DOI:10.19651/j.cnki.emt.2417496

多放大倍率掩码自编码器的乳腺癌图像分类*

司嘉龙[1] 贾 伟[1,2] 赵雪芬[1,2] 高宏娟[1,2]
(1.宁夏大学信息工程学院 银川 750021;2.宁夏"东数西算"人工智能与信息安全重点实验室 银川 750021)

摘 要:乳腺癌是对妇女健康构成严重威胁的疾病之一。早期诊断对于乳腺癌的治愈至关重要,计算机辅助乳腺癌分类诊断得到了广泛使用。虽然基于掩码自编码器的乳腺癌分类方法能够在乳腺癌病理图像已标注数据缺少的前提下进行模型性能的提升,但是现有的基于掩码自编码器的乳腺癌病理图像分类方法没有充分提取和融合不同放大倍率乳腺癌病理图像之间的特征信息。为了解决该问题,提出了一种基于多放大倍率掩码自编码器的乳腺癌病理图像分类方法。该方法在掩码自编码器的基础上结合放大独立和放大特异的优势。首先,设计了规则化噪音掩码模块来避免乳腺癌病理图像重要特征丢失。然后,将不同放大倍率乳腺癌病理图像块组合在一起输入到加入了交叉卷积映射的编码器中提取和融合不同放大倍率图像的特征。最后,在解码器中加入残差交叉注意力机制增强低放大倍率图像下细胞密度及排列顺序和高放大倍率图像下细胞纹理特征的融合。在 BreakHis 公共数据集上进行实验,与现有分类方法相比,该方法在 Top-1 Accuracy、精确率、召回率和 F1-Score 上至少提高了约 2%,说明该方法在良恶性乳腺癌病理图像准确分类方面表现出良好的性能。

关键词:乳腺癌病理图像;自监督学习;掩码自编码器;放大独立;放大特异
中图分类号:TP391;TN29 **文献标识码**:A **国家标准学科分类代码**:520.6

Breast cancer image classification based on multi-magnification mask autoencoders

Si Jialong[1] Jia Wei[1,2] Zhao Xuefen[1,2] Gao Hongjuan[1,2]
(1. School of Information Engineering, Ningxia University, Yinchuan 750021, China; 2. Ningxia Key Laboratory of Artificial Intelligence and Information Security for Channeling Computing Re-sources from the East to the West, Yinchuan 750021, China)

Abstract: Breast cancer is one of the diseases that pose a serious threat to women's health. Early diagnosis is crucial for the cure of breast cancer, and computer-aided breast cancer classification and diagnosis has been widely used. Although the mask autoencoders breast cancer classification method can improve the model performance under the premise of the lack of labeled data in breast cancer pathology images, the existing mask autoencoders breast cancer pathology image classification method does not adequately extract and fuse the feature information between breast cancer pathology images with different magnifications. To solve this problem, a multi-magnification mask autoencoders breast cancer pathology image classification method is proposed, which combines the advantages of magnification independence and magnification specificity on the basis of mask autoencoders. First, a uniform noise masked module is designed to avoid the loss of important features in breast cancer pathology images. Then, blocks of breast cancer pathology images with different magnifications are combined together and fed into an encoder incorporating cross convolution mapping to extract and fuse features from images with different magnifications. Finally, a residual cross attention mechanism is incorporated into the decoder to enhance the fusion of cell density and alignment order under low magnification images and cell texture features under high magnification images. Experiments on the BreakHis public dataset show that the proposed method improves at least about 2% in Top-1 Accuracy, Precision, Recall, and F1-Score compared to existing classification methods. The results demonstrate that the proposed method exhibits good performance in accurately classifying benign and malignant breast cancer pathology images.

Keywords: breast cancer pathology image; self-supervised learning; masked autoencoders; magnification-independence; magnification-specificity

0 引 言

依据国际癌症研究机构最新的全球数据估计,女性乳腺癌、肺癌以及结直肠癌这 3 种癌症类型的合计新诊断病例占比达到了全部癌症新发病例的 51%,其中乳腺癌单独占比即达到 32%,明显超越了其他癌症类型[1-2]。这意味

收稿日期:2024-12-01
* 基金项目:国家自然科学基金(62062057,12062021)、宁夏自然科学基金(2024AAC03063)、宁夏回族自治区重点研发项目(2023BDE03006)资助

着乳腺癌不仅在全球范围内成为了最普遍的癌症种类，其发病数量更是超越了包括肺癌在内的其他恶性肿瘤，凸显了乳腺癌防控的重要性。在中国国内，女性乳腺癌的新发病例数量紧随肺癌之后，呈现出不断攀升的态势，并且在女性癌症致死病因中稳居前五名[3]。面对乳腺癌和肺癌等癌症，采用图像分析方法，可以在癌症早期发现并增加治愈的癌症几率[4-6]，重视筛查和规范治疗是提高生存率、降低死亡率的关键因素。近几年，使用计算机辅助技术对乳腺癌图像分类成为乳腺癌诊断过程中的重要环节[7]。在深度学习技术的辅助下，病理医生可以快速且准确的得到分类结果，及时为患者制定治疗方案，进而降低乳腺癌的死亡率[8-9]。

掩码自编码器(mask autoencoders，MAE)[10]是一种自监督学习方法。其通过使用未掩码的图像块来还原被掩码的图像块，借助自监督方式提升编码器的特征提取能力，从而增强模型的分类准确度。近年来，在自然图像领域中，一系列基于 MAE 的方法[11-14]被提出，且已经取得了不错的效果。由于 MAE 具有在面对图像数据缺少时仍能有效提升模型性能的优势，一些学者对 MAE 在医学图像分类中展开研究。Wang 等[14]在 MAE 的基础上结合多实例学习对食管癌全切片组织病理图像进行自动分类，并使用注意力图辅助分析图像，最终实现图像自动分类。Bai 等[15]在基于编码器-解码器结构引入新的代理任务来识别小儿超声心动图的自监督方法，实现较高的分类准确率。Lu 等[16]提出一种结合文本信息和器官图像的多模态 MAE 方法，通过实验证明在下游任务中取得了先进的结果。但是乳腺癌病理图像有 4 种不同的放大倍率和复杂的特征，导致使用现有 MAE 方法进行乳腺癌病理图像分类时，还存在无法有效提取和融合不同放大倍率图像全局和局部特征的问题，导致分类效果不佳。例如，MAE 方法在乳腺 CT 图像中进行病理图像像素还原，在还原过程中，图像中的乳腺小叶的分布有所改变，但这些变化并未从根本上影响图像整体的类别识别，原因在于图像中的核心结构保持了一致性，图像中的乳腺管和输乳管窦并未发生太大变化。但是 MAE 方法在乳腺癌病理图像上进行像素还原时，细胞的排列模式及细胞边缘特征发生了显著的扰动，这些部分恰恰包含了对于诊断至关重要的癌细胞分布信息，从而影响最终的分类结果。

Benhammou 等[17]提出了放大特异(magnification-specificity，MS)方法，该方法通过对放大倍数归一化，忽略乳腺癌病理图像的放大倍数，完成对乳腺癌病理图像进行分类，在一些临床实验中取得了较好的分类效果。然而，实验证明该方法在不可见样本上表现较差[18]。Bayramoglu 等[19]提出了放大依赖(magnification-independence，MI)方法，通过混合不同倍率的乳腺癌病理图像进行图像分类，有效地融合了不同倍率图像的特征。但是在融合过程中不同倍率图像像素还原过程中会使得该方法无法有效地关注乳腺癌病理图像的重要分类特征，从而影响了方法的分类效果。

受到上述两种乳腺癌病理图像分类方法的启发，提出了一种基于多放大倍率掩码自编码器的乳腺癌病理图像分类方法(multi-magnification masked autoencoders，UMAE)的乳腺癌病理图像分类方法，该方法综合了 MS 和 MI 的优势。首先，在预训练阶段，该方法使用同一个编码器对所有放大倍率图像进行特征提取以进行 MI 分类，不同放大倍率使用同一个编码器进行特征提取。然后，在解码器阶段根据不同倍率进行图像还原，通过采用残差交叉注意力机制有效地加强模型特征共享和关注更多重要分类特征。最后，在微调阶段，使用在预训练阶段训练好的编码器对单一放大倍率图像进行 MS 分类。该编码器在预训练阶段综合了不同放大倍率的特征信息，在面对不可见样本时，也具有较好的分类效果。

综上所述，本文的贡献如下：

1)提出了一种融合 MI 和 MS 的乳腺癌病理图像分类方 UMAE。该方法融合 MI 和 MS 的各自优点，在预训练阶段使用同一个编码器融合不同放大倍率图像的特征信息，使得该编码器可以在微调阶段针对单一放大倍率图像实现高分类准确率。

2)在图像块特征提取过程中设计了规则化噪音掩码模块(uniform noise masked，UNM)，该模块可以在后续编码过程中保证细胞排列方式以及细胞纹理等重要特征信息不丢失的情况下，提高编码器对高级语义信息学习的能力。

3)在编码器中设计了交叉卷积映射(cross convolution projection，CCP)，提升模型特征提取和融合能力。同时，在解码器中提出了残差交叉注意力机制(residual cross attention mechanism，RCA)进一步增强低放大倍率图像下细胞密度及排列顺序和高放大倍率图像下细胞纹理特征的融合。

1　UMAE 方法

在本节中描述了 UMAE 的整体设计框架。UMAE 方法的整体框架如图 1 所示，该方法分为两个阶段：预训练阶段和微调阶段。在预训练阶段构建一个能够整合不同放大倍率图像特征的编码器。这一阶段的工作流程如下：首先，针对各不相同的放大倍率图像，运用规则化噪音掩码模块来生成分块图像数据。然后，各个放大倍率的图像块通过线性映射转化为图像序列，每个图像序列都称为令牌，这些令牌拼接后统一输入到编码器内，编码器内包含交叉卷积映射以便高效提取多放大倍率图像特征。编码器输出的令牌会依据各自的放大倍率进入单独的解码器中，解码器中包含残差交叉注意力机制以便高效实现像素还原。在微调阶段，采纳了以 vision transformer(ViT)[20]为核心架构的良恶性分类模型。将单一放大倍率的图像分解成多个图像块输入到经过预训练得到的编码器模型中进行良恶性分类。

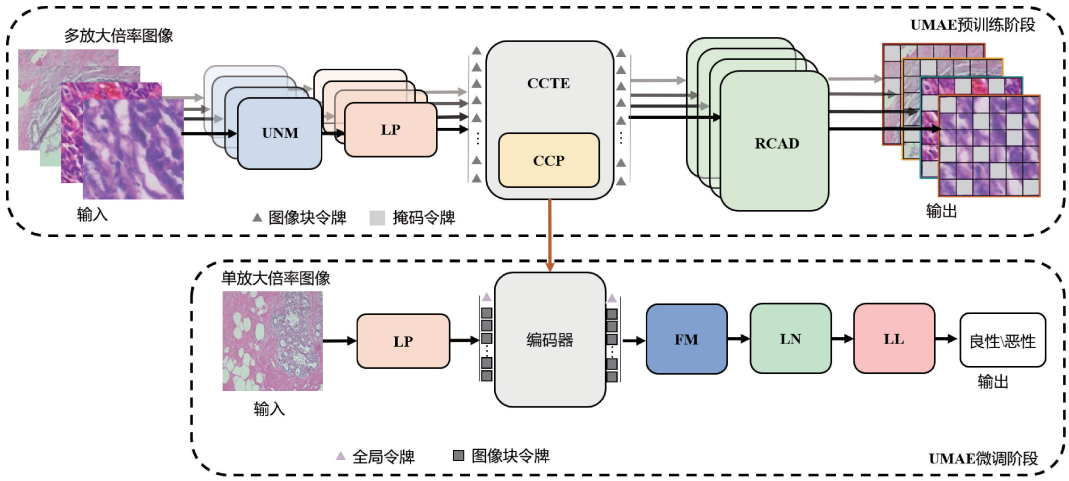


图 1 UMAE 整体框架

Fig. 1 The overall structure of UMAE

1.1 UMAE 的预训练阶段

UMAE 预训练阶段如图 2 所示,输入多放大倍率图像 $x_{40\times}$、$x_{100\times}$、$x_{200\times}$ 和 $x_{400\times}$ 经过 UNM 模块生成的图像块 $P^N_{40\times}$、$P^N_{100\times}$、$P^N_{200\times}$ 和 $P^N_{400\times}$。这些图像块分别经过线性映射层(linear projection,LP)把图像转换为令牌 $T^N_{40\times}$、$T^N_{100\times}$、$T^N_{200\times}$ 和 $T^N_{400\times}$,可表示为式(1):

$$\begin{cases} T^N_{40\times} = LP(P^N_{40\times}) \\ T^N_{100\times} = LP(P^N_{100\times}) \\ T^N_{200\times} = LP(P^N_{200\times}) \\ T^N_{400\times} = LP(P^N_{400\times}) \end{cases} \tag{1}$$

式中:$LP(\cdot)$ 表示线性映射层。然后,将所有的令牌进行连接,合并为一个令牌序列 T^N_{all},合并过程可表示为式(2):

$$T^N_{all} = Concat(T^N_{40\times}, T^N_{100\times}, T^N_{200\times}, T^N_{400\times}) \tag{2}$$

式中:$Concat(\cdot)$ 表示对内部的令牌进行拼接。T^N_{all} 输入到同一个交叉卷积 Transformer 编码器(cross convolution transformer encoder,CCTE)中进行特征学习生成令牌序列 z_l,接着令牌序列 z_l 按放大倍率类别分成不同的令牌,不同放大倍率的令牌都单独使用一个残差交叉注意力解码器(residual cross attention decoder,RCAD)进行像素还原得到还原后的图像块 $F_{40\times}$、$F_{100\times}$、$F_{200\times}$ 和 $F_{400\times}$。

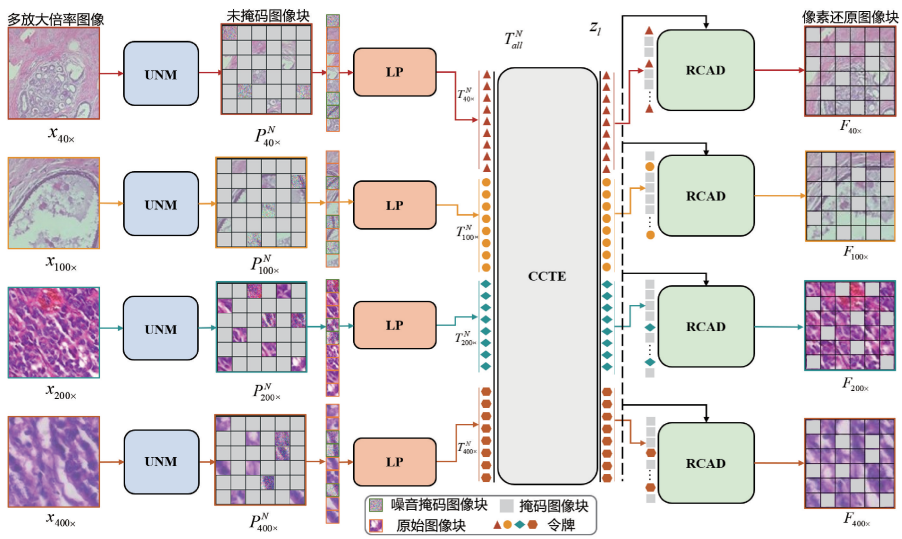


图 2 UMAE 预训练阶段

Fig. 2 UMAE pre-training stage

1)规则化掩码

Li 等[21]提出了将规则化掩码模块应用于金字塔结构网络,且在此方法中,使用了规则化掩码的比例为 75%。具体而言,在每 4 个相邻的块中,会掩码掉其中的 3 个块。尽管这一策略有助于网络在处理任务时保持一定的结构性,但其均匀性也导致了预训练任务的难度有所降低。因

为在这种设定下，网络可以相对容易地利用邻近区域的低级像素信息推断出缺失部分的像素，从而减弱了网络学习高级语义特征的能力[22]。受此启发，提出了规则化噪音掩码模块。该模块结合了均匀掩码和二次噪音掩码两种方法。在图 3 中，40×放大倍率图像 $x_{40\times}$ 均匀采样首先从每个 2×2 网格中严格采样几个随机图像块，使得图像被掩码掉 $\omega\%$。接下来，针对已经掩码的区域，采用二次噪音掩码进行进一步的随机噪音处理，噪音掩码的比例设定为 $\sigma\%$，图 3 中 3 个颜色框处的图像块进行了噪音处理。最后，均匀采样生成的图像块与二次噪音掩码处理后的图像块被结合在一起，形成 $P^N_{40\times}$。其他放大倍率图像也进行上述同样的操作，生成图像块 $P^N_{100\times}$、$P^N_{200\times}$ 和 $P^N_{400\times}$。

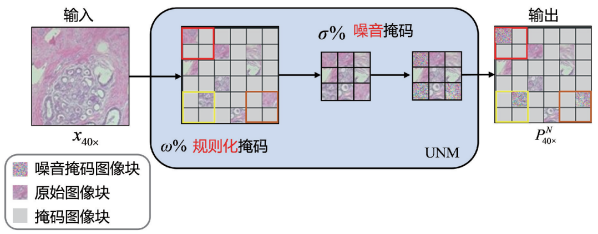


图 3　UNM 模块

Fig. 3　UNM module

4 种不同放大倍率图像 $x_{40\times}$、$x_{100\times}$、$x_{200\times}$ 和 $x_{400\times}$，分别经过切片操作变为图像块 $P^M_{40\times}$、$P^M_{100\times}$、$P^M_{200\times}$ 和 $P^M_{400\times}$，切片操作可表示为式(3)：

$$\begin{cases} P^M_{40\times} = rearrange(x_{40\times}) \\ P^M_{100\times} = rearrange(x_{100\times}) \\ P^M_{200\times} = rearrange(x_{200\times}) \\ P^M_{400\times} = rearrange(x_{400\times}) \end{cases} \tag{3}$$

式中：M 表示图像被分成图像块后的总数量，$rearrange(\cdot)$ 表示把图像转换为图像块。然后将所有的图像块进行 $\omega\%$规则化掩码和 $\sigma\%$二次噪音掩码，可表示为式(4)：

$$\begin{cases} P^N_{40\times} = (P^M_{40\times} \otimes I_l) \otimes' n_l, I_l \in \{0,1\}^{\alpha}, n_l \in \{0,1\}^{\beta} \\ P^N_{100\times} = (P^M_{100\times} \otimes I_l) \otimes' n_l, I_l \in \{0,1\}^{\alpha}, n_l \in \{0,1\}^{\beta} \\ P^N_{200\times} = (P^M_{200\times} \otimes I_l) \otimes' n_l, I_l \in \{0,1\}^{\alpha}, n_l \in \{0,1\}^{\beta} \\ P^N_{400\times} = (P^M_{400\times} \otimes I_l) \otimes' n_l, I_l \in \{0,1\}^{\alpha}, n_l \in \{0,1\}^{\beta} \end{cases} \tag{4}$$

式中：N 表示图像中未被掩码的图像块数量，$\otimes$ 表示规则化掩码，$\otimes'$ 表示二次噪音掩码，α 表示 I_l 中有 $\omega\%$ 的 0 和 $\sigma\%$ 的 1，β 表示 n_l 中有 $\sigma\%$ 的 0 和 $\omega\%$ 的 1。

2)交叉卷积 Transformer 编码器

CCTE 的结构如图 4 所示，T^N_{all} 增加编码后得到 z_0 和调整令牌顺序后的 z'_0，z_0 和 z'_0 输入 CCP 模块来生成对应的 query，key 和 value，并输入到 multi-head attention (MSA)中进行运算，经过残差链接将输出表述为输入的一个非线性变换的线性叠加，得到的结果经过线性归一化(layer normalization，LN)[23]层，利用神经网络中的隐藏层归一化为标准正态分布，加速收敛，最后经过多层感知机(multilayer perceptron，MLP)层，循环叠加运算 L 层生成。为了在后续微调阶段方便地使用普通的 ViT 模型进行分类，在令牌序列中添加了一个全局令牌，类似于 ViT 中的类令牌。由于所有放大倍率图像都具有二维结构，在 LP 层的初始化中加入了二维正余弦位置嵌入。该方法没有将每个放大倍率的所有令牌都输入到编码器中，而是选择了每个放大倍率的部分令牌进行组合输入。

在 T^N_{all} 基础上加入位置编码后生成的序列 z_0 的过程可表示为式(5)：

$$z_0 = [T_{global}, T_{all}E] + E_{pos} \tag{5}$$

式中：T_{global} 表示为全局令牌。$T_{all}E$ 可以表示 $T^1_{x_{40\times}}E, \cdots, T^m_{x_{40\times}}E, T^{m+1}_{x_{100\times}}E, \cdots, T^{m+n}_{x_{100\times}}E, T^{m+n+1}_{x_{200\times}}E, \cdots, T^{m+n+q}_{x_{200\times}}E, T^{m+n+q+1}_{x_{400\times}}E, \cdots, T^{m+n+q+k}_{x_{400\times}}E$，其中 $E \in \mathbb{R}(P^2 \cdot C) \times D$ 和 $E_{pos} \in \mathbb{R}(N+1) \times D$ 表示位置编码。m 表示 40×放大倍率令牌的数量，n 表示 100×放大倍率令牌的数量，q 表示 200×放大倍率令牌的数量，k 表示 400×放大倍率令牌的数量。

调整不同倍率图像令牌的位置，可以得到序列 z'_0 的过程可表示为式(6)：

$$z'_0 = [T_{global}, T^1_{x_{40\times}}E, \cdots, T^m_{x_{40\times}}E, T^{m+1}_{x_{200\times}}E, \cdots, T^{m+n}_{x_{200\times}}E, T^{m+n+1}_{x_{100\times}}E, \cdots, T^{m+n+q}_{x_{100\times}}E, T^{m+n+q+1}_{x_{400\times}}E, \cdots, T^{m+n+q+k}_{x_{400\times}}E] + E_{pos} \tag{6}$$

式中：m 表示 40×放大倍率令牌的数量，n 表示 200×放大倍率令牌的数量，q 表示 100×放大倍率令牌的数量，k 表示 400×放大倍率令牌的数量。

令牌序列 z_0 和 z'_0 输入到 CCP 模块中，利用该模块输出的 ${z_0}^q$，${z_0}^k$，${z_0}^v$ 和 ${z'_0}^q$，${z'_0}^k$，${z'_0}^v$ 计算得到的 $query_0$，key_0 和 $value_0$，计算过程可表示为式(7)：

$$\begin{cases} query_0 = \dfrac{(z_0^q + z'^q_0)}{2} \\ key_0 = \dfrac{({z_0}^k + {z'}^k_0)}{2} \\ value_0 = \dfrac{({z_0}^v + {z'}^v_0)}{2} \end{cases} \tag{7}$$

接着，在 MSA 运算得到的结果经过 LN 层结果输入到 MLP 层中进行运算得到解码器阶段的结果 z_l，可表示为式(8)：

$$z_l = MLP(LN(MSA(query_0, key_0, value_0) + z_{l-1})), \quad l \in 1, \cdots, L \tag{8}$$

式中：MSA 代表多头自注意力机制，LN 代表层归一化操作，MLP 代表多层感知机，L 取值为 12。

3)交叉卷积映射

相比于 Transformer，卷积神经网络能够更好的关注图像的局部特征[24]。受此启发，本文在使用 CCTE 获取全局特征的同时，为突出局部特征信息，在编码器中设计了

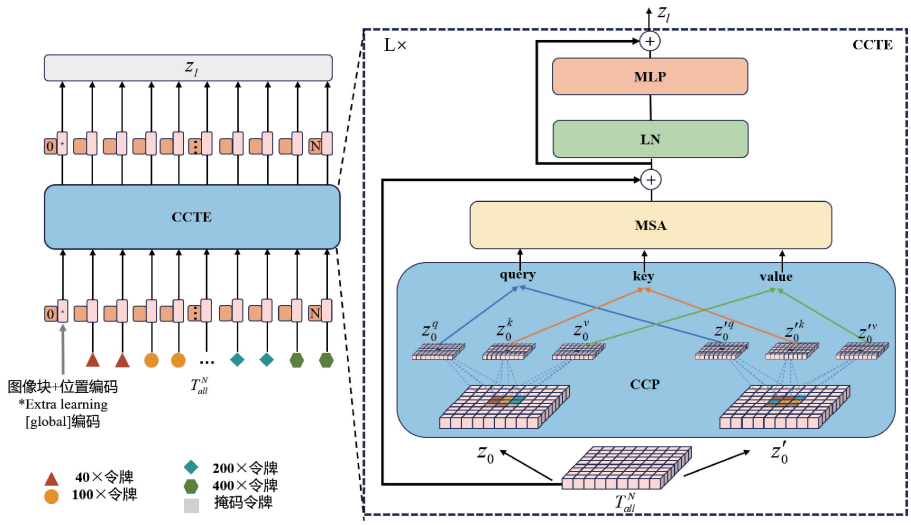


图 4 CCTE 结构

Fig. 4 Structure of CCTE

交叉卷积映射模块,从而融合全局特征和局部特征,提高分类准确度。受此启发,在编码器结构中加入交叉卷积映射模块,突出局部特征信息,结合 Transformer 对全局特征信息的关注,进一步融合图像全局特征和局部特征,提高分类准确度。

CCP 模块结构如图 5 所示。首先,输入 T_{all}^{N} 通过一系列操作被转换成对应的 z_0 和 z'_0。这一转换步骤包括对数据维度的调整,以便适应接下来的卷积操作。接着,经过 3×3 卷积核的卷积操作,得到与之对应的卷积令牌序列,这些卷积令牌序列分别为 z_0 和 z'_0 的输出。在 z_0 和 z'_0 的结果中,蓝色方框位置的令牌序列是有区别的。具体来说,在 z_0 生成的结果中,蓝色方框中的令牌序列所对应的卷积生成的黑色令牌,结合了 40×令牌和 200×令牌的特征。这里,40×和 200×令牌代表了不同尺度的信息,前者是更精细的局部特征,后者则涵盖了更大的范围或更高的语义层次。而在 z'_0 生成的结果中,蓝色方框中的令牌序列所对应的卷积生成的黑色令牌,结合的是 40×令牌和 100×令牌的特征。这种设计通过多尺度特征的融合,确保网络能够捕捉到更多种类的特征信息,增强了模型的表达能力。经过这些卷积操作后,输出序列通过展平操作得到 z_0^q,z_0^k,z_0^v 和 z'^q_0,z'^k_0,z'^v_0,即将卷积后的特征图展平成一维向量,准备进一步传入网络的后续层。在这个过程中,通过不同的组合方式和重塑操作,得到的 z_0^q,z_0^k,z_0^v 和 z'^q_0,z'^k_0,z'^v_0 在信息融合上有了更多的多样性。这种设计方案的关键优势在于能够在不同的尺度上融合更多不同的放大倍率信息。卷积操作关注局部特征,而展平操作则能够保留跨尺度的信息,从而在全局和局部之间实现更高效的特征融合。通过这种多尺度、多层次的信息融合,模型不仅增强了对局部特征的捕捉能力,也能够更好地处理复杂的全局信息,进一步提升了模型的特征融合能力和整体性能。z_0 和 z'_0 生成 $z_0{}^q$,$z_0{}^k$,$z_0{}^v$ 和 z'^q_0,z'^k_0,z'^v_0 可表示为式(9):

$$\begin{cases} z_0^q = Flatten(Conv2d(Reshape(z_0),s)) \\ z_0^k = Flatten(Conv2d(Reshape(z_0),s)) \\ z_0^v = Flatten(Conv2d(Reshape(z_0),s)) \\ z'^q_0 = Flatten(Conv2d(Reshape(z'_0),s)) \\ z'^k_0 = Flatten(Conv2d(Reshape(z'_0),s)) \\ z'^v_0 = Flatten(Conv2d(Reshape(z'_0),s)) \end{cases} \tag{9}$$

式中:$Flatten(\cdot)$ 表示把多维令牌序列展平为一维令牌序列,s 表示卷积核的大小,s 取值为 3。

4)残差交叉注意力解码器

在解码器重建图像块的过程中,为了从可见令牌中还原掩码令牌,对每个放大倍率图像引入了独立的解码器。每个解码器的输入包括来自正在重构的各自放大倍率的完整可见令牌集。采用与 MAE[7] 的相同方法,该方法在解码器中的可见令牌与一组掩码令牌一起被解码,其中掩码令牌的作用是充当占位符,用于解码器生成重建的图像块。

为了整合来自其他放大倍率的令牌信息,同时为了在注意力运算过程中避免信息丢失,在每个解码器中加入了残差交叉注意力机制,将单独一个倍率图像令牌作为 query,而其他倍率图像令牌则作为 key 和 value。与 MAE 相同,仅在掩码令牌上计算损失。

RCAD 的结构如图 6 所示,将来自 CCTE 输出的 z_l 令牌序列,经过 LP 层生成符合 RCAD 运算的 z_d 令牌序列,然后,z_d 经过 RCA 模块生成 $T'_{40\times}$,通过 MLP 层、三层 transformer blocks(TB)、LP 层,最后重构生成像素还原后的图像块 $F_{40\times}$。其他放大倍率图像也进行上述同样的操作,生成像素还原后的图像块 $F_{100\times}$、$F_{200\times}$ 和 $F_{400\times}$。

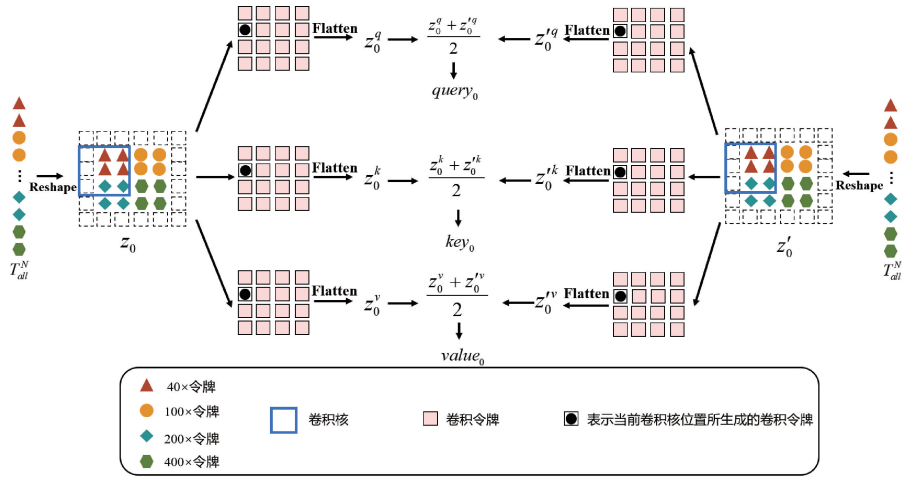


图 5　CCP 模块

Fig. 5　CCP module

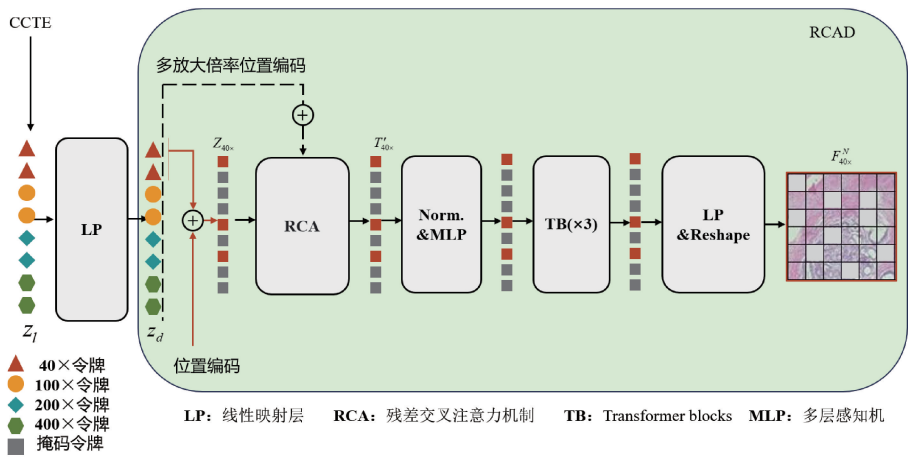


图 6　RCAD 结构

Fig. 6　Structure of RCAD

令牌序列 z_d 可表示为：

$$z_d = LP(z_l), d \in 1\cdots L, l \in 1,\cdots,L \tag{10}$$

根据式(10)中的顺序，把 z_d 序列分成不同倍率的序列 $z_{40\times}$、$z_{100\times}$、$z_{200\times}$ 和 $z_{400\times}$，同时也是不同倍率的 query、key 和 value，$Q_{40\times}$、$K_{40\times}$、$V_{40\times}$ 的值与 $z_{40\times}$ 相同，其中 $z_{40\times}$ 可以表示为 $z_d^{1\cdots m}$，$Q_{100\times}$、$K_{100\times}$、$V_{100\times}$ 的值与 $z_{100\times}$ 相同，其中 $z_{100\times}$ 可以表示为 $z_d^{m+1\cdots m+n}$，$Q_{200\times}$、$V_{200\times}$、$K_{200\times}$ 的值与 $z_{200\times}$ 相同，其中 $z_{200\times}$ 可以表示为 $z_d^{m+n+1\cdots m+n+q}$，$Q_{400\times}$、$K_{400\times}$、$V_{400\times}$ 的值与 $z_{400\times}$ 相同，其中 $z_{400\times}$ 可以表示为 $z_d^{m+n+q+1\cdots m+n+q+k}$。再经过 RCA 模块生成像素还原后的图像块 $F_{40\times}$、$F_{100\times}$、$F_{200\times}$ 和 $F_{400\times}$。最后进行损失函数的计算，可表示为式(11)：

$$\mathcal{L} = \frac{1}{J}\sum_{i=1}^{J} |x_{40\times} - F_{40\times}|^2 + \frac{1}{J}\sum_{i=1}^{J} |x_{100\times} - F_{100\times}|^2 + \frac{1}{J}\sum_{i=1}^{J} |x_{200\times} - F_{200\times}|^2 + \frac{1}{J}\sum_{i=1}^{J} |x_{400\times} - F_{400\times}|^2 \tag{11}$$

式中：$J = M - N$ 表示仅在掩码区域进行损失函数的计算。

5)残差交叉注意力机制

RCA 模块整合来自其他放大倍率令牌的信息，在注意力运算过程中加强特征融合。RCA 模块结构如图 7 所示。在该模块中，以 40×为例，输入包括 40×、100×、200×和 400×的令牌序列。首先，使用 40×的令牌作为 query 和 value，与 100×的令牌作为 key 进行注意力机制的计算。这一步的关键在于，40×的令牌代表了较精细的局部特征，而 100×的令牌代表了更广泛的上下文信息，通过这种方式，注意力机制能够在局部和全局之间建立联系，从而帮助模型更好地捕捉多尺度的信息。注意力机制的计算后，得到的结果与最初的 40×令牌相加。这种相加操作可以看作是对 40×令牌的加权融合，融入了 100×令牌所提供的上下文信息，进一步丰富了特征表示。随后，得到的结果与 200×的令牌继续进行注意力机制的计算，通过这

种方法,200×令牌所包含的更高层次的语义信息也被有效地融入到 40×令牌中,从而让网络能够捕捉到更广泛的尺度信息。在这个过程中,400×的令牌也经历相同的步骤,即与经过多次融合的 40×令牌进行注意力机制的计算。通过这一系列的计算,40×令牌最终融合了 100×、200×和 400×的特征信息。这个过程确保了 40×令牌不仅保留了局部细节,还结合了更大尺度上的信息,从而提升了模型对多尺度特征的感知能力。最后得到的融合特征就是 40×令牌 $T'_{40\times}$,这代表着融合了不同放大倍率的信息,并通过注意力机制精细地融合了多尺度特征,确保了模型能够有效地捕捉到全局和局部信息的平衡。这种设计的优势在于通过多尺度注意力机制,有效地将不同放大倍率的特征进行融合,增强了模型对不同尺度特征的感知能力,进而提升了模型的表达能力和性能。通过这种方式,RCA 模块不仅关注单一尺度的信息,还能够综合各个尺度的信息,从而使得模型在处理复杂任务时能够获得更为丰富和精准的特征表示。用公式表示为式(12):

$$\begin{cases} A^{40\times100\times}(Q_{40\times},K_{100\times},V_{40\times})=softmax\left(\dfrac{K_{100\times}Q_{40\times}^{T}}{\sqrt{d_v}}\right)V_{40\times} \\ T'_{40\times}=A^{40\times100\times}(Q_{40\times},K_{100\times},V_{40\times})+z_{40\times} \\ A^{T'_{40\times}200\times}(Q_{T'_{40\times}},K_{200\times},V_{T'_{40\times}})=softmax\left(\dfrac{K_{200\times}Q_{T'_{40\times}}^{T}}{\sqrt{d_v}}\right)V_{T'_{40\times}} \\ T'_{40\times}=A^{T'_{40\times}200\times}(Q_{T'_{40\times}},K_{200\times},V_{T'_{40\times}})+T'_{40\times} \\ A^{T'_{40\times}400\times}(Q_{T'_{40\times}},K_{400\times},V_{T'_{40\times}})=softmax\left(\dfrac{K_{400\times}Q_{T'_{40\times}}^{T}}{\sqrt{d_v}}\right)V_{T'_{40\times}} \\ T'_{40\times}=A^{T'_{40\times}400\times}(Q_{T'_{40\times}},K_{T'_{40\times}},V_{400\times})+A^{T'_{40\times}200\times}(Q_{T'_{40\times}},K_{200\times},V_{T'_{40\times}})+A^{40\times100\times}(Q_{40\times},K_{100\times},V_{40\times})+z_{40\times} \end{cases} \tag{12}$$

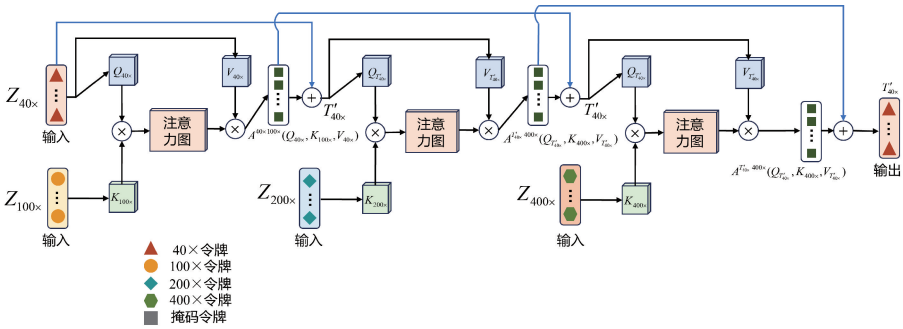


图 7 RCA 模块

Fig. 7 RCA module

其他倍率图像的计算也是如此,得到 $T'_{100\times}$,$T'_{200\times}$,$T'_{400\times}$,用公式表示为式(13):

$$\begin{cases} T'_{100\times}=A^{100\times40\times}(Q_{100\times},K_{40\times},V_{100\times})+A^{T'_{100\times}200\times}(Q_{T'_{100\times}},K_{200\times},V_{T'_{100\times}})+A^{T'_{100\times}400\times}(Q_{T'_{100\times}},K_{400\times},V_{T'_{100\times}})+z_{100\times} \\ T'_{200\times}=A^{200\times40\times}(Q_{200\times},K_{40\times},V_{200\times})+A^{T'_{200\times}100\times}(Q_{T'_{200\times}},K_{100\times},V_{T'_{200\times}})+A^{T'_{200\times}400\times}(Q_{T'_{200\times}},K_{400\times},V_{T'_{200\times}})+z_{200\times} \\ T'_{400\times}=A^{400\times40\times}(Q_{400\times},K_{40\times},V_{400\times})+A^{T'_{400\times}100\times}(Q_{T'_{400\times}},K_{100\times},V_{T'_{400\times}})+A^{T'_{400\times}200\times}(Q_{T'_{400\times}},K_{200\times},V_{T'_{400\times}})+z_{400\times} \end{cases} \tag{13}$$

在 RCA 模块之后,再经过一层 MLP 和三层 TB,最后再经过 LP,重塑成最初图像的大小 $F_{40\times}$。其他放大倍率图像也进行上述同样的操作,生成像素还原后的图像块 $F_{100\times}$、$F_{200\times}$ 和 $F_{400\times}$。生成过程可表示为式(14):

$$\begin{cases} F_{40\times}=Reshape(LP(TB(MLP(T'_{40\times})))) \\ F_{100\times}=Reshape(LP(TB(MLP(T'_{100\times})))) \\ F_{200\times}=Reshape(LP(TB(MLP(T'_{200\times})))) \\ F_{400\times}=Reshape(LP(TB(MLP(T'_{400\times})))) \end{cases} \tag{14}$$

1.2 UMAE 的微调阶段

微调阶段的结构如图 1 所示。在微调阶段,采用以 ViT 为基础框架的良恶性分类模型。将一张单一放大倍率的图像分解成多个图像块,并将它们输入到 UMAE 预训练编码器中。该编码器采用的是来自预训练阶段生成的编码器模型,使用该模型的权重进行初始化。接着,模型根据单一放大倍率的图像进行参数微调,生成相应的令牌序列。最后,将其中的全局令牌输入到特征映射层(feature map,FM)[25]中生成特征图,随后经过 LN 层进行正则化。最终,通过线性层(linear layer,LL)[26]将维度调整到分类维度,以进行良恶性分类。在整个框架中,UMAE 的预训练阶段采用了多放大倍率图像进行训练。由此得到的编码器预训练模型在微调阶段为微调阶段单放大倍率图像良恶性分类任务中的编码器进行了初始化。

这种设计使得在单放大倍率图像分类过程中使用的编码器能够有效融合多放大倍率图像的信息，进一步提升了分类准确度。通过这一模块，UMAE 方法融合了 MI 和 MS 的各自优点。

2　实验结果与分析

2.1　数据集介绍

实验中使用的乳腺癌组织病理学数据库 BreakHis 数据集[27]，该数据集由 82 例患者的 7 909 张乳腺肿瘤组织显微图像组成。其中包含 2 480 个良性样本和 5 429 个恶性样本，并且图像使用了不同的放大倍率，分别是 40×、100×、200×和 400×，低倍率(40×)能够模拟较为粗糙的图像特征，适合评估模型在整体轮廓或大尺度特征捕捉上的表现，而高倍率(400×)则更加注重细节信息的提取，适用于处理精细结构或微小差异的场景。100×和 200×作为中间的倍率值，能够提供一个过渡尺度，既能够捕捉到一定程度的细节信息，又不会过于精细到让噪声和微小误差显现。详细信息如表 1 所示。每张乳腺癌病理图像的大小为 700×460 像素。在实验中训练集、验证集和测试集的比例为 80%、10%和 10%。

表 1　BreakHis 数据集的详细信息

Table 1　Details of the BreakHis dataset

放大倍率	良性	恶性	总计
40×	652	1 370	1 995
100×	644	1 437	2 081
200×	623	1 390	2 013
400×	588	1 232	1 820
总计	2 480	5 429	7 909

2.2　预训练阶段

实验采用了 ViT-B[17]作为基础模型，patch size 被设置为 20×20 像素。在实验中，选择了 AdamW[28]优化器，基础学习率为 1×10^{-4}，同时将权重衰减设置为 0.05。学习率每经过 40 个 epochs 就下降 0.01，最终调整至 0。Input size 被设定为 460，而 batch size 则为 16。方法使用了 PyTorch 框架，版本为 1.12.0，并在搭建实验环境的服务器上使用了 NVIDIA GeForce RTX 3090 24G GPU。

2.3　评价标准

1)均方误差损失

均方误差损失(mean squared error loss，MSE Loss)是一种用于衡量模型预测值与实际观测值之间差异的损失函数。对于每个样本，MSE 计算模型的预测值与实际值之间的差异，将差异平方，并对所有样本取平均值。这样可以得到一个衡量模型整体性能的指标，其中差异越小，MSE 值越低，表示模型的预测越接近实际观测值，计算公式可表示为式(15)：

$$MSE\ Loss = \frac{1}{n}\sum_{i=1}^{n}(y_i - y'_i)^2 \tag{15}$$

2)Top-1 Accuracy

Top-1 Accuracy 是指在分类问题中，模型正确预测出测试集中每个样本的最可能的标签的比例。具体来说，对于测试集中的每个样本，模型预测的标签为 y'_i，真实的标签为 y_i，则 Top-1 Accuracy 的计算公式可表示为式(16)：

$$Top\text{-}1\ Accuracy = \frac{\sum_{i=1}^{N}\mathbb{I}(y'_i = y_i)}{N} \tag{16}$$

式中：$\mathbb{I}(\cdot)$为指示函数，当括号内的条件为真时取值为 1，否则为 0；N 为测试集中样本的总数。

3)精确率(Precision)

精确率是指在所有被分类为正类的样本中，真正为正类的样本所占的比例。可表示为式(17)：

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

式中：TP 是指被分类器正确地预测为正类的样本数量，FP 是指被分类器错误地预测为正类的负样本数量。

4)召回率(Recall)

召回率是指在所有真正为正类的样本中，分类器正确预测为正类的样本所占的比例。可表示为式(18)：

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

式中：TP 是指被分类器正确地预测为正类的样本数量，FN 是指被分类器错误地预测为负类的正样本数量。

5)F1-Score

F1-Score 是准确率和召回率的调和平均数，它将两个指标综合考虑，对两者都给予了相等的权重。可表示为式(19)：

$$F1\text{-}Score = 2\times\frac{Precision\times Recall}{Precision + Recall} \tag{19}$$

2.4　ω 和 σ 对 UNM 模块的影响

为了选择合适的超参数 ω 和 σ，本文在预训练阶段进行了系统的实验。具体地，通过计算 MSE Loss 来评估不同参数组合的效果，选择范围为[0，1]，步长为 0.1。这个步长的选择是为了在保证计算效率的同时，能够精细地探索每个参数的变化对模型性能的影响。此外，选择不同的放大倍率是为了观察在不同图像细节下，超参数对 MSE Loss 的影响。如表 2 所示，实验结果表明，在 40×放大倍率的条件下，设定 $\omega=0.7$ 和 $\sigma=0.3$ 时，MSE Loss 达到了最低值 0.242；在 100×放大倍率下，$\omega=0.8$ 和 $\sigma=0.2$ 时，MSE Loss 为 0.299；而在 200×放大倍率和 400×放大倍率下，最优组合仍为 $\omega=0.7$ 和 $\sigma=0.3$，分别得到的 MSE Loss 为 0.279 和 0.355。基于这些实验结果，本文注意到在不同的放大倍率下，$\omega=0.7$ 和 $\sigma=0.3$ 的组合表现稳定，因此选择了这个组合作为基础。然而，为了更全面地考虑各个条件的影响，本文最终选择了 $\omega=0.75$ 和 $\sigma=$

0.25 的组合作为最终模型的超参数设定,这一选择考虑了各倍率条件下的综合表现。

表 2 ω 和 σ 对 UNM 模块的影响

Table 2 The impact of ω and σ on the UNM module

放大倍率	ω	σ	MSE Loss
40×	0.1	0.9	0.806
	0.2	0.8	1.269
	0.3	0.7	0.774
	0.4	0.6	0.590
	0.5	0.5	0.339
	0.6	0.4	0.294
	0.7	0.3	**0.242**
	0.8	0.2	0.436
	0.9	0.1	0.887
100×	0.1	0.9	0.911
	0.2	0.8	0.856
	0.3	0.7	1.433
	0.4	0.6	1.364
	0.5	0.5	0.781
	0.6	0.4	0.634
	0.7	0.3	0.307
	0.8	0.2	**0.299**
	0.9	0.1	0.576
200×	0.1	0.9	0.672
	0.2	0.8	0.640
	0.3	0.7	0.553
	0.4	0.6	1.301
	0.5	0.5	0.406
	0.6	0.4	0.341
	0.7	0.3	**0.279**
	0.8	0.2	0.287
	0.9	0.1	0.439
400×	0.1	0.9	0.811
	0.2	0.8	0.940
	0.3	0.7	1.196
	0.4	0.6	0.481
	0.5	0.5	0.409
	0.6	0.4	0.390
	0.7	0.3	**0.355**
	0.8	0.2	0.401
	0.9	0.1	0.439

2.5 消融实验

使用消融实验来验证 UMAE 模型的有效性,实验结果如表 3 所示。在基础 MAE 架构上,加入了 UNM 模块以防止乳腺癌病理图像关键特征的丢失,该模块在 4 种不同放大倍数的图像中均提升了 Top-1 Accuracy,并且其他 3 个评价指标也取得了约 2%的增长。该方法所提出的 CCP 模块则更有效地提取并融合图像特征,尽管 MAE 中原有的随机掩码模块可能导致乳腺癌病理图像的部分特征丢失,但是在加入 CCP 模块后,所有放大倍率下的四个评价指标均实现了提升。为进一步优化融合各放大倍率图像特征信息,该方法设计了 RCA 模块,运用残差思想加以实现。实验结果显示,在 40×、200×及 400×放大倍率下,Top-1 Accuracy 稳定在 80%左右,其余 3 项指标都有约 1%的提高。在此基础结构之上再加入 RCA 模块后,Top-1 Accuracy 的整体提升幅度约为 1.5%,证明了 UNM 模块在保留图像关键特征的同时,能够协助 CCP 模块提取和融合更多的特征信息。更为显著的是,在此基础结构之上再加入 RCA 模块后,Top-1 Accuracy 的整体提升达到了 5%左右,特别是在 200×和 400×放大倍率的图像中,UMAE 方法的 Top-1 Accuracy 超过了 86%,其他 3 个指标也提升了大约 6%。通过对整体实验结果的分析,可以看出本模型融合了来自各种放大倍率的图像信息,充分证实了融合 MS 和 MI 优势的关键作用,从而进一步表明了该方法在处理乳腺癌病理图像时的优势。此外,在乳腺癌病理图像变化时,UNM 模块通过将均匀采样生成的图像块与二次噪音掩码处理后的图像块结合在一起,在保留细胞排列方式以及细胞纹理等重要特征信息的同时,能够获得更多的高级语义特征,这对于模型性能的提升起到了关键支撑作用。同时,CCP 模块和 RCA 模块的引入对 4 项评价指标产生了良好的提升效果。其中,对于编码器的 CCP 模块,其通过卷积操作获得局部特征以及卷积后的展平操作获得跨尺度的特征信息,能够将多尺度和多层次的病理特征信息进行融合,从而强化了模型对特征的提取和融合能力。而对于解码器中的 RCA 模块,其通过多尺度注意力机制,将不同放大倍率的令牌信息进行逐层整合,增强了低放大倍率图像下细胞密度及排列顺序和高放大倍率图像下细胞纹理特征的融合能力,使得模型具备了更好的多层次特征感知能力,从而使模型在对乳腺癌病理图像进行良恶性分类时体现出较高的精确度和可靠性。综上所述,UMAE 模型在 MAE 基础上集成 UNM、CCP 和 RCA 模块,提升了乳腺癌病理图像分类任务的能力。

2.6 实验结果对比

在 BreakHis 数据集中,对 UMAE 预训练模型在微调阶段的分类性能进行了评估,实验结果如表 4 所示。与 VGG[29]和 ResNet[30]相比,UMAE 采用了自监督学习的方法,这种学习方式不需要大规模的人工标注数据集,而是通过设计特定的预训练任务,使模型能够从无标签的数据中自动学习有用的特征表示。相反,VGG 和 ResNet 等传统模型则依赖于监督学习方式进行训练,需要大量的、经过精细标注的数据来指导模型的学习过程,以确保模型能够正确分类或识别图像中的对象。这种差异使得 UMAE

表 3　消融实验结果

Table 3　Results of ablation experiment

方法	放大倍率	Top-1 准确率/%	精确率/%	召回率/%	F1 分数/%
MAE	40×	78.52	93.66	83.64	88.37
	100×	77.10	92.74	82.45	87.29
	200×	80.27	95.11	82.76	88.51
	400×	79.87	93.88	84.79	89.10
MAE+ UNM	40×	78.64	94.26	84.17	88.93
	100×	77.18	93.11	83.26	87.91
	200×	80.55	95.30	83.00	88.73
	400×	80.09	94.09	85.19	89.42
MAE+CCP	40×	79.03	94.31	83.80	88.74
	100×	77.26	92.89	82.64	87.47
	200×	80.42	95.43	83.06	88.82
	400×	79.90	93.97	85.47	89.52
MAE+RCA	40×	79.17	94.87	84.07	89.14
	100×	77.69	93.87	83.44	88.35
	200×	80.81	96.09	82.96	89.04
	400×	80.06	95.36	85.46	90.14
MAE+UNM+CCP	40×	79.89	95.28	84.66	89.66
	100×	78.16	94.70	84.59	89.36
	200×	81.67	96.78	83.50	89.65
	400×	81.29	96.06	86.19	90.86
MAE+UNM+RCA	40×	79.11	94.23	84.10	88.88
	100×	77.68	93.03	83.59	88.06
	200×	82.10	96.50	86.94	91.47
	400×	81.55	94.57	88.64	91.51
MAE+CCP+RCA	40×	80.63	93.87	85.63	89.56
	100×	78.74	92.85	84.00	88.20
	200×	82.71	95.47	86.39	90.70
	400×	83.19	95.10	90.87	92.94
UMAE(ours)	40×	**81.90**	**95.45**	**87.59**	**91.35**
	100×	**80.21**	**93.94**	**86.50**	**90.07**
	200×	**86.29**	**97.48**	**93.03**	**95.20**
	400×	**86.26**	**96.16**	**92.90**	**94.54**

在处理乳腺癌病理图像分类任务时展现出优势。总之，UMAE 在不同放大倍率的图像分类中表现良好，能够更准确地识别病理图像特征和分类病理图像。MutilMAE[31]、MMAE[32]、SAME[33]和 GCMAE[34]都是基于 MAE[10]的自监督学习方法。MutilMAE 和 MMAE 方法核心思想是使用同一种图像的不同处理方法进行模型训练，UMAE 方法相较于 MutilMAE 和 MMAE 方法，UMAE 方法在 Top-1 Accuracy 指标上平均提高了约 2%，同时在其他 3 个评价指标上均有约 6%的数值增长。尽管 MutilMAE 在处理自然图像时表现出色，但在应用于乳腺癌病理图像分析时却未能达到预期效果；虽然 MMAE 尝试结合乳腺癌病理图像特点以及苏木素伊红染色进行模型训练，但是在应对多放大倍率图像问题时，其模型性能受限，导致 4 个指标均有所下滑。UMAE 方法相比于 SAME 方法，UMAE 方法在 4 个不同放大倍率的图像上，Top-1 Accuracy 提升了约 7%，其余 4 个指标提升高达 10%左右。这是因为 SAME 提出的小样本学习策略主要适用于 CT 图像场景，将其直接应用到乳腺癌病理图像分析时，其有效性受到限制。另一方面，GCMAE 方法结合了 MAE 和对比学习两种自监督学习机制，针对乳腺癌全切片图像分类显示出一定的性能提升，使得 UMAE 方法在各项指标上的提升幅度相对较小，仅约为 2%。然而，GCMAE 在选取特定切片图像进

行模型训练的过程中,面对已预先完成切片处理的乳腺癌病理图像时,其表现不如 UMAE 方法。综上所示,UMAE 提出的 UNM 模块不同于传统的 MAE 方法采用随机掩码模块,UNM 模块针对乳腺癌病理图像保证了关键诊断信息的保护,还通过适量引入噪声元素,提升了模型训练任务的复杂性,进而有效增强了模型的能力。此外,UMAE 还加入了 CCP 模块,该模块巧妙地结合了卷积神经网络与 Transformer 架构的优点,极大地提升了模型对全局上下文信息和局部细节特征的提取能力。同时,引入的 RCA 模块进一步增强低放大倍率图像下细胞密度及排列顺序和高放大倍率图像下细胞纹理特征的融合。

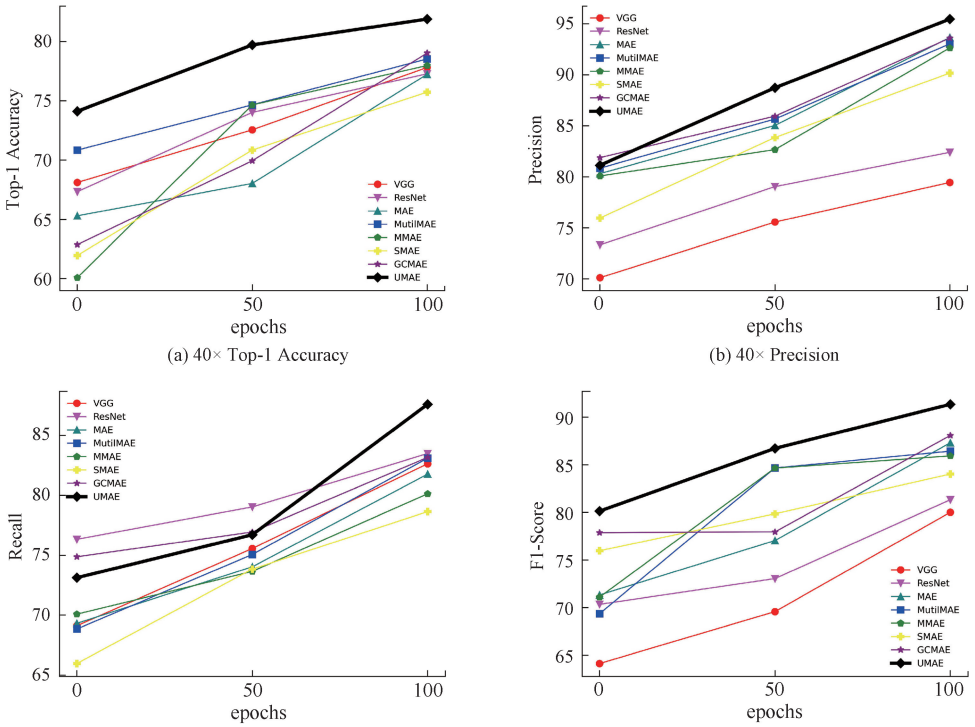
表 4 不同分类方法的实验结果

Table 4 Comparison of experimental results of different classification methods

方法	放大倍率	Top-1 准确率/%	精确率/%	召回率/%	F1 分数/%
VGG[29]	40×	77.82	79.44	82.64	76.48
	100×	78.94	80.62	83.72	77.75
	200×	80.56	82.44	84.98	80.05
	400×	79.37	81.49	84.25	78.91
ResNet[30]	40×	77.28	82.39	83.49	81.32
	100×	80.15	83.83	85.16	82.55
	200×	82.47	86.58	88.65	84.61
	400×	78.03	85.50	87.28	83.79
MAE[7]	40×	78.52	93.66	83.64	89.68
	100×	77.10	92.74	82.45	87.29
	200×	80.27	95.11	82.76	88.51
	400×	79.87	93.88	84.79	89.10
Mutil-MAE[31]	40×	79.84	94.51	84.11	89.01
	100×	79.00	94.01	83.94	88.69
	200×	81.59	95.27	85.46	90.10
	400×	80.88	93.41	84.80	88.90
MMAE[32]	40×	78.98	93.64	82.91	87.95
	100×	76.71	91.03	79.36	84.80
	200×	80.16	94.58	86.33	90.23
	400×	79.40	90.77	81.64	85.96
SMAE[33]	40×	76.10	91.50	80.64	85.73
	100×	75.84	90.30	78.00	83.70
	200×	78.66	90.51	80.24	85.07
	400×	77.74	89.87	80.11	84.71
GCMAE[34]	40×	80.74	94.11	85.77	89.75
	100×	79.03	91.78	83.02	87.18
	200×	83.61	95.61	89.74	92.58
	400×	84.96	96.11	91.03	93.50
UMAE(ours)	40×	**81.90**	**95.45**	**87.59**	**91.35**
	100×	**80.21**	**93.94**	**86.50**	**90.07**
	200×	**86.29**	**97.48**	**93.03**	**95.20**
	400×	**86.26**	**96.16**	**92.97**	**94.54**

UMAE 与现有分类方法的比较如图 8~11 所示。从图中可以看出,在不同放大倍率的乳腺癌病理图像分类任务中,UMAE 方法相较于 VGG 展示了显著的性能优势。具体而言,在 40×和 200×放大倍率的图像上,UMAE 方法的 Top-1 Accuracy、F1-Score、Precision 及 Recall 四项关键评价指标平均比 VGG 高出约 10%。对于 100×放大倍率的图像,尽管 UMAE 方法在训练初期的表现略差,但随着模型迭代次数的增加,其性能逐步提升,并最终超越了

VGG 和 ResNet 这两种经典的监督学习模型。在 400×放大倍率的图像，UMAE 方法不仅实现了更高的分类准确率，而且相对于 VGG 和 ResNet，其性能增长速率更快。在 40×放大倍率的图像中，Top-1 Accuracy 和 F1-Score，两项指标均持续高于所有的对比方法；而 Precision 和 Recall 两项指标综合来看，MutilMAE 方法与 UMAE 方法结果比较接近。对于 100×放大倍率的图像，在初始阶段 UMAE 方法在这 4 项指标上均低于对比方法，但是随着迭代次数的增加，UMAE 在四项指标的表现均优于其他分类方法。对于 200×放大倍率的图像，UMAE 在 Top-1 Accuracy 的表现优于其他分类方法在整个过程中领先与其他方法，其余 3 项指标与 MMAE 对比方法接近，但 UMAE 方法整体效果良好。而在 400×放大倍率的图像中，在初始阶段，UMAE 的整体表现略低于 GCMAE，而与 MutilMAE 分类效果接近，但是，随着迭代次数的增加，UMAE 的表现开始优于其他分类方法。由于该放大倍率图像的全局特征相对较少[35-36]，对模型提出了更高的要求，UMAE 方法在编码器中加入了 CCP，在多个尺度和多个层次上能够有效提取和融合局部和全局特征。并在解码器中加入了 RCA，通过多尺度注意力机制逐层融合不同放大倍率的局部和全局特征，建立了局部和全局之间的联系，从而具备较好的多倍率特征感应能力。此外，UNM 模块可以保留的细胞排列方式以及细胞纹理等重要特征信息，这为局部和全局的高级语义特征的提取提供了保障。尽管在此条件下 UMAE 方法优势并不突出，尤其是 Top-1 Accuracy 指标直到训练后期才略微超过其他方法，但仍能体现出一定的优势。综上所述，UMAE 方法在对不同放大倍率的乳腺癌病理图像进行分类时，具有较好的分类效果。这证明 UMAE 的多个模块能逐步整合并充分利用各放大倍率下的图像特征信息。

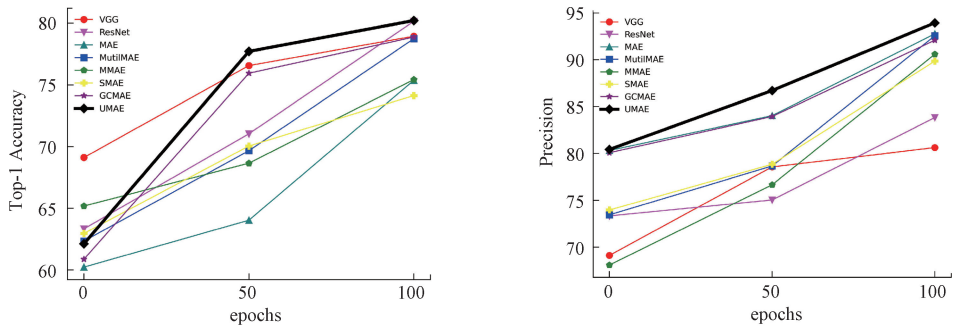


(a) 40× Top-1 Accuracy　(b) 40× Precision

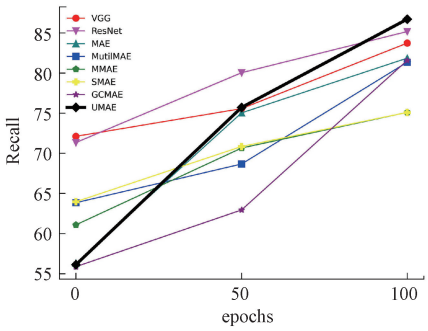
(c) 40× Recall　(d) 40× F1-Score

图 8　40×放大倍率图像的实验结果

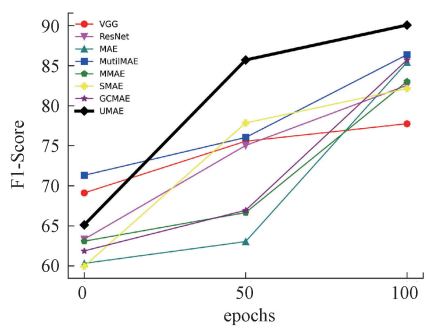
Fig. 8　Experimental results for 40× magnification images



(a) 100× Top-1 Accuracy　(b) 100× Precision



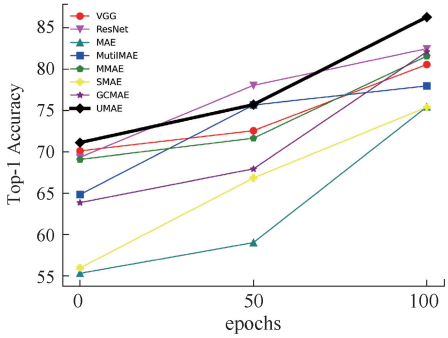
(c) 100× Recall



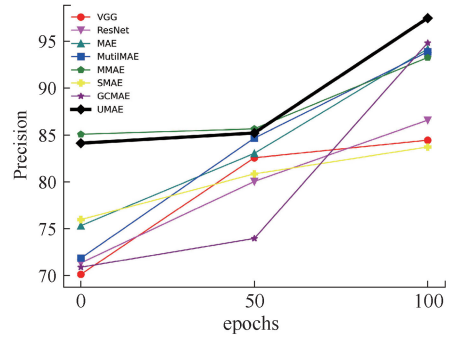
(d) 100× F1-Score

图 9 100×放大倍率图像的实验结果

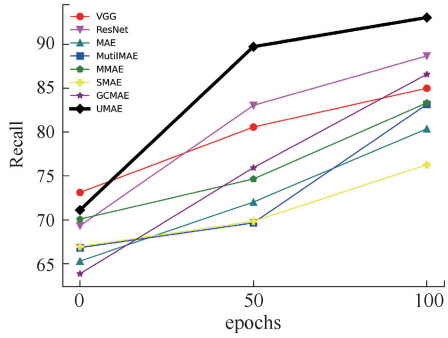
Fig. 9 Experimental results for 100× magnification images



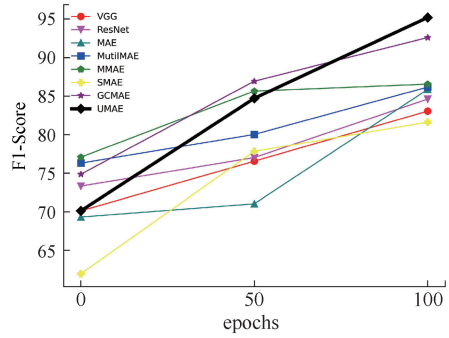
(a) 200× Top-1 Accuracy



(b) 200× Precision



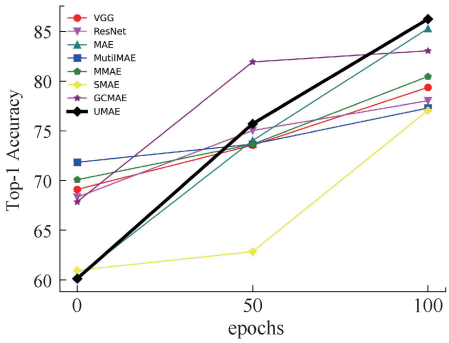
(c) 200× Recall



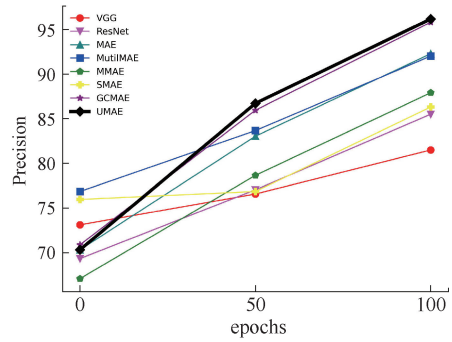
(d) 200× F1-Score

图 10 200×放大倍率图像的实验结果

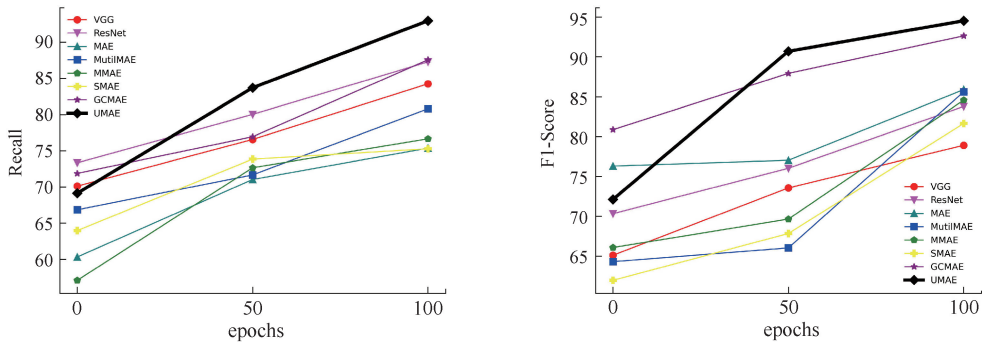
Fig. 10 Experimental results for 200× magnification images



(a) 400× Top-1 Accuracy



(b) 400× Precision



(c) 400× Recall　　(d) 400× F1-Score

图 11　400×放大倍率图像的实验结果

Fig. 11　Experimental results for 400× magnification images

2.7　验证 UNM 模块有效性的实验

图 12 是 UNM 模块和随机掩码模块在乳腺癌病理图像像素重建可视化对比图，通过图 12 清晰地展示了 UMAE 中的 UNM 模块在图像还原方面相较于采用随机掩码模块的方法更为卓越。当将随机掩码模块应用于乳腺癌病理图像时，如图 12 中红色方框所示，观察到图像中癌细胞的排列和细胞纹路发生了变化。这不仅导致缺乏对分类有益的信息，还会误导模型，使其在拟合过程中受到不正确信息的阻碍。UNM 首先通过均匀采样保证被掩码图像块附近有可以学习的未掩码图像块，接着使用二次噪音掩码提高任务难度，使得模型可以学习高级语义特征，成功地最大化对图像中重要信息的还原，特别是在癌细胞组织排列的还原方面取得了良好的效果。

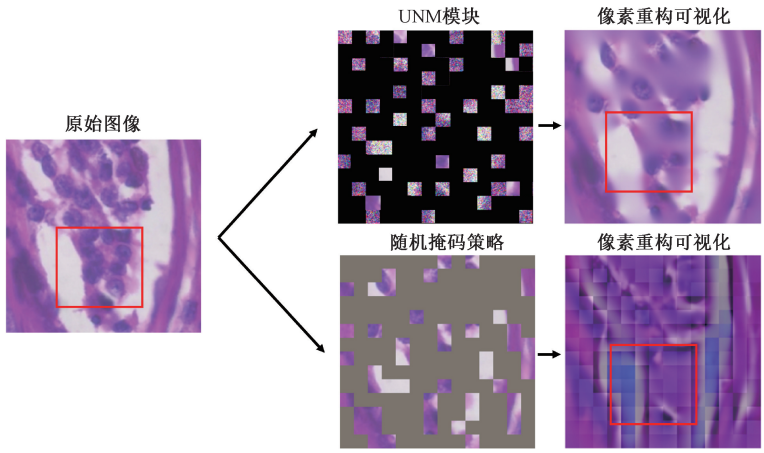


图 12　像素还原可视化对比

Fig. 12　Pixel reduction visualization comparison

2.8　验证 RCA 模块有效性的实验

验证 RCA 模块有效性的热力图对比如图 13 所示，通过观察热力图可以清晰地发现，在加入 RCA 模块后，模型通过提取来自不同放大倍率图像的特征信息，并将其融合在一起。RCA 采用逐渐融合结合残差方式，单放大倍率图像逐渐与其他 3 种放大倍率图像进行融合，在这个过程中残差加入中间结果，这种融合使得模型能够有效地关注更多区域的重要分类特征。在图 13 所示的 40×与 100×放大倍数图像内，黑色箭头指示的细胞排列及其密度特征受到了高度关注，相较于红色箭头标识的区域更为显著。而在 200×与 400×更高放大倍率的图像中，黑色箭头所指向的细胞纹理特征的关注度也明显超越了红色箭头所示区域，RCA 模块提高了相关区域的特征表示质量。这种提升在最终的准确分类中起到了重要作用，为模型的性能提升提供了基础。

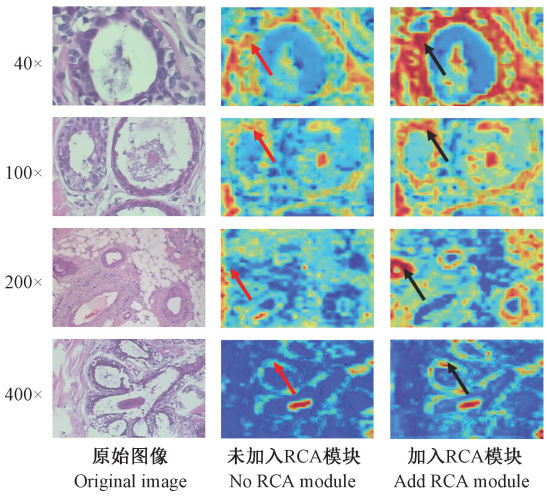


图 13 RCA 模块添加前后热力图对比图

Fig. 13 Comparison of heat map before and after RCA module addition

3 结 论

为进一步提升乳腺癌病理图像的分类效果,本文提出了一种有效的自监督方法 UMAE,该方法利用 MI 和 MS 融合不同放大倍率病理图像的信息,在图像切片过程中,使用 UNM 模块减少细胞排列顺序和细胞纹理等重要特征丢失,在编码器中使用 CCP 模块增强了模型提取和融合全局和局部特征的能力,在解码器中使用 RCA 模块进一步增强低放大倍率图像下细胞密度和高放大倍率图像下细胞纹理特征的融合。UMAE 方法在 BreakHis 数据集上表现良好,证明了其有效性。本文所提方法虽然采取了一些现有的数据增强方法,但是这些数据增强方法有可能丢失乳腺癌病理图像中的部分重要上下文信息,导致模型泛化能力不足的问题。本文现有的研究工作主要是针对分类方法的分类准确性进行研究,在未来的研究中,将探索在数据不平衡的情况下,如何提高分类方法的泛化能力,实现分类准确度的进一步提升,同时,在后续的研究中,将对分类方法的实时性进行评价,并研究在资源受限设备上的运行可靠性。

参考文献

[1] SIEGEL R L, GIAQUINTO A N, JEMAL A. Cancer statistics, 2024 [J]. CA: A Cancer Journal for Clinicians, 2024, 74(1): 12-49.

[2] ZHUANG J J, WU X H, MENG D D, et al. A swin transformer and residualnetwork combined model for breast cancer disease multi-classification using histopathological images[J]. Instrumentation, 2024, 11(1): 112-120.

[3] HAN B F, ZHENG R SH, ZENG H M, et al. Cancer incidence and mortality in China, 2022[J]. Journal of the National Cancer Center, 2024, 4(1): 47-53.

[4] 史韶杰, 韩建宁, 李大威,等. 基于多尺度曝光融合的医学 CT 图像增强方法[J]. 电子测量技术, 2022, 45(6): 106-111.
SHI SH J, HAN J N, LI D W, et al. Medical CT image enhancement method based on multi-scale exposure fusion [J]. Electronic Measurement Technology, 2022, 45(6): 106-111.

[5] 刘肇隆, 范馨月. 基于全尺度跳跃连接的 TransUNet 医学图像分割网络[J]. 国外电子测量技术, 2023, 42(11): 42-48.
LIU ZH L, FAN X Y. TransUNet medical image segmentation network based on full-scale jump connectivity [J]. Foreign Electronic Measurement Technology, 2023, 42(11): 42-48.

[6] 禹文明, 刘伟, 张其超. CT 图像融合专家知识的肺结节良恶性诊断方法[J]. 国外电子测量技术, 2023, 42(7):181-187.
YU W M, LIU W, ZHANG Q CH. CT image fusion expert knowledge-based method for benign [J]. Foreign Electronic Measurement Technology, 2023, 42(7): 181-187.

[7] 喻殿智, 张欣, 迟杏. 基于 CA-DenseNet 的乳腺癌病理图像识别[J]. 国外电子测量技术, 2022, 41(5): 137-143.
YU D ZH, ZANG X, CHI X. A pathological image recognition of breast cancer based on CA-DenseNet[J]. Foreign Electronic Measurement Technology, 2022, 41(5): 137-143.

[8] 李华, 杨嘉能, 刘凤, 等. 基于深度学习的乳腺癌病理图像分类研究综述[J]. 计算机工程与应用, 2020, 56(13):1-11.
LI H, YANG J N, LIU F, et al. Survey of breast cancer histopathology image classification based on deep learning [J]. Computer Engineering and Applications, 2020, 56(13):1-11.

[9] 陈余, 荆慧. 基于深度学习的超声影像组学在乳腺癌中的研究进展[J]. 肿瘤学杂志, 2022, 28(9):730-735.
CHEN Y, JING H. Advances in application of ultrasound-based deep learning radiomics in breast cancer[J]. Journal of Chinese Oncology, 2022, 28(9): 730-735.

[10] HE K M, CHEN X L, XIE S N, et al. Masked autoencoders are scalable vision learners[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 16000-16009.

[11] PANIGRAHY S, KARMAKAR S. Hydrophobicity classification of polymeric insulators using a masked

autoencoder model in vision transformer [J]. Computers and Electrical Engineering, 2024, 116: 109165-109177.

[12] SHABAN W M. Breast cancer classification based on artificial intelligence: state of the art[J]. Nile Journal of Communication and Computer Science, 2023, 6: 1-28.

[13] SUN L C, LIAN ZH, LIU B, et al. HiCMAE: Hierarchical contrastive masked autoencoder for self-supervised audio-visual emotion recognition[J]. ArXiv preprint arXiv: 2401.05698,2024.

[14] WANG Y, NI D, HUANG ZH Y, et al. A self-supervised learning framework based on masked autoencoder for complex wafer bin map classification[J]. Expert Systems with Applications, 2024, 249: 123601-123612.

[15] BAI Y H, LI W Q, AN J P, et al. Masked autoencoders with handcrafted feature predictions: Transformer for weakly supervised esophageal cancer classification[J]. Computer Methods and Programs in Biomedicine, 2024, 244: 107936-107948.

[16] LU M K, WANG T Y, XIA Y. Multi-modal pathological pre-training via masked autoencoders for breast cancer diagnosis[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2023: 457-466.

[17] BENHAMMOU Y, ACHCHAB B, HERRERA F, et al. BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights[J]. Neurocomputing, 2020, 375: 9-24.

[18] SEPAHVAND M, ABDALI-MOHAMMADI F. Overcoming limitation of dissociation between MD and MI classifications of breast cancer histopathological images through a novel decomposed feature-based knowledge distillation method [J]. Computers in Biology and Medicine, 2022, 145: 105413-105426.

[19] BAYRAMOGLU N, KANNALA J, HEIKKILÄ J. Deep learning for magnification independent breast cancer histopathology image classification [C]. 2016 23rd International Conference on Pattern Recognition, 2016: 2440-2445.

[20] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale [J]. ArXiv preprint arXiv:2010.11929,2020.

[21] LI X, WANG W H, YANG L F, et al. Uniform masking: Enabling mae pre-training for pyramid-based vision transformers with locality[J]. ArXiv preprint arXiv:2205.10063,2022.

[22] WANG W H, XIE E Z, LI X, et al. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions [C]. IEEE/CVF International Conference on Computer Vision, Montreal, 2021: 568-578.

[23] BA J L, KIROS J R, HINTON G E. Layer normalization [J]. ArXiv preprint arXiv: 1607.06450,2016.

[24] WU H P, XIAO B, CODELLA N, et al. CVT: Introducing convolutions to vision transformers[C]. IEEE/CVF International Conference on Computer Vision, Montreal, 2021: 22-31.

[25] DENG W Y, LENDASSE A, ONG Y S, et al. Domain adaption via feature selection on explicit feature map [J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 30(4): 1180-1190.

[26] XIANG Z J, ZENG X Y, LIN D, et al. Optimizing implementations of linear layers [J]. IACR Transactions on Symmetric Cryptology, 2020, (2): 120-145.

[27] SPANHOL F A, OLIVEIRA L S, PETITJEAN C, et al. A dataset for breast cancer histopathological image classification [J]. IEEE Transactions on Biomedical Engineering, 2015, 63(7): 1455-1462.

[28] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization [J]. ArXiv preprint arXiv: 1711.05101,2019.

[29] LI Y ZH, DENG K Q. Breast cancer identification study using improved VGG[C]. 2023 8th International Conference on Cloud Computing and Big Data Analytics, 2023: 467-470.

[30] BEHAR N, SHRIVASTAVA M. ResNet50-based effective model for breast cancer classification using histopathology images[J]. CMES-Computer Modeling in Engineering & Sciences, 2022, 130(2):823-839.

[31] BACHMANN R, MIZRAHI D, ATANOV A, et al. Multimae: Multi-modal multi-task masked autoencoders[C]. European Conference on Computer Vision, 2022: 348-367.

[32] IKEZOGWO W O, SEYFIOGLU M S, SHAPIRO L. Multi-modal masked autoencoders learn compositional histopathological representations[J]. ArXiv preprint arXiv:2209.01534,2022.

[33] XU J SH, STIRENKO S. Self-supervised model based on masked autoencoders advance CT scans classification[J].

ArXiv preprint arXiv:2210.05073,2022.

[34] QUAN H, LI X Y, CHEN W X, et al. Global contrast masked autoencoders are powerful pathological representation learners [J]. ArXiv preprint arXiv:2205.09048,2023.

[35] SHEIKH T S, LEE Y H, CHO M Y. Histopathological classification of breast cancer images using a multi-scale input and multi-feature network[J]. Cancers, 2020, 12(8): 2031.

[36] KHAN S U R, ZHAO M, ASIF S, et al. GLNET: Global-local CNN′s-based informed model for detection of breast cancer categories from histopathological slides [J]. The Journal of Supercomputing, 2023,80:7316-7348.

作者简介

司嘉龙,硕士,主要研究方向为医学图像处理与分析。

E-mail:s990318@163.com

贾伟(通信作者),博士,副教授,主要研究方向为医学图像处理与分析、计算病理学。

E-mail:jiawnx@163.com

赵雪芬,博士,副教授,主要研究方向为医学图像处理与分析、机器学习。

E-mail:snownfen@163.com

高宏娟,博士,副教授,主要研究方向为图像处理与分析、计算机视觉、机器学习。

E-mail:4349996@qq.com