, 40(4): 423-434.

基于DINOv2到EfficientViT的多损失蒸馏 Patch特征少样本异常检测

韩 信¹, 邵星灵^{1*}, 李秀源², 闫佳乐², 邓瑞祥¹

(1. 中北大学 电气与控制工程学院, 山西 太原 030051; 2. 中北大学 仪器与电子学院, 山西 太原 030051)

摘要: 针对少样本工业异常检测任务中大规模模型部署困难与轻量级模型性能不足的问题, 提出一种基于知识蒸馏的轻量化异常检测方法。以DINOv2作为教师模型, EfficientViT作为学生模型, 通过知识迁移实现在资源受限环境下的高性能检测。设计特征对齐适配器以弥合师生模型特征差异, 并构建结合Huber损失、余弦相似度损失与Gram矩阵损失的多目标蒸馏函数, 从数值、方向和空间统计关系多维度约束学生模型学习。在MVTec AD数据集上的实验表明, 仅使用正常样本训练即可显著提升学生模型性能: 蒸馏后的学生模型与未蒸馏基准模型相比, 图像级异常检测AUROC提升0.170; 与教师模型相比, 在保持相当性能的同时, 参数量减少约23%, 推理速度提升4倍。消融实验与可视化分析验证了模型各组件的有效性。

关键词: 异常检测; 知识蒸馏; 轻量化模型; 视觉Transformer; 少样本学习

中图分类号: TP183; TP391.41 文献标识码: A DOI: 10.62756/csjs.1671-7449.2026060

引用格式: 韩信, 邵星灵, 李秀源, 等. 基于DINOv2到EfficientViT的多损失蒸馏Patch特征少样本异常检测[J]. 测试技术学报, 2026, 40(4): 423-434.

Han Xin, Shao Xingling, Li Xiuyuan, et al. Multi-loss distillation of patch features from DINOv2 to EfficientViT for few-shot anomaly detection[J]. Journal of Test and Measurement Technology, 2026, 40(4): 423-434.

Multi-Loss Distillation of Patch Features from DINOv2 to EfficientViT for Few-Shot Anomaly Detection

Han Xin¹, Shao Xingling^{1*}, Li Xiuyuan², Yan Jiale², Deng Ruixiang¹

(1. School of Electrical and Control Engineering, North University of China, Taiyuan 030051, China;
2. School of Instrument and Electronics, North University of China, Taiyuan 030051, China)

Abstract: To address the problem in few-shot industrial anomaly detection where large-scale models are difficult to deploy and lightweight models show limited performance, a lightweight anomaly detection method based on knowledge distillation is proposed. DINOv2 as the teacher model and EfficientViT as the student model, transferring knowledge are used to achieve high detection performance under resource constraints. A feature alignment adapter is designed to reduce the representation gap between the two models, and a multi-objective distillation loss combining Huber loss, cosine similarity loss, and Gram matrix loss guides the student

【收稿日期】2025-10-15

【修回日期】2025-12-25

【录用日期】2025-12-30

【基金项目】国家自然科学基金资助项目(62203404)

【作者简介】韩 信(2001—), 男, 硕士生, 主要从事异常检测和图像处理研究。E-mail: 2336868413@qq.com。

*【通信作者】邵星灵(1988—), 男, 教授, 博士, 主要从事导航制导与控制、无人机视觉异常自主安全控制等研究。E-mail: shaoxl@nuc.edu.cn。

from multiple dimensions of magnitude, direction, and spatial correlation. Experiments on the MVTec AD dataset show that the student model, trained only with normal samples, achieves a 0.170 improvement in image-level AUROC over the baseline model without distillation, while maintaining comparable performance to the teacher model with about 23% fewer parameters and 4 times faster inference. The effectiveness of the proposed framework is confirmed by ablation and visualization studies.

Key words: anomaly detection; knowledge distillation; lightweight model; vision Transformer; few-shot learning

0 引言

异常检测是指识别与预期模式或行为显著不同的数据点的过程,其核心目标在于发现系统运行中潜在的不寻常事件、异常行为或趋势^[1]。作为机器学习和计算机视觉中的重要研究方向,异常检测已被广泛应用于工业检测^[2-3]、视频监控^[4]、医疗诊断^[5],以及网络安全^[6]等领域。然而,在大多数实际场景中,异常数据往往稀缺、分布多样且难以提前获取,这使得传统依赖于有监督样本的检测方法面临挑战。

为应对异常样本不足的问题,少样本异常检测逐渐受到广泛关注^[7]。少样本异常检测的核心问题是如何建立能够泛化到未知异常模式的表征体系。近年来,Transformer架构的发展为计算机视觉领域的应用提供了重要启示^[8],也为异常检测中的表征学习提供了丰富的技术思路。邹琦萍等^[9]提出了基于多层级视频Transformer的视觉自动定位方法(MLVT),通过多级标记划分与逐层可变形注意力机制,在自建视频数据集上的分类准确率达88.9%,有效实现了从模糊到清晰序列中精准选出最清晰帧。薛振华等^[10]提出了SSMOD-Net,借助视觉大模型架构,设计了状态空间模型驱动的提示编码器,为SAM生成动态提示,实现了无需微调的高效工业异常检测,提升了模型对未知异常的泛化能力。Zhu等^[11]提出了一种基于双层路由注意力机制的视觉Transformer模型BiFormer,在ImageNet-1K图像分类任务中达到83.2%的Top-1准确率。金泽辉等^[12]提出了视觉语言模型CRM,通过RGFormer将CLIP图像特征映射至文本空间,在COCO Karpathy测试集上得到145.3的CIDEr分数,性能优于基线方法,兼具高效性与优异的跨模态生成能力。以上研究体现了一个趋势:高质量的注意力机制设计、有效的跨尺度/跨模态特征对齐,以及大模型蕴含的强泛化先验,是构建鲁棒异常检测系统的关键要素^[13]。

视觉大模型在异常检测中的应用主要基于其强

大的特征表示能力和分布外检测潜力。基于Transformer架构的视觉大模型是通过自注意力机制实现全局感受野,突破了传统卷积网络的局部性限制^[14]。刘势杰等^[15]提出了改进U-Net与跨模态自蒸馏的医学图像融合方法,通过双分支编码和注意力门机制,有效提升融合图像的细节保留与跨模态特征表达能力。Guo等^[16]提出了一种结合视觉Transformer与自监督学习的皮肤病分类方法,在ISIC 2018数据集上AUC、F1分数等指标显著优于传统CNN模型,特征空间呈现明显聚类,训练收敛快、稳定性高,并且减少了对大规模标注数据的依赖。Yang等^[17]提出了一种基于Vision Transformer自编码器的无监督工业图像异常检测方法,通过引入全局上下文建模、记忆模块和坐标注意力机制,在MVTec AD和BeanTech AD数据集上图像级准确度平均提升约20%,显著优于传统卷积编码器。当测试样本与训练分布存在偏差时,模型会产生不一致的特征响应或注意力模式,这些信号为异常识别提供了有效依据^[18]。然而,视觉大模型的巨大参数量和计算需求使其难以在资源受限的边缘设备上部署,且模型的特征表示可能存在冗余^[19]。因此,如何有效提取和利用大模型中的异常敏感特征仍需深入研究。

知识蒸馏作为一种有效的模型压缩与迁移学习方法,已成为实现轻量化少样本异常检测的重要途径^[20]。宋建华等^[21]提出了一种基于结构化知识蒸馏的轻量级伪装目标检测方法,通过优化软硬标签损失和学习率策略,提升了蒸馏效果与训练稳定性。王莉莉等^[22]提出了一种基于轻量化ViT的面部疲劳监测算法,在WiderFace数据集上3个难度等级的平均精度分别达到92.3%、90.2%和81.3%,检测速度为每帧68 ms,兼顾高精度与实时性,且提升了疲劳检测性能。崔智等^[23]提出了一种基于多教师模型引导知识蒸馏的图像去雾算法,在SOTS室外数据集上峰值信噪比达23.57 dB、结构相似度达0.934,提升了去雾图像的质量,有效缓解了颜色失真问题。

谢奕涛等^[24]提出了面向目标类别的掩码蒸馏方法(MDTC),在多个图像分类数据集上实现高效知识提取,40%以上学生模型准确率超越教师模型,且训练耗时更少。知识迁移主要包括3种形式:响应知识^[25]、特征知识^[26]和关系知识^[27]。其中,特征知识与关系知识能够更全面地传递教师模型的结构化表征能力,因此在视觉任务中表现出更强的迁移效果^[28]。

通过联合优化多类型蒸馏损失,学生模型能够在参数量大幅减少的情况下仍保持较高的表征能力^[29-30]。以自监督预训练模型为教师模型,将其对正常样本的强判别特征作为监督信号,借助特征级与关系级对齐损失,将这种能力压缩至轻量化学生模型。最终,学生模型可脱离教师模型独立执行高精度异常检测,并在终端设备上实现检测性能与推理效率的有效平衡^[31]。

本研究针对轻量化模型在少样本异常检测中的性能不足问题,以轻量化视觉 Transformer——EfficientViT^[32]为学生模型,通过从大规模预训练模型蒸馏知识,得到一种适用于少样本异常检测的高效检测模型。尽管 EfficientViT 作为一种轻量化视觉 Transformer 模型在参数规模和计算开销上具有显著优势,但其在少样本甚至单一样本条件下的异常检测能力有限,难以有效识别复杂场景中的异常。相比之下,大规模预训练模型 DINOv2^[33]在少样本异常检测中展现了优异的特征提取能力,但其庞大的参数量使其难以在资源受限的场景中直接应用。基于上述内容,本文的主要工作可分为以下4点:

1) 构建基于 DINOv2 与 EfficientViT 的知识蒸馏框架:针对轻量化模型在少样本异常检测中的性能问题,设计以冻结的 DINOv2 为教师模型、可训练的 EfficientViT 为学生模型的蒸馏架构。

2) 设计特征对齐适配器用于解决特征不匹配问题:针对教师与学生模型特征在空间分辨率和通道维度上的差异,提出包含可学习位置编码、基于内容的动态聚合机制和线性投影的三阶段适配器,以实现跨模型特征的精细化空间与通道维度对齐。

3) 提出融合多维度对齐目标的复合损失函数:设计结合 Huber 损失、余弦相似度损失和 Gram 矩阵损失的复合目标函数,分别从特征数值、特征方向和空间关系3个维度约束学生模型,全面学习教师模型的判别性表征。

4) 开展系统性实验验证该方法有效性:在

MVTec AD 标准数据集上,通过设置单样本和多样本,全面评估验证本文所提方法的性能表现。消融实验进一步证实特征适配器和复合损失函数对各模块贡献的有效性,使用可视化分析从定性角度展示方法在异常检测上的优化。

1 实验方法

1.1 少样本异常检测方法框架

少样本异常检测旨在解决工业质检、医疗影像分析等领域中异常样本稀缺或获取成本高的问题。该任务的核心目标是仅利用极少量正常样本,使模型能够快速适应新任务,并对测试样本中的未知异常实现精准识别与定位。

本文提出一种基于预训练特征比对的异常检测方法,其流程如图 1(a) 所示。该方法的有效性基于两个基本假设:首先,在大规模数据集上预训练的神经网络,其提取的特征具有强大的通用表征能力,能够捕获图像中与异常相关的判别性信息^[34];其次,在深度特征空间中,正常样本的特征倾向于聚集在紧凑的子空间内,形成正常原型,而异常样本的特征则会因语义或纹理的未知性与多样性而显著偏离该子空间。该类方法包含两个主要步骤:

1) 构建特征内存库

对于特定的检测类别,该方法首先利用一个或少量正常样本构建一个参考基准,即将正常样本输入一个固定的预训练特征提取模型,获取其多层次的深度特征图。这些特征图中蕴含从细微纹理到高级语义的丰富信息。随后,通过对这些特征进行聚合,构建一个特征内存库 $M = \{m_1, m_2, \dots, m_k\}$, m_k 为特征内存向量。该内存库代表了此类别正常模式在特征空间中的分布,可作为后续比对的原型字典。

2) 异常检测与定位

在检测阶段,待测样本经同一个预训练特征提取器,获得其深度特征图 F_q 。异常检测的核心是通过计算查询特征与内存库中正常原型的差异实现的。对于 F_q 上的每一个局部特征向量 $f_q^{(i,j)}$,计算其与内存库 M 中所有原型特征之间的最小距离为

$$d^{(i,j)} = \min_{m_k \in M} \|f_q^{(i,j)} - m_k\|。 \quad (1)$$

该距离 $d^{(i,j)}$ 可定量反映当前区域特征与正常模式的偏离程度。距离值越大,该区域属于异常的概率越高。最终,将所有空间位置的距离值汇总并上

采样,生成一张高分辨率的异常分数热力图,通过设定阈值可实现像素级的异常区域分割与定位。

1.2 知识蒸馏流程

本文提出了一种基于知识蒸馏的轻量化特征提取器方法,核心思想是以 DINOv2 作为教师模型,从仅包含正常样本的训练集中提取高质量的特征,并将该表征能力迁移到轻量化学生模型

EfficientViT 上,使其兼顾检测性能与推理效率。知识蒸馏框架如图 1(b) 知识蒸馏所示,训练时,输入图像被同时送入教师与学生模型的编码路径:教师模型采用其自身的预处理与编码器直接对原始图像进行特征提取,学生模型则在其前端采用 Overlap PatchEmbed 对图像进行重叠分块以保留细粒度信息,随后通过 EfficientViT 的若干高效 Transformer Block 逐层提取特征。

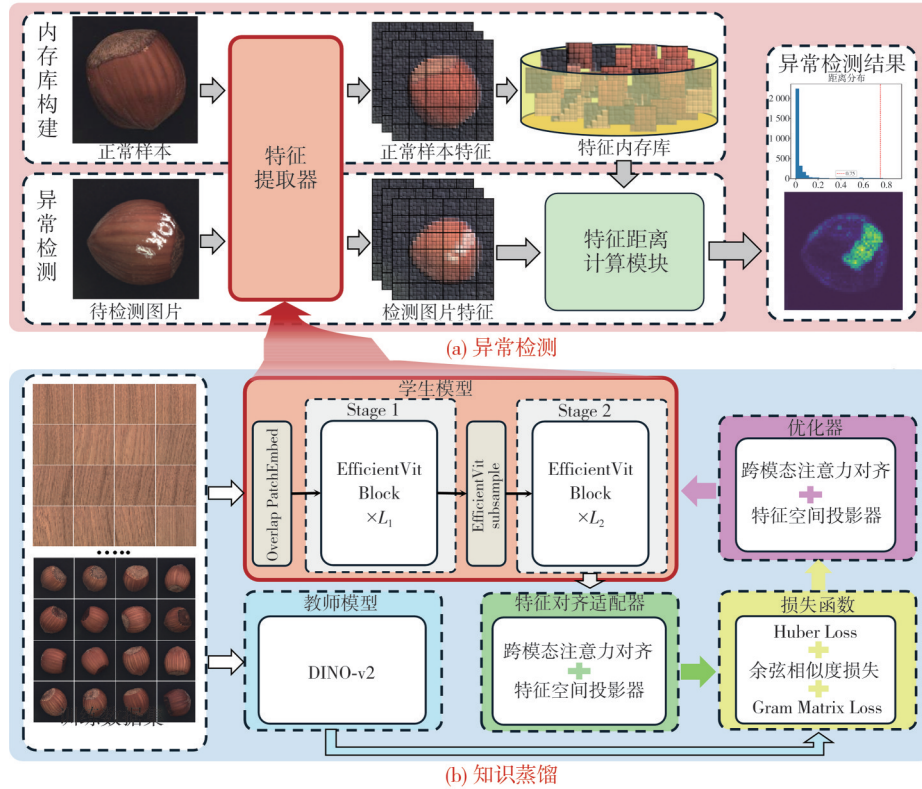


图 1 少样本异常检测框架与知识蒸馏流程图

Fig. 1 Framework for few-shot anomaly detection and knowledge distillation process flow

为缩小教师模型与学生模型在表征空间和维度上的差异,本文引入了特征对齐适配器,包含跨模态注意力对齐模块和特征空间投影器,用于将教师特征映射到与学生特征可比较的表示空间,并进行精细对齐。学生模型的训练目标由多项蒸馏损失共同约束: Huber Loss 提升回归鲁棒性,余弦相似度损失保持特征方向一致性, Gram Matrix Loss 刻画并对齐特征分布的高阶统计特性。需要说明的是,训练数据集中仅含正常样本,蒸馏过程的目标是使学生模型在仅在正常样本的条件下尽可能复现教师对正常样本的判别性表征,训练过程中仅更新学生模型参数,教师模型保持冻结。

1.3 特征适配器

为了将来自学生模型的特征图有效地转换并

对齐到教师模型提取特征的表示空间中,本文设计了一个轻量级的特征对齐适配器模块。如图 2 所示,该模型包含 3 个连续的数据变换阶段,其核心目标是将具有特定空间分辨率和通道维度的输入特征适配为具有目标通道维度且空间信息被有效聚合的表示形式。

该适配器接收形状为 $[B, 3 \times 136, 32]$ 的输入特征,对应于批大小为 B , 56×56 空间分辨率, 32 通道的特征图,输出形状为 $[B, 1 \times 024, 384]$ 的标准化特征表示。该适配器的核心设计包含 3 个连续的数据变换阶段:

首先,为保留输入特征的相对空间关系,引入可学习的位置编码。该过程定义为

$$X' = X + P, \quad (2)$$

式中: $P \in \mathbb{R}^{1 \times 3136 \times 32}$ 为位置编码参数; X' 为通过广播

机制与输入特征相加,得到的空间信息增强的特征

表示。

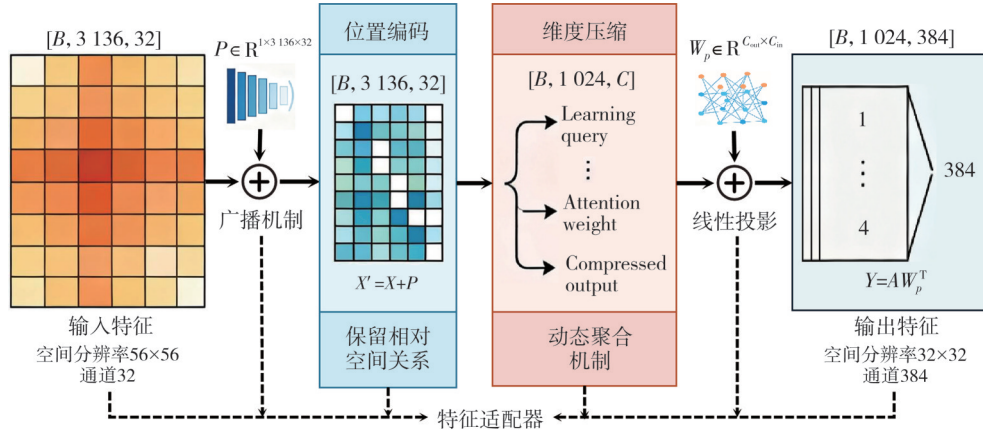


图 2 特征适配器模块示意图

Fig. 2 Schematic diagram of feature adapter module

其次,为实现自适应的空间维度压缩,本文采用基于内容的动态聚合机制,适配器在该机制中引入一组可学习的查询向量 $Q \in \mathbb{R}^{1 \times 1024 \times 32}$,作为聚合的锚点。对于每个输入样本,通过复制扩展得到 $Q_b \in \mathbb{R}^{B \times 1024 \times 32}$ 。聚合过程通过注意力机制实现,其核心是计算查询向量与所有空间位置间的相关性权重,并据此进行加权汇总为

$$A = \text{Attention}(Q_b, X', X'). \quad (3)$$

该操作将高维空间特征压缩至目标维度,同时根据内容相关性保留最重要的空间上下文信息,生成紧凑的中间表示 A 。

最后,通过线性投影完成通道维度的对齐,

$$Y = A W_p^T, \quad (4)$$

式中: $W_p \in \mathbb{R}^{C_{out} \times C_{in}}$ 为投影矩阵, C_{in} 和 C_{out} 分别为输入特征维度与输出特征维度,可将特征从 32 维映射到目标维度 384。该特征对齐适配器通过序列化的位置增强、内容感知的聚合和线性投影,实现了跨模态特征在空间和通道维度上的统一表示。输出特征 Y 可直接作为后续 Transformer 解码器等模块的输入。

1.4 特征对齐损失函数

为实现学生模型特征与教师模型特征在数值、方向和空间关系上的全面对齐,本文设计了一种复合损失函数。该函数通过加权整合 3 种不同目标的损失,引导学生模型高效地蒸馏教师模型的知识。

1) 特征值对齐

为确保特征激活值的匹配,本研究采用 Huber 损失。该损失函数结合了 L1 与 L2 损失的优点:在误差较小时,具备 L2 损失的光滑性,有

利于优化收敛;在误差较大时,则表现出 L1 损失对离群值的鲁棒性。这一特性使其能有效处理特征蒸馏中可能存在的尺度差异与异常激活,有利于保障训练稳定性。该损失项定义为

$$L_{\text{value}} = \frac{1}{BN} \sum_{i=1}^B \sum_{j=1}^N L_{\text{Huber}}(A_s^{(i,j)} - T^{(i,j)}), \quad (5)$$

式中: L_{value} 为 Huber 损失; T 为教师特征; A_s 为学生特征; B 为训练批次; N 为特征总数。Huber 函数 L_{Huber} 的具体形式为

$$L_{\text{Huber}}(x) = \begin{cases} 0.5x^2, & |x| \leq 1, \\ |x| - 0.5, & \text{otherwise.} \end{cases} \quad (6)$$

2) 特征方向对齐

特征向量的方向一般比其绝对值承载更丰富的语义信息。为约束学生特征与教师特征在高维空间中的朝向一致性,如图 3 所示,本文引入余弦相似度损失。该损失通过对特征向量进行 L2 归一化,消除幅值差异的影响,从而专注于学习特征空间中的语义分布模式,这对于下游任务的语义理解至关重要。其计算公式为

$$L_{\text{direction}} = 1 - \frac{1}{BN} \sum_{i=1}^B \sum_{j=1}^N \left(\frac{A_s^{(i,j)}}{\|A_s^{(i,j)}\|_2} \cdot \frac{T^{(i,j)}}{\|T^{(i,j)}\|_2} \right), \quad (7)$$

式中: $L_{\text{direction}}$ 为 Cosine 方向损失。

3) 空间关系对齐

为使学生模型不仅学习点对点的特征,更能模仿教师特征中蕴含的高级空间结构关系,本文采用基于 Gram 矩阵的损失。Gram 矩阵通过计算特征通道间的内积来表征它们之间的空间相关性,能够有效捕捉特征的纹理风格信息。最小化 Gram 矩阵的差异有助于保持特征图整体视觉结构的一致性。该

损失定义为

$$L_{\text{relation}} = \frac{1}{BD^2} \sum_{i=1}^B \|G(A_s^{(i)}) - G(T^{(i)})\|_F^2, \quad (8)$$

式中: L_{relation} 为 Gram 关系损失值; D 为特征维度; $G(F) = F^T F \in \mathbb{R}^{D \times D}$ 为特征 F 的 Gram 矩阵; $\|\cdot\|_F$ 为 Frobenius 范数。

4) 复合损失函数

最终的训练目标由上述 3 种损失函数加权组合而成,

$$L_{\text{align}} = L_{\text{value}} + L_{\text{direction}} + \lambda L_{\text{relation}}. \quad (9)$$

由于 L_{value} 和 $L_{\text{direction}}$ 分别基于逐元素差异和单位向量夹角计算, 具有较好的数值稳定性与尺度一致性, 因此无需额外权重调节。而 L_{relation} 作为 Gram 矩阵损失, 反映的是特征通道间的二阶统计关系, 其值域较大且对特征激活强度敏感, 若直接加入可能导致训练过程中该项主导优化过程。为此, 式(9)引入可学习或可调的平衡系数 λ 控制其相对贡献, 确保多维度特征对齐的协同优化。本文通过实验确定权重系数 $\lambda = 0.5$, 该配置能有效平衡不同对齐目标的影响, 且该复合损失函数通过多层次约束, 共同确保了特征对齐的全面性与有效性, 适用于所有物体类别的异常检测。

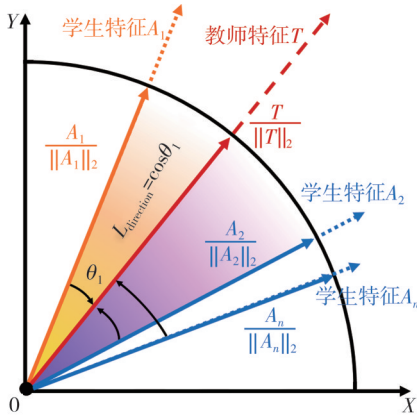


图3 特征方向一致性与余弦损失示意图

Fig. 3 Schematic diagram of feature direction consistency and cosine loss

2 实验结果与分析

2.1 数据集与评估指标

2.1.1 数据集

本研究基于 MVTec AD 数据集^[35-36]进行验证, 该数据集由工业界领先的机器视觉公司 MVTec 构建, 是工业异常检测领域的权威基准数据集。数据集中共包含 15 个不同的工业产品与纹

理类别, 图像总数超过 5 000 张。

本研究严格遵循数据集官方设定的协议, 仅使用各类别的训练集中仅含正常样本进行模型训练。为深入探究所提方法在复杂纹理与结构性缺陷上的检测能力, 本文选取了榛子与木材两个类别作为核心分析案例, 如图 4 所示。该选择基于以下考量: 榛子类别包含了表面纹理变化与微划痕、凹陷等挑战, 而木材类别则代表了具有天然不规则纹理背景下的结构性缺陷检测难题。



(a) 榛子正常样本



(b) 木材正常样本

图4 MVTec AD数据集

Fig. 4 MVTec AD dataset

2.1.2 评估指标

为全面评估本文所提的异常检测方法的性能, 采用以下指标:

1) 图像级准确度 (Area Under the Receiver Operating Characteristic Curve, AUROC): 评估模型区分正常与异常图像的整体能力, 值越接近 1, 性能越优, 计算为 ROC 曲线下面积,

$$AUROC_{\text{image}} = \int_0^1 TPR(FPR) dFPR, \quad (10)$$

式中: TPR 和 FPR 分别为真阳性率和假阳性率, $TPR = \frac{TP}{TP + FN}$, $FPR = \frac{FP}{FP + TN}$, TP 为异常且被模型正确预测为异常的样本数量, FP 为正常

但被模型错误预测为异常的样本数量, TN 为正常且被模型正确预测为正常的样本数量, FN 为异常但被模型错误预测为正常的样本数量。

2) 模型参数量(Parameter)是衡量深度学习模型复杂度的重要指标,反映了网络中可训练参数的总数,包括卷积核权重、偏置项等。参数量越少,通常意味着模型更轻量化,所需的存储空间更小,也更有利于在资源受限的设备上部署。因此,在保证检测性能的前提下,较低的参数量是模型设计的一个重要优化目标。

3) 推理速度(Frames Per Second, FPS)指模型在单位时间内能够处理的图像帧数,是评估模型推理效率的关键指标。较高的FPS值表明模型具有更快的前向推理速度,适用于实时性要求较高的工业检测或在线监控场景。

2.2 训练设置

1) 环境配置

本研究所有实验均在统一的硬件与软件环境下进行,具体配置如表1所示。该计算平台为模型训练与评估提供稳定的支持,可提升实验结果的可靠性与可复现性。

表1 实验环境配置

Tab. 1 Experimental setup

配置类别	具体参数
GPU	NVIDIA GeForce RTX 3090 (显存 24 GB)
CPU	AMD EPYC 7642 48 核
内存	90 GB
操作系统	Ubuntu 22.04 LTS
深度学习框架	PyTorch 2.5.1
CUDA 版本	12.4

2) 优化器与超参数

采用Adam优化器^[37-38]进行模型训练,其参数更新遵循

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}, \quad (11)$$

式中: θ_{t+1} 为第 $t+1$ 步的模型参数; θ_t 为第 t 步的模型参数; η 为学习率; $\hat{m}_t$ 和 $\hat{v}_t$ 分别为经过偏差校正后的一阶矩估计与二阶矩估计; ϵ 为防止分母为零的微小常数。

学生模型主干网络的学习率设置为 1×10^{-4} , 特征对齐适配器模块因其任务特性采用较高的学习率 1×10^{-3} , 以快速适应教师模型的特征分布。训练批大小设置为 16, 总共训练 150 个周期。学习率调度采用步长衰减策略, 每 10 个周期乘以因

子 0.5 进行衰减。

3) 数据增强策略

为EfficientViT能充分学习到DINOv2的特征提取能力,并构建一个稳定且具泛化性的训练环境,训练过程采用了以下数据增强策略:

① 多角度旋转增强:对输入图像依次进行 0° 、 90° 、 180° 及 270° 旋转,有效扩充训练数据量,提升模型对空间变化的鲁棒性;

② 色彩空间转换:将图像从BGR色彩空间转换至RGB色彩空间,以适应预训练模型的输入要求;

③ 标准化处理:使用ImageNet数据集的标准均值和标准差进行归一化处理,确保输入数据分布的一致性。

2.3 实验结果与分析

为验证本文所提基于知识蒸馏的轻量化异常检测方法的有效性,在MVTec AD数据集上进行了深入的实验评估与分析。本实验的核心在于评估不同模型作为少样本异常检测的特征提取器的性能,使用本文方法蒸馏后得到的特征提取模型EfficientViT(本文方法)与教师特征提取模型DINOv2以及未经蒸馏的直接使用的EfficientViT(Baseline)进行了详尽的对比。

2.3.1 定量结果分析

表2展示了基于3种不同特征提取网络的单样本异常检测在AUROC、Parameter及FPS上的性能对比。表2清晰地反映出不同模型在检测精度与推理效率之间的权衡关系,为进一步分析蒸馏效果提供了直观支持。

表2 不同方法在榛子与木材上的平均性能对比

Tab. 2 Comparison of average performance of different methods on hazelnut and wood

模型	AUROC	参数量/M	FPS/M
DINOv2 (Teacher)	0.912	86	30
EfficientViT (Baseline)	0.655	66	120
EfficientViT (本文方法)	0.827	66	120

相较于大型教师模型,该方法在精度有差距但显著提升效率的前提下实现了有效蒸馏。成功地将教师模型DINOv2的部分知识蒸馏到了轻量化的EfficientViT中。

如表2所示,在榛子和木材类别上,本文方法的特征提取模型在图像级检测精度上的AUROC为0.827,仍低于教师模型DINOv2的0.912,但显著

优于直接使用的 Baseline 模型的 0.655。与此同时,模型 *Parameter* 约为 DINOv2 的 77%,推理速度是其 4 倍。这表明本文提出的蒸馏框架在保持较高检测性能的同时,实现了显著的效率提升。相较于轻量基准模型,在效率完全相当的前提下实现精度显著提升。与未经蒸馏的 EfficientViT (Baseline) 相比,本文所提方法的网络结构未变,因此, *Parameter* 和 *FPS* 保持在同一水平,均为 66 M 和 120 M。但通过本文设计的损失函数与适配器,特征提取模型 EfficientViT (本文方法) 在图像级检测精度上的 *AUROC* 从 0.655 提升至 0.827。

除单样本异常检测外,本文还对 2、4、8 和 10 个样本分别进行了异常检测实验,实验结果见表 3。同样,为直观分辨不同样本数量对异常检测结果的性能差异,将实验结果绘制了折线图(见图 5),增加数据的可视化。

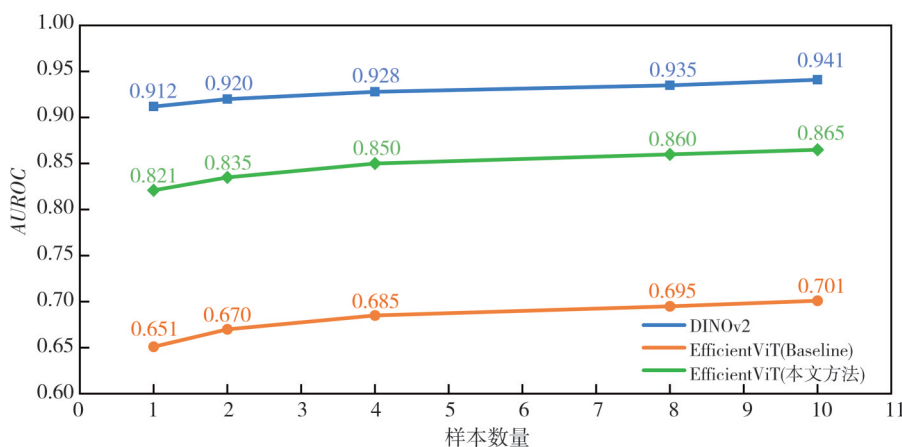


图 5 不同样本数的异常检测指标结果

Fig. 5 Anomaly detection metrics under different sample sizes

从图 5 的性能曲线可以观察到,本文方法从 1 个样本到 10 个样本实现了 0.044 的 *AUROC* 提升,而基线方法提升 0.050,两者提升幅度相近,但本文的方法始终保持在更高的性能水平。这种表现说明本文方法在样本增加时能够保持相对优势。图 5 清晰展示了 3 条性能曲线的对比关系:本文的方法曲线始终位于基线上方,且由于模型的轻量化使其和教师模型始终有差距,但是该差距保持稳定。该结果说明了本文方法在任何样本数量下都优于基线,同时在与高性能教师模型 的比较中展现了良好的性能接近度。

2.3.2 消融实验

为了验证本文所提各模块的必要性,在榛子、

表 3 不同样本数的异常检测的性能对比

Tab. 3 Performance comparison of anomaly detection under different sample sizes

模型	AUROC				
	1 个样本	2 个样本	4 个样本	8 个样本	10 个样本
DINOv2 (Teacher)	0.912	0.920	0.928	0.935	0.941
EfficientViT (Baseline)	0.651	0.670	0.685	0.695	0.701
EfficientViT (本文方法)	0.821	0.835	0.850	0.860	0.865

由图 5 可知,本文的方法在 2、4、8 和 10 个样本的异常测试性能均优于基线方法。本文方法的单样本的 *AUROC* 比基线方法高 0.170,10 个样本的 *AUROC* 比基线方法高 0.164。该多样本的异常检测实验结果证明了本文方法不依赖于特定样本数量的特性。

木材、皮革类别上分别开展了基于单样本学习的异常检测消融实验,以 *AUROC* 为性能指标,结果如表 4 和图 6 所示。

表 4 消融实验结果

Tab. 4 Ablation experiment results

实验配置	AUROC		
	木材	皮革	榛子
Baseline (无蒸馏 EfficientViT)	0.652	0.647	0.661
+Huber Loss	0.665	0.663	0.681
+Huber+Cosine Loss	0.718	0.731	0.724
+Huber+Cosine+Gram	0.782	0.767	0.773
Full(all loss+Feature Adapter)	0.826	0.818	0.834

初始 Baseline 采用未经蒸馏的 EfficientViT 作为学生模型,在木材、皮革和榛子 3 类数据上的 *AUROC* 分别为 0.652、0.647 和 0.661,表明轻量

级网络难以有效捕捉复杂纹理中的异常数据。引入 Huber 损失后模型的 *AUROC* 略有提升, 分别

达到 0.665、0.663 和 0.681, 说明 Huber 损失有助于稳定训练并缓解梯度震荡。

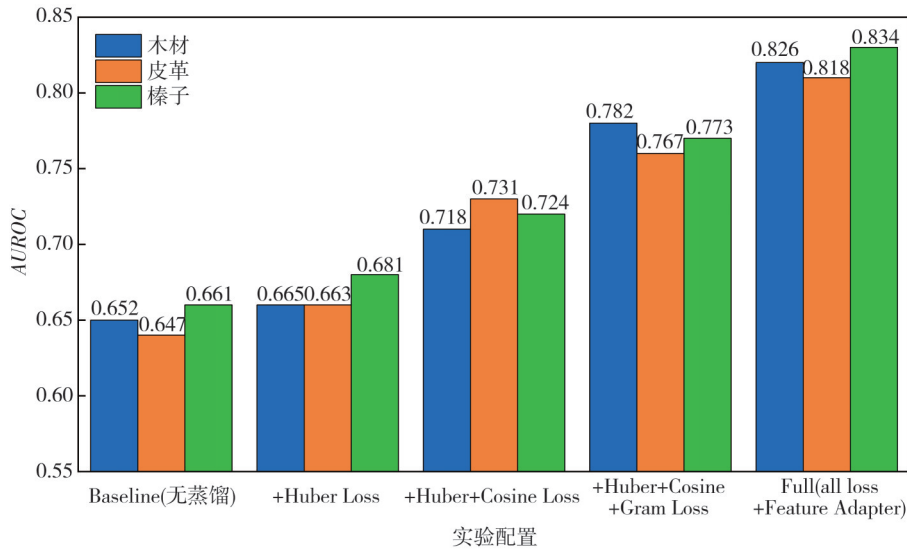


图 6 消融实验结果

Fig. 6 Ablation experiment results

进一步加入余弦相似性损失 (Cosine Loss) 后, 模型性能显著提升, 在木材、皮革和榛子 3 类数据上的 *AUROC* 分别达到 0.718、0.731 和 0.724。该结果表明, 通过增强特征结构对齐, 可更有效地迁移语义信息, 提升对细微差异的鲁棒性。然后, 再引入 Gram 矩阵损失, 其捕捉特征中的二阶统计特性, 此时模型在这 3 类数据的 *AUROC* 进一步提升至 0.782、0.767 和 0.773。该损失在方向性明显的纹理上表现突出, 有助于建模长程依赖和区分真实缺陷。

最终, 模型中加入可学习的 Feature Adapter, 实现动态特征变换, 整体性能达到最优, 3 类数据的 *AUROC* 分别为 0.826、0.818 和 0.834。说明该模块能有效缓解容量差异, 实现更高效的语义知识迁移。

轻量化的 EfficientViT 网络能够提取出与大型教师模型质量相当、且显著优于其原始状态的判别性特征。

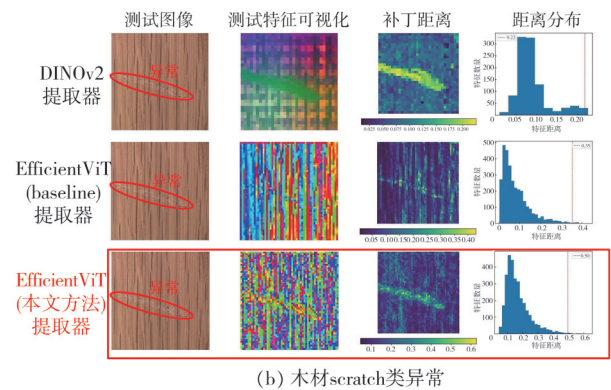
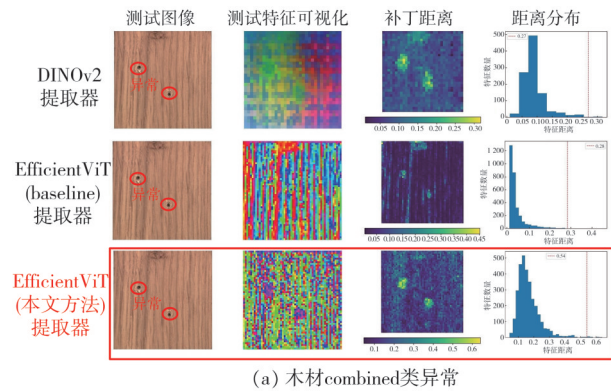


图 7 不同提取器下木材异常检测结果

Fig. 7 Wood anomaly detection results with different extractors

2.3.3 可视化分析

如图 7 和图 8 所示, 本文方法生成的连续值异常热力图, 未经过阈值二值化, 与真实异常区域的空间分布高度吻合, 其对齐程度优于 Baseline 特征的结果, 且与基于 DINOv2 教师网络的热力图效果非常接近。特别地, 图 7 和图 8 中距离分布直方图取所有特征距离中前 1% 为阈值, 用于凸显高异常响应区域的统计特性, 避免正常背景噪声主导整体分布。

这从视觉层面直接印证了经过知识蒸馏后,

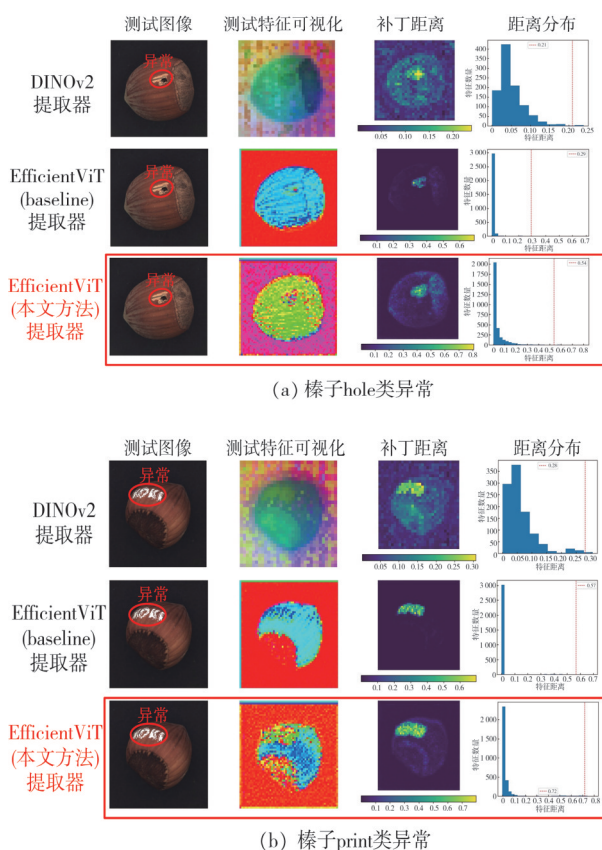


图8 不同提取器下榛子异常检测结果

Fig. 8 Hazelnut anomaly detection results with different extractors

3 结论

本文提出了一种基于知识蒸馏的轻量化异常检测方法,旨在将大规模视觉模型 DINOv2 强大的表征能力高效地迁移至轻量级网络 EfficientViT 中,以实现在资源受限场景下的高性能异常检测。通过在 MVTec AD 数据集的挑战性类别上进行实验,验证了所提方法的有效性。

实验结果表明,相较于庞大的教师模型,本文方法在仅牺牲少量检测精度的情况下,实现了显著的效率提升,模型参数量降至教师模型的约 77%,推理速度提升至其 4 倍;相较于未经蒸馏的轻量基准模型,在模型结构与计算效率完全不变的前提下,将检测性能提升了 26%。消融实验进一步证实了设计的多目标损失函数与可学习的特征适配器模块均为性能提升做出了关键性贡献,以及从不同维度有效地约束和增强了学生模型的特征学习能力。可视化分析表明,蒸馏后的学生模型所产生的异常定位热力图与教师模型效果高度接近,优于基线模型,验证了其生成特征的高质量。

本文工作为在边缘计算等资源受限环境中部署高效、准确的工业异常检测系统提供了一种有效的解决方案。

利益冲突声明/Conflict of Interests

所有作者声明不存在利益冲突。

All authors declare no relevant conflict of interests.

参考文献

- [1] Hojjati H, Ho T K K, Armanfard N. Self-supervised anomaly detection in computer vision and beyond: a survey and outlook[J]. Neural Networks, 2024, 172: 106106.
- [2] 解琳, 韩跃平, 翟倩, 等. 基于改进 YOLOv7-tiny 的 PCB 表面缺陷检测[J]. 测试技术学报, 2025, 39(1): 81-87.
Xie Lin, Han Yueping, Zhai Qian, et al. PCB surface defect detection based on improved YOLOv7-tiny[J]. Journal of Test and Measurement Technology, 2025, 39(1): 81-87. (in Chinese)
- [3] Ren C, Liu D. Anomaly detection and segmentation based on multi-student teacher network[J]. Journal of Measurement Science and Instrumentation, 2022, 13(2): 235-241.
- [4] 李石峰, 罗晰, 刘晓茹, 等. 基于 Transformer 架构的端到端视频异常检测方法[J]. 计算机技术与发展, 2025, 35(6): 49-55.
Li Shifeng, Luo Xi, Liu Xiaoru, et al. An end-to-end video anomaly detection method based on transformer architecture[J]. Computer Technology and Development, 2025, 35(6): 49-55. (in Chinese)
- [5] Zhang X, Xu M, Qiu D, et al. MediCLIP: adapting CLIP for few-shot medical image anomaly detection[C]// Medical Image Computing and Computer Assisted Intervention - MICCAI 2024. Cham: Springer, 2024: 458-468.
- [6] 杨维荣. 基于智能边缘计算的网络安全防护与威胁检测分析[J]. 电子技术, 2024, 53(10): 256-257.
Yang Weirong. Analysis of network security protection and threat detection based on intelligent edge computing [J]. Electronic Technology, 2024, 53(10): 256-257. (in Chinese)
- [7] 郭新茹, 宋丽娟, 朱文倩, 等. 面向低样本的工业图像异常检测综述[J]. 计算机工程与应用, 2025, 61(13): 26-45.
Guo Xinru, Song Lijuan, Zhu Wenqian, et al. Review of low-shot industrial image anomaly detection [J].

- Computer Engineering and Applications, 2025, 61(13): 26-45. (in Chinese)
- [8] Han K, Wang Y, Chen H, et al. A survey on vision transformer[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1): 87-110.
- [9] 邹琦萍, 李博涛, 陈赛安, 等. 基于多层级视频 Transformer 的视觉自动定位方法[J]. 工程科学与技术, 2024, 56(6): 34-43.
Zou Qiping, Li Botao, Chen Saian, et al. Visual automatic localization method based on multi-level video transformer [J]. Advanced Engineering Sciences, 2024, 56(6): 34-43. (in Chinese)
- [10] 薛振华, 李强, 黄超. 视觉基础模型驱动的像素级图像异常检测方法[J]. 计算机应用, 2025, 45(3): 823-831.
Xue Zhenhua, Li Qiang, Huang Chao. Vision foundation model-driven pixel-level image anomaly detection method[J]. Journal of Computer Applications, 2025, 45(3): 823-831. (in Chinese)
- [11] Zhu L, Wang X, Ke Z, et al. BiFormer: vision transformer with bi-level routing attention [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 10323-10333.
- [12] 金泽辉, 陈凯, 唐敦兵, 等. 基于改进 Transformer 的视觉语言模型[J/OL]. 北京航空航天大学学报, 2025-03-19. <https://link.cnki.net/doi/10.13700/j.bh.1001-5965.2024.0656>.
Jin Zehui, Chen Kai, Tang Dunbing, et al. Visual language model based on improved transformer[J/OL]. Journal of Beijing University of Aeronautics and Astronautics, 2025-03-19. <https://link.cnki.net/doi/10.13700/j.bh.1001-5965.2024.0656>. (in Chinese)
- [13] Kaplan J, Mc Candlish S, Henighan T, et al. Scaling laws for neural language models[PP/OL]. arXiv(2020-01-23) [2025-12-11]. <https://doi.org/10.48550/arXiv.2001.08361>.
- [14] Khan A, Rauf Z, Sohail A, et al. A survey of the vision transformers and their CNN-transformer based variants [J]. Artificial Intelligence Review, 2023, 56(3): 2917-2970.
- [15] 刘势杰, 王丽芳, 郁晓庆. 基于改进 U-Net 和跨模态自蒸馏的医学图像融合[J]. 测试技术学报, 2024, 38(6): 686-694.
Liu Shijie, Wang Lifang, Yu Xiaoqing. Medical image fusion based on improved U-Net and cross-modal self-distillation [J]. Journal of Test and Measurement Technology, 2024, 38(6): 686-694. (in Chinese)
- [16] Guo F, Wu X, Zhang L, et al. A self-supervised vision transformer approach for dermatological image analysis [EB/OL]. [2025-12-11]. <https://doi.org/10.13140/RG.2.2.23655.76967>.
- [17] Yang Q, Guo R, Yang Q, et al. An unsupervised method for industrial image anomaly detection with vision transformer-based autoencoder [J]. Sensors, 2024, 24(8): 2440.
- [18] Iqbal E, Khan S U, Javed S, et al. Multi-scale feature reconstruction network for industrial anomaly detection [J]. Knowledge-Based Systems, 2024, 305: 112650.
- [19] 张玉博, 杨帆, 郭亚, 等. 基于视觉大模型的垃圾分类轻量化算法研究[J]. 计算机工程, 2025, 51(7): 140-151.
Zhang Yubo, Yang Fan, Guo Ya, et al. Research on lightweight algorithm for garbage classification based on visual large model [J]. Computer Engineering, 2025, 51(7): 140-151. (in Chinese)
- [20] 李通, 羊红光, 刘康, 等. 大模型知识蒸馏方法研究进展[J]. 河北省科学院学报, 2025, 42(2): 94-96.
Li Tong, Yang Hongguang, Liu Kang, et al. Knowledge distillation methods for large models: a research review[J]. Journal of the Hebei Academy of Sciences, 2025, 42(2): 94-96. (in Chinese)
- [21] 宋建华, 刘淳, 张龔. 基于结构化知识蒸馏的轻量化伪装目标检测模型[J]. 计算机科学, 2026, 53(4): 299-307.
Song Jianhua, Liu Chun, Zhang Yan. Lightweight camouflaged object detection model based on structured knowledge distillation[J]. Computer Science, 2026, 53(4): 299-307. (in Chinese)
- [22] 王莉莉, 殷硕峰, 马也. 基于轻量化 VIT 的面部疲劳监测算法研究[J]. 计算机仿真, 2025, 42(1): 317-322.
Wang Lili, Yin Shuofeng, Ma Ye. Research on facial fatigue monitoring algorithm based on lightweight VIT [J]. Computer Simulation, 2025, 42(1): 317-322. (in Chinese)
- [23] 崔智高, 兰云伟, 王念, 等. 基于多教师引导的知识蒸馏图像去雾算法[J]. 火箭军工程大学学报, 2024(5): 35-43.
Cui Zhigao, Lan Yunwei, Wang Nian, et al. An image dehazing algorithm based on multi-teacher guided knowledge distillation [J]. Journal of Rocket Force University of Engineering, 2024(5): 35-43. (in Chinese)
- [24] 谢奕涛, 苏鹭梅, 杨帆, 等. 面向目标类别分类的无数据知识蒸馏方法[J]. 中国图象图形学报, 2024, 29(11): 3401-3416.

- Xie Yitao, Su Lumei, Yang Fan, et al. Data-free knowledge distillation for target class classification[J]. Journal of Image and Graphics, 2024, 29(11): 3401-3416. (in Chinese)
- [25] Song L, Gong X, Zhou H, et al. Exploring the knowledge transferred by response-based teacher-student distillation[C]//Proceedings of the 31st ACM International Conference on Multimedia, 2023: 2704-2713.
- [26] Yang Z, Li Z, Zeng A, et al. ViTKD: feature-based knowledge distillation for vision transformers[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2024: 1379-1388.
- [27] Sun S, Ren W, Li J, et al. Logit standardization in knowledge distillation[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 15731-15740.
- [28] Yang C, Yu X, An Z, et al. Categories of response-based, feature-based, and relation-based knowledge distillation[M]//Advancements in Knowledge Distillation: Towards New Horizons of Intelligent Systems. Cham: Springer International Publishing, 2023: 1-32.
- [29] 王改华, 李柯鸿, 龙潜, 等. 基于知识蒸馏的轻量化 Transformer 目标检测[J]. 系统仿真学报, 2024, 36(11): 2517-2527.
- Wang Gaihua, Li Kehong, Long Qian, et al. Object detection of lightweight transformer based on knowledge distillation [J]. Journal of System Simulation, 2024, 36(11): 2517-2527. (in Chinese)
- [30] 杨文茵, 黄嵩权, 谭振林, 等. 融合相似交叉熵和知识蒸馏的人脸年龄估计方法[J]. 计算机技术与发展, 2025, 35(4): 113-120.
- Yang Wenyin, Huang Aiquan, Tan Zhenlin, et al. Age estimation of face images by fusing similarity cross entropy and knowledge distillation [J]. Computer Technology and Development, 2025, 35(4): 113-120. (in Chinese)
- [31] 钟锐, 王晨, 宋亚锋, 等. 基于样本均衡蒸馏的少样本人脸识别方法[J]. 计算机工程与设计, 2024, 45(11): 3457-3462.
- Zhong Rui, Wang Chen, Song Yafeng, et al. Face recognition method with few samples based on sample balance distillation [J]. Computer Engineering and Design, 2024, 45(11): 3457-3462. (in Chinese)
- [32] Liu X, Peng H, Zheng N, et al. EfficientViT: memory efficient vision transformer with cascaded group attention[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 14420-14430.
- [33] Oquab M, Darcet T, Moutakanni T, et al. Dinov2: learning robust visual features without supervision[PP/OL]. arXiv(2023-04-14)[2025-12-11]. <https://doi.org/10.48550/arXiv.2304.07193>.
- [34] Damm S, Laszkiewicz M, Lederer J, et al. Anomaly DINO: boosting patch-based few-shot anomaly detection with DINOv2[C]//2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2025: 1319-1329.
- [35] Bergmann P, Batzner K, Fauser M, et al. The MVTec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection [J]. International Journal of Computer Vision, 2021, 129(4): 1038-1059.
- [36] Bergmann P, Fauser M, Sattlegger D, et al. MVTec AD: a comprehensive real-world dataset for unsupervised anomaly detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 9584-9592.
- [37] 闫佳乐. 基于 EIS 的退役锂离子电池 SOH 估计研究[D]. 合肥: 合肥学院, 2023: 22.
- Yan Jiale. Research on State of health estimation of retired lithium-ion batteries based on EIS[D]. Hefei: Hefei University, 2023: 22. (in Chinese)
- [38] 陈万强, 陆永华, 朱赞, 等. 基于深度学习图像配准的燃调表面缺陷检测算法[J]. 测试技术学报, 2024, 38(6): 586-592.
- Chen Wanqiang, Lu Yonghua, Zhu Yun, et al. Research on surface defect detection algorithm of fuel pump regulator based on deep learning image registration[J]. Journal of Test and Measurement Technology, 2024, 38(6): 586-592. (in Chinese)