文章编号:2096-1472(2025)10-0072-07
DOI:10.19644/j.cnki.issn2096-1472.2025.010.015

基于多组师生网络分层特征融合框架的异常检测模型

董浩然[1]，业 宁[1]，周小亮[2]

(1.南京林业大学信息科学技术学院，江苏 南京 210037；
2.南京理工大学计算机科学与工程学院，江苏 南京 210094)
✉ 1626835594@qq.com; yening@njfu.edu.cn; xiaoliangzhou@njust.edu.cn

摘 要:为了提升图像异常检测模型对复杂异常的识别率以及对多尺寸异常的检测稳定性,提出了基于多组师生网络分层特征融合框架的异常检测模型(Multi-group Teacher-student Network Hierarchical Feature Fusion Framework, MTHF)。通过构建3组不同的师生检测模块,在每组学生网络的蒸馏层后引入特征融合模块,提升多尺度异常检测精度,采用两阶段融合策略实现异常区域全面覆盖的同时保证边缘清晰。在MVTec AD数据集上进行实验,结果表明,MTHF模型的平均检测精度相比当前最先进方法提升了7.62%,充分证明了模型在图像异常检测任务中的有效性。

关键词:异常检测;知识蒸馏;特征融合;注意力机制
中图分类号:TP391.4 **文献标志码**:A

Anomaly Detection Model Based on Hierarchical Feature Fusion Framework of Multi-group Teacher-student Network

DONG Haoran[1], YE Ning[1], ZHOU Xiaoliang[2]

(1. *College of information science and technology & Artifical Intelligence, Nanjing Forestry University, Nanjing* 210037, *China*;
2. *School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing* 210094, *China*)
✉ 1626835594@qq.com; yening@njfu.edu.cn; xiaoliangzhou@njust.edu.cn

Abstract: To enhance the recognition rate of image anomaly detection models for complex anomalies and improve detection stability across multi-scale anomalies, this paper proposes an anomaly detection model based on a Multi-group Teacher-student Network Hierarchical Feature Fusion Framework (MTHF). By constructing three distinct teacher-student detection modules and introducing feature fusion modules after the distillation layers of each student network, the model enhances multi-scale anomaly detection accuracy. A two-stage fusion strategy achieves comprehensive coverage of anomalous regions while preserving sharp boundaries. Experiments on the MVTec AD dataset demonstrate that the MTHF model improves the average detection accuracy by 7.62% compared to state-of-the-art methods, validating its effectiveness in image anomaly detection tasks.

Key words: anomaly detection; knowledge distillation; feature fusion; attention mechanism

0 引言(Introduction)

近年来,知识蒸馏技术为异常检测带来了新的解决思路。但是现有模型存在复杂异常识别精度低和不同尺度异常识别精度波动大的问题。针对以上问题,本文提出了基于多组师生网络分层特征融合框架的异常检测模型(MTHF)。首先,模型基于ResNet50、VGG16和ResNet34构建了3组不同架构的师生网络,利用不同网络对特征的捕捉优势,弥补了单一架构的局限性,用以增强对复杂异常的特征提取能力。进一步的,在每组学生网络的蒸馏层后引入特征融合模块,用以充分融合浅层和深层特征,增强对不同尺寸异常的检测能力,解决在多尺

收稿日期:2024-12-12

度异常检测中的不足。最后,模型对3组检测结果采用两阶段融合策略,优化异常区域的主体定位和边缘界定。同时保留边缘精准性和定位完整性,生成更加精确的异常定位图。为了验证MTHF模型有效性,在MVTec AD数据集上进行实验。实验结果表明,所提出模型的检测精度相比当前最先进方法提升了7.62%。

1 相关工作(Related work)

1.1 基于知识蒸馏的异常检测

知识蒸馏模拟人类教育过程[1-3],其中具备强大特征提取能力的教师网络指导较简单的学生网络学习。教师网络的知识可以是其输出层的决策信息或中间层的特征表示,分别对应输出层知识蒸馏与特征层知识蒸馏。学生网络通过模仿教师网络的输出和特征激活模式,实现对教师网络知识的继承。知识蒸馏的核心在于将教师网络学习到的特征和决策规则转移给学生网络,使后者在对蒸馏知识的学习过程中达到或接近教师的水准。模仿的过程通过损失函数的迭代来实现,在特征层蒸馏中,损失函数L_{FeaD}如式(1)所示:

$$L_{\mathrm{FeaD}}(f_t(x),f_s(x))=L_F(\Phi_t(f_t(x)),\Phi_s(f_s(x))) \quad (1)$$

其中:$f_t(x)$和$f_s(x)$分别代表教师网络和学生网络对于输入数据x生成的特征图,$\Phi_t(\cdot)$和$\Phi_s(\cdot)$表示对特征图的投影变换函数,用来将师生特征调整至同一维度,L_F表示损失计算方法,通常可以使用均方误差、l_2距离损失。

知识蒸馏技术在异常检测以外的其他领域也被广泛应用。XU等[4]提出了一种使用知识蒸馏构建深度脉冲神经网络模型,通过联合训练算法合理地构建更高效的深度脉冲结构,对各类人工噪声和自然信号具有极好的抗噪能力,应用于实际场景的轻量级、高效的类脑计算。张晓涌等[5]使用更适合生成语言模型上的反向蒸馏,以防止学生模型高估教师分布的低概率区域。生成的响应更精确,总体质量更高,曝光偏差更低,校准更好,长文本生成性能更高。SHU等[6]将每个通道的激活图归一化,得到软概率图。通过简单地最小化两个网络的通道概率图之间的KL散度,蒸馏过程更加关注每个通道最突出的区域,从而完成密集的预测任务。

1.2 混合注意力机制

在注意力机制的多样化实现中,两大主流方法分别是空间注意力和通道注意力,它们各自聚焦于输入特征的不同维度来优化信息处理。混合注意力机制(Fusion Attention Model,FAM)包含通道注意力、空间注意力、交叉融合模块3个部分,结构如图1所示。

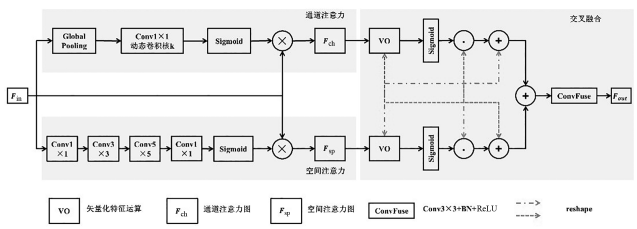


图1　混合注意力模块结构示意图

Fig. 1 Schematic diagram of the structure of the hybrid attention module

如图1所示,先分别提取输入特征图的通道注意力特征和空间注意力特征,再将其交叉融合传递到最终特征图F_{out}中。交叉融合模块用于将通道注意力特征和空间注意力特征融合形成交叉注意力特征。交叉融合模块运算如式(2)~式(4)所示:

$$\bar{F}'_{\mathrm{sp}}=\mathrm{Softmax}(\bar{F}_{\mathrm{sp}}(\bar{F}_{\mathrm{ch}})^{\mathrm{T}})\bar{F}_{\mathrm{ch}} \quad (2)$$

$$\bar{F}'_{\mathrm{ch}}=\mathrm{Softmax}(\bar{F}_{\mathrm{ch}}(\bar{F}_{\mathrm{sp}})^{\mathrm{T}})\bar{F}_{\mathrm{sp}} \quad (3)$$

$$F_{\mathrm{out}}=\mathrm{ReLU}(\mathrm{BN}(F_{\mathrm{ch}}+\bar{F}'_{\mathrm{ch}}+F_{\mathrm{sp}}+\bar{F}'_{\mathrm{sp}})) \quad (4)$$

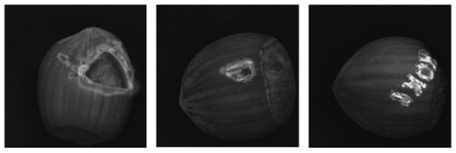
其中:F_{ch}和F_{sp}分别表示通道注意力特征图和空间注意力特征图,$\bar{F}_{\mathrm{ch}}$和$\bar{F}_{\mathrm{sp}}$分别表示F_{ch}和F_{sp}的矢量化特征,$\bar{F}'_{\mathrm{ch}}$和$\bar{F}'_{\mathrm{sp}}$表示交互矢量特征图。

1.3 异常检测

通过教师学生模型架构,隐式学习正常数据的特征分布,并利用预训练模型实现高效的无监督学习。BERGMANN等[1]首次将此框架应用于异常检测,通过比较学生网络和教师网络的输出差异,有效提升了异常的识别能力。SALEHI等[2]提出的多分辨率知识蒸馏策略,旨在通过从多个中间层提取特征,加强学生网络对教师网络正常图像特征的学习。在此基础上,MA等[3]使用基于自动编码器的重构过程来更新教师重要性权重。通过学生网络的重构误差来检测异常。

然而,现有方法依赖单一骨干检测网络,限制了对异常样本特征的全面提取,进而影响了对复杂异常的识别精度。如图2所示,在榛子类样本测试结果中,虽然破口类和孔洞类异常的检测精度较为稳定,但在处理形态较为复杂的喷涂类异常时,检测区域覆盖不够全面,检测精度明显波动。

(a)原始图像



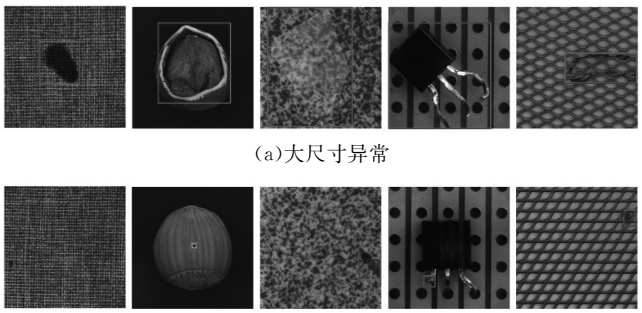
(b)测试结果

(c)基准答案

图2　榛子样本中不同形态异常检测结果

Fig. 2 Detection results of different morphological abnormalities in hazelnut samples

此外,此类模型在应对尺度多变的异常时也存在局限性。随着卷积层深度的增加,感受野发生变化,不同尺度的异常特征提取能力出现差异。而现实情况中,检测样本的异常尺寸往往差异很大,如图3所示。



(a)大尺寸异常

(b)小尺寸异常

图 3 不同尺寸异常样本

Fig. 3 Abnormal samples of different sizes

2 MTHF 模型(MTHF model)

2.1 模型总体架构

针对骨干网络结构单一,且未能充分融合各层级特征,导致现有模型对新型异常的识别率较低,并且在应对不同尺寸的异常时检测精度波动较大的问题,本文提出了基于多组师生网络分层特征融合框架的异常检测模型(MTHF),框架如图 4 所示。

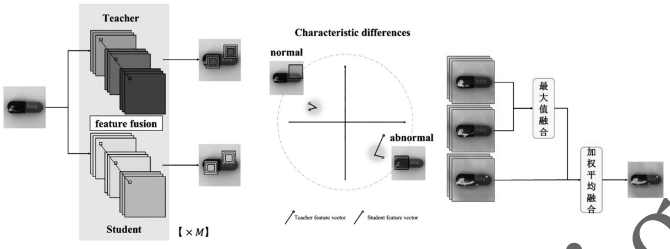


图 4 模型总体结构

Fig. 4 Overall structure of model

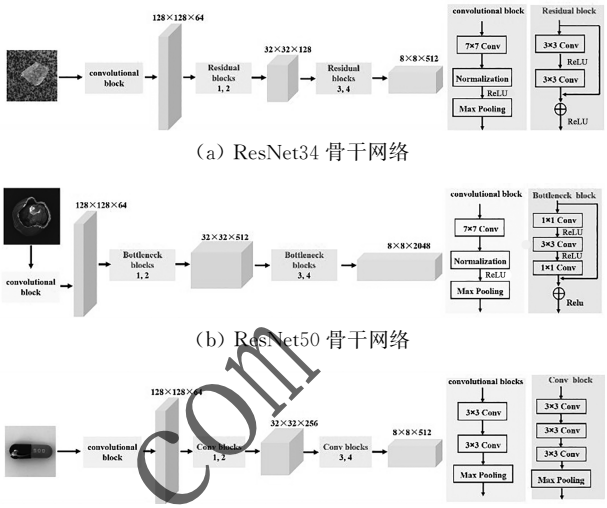
如图 4 所示,MTHF 采用 M 组(取 $M=3$)并联的师生网络结构,分别基于 ResNet34、ResNet50 和 VGG16,以捕获多样化的特征表示。每组网络的教师网络均在 ImageNet 上进行预训练,迁移在大规模数据上学到的边缘、纹理等特征,避免特征提取不足问题,同时提升了模型的泛化能力,更好地应对训练集中缺失的异常样本。学生网络采用与对应教师网络相同的架构,通过训练阶段完善的模仿教师网络对于正样本的特征表征。并在 3 层蒸馏层后加入特征融合模块,将浅层和深层特征进行融合,以更好地处理不同尺度的异常。

输入图像经过师生组网络,分别提取出教师特征和学生特征。对于同一块图像区域,未出现异常时,教师和学生的特征向量在角度和方向上高度一致,反之,对于异常区域通过向量间差异完成定位。通过特征差异,每组师生网络分别得到一张异常定位图。

3 组师生检测模块结果的融合策略分为两阶段。第一阶段是对两组 ResNet 师生网络输出的检测结果进行最大值融合,以提高对局部异常的检测能力;第二阶段则对一阶段融合的输出结果和 VGG 组检测结果进行加权平均融合,生成最终的异常检测图,确保不同特征层次的有效整合和检测结果的鲁棒性。

2.2 师生检测模块

MTHF 在 3 组师生检测模块中分别使用了 ResNet34、ResNet50 和 VGG16 做为骨干网络,用来平衡全局与边缘特征的提取能力。在骨干网络中选取第 1、3、5 层特征层作为蒸馏层进行特征蒸馏以提供完整的特征表示,从而提升蒸馏效果,如图 5 所示。



(a) ResNet34 骨干网络

(b) ResNet50 骨干网络

(c) VGG16 骨干网络

图 5 3 组骨干网络结构图

Fig. 5 Three sets of backbone network structure diagrams

如图 5(a)所示,ResNet34 是一个 34 层的相对较浅层网络,使用标准的残差块(Residual Block),每块包含两个 3×3 卷积层,通过跳跃连接解决深层网络的梯度消失问题。因为较浅的结构可以更集中于细粒度的特征提取,而不会像更深的网络那样过度平滑或者丢失细节,所以 ResNet34 能够更好地捕捉图像的边缘细节和局部特征,特别是在纹理类异常检测中表现突出。

同时为了在物体类样本检测中保持精度,使用 ResNet50 网络。如图 5(b)所示,ResNet50 是一个 50 层的深层网络,使用瓶颈块(bottleneck block),每块包含 3 个卷积层(1×1,3×3,1×1)。用于降维和恢复维度,以减轻计算负担。更深的网络使它能够学习更抽象和高级的特征,尤其适用于物体类异常检测。但瓶颈块的降维操作也导致其对细节的关注减少,ResNet50 在处理细微局部异常或复杂纹理时表现不够敏感。

在 ResNet 系列网络的检测结果中,存在异常区域漏检的问题,为此引入 VGG16 网络。如图 5(c)所示,VGG16 是一个 16 层的浅层网络,它使用多层 3×3 的小卷积核结构,通过多个卷积层堆叠,增大感受野,同时保留空间细节,对异常区域进行更加全面的覆盖。其在全局特征提取上表现更为稳定,尤其适用于复杂纹理类异常或局部特征不明显的场景。

综合使用 3 组网络,可以做到对异常的全面覆盖和检测边缘清晰。VGG16 能够覆盖整个异常区域,确保复杂异常区域不会被漏检,弥补了 ResNet 系列网络在异常区域覆盖上的不足。两种 ResNet 网络擅长边缘检测,能够捕捉细微的局部异常,与 VGG16 网络形成互补。分别将 3 组骨干网络构建为教

师学生网络,在特征层之间进行知识蒸馏,具体流程如图 6 所示。

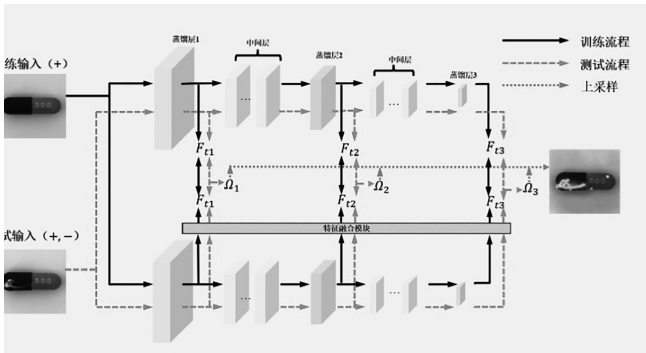


图 6　师生组网络结构图

Fig. 6 Network structure diagram of teacher-student group

在训练阶段,教师加载预训练参数并保持固定,学生网络加载随机参数。在两组网络中选取相同层级的卷积层作为中间层实施知识蒸馏过程。将正样本分别输入教师网络和学生网络,计算其通过教师网络和学生网络所得特征表示之间的差异,得到样本的特征损失。通过损失函数回归分析,优化出学生网络对教师网络正样本特征表示的最佳模仿参数。

在测试阶段,将测试样本输入教师网络和学生网络,利用教师网络和学生网络所得特征表示之间的差异进行异常检测。学生网络对于正常数据得到的特征表示与教师网络相同,对于异常数据,学生网络与教师网络得到的特征表示有较大差异,据此,将总损失用作异常得分。对于判断为异常的样本,计算教师网络和学生网络各层所得特征表示之间的差异,得到差异矩阵,并进行上采样以获得异常图。

2.3　特征融合模块

为了充分融合浅层和深层特征,增强对不同尺度异常的检测能力。在学生网络的蒸馏层上,MTHF 通过构建特征金字塔结构,实现特征融合,结构如图 7 所示。

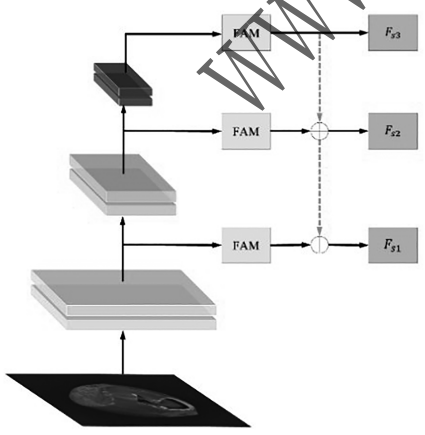


图 7　特征融合模块示意图

Fig. 7 Schema of the feature fusion module

如图 7 所示,特征融合模块将学生网络的 3 层蒸馏层的输出特征图来进行两两融合。具体而言,最高层的输出特征图首先通过混合注意力模块 FAM 处理得到 F_{t3},随后,将经过 FAM 处理的前一层特征图与 F_{s3} 相加生成 F_{s2},遵循同样模式得到 F_{s1}。为了匹配尺寸,对高层特征图实施上采样操作与低层特征图融合。

通过限定融合顺序,确保低层特征仅整合了紧邻层特征,而非直接包含高层的特征信息,避免了跨尺度特征融合时容易引发的混淆效应,从而抑制了定位精度下降。F_{s1}、F_{s2} 及 F_{s3} 各自集成不同层级的特征差异,有效放大了学生网络与教师网络在特征表示上的差异。同时,相邻层间的两两融合不仅维持了不同分辨率特征图对异常尺寸变化的敏感度,还确保了各层在特征表达上的相对独立性,避免了过度融合可能引起的特征模糊。

2.4　损失函数

MTHF 通过比较学生网络与教师网络对于异常样本的特征差异,实现异常检测。因此,在正样本上需要实现师生特征提取达到高度一致,设置蒸馏损失函数的目标是在训练阶段获得一个对教师模型在正常样本上特征表征有很好模仿效果的学生模型。为了充分利用特征向量间的距离和角度差异,MTHF 中每层蒸馏层将常规的距离损失函数(L_{dir})和余弦相似性度函数(L_{val})相结合,如式(5)和式(6)所示:

$$L_{\mathrm{val}}=\frac{1}{N_i}\sum_{i=1}^{N_i}(F_{ti}-F_{si})^2 \tag{5}$$

$$L_{\mathrm{dir}}=1-\sum_{i=1}^{N_i}(F_{ti}-F_{si})\frac{V(F_{si})^T\cdot V(F_{ti})}{\|V(F_{si})\|\,\|V(F_{ti})\|} \tag{6}$$

其中:N_i 为第 i 个蒸馏层所有神经元个数。F_{ti} 和 F_{si} 分别表示教师网络和学生网络第 i 层的特征输出。$V(\boldsymbol{y})$ 表示将任意维度 $\boldsymbol{y}$ 向量矩阵转化为 1 维向量,$\|y\|$ 表示 $\boldsymbol{y}$ 向量的范数。L_i 表示第 i 个蒸馏层的损失函数。由距离损失和方向损失相加得到,如式(7)所示:

$$L_i=L_{\mathrm{val}}+L_{\mathrm{dir}} \tag{7}$$

ResNet34,ResNet50,VGG16 为骨干网络的师生组检测模块训练阶段各自独立,损失函数分别对应为 L_{ResNet34},L_{ResNet50},L_{VGG16}。对于每组检测模块,共设置了 3 层特征蒸馏,考虑到不同网络在不同特征层之间的特征提取侧重有所区别,故每组总损失由每层损失加权相加得出,则总损失表示为 L_{GroupFea},GroupFea∈{ResNet34,ResNet50,VGG16},如式(8)所示:

$$L_{\mathrm{GroupFea}}=\sum_{i=1}^{3}\theta_i\cdot L_i \tag{8}$$

其中:θ_i 表示第 i 层特征对应的权重。

2.5　两阶段融合

在测试阶段,对于判断为异常的样本,计算教师网络和学生网络各层所得特征表示之间的差异,得到差异矩阵,并进行上采样以获得异常图。对于一张样本图片,3 组师生检测网络共得到 3 张结果图。

首先,通过最大值融合充分利用两组 ResNet 网络在边缘检测中的优势,提取异常区域的最显著特征。随后,采用加权平均的方式,将 VGG16 的更为全面的特征与 ResNet 的局部边缘特征相结合,形成兼具全面覆盖和边缘精准的融合结果。缓解 VGG16 在边缘检测上的模糊问题,并增强 ResNet 网络在复杂异常区域的覆盖能力。

第一阶段从 ResNet34、ResNet50 两个模型输出中，逐像素选择对最大的异常值作为融合后的对应异常值，以此填补漏检测区域，放大异常。对于任意像素位置(i,j)，最大值融合后的结果图在该位置的异常值$\Omega_{\max}(i,j)$定义为

$$\Omega_{\max}(i,j)=\max(\Omega_{\text{ResNet34}}(i,j),\Omega_{\text{ResNet50}}(i,j)) \quad (9)$$

其中：$\Omega_{\text{ResNet34}}(i,j)$和$\Omega_{\text{ResNet50}}(i,j)$分别为对应师生组检测结果图在该位置的异常值。

第二阶段在第一阶段结果的基础上，将$\Omega_{\max}$与 VGG16 模型输出的Ω_{VGG16}进行融合，引入权重因子α和β（其中，$\alpha+\beta=1$），用于平衡最大值融合结果$\Omega_{\max}(i,j)$与 VGG16 模型输出Ω_{VGG16}的重要性。逐像素的进行加权平均融合，融合公式为

$$\Omega_{\text{final}}(i,j)=\alpha\cdot\Omega_{\max}(i,j)+\beta\cdot\Omega_{\text{VGG16}}(i,j) \quad (10)$$

3 实验(Experiment)

3.1 实验设置

使用 MVTec AD 数据集进行实验，该数据集涵盖了 15 种类别的共 5354 张高分辨率彩色图像。图像被划分为两类，其中：纹理类别包括 carpet、grid、leather、tile、wood 5 种；物体类别包括 bottle、cable、capsule、hazelnut、metal-nut、pill、screw、toothbrush、transistor、zipper 10 种。训练集包括 3629 张正常样本图像，测试集包含正常样本和各种类型的异常样本及其对应的二值异常图，用来进行结果评估。在预处理阶段，所有输入图像被统一调整至 256×256 像素尺寸，以确保模型输入的一致性。

训练配置方面，每组模型均历经了 500 轮迭代训练，选用随机梯度下降(Stochastic Gradient Descent，SGD)作为优化器，动量参数设为 0.9，学习率设为 0.4，批次大小为 32。

3.2 实验结果

为了验证模型的有效性，MTHF 对比了以下先进模型：

SSIM-AE[7]：采用了基于结构相似性指数的感知损失函数，通过量化图像局部区域间的内在联系，同时兼顾亮度、对比度及结构特征，来评估图像质量。

Patch-SVDD[8]：将深度支持向量数据描述框架拓展至基于图像块的自监督学习范式，模型能进一步实现对异常区域的精确分割。通过将 DSDD 与图像块分析相结合，模型得以从局部视图中学习到丰富的特征表示。

FCCD[9]：利用编码-解码架构捕捉正常图像的特征并重构它们。模型通过比较输入图像与其重构版本之间的差异来检测异常，差异较大的区域被标记为可能的异常点，从而实现异常的检测与定位。

CNN-Dict[10]：通过提取训练集中图像的局部子图来构建一个描述正常图像特征的字典，将待测图像的子图与字典中的正常子图进行对比，评估它们之间的相似度。

RDOE[11]：提出了一个由教师编码器和学生解码器组成的新型模型，并相应地引入了一个简单但有效的“逆蒸馏”范式。引入了一个可训练的一类瓶颈嵌入模块。获得的紧凑嵌入有效地保留了正常模式的基本信息。

实验中，图像级精度可以在宏观上判断模型是否漏检或误检。像素级精度则关注模型在微观上的定位能力。

使用 AUCROC（Area Under the Receiver Operating Characteristic Curve)作为精度评估指标。同时，各模型检测精度在不同样本类别上由于图像特征复杂度差异表现出不同的浮动。使用变异系数(Coefficient of Variation，CV)作为精度稳定性的衡量标准，通过标准差与均值的比率，提供无量纲的波动性指标，使得在不同类别之间的相对波动量化，准确地衡量方法稳定性。实验像素级精度结果见表 1。

表1 各常见模型在 MVTec AD 数据集中异常定位像素级 AUCROC 值

Tab. 1 Pixel-level AUCROC values of various models for anomaly localization in the MVTec AD dataset

类别	SSIM-AE/%	CNN-Dict/%	Patch-SVDD/%	FCCD/%	RDOE/%	MTHF/%
地毯	87.44	78.39	93.29	97.75	98.27	**99.73**
铁网	93.85	63.69	96.31	95.26	94.45	**97.33**
皮革	78.62	87.74	97.54	98.10	98.02	**99.59**
地砖	60.17	93.06	91.36	93.62	91.43	**98.81**
木板	72.79	91.86	90.28	95.92	93.22	**99.17**
瓶子	93.20	78.24	98.45	96.42	98.45	**99.42**
电缆	85.36	79.21	**97.46**	94.60	94.71	96.75
药丸	94.22	84.35	95.24	95.25	98.34	**99.91**
榛子	97.61	72.38	97.23	97.58	96.04	**98.35**
螺母	89.63	82.98	98.73	98.37	96.28	**99.33**
药片	90.98	68.13	94.42	97.41	94.34	**98.27**
螺丝	96.38	87.78	95.66	93.61	97.73	**98.41**
牙刷	92.25	90.14	98.28	95.73	98.14	**99.58**
晶体管	90.57	66.55	94.51	95.03	**97.87**	97.28
拉链	80.26	76.87	95.11	97.16	**98.44**	97.83
平均	86.96	80.06	95.59	96.12	96.38	**98.65**
CV	11.71	11.62	2.65	1.63	2.34	**1.02**

由表 1 可知，MTHF 在全部 15 个类别中的 12 个上体现出最优精度，这是由于模型在学生网络中加入了特征融合模块，充分融合了浅层和深层特征，增强对不同尺度异常的检测能力。同时进一步放大了教师网络与学生网络在异常样本上的特征差异，提高了检测精度。而在剩余的电缆、晶体管、拉链 3 个类别上检测精度低于最优值，这是由于以上样本中的部分异常与正常区域颜色信息较为接近，结构复杂且限于某些小的局部区域，需要更精确的局部信息，而加权融合策略在捕捉这些微小异常时不够敏感。

同时，MTHF 对于不同检测类别的检测精度波动较现有模型明显缩小，精度最高的药丸类和最低的电缆类精度差值只有 3.16%，总体精度变异系数也为最优值 1.02%，可以看出，MTHF 的精度波动最小。这是由于模型设置了 3 组不同的师生组进行检测，并采用了两阶段的结果融合，利用了不同骨干网络对特征的捕捉优势，弥补了单一网络架构的局限性，增强了对正常样本多样化特征的提取能力，并优化了异常区域的主体定位和边缘界定。综上可以得出，本文方式实现了稳定性和精准性的提升。

实验对于不同样本检测均值精度的图像级和像素级评估结果如图 8 所示。

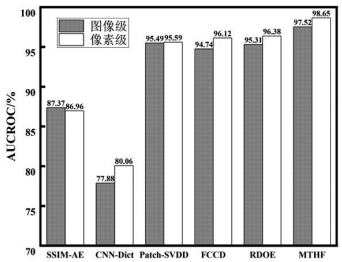
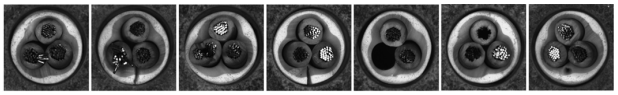


图 8　常见模型在 MVTec AD 数据集中异常定位精度

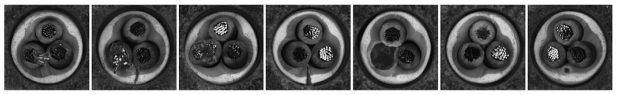
Fig. 8 Accuracy for anomaly localization of various models in the MVTec AD dataset

由图 8 所示,MTHF 的均值精度在图像级和像素级结果评估中都为最优值,分别高于次优值 2.03%和 2.27%,高于各模型均值(不含 MTHF)7.36%和 7.62%。

MTHF 对于同一测试主体(以电缆类为例,MTHF 对于电缆类样本的检测精度在全部 15 类中最低)的不同异常类型,包括不同尺寸的异常的检测可视化结果如图 9 所示。



(a)原始图像



(b)检测结果

(c)基准答案

图 9　电缆类样本各类异常检测结果图

Fig. 9 Diagram of various abnormal detection results for cable samples

由图 9 所示,在缺金属线类这种异常区域临近像素存在混淆性较高的正常区域的情况下,模型会出现误检的可能。但检测精度在尺寸变化和类型变化时还是保持了较高的稳定性。

3.3　消融实验

3.3.1　师生组检测模块和结果融合模块有效性验证

为了验证 MTHF 中师生组检测模块和结果融合模块的有效性,对师生组检测模块中 ResNet34、ResNet50 和 VGG16 3 组网络采用不同的组合选取和结果融合方式进行比较。实验中将最大值融合记为 MF,加权平均融合记为 WAF,试验分组及结果分别如表 2、图 10 所示。

表 2　师生组检测模块和结果融合模块消融实验

Tab. 2 Ablation study on teacher-student group detection modules and result fusion module

Group	ResNet34	ResNet50	VGG16	MF	WAF
A	√	×	×	×	×
B	×	√	×	×	×
C	×	×	√	×	×
D	√	√	×	√	×

续表

Group	ResNet34	ResNet50	VGG16	MF	WAF
E	√	×	√	×	√
F	×	√	√	×	√
G	√	√	√	√	√

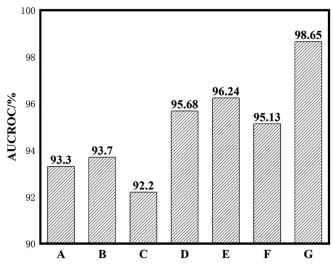


图 10　师生组消融实验结果图

Fig. 10 Results of ablation experiment on teacher-student groups

由图 10 可知,相较于未进行多组师生融合检测的 A、B、C 3 组,使用融合策略的 4 组检测精度均有上升。MTHF 采用的 G 组,同时使用 3 组网络,并先对 ResNet34 组和 ResNet50 组进行最大值融合,再对 VGG16 组进行加权平均融合时,平均检测精度达到最高的 98.65%。这是由于 3 组网络结构不同,检测侧重不同,并行使用可以更好地实现对异常的全面检测。而在边缘清晰整体有缺漏的 ResNet34 组和 ResNet50 组之间进行最大值融合,保持了低误检率的同时,提高了精度。在此基础上将一阶段融合结果与 VGG 组进行加权平均融合,进一步结合了 VGG 网络检测结果的全面性。

3.3.2　特征融合模块有效性验证

为了验证特征融合模块的有效性,将不添加特征融合的检测结果,与仅进行相邻层上采样的特征融合以及 MTHF 使用的添加了 FAM 的特征融合进行精度比较,结果见表 3。

表 3　特征融合模块消融实验

Tab. 3 Ablation experiments of feature fusion module

特征融合模块	FAM	AUCROC/%
×	×	94.83
√	×	96.13
×	√	95.02
√	√	98.65

由表 3 可知,单独添加不含 FAM 的特征融合模块使均值精度提升了 1.3%,在特征融合模块中引入 FAM 使均值精度继续提升 2.52%。这是由于混合注意力模块提升了学生网络对于特征的提取能力,增大了其与教师网络间的对异常样本提取特征的差异,降低漏检几率。而特征融合模块的引入充分融合了浅层和深层特征,解决在多尺度异常检测中的不足。

4　结论(Conclusion)

本文提出了一种基于多组师生网络分层特征融合框架的异常检测模型(MTHF),通过结合多组不同架构的师生网络与多层次特征融合策略,提升了异常检测的精度和鲁棒性。模型中的多组师生网络架构通过捕捉不同层次的特征,提升了对各种异常的感知能力,而分层次的特征融合策略则进一步增强了

模型对不同尺度异常的适应性。此外，混合注意力机制有效提高了模型对异常区域的敏感度，通过最大值融合与加权平均融合的双重结果融合策略，模型能够更加精准地定位异常区域。结果表明，MTHF 模型在检测精度和稳定性方面均优于现有的最先进方法。

尽管 MTHF 模型在实验中表现出色，但也存在一些局限性，如多组网络架构对计算资源的需求较高，训练和推理的效率需要进一步优化。未来的工作将致力于提高模型的计算效率，借助自适应学习或生成对抗网络等技术，增强模型对新型异常的检测能力。增强其对新型异常的适应性，并探索其在更多应用领域中的潜力，为异常检测技术的发展提供新的思路。

参考文献(References)

[1] BERGMANN P，LWE S，FAUSER M，et al. Improving unsupervised defect segmentation by applying structural similarity to autoencoders[C]// SCITEPRESS. Proceedings of the 14th International Conference on Computer Vision Theory and Applications (VISAPP). Prague: SCITEPRESS，2019：372-380.

[2] SALEHI M，SADJADI N，BASELIZADEH S，et al. Multiresolution knowledge distillation for anomaly detection [C]// IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE，2021：1466-1475.

[3] MA X，WU J，XUE S，et al. A comprehensive survey on graph anomaly detection with deep learning[J]. IEEE transactions on knowledge and data engineering，2021，35(12)：12012-12038.

[4] XU Q，LI Y，SHEN J，et al. Constructing deep spiking neural networks from artificial neural networks with knowledge distillation[C] // IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE，2023：7886-7895.

[5] 张晓涌，王黎明，李璇，等. 基于反向知识蒸馏的工件异常检测算法研究[J]. 激光杂志，2024，45(8)：92-97.

[6] SHU C，LIU Y，GAO J，et al. Channel-wise knowledge distillation for dense prediction[C]// IEEE. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Montreal：IEEE，2021：5311-5320.

[7] BAKUROV I，BUZZELLI M，SCHETTINI R，et al. Structural similarity index (SSIM) revisited: a data-driven approach[J]. Expert systems with applications，2022，189：116087.

[8] YI J，YOON S. Patch SVDD: patch-level SVDD for anomaly detection and segmentation[C]// Springer. Proceedings of the European Conference on Computer Vision (ECCV). Glasgow，UK：Springer，2021：336-352.

[9] SELVARAJU R R，COGSWELL M，DAS A，et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[J]. International journal of computer vision，2020，128(2)：336-359.

[10] NAPOLETANO P，PICCOLI F，SCHETTINI R. Anomaly detection in nanofibrous materials by CNN-based self-similarity[J]. Sensors，2018，18(1)：209-209.

[11] DENG H，LI X. Anomaly detection via reverse distillation from one-class embedding[C]// IEEE. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE，2022：9737-9746.

作者简介：

董浩然(1999-)，男(汉族)，合肥，硕士生。研究领域：图像异常检测，深度学习。

业　宁(1967-)，男(汉族)，南京，教授，博士。研究领域：数据挖掘，机器学习。本文通信作者。

周小亮(1997-)，男(汉族)，南京，硕士。研究领域：大模型知识蒸馏，机器学习。

(上接第 66 页)

[10] LIU S C，FENG F，HUANG X G，et al. Novel neuro-coupling matrix technique for parametric modeling of microwave filters[J]. IEEE microwave and wireless technology letters，2024，34(5)：871-874.

[11] JIN J，FENG F，NA W，et al. Advanced cognition-driven EM optimization incorporating transfer function-based feature surrogate for microwave filters[J]. IEEE transactions on microwave theory and techniques，2021，69(1)：15-28.

[12] WANG R，XU J. Extracting coupling matrix and unloaded Q from scattering parameters of lossy filters[J]. Progress in electromagnetics research-pier，2011，115：303-315.

[13] RAYAS-SÁNCHEZ J. EM-based optimization of microwave circuits using artificial neural networks: the state-of-the-art[J]. IEEE transactions on microwave theory and techniques，2003，52(3)：555-567.

作者简介：

刘伶龙(2000-)，男(汉族)，西安，硕士生。研究领域：无源滤波器建模与优化。

刘文远(1989-)，女(汉族)，西安，讲师，博士。研究领域：高频电路与器件建模及优化方法。本文通信作者。

延　博(1998-)，男(汉族)，榆林，硕士生。研究领域：无源滤波器建模与优化。