中图法分类号:TP391　　文献标识码: A　　文章编号: 1006-8961(2025)10-3215-15

论文引用格式:Liu J M and Zhuang W K. 2025. Industrial anomaly detection by combining visual Mamba and patch feature distribution. Journal of Image and Graphics, 30(10):3215-3229(刘建明, 庄维宽. 2025. 结合视觉Mamba和块特征分布的工业异常检测. 中国图象图形学报, 30(10):3215-3229)[DOI:10.11834/jig.240594]

结合视觉Mamba和块特征分布的工业异常检测

刘建明[1,2*],庄维宽[1]

1. 江西师范大学数字产业学院,上饶　334000; 2. 江西师范大学计算机信息工程学院,南昌　330000

摘　要: 目的　工业异常检测在现代工业生产中具有至关重要的作用,现有的工业异常检测方法主要是基于卷积神经网络(convolutional neural network,CNN)或视觉变换器(vision Transformer,ViT)网络来实现。然而,CNN存在难以处理长距离依赖关系的不足,而ViT又面临时间复杂度高的问题。基于此,提出一种结合视觉Mamba和块特征分布的无监督工业异常检测模型。**方法**　该模型包含两个互补分支网络:块特征分布估计网络和基于视觉Mamba的自编码重建网络。块特征分布估计网络主要依赖局部块特征进行异常检测,通过融合高效的预训练块特征描述网络以及视觉Mamba编码器提取的正常样本的块特征,学习一个高斯混合密度网络来估计正常样本局部块特征的分布。在测试阶段利用高斯混合密度网络估计异常图像的各个位置的异常得分,从而得到一个局部异常得分图(local anomaly map,LAM);基于视觉Mamba的自编码重构网络则利用视觉Mamba编码器来捕捉长距离关联特征,增强对跨不同类别和形态的复杂异常图像的全局建模能力,在测试阶段利用重建误差估计异常图像的全局异常得分图(global anomaly map,GAM);最后,合并LAM和GAM得到最终检测结果。**结果**　在MvTec AD(MvTec anomay detection dataset)、VisA和BTAD(bean tech anomaly detection)等公开数据集上与其他先进算法进行了比较,取得了有竞争力的结果。在MvTec AD数据集上所提模型相比性能第2的模型在像素级上AU-ROC(area under the receiver operating characteristic curve)指标提升了0.9%,在图像级上AU-ROC指标提升了2.4%。在BTAD数据集所提模型相比性能第2的模型在图像级上AU-ROC提升0.4%。在VisA数据集上模型相比性能第2的模型在像素级上AU-ROC指标提升了0.6%。**结论**　将视觉状态空间用于图像重建检测图像异常是可行的,检测效果具有竞争力。

关键词: 异常检测;异常分割;视觉状态空间模型(SSM);高斯密度混合网络;异常数据集

Industrial anomaly detection by combining visual Mamba and patch feature distribution

Liu Jianming[1,2*], Zhuang Weikuan[1]

1. School of Digital Industry, Jiangxi Normal University, Shangrao　334000, China;

2. School of Computer and Information Engineering, Jiangxi Normal University, Nanchang　330000, China

Abstract: Objective Industrial image anomaly detection plays a crucial role in modern industrial production because it can timely detect defects in products, effectively improve product qualification rate, enhance industrial productivity, and reduce production costs. Traditional anomaly detection algorithms often show certain limitations when dealing with new

收稿日期:2024-10-21;**修回日期:**2025-01-18;**预印本日期:**2025-01-25

*通信作者:刘建明　liujianming@jxnu.edu.cn

基金项目:国家自然科学基金项目(62266022);江西省自然科学基金项目(20242BAB25110)

Supported by: National Natural Science Foundation of China (62266022); Natural Science Foundation of Jiangxi Province, China (20242BAB25110)

types of anomalies, especially complex issues such as logical anomalies. Thus, they have difficulty meeting the demand for high-precision and efficient detection in industrial production. Therefore, this study is committed to exploring the potential application of visual state space in the field of image processing and anomaly detection. The aim is to find a more effective method for addressing the shortcomings of traditional algorithms in detecting new types of anomalies, especially the limitations in handling logical anomalies. The reconstruction-based method is considered capable of addressing logical anomalies caused by factors such as object quantity, structure, position, and arrangement order because using only normal images to train the model will result in significant errors in the reconstructed output compared with images containing logical anomalies. Existing reconstruction-based anomaly detection methods are mainly based on convolutional neural networks (CNNs) or vision Transformer (ViT) networks. However, CNN exhibits difficulty in handling long-distance dependencies, while ViT presents high time complexity. The latest research shows that state space models represented by Mamba can effectively model long dependencies while maintaining linear complexity. We have explored the potential application of visual state space in anomaly detection and aspire to develop a more precise and efficient image anomaly detection technology by leveraging its advantages to meet strict quality control requirements in industrial production. This endeavor will drive industrial production toward intelligent automation direction while improving overall efficiency and competitiveness. **Method** A novel unsupervised industrial anomaly detection model combining visual Mamba and patch feature distribution is proposed. This model consists of two complementary branch networks: a patch feature distribution estimation network and a self-encoding reconstruction network based on visual Mamba. The patch feature distribution estimation network primarily relies on local patch features for anomaly detection. It fuses local patch features of normal samples through the Vision Mamba encoder and pretrained efficient patch description network and learns a Gaussian mixture density network to estimate the distribution of these features. During the testing phase, this Gaussian mixture density network is used to estimate anomaly scores at various positions in the anomalous images, which produces a local anomaly map (LAM). Meanwhile, the self-encoding reconstruction network based on visual Mamba utilizes a visual Mamba encoder to capture long-range associated features, which enhances the global modeling capability for complex anomaly detection across different categories and forms. In the testing phase, reconstruction errors are used to estimate a global anomaly map (GAM) for the anomalous images. Finally, LAM and GAM are combined to obtain the final detection results. For the dataset, we conducted detailed preprocessing and clipped the images to appropriate sizes according to the requirements of different models. For example, the size of the input image was 256 × 256 pixels. We carefully adjusted the number of coding blocks in the encoder of the visual state space in the reconstruction method to achieve the best anomaly detection performance and maximize the overall performance of the model. The experiments in this study were conducted on a desktop computer equipped with an Intel Core i5, 2.5 GHz CPU, GeForce GTX 3060Ti GPU with 12 GB memory, 32 GB RAM, and Ubuntu18.04 as the operating system. According to our experimental observations, we set the learning rate to 0.001, configured the model to run for 200 epochs, and determined a batch size of 48. Regarding the selection of image blocks, in the PDN method combined with Patchsize, we chose a value of 32. **Result** We compared our model with other advanced algorithms on publicly available datasets such as MvTec AD, VisA, and BTAD, and our model demonstrated highly competitive performance. On the MvTec AD dataset, our model improved the pixel-level AU-ROC metric by 0.9% to reach 93.9%, and the image-level AU-ROC metric by 2.4% to reach 93.8%, compared with the second-best performing model. On the BTAD dataset, our model achieved a 0.4% improvement in image-level AU-ROC (reaching 92.6%) compared with the second-best performing model. On the VisA dataset, our model achieved a 0.6% improvement in pixel-level AU-ROC (reaching 96.6%) compared with the second-best performing model. According to visualizations of anomaly localization in our study on MvTec and VisA datasets, the anomaly localization of our model is more accurate than those of other models. **Conclusion** The application of visual state space to image reconstruction for detecting image anomalies is a feasible and effective method, and its anomaly localization effect has significant competitiveness. This study believes that aggregating features in the middle of the extraction model will be more helpful for adapting to anomaly detection tasks. The setting of the number of image block vectors may be helpful for the localization and detection of anomalies because more image block descriptor vectors can represent more detailed information. The two points are worth further research in the future. This study organically combines two popular methods in the industrial anomaly detection field while integrating visual state space into the model, which sup-

ports its application in the field of anomaly detection.

Key words: anomaly detection; anomaly segmentation; vision state space model (SSM); Gaussian density approximation network; anomaly dataset

0 引 言

工业图像缺陷自动检测是生产智能化的关键问题之一。基于视觉的自动缺陷检测已广泛应用于半导体制造、纺织工业以及航空航天等领域，及时发现缺陷对于生产过程中的质量控制具有重要意义。异常检测任务可以有监督或无监督方式进行。基于监督学习的异常检测方法（Pang 等，2022）在检测精度和鲁棒性方面都有了显著的提高。但是，有监督检测方法不可避免地需要大量的标记数据训练模型，以学习有效的特征表示，从而使模型具有更高的泛化能力。然而，实际工业生产中可用的异常数据稀缺，很难收集到足够的缺陷样本。

近年来，无监督异常检测方法不断发展（Batzner 等，2024；Deng 和 Li，2022；Roth 等，2022；Bae 等，2023；Kim 等，2023；Cohen 和 Hoshen，2020）。基于特征嵌入的异常检测算法（Cohen 和 Hoshen，2020）使用记忆库对卷积神经网络（convolutional neural network，CNN）提取的正常图像特征进行存储，用正常特征与异常特征之间的距离衡量异常程度，但是记忆库需要大量存储空间。为了缓解这个问题，对特征分布建模的方法被提出（Roth 等，2022；Defard 等，2021），这些方法利用多元高斯分布等参数化模型进行建模或对内存库二次采样进行优化。然而，上述方法由于只利用了局部信息，对逻辑异常（丢失、放错或多余的物体，或违反几何约束，例如螺丝的长度等）检测的效果并不理想（Liu 等，2023）。

利用自编码器进行图像重建也是图像异常检测的解决方案之一，它可以用于解决一些逻辑异常的问题，如对象的错误排序（Batzner 等，2024）。基于自动编码器重建的方法依赖于正常图像的准确重建和异常图像的不准确重建，这使得通过将重建图像与输入图像进行比较来检测异常成为可能。早期的方法（Zhou 等，2020；Bergmann 等，2019a）采用基于 CNN 的自编码重建模型进行异常检测。视觉 Transformer 能够有效捕捉长距离关联特征，被用于替换 CNN 实现图象的重建（Lee 和 Kang，2022；Pirnay 和 Chai，2022；Yang 和 Guo，2024），取得了不错的效果。然而，视觉 Transformer 计算量较大，训练和部署成本高昂。近年来，基于知识蒸馏的师生网络（Bergmann 等，2020；Zhang 等，2023；Batzner 等，2024）在公开可用的数据集上表现良好，图像级检测指标已接近饱和。然而，像素级异常检测仍然非常具有挑战性。

最近，基于状态空间模型（state space model，SSM）的视觉 Mamba（vision Mamba，ViM）架构（Zhu 等，2025）被提出，它在维持线性计算复杂度的同时，能够有效建模长距离依赖关系。这对处理图像的全局信息和理解图像内容位置关系很有帮助，这些信息可以更好地进行图像重建，并且它所具有的线性复杂度可以在处理高分辨率图像时进行更高效的计算。

基于以上原因，本文提出一种结合视觉 Mamba 和块特征分布的无监督工业异常检测方法（industrial anomaly detection by combining visual Mamba and patch feature distribution，VMPFD-AD）。一方面，通过高斯混合密度模型建模正常样本的块特征分布，利用局部特征概率密度估计异常块的得分；另一方面，为了解决传统方法对逻辑异常检测不敏感的问题，提出一种基于视觉 Mamba 自编码重建的异常检测互补网络。利用视觉 Mamba 编码器捕捉长距离关联特征，增强跨不同类别和形态的复杂异常图像的全局建模能力。最后，合并两个互补分支网络的结果以提高异常检测性能。在公开数据集的实验表明，本文提出的 VMPFD-AD 方法达到了最先进性能。

1 相关研究

1.1 工业异常检测

近年来，深度学习技术在工业异常检测领域发展迅猛，众多基于深度学习的无监督方法相继提出并应用。按照核心理论不同，大致可分为以下几类：1）基于记忆库方法（Roth 等，2022；Bae 等，2023；Kim 等，2023）。这类方法通过利用神经网络强大的特征提取能力，提取图像的块特征，并将其存储到记忆库

中，后续将测试图像特征与之对比判断是否异常。然而，这类方法需要大量的存储空间，对大规模的图像特征处理可能会有困难。2）基于特征分布建模的方法（Rudolph 等，2021；Defard 等，2021；Gudovskiy 等，2022）。这类方法对神经网络提取的特征利用多元高斯分布等参数化模型进行特征分布建模。然而，该类方法对数据的分布情况非常敏感，如果数据分布发生变化，可能会导致模型的性能下降。3）师生模型（Bergmann 等，2020；Zhang 等，2023a；Gu 等，2023；Batzner 等，2024）。通过使用预训练模型充当教师网络在正常图像下训练学生网络，模型在异常图像推理下学生网络输出有别于教师网络。这类方法充分利用了教师网络在大规模数据上学习到的特征表示和知识，但是如何选择教师网络、学生网络的设计以及模型训练的整体思路等都会影响模型的整体性能。4）基于图像重建的方法（Zhou 等，2020；Bergmann 等，2019a；Lee 和 Kang，2022；Pirnay 和 Chai，2022）。这类方法期望模型能够将异常图像重建为正常图像，借助重建图像与输入的异常图像之间的差异实现异常检测。然而，图像重建的精度对异常检测的结果影响重大，如果重建图像与输入图像之间的差异较小，可能会导致误判。5）结合不同的方法解决异常检测问题。最近的 EfficientAD（Batzner 等，2024）结合了基于重建和师生网络的方法来解决包括逻辑异常在内的各种异常检测问题。Sugawara 和 Imamura（2024）在上述方法的基础上加上基于特征分布建模的方法，利用多元高斯分布拟合特征分布，并用马氏距离（Rippel 等，2021）计算异常分数。王素琴等人（2024）针对工业产品表面相似特征缺陷的检测问题，提出一种基于 YOLOv5（you only look once）的差异化检测网络 YOLO-Differ。这类方法为本文方法提供了参考。

1.2 小样本异常检测

RegAD（registration based few-shot anomaly detection）（Huang 等，2022）是最早探索少量样本异常检测方法的研究之一，PatchCore（Roth 等，2022）和 DifferNet（Rudolph 等，2021）也在少量样本场景中展现了优异性能。WinCLIP（Jeong 等，2023）将 CLIP（contrastive language-image pre-training）模型应用于异常检测，大幅提升了少量样本场景下的检测性能。最新的 FastRecon（fast feature reconstruction）（Fang 等，2023）利用分布正则化回归重建异常特征，表现出卓越性能。

1.3 视觉状态空间模型

状态空间模型因为处理长语言序列建模方面的有效性而受到广泛关注。Gu 等人（2022）提出 S4（structured state space sequence model）模型。这是一种 Transformer 的替代方案，用于建模长距离依赖关系。Smith 等人（2023）通过在 S4 层引入 MIMO（multiple-input multiple-output）SSM 和高效并行扫描，提出新的 S5 层。Gu 和 Dao（2023）提出数据依赖的 SSM 层，并构建了一个通用语言模型骨干 Mamba，它在各种数据集上的表现优于各种规模的 Transformer。Mamba 的成功引发其在计算机视觉的广泛研究。Zhu 等人（2025）提出的 ViM 采用双向序列建模，在自然图像分类任务上取得巨大成功。

2 本文方法

本模型结合了局部特征和全局特征，其结构如图 1 所示。整个网络分为两个子网络：块特征分布估计网络和视觉 Mamba 自编码重建网络。在训练阶段，给定图像 $\boldsymbol{X} \in \mathbf{R}^{H \times W \times C}$（$C$ 是通道数，H 和 W 分别是高度和宽度）输入到这两个子网络：一方面，块特征分布估计网络首先通过预训练的高效块特征提取网络提取正常样本的局部块特征，然后融合视觉 Mamba 编码器提取的全局特征得到增强后的特征，最后，利用上述增强特征训练高斯混合概率密度估计网络，在测试阶段，该网络用于估计异常图像各个位置的异常得分，从而生成局部异常得分图（local anomaly map，LAM）；另一方面，视觉 Mamba 自编码重建网络首先对输入图像 $\boldsymbol{X}$ 进行分块，得到一系列图像块 $\boldsymbol{X}_p \in \mathbf{R}^{N \times (P \times P \times C)}$，$(P, P)$ 是图像块维度，N 是图像块的数量，由 $N = HW/P^2$ 计算所得，然后通过线性层对这些图像块进行特征嵌入，并在图像块向量中添加位置向量以保留位置信息，接着使用 4 层视觉 Mamba 对其进行编码，将得到的特征输入由 6 层转置卷积模块构成的解码器以重建原始图像。在测试阶段，异常图像和重建图像的误差用于估计全局异常的得分图（global anomaly map，GAM）。最后，将 LAM 和 GAM 加权融合，获得最终检测结果。

2.1 基于视觉 Mamba 的自编码重建网络

传统自编码器利用卷积神经网络进行图像重

建，由于视觉 Mamba 能够在线性复杂度下捕捉长距离关联特征，它增强了在不同类别和形态下复杂异常检测的全局建模能力，因此，本文将视觉 Mamba 用于自编码器的图像重建，采用 4 层视觉 Mamba 块作为编码器，6 层转置卷积块作为解码器。在测试阶段，根据重构误差生成全局异常得分图 GAM。具体结构如图 1(a)所示。

2.1.1　视觉状态空间块 ViM

原始的 Mamba 模块是为一维序列设计的，它并不适合视觉任务。ViM 模块(Zhu 等，2025)为了满足视觉任务的需求引入了双向序列建模。ViM 块的操作流程如图 1(c)所示。输入标记序列 $\boldsymbol{T}_{l-1}$ 首先经过归一化层。接下来，将归一化序列线性投影到维度为 E 的 $\boldsymbol{x}$ 和 $\boldsymbol{z}$，$\boldsymbol{x}$ 和 $\boldsymbol{z}$ 分别是经过线性投影得到的主信号和门控信号。然后，对 $\boldsymbol{x}$ 进行前向和后向处理。对于每个方向，首先将 1-D 卷积应用于 $\boldsymbol{x}$ 并获得 $\boldsymbol{x}'_{\mathrm{o}}$。然后，分别将 $\boldsymbol{x}'_{\mathrm{o}}$ 线性变换得到 $\boldsymbol{B}_{\mathrm{o}}, \boldsymbol{C}_{\mathrm{o}}, \boldsymbol{\Delta}_{\mathrm{o}}$。再使用 $\boldsymbol{\Delta}_{\mathrm{o}}$ 分别变换出 $\overline{\boldsymbol{A}}_{\mathrm{o}}, \overline{\boldsymbol{B}}_{\mathrm{o}}$。最后，通过输入 $\overline{\boldsymbol{A}}_{\mathrm{o}}, \overline{\boldsymbol{B}}_{\mathrm{o}}, \boldsymbol{C}_{\mathrm{o}}, \boldsymbol{x}'_{\mathrm{o}}$ 到 SSM 中得到两个值，记为 $\boldsymbol{y}_{\mathrm{forward}}$ 和 $\boldsymbol{y}_{\mathrm{backward}}$。这两个值通过与 $\boldsymbol{z}$ 相乘进行门控并相加，最后通过线性层且与初始 $\boldsymbol{T}_{l-1}$ 相加获得输出标记序列，记为 $\boldsymbol{T}_{l}$。

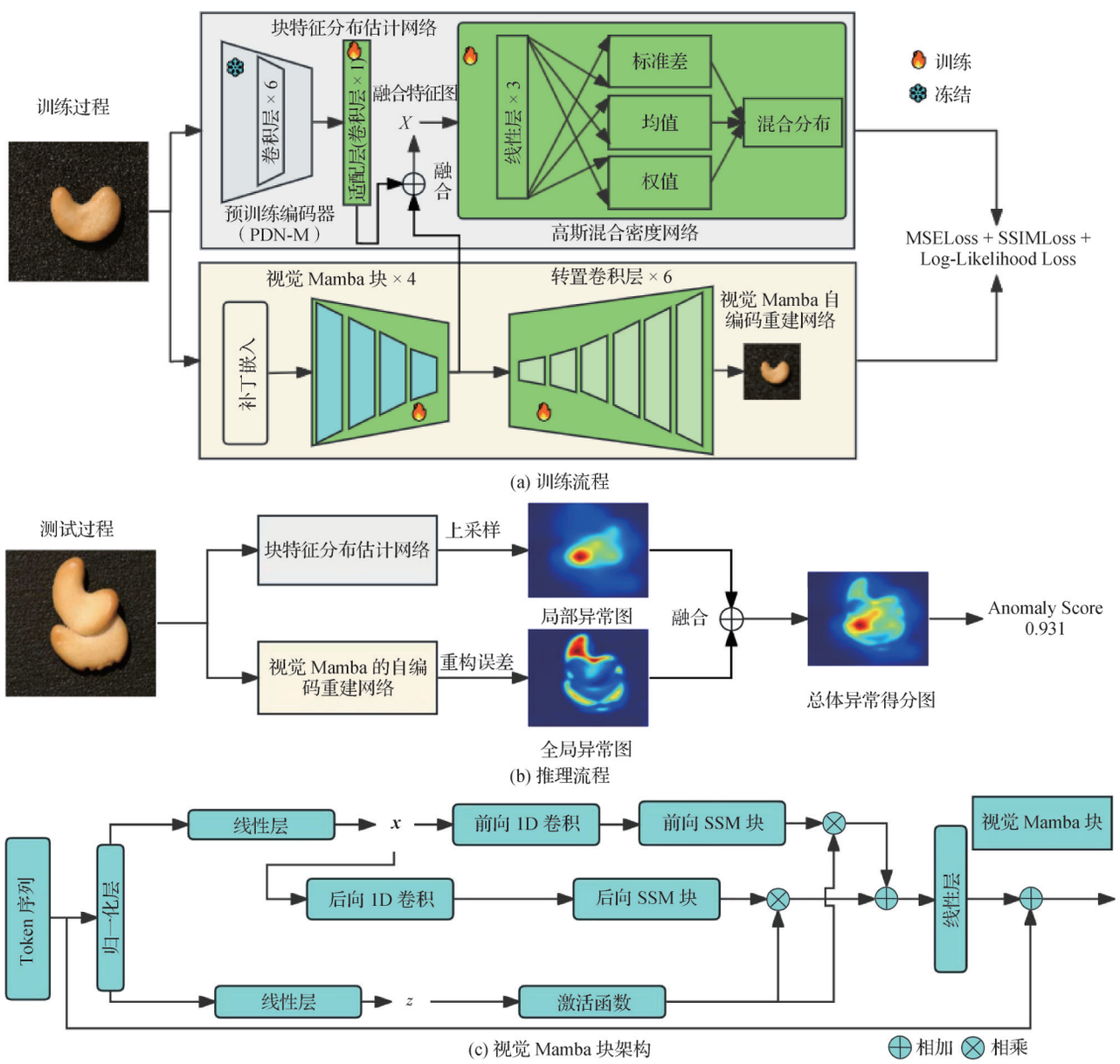

图 1　本文方法框架

Fig. 1　The framework of the proposed method((a)training pipeline;(b)inference pipeline;(c)architecture of vision Mamba module)

2.1.2 解码器

解码器用于将重构向量解码回原始图像形状。在本文的实验中，使用了6个转置卷积层，中间使用批量归一化层和ReLU（rectified linear unit）激活函数，最后一层使用tanh作为最终的非线性激活函数，具体结构如图1（a）所示。

2.2 块特征分布估计网络

本文利用补丁描述网络（patch descriptors network，PDN）（Batzner等，2024）作为块特征提取器，提取正常图像的特征，该特征经过一个可学习的卷积层以适应异常检测任务，最后将输出的特征用于训练高斯混合密度网络（Gaussian mixture density network，GMDN）（Bishop，1994），以获得正常样本局部特征的概率密度分布。GMDN的作用是估计密度模型的条件分布$p(\boldsymbol{y}|\boldsymbol{x})$，其中$\boldsymbol{x}$是图像的低维特征，$\boldsymbol{y}$是目标变量。本文密度模型采用具有满协方差矩阵$\boldsymbol{\Sigma}_k$的高斯混合模型（Gaussian mixture model，GMM）。GMM的概率密度函数$\hat{p}(\boldsymbol{y}|\boldsymbol{x})$由$K$个高斯函数的加权和构成，具体为

$$\hat{p}(\boldsymbol{y}|\boldsymbol{x})=\sum_{k=1}^{K}w_k(x;\theta)\mathcal{N}\left(\boldsymbol{y}|\boldsymbol{\mu}_k(\boldsymbol{x};\theta),\sigma_k^2(\boldsymbol{x};\theta)\right) \tag{1}$$

式中，$\mathcal{N}$代表高斯分布，$w_k(\boldsymbol{x};\theta)$表示第$k$个高斯分布的权重，$\boldsymbol{\mu}_k(\boldsymbol{x};\theta)$是均值，$\sigma_k^2(\boldsymbol{x};\theta)$是第$k$个高斯的方差。这3个GMM参数使用3层神经网络进行估计，以得到正常特征的流形。使用softmax函数对权重估计的输出进行归一化处理，具体为

$$w_k(\boldsymbol{x};\theta)=\frac{\exp\left(a_k^w(\boldsymbol{x};\theta)\right)}{\sum_{i=1}^{K}\exp\left(a_i^w(\boldsymbol{x};\theta)\right)} \tag{2}$$

式中，$a_k^w(\boldsymbol{x})\in\mathbf{R}$是神经网络输出的权重logit分数。$\sigma_k(\boldsymbol{x};\theta)$为标准偏差，利用softmax激活函数进行估计，具体为

$$\sigma_k(\boldsymbol{x};\theta)=\log(1+\exp(\beta\times\boldsymbol{x})) \tag{3}$$

式中，β是缩放因子，本文实验中设为1，均值$\boldsymbol{\mu}_k(\boldsymbol{x};\theta)$没有约束，因此只需使用线性层输出。

2.3 损失函数

基于重建的部分，采用了两种损失的组合：均方误差（mean squared error，MSE）和结构相似指数（structural similarity index，SSIM）。

均方误差（MSE）是一个像素级损失，它假设像素之间相互独立。MSE损失计算为两幅图像的像素级差的平方的平均值，用Frobenius范数正式定义为

$$f_{\mathrm{MSE}}(\boldsymbol{x},\hat{\boldsymbol{x}})=\frac{1}{WH}\|\boldsymbol{x}-\hat{\boldsymbol{x}}\|_{\mathrm{F}}^2 \tag{4}$$

式中，$\boldsymbol{x}$为输入，$\hat{\boldsymbol{x}}$为解码器网络的输出（分别为原始图像和重建图像张量），W和H分别为图像宽度和高度，F表示Frobenius范数。

结构相似指数（SSIM）通过考虑标准MSE方法中丢失的视觉属性来度量图像相似度（Bergmann等，2019b），具体为

$$f_{\mathrm{SSIM}}(\boldsymbol{x},\hat{\boldsymbol{x}})=\frac{\left(2\mu_x\mu_{\hat{x}}+c_1\right)\left(2\sigma_{x\hat{x}}+c_2\right)}{\left(\mu_x^2+\mu_{\hat{x}}^2+c_1\right)\left(\sigma_x^2+\sigma_{\hat{x}}^2+c_2\right)} \tag{5}$$

SSIM的取值范围为[-1，1]，值越接近1，两个图像越相似。其中，μ_x和$\mu_{\hat{x}}$是输入图像和重建图像的平均值，σ_x^2和$\sigma_{\hat{x}}^2$是输入图像和重建图像的方差，$\sigma_{x\hat{x}}$是它们的协方差，c_1和c_2是用于数值稳定性的两个常数。

对于高斯混合密度网络训练，使用对数似然损失（log-likelihood loss）。通过最大似然估计拟合高斯估计网络的参数θ，具体为

$$\theta^*=-\arg\min_\theta\sum_{i=1}^{n}\log p_\theta\left(\boldsymbol{y}_i|\boldsymbol{x}_i\right) \tag{6}$$

将高斯混合模型的概率密度函数式（1）代入式（6），得到目标损失函数。具体为

$$LL=-\arg\min_\theta\sum_{i=1}^{n}\log\sum_{k=1}^{K}w_k(\boldsymbol{x}_i;\theta)\cdot\mathcal{N}\left(\boldsymbol{y}_i|\boldsymbol{\mu}_k(\boldsymbol{x}_i;\theta),\sigma_k^2(\boldsymbol{x}_i;\theta)\right) \tag{7}$$

最终的损失函数是上述3个损失的加权和，具体为

$$L(\boldsymbol{X})=LL+\lambda_1 f_{\mathrm{MSE}}(\boldsymbol{X},\hat{\boldsymbol{X}})+(1-f_{\mathrm{SSIM}}(\boldsymbol{X},\hat{\boldsymbol{X}})) \tag{8}$$

式中，本实验对所有数据集设置$\lambda_1=5$。$\boldsymbol{X}$是输入，$\hat{\boldsymbol{X}}$是重构向量，$1-f_{\mathrm{SSIM}}(\boldsymbol{X},\hat{\boldsymbol{X}})$的目的是为了满足损失函数的目标一致性要求。

2.4 测试阶段异常得分计算

定义$\boldsymbol{x}_{\mathrm{i}}\in\boldsymbol{X}_{\mathrm{test}}$为测试图像，$\boldsymbol{X}_{\mathrm{test}}$为测试集，本文的局部异常得分图$\boldsymbol{LAM}$是将$\boldsymbol{x}_{\mathrm{i}}$输入到视觉Mamba编码器$F_\phi$，得到输出的特征图$\boldsymbol{F}_{h,w}^i$。同时也将$\boldsymbol{x}_i$输入到预训练网络PDN，然后经过适配层$G_\theta$得到输出的特征$\boldsymbol{G}_{h,w}^i$，通过融合$\boldsymbol{F}_{h,w}^i$和$\boldsymbol{G}_{h,w}^i$得到总体特征图$\boldsymbol{T}_{h,w}^i$。具体为

$$\boldsymbol{F}_{h,w}^i=F_\phi\left(\boldsymbol{x}_i\right) \tag{9}$$

$$\boldsymbol{G}_{h,w}^i=G_\theta\left(f_{\mathrm{PDN}}\left(\boldsymbol{x}_i\right)\right) \tag{10}$$

$$\boldsymbol{T}_{h,w}^i=f_{\mathrm{fusion}}\left(\boldsymbol{F}_{h,w}^i,\boldsymbol{G}_{h,w}^i\right) \tag{11}$$

之后，将 $\boldsymbol{T}_{h,w}^{i}$ 输入到高斯密度估计网络 GMDN 输出混合系数 π、均值 μ 和方差 σ，然后计算高斯混合模型（GMM）的负对数似然损失进行异常预测，得到异常得分图 $\boldsymbol{S}_{h,w}^{i}$，具体为

$$\pi,\mu,\sigma = f_{\mathrm{MDN}}\left(\boldsymbol{T}_{h,w}^{i}\right) \tag{12}$$

$$\boldsymbol{S}_{h,w}^{i} = f_{\mathrm{mdn\_loss}}\left(\boldsymbol{T}_{h,w}^{i},\ \mu,\sigma,\pi\right) \tag{13}$$

接着，将其上采样到原始图像大小，得到局部异常得分图 $\boldsymbol{LAM}$，具体为

$$\boldsymbol{LAM}\left(x_i\right) = f_{\mathrm{upsample}}\left(\boldsymbol{S}_{h,w}^{i}\right) \tag{14}$$

输入 $\boldsymbol{x}_i$ 到视觉 Mamba 自编码器 A_β 生成重构向量 $\boldsymbol{A}_{h,w}^{i}$，计算其与原始图像张量 $\boldsymbol{X}_{h,w}^{i}$ 之间的像素级重构误差，从而进一步计算全局异常得分图 $\boldsymbol{GAM}$。

本文还需要将两个异常图归一化到相同的尺度下，通过计算分位数然后用其进行归一化，最后融合得到总异常得分图 $\boldsymbol{TAM}$。总体的异常得分 $Score_{AD}$ 是通过取融合后的异常图 $\boldsymbol{O}_{h,w}^{i}$ 的最大值来进行计算。对于局部异常得分图（$\boldsymbol{LAM}$）的分位数在验证集上的计算为

$$q_{\mathrm{LAMstart}} = f_{\mathrm{Quantile}}(\boldsymbol{LAM}, 0.9) \tag{15}$$

$$q_{\mathrm{LAMend}} = f_{\mathrm{Quantile}}(\boldsymbol{LAM}, 0.995) \tag{16}$$

$$\boldsymbol{LAM}_{\mathrm{norm}} = \frac{\boldsymbol{LAM} - q_{\mathrm{LAMstart}}}{q_{\mathrm{LAMend}} - q_{\mathrm{LAMstart}}} \tag{17}$$

式中，f_{Quantile} 是一个计算分位数的函数，0.9 与 0.995 是第 90 百分位数和第 99.5 百分位数。式（17）是归一化操作。对于全局异常得分图（$\boldsymbol{GAM}$）的分位数在验证集上的计算为

$$q_{\mathrm{GAMstart}} = f_{\mathrm{Quantile}}(\boldsymbol{GAM}, 0.9) \tag{18}$$

$$q_{\mathrm{GAMend}} = f_{\mathrm{Quantile}}(\boldsymbol{GAM}, 0.995) \tag{19}$$

$$\boldsymbol{GAM}_{\mathrm{norm}} = \frac{\boldsymbol{GAM} - q_{\mathrm{GAMstart}}}{q_{\mathrm{GAMend}} - q_{\mathrm{GAMstart}}} \tag{20}$$

总异常得分图 $\boldsymbol{TAM}$ 通过融合 $\boldsymbol{LAM}_{\mathrm{norm}}$ 和 $\boldsymbol{GAM}_{\mathrm{norm}}$ 得到。最后，异常得分取 $\boldsymbol{TAM}$ 的最大值，具体为

$$Score_{AD}\left(x_i\right) = \max_{(h,w)\in W_0\times H_0} \boldsymbol{TAM}_{h,w}^{i} \tag{21}$$

3　实验与分析

3.1　实验数据与实验设置

为了研究模型的有效性，在 3 个工业异常检测数据集上进行了实验。

1）MvTec AD（MvTec anomay detection dataset）数据集（Bergmann 等，2019a）。该数据集是一个广泛应用于异常检测领域的真实世界异常检测数据集，涵盖了 5 354 幅图像，包括灰度图像和 RGB 图像，涉及多种纹理和对象类别。该数据集既包含正常图像，也包含异常图像，展现了 70 种不同类型的现实世界异常产品。

2）BTAD（bean tech anomaly detection）数据集（Mishra 等，2021）。该数据集涵盖了 3 种工业制品的彩色图像，且在测试集中，每一幅异常图像均配备了精确到像素级别的真实掩码。其中，产品 1 的像素分辨率为 1 600 × 1 600 像素，产品 2 的像素分辨率为 600 × 600 像素，产品 3 的像素分辨率为 800 × 600 像素。产品 1、2 和 3 的训练图像数量分别为 400、1 000 和 399 幅。

3）VisA 数据集（Zou 等，2022）。该数据集是一个用于视觉异常检测和分割的数据集。它有 12 种不同的对象，如不同类型的印刷电路板、胶囊、蜡烛、通心粉、腰果和口香糖等。总共包含 10 821 幅图像，其中，9 621 幅为正常样本，1 200 幅为异常样本。异常图像包含各种缺陷，如表面划痕、凹痕、色斑或裂缝，以及结构缺陷如错位或缺失部件。

3.2　实验参数的设置与分析

训练的超参数设置如表 1 所示。为了让数据集更好地适配相关模型，对数据集进行了预处理，所有图像都在传递给模型之前裁剪为 256 × 256 像素，在这个设置下自编码器使用的图像块大小（patch size）设置为 32。除此之外，还对用于重建的 ViM 块的数量进行了调整。在层数设为 4 的情况下，模型的表现效果是最好的，如图 2 所示。

表 1　超参数表

Table 1　Hyperparameter table

超参数	大小
学习率	0.001
批量大小	48
训练轮数	200
权重衰减	0.000 1

3.3　实验结果对比与分析

为了对模型性能进行全面的分析，本文与最新的工业异常检测模型在 MvTec AD、BTAD 和 VisA 3 个数据集上进行实验对比。如表 2 所示，相比于其他 4 个基于 Transformer 的模型，本文方法在 MvTec

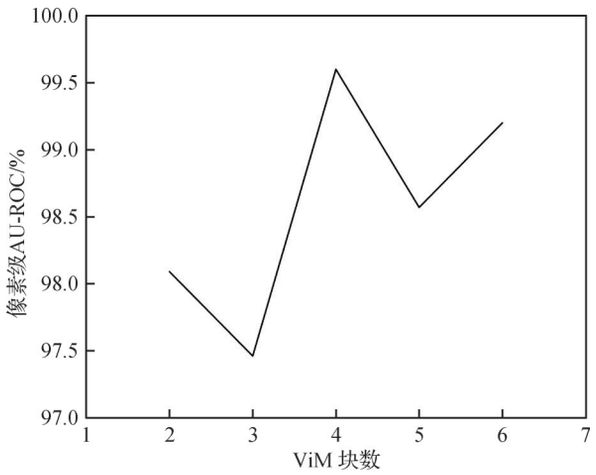

图 2　Bottle类下ViM块数对异常检测精度的影响

Fig. 2　The impact of the number of ViM blocks on anomaly detection accuracy in the bottle class

AD数据集上平均的像素级AU-ROC和图像极AU-ROC结果均是最佳，并且在其他所有类中，5个类实现了最高的像素级AU-ROC和图像级AU-ROC。根据表2可以发现，本文模型在其他模型表现不好的类，如Carpet、Capsule和Metal Nut类上，图像级AU-ROC表现优秀，分别达到了97.3%、96.5%和96.5%。但模型在Cable类和Screw类上的图像级的异常检测精度表现不佳，在3.5小节结合可视化图像对这些原因进行了详细分析。

表3是在VisA数据集上的实验效果，本文模型像素级AU-ROC指标的平均效果达到了最优，超过第2模型0.6%，达到96.6%，并且在Cashew、pcb1、pcb4类上像素级AU-ROC指标超过所有对比模型。然而模型面对Macaroni2类别，图像级别的异常检测精度低于其他模型，对其原因的详细分析在3.5小节说明。

表2　不同方法在MvTec AD数据集的图像级分类精度和像素级分类精度

Table 2　Summary table of different models' results on the MvTec AD dataset: image-level AU-ROC and pixel-level AU-ROC

/%

方法	Carpet	Grid	Leather	Tile	Wood	Bottle	Cable
VT-ADL (Mishra等, 2021)	–/77.3	–/87.1	–/72.8	–/79.6	–/78.1	–/94.9	–/77.6
AnoVIT (Lee和Kang, 2022)	50.0/65.0	52.0/83.0	85.0/89.0	89.0/57.0	95.0/85.0	83.0/86.0	74.0/89.0
HaloAE(Mathian等, 2023)	69.7/89.4	94.9/83.1	97.8/**98.5**	95.7/78.5	**100.0**/91.1	**100.0**/91.9	84.6/87.6
ViT-MCA (Yang和Guo, 2024)	85.0/88.4	89.6/**97.2**	92.0/96.6	92.8/92.8	95.3/**91.4**	94.0/**95.1**	**93.0/92.6**
本文	**97.3/96.9**	**98.5**/91.8	**100.0**/97.7	**99.6/95.3**	99.6/90.8	99.6/93.4	82.5/92.1

方法	Capsule	Hazelnut	Metal Nut	Pill	Screw	Toothbrush	Transistor	Zipper	平均
VT-ADL (Mishra等, 2021)	–/67.2	–/89.7	–/72.6	–/70.5	–/92.8	–/90.1	–/79.6	–/80.8	–/80.7
AnoVIT (Lee和Kang, 2022)	73.0/91.0	88.0/94.0	86.0/88.0	72.0/86.0	**100.0**/92.0	74.0/90.0	83.0/80.0	73.0/76.0	78.0/83.0
HaloAE(Mathian等, 2023)	88.4/**97.8**	99.8/97.8	88.4/85.2	90.1/91.5	89.6/**99.0**	92.9/**97.2**	84.4/87.5	**99.7/96.0**	91.4/91.2
ViT-MCA (Yang和Guo, 2024)	83.7/93.1	**100.0/98.2**	89.5/91.0	86.3/92.6	**100.0**/97.7	93.2/89.4	**86.8/95.0**	89.7/93.2	91.2/93.0
本文	**96.5**/96.8	99.0/95.4	**96.5/94.2**	**93.1/97.0**	70.6/93.7	**97.8**/96.8	84.6/84.6	97.3/91.9	**93.8/93.9**

注：加粗字体表示各列各组最优结果。“–”表示数据为空。“/”前后分别为图像级和像素级的分类精度。

图3是不同模型在NVIDIA RTX A6000 GPU和VisA数据集下延迟与平均AU-ROC的对比图，实验数据参考Batzner等人(2024)的工作。

如图3所示，本文模型在延迟低于15 ms的模型中准确率最高，对比SimpleNet (a simple network for image anomaly detection)(Liu等，2023)和DSR(dual subspace re-projection network)(Zavrtanik等，2022)，本文的准确率有领先且延迟相近，对比AST(asymmetric student teacher networks for industrial anomaly detection)(Rudolph等，2023)、PatchCore(Roth等，2022)等模型，本文模型准确率有待提高但是延迟最低。与GCAD(global context anomaly detection)(Bergmann等，2022)相比，本文模型准确率有明显优势，延迟却相差不大。该实验表明了本文模型具有不错的检测效率，但后期还需提升模型的准确率。

表4的结果显示，本文方法在BTAD数据集上虽

表 3　不同方法在 VisA 数据集的图像级分类精度和像素级分类精度

Table 3　Summary table of different models' results on the VisA dataset: image-level AU-ROC and pixel-level AU-ROC

/%

方法	Candle	Capsules	Cashew	Chewinggum	Fryum	Macaroni1
FastFlow (Yu 等,2021)	92.8/94.9	71.2/75.3	91.0/91.4	91.4/98.6	88.6/97.3	98.3/97.3
DRAEM (Zavrtanik 等,2021)	94.4/97.3	76.3/**99.1**	90.7/88.2	94.2/97.1	**97.4**/92.7	95.0/**99.7**
RD4AD (Deng 和 Li,2022)	92.2/97.9	90.1/89.5	**99.6**/95.8	**99.7**/**99.0**	96.6/94.3	**98.4**/97.7
OmniAL (Zhao,2023)	**96.6**/**98.7**	**99.4**/83.2	96.9/98.4	97.4/98.5	96.9/**98.9**	89.9/99.1
本文	89.2/94.2	81.0/93.7	90.9/**98.5**	95.3/98.2	92.5/97.1	83.3/96.4

方法	Macaroni2	Pcb1	Pcb2	Pcb3	Pcb4	Pipe_fryum	平均
FastFlow (Yu 等,2021)	86.3/89.2	77.4/75.2	61.9/67.3	74.3/94.8	80.9/89.9	72.0/87.3	82.2/88.2
DRAEM (Zavrtanik 等,2021)	96.2/**99.9**	54.8/90.5	77.8/90.5	94.5/98.6	93.4/88.0	**99.4**/90.9	88.7/94.4
RD4AD (Deng 和 Li,2022)	**97.6**/87.7	**97.6**/75.0	91.1/64.8	**95.5**/95.5	96.5/92.8	97.0/92.0	**96.0**/90.1
OmniAL (Zhao,2023)	87.9/98.6	85.1/90.5	**97.1**/**98.9**	94.9/**98.7**	**97.0**/89.3	91.4/**99.1**	94.2/96.0
本文	82.7/92.8	93.7/**98.5**	96.2/95.9	94.3/97.4	90.0/**97.2**	89.5/99.0	89.9/**96.6**

注:加粗字体表示各列各组最优结果。

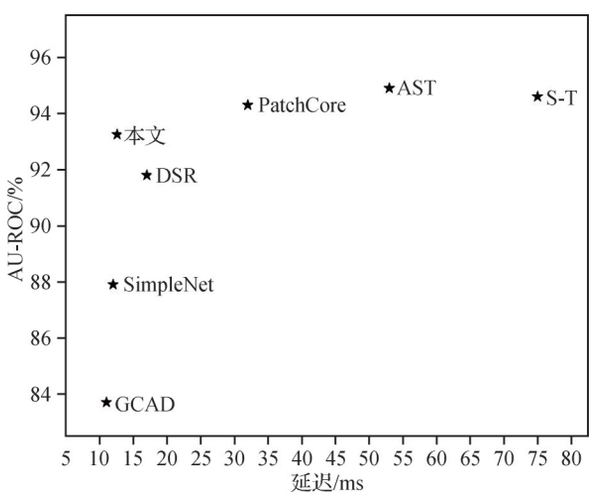

图 3　不同模型在 VisA 数据集的平均 AU-ROC 与延迟的对比图

Fig. 3　Comparison of average AU-ROC and latency for different models on the VisA dataset

然图像级 AU-ROC 的结果超过第 2 模型 0.4%,达到 92.6%,但是像素级别的异常检测效果还有待提高。通过观察 02 类别,由图 5 的真实掩码可知,该类别仅将划痕作为异常,而纹路的变化不设为异常,但是通常人的直觉也会将其判定为异常,这导致本文重建模块生成的热力图异常区域偏离真实掩码对应区域,大幅降低了模型异常检测以及定位的性能,导致本文方法不能完全发挥。

3.4　消融实验

为了研究视觉状态空间在异常检测任务中的有效性,本文在 BTAD 数据集上进行消融实验研究。在重建模块中,对基于 CNN 的自编码器和基于 ViM 的自编码器进行对比。整体模型的结果如表 5 所示,模型以 ViM 作为编码器进行图像重建,在 BTAD 数据集的图像级 AU-ROC、像素级 AU-ROC 平均值均比传统的基于 CNN 的自编码器精度更高。图像级 AU-ROC 提升 1.6%,像素级 AU-ROC 提升 3.7%。

此外,为了研究各模块在工业图像异常检测中的有效性,做了进一步的消融实验,结果如表 6 所示。在 BTAD 数据集下,视觉 Mamba 自编码重建网络(visual Mamba autoencoder reconstruction network, VMARN)的使用能够提高模型的整体像素级以及图像级的异常检测精度。该模块结合块特征分布建模模块(patch feature distribution modeling module, PFDMM),比仅使用 PFDMM 模块,图像级 AU-ROC 和像素级 AU-ROC 分别提升 1.8% 和 1.2%,表明结合重建的方法在本文框架下能有效提高图像异常检测的效果。这是因为基于视觉 Mamba 重建的方法能够捕捉长距离关联特征,具有全局建模能力,且本文认为结合重建方法是本文模型检测逻辑异常的关键,观察图 1(b)在一幅具有逻辑数量异常的图像上,VMARN 模块的全局异常图 GAM 表现的异常范围远大于 PFDMM 模块的局部异常图 LAM,且异常区域更接近真实的异常范围,异常分割效果也更好,这证明了模型在检测逻辑异常上的潜力。而块特征

表4　不同方法在BTAD数据集的图像级分类精度和像素级分类精度

Table 4　Summary table of different models' results on the BTAD dataset: image-level AU-ROC and pixel-level AU-ROC

/%

方法	01类别	02类别	03类别	平均
VT-ADL	-/92	-/89	-/86	-/90.0
P-SVDD (Yi和Yoon,2021)	95.7/91.6	72.1/93.6	82.1/91.0	83.3/92.1
PatchCore (Roth等,2022)	90.9/**95.5**	79.3/94.7	**99.8**/**99.3**	90.0/96.5
InReaCh (McIntosh和Albu,2023)	-/-	-/-	-/-	90.3/**96.9**
ViT-MCA	**98.7**/93.2	**85.0**/89.4	93.0/93.1	92.2/91.9
本文	97.7/86.3	82.9/**95.2**	97.1/95.7	**92.6**/92.4

注：加粗字体表示各列各组最优结果，"-"表示数据为空。

表5　重建模块中不同类型编码器在BTAD数据集上的图像级平均分类精度和像素级平均分类精度

Table 5　Image-level and pixel-level average classification accuracy of different encoder types in the reconstruction module on the BTAD dataset

/%

序号	CNN	vision Mamba	平均
1	√	-	91.0/88.7
2	-	√	**92.6/92.4**

注：加粗字体表示最优结果。"-"为不使用，"√"为使用。

分布建模方法只考虑局部特征，减少了其他背景的干扰，可以检测大多数重建方法检测不到的异常，通过将它们的异常图归一化融合，有助于提升整体的异常检测效果。

该实验尝试融合预训练网络PDN和视觉Mamba编码器的低维特征，获得融合特征图，然后输入到高斯密度估计网络进行训练。为了验证融合实验的有效性，进行了消融实验，结果如表7所示。模型融合不同编码器特征，比仅使用PDN特征时模型

表6　不同模块在BTAD数据集上的图像级平均分类精度和像素级平均分类精度

Table 6　Image-level AU-ROC and pixel-level AU-ROC of different modules on the BTAD dataset

/%

序号	VMARN	PFDMM	平均
1	√	-	71.0/85.0
2	-	√	90.8/91.2
3	√	√	**92.6/92.4**

注：加粗字体表示最优结果。"-"为不使用，"√"为使用。

表7　不同编码器特征在BTAD数据集上的图像级平均分类精度和像素级平均分类精度

Table 7　Image-level AU-ROC and pixel-level AU-ROC of different encoder features on the BTAD dataset

/%

序号	PDN	vision Mamba	平均
1	√	-	92.1/91.4
2	-	√	79.0/83.4
3	√	√	**92.6/92.4**

注：加粗字体表示最优结果。"-"为不使用，"√"为使用。

在BTAD数据集上像素级AU-ROC和图像级AU-ROC分别提升1%和0.5%，证明了视觉Mamba编码器特征的有效性。

3.5　可视化

图4是不同模型在MvTec AD数据集以及VisA数据集上的异常区域可视化对比结果。可以看出，在牙刷以及塑料胶囊上，本文模型比DRAEM(Zavrtanik等，2021)模型异常定位的范围更加精确，但是在塑料类别上与异常塑料位置有轻微的偏离，这是对低维特征图进行异常评估然后上采样导致定位偏离。原因是使用的特征图只有64个块，而这些块的异常情况并不能完全代表256×256像素的具体位置的异常情况。用更多含有图像块描述向量的特征图可以缓解这个问题。

为了进一步体现ViM自编码器重建效果的优越性，分别对其以及基于CNN的自编码器的重建结果进行可视化，如图5所示。在VisA数据集上的腰果类别上，用了一个数量变化的逻辑异常图像，可以发现使用ViM编码器，能够更好地重建出正常的图像，

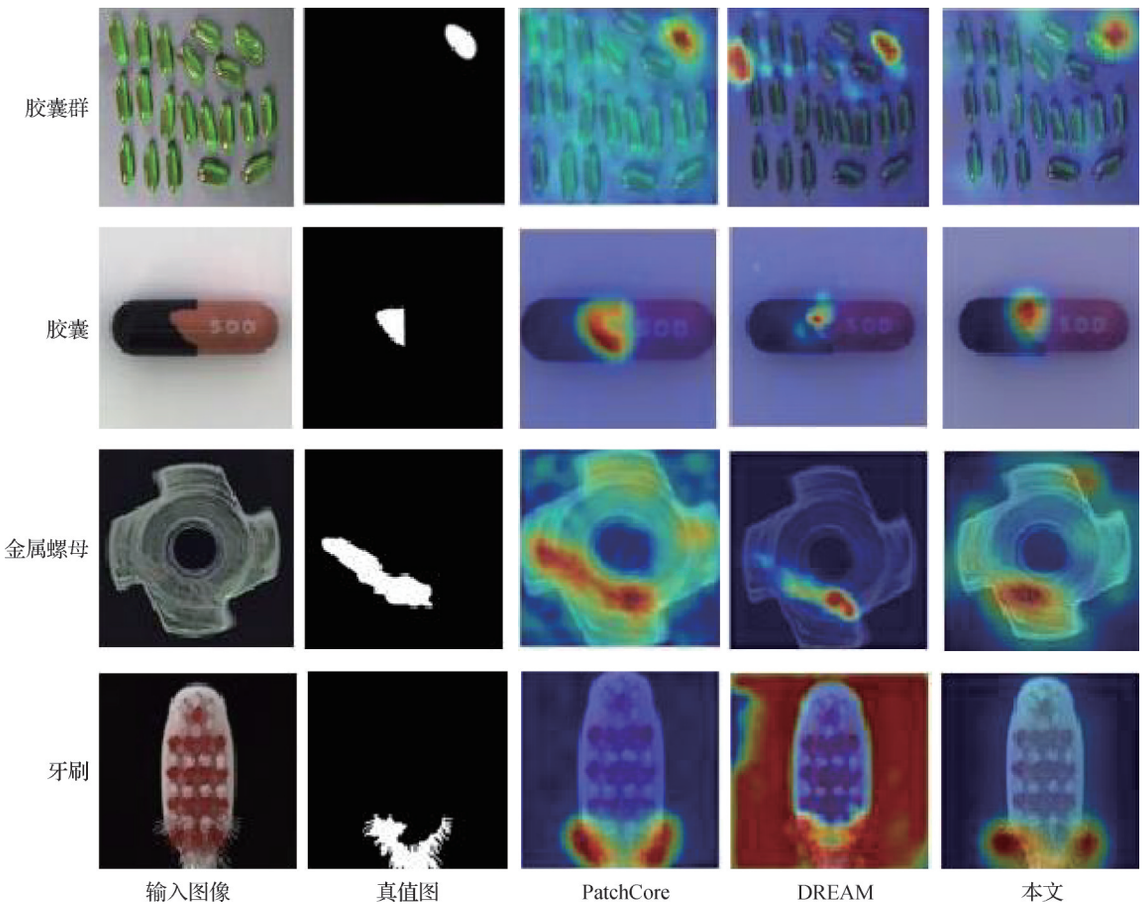

图 4　MvTec AD 和 VisA 数据集下的定性实验结果

Fig. 4　Qualitative results on the MvTec AD and VisA datasets

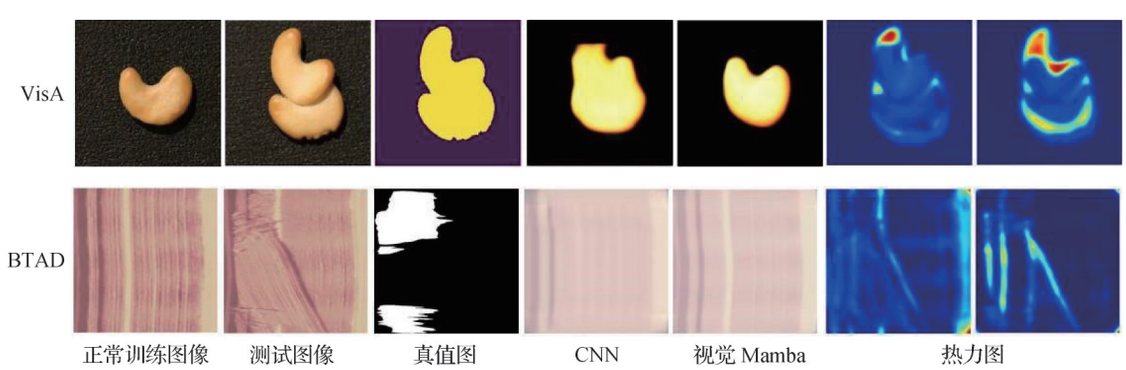

图 5　BTAD 和 VisA 数据集下不同编码器的重建实验结果

Fig. 5　Experimental results of reconstruction with different encoders on the BTAD and VisA datasets

而 CNN 编码器重建出的图像对图像轮廓、边缘信息学习的不够透彻。观察它们的重构误差热力图也可以发现 ViM 自编码器的异常区域以及异常强度都比传统的基于 CNN 的自编码器更精确。图 5 第 2 行是在 BTAD 数据集 02 类别上进行重建实验的可视化效果，可以发现模型对木板条纹颜色、形状、数量和分布结构的学习程度，ViM 的重建结果均强于 CNN 重建的结果，这说明用 ViM 的重建方法重建出的正常图像能够比 CNN 的方法捕捉到更多的细节信息。

本文融合了两个编码器的特征图，并且用融合特征图训练模型，提升了模型的性能，两个编码器的特征图可视化如图 6 所示。观察两者的特征图可知，对于一幅图像，它们提取了图像不同的特征。根据特征图的内容，本文认为视觉 Mamba 编码器捕获了 PDN 网络获取不到的图像块特征，猜测可能是位置信息以及视觉 Mamba 编码器捕获的某些长距离

关联特征，视觉Mamba特征图的内容更能代表图像的形状，以及纹理、位置信息。这些特征可能有利于异常的检测，所以融合特征图提升了模型进行异常检测的性能。

图7是本文模型单独的一些异常定位可视化图像。对于大部分图像，模型可以准确地定位到异常区域，但是对于少部分类别，模型会将正常区域也判断为异常，观察螺丝（Screw）类别可知模型在螺丝头部的正常区域进行了误判，该现象可能是导致本文模型在MvTec AD数据集的Screw类别上异常检测准确率低的原因。通过观察发现在VisA数据集的蜡烛（Candle）类别上也存在误判的情况。这种情况还发生在通心粉（Macaroni）的类别上，该类别上模型认为影子区域属于轻微异常，这说明了模型倾向于认为影子形状变化属于异常，但它们实际上是正常的。

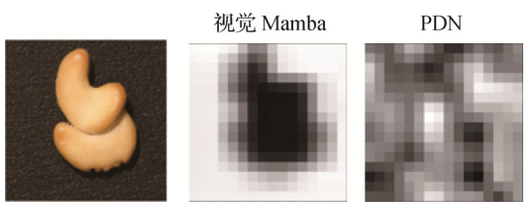

图6　不同编码器输出的低维特征图

Fig. 6　Low-dimensional feature maps produced by different encoders

本文模型虽然可以快速定位到异常的区域，但是精准性尚有待提高，观察图7不难发现，模型对那些形状奇怪的异常难以完全定位，这是可能因为模型最后的输出只含64个图像块，并不能完全描述整幅图像的情况，要更精确地分割异常区域需要模型利用更多的图像块描述向量，然而这又会加大模型计算复杂度。

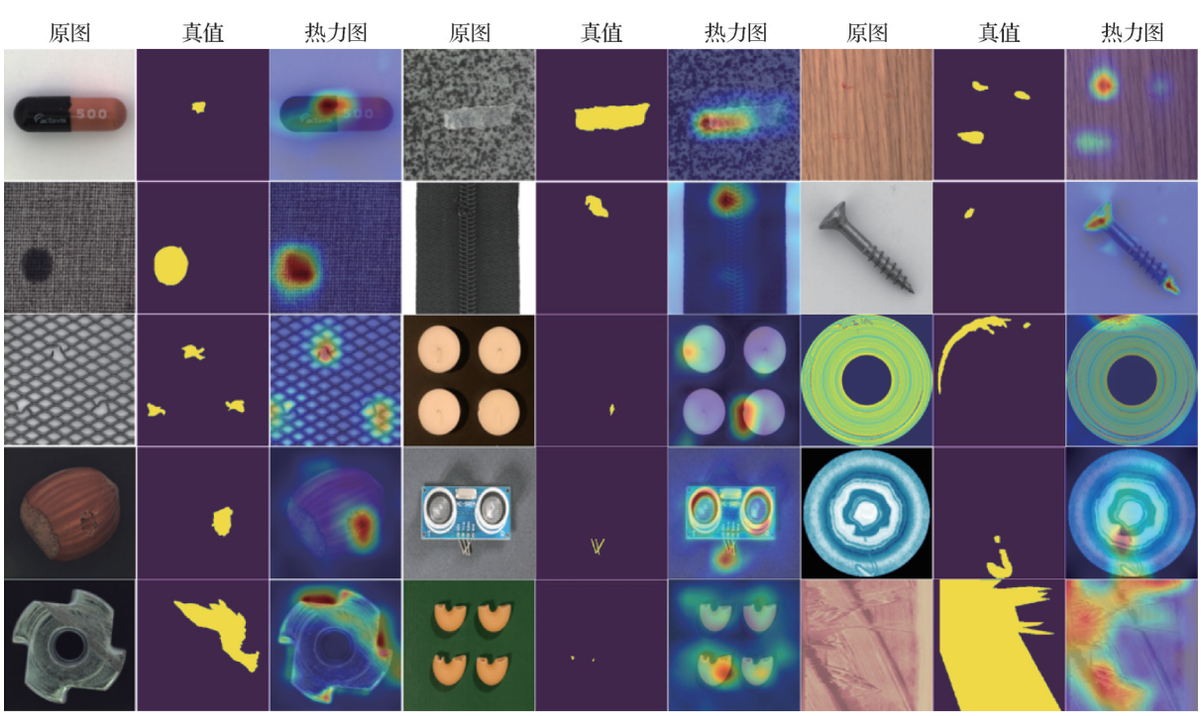

图7　本文模型在MvTec AD、BTAD和VisA数据集上的一些随机定性结果

Fig. 7　The model in this paper presents some random qualitative results on the MvTec AD, BTAD, and VisA datasets

4　结　论

本文提出一个结合视觉Mamba和块特征分布的无监督工业异常检测框架，该框架融合了图像重建和局部特征嵌入的方法，在异常检测性能和计算效率上均表现出色，为结构性和逻辑性异常的检测设定了新的标准。与其他异常检测模型相比，本文的方法取得了最好的效果。目前的模型还需要大量正常样本进行训练，而在实际工业生产中，针对特定新产品，收集足够多的正常样本往往比较困难，并且现实中采集的正常图像样本通常存在噪声，这往往

会影响模型的准确性。针对上述问题，零样本、小样本以及弱监督的工业异常检测逐渐引起关注，成为当前的研究主流。下一步研究中，将扩展现有模型到小样本甚至零样本工业异常检测中。

参考文献（References）

Bae J，Lee J H and Kim S. 2023. PNI：industrial anomaly detection using position and neighborhood information//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision（ICCV）. Paris，France：IEEE：6350-6360［DOI：10.1109/ICCV51070.2023.00586］

Batzner K，Heckler L and König R. 2024. EfficientAD：accurate visual anomaly detection at millisecond-level latencies//Proceedings of 2024 IEEE/CVF Winter Conference on Applications of Computer Vision（WACV）. Los Alamitos，USA：IEEE Computer Society：127-137［DOI：10.1109/WACV57701.2024.00020］

Bergmann P，Batzner K，Fauser M，Sattlegger D and Steger C. 2022. Beyond dents and scratches：logical constraints in unsupervised anomaly detection and localization. International Journal of Computer Vision，130（4）：947-969［DOI：10.1007/s11263-022-01578-9］

Bergmann P，Fauser M，Sattlegger D and Steger C. 2019a. MVTec AD — a comprehensive real-world dataset for unsupervised anomaly detection//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. Los Alamitos，USA：IEEE Computer Society：9584-9592［DOI：10.1109/CVPR.2019.00982］

Bergmann P，Fauser M，Sattlegger D and Steger C. 2020. Uninformed students：student-teacher anomaly detection with discriminative latent embeddings//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. Los Alamitos，USA：IEEE Computer Society：4182-4191［DOI：10.1109/CVPR42600.2020.00424］

Bergmann P，Löwe S，Fauser M，Sattlegger D and Steger C. 2019b. Improving unsupervised defect segmentation by applying structural similarity to autoencoders//Proceedings of the 14th International Joint Conference on Computer Vision，Imaging and Computer Graphics Theory and Applications（VISIGRAPP 2019）—Volume 5：VISAPP. Prague，Czech Republic：SciTePress：372-380［DOI：10.5220/0007364503720380］

Bishop C M. 1994. Mixture Density Networks. Technical Report No. NCRG/94/004. Aston University

Cohen N and Hoshen Y. 2020. Sub-image anomaly detection with deep pyramid correspondences［EB/OL］.［2024-10-21］. https://arxiv.org/pdf/2005.02357.pdf

Defard T，Setkov A，Loesch A and Audigier R. 2021. PaDiM：a patch distribution modeling framework for anomaly detection and localization//Proceedings of 2021 International Conference on Pattern Recognition（ICPR）. Milano，Italy：Springer：475-489［DOI：10.1007/978-3-030-68799-1_35］

Deng H Q and Li X Y. 2022. Anomaly detection via reverse distillation from one-class embedding//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. New Orleans，USA：IEEE：9727-9736［DOI：10.1109/CVPR52688.2022.00951］

Fang Z，Wang X Y，Li H C，Liu J J，Hu Q G and Xiao J M. 2023. FastRecon：few-shot industrial anomaly detection via fast feature reconstruction//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision（ICCV）. Paris，France：IEEE：17435-17444［DOI：10.1109/ICCV51070.2023.01603］

Gu A and Dao T. 2023. Mamba：linear-time sequence modeling with selective state spaces［EB/OL］.［2024-10-21］. https://arxiv.org/pdf/2312.00752.pdf

Gu A，Goel K and Ré C. 2022. Efficiently modeling long sequences with structured state spaces［EB/OL］.［2024-10-21］. https://arxiv.org/pdf /2111.00396.pdf

Gu Z H，Liu L，Chen X，Yi R，Zhang J N，Wang Y B，Wang C G，Shu A N，Jiang G N and Ma L Z. 2023. Remembering normality：memory-guided knowledge distillation for unsupervised anomaly detection//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision（ICCV）. Paris，France：IEEE：16355-16363［DOI：10.1109/ICCV51070.2023.01503］

Gudovskiy D，Ishizaka S and Kozuka K. 2022. CFLOW-AD：real-time unsupervised anomaly detection with localization via conditional normalizing flows//Proceedings of 2022 IEEE/CVF Winter Conference on Applications of Computer Vision（WACV）. Waikoloa，USA：IEEE：1819-1828［DOI：10.1109/WACV51458.2022.00188］

Huang C Q，Guan H Y，Jiang A F，Zhang Y，Spratling M and Wang Y F. 2022. Registration based few-shot anomaly detection//Proceedings of 17th European Conference on Computer Vision（ECCV 2022）. Tel Aviv，Israel：Springer：303-319［DOI：10.1007/978-3-031-20053-3_18］

Jeong J，Zou Y，Kim T，Zhang D Q，Ravichandran A and Dabeer O. 2023. WinCLIP：zero-/few-shot anomaly classification and segmentation//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. Los Alamitos，USA：IEEE Computer Society：19606-19616［DOI：10.1109/CVPR52729.2023.01878］

Kim D，Park C，Cho S and Lee S. 2023. FAPM：fast adaptive patch memory for real-time industrial anomaly detection//Proceedings of 2023 IEEE International Conference on Acoustics，Speech and Signal Processing（ICASSP）. Rhodes Island，Greece：IEEE：1-5［DOI：10.1109/ICASSP49357.2023.10096400］

Lee Y and Kang P. 2022. AnoViT：unsupervised anomaly detection and localization with vision transformer-based encoder-decoder. IEEE

Access, 10: 46717-46724 [DOI: 10.1109/ACCESS. 2022. 3171559]

Liu Z K, Zhou Y M, Xu Y S and Wang Z L. 2023. SimpleNet: a simple network for image anomaly detection and localization//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE Computer Society: 20402-20411 [DOI: 10.1109/CVPR52729.2023.01954]

Mathian E, Liu H D, Fernandez-Cuesta L, Samaras D, Foll M and Chen L M. 2023. HaloAE: a local transformer auto-encoder for anomaly detection and localization based on haloNet//Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023). Lisbon, Portugal: SciTePress: 325-337 [DOI: 10.5220/0011865900003417]

McIntosh D and Albu A B. 2023. Inter-realization channels: unsupervised anomaly detection beyond one-class classification//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 6262-6272 [DOI: 10.1109/ICCV51070.2023.00578]

Mishra P, Verk R, Fornasier D, Piciarelli C and Foresti G L. 2021. VT-ADL: a vision transformer network for image anomaly detection and localization//Proceedings of the 30th IEEE International Symposium on Industrial Electronics (ISIE). Kyoto, Japan: IEEE: 1-6 [DOI: 10.1109/isie45552.2021.9576231]

Pang G S, Shen C H, Cao L B and Van Den Hengel A. 2022. Deep learning for anomaly detection: a review. ACM Computing Surveys, 54(2): #38 [DOI: 10.1145/3439950]

Pirnay J and Chai K. 2022. Inpainting transformer for anomaly detection//Proceedings of the 21st International Conference on Image Analysis and Processing. Lecce, Italy: Springer-Verlag: 394-406 [DOI: 10.1007/978-3-031-06430-2_33]

Rippel O, Mertens P and Merhof D. 2021. Modeling the distribution of normal data in pre-trained deep features for anomaly detection//Proceedings of the 25th International Conference on Pattern Recognition (ICPR). Milan, Italy: IEEE Computer Society: 6726-6733 [DOI: 10.1109/ICPR48806.2021.9412109]

Roth K, Pemula L, Zepeda J, Schölkopf B, Brox T and Gehler P. 2022. Towards total recall in industrial anomaly detection//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE Computer Society: 14298-14308 [DOI: 10.1109/CVPR52688.2022.01392]

Rudolph M, Wandt B and Rosenhahn B. 2021. Same same but differNet: semi-supervised defect detection with normalizing flows//Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa, USA: IEEE Computer Society: 1906-1915 [DOI: 10.1109/WACV48630.2021.00195]

Rudolph M, Wehrbein T, Rosenhahn B and Wandt B. 2023. Asymmetric student-teacher networks for industrial anomaly detection//Proceedings of 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, USA: IEEE Computer Society: 2591-2601 [DOI: 10.1109/WACV56688.2023.00262]

Smith J T H, Warrington A and Linderman S W. 2023. Simplified state space layers for sequence modeling [EB/OL]. [2023-03-03]. https://arxiv.org/pdf/2208.04933.pdf

Sugawara S and Imamura R. 2024. PUAD: frustratingly simple method for robust anomaly detection//Proceedings of 2024 IEEE International Conference on Image Processing (ICIP). Abu Dhabi, United Arab Emirates: IEEE: 842-848 [DOI: 10.1109/ICIP51287.2024.10647438]

Wang S Q, Cheng C, Shi M and Zhu D M. 2024. Defect detection method for industrial product surfaces with similar features by combining frequency and ViT. Journal of Image and Graphics, 29(10): 3074-3089 (王素琴, 程成, 石敏, 朱登明. 2024. 结合频率和ViT的工业产品表面相似特征缺陷检测方法. 中国图象图形学报, 29(10): 3074-3089) [DOI: 10.11834/jig.230532]

Yang Q Y and Guo R Z. 2024. An unsupervised method for industrial image anomaly detection with vision transformer-based autoencoder. Sensors, 24(8): #2440 [DOI: 10.3390/s24082440]

Yi J H and Yoon S. 2021. Patch SVDD: patch-level SVDD for anomaly detection and segmentation//Proceedings of the 15th Asian Conference on Computer Vision (ACCV) 2020. Kyoto, Japan: Springer: 375-390 [DOI: 10.1007/978-3-030-69544-6_23]

Yu J W, Zheng Y, Wang X, Li W, Wu Y S, Zhao R and Wu L W. 2021. FastFlow: unsupervised anomaly detection and localization via 2D normalizing flows [EB/OL]. [2024-10-21]. https://arxiv.org/pdf/2111.07677.pdf

Zavrtanik V, Kristan M and Skočaj D. 2021. DRÆM — a discriminatively trained reconstruction embedding for surface anomaly detection//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 8310-8319 [DOI: 10.1109/ICCV48922.2021.00822]

Zavrtanik V, Kristan M and Skočaj D. 2022. DSR — a dual subspace re-projection network for surface anomaly detection//Proceedings of the 17th European Conference on Computer Vision (ECCV). Tel Aviv, Israel: Springer: 539-554 [DOI: 10.1007/978-3-031-19821-2_31]

Zhang X, Li S Y, Li X, Huang P, Shan J L and Chen T. 2023. DeSTSeg: segmentation guided denoising student-teacher for anomaly detection//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 3914-3923 [DOI: 10.1109/CVPR52729.2023.00381]

Zhao Y. 2023. OmniAL: a unified CNN framework for unsupervised anomaly localization//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 3924-3933 [DOI: 10.1109/CVPR52729.2023.00382]

Zhou K, Xiao Y T, Yang J L, Cheng J, Liu W, Luo W X, Gu Z W, Liu J and Gao S H. 2020. Encoding structure-texture relation with

p-net for anomaly detection in retinal images//Proceedings of the 16th European Conference on Computer Vision (ECCV). Glasgow, UK: Springer: 360-377 [DOI: 10.1007/978-3-030-58565-5_22]

Zhu L H, Liao B C, Zhang Q, Wang X L, Liu W Y and Wang X G. 2025. Vision Mamba: efficient visual representation learning with bidirectional state space model//Proceedings of the 41st International Conference on Machine Learning (ICML). Vienna, Austria: JMLR.org: #2584 [DOI: 10.5555/3692070.3694654]

Zou Y, Jeong J, Pemula L, Zhang D Q and Dabeer O. 2022. SPot-the-difference self-supervised pre-training for anomaly detection and segmentation//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 392-408 [DOI: 10.1007/978-3-031-20056-4_23]

作者简介

刘建明，男，副教授，硕士生导师，主要研究方向为工业图像异常检测和医学图像研究。E-mail：liujianming@jxnu.edu.cn

庄维宽，男，硕士研究生，主要研究方向为工业图像异常检测。E-mail：201820210127@ecut.edu.cn