中图法分类号:TP18;TP391.41　　**文献标识码**:A　　**文章编号**:1006-8961(2026)10-3763-16

论文引用格式:Zhao Y C, Zhu N, Song X G, Hei X H and Shi Z H. 2026. Cross-domain self-supervised representation learning for medical image anomaly detection. Journal of Image and Graphics, 31(10):3763-3778(赵映程, 朱凝, 宋霄罡, 黑新宏, 石争浩. 2026. 面向医学图像异常检测的跨域自监督表征学习. 中国图象图形学报, 31(10):3763-3778)[DOI:10.11834/jig.250599]

面向医学图像异常检测的跨域自监督表征学习

赵映程[1],朱凝[1],宋霄罡[1,2,3],黑新宏[1,2,3],石争浩[1,2*]

1. 西安理工大学计算机科学与工程学院,西安　710048; 2. 陕西省网络计算与安全技术重点实验室,西安　710048;
3. 人机共融智能机器人陕西省高校工程研究中心,西安　710048

摘　要:**目的**　医学图像异常检测旨在以无监督方式识别临床影像中的病变区域,但现有方法在迁移自然图像预训练模型至医学域时存在语义差异问题,导致模型对解剖结构差异敏感且对早期病变识别能力不足。现有方法若对预训练编码器进行微调,易因模式崩溃导致失败;而若保持冻结,则难以充分适配医学域。为此,提出一种跨域自监督表征学习(cross-domain self-supervised representation learning,CSRL)框架,以提升医学图像异常检测的准确性与鲁棒性。**方法**　CSRL框架包含两个阶段:第1阶段通过域适应对比学习(domain-adaptive contrastive learning,DACL)网络,采用在线—目标网络双路径框架实现预训练编码器从自然图像域到医学图像域的稳健迁移;第2阶段构建特征重建网络,引入协同注意力增强模块(synergistic attention enhancement,SAE)以增强病变特征表示,并结合多尺度异常融合模块(multi-scale anomaly fusion,MAF)实现跨层级异常响应的动态融合。**结果**　实验在APTOS(Asia Pacific Tele-Ophthalmology Society)、ISIC(International Skin Imaging Collaboration)和BR35H 3个公开医学数据集上与10种代表性方法进行了对比。实验表明,本文方法在图像级异常检测任务中取得最优性能。在BR35H数据集上,受试者工作特征曲线下面积(area under curve,AUC)达到99.90%;在ISIC数据集上,AUC为91.79%;在APTOS数据集上,AUC为97.71%。相较于当前先进方法,本文方法的最佳AUC值在APTOS、ISIC和BR35H数据集上分别提升了0.35%、3.52%和0.02%。消融实验验证了各模块的有效性,可视化结果进一步表明本文方法在异常定位方面具有更优的空间一致性与临床可解释性。**结论**　CSRL框架通过跨域自监督表征学习、协同注意力增强与多尺度融合机制,有效缓解了预训练模型在医学图像中的语义差异问题,提升了异常检测的判别能力与定位精度,为医学图像无监督异常检测提供了可靠的解决方案。

关键词:医学图像分析;异常检测;自监督学习;跨域迁移;特征重建;注意力机制;多尺度融合

Cross-domain self-supervised representation learning for medical image anomaly detection

Zhao Yingcheng[1], Zhu Ning[1], Song Xiaogang[1,2,3], Hei Xinhong[1,2,3], Shi Zhenghao[1,2*]

1. School of Computer Science and Engineering, Xi'an University of Technology, Xi'an　710048, China;
2. Shaanxi Key Laboratory for Network Computing and Security Technology, Xi'an　710048, China; 3. Human Machine Integration Intelligent Robot Shaanxi Provincial University Engineering Research Center, Xi'an　710048, China

Abstract: **Objective**　Medical image anomaly detection represents a fundamental challenge in clinical diagnostics, aiming

收稿日期:2025-12-05;**修回日期**:2026-02-12;**预印本日期**:2026-02-19

*通信作者:石争浩　362743337@qq.com

基金项目:国家自然科学基金项目(U2568225);陕西省自然科学基础研究计划基金资助(2025JC-YBMS-755, 2025JC-YBQN-934)

Supported by: National Natural Science Foundation of China(U2568225); Natural Science Basic Research Program of Shaanxi Province(2025JC-YBMS-755, 2025JC-YBQN-934)

to automatically identify and localize potential lesion regions without manual annotations. Current reconstruction-based anomaly detection methods predominantly rely on features extracted from models pretrained on natural image datasets such as ImageNet. While these approaches have demonstrated promising results, they face considerable limitations due to the substantial domain shifts between natural and medical images. These domain discrepancies lead to semantic misalignments during feature reconstruction, resulting in two primary issues: oversensitivity to normal anatomical variations and insufficient discriminative capability for detecting subtle or early-stage lesions. Such limitations substantially hinder the clinical applicability of existing methods, particularly in screening scenarios where early detection is crucial. To address these challenges and to overcome the dilemma where fine-tuning pretrained encoders often leads to pattern collapse while keeping them frozen limits domain adaptation, this study proposes a cross-domain self-supervised representation learning (CSRL) framework, specifically designed to enhance both the accuracy and robustness of unsupervised medical image anomaly detection through effective domain adaptation and multi-scale feature optimization. **Method** The proposed CSRL framework operates through two carefully designed sequential learning stages, each targeting distinct aspects of the anomaly detection pipeline while maintaining architectural coherence. In the first stage, a domain-adaptive contrastive learning (DACL) network is constructed on the basis of the bootstrap your own latent architecture. This module employs an innovative dual-path online-target network framework to facilitate robust cross-domain transfer of pretrained encoders. The online and target networks process different strongly augmented views of the same medical image through random cropping, color jittering, and Gaussian blurring operations. The target network parameters are updated via an exponential moving average of the online network, with a momentum coefficient of 0.99 to ensure training stability. A symmetric negative cosine similarity loss function is employed to maximize feature consistency between the two views, effectively driving the encoder to learn augmentation-invariant and domain-adaptive representations. This sophisticated learning mechanism enables the model to capture essential medical image characteristics while discarding domain-specific artifacts from natural images, thereby establishing a solid foundation for subsequent anomaly detection tasks. In the second stage, we construct an end-to-end feature reconstruction network specifically optimized for anomaly scoring and precise localization. The encoder component is initialized with the adapted weights from the DACL module and remains completely frozen during this training phase to preserve the learned domain-invariant representations. The network extracts multi-scale deep feature maps at three different resolutions, which are subsequently processed by a specially designed bottleneck layer comprising three residual blocks with skip connections for effective feature fusion. A symmetric decoder architecture with transposed convolutional layers then reconstructs the features at corresponding scales, maintaining spatial integrity throughout the reconstruction process. To enhance the model's focus on potential lesion regions further, we introduce a synergistic attention enhancement (SAE) module that innovatively integrates three complementary attention branches: coordinate attention for precise spatial localization, channel attention for feature recalibration, and spatial attention for regional emphasis. These attention mechanisms are dynamically aggregated through learnable weighting coefficients to refine feature responses in an adaptive manner. Additionally, a multi-scale anomaly fusion (MAF) module computes cosine distance-based anomaly maps at different scales and employs learnable weights combined with spatial attention to generate the final optimized anomaly map. The entire framework is optimized using a joint loss function that combines reconstruction error, fusion consistency, and sparsity constraints, ensuring comprehensive training supervision across multiple objectives. **Result** Comprehensive experiments were conducted on three publicly available medical image datasets covering diverse modalities, i.e., Asia Pacific Tele-Ophthalmology Society (APTOS), International Skin Imaging Collaboration (ISIC), and BR35H, to evaluate the proposed method rigorously under various clinical scenarios. The CSRL framework was compared with 10 state-of-the-art methods, including autoencoder-based models (AE, VAE), generative approaches (GANomaly, f-AnoGAN), memory-based models (PaDiM), and recent pretraining-based methods (STFPM, MKD, EDC, EA2D). All experiments were performed using consistent evaluation protocols and five independent runs to ensure statistical significance. Quantitative results demonstrate that our method achieves optimal performance in image-level anomaly detection tasks. On the BR35H dataset, CSRL achieved an AUC of 99.90%; on the ISIC dataset, an AUC of 91.79%; and on the APTOS dataset, an AUC of 97.71%. Compared with the current state-of-the-art method, the best AUC of our method is improved by 0.35%, 3.52%, and 0.02% on the APTOS, ISIC, and BR35H datasets, respectively. Extended ablation studies provided deeper insights

into the individual contributions of each module. The DACL module alone improved AUC from 88.23% to 91.72% on the ISIC dataset, highlighting its crucial role in domain adaptation. The SAE module contributed considerably to precise localization, while the MAF module effectively integrated multi-scale information for comprehensive anomaly detection. When combined, the full model achieved 94.74% AUC in its best run, demonstrating the synergistic effect of all components. Qualitative visualization of anomaly localization maps revealed that CSRL produces highly concentrated and clinically interpretable heatmaps that align well with actual lesion contours, outperforming comparison methods in spatial consistency and anatomical plausibility. The generated activation maps show precise correspondence with pathological regions while maintaining minimal responses in healthy tissues, substantially improving clinical interpretability for medical practitioners.

Conclusion This study presents a novel CSRL framework that effectively addresses the fundamental challenge of semantic discrepancy in transferring pretrained models to medical image anomaly detection. By systematically integrating domain-adaptive contrastive learning, synergistic attention enhancement, and intelligent multi-scale anomaly fusion, the framework achieves remarkable improvements in both discriminative capability and spatial precision of anomaly detection. Extensive experimental validation across multiple medical imaging modalities confirms the effectiveness and generalizability of the proposed approach, establishing new state-of-the-art performance while maintaining clinical interpretability. The framework provides a reliable and scalable solution for unsupervised anomaly detection in medical images, with immediate potential applications in early diagnosis and clinical screening programs. Future work will focus on three key directions: developing more efficient cross-domain adaptation strategies for few-shot and limited-data scenarios, optimizing model complexity and inference speed for practical clinical deployment, and enhancing sensitivity to subtle anomalies through advanced feature alignment mechanisms. These research directions aim to strengthen both the theoretical foundations and practical applicability of self-supervised learning in medical imaging, ultimately contributing to improved healthcare outcomes through more accessible and accurate diagnostic tools.

Key words: medical image analysis; anomaly detection; self-supervised learning; cross-domain transfer; feature reconstruction; attention mechanism; multi-scale fusion

0 引　言

无监督异常检测技术已广泛应用于工业生产（刘建明和庄维宽，2025）、视频监控（汪洋 等，2025）等多个领域，该类任务通常被视为一类离群值检测问题，首先利用大量正常样本学习正常图像的模式，并在推断阶段通过评估输入样本与正常分布之间的偏离度进行异常判定。在医学图像领域，异常检测技术旨在识别与定位临床影像中的健康隐患，由于其能够避免监督学习对高质量标注数据的依赖，在提升疾病筛查效率及罕见病诊断方面具有重要价值。

为量化样本离群程度，现有的医学异常检测研究形成了多类代表性范式。早期方法大多直接在数据空间或浅层特征空间中，依据样本点的距离（Angiulli 等，2006）、局部密度（Breunig 等，2000）或聚类归属（Yang 等，2009）等统计方法量化其离群程度。随着深度学习技术的快速发展，基于深度神经网络的检测方法已成为当前研究的主流，包括基于异常合成的方法（Tan 等，2020；Müller 等，2023），通过向正常样本添加人工异常使模型学习区分正常与合成异常模式，继而泛化到真实异常场景；基于记忆库的方法（Zhou 等，2022a；Roth 等，2022），存储正常样本的特征原型并通过比对匹配度识别异常；基于自监督学习的方法（Grill 等，2020；Cai 等，2023），通过设计代理任务学习正常特征的鲁棒表示；以及基于重建的检测方法（Schlegl 等，2017；Baur 等，2021；Zhang 等，2022），利用生成模型学习正常样本分布，通过计算输入与重建结果的差异量化异常。其中，基于重建的异常检测方法因其简便易行、可解释性强且能灵活结合其他方法，成为领域中应用最广泛的范式。

基于重建的异常检测方法经历了从数据重建到特征重建的发展历程。初期研究主要依赖自编码器、生成对抗网络等模型在图像空间进行像素级重建并识别异常，但易因模型泛化能力过强引发过度重建问题。例如 Baur 等人（2021）的基准性研究系统性归纳及评估了多种自编码器模型在脑磁共振图像上的表现；Bercea 等人（2023）引入密集变形场约束，通过几何变换增强重建过程的判别能力；Kasce-

nas等人(2023)系统性探索了去噪自编码器在医学图像异常检测中的性能;Schlegl等人(2017)最先将生成对抗网络引入医学异常检测并提出AnoGAN(anomaly detection with generative adversarial network),其迭代过程计算成本极高,后续研究者针对该问题提出了GANomaly(Akcay等,2018)和f-AnoGAN(Schlegl等,2019)等更高效的方法;Han等人(2021)提出的MADGAN(medical anomaly detection generative adversarial network)针对三维脑部磁共振成像(magnetic resonance imaging, MRI)数据特点,利用多个相邻切片重建以捕捉空间上下文信息;Behrendt等人(2024)提出基于图像块扩散模型的方法,通过利用局部上下文信息有效抑制异常区域的重建质量。近期,研究重点逐渐转向特征重建方法(Rippel等,2021;You等,2022b),即利用大规模数据集的预训练模型提取图像的高层语义特征,在特征空间中进行重建与差异度量。例如Rippel等人(2021)对预训练特征进行多元高斯分布拟合,依据马氏距离计算异常分数;Wang等人(2021)提出的学生—教师特征金字塔匹配(student-teacher feature pyramid matching, STFPM)方法构建师生特征金字塔匹配机制;Salehi等人(2021)提出的多分辨率知识蒸馏(multiresolution knowledge distillation, MKD)方法通过多分辨率知识蒸馏策略迁移预训练特征;You等人(2022a)提出的异常检测Transformer(anomaly detection Transformer, ADTR)框架探索了Transformer结构在特征重建中的潜力;AE-Flow(autoencoders with normalizing flows)(Zhao等,2023)在编码器和解码器之间加入归一化流模块;UniAD(unified anomaly detection)(You等,2022b)设计统一的多类别检测框架,自适应聚合多尺度预训练特征;Deng和Li(2022)提出的反向蒸馏单类嵌入异常检测(reverse distillation from one-class embedding for anomaly detection, RD4AD)模型采用反向知识蒸馏架构,实现端到端的特征重建与异常检测。多项研究表明,预训练特征能够提供高判别性与无噪声的图像表示,有效缓解过度重建问题,但为防止端到端优化中预训练模型出现模式崩溃(Guo等,2024;Tang等,2025),训练过程中其参数通常会被冻结。由于自然图像与医学图像间存在显著语义差异,该做法往往导致预训练模型无法通过微调适配至医学域,制约检测性能,主要体现为将组织解剖结构差异误判为病灶异常,以及对早期病变等细微异常敏感性不足。此外,近期的一项综述性研究也指出,现有方法在跨数据集评估时普遍面临泛化能力不足与评估基准不一致的问题,进一步制约了其在真实临床场景中的应用可靠性(Cai等,2025)。

为突破编码器冻结带来的性能瓶颈,Guo等人(2024)提出编码器—解码器对比方法EDC(encoder-decoder contrast),通过联合优化编码器与解码器,结合停止梯度操作与全局余弦损失缓解模式崩溃,实现编码器的稳定微调;Tang等人(2025)提出的EA2D(encoder-attention-2decoder)方法进一步整合转换一致性对比学习与双解码器重建机制,显式优化预训练编码器以缩小自然图像与医学图像之间的语义差异。此外,基于异常合成的方法中,CutPaste方法(Li等,2021)通过裁剪和旋转图像块合成异常,Tan等人(2021)提出泊松图像插值方法生成更真实的异常区域,Sato等人(2023)提出解剖感知粘贴策略,利用肺部分割掩膜指导胸部X光片中的异常合成;基于记忆库匹配的方法中,MemAE(Gong等,2019)在自编码器中引入可学习的记忆模块,ProxyAno(Zhou等,2022a)以超像素作为代理任务,MemSTC-Net(Zhou等,2022b)通过记忆正常图像的结构—纹理对应关系,PaDiM(patch distribution modeling)方法(Defard等,2021)建立参数化的多元高斯分布建模正常样本统计特征;基于自监督学习的方法中,SALAD(self-supervised and translation-consistent features for anomaly detection)(Zhao等,2021)在图像和特征空间构建双向重建路径,PatchCL-AE(Lu等,2024)将重建图像的局部块与非对应块作为正负样本,增强模型对局部语义一致性的建模能力。

针对现有方法中预训练模型跨域迁移的语义差异问题,本文提出一种基于跨域自监督表征学习的端到端重建框架,用于解决预训练模型在医学图像异常检测中的迁移适应性问题。该框架包含两个关键阶段:第1阶段基于BYOL(bootstrap your own latent)自监督学习实现预训练编码器的医学域适应性迁移;第2阶段构建端到端的特征重建网络,利用跨域适应的编码器提取判别性特征,并通过解码器重建与多模块协同实现异常区域的精准检测与定位。

本文的贡献总结如下:1)构建域适应对比学习模型,有效解决自然图像预训练模型在医学域

的迁移适应性问题，缓解语义差异导致的性能制约；2）设计一种协同注意力策略，通过坐标、通道与空间注意力的联合嵌入，增强模型对病变特征的判别性表示能力；3）提出自适应多尺度异常图融合模块，实现跨层级异常响应的动态加权与空间优化。

1　方　法

本文提出一种面向医学图像异常检测的跨域自监督表征学习（cross-domain self-supervised representation learning，CSRL）框架。该框架旨在解决自然图像预训练模型在迁移至医学领域时存在的语义差异问题，并通过端到端的特征重建与多模块协同，实现精准的异常检测与定位。方法的整体流程如图 1 所示，其主要由两个阶段构成，第 1 阶段通过构建域适应对比学习（domain-adaptive contrastive learning，DACL）网络进行跨域自监督迁移，旨在将预训练编码器适配至医学图像域以优化其特征表示；第 2 阶段为基于特征重建的异常检测，利用域适配后的编码器—解码器架构，结合协同注意力增强模块（synergistic attention enhancement，SAE）用于聚焦并增强潜在病变的特征响应，并引入多尺度异常融合模块（multi-scale anomaly fusion，MAF）以整合不同层级的语义信息，最终实现对异常区域的精准量化与定位。下文将对各模块进行详细阐述。

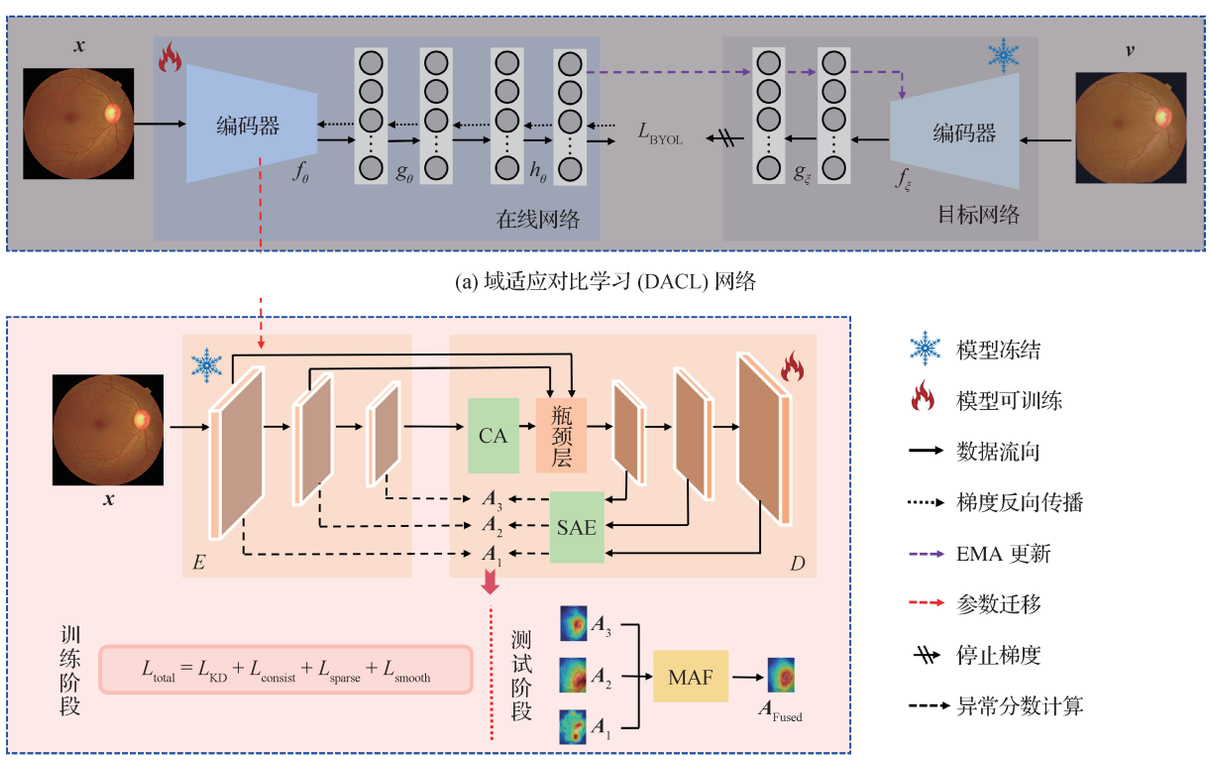


图 1　跨域自监督表征学习框架整体架构

Fig. 1　The overall architecture of the cross-domain self-supervised representation learning framework ((a) domain-adaptive contrastive learning (DACL) network; (b) feature reconstruction network)

1.1　域适应对比学习网络

为缓解直接迁移自然图像预训练模型所导致的域偏移问题，以增强特征编码器对医学信息的判别性感知能力，本文设计了域适应对比学习（DACL）网络。该模块采用 BYOL 的在线—目标网络双路径框架，通过最大化在线与目标网络对同一图像不同增强视图的表征一致性，以驱动预训练编码器进行医学域适应性微调。

如图 1（a）所示，DACL 通过构建两个结构相同但参数更新方式不同的神经网络学习稳健的医学特征表示，其分别称为在线网络和目标网络，其中在线网络的参数记做 $\boldsymbol{\theta}$，目标网络参数则记做 $\boldsymbol{\xi}$。给定一个包含 N 个正常样本的训练集 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N\}$，$\boldsymbol{x}_i\in\mathbf{R}^{H\times W\times 3}$，首先对每个输入样本 $\boldsymbol{x}$，

分别进行随机数据增强操作以得到与原始图像不同的视图，具体为

$$\boldsymbol{v} = t(\boldsymbol{x}) \tag{1}$$

式中，$t \sim \boldsymbol{T}$表示数据增强操作t从预定义的$\boldsymbol{T}$中随机采样，$\boldsymbol{T}$为本研究中针对表征学习设计的增强策略集合。具体而言，采用的增强操作包括随机缩放裁剪、水平翻转、颜色抖动和高斯模糊等，以从多个维度提升模型对医学图像常见变化的鲁棒性，并迫使模型学习具有变换无关性的核心特征。

在线网络与目标网络均包括一个编码器和一个投影头，其中编码器的主干网络为Wide-ResNet50-2模型，投影头则由一个多层感知机（multilayer perceptron，MLP）网络构成，在线网络额外包含一个MLP预测头。视图$\boldsymbol{x}$和$\boldsymbol{v}$分别通过上述两个网络转化为预测向量$\boldsymbol{q}_\theta$和投影向量$\boldsymbol{z}'_\xi$，表示为

$$\boldsymbol{y}_\theta = f_\theta(x), \boldsymbol{z}_\theta = g_\theta(y_\theta), \boldsymbol{q}_\theta = h_\theta(z_\theta) \tag{2}$$

$$\boldsymbol{y}'_\xi = f_\xi(\boldsymbol{v}), \boldsymbol{z}'_\xi = g_\xi(\boldsymbol{y}'_\xi) \tag{3}$$

式中，f代表编码器，g代表投影头，h代表预测头。目标网络的参数$\boldsymbol{\xi}$不通过梯度下降直接优化，而是作为在线网络参数θ的指数移动平均（exponential moving average，EMA）进行缓慢更新，表示为

$$\boldsymbol{\xi} \leftarrow \tau\boldsymbol{\xi} + (1 - \tau)\theta \tag{4}$$

式中，$\tau \in [0, 1]$是动量衰减系数，EMA的间接更新机制旨在提升训练过程的稳定性，从而避免模式坍塌。

DACL的目标是在正常医学图像模式的特征空间中，最大化在线网络对视图$\boldsymbol{x}$的预测$\boldsymbol{q}_\theta$与目标网络对视图$\boldsymbol{v}$的$\boldsymbol{z}'_\xi$投影之间的一致性，以迫使编码器专注于学习稳健的解剖结构表征。为此，定义损失函数为BYOL对称负余弦相似度，具体为

$$L_{\mathrm{BYOL}}^{\theta,\xi} = 2 - 2 \cdot \frac{\left\langle \boldsymbol{q}_\theta, \boldsymbol{z}'_\xi \right\rangle}{\| \boldsymbol{q}_\theta \|_2 \cdot \| \boldsymbol{z}'_\xi \|_2} \tag{5}$$

式中，$\langle \cdot, \cdot \rangle$表示向量内积。随后，通过交换$\boldsymbol{x}$和$\boldsymbol{v}$输入两个网络，得到一组对称的损失$\tilde{L}_{\mathrm{BYOL}}^{\theta,\xi}$。该阶段的总损失函数为二者之和，即

$$L_{\mathrm{BYOL}} = L_{\mathrm{BYOL}}^{\theta,\xi} + \tilde{L}_{\mathrm{BYOL}}^{\theta,\xi} \tag{6}$$

通过最小化该损失，编码器f_θ能够学习到对医学图像内容敏感、且对数据增强不变的特征表示，从而缩小自然图像与医学图像之间的语义差异。同时，在线网络的编码器将被冻结，并用于后续的特征重建网络。

1.2 特征重建网络

第2阶段的任务为构建一个端到端的特征重建网络，用于实现对图像异常的量化评分与空间定位。如图1(b)所示，其采用编码—解码结构，并引入了两个核心模块以优化异常特征的表示与定位。

1.2.1 网络架构组成

1）编码器。编码器E采用完成跨域自适应训练的DACL在线编码器f_θ的参数作为初始化，并在该阶段的训练过程中保持冻结状态，以规避模式崩溃风险。给定输入图像$\boldsymbol{x}$，编码器分别提取3个不同尺度的深层特征图，即

$$\{\boldsymbol{F}_1, \boldsymbol{F}_2, \boldsymbol{F}_3\} = E(\boldsymbol{x}) \tag{7}$$

2）瓶颈层。编码器输出的多尺度特征共同被送入一个瓶颈层（bottleneck layer，BL）。如图2所示，该模块用于统一不同层次特征图的尺度并对其信息进行融合，从而为解码器提供一个紧凑且信息丰富的特征表示，具体为

$$\varphi = BL(\boldsymbol{F}_1, \boldsymbol{F}_2, \boldsymbol{F}_3) \tag{8}$$

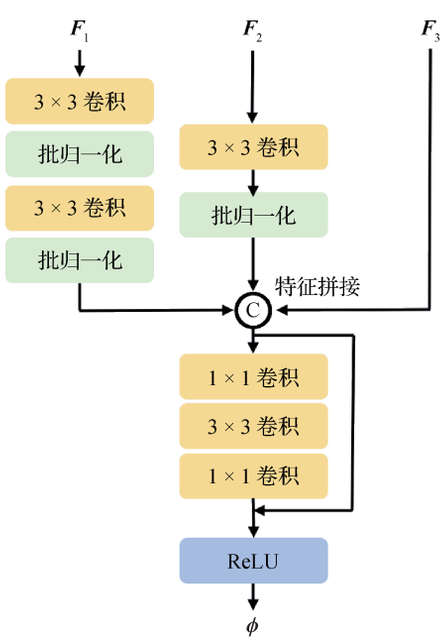


图2 瓶颈层结构示意图

Fig. 2 Schematic diagram of the bottleneck layer

3）解码器。解码器D采用与编码器近似对称的反卷积网络de-Wide-ResNet50-2，由多个上采样块组成。每个上采样块通常包含转置卷积或双线性插值上采样，后接常规卷积、批归一化和修正线性单元（rectified linear unit，ReLU）激活。解码器的目标是将瓶颈层输出的融合特征$\boldsymbol{\phi}$逐步上采样，并重建出

与编码器多尺度特征图空间尺寸相对应的特征，具体为

$$\{\hat{\boldsymbol{F}}_1, \hat{\boldsymbol{F}}_2, \hat{\boldsymbol{F}}_3\} = D(\boldsymbol{\phi}) \tag{9}$$

1.2.2　协同注意力增强模块

为了进一步提升模型对潜在异常区域的聚焦能力，本文在特征重建网络中引入协同注意力增强(SAE)模块，其结构如图3所示。该模块并行地集成了3个注意力分支，以提升模型对潜在异常区域的聚焦能力。

1)坐标注意力(coordinate attention，CA)分支。该分支通过分解坐标信息，捕获像素间精确的行、列依赖关系。

CA模块首先作用于编码器输出特征 $\boldsymbol{F}_3 \in \mathbf{R}^{C\times H\times W}$ 以增强下采样过程中损失的空间结构信息，其使用两个一维全局平均池化(global average pooling，GAP)核，分别沿水平坐标和垂直坐标对每个通道进行编码，得到一对方向感知的特征向量。具体为

$$z_c^h = \frac{1}{W}\sum_{0\leqslant j<W} \boldsymbol{F}_3(c,i,j) \tag{10}$$

$$z_c^w = \frac{1}{H}\sum_{0\leqslant i<H} \boldsymbol{F}_3(c,i,j) \tag{11}$$

式中，$\boldsymbol{F}_3(c,i,j)$ 代表在位置 (i,j) 处，第 c 个通道的特征图，H 与 W 分别为特征图的高度与宽度。随后，将上述两个特征图进行拼接后送入一个通道压缩比为 r 的 1×1 卷积变换函数，通过非线性激活，生成中间特征图 $\boldsymbol{F}\in\mathbf{R}^{C/r\times1\times(H+W)}$，并沿空间维将 $\boldsymbol{F}$ 分割为 $\boldsymbol{F}^h\in\mathbf{R}^{C/r\times H\times1}$ 和 $\boldsymbol{F}^w\in\mathbf{R}^{C/r\times1\times W}$ 两个独立张量。最后，分别使用两个 1×1 卷积 C^h 与 C^w，将 $\boldsymbol{F}^h$ 与 $\boldsymbol{F}^w$ 调整为与原特征图 $\boldsymbol{F}_3$ 相同的通道数，并应用sigmoid激活函数得到水平方向和垂直方向的注意力权重，具体为

$$\boldsymbol{G}_h = \sigma(C_h(\boldsymbol{F}^h)),\ \boldsymbol{G}_w = \sigma(C_w(\boldsymbol{F}^w)) \tag{12}$$

最终，坐标注意力模块的输出通过对原始输入进行加权，得到

$$\boldsymbol{F}_3' = \boldsymbol{F}_3 \otimes \boldsymbol{G}_h \otimes \boldsymbol{G}_w \tag{13}$$

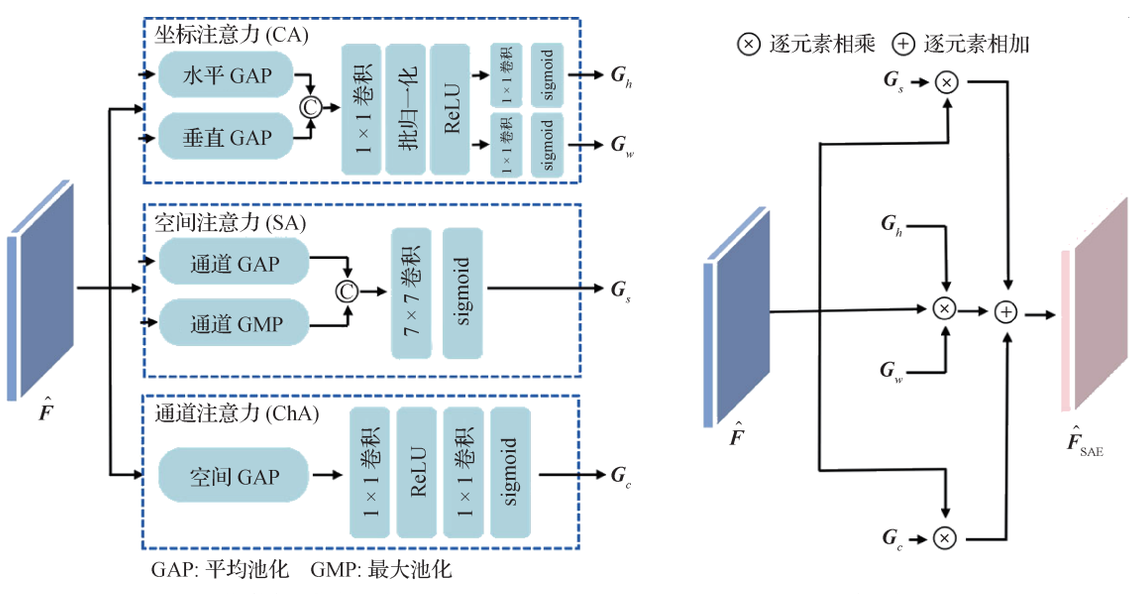
式中，⊗表示逐元素相乘。该模块输出 $\boldsymbol{F}_3'$，在保留原始语义信息的同时，进一步增强了在空间位置上的判别能力。经CA增强的特征 $\boldsymbol{F}_3'$ 与两个中低层特征 $\boldsymbol{F}_1$，$\boldsymbol{F}_2$ 共同被送入瓶颈层。

此外，对于不同尺度的解码器重建特征 $\hat{\boldsymbol{F}}\in\mathbf{R}^{C\times H\times W}$，同样利用CA模块以增强其异常定位，该过程表示为 $\hat{\boldsymbol{F}}_{\mathrm{ca}} = \mathrm{CA}(\hat{\boldsymbol{F}})$。

2)通道注意力(channel attention，ChA)分支。该分支采用近似于挤压—激励网络的结构，旨在建模通道间的相互依赖关系，并自适应地重新校准通道维度的特征响应。首先，通过全局平均池化将空间信息压缩为一个通道描述符。即

$$z_c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\hat{\boldsymbol{F}}_c(i,j) \tag{14}$$

随后，通过两个 1×1 卷积层和一个sigmoid激



(a) 注意力权重生成机制　　(b) 特征增强融合过程

图3　协同注意力增强模块结构示意图

Fig. 3　Schematic diagram of the synergistic attention enhancement module

((a) attention weight generation mechanism; (b) feature enhancement and fusion process)

活函数，为每个通道生成一个权重，具体为

$$\boldsymbol{G}_c = \sigma(\boldsymbol{W}_2\delta(\boldsymbol{W}_1 z_c)) \tag{15}$$

式中，$\boldsymbol{W}_1 \in \mathbf{R}^{C/r \times C}$，$\boldsymbol{W}_2 \in \mathbf{R}^{C \times C/r}$为可学习参数，$\delta$为ReLU激活函数，$r$为通道压缩比。最终，通道注意力的输出为$\hat{\boldsymbol{F}}_{ch} = \hat{\boldsymbol{F}} \otimes \boldsymbol{G}_c$。

3）空间注意力（spatial attention，SA）分支。该分支专注于发现特征图中最具信息量的空间区域。它通过合并跨通道的最大池化（MaxPool）和平均池化（AvgPool）信息生成一个空间注意力图，具体为

$$\boldsymbol{G}_s = \sigma(Conv_{7\times7}([AvgPool(\hat{\boldsymbol{F}}); MaxPool(\hat{\boldsymbol{F}})])) \tag{16}$$

式中，$[\cdot;\cdot]$表示沿通道维的拼接，$Conv_{7\times7}$是一个核大小为7×7的卷积层，用于融合上下文信息。空间注意力的输出为$\hat{\boldsymbol{F}}_{sp} = \hat{\boldsymbol{F}} \otimes \boldsymbol{G}_s$。

4）动态聚合。将3个注意力分支的输出进行加权求和，得到SAE模块的最终输出，具体为

$$\hat{\boldsymbol{F}}_{\mathrm{SAE}} = \alpha \cdot \hat{\boldsymbol{F}}_{ca} + \beta \cdot \hat{\boldsymbol{F}}_{ch} + \gamma \cdot \hat{\boldsymbol{F}}_{sp} \tag{17}$$

式中，α,β,γ是可学习的标量权重。为增强训练稳定性并避免权重失衡风险，通过softmax函数对这些权重进行归一化约束，确保其非负且和为1。并在训练过程中通过反向传播进行优化，从而使得网络能够自适应调整三分支优先级。

将SAE模块分别应用于解码器输出的3个重建特征$\hat{\boldsymbol{F}}_1, \hat{\boldsymbol{F}}_2, \hat{\boldsymbol{F}}_3$，即可得到经过注意力增强的重建特征$\hat{\boldsymbol{F}}_1^a, \hat{\boldsymbol{F}}_2^a, \hat{\boldsymbol{F}}_3^a$。

1.3 多尺度异常融合模块

在推理阶段，通过计算编码器输出特征与经SAE增强后的解码器重建特征之间的差异来生成异常图，用于定量指标计算及可视化异常定位。本文采用余弦距离进行度量，主要是因其对特征幅度的变化不敏感，能更好地反映语义上的偏离。

分别在不同尺度的特征上进行计算，对于第k个尺度，其异常图$\boldsymbol{A}_k$计算为

$$\boldsymbol{A}_k = 1 - \mathrm{Cos}_{\mathrm{Similarity}}(\boldsymbol{F}_k, \hat{\boldsymbol{F}}_k^a), \quad k \in \{1,2,3\} \tag{18}$$

式中，$\boldsymbol{F}_k$表示该尺度的输入特征，$\hat{\boldsymbol{F}}_k^a$表示对应的重建特征，$\mathrm{Cos}_{\mathrm{Similarity}}$表示余弦相似度计算，$\boldsymbol{A}_k$值越高，表示该位置与正常模式的偏离程度越大。随后，将所有$\boldsymbol{A}_k$通过双线性插值上采样恢复至输入图像的空间尺寸$H \times W$。

k个不同尺度的异常图在信息上存在互补性，其中浅层特征图空间细节丰富，利于精确定位；而深层特征图则语义信息强，更擅长于识别高级别的语义异常。为了针对不同模态的医学图像自适应地利用这些互补信息，本文设计了一种多尺度异常融合（MAF）模块，通过动态加权聚合来优化最终的异常图输出，其流程如图4所示。

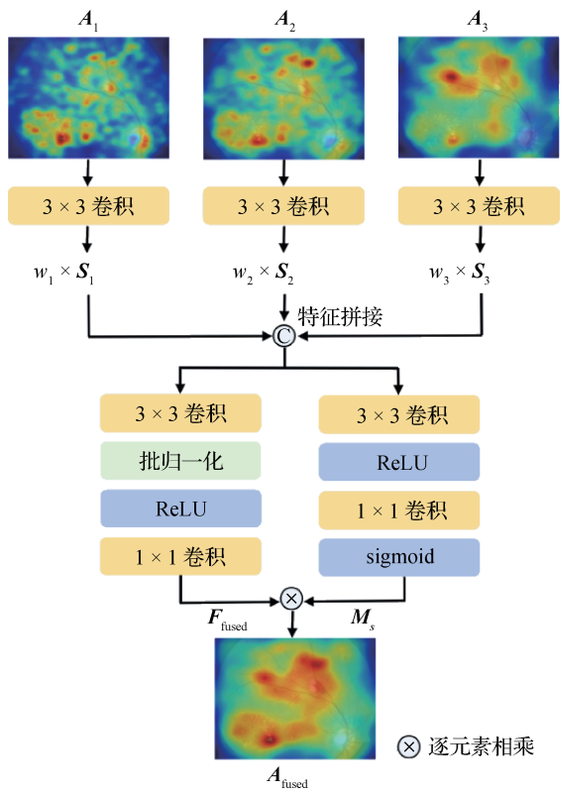


图4 多尺度异常融合模块流程示意图

Fig. 4 Schematic diagram of the multi-scale anomaly fusion module workflow

首先，将多尺度异常图通过双线性插值统一至目标空间分辨率。随后，使用独立的3×3卷积核将每个上采样后的异常图$\boldsymbol{A}_k$映射到一个共享的特征空间，得到$\boldsymbol{S}_k \in \mathbf{R}^{C_f \times H \times W}$。同时，通过可学习的尺度权重参数$\lambda \in \{\lambda_1, \lambda_2, \lambda_3\}$，经softmax归一化后得到各尺度的融合权重$w_k$，具体为

$$w_k = \frac{\exp(\lambda_k)}{\sum_{j=1}^{3}\exp(\lambda_j)}, \quad k \in \{1,2,3\} \tag{19}$$

尺度加权特征计算如下

$$\tilde{\boldsymbol{F}}_k = w_k \cdot \boldsymbol{S}_k \tag{20}$$

随后，为了进一步细化融合结果，首先将加权后的多尺度特征沿通道维度拼接，即

$$\boldsymbol{F}_{\mathrm{cat}} = Concat[\tilde{\boldsymbol{F}}_1, \tilde{\boldsymbol{F}}_2, \tilde{\boldsymbol{F}}_3] \tag{21}$$

然后，对$\boldsymbol{F}_{\mathrm{cat}}$进行深度特征融合，即

$$\boldsymbol{F}_{\text{fused}} = Conv_{1\times1}\left(ReLU\left(BN(Conv_{3\times3}(\boldsymbol{F}_{\text{cat}}))\right)\right) \quad (22)$$

此外，将 $\boldsymbol{F}_{\text{cat}}$ 输入一个轻量的空间注意力单元生成一个空间权重图 $\boldsymbol{M}_s \in \mathbf{R}^{1\times H\times W}$，具体为

$$\boldsymbol{M}_s = \sigma\left(Conv_{1\times1}\left(ReLU(Conv_{3\times3}(\boldsymbol{F}_{\text{cat}}))\right)\right) \quad (23)$$

式中，σ 表示 sigmoid 函数。最终的多尺度融合异常图 $\boldsymbol{A}_{\text{fused}}$ 通过对 $\boldsymbol{F}_{\text{fused}}$ 应用空间注意力权重 $\boldsymbol{M}_s$ 得到，具体为

$$\boldsymbol{A}_{\text{fused}} = \boldsymbol{F}_{\text{fused}} \odot \boldsymbol{M}_s \quad (24)$$

$\boldsymbol{A}_{\text{fused}}$ 中的每个像素值代表了该位置为异常的概率，$\odot$ 表示逐元素相乘。

作为对 MAF 模块的约束，以保证融合异常图能够在保留各尺度重建信息的基础上，生成高判别性且与空间分布一致的异常定位结果，本文采用了如下的联合损失函数，具体为

$$L_{\text{consist}} = \frac{1}{3}\sum_{k=1}^{3}\| \boldsymbol{A}_{\text{fused}} - Resize(\boldsymbol{A}_k)\|_2^2 \quad (25)$$

$$L_{\text{sparse}} = \| \boldsymbol{A}_{\text{fused}}\|_1 \quad (26)$$

式中，L_{consist} 为融合一致性损失，*Resize* 表示调整尺寸到融合异常图大小，该项度量融合后的异常图 $\boldsymbol{A}_{\text{fused}}$ 与每一个尺度的初始异常图之间的差异。L_{sparse} 为稀疏性约束损失，通过对融合后的异常图施加 L_1 正则化，引导模型产生更加集中的异常激活区域。

除上述损失外，本文还构建了重建损失用于约束整体特征重建过程。该项作为异常检测任务的核心，旨在引导解码器准确重构各尺度的正常特征，从而迫使异常区域产生高重建误差，其目标为最小化编码器原始特征与解码器重建特征之间的差异，表示为

$$L_{\text{recon}} = \frac{1}{3}\sum_{k=1}^{3}\boldsymbol{E}[\boldsymbol{A}_k] \quad (27)$$

最终，第 2 阶段训练的总损失函数为上述 3 项的加权和，即

$$L_{\text{total}} = L_{\text{recon}} + L_{\text{consist}} + \lambda L_{\text{sparse}} \quad (28)$$

式中，λ 作为平衡各项损失贡献的权重因子。通过联合优化该损失函数，从而使模型既能准确地重建正常模式，同时能够通过多尺度融合和空间约束，生成高精度的异常定位图。

2　实验结果与分析

2.1　数据集

为了全面评估本文所提出 CSRL 框架的检测性能，在 3 个公开医学影像数据集上进行实验。

1）APTOS（Asia Pacific Tele-Ophthalmology Society）数据集（Maggie 和 Dane，2019）。该数据集由 2019 年 APTOS 盲症检测挑战赛官方发布，共包含 3 662 幅糖尿病性视网膜眼底 RGB 图像，健康（正常）图像 1 805 幅；病变（异常）图像 1857 幅。本文从正常样本中随机选取 1 000 幅图像作为训练集，其余 2 662 幅正常与异常图像则共同组合为测试集。在预处理过程中，图像统一调整为 256 × 256 像素的尺寸。

2）ISIC（International Skin Imaging Collaboration）数据集（Codella 等，2019）。该数据集由 ISIC2018 挑战赛的任务 3 发布，包含 6 898 幅皮肤 RGB 图像，分为 7 种类别，本文以其中的色素痣皮肤图像作为正常样本，其余 6 个不同类别的疾病皮肤图像作为异常样本。官方训练集包含 6 705 幅正常图像，官方验证集用做测试集，其中包括 123 幅正常图像与 70 幅异常图像。在预处理过程中，图像调整为 256 × 256 像素，并中心裁剪至 224 × 224 像素。

3）BR35H 数据集（Hamada，2020）。该数据集发布于 Kaggle 挑战赛，由从 3D 模态图像中截取的 2D 脑部 MRI 灰度图像切片构成，其中包含 1 500 幅作为异常样本的肿瘤图像以及 1 500 幅作为正常样本的健康图像。本文从正常图像中随机选取 1 000 幅图像作为训练集，其余 2 000 幅图像则作为测试集，所有图像调整为 64 × 64 像素。

2.2　实验设置

本文所有实验均基于 PyTorch 框架，在单张 NVIDIA GeForce RTX 4090 显卡上进行。模型使用 Adam 优化器进行训练，在 APTOS、ISIC 及 BR35H 这 3 个数据集中，BYOL 自监督预训练阶段模型的学习率分别均设置为 5×10^{-5}，网络训练批尺寸均设置为 32，分别训练 50、50 和 100 个轮次；特征重建阶段模型的学习率分别为 $\{1\times10^{-2},\ 1\times10^{-3},\ 5\times10^{-3}\}$，网络训练批尺寸均设置为 16，分别训练 75、15 和 100 个轮次。APTOS 与 ISIC 数据集的病变区域分布较为弥散，而 BR35H 数据集的病灶则较为集中。因此，对于前两者采用 $\boldsymbol{A}_{\text{fused}}$ 的均值作为图像级的异常评分，以反映整体异常；而对于后者则采用 $\boldsymbol{A}_{\text{fused}}$ 的最大值作为图像级异常评分，以聚焦于其中的局部异常。

2.3 评价指标

本研究采用的评估指标包括受试者工作特征曲线下面积(area under curve, AUC)、F1分数(F1-score)、准确率(accuracy, ACC)、敏感性(sensitivity, SEN)和特异性(specificity, SPE)。其中,AUC、F1和ACC能够对整体性能进行评估,而SEN和SPE则分别侧重于正样本和负样本,模型的分类阈值由F1的最优值确定。作为图像级检测任务,本研究将单幅图像作为样本单位,统计真阳性、真阴性、假阳性和假阴性的样本量并用于计算上述各项指标。

2.4 对比实验

将本文提出的CSRL方法与经典方法及当前最先进技术(state-of-the-art, SOTA)在3个医学数据集上进行了对比。其中包括基于像素重建的基础方法自动编码器(autoencoder, AE)与变分自编码器(variational autoencoder, VAE);基于生成对抗网络(generative adversarial network, GAN)的经典像素重建方法GANomaly(Akcay等, 2018)、f-AnoGAN(Schlegl等, 2019)和SALAD(Zhao等, 2021);采用基于记忆库匹配机制的PaDiM方法(Defard等, 2021);采用预训练编码器的方法,如基于特征蒸馏的STFPM(Wang等, 2021)和MKD(Salehi等, 2021)方法,以及目前表现出最佳性能的反向蒸馏重建类方法EDC(Guo等, 2024)与EA2D(Tang等, 2025)。对比实验的结果来自于已发表论文的公开数据,或通过复现其开源代码获得。为与现有研究进行全面对比,本文同时展示了多次实验的平均结果及训练中出现的最佳结果。其中,本文方法的平均结果为5次独立实验最终迭代结果的平均值;最佳结果(以†标注)则记录了训练周期内最高AUC值所对应的各项指标。

2.4.1 量化对比结果

1)BR35H数据集中作为异常的肿瘤区域显著性较高,属于易检测场景。实验结果如表1所示,其中基于预训练编码器的方法性能显著优于从头训练的像素重建方法。例如,AE的AUC仅为89.57%,而多数预训练类方法则均取得大于98%的高AUC,当前性能领先的方法更是进一步将性能提升至99%以上。这一结果凸显了利用大规模数据集预训练所获得的特征提取模型在医学异常检测任务中具有极高的迁移价值。相较于上述方法,本文CSRL方法在图像级AUC、F1、ACC与SPE指标上的平均结果与最佳结果均已达到当前的最优值,分别为99.90%(99.97%)、99.63%(99.70%)、99.45%(99.55%)与98.16%(99.20%)。这些量化指标显示,本文方法在该数据集上已趋近于理想状况下的检测性能。

表1 在BR35H数据集上的异常检测性能对比

Table 1 Comparison of anomaly detection performance on BR35H dataset

/%

方法	来源	AUC	F1	ACC	SEN	SPE
AE	–	89.57	89.98	84.20	94.66	52.80
VAE	–	91.36	91.24	86.80	91.73	72.00
GANomaly	ACCV'18	93.78	92.85	89.10	94.40	73.20
f-AnoGAN	MIA'19	94.80	93.76	90.45	95.66	74.80
SALAD	TMI'21	97.69	96.83	95.15	98.86	96.10
STFPM	BMVC'21	93.11	93.46	89.80	97.20	67.60
MKD	CVPR'21	98.70	94.30	96.26	98.06	83.00
PaDiM	ICPR'21	98.19	98.00	96.95	99.80	88.40
EDC	TMI'24	99.85	99.57	99.35	99.95	97.40
EA2D	TMI'25	99.83	99.62	99.44	**99.99**	97.80
CSRL(本文)	–	**99.90**	**99.63**	**99.45**	99.88	**98.16**
EDC†	TMI'24	99.94	99.50	99.24	**99.93**	97.20
EA2D†	TMI'25	99.95	**99.70**	**99.55**	99.80	98.80
CSRL(本文)†	–	**99.97**	**99.70**	**99.55**	99.70	**99.20**

注:加粗字体分别表示平均结果组和最佳结果组中各列的最优结果。"†"代表训练过程中取得的最佳结果。"–"表示无具体来源,ACCV: Asian Conference on Computer Vision; MIA: Medical Image Analysis; TMI: IEEE Transactions on Medical Imaging; BMVC: British Machine Vision Conference; CVPR: IEEE/CVF Conference on Computer Vision and Pattern Recognition; ICPR: International Conference on Pattern Recognition。

2)ISIC数据集的特殊性在于其正常样本(色素痣)属于一种良性病变,与异常样本的视觉差异较小,因此整体检测难度较高。表2中的实验结果表明,多数现有方法在此挑战下表现不佳,其平均AUC普遍未达到90%,综合反映检测精度的F1分数也多低于70%。相较之下,EDC、EA2D与本文提出的CSRL方法具有较大的性能优势,各项指标均保持在80%以上,展现出在该任务中的可靠性。其中,本文提出的CSRL方法表现突出,在平均性能上,其AUC、F1、ACC与SPE这4项指标分别达到91.79%、83.46%、88.08%与90.86%,在全部方法

表 2　在 ISIC 数据集上的异常检测性能对比

Table 2　Comparison of anomaly detection performance on ISIC dataset

/%

方法	来源	AUC	F1	ACC	SEN	SPE
AE	–	80.72	69.23	75.13	77.14	62.79
VAE	–	80.49	67.02	74.96	70.00	77.78
GANomaly	ACCV’18	71.60	61.91	67.53	72.38	64.77
f-AnoGAN	MIA’19	75.90	65.06	69.95	77.14	65.85
SALAD	TMI’21	72.62	61.90	66.84	74.29	62.60
PaDiM	ICPR’21	82.13	72.04	73.06	**95.71**	60.16
EDC	TMI’24	89.33	80.10	84.28	84.76	84.01
EA2D	TMI’25	90.98	83.27	87.25	87.43	87.15
CSRL(本文)	–	**91.79**	**83.46**	**88.08**	83.21	**90.86**
EDC†	TMI’24	90.53	81.94	86.53	84.29	87.80
EA2D†	TMI’25	91.22	83.56	87.56	87.14	87.80
CSRL(本文)†	–	**94.74**	**88.57**	**91.71**	**88.57**	**93.50**

注：加粗字体表示平均结果组和最佳结果组中各列的最优结果。“†”代表训练过程中取得的最佳结果。“–”表示无具体来源。

表 3　在 APTOS 数据集上的异常检测性能对比

Table 3　Comparison of anomaly detection performance on APTOS dataset

/%

方法	来源	AUC	F1	ACC	SEN	SPE
AE	–	76.91	83.61	74.07	94.83	26.21
VAE	–	77.21	85.04	77.12	93.26	39.87
GANomaly	ACCV’18	83.72	85.85	79.37	89.71	55.52
f-AnoGAN	MIA’19	89.64	89.71	85.23	92.29	68.94
SALAD	TMI’21	78.46	85.06	77.35	78.76	42.48
STFPM	BMVC’21	93.59	91.86	88.35	94.23	74.78
MKD	CVPR’21	88.51	81.70	87.84	94.72	51.67
PaDiM	ICPR’21	77.17	85.47	77.24	95.96	34.04
EDC	TMI’24	95.84	93.49	90.93	93.34	85.36
EA2D	TMI’25	97.53	**95.95**	**94.35**	**95.91**	**90.95**
CSRL(本文)	–	**97.71**	95.22	93.78	95.75	87.63
EDC†	TMI’24	96.13	93.89	91.47	93.91	85.84
EA2D†	TMI’25	97.59	**96.01**	**94.40**	**96.55**	89.44
CSRL(本文)†	–	**97.94**	95.36	93.54	95.15	**89.81**

注：加粗字体平均结果组和最佳结果组中各列的最优结果。“†”代表训练过程中取得的最佳结果。“–”表示无具体来源。

中均取得最优结果；在最佳性能方面，CSRL 实现全面领先，其 AUC、F1、ACC、SEN 与 SPE 分别进一步提升至 94.74%、88.57%、91.71%、88.57% 与 93.50%，相较于次优方法 EA2D 提升较为显著，分别增加 3.52%、5.01%、4.15%、1.43% 与 5.70%，表明了 CSRL 方法在复杂检测任务中的鲁棒性。

3）APTOS 数据集的异常为糖尿病视网膜病变，其常表现为细微的血管病变或渗出物，部分轻度病变区域较为隐蔽，因此对模型的细粒度特征捕捉能力要求较高。实验结果如表 3 所示，在最具综合代表性的 AUC 指标方面，传统方法的平均 AUC 值普遍低于 90%，而 SOTA 方法 EDC 与 EA2D 则能够提升至 95% 以上。与之相比，本文提出的 CSRL 方法取得了所有方法中的最优 AUC 结果，平均值达到 97.71%，超越此前最高值 0.18%；而最佳值则达到 97.94%，提升 0.35%。以上结果证明了 CSRL 方法在 APTOS 数据集上具备可靠的检测能力。

2.4.2　异常定位可视化对比结果

为评估方法的可解释性并直观展示其病灶定位性能，绘制了本文 CSRL 方法与 EDC、EA2D 方法对 3 种数据集异常样本的可视化结果以进行对比分析。

为生成异常定位热图，将本文方法输出的每个 $\boldsymbol{A}_{\text{fused}}$ 矩阵归一化至[0，1]区间，并上采样至原始图像尺寸后叠加于输入图像之上。作为对比，EDC 与 EA2D 方法的热图则依据其开源代码的默认流程生成。结果如图 5 所示，CSRL 方法能够在病变区域生成高热度值的响应，不仅在空间分布上与病灶的真实轮廓表现出较好的一致性，且其显著性激活区域对病灶的聚焦程度也优于对比方法。可视化对比结果表明，CSRL 方法在不同形态的医学图像中均能生成更加集中、更具临床参考价值的异常定位结果，在像素级异常定位任务上具有显著优势。

2.5　消融实验

为了综合验证本文所提出 CSRL 框架的有效性，在所有数据集上进行消融实验，以评估 CSRL 框架中 DACL、SAE 及 MAF 这 3 个核心模块的贡献。实验记录了各配置在训练过程中取得的最佳性能，结果如表 4 所示。实验结果表明，每个模块均对模型检测性能的提升发挥出作用，且三者协同工作时效果普遍更佳。

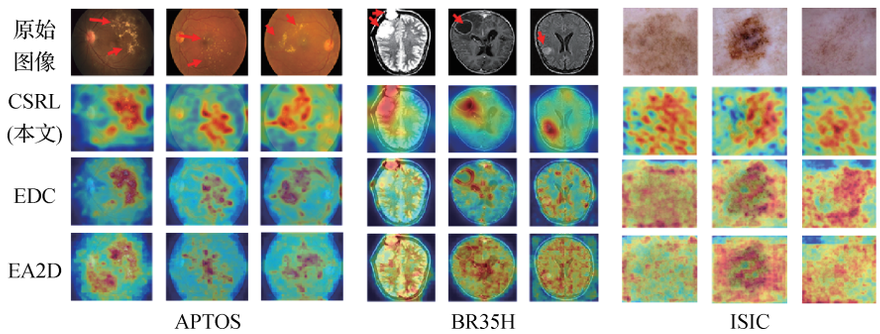


图5　不同方法异常定位结果可视化对比

Fig. 5　Visual comparison of anomaly localization results by different methods

表4　消融实验结果

Table 4　Results of ablation studies

/%

DACL	SAE	MAF	APTOS			ISIC			BR35H		
			ACC↑	F1↑	AUC↑	ACC↑	F1↑	AUC↑	ACC↑	F1↑	AUC↑
–	–	–	92.30	94.52	96.94	81.35	75.68	88.23	99.50	99.67	99.89
–	√	–	92.40	94.61	97.11	85.49	80.82	91.16	99.50	99.67	99.93
√	–	–	93.00	94.99	97.60	86.53	82.67	91.72	**99.60**	**99.73**	99.96
–	–	√	92.11	94.41	97.23	82.90	78.71	88.52	99.40	99.60	99.58
√	√	–	89.20	92.58	95.66	87.05	83.44	93.43	99.40	99.63	99.95
–	√	√	91.32	93.95	97.12	81.87	77.71	89.19	99.55	99.70	99.84
√	–	√	92.86	94.93	97.17	89.64	85.51	92.31	99.30	99.54	99.90
√	√	√	**93.78**	**95.22**	**97.71**	**91.71**	**88.57**	**94.74**	99.55	99.70	**99.97**

注：加粗字体表示各列最优结果。“†”代表训练过程中的最佳结果。“√”表示使用对应模块，“–”表示未使用。

在APTOS数据集上，在基础模型中引入SAE模块后，AUC提升至97.11%，表明注意力机制有效增强了病变特征表示。单独加入DACL模块后，AUC进一步提升至97.60%，验证了跨域自监督预训练对特征适应性的改善效果。MAF模块单独使用时，AUC达到97.23%，证明多尺度融合有助于整合不同层级的语义信息。当3个模块共同工作时，模型取得了最优的AUC 97.71%和ACC 93.78%，表明DACL、SAE与MAF三者能够有效协同，共同提升对细微病变的检测与定位能力。

在ISIC数据集上，各模块的贡献更为显著。基础模型的AUC最佳值为88.23%，单独加入SAE或DACL后，AUC分别提升至91.16%和91.72%，显示出在复杂视觉差异任务中，特征增强与域适应的重要作用。MAF单独使用亦带来提升，AUC达到88.52%。DACL与SAE共同作用时，AUC进一步提升至93.43%，表明两者在该任务中具有明显的协同效应。3个模块结合后，模型取得了最佳的AUC 94.74%和ACC 91.71%，验证了CSRL框架在困难检测场景中的有效性。

在BR35H数据集上，基础模型已具备优异的性能，AUC最佳值达到99.89%。各模块的引入均带来稳定提升，其中，3个模块联合后取得了最高的AUC 99.97%，证明CSRL框架在不同难度任务中均能保持领先的检测性能。

综合分析3个数据集的消融实验结果可见，基于BYOL的跨域自监督预训练模块（DACL）在提升模型性能方面展现出显著优势。在APTOS和ISIC数据集上，DACL分别带来0.66%和3.49%的AUC提升，其提升幅度在3个模块中表现最为突出。特别是在具有挑战性的ISIC数据集上，DACL与SAE模块的协同作用进一步将性能提升至93.43%，显示出其作为基础特征提取模块的重要价值。实验结果充分表明了跨域自适应表征学习能够通过将自然图

像中的通用视觉知识有效迁移至医学领域，从而显著提升了模型对医学异常特征的判别能力。

由于在异常检测任务中，异常定位的准确性对评估模型临床价值至关重要，为直观呈现多尺度特征融合在该方面的优势，本文还对MAF模块进行了可视化消融分析。

如图6所示，本文对比了MAF模块融合后的异常热图与3个不同尺度的初始异常图像。从视觉效果来看，浅层特征生成的异常图能够精确定位病灶核心区域，空间细节丰富，背景干扰较少，但对病灶边界的描述能力有限，覆盖范围明显不足；中层特征异常图在定位准确性和范围覆盖之间取得了一定的平衡，但部分区域仍存在热度弥散现象；深层特征异常图虽然能够覆盖完整的病灶区域，但空间聚焦性较弱，热度分布较为分散。经过MAF模块融合后，生成的异常图有效综合了各尺度的优势，不仅保持了高度的空间聚焦性，同时实现了完整的病灶范围覆盖，热力分布与病灶的真实解剖轮廓高度吻合。可视化结果表明，MAF模块通过自适应融合多尺度特征，在保持精确定位的同时较好地展现了病灶轮廓，从而提升了异常检测的准确性与临床可解释性。

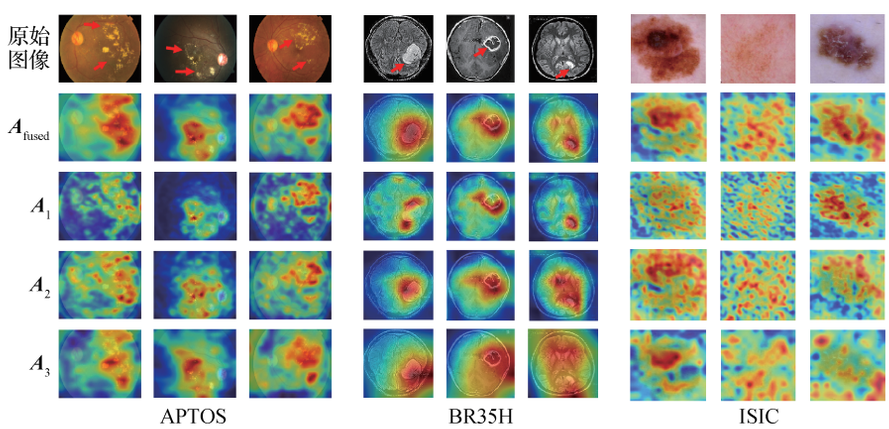


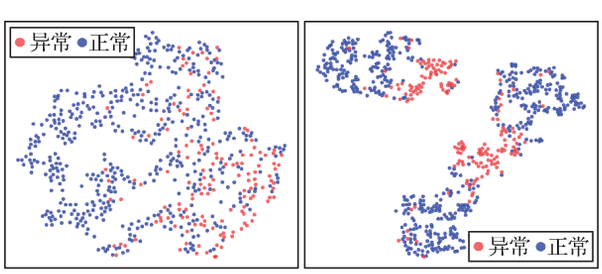
图6　多尺度异常融合模块定位效果可视化

Fig. 6　Visualization of localization performance with the multi-scale anomaly fusion module

2.6　特征可视化分析

为了进一步探究本文所提出CSRL框架的表征学习能力，使用t-SNE（t-distributed stochastic neighbor embedding）对ISIC数据集中正常与异常样本的特征分布进行了可视化分析。图7展示了经过DACL模块微调前后，编码器提取特征在二维空间中的分布情况，其中每个点对应于测试集中正常或异常样本的特征。可以看出，微调前特征空间中正常与异常样本存在较大重叠，表明自然图像预训练模型难以有效区分医学图像中的正常与病变模式；而经过DACL微调后，两类样本在特征空间中呈现出更清晰的聚类结构，且异常样本的分布更加紧凑，与正常样本间的边界更为明显。这进一步证明DACL模块能够有效缩小自然图像与医学图像间的语义差异，提升模型对异常特征的判别能力。

综合量化与可视化分析，消融实验充分证明了DACL、SAE与MAF这3个模块在CSRL框架中的必要性与互补性。DACL模块通过跨域自适应提升了特征表征的判别能力，SAE模块通过协同注意力机制增强了病灶区域的特征响应，而MAF模块则通过多尺度融合优化了异常定位的空间精度。通过多模块的协同作用，共同提升了医学图像异常检测的综合性能。



(a) DACL 微调前特征分布　(b) DACL 微调后特征分布

图7　基于t-SNE的跨域自监督学习模块特征可视化

Fig. 7　t-SNE-based feature visualization of the cross-domain self-supervised learning module ((a) feature distribution before DACL fine-tuning; (b) feature distribution after DACL fine-tuning)

3 结 论

本文提出了一种面向医学图像异常检测的跨域自监督表征学习框架(CSRL),旨在解决自然图像预训练模型在迁移至医学图像时存在的语义差异问题。该框架通过域适应对比学习(DACL)模块实现预训练编码器的跨域稳健迁移,利用协同注意力增强(SAE)模块强化病变区域的特征表示,并借助多尺度异常融合(MAF)模块优化异常定位的空间精度。

在APTOS、ISIC和BR35H这3个公开数据集上的实验结果表明,本文方法在图像级异常检测任务中均取得了领先的性能,显著优于现有的多种代表性方法。消融实验进一步验证了DACL、SAE和MAF这3个模块在提升模型判别能力与定位精度方面的有效性与互补性。可视化分析表明,本文方法生成的异常热图具有更高的空间一致性与临床可解释性,能够更准确地覆盖病灶区域。

尽管本文提出的CSRL框架在多个数据集上验证了其有效性,但仍存在若干局限性与值得探讨之处,为未来研究指明了方向:

1)对数据增强策略的敏感性。框架的DACL模块依赖于特定的强数据增强组合来实现有效的域不变特征学习。虽然本文采用的增强策略,如随机裁剪、颜色抖动、高斯模糊等在实验数据集中表现良好,但其普适性可能受到不同医学成像模态特有噪声与伪影的挑战。如何设计或自适应选择与目标模态特性相匹配的数据增强策略,以进一步提升跨域适应的鲁棒性,是未来值得探索的方向。

2)模型复杂性与临床部署考量。本文框架采用两阶段训练模式,并引入了包含多分支注意力的SAE模块以及多尺度融合的MAF模块,这无疑增加了模型的总体参数量与计算开销。因此,在追求高检测性能的同时,如何对模型进行轻量化设计、知识蒸馏或采用动态推理机制,以满足临床实时诊断或边缘设备部署的需求,是迈向实际应用的关键步骤。

3)对细微及早期异常的判别能力。实验结果显示,模型在综合评估指标AUC上优势明显,但在某些场景下敏感性(SEN)指标的提升相对有限。这表明模型对于对比度极低、边界模糊或处于非常早期阶段的病变,其判别能力仍有提升空间。未来的工作可以探索引入更精细的特征对齐损失、结合病变区域的先验形状知识,或利用极少量弱标注信息进行引导,以增强模型对此类困难样本的感知能力。

4)领域泛化与跨中心评估。本研究均在公开的、质量较高的数据集上进行训练与测试。在实际临床环境中,医学影像数据往往来源于不同厂家设备、不同采集协议以及不同人群分布,存在显著的域偏移。本研究尚未在跨中心、多设备的泛化性能上进行系统评估。构建更全面的跨域评估基准,并研究如何使模型具备更强的领域泛化能力,对于推动其真正的临床转化具有重要意义。

综上所述,本文工作为医学图像无监督异常检测提供了一种有效的跨域表征学习思路,未来研究将在保持性能优势的基础上,着力于模型的鲁棒性、高效性与泛化能力,推动此类技术向临床可靠应用迈进。

参考文献(References)

Akcay S, Atapour-Abarghouei A and Breckon T P. 2018. GANomaly: semi-supervised anomaly detection via adversarial training//Proceedings of the 14th Asian Conference on Computer Vision. Perth, Australia: Springer: 622-637 [DOI: 10.1007/978-3-030-20893-6_39]

Angiulli F, Basta S and Pizzuti C. 2006. Distance-based detection and prediction of outliers. IEEE Transactions on Knowledge and Data Engineering, 18(2): 145-160 [DOI: 10.1109/TKDE.2006.29]

Baur C, Denner S, Wiestler B, Navab N and Albarqouni S. 2021. Autoencoders for unsupervised anomaly segmentation in brain MR images: a comparative study. Medical Image Analysis, 69: #101952 [DOI: 10.1016/j.media.2020.101952]

Behrendt F, Bhattacharya D, Krüger J, Opfer R and Schlaefer A. 2024. Patched diffusion models for unsupervised anomaly detection in brain MRI//Proceedings of the Medical Imaging with Deep Learning. Nashville, USA: PMLR: 1019-1032

Bercea C I, Rueckert D and Schnabel J A. 2023. What do AEs learn? Challenging common assumptions in unsupervised anomaly detection//Proceedings of the 26th International Conference on Medical Image Computing and Computer-Assisted Intervention. Vancouver, Canada: Springer: 304-314 [DOI: 10.1007/978-3-031-43904-9_30]

Breunig M M, Kriegel H P, Ng R T and Sander J. 2000. LOF: identifying density-based local outliers//Proceedings of 2000 ACM SIGMOD International Conference on Management of Data. Dallas, America: ACM: 93-104 [DOI: 10.1145/342009.335388]

Cai Y, Chen H, Yang X, Zhou Y and Cheng K T. 2023. Dual-

distribution discrepancy with self-supervised refinement for anomaly detection in medical images. Medical Image Analysis, 86: #102794 [DOI: 10.1016/j.media.2023.102794]

Cai Y, Zhang W W, Chen H and Cheng K T. 2025. MedIAnomaly: a comparative study of anomaly detection in medical images. Medical Image Analysis, 102: #103500 [DOI: 10.1016/j. media. 2025. 103500]

Codella N, Rotemberg V, Tschandl P, Celebi M E, Dusza S, Gutman D, et al. 2019. Skin lesion analysis toward melanoma detection 2018: a challenge hosted by the international skin imaging collaboration (ISIC) [EB/OL]. [2025-11-23]. https://arxiv.org/pdf/1902.03368.pdf

Defard T, Setkov A, Loesch A and Audigier R. 2021. PaDiM: a patch distribution modeling framework for anomaly detection and localization//Proceedings of the Pattern Recognition. ICPR International Workshops and Challenges. Cham, Switzerland: 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. New Orleans, USA: IEEE: 9727-9736 [DOI: 10.1109/CVPR52688.2022.00951]

Gong D, Liu L Q, Le V, Saha B, Mansour M R, Venkatesh S, et al. 2019. Memorizing normality to detect anomaly: memory-augmented deep autoencoder for unsupervised anomaly detection//Proceedings of 2019 IEEE/CVF international conference on computer vision. Seoul, Korea (South): 1705-1714 [DOI: 10.1109/ICCV. 2019. 00179]

Grill J B, Strub F, Altché F, Tallec C, Richemond P H, Buchatskaya E, et al. 2020. Bootstrap your own latent a new approach to self-supervised learning//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #1786

Guo J, Lu S, Jia L Z, Zhang W H and Li H Q. 2024. Encoder-decoder contrast for unsupervised anomaly detection in medical images. IEEE Transactions on Medical Imaging, 43(3): 1102-1112 [DOI: 10.1109/TMI.2023.3327720]

Hamada A. 2020. Br35H: brain tumor detection 2020 [EB/OL]. [2025-11-23]. https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection

Han C, Rundo L, Murao K, Noguchi T, Shimahara Y, Milacski Z Á, et al. 2021. MADGAN: unsupervised medical anomaly detection GAN using multiple adjacent brain MRI slice reconstruction. BMC Bioinformatics, 22(2): #31 [DOI: 10.1186/S12859-020-03936-1]

Kascenas A, Sanchez P, Schrempf P, Wang C Y, Clackett W, Mikhael S S, et al. 2023. The role of noise in denoising models for anomaly detection in medical images. Medical Image Analysis, 90: #102963 [DOI: 10.1016/j.media.2023.102963]

Kim T, Lee Y G, Jeong I, Ham S Y and Woo S S. 2024. Patch-wise vector quantization for unsupervised medical anomaly detection. Pattern Recognition Letters, 184: 205-211 [DOI: 10.1016/j. patrec. 2024.06.028]

Li C L, Sohn K, Yoon J and Pfister T. 2021. CutPaste: self-supervised learning for anomaly detection and localization//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 9659-9669 [DOI: 10.1109/CVPR46437.2021.00954]

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]

Lu S, Zhang W H, Guo J, Liu H R, Li H Q and Wang N L. 2024. PatchCL-AE: anomaly detection for medical images using patch-wise contrastive learning-based auto-encoder. Computerized Medical Imaging and Graphics, 114: #102366 [DOI: 10.1016/j.compmedimag.2024.102366]

Maggie K and Dane S. 2019. APTOS 2019 blindness detection [EB/OL]. [2025-11-23]. https://kaggle.com/competitions/aptos2019-blindness-detection

Müller J P, Baugh M, Tan J, Dombrowski M and Kainz B. 2023. Confidence-aware and self-supervised image anomaly localisation//Proceedings of the 5th International Workshop on Uncertainty for Safe Utilization of Machine Learning in Medical Imaging. Vancouver, Canada: Springer: 177-187 [DOI: 10.1007/978-3-031-44336-7_18]

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: 6726-6733 [DOI: 10.1109/ICPR48806.2021.9412109]

Roth K, Pemula L, Zepeda J, Scholkopf 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. New Orleans, USA: IEEE: 14298-14308 [DOI: 10.1109/CVPR52688.2022.01392]

Salehi M, Sadjadi N, Baselizadeh S, Rohban M H and Rabiee H R. 2021. Multiresolution knowledge distillation for anomaly detection//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 14897-14907 [DOI: 10.1109/CVPR46437.2021.01466]

Sato J, Suzuki Y, Wataya T, Nishigaki D, Kita K, Yamagata K, et al. 2023. Anatomy-aware self-supervised learning for anomaly detection in chest radiographs. iScience, 26(7): #107086 [DOI: 10.1016/j.isci.2023.107086]

Schlegl T, Seeböck P, Waldstein S M, Langs G and Schmidt-Erfurth U. 2019. f-AnoGAN: fast unsupervised anomaly detection with generative adversarial networks. Medical Image Analysis, 54: 30-44 [DOI: 10.1016/j.media.2019.01.010]

Schlegl T, Seeböck P, Waldstein S M, Schmidt-Erfurth U and Langs G. 2017. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery//Proceedings of the 25th International Conference on Information Processing in Medical Imaging. Boone, USA: Springer: 146-157 [DOI: 10.1007/978-3-319-59050-9_12]

Tan J, Hou B, Batten J, Qiu H Q and Kainz B. 2020. Detecting outliers with foreign patch interpolation [EB/OL]. [2025-11-23]. https://arxiv.org/pdf/2011.04197.pdf

Tan J, Hou B, Day T, Simpson J, Rueckert D and Kainz B. 2021. Detecting outliers with Poisson image interpolation//Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention. Strasbourg, France: Springer: 581-591 [DOI: 10.1007/978-3-030-87240-3_56]

Tang P, Yan X X, Hu X B, Wu K, Lasser T and Shi K Y. 2025. Anomaly detection in medical images using encoder-attention-2Decoders reconstruction. IEEE Transactions on Medical Imaging, 44(8): 3370-3382 [DOI: 10.1109/TMI.2025.3563482]

Wang G D, Han S M, Ding E R and Huang D. 2021. Student-teacher feature pyramid matching for anomaly detection [EB/OL]. [2025-11-23]. https://arxiv.org/pdf/2103.04257.pdf

Wang Y, Zhou J G, Yan J and Guan J H. 2025. Survey of anomaly detection methods in surveillance videos based on deep learning. Journal of Image and Graphics, 30(3): 615-640 (汪洋, 周脚根, 严俊, 关佶红. 2025. 基于深度学习的监控视频异常检测方法综述. 中国图象图形学报, 30(3): 615-640) [DOI: 10.11834/jig.240329]

Yang X W, Latecki L J and Pokrajac D. 2009. Outlier detection with globally optimal exemplar-based GMM//Proceedings of 2009 SIAM International Conference on Data Mining. Sparks, USA: Society for Industrial and Applied Mathematics: 145-154 [DOI: 10.1137/1.9781611972795.13]

You Z Y, Cui L, Shen Y J, Yang K, Lu X, Zheng Y, et al. 2022a. A unified model for multi-class anomaly detection//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc.: #330

You Z Y, Yang K, Luo W H, Cui L, Zheng Y and Le X Y. 2022b. ADTR: anomaly detection transformer with feature reconstruction//Proceedings of the 29th International Conference on Neural Information Processing. Cham, Switzerland: Springer: 298-310 [DOI: 10.1007/978-3-031-30111-7_26]

Zhang H B, Guo W P, Zhang S Q, Lu H S and Zhao X M. 2022. Unsupervised deep anomaly detection for medical images using an improved adversarial autoencoder. Journal of Digital Imaging, 35(2): 153-161 [DOI: 10.1007/s10278-021-00558-8]

Zhao H, Li Y X, He N J, Ma K, Fang L Y, Li H Q, et al. 2021. Anomaly detection for medical images using self-supervised and translation-consistent features. IEEE Transactions on Medical Imaging, 40(12): 3641-3651 [DOI: 10.1109/TMI.2021.3093883]

Zhao Y Z, Ding Q Q and Zhang X Q. 2023. AE-FLOW: autoencoders with normalizing flows for medical images anomaly detection//Proceedings of the 11th International Conference on Learning Representations. Kigali, Republic of Rwanda: #ICLR

Zhou K, Li J, Luo W X, Li Z X, Yang J L, Fu H Z, et al. 2022a. Proxy-bridged image reconstruction network for anomaly detection in medical images. IEEE Transactions on Medical Imaging, 41(3): 582-594 [DOI: 10.1109/TMI.2021.3118223]

Zhou K, Li J, Xiao Y T, Yang J L, Cheng J, Liu W, et al. 2022b. Memorizing structure-texture correspondence for image anomaly detection. IEEE Transactions on Neural Networks and Learning Systems, 33(6): 2335-2349 [DOI: 10.1109/TNNLS.2021.3101403]

作者简介

赵映程,男,讲师,主要研究方向为计算机视觉和医学影像处理。E-mail: zyc@xaut.edu.cn

石争浩,通信作者,男,教授,主要研究方向为机器视觉、医学图像处理及机器学习。E-mail: 362743337@qq.com

朱凝,男,硕士研究生,主要研究方向为计算机视觉。E-mail: 2251221121@stu.xaut.edu.cn

宋霄罡,男,教授,主要研究方向为计算机视觉和无人系统自主导航。E-mail: songxg@xaut.edu.cn

黑新宏,男,教授,主要研究方向为智能系统与安全关键系统、人工智能、大数据及其在轨道交通等系统中的应用。E-mail: heixinhong@xaut.edu.cn