

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2024)09-2825-12

论文引用格式: Zhang J X, Gao C Y, Sun J, Ding X Y and Liu B. 2024. Whole slide pathological image classification of breast cancer based on mixed supervision learning. Journal of Image and Graphics, 29(09):2825-2836(张建新, 高程阳, 孙鉴, 丁雪妍, 刘斌. 2024. 混合监督学习的乳腺癌全切片病理图像分类. 中国图象图形学报, 29(09):2825-2836)[DOI:10.11834/jig.230343]

混合监督学习的乳腺癌全切片病理图像分类

张建新¹, 高程阳¹, 孙鉴¹, 丁雪妍^{1*}, 刘斌²

1. 大连民族大学计算机科学与工程学院, 大连 116600; 2. 大连理工大学国际信息与软件学院, 大连 116620

摘要: 目的 自监督与弱监督学习是解决乳腺癌全切片病理图像分类标注困难的有效方式。然而, 由于组织病理图像的复杂性与多样性, 仅依靠自监督学习生成的伪标签可能无法准确反映图像真实类别信息; 同时, 单一弱监督学习方法又存在标签信息匮乏等问题, 在病理图像学习过程中易受干扰而导致预测结果不稳定。为此, 提出了一种混合监督学习的乳腺癌全切片病理图像分类方法。方法 首先, 使用基于 MoBY 自监督框架进行训练, 通过对比学习方式深入挖掘乳腺癌病理图像内在结构信息; 然后, 采用弱监督多示例学习方法进一步优化自监督模型, 来获得更精准的判别示例; 最后, 从每幅全切片中筛选出具有代表性的乳腺癌病理图像关键示例, 并借助 Transformer 编码器实现关键示例的特征融合以增强不同病理图像块之间的关联性, 从而实现乳腺癌全切片病理图像的高精度分类。结果 在公开的 Camelyon-16 乳腺癌病理图像数据集上进行实验评估, 相比于该数据集上既有最优弱监督和自监督方法, 本文方法的曲线下面积值分别可提升 2.34% 和 2.74%, 验证了所提出混合监督学习方法的有效性。此外, 在 MSK (Memorial Sloan-Kettering) 腺癌病理外部验证数据集上较有监督方法取得了 6.26% 的性能提升, 表明了本文方法的良好泛化能力。结论 提出了混合监督学习的乳腺癌全切片病理图像分类方法, 通过集成 MoBY 自监督对比学习与 Transformer 弱监督多示例学习, 实现了乳腺癌全切片病理图像的更准确分类。

关键词: 乳腺癌全切片病理图像; 分类; 混合监督学习; 特征融合; Transformer

Whole slide pathological image classification of breast cancer based on mixed supervision learning

Zhang Jianxin¹, Gao Chengyang¹, Sun Jian¹, Ding Xueyan^{1*}, Liu Bin²

1. School of Computer Science and Engineering, Dalian Minzu University, Dalian 116600, China;

2. International School of Information Science and Engineering, Dalian University of Technology, Dalian 116620, China

Abstract: **Objective** Breast cancer belongs to the most common malignant tumors among women, and its early diagnosis and accurate classification bear great importance. Breast cancer whole slide pathological images serve as important auxiliary diagnostic means, and their classification can assist doctors in the accurate identification of tumor types. However, given the complexity and huge data volume of breast cancer whole slide pathological images, manual annotation of the label of each image becomes time consuming and labor intensive. Therefore, researchers have proposed various automated meth-

收稿日期: 2023-06-21; 修回日期: 2023-12-23; 预印本日期: 2024-01-04

* 通信作者: 丁雪妍 dingxueyan@dlmu.edu.cn

基金项目: 国家自然科学基金项目(61972062); 辽宁省基础研究项目(2023JH2/101300191); 辽宁省博士科研启动基金项目(2023-BS-078); 中央高校基本科研业务费专项资金项目(04442024042)

Supported by: National Natural Science Foundation of China(61972062); Basic Research Project of Liaoning Province(2023JH2/101300191); Doctoral Research Startup Fund of Liaoning Province(2023-BS-078); Fundamental Research Funds for the Central Universities(04442024042)

ods to address the issue encountered in the classification of breast cancer whole slide pathological images. Self- and weakly supervised learning effectively tackling the challenge of breast cancer whole slide pathological image classification. Self-supervised learning is a type of machine learning method that skips the manual annotation of labels. This method design tasks that enable the model to learn feature representations from unlabeled data. Self-supervised learning has achieved remarkable progress in the field of computer vision, but it still faces certain challenges in breast cancer whole slide pathological image classification. Given the complexity and diversity of pathological images, relying solely on the pseudo labels generated by self-supervised learning may fail to accurately reflect the true classification information, which affects the classification performance. On the other hand, weakly supervised learning leverages information from unlabeled image data through various methods, such as multiple instance learning or label propagation. However, the associated models encounter challenges, such as limited label information and noise, which affect the model's stability during the learning process and thus the stability of prediction results. To overcome the limitations of self- and weakly supervised learning, this paper proposes a mixed supervised learning method for breast cancer whole slide pathological image classification. The integration of MoBY self-supervised contrastive learning with weakly supervised multi-instance learning combines the advantages of these learning architectures and makes full use of unlabeled and noisy labeled data. In addition, such combination improves the classification performance through feature selection and spatial correlation enhancement, which results in increased robustness. **Method** First, the self-supervised MoBY was used to train the model on unlabeled pathological image data. MoBY, can learn key feature representations from images, is a self-supervised learning method based on self-reconstruction and contrastive learning. This process enables the model to extract useful feature information from unlabeled data and provide better initialization parameters for subsequent classification tasks. Then, a weakly supervised learning approach based on multiple instance learning was used for further model optimization. Multiple instance learning utilizes information from unlabeled image data for model training. In breast cancer whole slide pathological image classification, the accurate annotation of each image category often presents a challenge. This type of learning divides images into positive and negative instances based on instance-level labels to train the model. This approach partially contributes to solving the problem of limited label information and improves a model's robustness and generalization capability. For the feature selection stage, representative feature vectors were selected from each whole slide image to reduce redundancy and noise, extract the most informative features, and improve the model's focus and discriminative capability toward key regions. In addition, the paper leverages a Transformer encoder to improve the correlation among various image patches. The Transformer encoder is a powerful tool for modeling global contextual information in images, and it captures semantic relationships between different regions of an image to further increase the classification accuracy. The introduction of the Transformer encoder into breast cancer whole slide pathological image classification enables the improved utilization of global image information and further understanding of a model's image structure and context. Comprehensive application of methods, such as self- and weakly supervised learning, resulted in the high accuracy and robustness of the proposed mixed supervised learning approach for the classification of breast cancer whole slide pathological images in this paper. In experiments, this method achieved excellent classification results on a dataset of breast cancer whole slide pathological images. This approach serves as a powerful tool and technical support for the early diagnosis and accurate classification of breast cancer. **Result** The effectiveness of the mixed supervised model was validated through evaluation experiments conducted on the publicly available Camelyon-16 breast-cancer pathological image dataset. Compared with the state-of-the-art weakly and self-supervised models of this dataset, the proposed model achieved evident improvements of 2.34% and 2.74% in the area under the receiver operating characteristic, respectively. This finding indicates that the proposed method outperformed the other models in terms of breast cancer whole slide pathological image classification tasks. To further validate its generalization capability, we performed experiments on an external validation dataset of MSK. The proposed model for this validation dataset demonstrated a great performance improvement of 6.26%, which further confirms its strong generalization capability and practicality. **Conclusion** The proposed breast cancer whole slide pathological image classification method based on mixed supervision achieved remarkable results in addressing the related challenge. By leveraging the advantages of self-supervised learning, weakly supervised learning, and spatial correlation enhancement, the given model demonstrated improved classification performance on public and external validation datasets. This method exhibits a good

generalization capability and offers a viable solution for the early diagnosis and treatment of breast cancer. Future research should further refine and optimize the proposed method to increase its accuracy and robustness in breast cancer whole slide pathology image classification. This paper will address the challenges in breast cancer pathological image classification and contribute to the development of early breast cancer diagnosis and treatment.

Key words: breast cancer whole slide pathology image; classification; mixed supervised learning; feature fusion; Trans-former

0 引言

乳腺癌作为一种常见恶性肿瘤,严重威胁到女性生活与健康(Sung等,2021)。临床显示,早期诊断与治疗对提高乳腺癌患者生存率至关重要(Anderson等,2021)。然而,传统病理图像诊断效率相对较低,需经验丰富的医生在高倍显微镜下仔细观察(Wang等,2020)。全切片数字病理图像(whole slide pathology image, WSI)的出现极大提高了病理诊断效率,但仍存在对医生专业水平要求较高和准确率待提升等问题。随着乳腺癌发病数量的不断上升,病理图像的高效且高精度自动识别对于提高患者生存率具有更重要的现实意义(Cheng等,2016;金旭等,2020)。近年来,深度学习在医学图像分析领域取得了巨大成功(田娟秀等,2018;赵樱莉等,2023),研究者也尝试将其引入到WSI病理图像辅助诊断中。特别地,为解决WSI病理图像超大分辨率(十万以上像素)带来的巨大计算量和精细标注成本等问题,研究者引入了具有更低标注代价的弱监督和自监督深度学习方法,有效适配了该医学任务并逐渐受到更多关注。

弱监督多示例学习(multiple instance learning, MIL)方法将WSI病理图像切分的局部病理图像块作为单个示例,将整幅WSI病理图像作为示例的包,较好地适应于病理图像的超大像素尺寸和大规模图像级标注数据特点。根据训练方式,目前基于深度多示例学习的病理图像分类方法主要可分为示例级和嵌入级方法。示例级方法通过训练一个鲁棒的示例分类器来预测包中每个示例的潜在标签,并利用多示例池化来综合包中示例预测结果,从而实现对WSI病理图像的有效分类(Wang等,2018;Yan等,2018)。示例级方法虽然能够直接对示例进行分类并找到关键示例,但由于关键示例数量限制,会因伪标签噪声导致示例分类器训练不稳定而限制方法性

能,为获得更优分类器,则需使用大量病理图像数据更新迭代关键示例和分类器(Campanella等,2019)。此外,考虑到基于池化的聚合方法缺乏灵活性,研究者也尝试结合嵌入级训练方法融合示例来提高分类性能(Campanella等,2019;Qu等,2022)。嵌入级方法将包中的全部示例映射为固定长度特征向量,通过聚合特征向量得到包级别特征表示,进而采用有监督方式训练包级别分类器来实现包级预测。嵌入级方法通常更加关注示例特征融合方式,目前病理图像分类的嵌入级聚合方式主要包括注意力聚合(Lu等,2021;Su等,2023)、图卷积神经网络聚合(Guan等,2022)和基于Transformer的聚合(Shao等,2021;Chikontwe等,2022),从不同角度改进了传统池化方法无法自适应训练的弊端,有效提高了病理图像的分类精度。然而,由于嵌入级方法更加关注包级分类器的训练,在特征提取阶段则直接使用ImageNet预训练模型,忽视了自然图像与病理图像的差异;此外,特征融合的聚集函数也缺少一定灵活性,难以准确显示不同示例对包级分类器的贡献(Lu等,2021;Shao等,2021)。

另外,受近年自监督学习在自然图像领域取得巨大成功的启发,研究者也开始关注利用未标记数据自身生成的监督信号来学习有价值的病理图像特征表达,以此应对缺少精细标注的病理图像分类任务。目前,该领域已有自监督学习方法主要集中在自监督特征提取器的引入和优化(Chen和Krishnan,2022;Wang等,2022)、以及辅助任务的设计(Koohbanani等,2021;Srinidhi等,2022)等方面。Chen和Krishnan(2022)通过训练多种自监督模型,验证其在弱监督和补丁级任务中的有效性,以寻找适用于病理学的有效表示方法。Srinidhi等人(2022)通过构建分辨率序列预测的辅助任务,利用不同分辨率序列的组合来训练网络,使得模型能够学习不同分辨率序列之间的关系,并从中提取出具有价值的病理图像特征。然而,在乳腺癌全切片病理图像中存

在阳性样本相对稀缺问题,这使得传统的自监督算法更加偏向处理阴性样本,而对阳性样本的学习效果较差。因此,仅依赖自监督学习方法生成的监督信号可能无法准确地反映真实的类别信息。此外,为获得更精准的特征提取器,在下游任务中常需要使用部分有标签的数据进行微调,但这些数据仍需要通过人工标注的方式获取,又增加了额外的标注成本。

针对以上问题,本文提出一种基于混合监督的多示例乳腺癌全切片病理图像分类方法(mixed supervised MIL, MSMIL)。通过融合 MoBY 监督对比学习与 Transformer 弱监督多示例学习,从而实现乳腺癌全切片病理图像的高精度分类。具体来说,首先利用 MoBY 自监督框架(Xie 等, 2021)从大量未标记的 WSI 乳腺癌病理图像示例中学习语义丰富的特征表达,为病理分析任务提供先验知识,同时也降低后续弱监督方法所产生的伪标签噪声的影响。然后,采用基于 MIL 的弱监督方法筛选出关键示例并赋予相应的包级伪标签,通过有监督的微调进一步提高分类器的准确性。在特征融合阶段,采用视觉 Transformer(vision Transformer, ViT)(Dosovitskiy 等, 2021)聚集示例特征,以更好地捕捉示例之间的全局关系和语义信息。与已有的基于 Transformer 特征融合的方法相比,本研究仅选择关键示例进行特征融

合,通过学习和利用这些关键示例的特征,增强了对阳性样本的关注,减少了因类别不平衡产生的影响,进而提高对乳腺癌病理图像不同类别的分类能力。

本文的主要贡献如下:1)通过集成 MoBY 自监督对比学习与 Transformer 弱监督多示例学习,提出一种基于混合监督学习的乳腺癌全切片病理图像分类新方法。2)将基于 MoBY 的自监督框架与基于 MIL 的弱监督框架相结合,建立更为高效准确的特征提取器机制,以充分利用 WSI 乳腺癌病理图像数据并选出关键示例;借助 Transformer 编码器对关键示例进行融合,以加强不同示例之间的相关性,完成 WSI 乳腺癌病理图像的分类。同时,Transformer 中注意力机制的使用在特征融合过程中能够提供更好的可解释性和可视化能力,帮助理解模型的决策过程和注意力分布。3)在公开的 Camelyon-16 和 MSK (Memorial Sloan-Kettering)数据集上进行了大量实验,以评估所提出方法的有效性。实验结果表明,本文方法在 WSI 乳腺癌病理图像分类任务上具有较高的准确性,与既有弱监督和自监督学习方法相比体现出较好的竞争力。

1 基于混合监督学习的分类方法

如图 1 所示,MSMIL 模型主要分为 MoBY 自监

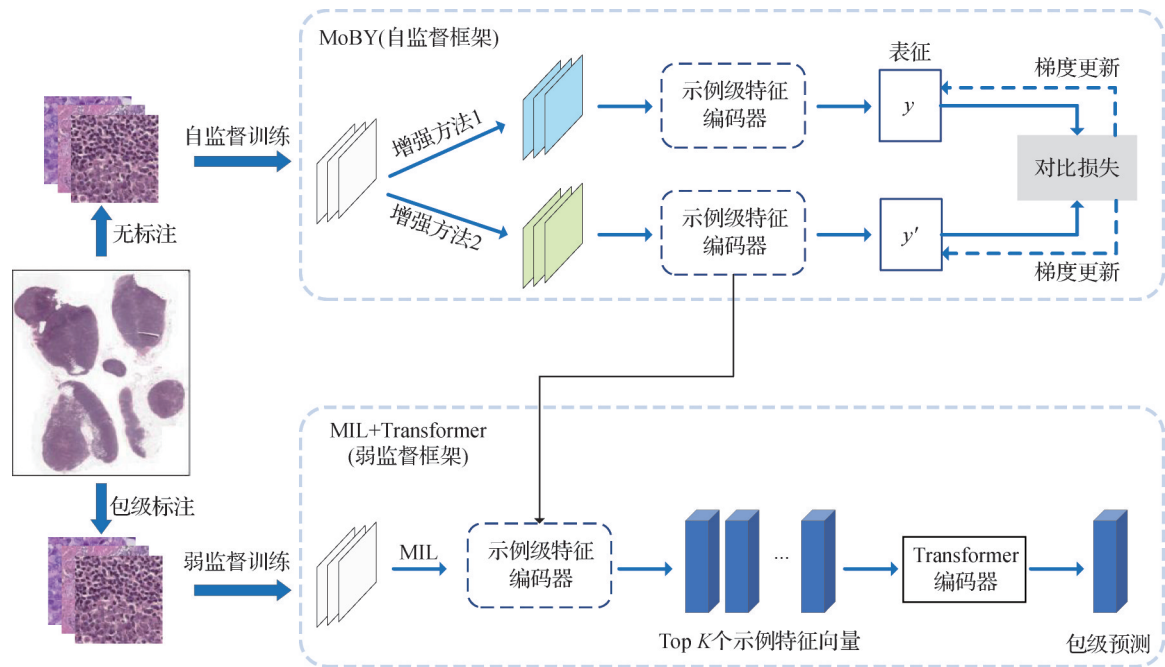


图1 MSMIL模型的整体架构

Fig. 1 The overall architecture of the MSMIL model

督对比学习和基于Transformer的弱监督多示例学习两个阶段。在第1阶段,将WSI切分的全部图像块输入MoBY自监督框架中,通过对比学习的方式学习不同乳腺癌病理图像之间的差异,从而提取具有判别性的形态结构信息。在第2阶段,采用MIL方法筛选WSI中的关键示例,对具备一定病理知识的编码器进一步优化,以提取更具区分性与泛化性的病理图像块示例特征。最后通过模型筛选出 K 个具有代表性的示例特征,输入Transformer编码器中进行特征融合,实现乳腺癌病理图像包级分类。

1.1 基于MoBY的自监督框架

由于病理图像与自然图像在独特的特征和上下文信息上具有较大差异,直接使用ImageNet预训练模型可能无法充分捕捉到病理图像的关键信息。因此,为了更有效地提取病理图像中的特征信息,并考虑到乳腺癌病理图像数据集通常缺乏精细标注,本文引入了MoBY自监督框架,该框架采用对比学习策略利用病理图像的自身特征进行训练,提高了特征提取的能力并减少了对大量标注数据的依赖,为后续病理图像的预测奠定基础。该框架综合了MoCov2(momentum contrast)(Chen等,2020)版本的动量设计、特征队列和对比损失,以及BYOL(bootstrap your own latent)(Grill等,2020)的非对称编码器、非对称数据扩充和动量调度器的设计思想,降低计算成本的同时,提供了更强的特征提取能力,以更高效的方式完成自监督训练。

MoBY具体流程图如图2所示。通过比较样本间的相似性和差异性,使学习到的特征能够更好地区分不同类别样本。该框架主要由以下4个部分组成:

1)数据扩充部分。将输入图像应用两组不同的数据增强方法来生成一对同类样本视图 v 和 v' ,其中第1组数据增强包括随机中心裁剪、随机水平翻转、色彩抖动、随机灰度化、高斯模糊和归一化操作;第2组数据增强在第1组的基础上添加了曝光操作,这对视图被视为具有相同类别标签的正样本对。与这对正样本相比,其他图像被视为负样本。针对病理医生、设备和环境等因素导致病理图像染色差异,本文加强了色彩抖动、灰度化和曝光等和颜色相关数据增强方法的概率,来消除明显颜色差异导致的错误分类。

2)编码器部分。采用Swin Transformer模型(Liu等,2021)作为编码器从生成视图中提取特征向量。MoBY模型由在线编码器和目标编码器组成,在线编码器负责处理当前输入图像并生成相应的特征表示,通过梯度更新来进行优化。目标编码器则生成一组固定的特征表示,其更新采用渐进式动量更新策略,以提供更稳定的特征学习信号,更新策略计算为

$$\theta_k \leftarrow m\theta_k + (1-m)\theta_q \quad (1)$$

式中, θ_k 表示目标编码器的参数, θ_q 表示在线编码器的参数, m 为动量系数,默认初始值为0.99,在训练过程中,动量的值逐渐增加到1。

3)在编码器之后,引入了一个具有两层多层感知机(multilayer perceptron, MLP)的映射器,用以将输入图像转换到一个丰富的潜在特征空间,以提取更深层次的语义内容。同时在目标编码器中,将生成的特征向量添加到特征队列中,用于后续的对比计算。此外,为了增强特征的预测能力,在线编码器中引入了额外预测器,用于对潜在特征空间内的向

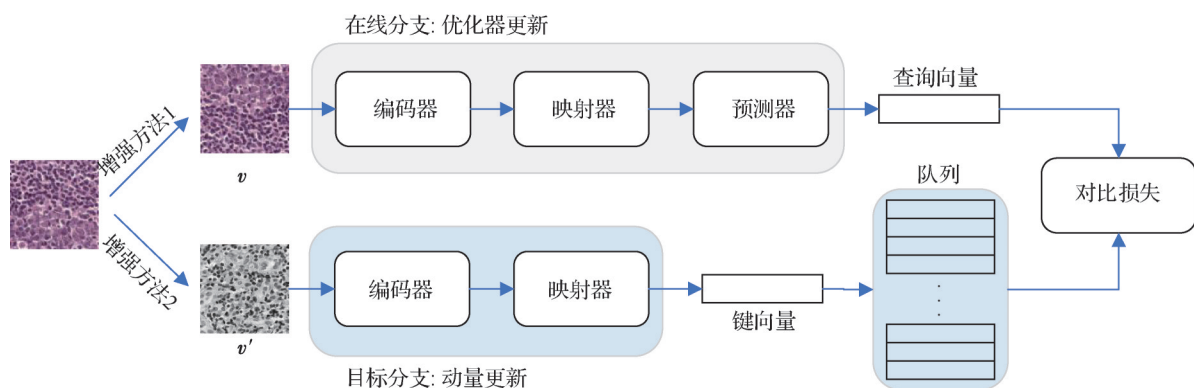


图2 MoBY自监督框架

Fig. 2 Self supervision framework of the MoBY

量进行预测,其也采用两层 MLP 结构,旨在模拟目标映射器的潜在特征表示,以此来最小化在线网络与目标网络输出间的表征差异。

4) 计算对比损失。在线编码器生成的特征向量与来自特征队列的样本进行对比,通过计算正样本对之间的相似性和负样本对之间的差异性,得到对比损失值,进而实现反向传播,更新在线编码器的参数,使得模型学习到更具代表性和判别性的图像特征表示。对于在线视图 q ,对比损失计算为

$$L_q = -\log \frac{\exp(\frac{q \times k_o}{\gamma})}{\sum_{i=0}^N \exp(\frac{q \times k_i}{\gamma})} \quad (2)$$

式中, k_o 表示来自同一图像的其他视图的目标特征; k_i 是关键字队列中的目标特征; γ 是温度系数; N 是关键字队列的大小(默认为 4 096)。通过 MoBY 的对比训练可获得性能较佳的特征提取器,用于后续的特征提取任务。

1.2 基于 MIL 与 Transformer 的弱监督学习

通过上述自监督方法充分利用无标签的乳腺癌病理图像数据,使得模型具备了较好的特征提取能力。为进一步提升 WSI 乳腺癌病理图像上的分类效果,使用基于 MIL 的弱监督方法,该方法利用仅有的 WSI 级标签对自监督模型进行优化,以实现关键示例的分类和筛选。最后,通过采用 Transformer 编码器对关键示例进行聚合,增强不同示例之间的相关性,实现 WSI 级乳腺癌病理图像的高精度分类。

1.2.1 示例选取

在病理图像中,通常只能获得整幅 WSI 的标签,而无法获得每个细胞或区域的具体标签。因此,无法直接将每个细胞或区域与其对应的标签进行匹配。为解决这一问题,MIL 提供了一种有效的解决方案,将整幅 WSI 视为一个包,每个包对应一个标签。将 WSI 包中切分出的小块作为包的示例,如果示例中存在阳性示例,则认为该幅 WSI 为阳性;反之,则为阴性。将全部 WSI 的标签用集合 Y 表示,其中 $Y = \{Y_1, Y_2, \dots, Y_n\}$,对应关系为

$$Y_i = \begin{cases} 1 & \exists y_{ij} = 1 \\ 0 & \forall y_{ij} \neq 1 \end{cases} \quad (3)$$

式中, Y_i 为第 i 幅 WSI 的标签, y_{ij} 为第 i 幅 WSI 中第 j 个示例的标签。

本文采用 Swin Transformer-T(Liu 等, 2021)作为

骨干网络,其独特之处在于引入了分层注意力机制和窗口化的注意力机制。分层结构使其能够同时捕捉全局和局部乳腺癌病理图像信息,拓展了编码器的感知范围和表征能力;窗口化的注意力机制使 Swin Transformer 在保持全局感知的同时,能精准捕捉乳腺癌病理图像的局部细节信息,从而提高编码的准确性。此外, Swin Transformer 引入了可变形注意力机制,进一步提升了模型的感知能力和定位能力。乳腺癌病理图像中的细胞核形态和分布具有一定的变异性,传统的固定感受野可能无法充分捕捉这种变异性。可变形注意力机制通过自适应地对不同位置进行建模,能够更好地适应图像中的变化,提高对细胞核的感知和定位能力,提升了模型的鲁棒性。使用 Swin Transformer-T 的官方默认参数进行实验。

在多示例学习的实施中,首先将包内的所有实例输入经过自监督学习模型训练的编码器,利用线性评估得到每个实例的预测概率。根据这些预测概率进行排序,并筛选出预测概率最高的前 K 个实例作为关键实例,同时为其分配包级别的伪标签。接下来,利用这些关键示例及其伪标签对编码器进行迭代优化,目的是改善模型对实例的识别能力。最后将获取到的关键示例用于后续的包级预测任务。其中,超参数 K 的值设定为 15。

1.2.2 包级预测

包级预测部分旨在将前一阶段筛选的关键示例进行特征融合,从而获得 WSI 级标签。然而,传统的基于池化的方法(Wang 等, 2018; Yan 等, 2018)以及基于循环神经网络(recurrent neural network, RNN)的方法(Vaswani 等, 2017)通常不够灵活,并且在处理整个图像的全局信息以及长距离数据依赖关系方面具有局限性。因此,为了确保模型具备良好的灵活性和全局感知能力,本文采用 Transformer 编码器作为包级预测的核心结构。

包级预测流程如图 3 所示。具体来讲,为了保留序列中关键示例的位置信息,本文为每个关键示例学习一个特定的位置编码,并将其与所得到的 K 个特征向量一同输入到编码器中,以精确捕捉序列的整体结构和依赖关系。编码器由多头自注意力块(multi-head self-attention, MSA)和多层 MLP 交替组合而成,在模块前后分别引入了层归一化(layer norm, LN)和残差连接,以提高模型训练的稳定性和收敛速度。MLP 层由两个全连接层组成,其中引入

了 GELU (Gaussian error linear units) 激活函数增强非线性表达能力。自注意力 (self-attention, SA) 通过动态生成权重来捕捉长距离的关联性。其计算可分解为以下几个步骤: 首先, 将输入特征向量 X 进行线性投影, 通过与 3 个不同的权重矩阵 W_q 、 W_k 和 W_v 进行点积运算, 得到相应的查询向量 Q 、键向量 K 和值向量 V , 计算式为

$$\begin{aligned} Q &= W_q X \\ K &= W_k X \\ V &= W_v X \end{aligned} \quad (4)$$

式中, 查询向量 Q 用于度量输入序列中的每个元素与其他元素的相关性, 而键向量 K 和值向量 V 则提供了计算注意力的权重信息。之后, 通过点积操作计算 Q 和 V 之间的相关性并应用 softmax 操作。为了确保梯度的稳定性, 将注意力权重与对应的 V 相乘,

得到输入向量的自注意力得分。具体计算式为

$$SA(Q, K, V) = f_{\text{softmax}} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (5)$$

式中, d_k 是 Q 和 K 的维度, 通过将乘积结果除以 $\sqrt{d_k}$ 进行归一化, 避免输入维度过长时, 乘积结果太大影响梯度传播。多头自注意力则是将输入向量传递给 h 个不同的自注意力单元, 分别获取输入特征向量中的关键信息, 最后将每个自注意力单元计算得到的向量拼接在一起, 投影到线性层中, 得到最终的结果为

$$f_{\text{MultiHead}}(Q, K, V) = f_{\text{Concat}}(SA_1, \dots, SA_h) W_o \quad (6)$$

式中, W_o 是投影矩阵, h 表示特征空间的个数。此外, 鉴于本文中仅对选取到的关键示例进行融合, 因此在包级预测阶段仅采用一层 Transformer 编码模块, 其中隐藏层大小设为 768, 注意力头数量为 16。

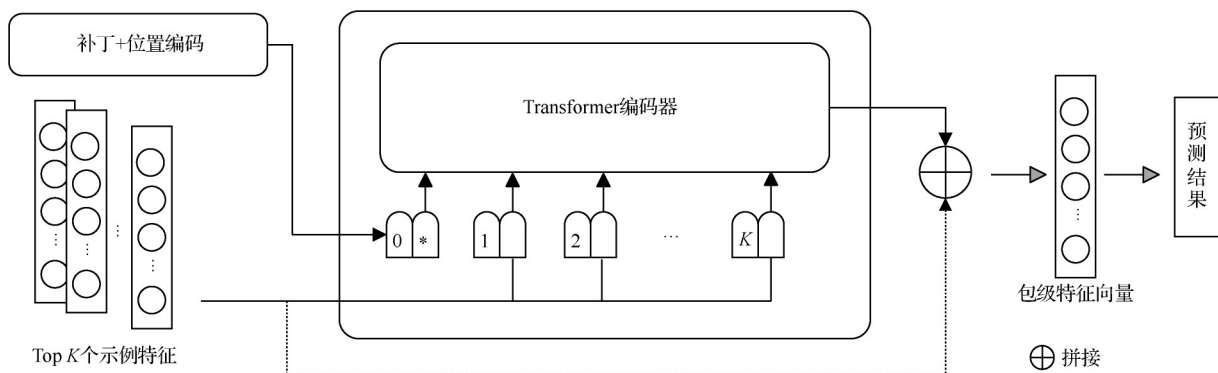


图3 特征融合阶段

Fig. 3 Feature fusion stage

2 实验与结果分析

2.1 数据集与实验设定

本文使用 Camelyon-16 公开数据集 (Bejnordi 等, 2017) 来评估所提出的 MSMIL 模型。该数据集包含 399 幅苏木精和伊红 (Hematoxylin and Eosin, H&E) 染色的乳腺癌前哨淋巴结全切片图像, 用于乳腺癌转移检测, 其中 159 幅图像为转移阳性, 其余 240 幅为阴性。官方指定了其中的 270 幅作为训练集, 129 幅为测试集。为了验证本文提出的 MSMIL 模型的有效性, 仅使用 WSI 级标签进行实验。

为了进一步验证本模型的泛化能力, 在 Campanella 等人 (2019) 公开的 MSK 腺癌转移数据集进行泛化性测试。该数据集包括来自 78 名患者腋窝

淋巴结样本的 130 幅 WSI, 使用苏木精和伊红染色进行组织区域标记。关于数据集划分的详细信息参见表 1。

表1 数据集划分描述

Table 1 The division of experimental datasets

数据集	训练集		测试集	
	正常	肿瘤	正常	肿瘤
Camelyon-16	159	111	80	49
MSK	—	—	94	36

注: “—”表示无相关数据。

本文在自监督训练过程中采用了 MoBY 官方的设定。在弱监督训练过程中使用了 PyTorch 框架 (Paszke 等, 2019), Adam 优化器, 初始学习率设为 0.000 1, 采用 64 的小批量迭代了 35 次。所有实验

均在一台配备有两张英伟达 GTX 3090 GPU 的服务器上进行。

2.2 数据预处理

完整的 WSI 病理图像中存在大量的背景区域, 该区域对于模型训练来说是不必要的。因此, 为提高模型性能并减少计算资源的消耗, 本文通过高斯模糊变换与反二进制阈值化操作对背景区域进行过滤, 并保留组织区域占据图像面积 40% 以上的部分, 以突出关键组织区域, 并从该区域中生成若干个尺寸为 512×512 像素的不重叠的图像切片用于模型训练。分割过程如图 4 所示。

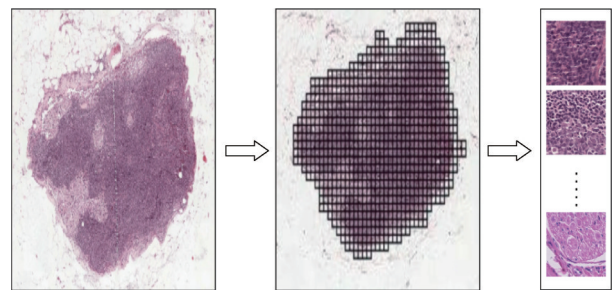


图 4 WSI 分割过程
Fig. 4 WSI segmentation process

2.3 消融实验

为评估 MSMIL 方法中不同方法的贡献, 在 Camelyon-16 数据集 40 倍率下进行消融实验, 并给出每一阶段的准确率 (accurate, ACC) 和曲线下面积 (area under curve, AUC) 等指标。其中 ACC 用来衡量模型的准确性; AUC 则通过衡量模型在不同阈值下的真阳率与假阳率之间的权衡关系, 以全面评估模型的分类性能。

消融实验结果如表 2 所示, 方法 A 采用 MIL 弱监督学习方法进行特征提取, 利用循环神经网络 (RNN) 方法聚合关键示例特征, 并将其作为基准。实验结果显示, 该方法的 ACC 和 AUC 分别达到了 94.57% 和 94.44%。方法 B 对特征聚合部分进行了改进, 本文将 RNN 替换为 Transformer 模型, 其 ACC 与 AUC 分别上涨了 0.78% 与 1.02%。上述实验表明, Transformer 可以加强所选示例之间的相关性, 进一步提升模型的效果。此外, 方法 C 使用了 MoBY 自监督框架提取示例特征, 在不使用任何标签的情况下, 模型仍能达到较高的精度, 但相比 MIL 弱监督方法结果略有不足。最后方法 D (本文方法) 将自监督方法与弱监督方法相结合, 并引入 Transformer 方

法完成特征聚合。实验结果表明, 与其他方法相比, 本文在精度方面取得了最佳模型表现。与基准模型相比, 本文方法在 ACC 和 AUC 方面分别提高了 3.1% 和 2.5%。这些结果验证了本文方法的有效性。

表 2 模型在 Camelyon-16 测试集上的消融实验结果
Table 2 The ablation experimental results of the model on the Camelyon-16 test set

方法	示例选取	包级预测	ACC/%	AUC/%
A	MIL	RNN	94.57	94.44
B	MIL	Transformer	95.35	95.46
C	MoBY	Transformer	93.02	91.61
D	MoBY+MIL	Transformer	97.67	96.94

注: 加粗字体表示各列最优结果, 示例选取阶段均采用 Swin Transformer 为骨干网络。

此外, 为了更直观地观察模型分类错误的细节, 本文利用混淆矩阵对上述方法的分类结果进行了可视化, 如图 5 所示, 颜色越深数量越多, 数值为具体数量。仅使用弱监督方法或自监督方法时, 模型获取到的关键示例特征存在一定的局限性, 导致假阳性和假阴性的数量无法进一步降低。然而, 在采用本文提出的混合监督方法后, 模型的假阳性和假阴性明显减少, 其中假阳性的数量甚至为零。这一结果表明, 混合监督方法进一步捕获了病理图像深层次的纹理特征, 从而实现了更高的准确性, 使得判别示例更加准确。

2.4 与其他优秀方法比较

为了进一步证明本文模型在 WSI 分类任务上的强大竞争力, 本文与目前具有代表性的弱监督与自监督方法进行了比较, 结果如表 3 所示。

从表 3 可以看出, 在弱监督方法中, 相较于 Campanella 等人 (2019) 提出的采用基于 ResNet-34 的 MIL 弱监督学习方法进行特征提取并利用 RNN 实现特征融合的方法, 本文方法在 ACC 和 AUC 上分别取得了 6.2% 与 5% 的性能提升。Attention2Majority (Su 等, 2023) 方法通过将正常病理图像的表征与所有输入图像块的特征进行比较, 以推断出输入全切片病理图像中的关键示例。DTFD-MIL (Zhang 等, 2022) 方法通过引入伪包来扩大包的数量, 并在此基础上建立双层 MIL 框架来有效地利用固有特征。虽

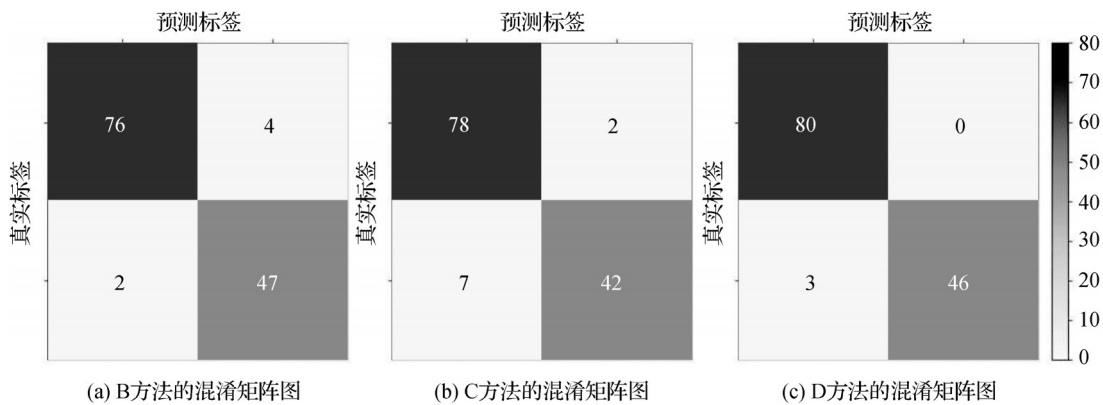


图 5 混淆矩阵结果图

Fig. 5 Confusion matrix result graph

((a) confusion matrix of method B; (b) confusion matrix of method C; (c) confusion matrix of method D)

表 3 Camelyon-16测试集的对比实验结果

Table 3 Comparison results of Camelyon-16 testset

方法	/%	
	ACC	AUC
MIL-RNN(Campanella 等, 2019)	91.47	91.94
Clam(Lu 等, 2021)	—	93.60
TransMIL(Shao 等, 2021)	88.37	93.09
DSMIL(Li 等, 2021)	89.92	91.65
Self-path(Koohbanani 等, 2021)	—	93.70
DTFD-MIL(Zhang 等, 2022)	90.80	94.60
Transpath+CLAM-SB(Wang 等, 2022)	92.20	94.20
Attention2Majority(Su 等, 2023)	86.50	89.10
CWC-Transformer(Wang 等, 2023)	89.14	93.85
MSMIL(本文)	97.67	96.94

注:加粗字体表示各列最优结果。“—”表示所引论文未提供相应实验结果。

然 Attention2Majority 和 DTFD-MIL 方法在关键区域的选取方面有所加强,但对病理图像数据的利用仍不够充分,与 MSMIL 相比,在 AUC 上仍分别存在 7.84% 和 2.34% 的差距。在 Clam(Lu 等, 2021)方法中,通过引入注意力机制来识别具有较高价值的区域进行融合。TransMIL(Shao 等, 2021)和 CWC-Transformer(Wang 等, 2023)等方法从加强不同示例之间的相关性着手,引入基于 Transformer 的方法加强对示例之间形态和空间信息的关注,不同的是,前者用来聚合形态信息,后者用来描述全局信息。尽管 Lu 等人(2021)、Shao 等人(2021)和 Wang 等人(2023)通过引入注意力的方式来更好地融合关键区

域,但其 AUC 仍比本文模型低 3.34%、3.85% 和 3.09%。此外,在自监督方法中,DSMIL(Li 等, 2021)引入对比学习来模拟示例间关系,并基于示例间的余弦距离计算出注意力分数来获取关键示例,然而,他们的 ACC 和 AUC 仍然低于本文模型 7.75% 和 5.29%。Self-path(Koohbanani 等, 2021)通过构建多个辅助任务以获取更丰富的病理特征,在仅使用 20% 的标签的情况下实现了 93.7% 的 AUC,但该过程仍需要人工标注的参与。Transpath(Wang 等, 2023)方法中引入了语义相关对比学习策略,来挖掘更多的阳性样本对,以加强组织病理图像样本的多样性,但实验结果表明,仅通过自监督方法无法实现更高的精度。

综上所述,与以上现有技术相比,本文提出的 MSMIL 模型在 WSI 乳腺癌病理图像分类任务上表现出显著的优势。该模型采用了自监督与弱监督相结合的策略,有效地利用了无标签数据,并成功地避免了人工标注的需求,从而进一步增强了对病理图像特征的提取能力;同时利用 Transformer 编码器从不同的病理补丁中捕捉特征从而获取到全局信息,更好地探索了不同示例之间的相关性,从而加强了特征的融合。这些改进使得模型在 WSI 乳腺癌病理图像分类任务中取得了显著的性能提升。

2.5 泛化性实验

为了评估 MSMIL 模型的泛化能力和通用性,本文选择了 MSK 数据集作为外部验证,进一步评估模型性能。该数据集提供了不同来源和特征的病理图像样本,可以更全面地验证本文模型在不同数据集上的性能表现。需注意的是,与 Camelyon-16 相比,

MSK数据集是在20倍率下进行扫描的。此外,MSK数据集与Camelyon-16数据集来自不同的机构,在染色、处理和扫描等操作方面存在一定差异。这导致MSK数据集的性能评估结果普遍略低于Camelyon-16数据集的评估结果。

评估结果如图6所示。左侧两个柱展示了Wang等人(2016)以有监督学习方式在Camelyon-16训练集上训练后,分别在Camelyon-16测试集和MSK数据集上的测试结果;右侧两个柱为本文模型在Camelyon-16训练集上训练后,分别在Camelyon-16测试集和MSK数据集上的测试结果。从图中可以观察到,在WSI乳腺癌病理图像分类任务中,相较于传统的监督模型,本文模型AUC评估指标分别提升了3.74%和6.26%。实验结果表明,本文模型不仅在本数据集上表现出良好的性能,且在外部测试集上也展现出较好的泛化性。

2.6 可视化结果

本文采用注意力机制直观地展示MSMIL模型在乳腺癌全切片图像中的关注区域,如图7所示。为了呈现不同示例的关注程度,对注意力权重进行了归一化处理,将其限定在0~1之间。其中明亮的区域对应具有较高贡献度的示例。观察可视化结果,可发现在预测过程中,模型将注意力主要集中在

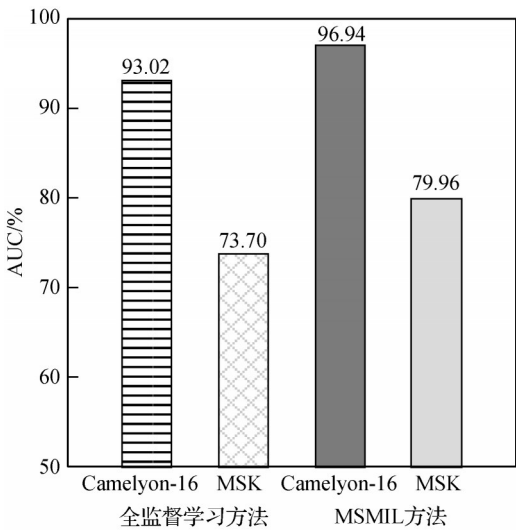


图6 外部验证实验结果的对比图
Fig. 6 Comparison of external verification experiment results

病变区域,进一步证明了模型具有准确的识别能力,并提供了一定的可解释性。

3 结 论

针对自监督与弱监督学习在乳腺癌全切片病理图像分类上面临的标签不足导致性能受限等问题,本文提出了一种基于混合监督的乳腺癌全切片病理

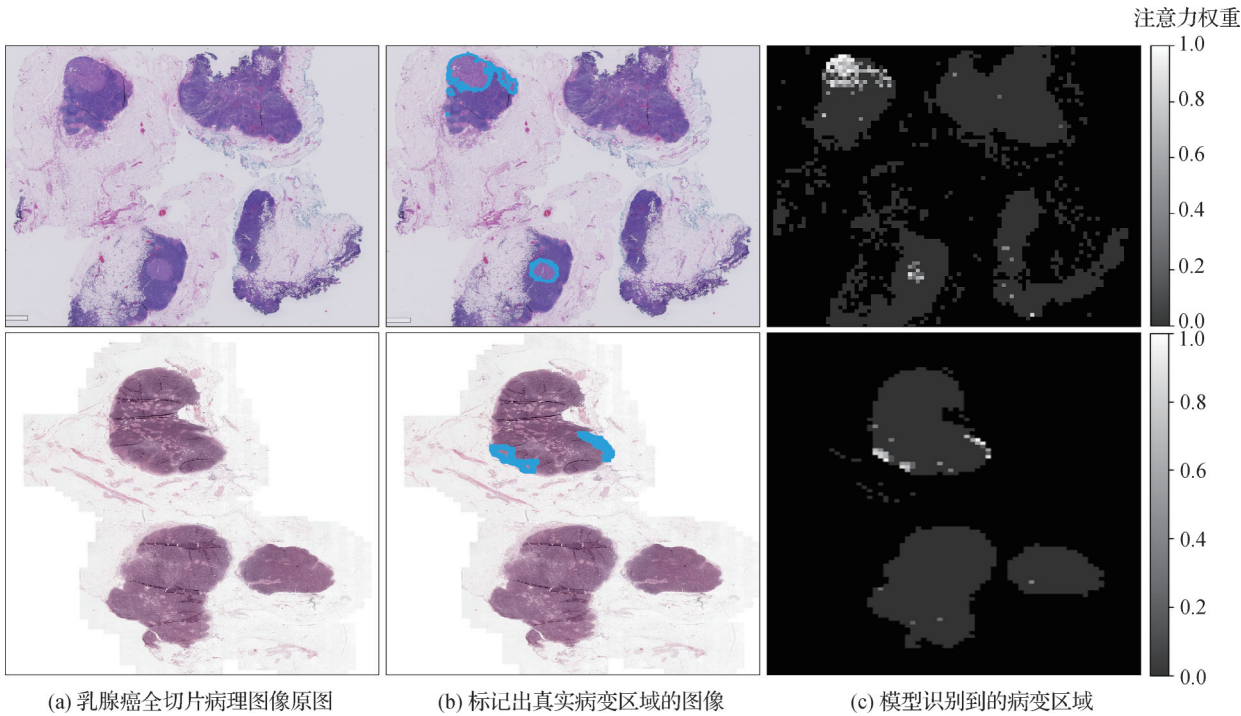


图7 模型可视化结果图

Fig. 7 The visualization results of the model ((a) original breast cancer whole slide pathology images; (b) images with actual lesion area marked; (c) the lesion area identified by the model)

图像分类方法。以 MoBY 自监督学习框架为基础,从未标记病理图像数据中学习特征表示,并结合弱监督多示例学习选择关键示例来优化模型,解决了弱监督学习方法中关键示例的伪标签噪声问题。同时,通过 Transformer 编码器实施关键示例特征融合,进一步实现了 WSI 级别乳腺癌病理图像的高精度分类。在两个公开的乳腺癌组织病理图像数据集上的实验结果证明了所提出混合监督学习方法的有效性,与既有弱监督和自监督学习方法相比体现出较好的竞争力。在未来研究中,将进一步探索混合监督学习方法的潜力,并结合更先进的自监督方法,通过设计更有效的混合策略和训练算法,以进一步提高乳腺癌全切片病理图像分类的性能。

参考文献 (References)

- Anderson B O, Ilbawi A M, Fidarova E, Weiderpass E, Stevens L, Abdel-Wahab M and Mikkelsen B. 2021. The global breast cancer initiative: a strategic collaboration to strengthen health care for non-communicable diseases. *The Lancet Oncology*, 22 (5): 578-581 [DOI: 10.1016/S1470-2045(21)00071-1]
- Bejnordi B E, Veta M, van Diest P J, van Ginneken B, Karssemeijer N, Litjens G, van der Laak J A W M, Hermesen M, Manson Q F, Balkenhol M, Geessink O, Stathonikos N, van Dijk M C, Bult P, Beca F, Beck A H, Wang D Y, Khosla A, Gargeya R, Irshad H, Zhong A X, Dou Q, Li Q Z, Chen H, Lin H J, Heng P A, Haß C, Bruni E, Wong Q, Halici U, Öner M Ü, Cetin-Atalay R, Berseth M, Khvatkov V, Vylegzhanin A, Kraus O, Shaban M, Rajpoot N, Awan R, Sirinukunwattana K, Qaiser T, Tsang Y W, Tellez D, Annuscheit J, Hufnagl P, Valkonen M, Kartasalo K, Latonen L, Ruusuvaari P, Liimatainen K, Albarqouni S, Mungal B, George A, Demirci S, Navab N, Watanabe S, Seno S, Takenaka Y, Matsuda H, Ahmady Phoulady H, Kovalev V, Kalinovskiy A, Liauchuk V, Bueno G, Fernandez-Carrobles M M, Serrano I, Deniz O, Racoceanu D and Venâncio R. 2017. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*, 318(22): 2199-2210 [DOI: 10.1001/jama.2017.14585]
- Campanella G, Hanna M G, Geneslaw L, Miraflor A, Silva V W K, Busam K J, Brogi E, Reuter V E, Klimstra D S and Fuchs T J. 2019. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine*, 25(8): 1301-1309 [DOI: 10.1038/s41591-019-0508-1]
- Chen R J and Krishnan R G. 2022. Self-supervised vision Transformers learn visual concepts in histopathology [EB/OL]. [2023-05-20]. <https://arxiv.org/pdf/2203.00585.pdf>
- Chen X L, Fan H Q, Girshick R B and He K M. 2020. Improved baselines with momentum contrastive learning [EB/OL]. [2023-05-20]. <https://arxiv.org/pdf/2003.04297.pdf>
- Cheng C L, Azhar R, Sng S H A, Chua Y Q, Hwang J S G, Chin J P F, Seah W K, Loke J C L, Ang R H L and Tan P H. 2016. Enabling digital pathology in the diagnostic setting: navigating through the implementation journey in an academic medical centre. *Journal of Clinical Pathology*, 69 (9): 784-792 [DOI: 10.1136/jclinpath-2015-203600]
- Chikontwe P, Nam S J, Go H, Kim M, Sung H J and Park S H. 2022. Feature re-calibration based multiple instance learning for whole slide image classification//Proceedings of the 25th International Conference on Medical Image Computing and Computer Assisted Intervention. Singapore, Singapore: Springer: 420-430 [DOI: 10.1007/978-3-031-16434-7_41]
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Housby N. 2021. An image is worth 16 × 16 words: Transformers for image recognition at scale [EB/OL]. [2023-05-20]. <https://arxiv.org/pdf/2010.11929.pdf>
- Grill J B, Strub F, Altché F, Tallec C, Richemond P H, Buchatskaya E, Doersch C, Pires B A, Guo Z D, Azar M G, Piot B, Kavukcuoglu K, Munos R and Valko M. 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.: 21271-21284
- Guan Y H, Zhang J, Tian K, Yang S, Dong P, Xiang J X, Yang W, Huang J Z, Zhang Y Y and Han X. 2022. Node-aligned graph convolutional network for whole slide image representation and classification//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 18813-18823 [DOI: 10.1109/CVPR52688.2022.01825]
- Jin X, Wen K, Lyu G F, Shi J, Chi M X, Wu Z and An H. 2020. Survey on the applications of deep learning to histopathology. *Journal of Image and Graphics*, 25(10): 1982-1993 (金旭, 文可, 吕国锋, 石军, 迟孟贤, 武铮, 安虹. 2020. 深度学习在组织病理学中的应用综述. *中国图象图形学报*, 25(10): 1982-1993) [DOI: 10.11834/jig.200460]
- Koohbanani N A, Unnikrishnan B, Khurram S A, Krishnaswamy P and Rajpoot N. 2021. Self-path: self-supervision for classification of pathology images with limited annotations. *IEEE Transactions on Medical Imaging*, 40(10): 2845-2856 [DOI: 10.1109/TMI.2021.3056023]
- Li B, Li Y and Eliceiri K W. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 14318-14328 [DOI: 10.1109/cvpr46437.2021.01409]
- Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, Lin S and Guo B N. 2021. Swin Transformer: hierarchical vision Transformer using shifted windows//Proceedings of 2021 IEEE/CVF International Con-

- ference on Computer Vision. Montreal, Canada: IEEE: 10012-10022. [DOI: 10.1109/ICCV48922.2021.00986]
- Lu M Y, Williamson D F K, Chen T Y, Chen R J, Barbieri M and Mahmood F. 2021. Data-efficient and weakly supervised computational pathology on whole slide images. *Nature Biomedical Engineering*, 5(6): 555-570 [DOI: 10.1038/s41551-020-00682-w]
- Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z M, Gimelshein N, Antiga L, Desmaison A, Köpf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J J and Chintala S. 2019. PyTorch: an imperative style, high-performance deep learning library//*Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Vancouver, Canada: Curran Associates Inc.: 8026-8037
- Qu L H, Luo X Y, Wang M N and Song Z J. 2022. Bi-directional weakly supervised knowledge distillation for whole slide image classification//*Proceedings of the 36th Conference on Neural Information Processing Systems*. New Orleans, USA: Curran Associates Inc.: 15368-15381
- Shao Z C, Bian H, Chen Y, Wang Y F, Zhang J, Ji X Y and Zhang Y B. 2021. TransMIL: Transformer based correlated multiple instance learning for whole slide image classification//*Proceedings of the 34th Conference on Neural Information Processing Systems*. Sydney, Australia: Curran Associates Inc.: 2136-2147
- Srinidhi C L, Kim S W, Chen F D and Martel A L. 2022. Self-supervised driven consistency training for annotation efficient histopathology image analysis. *Medical Image Analysis*, 75: #102256 [DOI: 10.1016/j.media.2021.102256]
- Su Z Y, Rezapour M, Sajjad U, Gurcan M N and Niazi M K K. 2023. Attention2Minority: a salient instance inference-based multiple instance learning for classifying small lesions in whole slide images. *Computers in Biology and Medicine*, 167: #107607 [DOI: 10.1016/j.combiomed.2023.107607]
- Sung H, Ferlay J, Siegel R L, Laversanne M, Soerjomataram I, Jemal A and Bray F. 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3): 209-249 [DOI: 10.3322/caac.21660]
- Tian X J, Liu G C, Gu S S, Ju Z J, Liu J G and Gu D D. 2018. Deep learning in medical image analysis and its challenges. *Acta Automatica Sinica*, 44(3): 401-424 (田娟秀, 刘国才, 谷珊珊, 鞠忠建, 刘劲光, 顾冬冬. 2018. 医学图像分析深度学习方法研究与挑战. *自动化学报*, 44(3): 401-424) [DOI: 10.16383/j.aas.2018.c170153]
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I. 2017. Attention is all you need//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, USA: Curran Associates Inc.: 6000-6010
- Wang D Y, Khosla A, Gargeya R, Irshad H and Beck A H. 2016. Deep learning for identifying metastatic breast cancer [EB/OL]. [2023-05-20]. <https://arxiv.org/pdf/1606.05718.pdf>
- Wang X, Chen H, Gan C X, Lin H J, Dou Q, Tsougenis E, Huang Q T, Cai M Y and Heng P A. 2020. Weakly supervised deep learning for whole slide lung cancer image analysis. *IEEE Transactions on Cybernetics*, 50(9): 3950-3962 [DOI: 10.1109/TCYB.2019.2935141]
- Wang X G, Yan Y L, Tang P, Bai X and Liu W Y. 2018. Revisiting multiple instance neural networks. *Pattern Recognition*, 74: 15-24 [DOI: 10.1016/j.patcog.2017.08.026]
- Wang X Y, Yang S, Zhang J, Wang M H, Zhang J, Yang W, Huang J Z and Han X. 2022. Transformer-based unsupervised contrastive learning for histopathological image classification. *Medical Image Analysis*, 81: #102559 [DOI: 10.1016/j.media.2022.102559]
- Wang Y W, Guo J, Yang Y, Kang Y, Xia Y L, Li Z H, Duan Y C and Wang K L. 2023. CWC-Transformer: a visual Transformer approach for compressed whole slide image classification. *Neural Computing and Applications*, 1-13 [DOI: 10.1007/s00521-022-07857-3]
- Xie Z D, Lin Y T, Yao Z L, Zhang Z, Dai Q, Cao Y and Hu H. 2021. Self-supervised learning with swin Transformers [EB/OL]. [2023-05-20]. <https://arxiv.org/pdf/2105.04553.pdf>
- Yan Y L, Wang X G, Guo X J, Fang J M, Liu W Y and Huang J Z. 2018. Deep multi-instance learning with dynamic pooling//*Proceedings of the 10th Asian Conference on Machine Learning*. Beijing, China: PMLR: 662-677
- Zhao Y L, Ding W L, You Q H, Zhu F L, Zhu X J, Zheng K and Liu D D. 2023. Classification of whole slide images of breast histopathology based on spatial correlation characteristics. *Journal of Image and Graphics*, 28(4): 1134-1145 (赵樱莉, 丁维龙, 游庆华, 朱峰龙, 朱筱婕, 郑魁, 刘丹丹. 2023. 融合空间相关性特征的乳腺组织病理全切片分类. *中国图象图形学报*, 28(4): 1134-1145) [DOI: 10.11834/jig.211133]
- Zhang H R, Meng Y D, Zhao Y T, Qiao Y H, Yang X Y, Coupland S E and Zheng Y L. 2022. DTFD-MIL: double-tier feature distillation multiple instance learning for histopathology whole slide image classification//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 18802-18812 [DOI: 10.1109/CVPR52688.2022.01824]

作者简介

张建新,男,教授,主要研究方向为医学影像分析和计算机视觉。E-mail:jxzhang0411@163.com

丁雪妍,通信作者,女,讲师,主要研究方向为机器视觉和图像分析与理解。E-mail:dingxueyan@dlnu.edu.cn

高程阳,男,硕士研究生,主要研究方向为深度学习和乳腺癌全切片病理图像分类。E-mail:m15648165880@163.com

孙鉴,男,讲师,主要研究方向为多智能体、集群算法和图像分析与理解。E-mail:m_sunjian@163.com

刘斌,男,教授,主要研究方向为医学影像处理与三维重建、计算机视觉与图形图像、计算机辅助术前规划与仿真。

E-mail:liubin@dlut.edu.cn