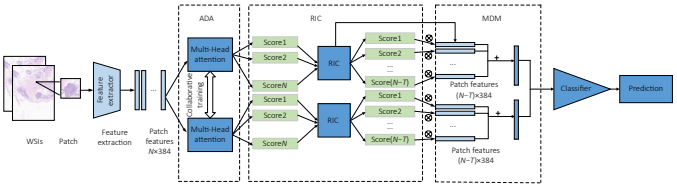
OEE | 光电工程 Opto-Electronic Engineering

Article

2026 年 3 月, 第 53 卷, 第 3 期

基于混合蒸馏的弱监督病理图像识别研究



陈金令[1], 刘　蓉[1], 唐卓葳[2*], 柯　琦[3], 季语祝[3], 高子清[3]

[1]西南石油大学电气信息学院，四川 成都 610500；
[2]电子科技大学医学院附属绵阳医院，绵阳市中心医院乳腺外科，四川 绵阳 621000；
[3]电子科技大学医学院附属绵阳医院，绵阳市中心医院病理科，四川 绵阳 621000

摘要：在医学图像智能诊断领域中，深度学习模型的检测性能往往依赖大规模高质量的人工标注数据集。鉴于数字病理全切片扫描图像 (whole slide image, WSI) 多尺度、大尺寸和标注困难等特点，现有算法大多基于多示例学习 (multiple instance learning, MIL) 方法，该方法只需图像级标签就能实现 WSI 分类。然而，MIL 依然面临着注意力分配过于集中和模型过拟合等问题。本文提出了一种基于混合蒸馏训练策略的多示例学习模型。首先基于自监督学习的 Vision Transformer-S/16 进行预训练，降低对标注数据的依赖。然后提出了一种双路自适应注意力和随机示例屏蔽相结合的混合蒸馏模型：该模型利用自适应注意力模块实现了对纹理特征关注程度的动态调整，并对纹理间的相关性进行分析；采用随机示例屏蔽方法，随机屏蔽具有高注意力分数的示例，提高了对关键示例的关注程度；引入混合蒸馏机制，该机制融合随机屏蔽后剩余示例的特征信息。最后，利用分类器对整个包的阴阳性进行预测，实现乳腺癌全切片病理图像的高精度分类。实验结果表明，该模型在 Camelyon16 数据集上，其 Acc、AUC、F1-score 值分别达到 98.445%、0.984、0.983，在 BRACS 数据集上达到 77.852%、0.905、0.764，显著提升了检测性能。

关键词：图像分类；自监督学习；多示例学习；随机示例屏蔽；自适应注意力；混合蒸馏

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250258 | CSTR: 32245.14.oee.2026.250258 | 中图分类号：TP391.4 | 文献标识码：A

陈金令，刘蓉，唐卓葳等. 基于混合蒸馏的弱监督病理图像识别研究 [J]. 光电工程，2026，**53**(3): 250258

Chen J L, Liu R, Tang Z W et al. Research on whole slide image recognition based on weakly supervised learning and mixed distillation mechanism[J]. *Opto-Electron Eng*, 2026, **53**(3): 250258

1　引　言

癌症已成为全球范围内人类死亡的主要原因之一，其中乳腺癌是女性中常见的癌症，据世界卫生组织的调查研究显示，每年全球大约有 200 万女性乳腺癌新增病例, 占癌症病例总数的 11.7%，占女性癌症病例总数的 24.5%。乳腺癌也是导致女性死亡人数最多的癌症，占女性每年癌症死亡总数的 15.5%[1-2]。因此乳腺癌的早期预防、诊断以及治疗就显得尤为重要。传统的乳腺癌诊断方法依赖病理医生通过显微镜观察切片图像来定位病

收稿日期：2025-08-30；修回日期：2025-11-26；录用日期：2025-11-28；网络出版日期：2026-03-19

基金项目：华南恶性肿瘤防治全国重点实验室开放课题 (HN2025-14)；四川省重点研发计划 (重大科技专项) 项目 (2022YFS0020)；四川省卫生健康委员会研究者发起的临床研究项目 (23LCYJ020)；绵阳市中心医院临床专项基金 (2025LC014)；南充市 2022 年市校科技战略合作专项 (22SXQT0292)

*通信作者：唐卓葳，tungtung120928@163.com

版权所有©2026 中国科学院光电技术研究所

灶区域，并进行病理分类。然而这种传统方式存在很多局限性：首先它对于病理医生的专业经验和技能要求较高，而具备这些资质的病理医生相对稀缺；其次，由于全切片扫描图像 (whole slide image, WSI) 的尺寸和分辨率通常较大，整个诊断过程耗时耗力；最后，仅通过人工观测进行诊断与分类具有一定的主观性并且存在误诊的风险。在乳腺癌病例数量持续增长的趋势下，大众对于引入更为高效、精准的计算机辅助诊断工具的需求愈发迫切。

近年来，深度学习在计算机视觉领域中有了显著的发展，特别是在临床癌症诊断方面，通过计算机辅助 WSI 分类受到很多病理学家的关注[3]。目前有许多研究者致力于研究不同的计算机辅助诊断方法以辅助病理学家进行更高效的疾病诊断。然而 WSI 通常缺乏像素级的详细标注，限制了深度学习模型在图像块级上的直接训练和应用。为了解决这一问题，弱监督多示例学习成为分析 WSI 的主流方法[4-7]。其中基于注意力的多示例方法也是本文研究的内容之一，其通过引入注意力机制来增强模型对数据中关键特征的识别能力。例如 Li 等[8] 将局部注意力引入模型中提出了 DSMIL (dual-stream multiple instance learning networ)，该模型通过测量每个示例与得分最高的示例之间的距离，进而为每个示例分配不同的注意力；Shao 等[9] 提出的 TransMIL 是基于 Transformer 的分类模型，该模型首次考虑了在同一个包中不同示例之间的关联性；ABMIL (Attention-based multiple instance learning) 通过聚合每个示例的注意力分数从而表示整个切片的特征[10]，这种方法取得了不错的效果；此外，还有研究如 CLAM[11] 等，通过为示例评分并将注意力分数与特征结合进一步优化了模型的性能。这些研究工作极大推动了病理图像分类技术的发展。

然而目前主流的弱监督多示例方法仍然面临着以下挑战： 1) 在设计多示例分类模型时通常假设示例间是相互独立且同分布的，但在许多情况下，这种假设并不总是成立[10]，因为示例间是相互联系且具有相关性。这种设计上的简化可能导致模型在处理不同类型示例时，无法有效捕捉它们之间的关系，进而忽视了单个区域与其周围环境之间以及不同区域之间的相互关联信息； 2) 多数现有模型在预测时往往只关注部分简单的示例特征。然而 WSI 因其庞大的尺寸、广泛的覆盖范围和极高的分辨率，包含了极为复杂和丰富的信息。如果模型仅基于简单示例的特征进行训练，它们可能无法适应各种复杂的示例，容易出现过拟合现象[12-13]。这种复杂性要求模型不仅要关注局部特征，还要能够理解和整合全局信息，以提高泛化能力。否则，分类模型的泛化能力将受到限制，增加了过拟合的风险。

综上，本文提出了基于混合蒸馏策略的多示例学习模型 (mixed distillation mechanism multi-instance learning, MDM-MIL)。相比现有的分类算法进行了多项改进与创新设计，主要体现在以下三部分：

1) 自适应注意力 (Adaptive attention, ADA) 模块，该模块包括自适应门控注意力子模块和自适应多头注意力子模块，能够分析输入特征的相关性，捕获边界纹理特征，确保不同的分支学习到不同的纹理特征，提高了预测结果的准确性。

2) 随机示例屏蔽 (Random instance concealing, RIC)，该方法能将注意力更均匀地分配给没有被关注到的示例，提高模型对全局信息的理解能力。

3) 混合蒸馏机制 (Mixed distillation mechanism, MDM)，该机制融合两个分支经过 RIC 方法之后剩余示例的特征信息，提高了模型的学习效率和分类准确性。

2 研究方法

本文提出的 MDM-MIL 总体结构如图 1 所示。首先采用自监督学习 (self-supervised learning, SSL) 方法[14-15] 预训练的 Vision Transformer-S/16 (ViT-S/16) 网络提取图像块的特征，生成相应的特征向量。然后利用本文提出的 MDM-MIL 算法对这些特征向量进行处理： 1) 通过 ADA 模块对不同纹理的示例特征进行评分； 2) 使用 RIC 方法将注意力分数排在前 T 的示例进行随机屏蔽，并将剩余示例的特征信息进行聚合； 3) 利用混合蒸馏策略融合双路特征信息，利用分类器对整个包的阴阳性进行预测。

2.1 数据预处理

在病理切片图像分类任务中，WSI 尺寸巨大且包含大量区域，这种大尺寸的图像在当前硬件条件下是无法直接被处理的，因此基于图像块的处理方法[3] 成为高维病理图像的主流做法。基于此，本文的数据预处理过程分为以下三步：首先是前景分割，采用 OTSU 算法从 WSI 中分割出病理学家感兴趣的区域；其次对分割出的感兴趣区域进行裁剪，每个 WSI 被裁剪成尺寸为 224×

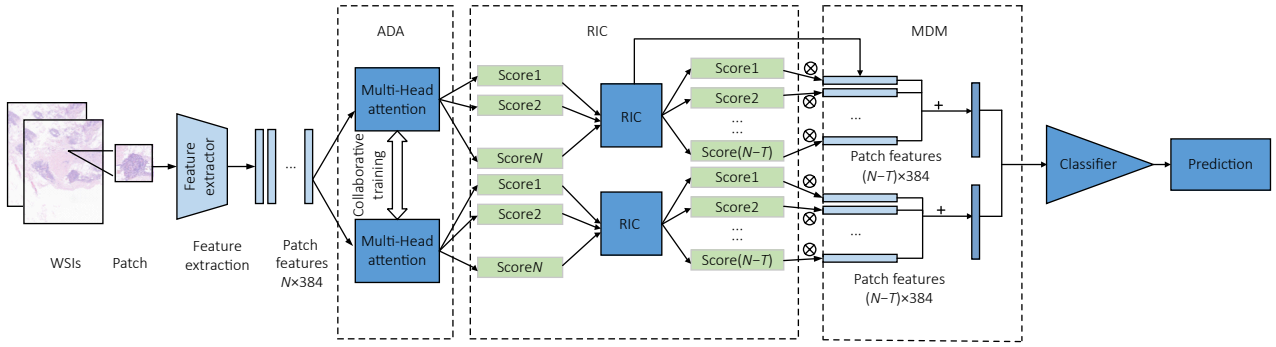


图 1 MDM-MIL 的总体结构图

Fig. 1 Overall structure of the MDM-MIL model

224 不重叠的图像块[16]。本文将 WSI 视为多示例学习中的一个包，将图像块视作一个示例。

在二元分类问题中，$\boldsymbol{B}=\{(x_1,y_1),\cdots,(x_n,y_n)\}$是一个包，其中$x_n\in x$，是标签$y_n\in\{0,1\}$的示例，包的标签 Y 定义如下

$$Y=\begin{cases}0, & \sum_{i=1}^{N}y_n=0\\ 1, & \text{otherwise}\end{cases}. \tag{1}$$

这里数据集的标签是针对整张图片的，因此提取出的图像块并没有标签。最后，使用特征提取器来提取示例的特征，本文学习示例特征的模型是基于 SSL 框架 DINO 预训练的 VIT-S/16 模型。

2.2 自适应注意力模块

在病理图像分析中，图像的纹理特征具有显著的多样性，不同的纹理特征能够反映出不同的病理状态、组织类型或病变阶段。然而神经网络在训练过程中，为了最小化训练损失，更倾向于学习相对简单的示例特征，这可能导致模型在处理复杂纹理或结构特征的示例时表现不佳[17-19]，除此之外，许多模型没有考虑示例间的相关性。因此本文提出了 ADA 模块，分析输入特征信息的相关性，避免学习到重复的特征信息，更好地捕获边界纹理特征。

如图 2 所示，本文提出的 ADA 模块核心思想可以概括为以下几个部分：

1) 双路特征识别：第一路采用自适应门控注意力子模块，两个自适应门控注意力分支允许模型动态调整对纹理特征的关注程度，从而更准确地区分病变组织与正常组织的交界处的纹理特征，如边界的不规则性、病变区域的侵入性等；第二路采用自适应多头注意力子模块，处理自适应多头注意力分支信息，分析特征信息的相关性。这种自适应方法不仅增强了模型对局部特征的识别能力，还显著提升了模型对全局上下文的理解。具体来说，每个注意力头独立地计算其对应的查询 (query，$\boldsymbol{Q}$)、键 (key，$\boldsymbol{K}$) 和值 (value，$\boldsymbol{V}$)，然后通过动态调整的权重将这些结果合并，最终获得综合的特征表示。

2) 特征嵌入与聚合：将不同自适应注意力分支获取到的特征转换成低维嵌入向量，并将这些嵌入向量有效聚合，可以得到包嵌入，整合来自不同分支的信息。

3) 语义正则化：为增强模型对特征的理解，在每个分支的嵌入表示后引入一个多层感知机 (MLP) 层，实现

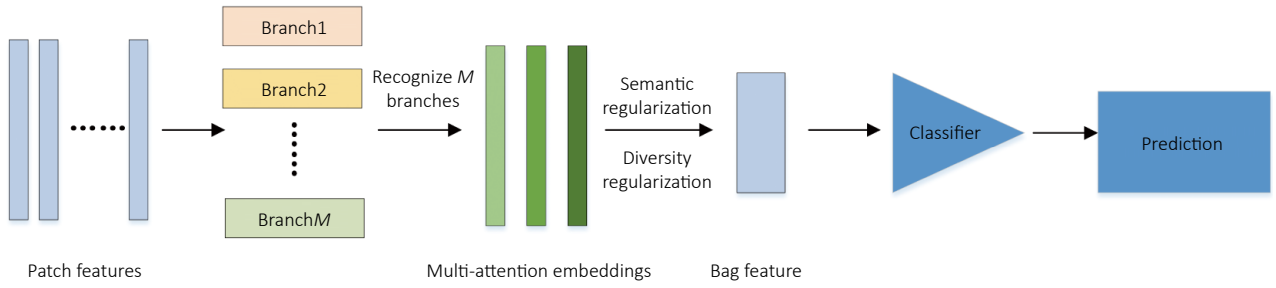


图 2 自适应注意力的结构图

Fig. 2 Structure of the adaptive attention module

语义的正则化，进一步提炼特征信息。

4) 包的注意力分数计算：为汇总所有分支的预测，将所有分支的注意力分数取平均，作为整个包的分数。

$$a=\frac{1}{M}\sum_{i=1}^{M}a_i\,, \tag{2}$$

式中：M 表示自适应注意力分支数；a_i 表示第 i 个分支的所有关注值，$a_i=\{a_{i1},a_{i2},\cdots,a_{in}\}$；$a$ 表示整个包的分数，维度为 N。

5) 多损失函数联合优化：模型的训练采用多个损失函数联合优化的方式进行。

交叉熵损失函数：用来衡量预测与实际标签之间的差异，促使模型学习准确的分类边界：

$$L_{\mathrm{p}}=-\frac{1}{M}\sum_{i=1}^{M}Y\log_2\hat{Y}_i+(1-Y)\log_2\left(1-\hat{Y}_i\right), \tag{3}$$

式中：$\hat{Y}=g_i(z_i)$ 是对第 i 个分支嵌入 z_i 的预测。

多样性损失函数：鼓励模型学习不同纹理的特征并挖掘更多的预测信息，增强模型对多样化信息的敏感性。该函数通过度量不同注意力分支之间关注值的相似性来避免学习到相似的特征。

$$L_{\mathrm{d}}=\frac{2}{M(M-1)}\sum_{i=1}^{M}\sum_{j=i+1}^{M}\cos(a_i,a_j)\,, \tag{4}$$

式中：a_j 表示第 j 个分支的所有关注值；$\cos(\cdot)$ 函数用于度量注意力分支的分数值。

包分类器损失函数：可以确保在汇总分支预测时能够获得准确的包分数。

$$L_{\mathrm{b}}=-Y\log\hat{Y}+(1-Y)\log(1-\hat{Y})\,. \tag{5}$$

模型的训练是通过综合考虑上述三个损失函数来进行的，即

$$L=L_{\mathrm{p}}+L_{\mathrm{d}}+L_{\mathrm{b}}\,. \tag{6}$$

2.3 随机示例屏蔽方法

在目前的研究中，注意力机制主要关注注意力分数非常高的示例，为了缓解这种情况，本文提出 RIC 方法。该方法是对关注度高的部分示例进行随机屏蔽操作，然后将注意力分配给其它示例，提高模型对整个 WSI 图像特征的学习。

如图 3 所示，RIC 方法的具体实施步骤如下：

1) 注意力分数的生成：首先，利用 ADA 模块生成每个示例的注意力分数。

2) 示例的排序与屏蔽：生成注意力分数后，本文将所有示例的注意力值从高到低进行排序，排在前 T 的示例的关注值随机设置为 0，概率为 p，此过程可以表示为

$$a_n=\begin{cases}0, & \text{前}T\text{个任意示例的值, 概率为}p\\ a_n, & \text{其它}\end{cases}, \tag{7}$$

式中：p 和 T 是两个控制屏蔽强度的超参数。这一操作的目标是降低这些示例的关注度，以便将注意力转移到其他示例上。

3) 注意力分数的重新调整：在执行随机屏蔽操作之后，为了保持所有示例的关注度之和为 1，即 $\sum_{n=1}^{N}a_n=1$，对剩余的关注度进行归一化处理：$a_n\to1/\left(\sum_{n=1}^{N}a_n\right)a_n$，从而确保维护模型在关注度分配上的一致性。

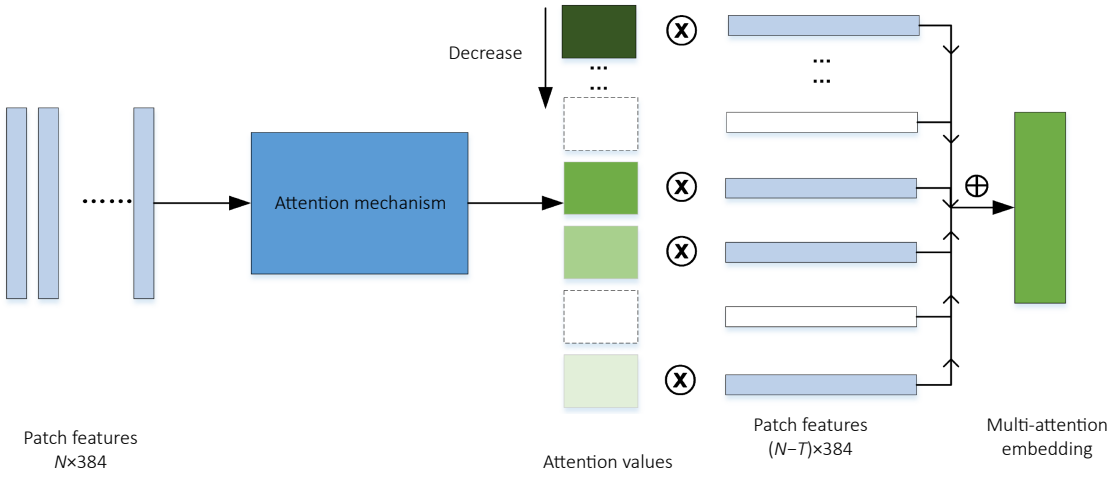


图 3 随机示例屏蔽的结构图

Fig. 3 Structure of the random instance concealing module

2.4 混合蒸馏策略

如图 1 所示，MDM-MIL 由上下两路构成，第一路由自适应门控注意力子模块计算出所有示例的注意力分数，然后对经过随机示例屏蔽调整后的注意力分数进行聚合操作，整合剩余的示例信息，这个过程可以表示为

$$z_1 = \sum_{n=1}^{N-T} a_n h_n \,, \tag{8}$$

式中：a_n代表第 n 个示例h_n的注意力分数。注意力分数a_n的计算方法具体如下：

$$a_n = \frac{\exp\{\boldsymbol{w}^{\mathrm{T}}(tanh(\boldsymbol{V}_1 h_n) \odot sigm(\boldsymbol{V}_2 h_n))\}}{\sum_{j=1}^{N} \exp\{\boldsymbol{w}^{\mathrm{T}}(tanh(\boldsymbol{V}_1 h_j) \odot sigm(\boldsymbol{V}_2 h_j))\}}, \tag{9}$$

式中：$\boldsymbol{w}$、$\boldsymbol{V}_1$、$\boldsymbol{V}_2$表示参数矩阵，$\boldsymbol{w} \in \mathbb{R}^{L\times 1}$，$\boldsymbol{V}_1$、$\boldsymbol{V}_2 \in \mathbb{R}^{L\times M}$；$\odot$表示元素级乘法；$tanh(\cdot)$和$sigm(\cdot)$表示非线性激活函数；上标 T 表示矩阵转置操作。

第二路采用自适应多头注意力子模块计算出所有示例的注意力分数，然后对经过随机示例屏蔽调整后的注意力分数进行聚合操作，整合剩余的示例信息，这个过程可以表示为

$$\boldsymbol{A} = softmax\left(\frac{\boldsymbol{Q}\cdot\boldsymbol{K}^{\mathrm{T}}}{\|\boldsymbol{Q}\|\,\|\boldsymbol{K}\|} \Big/ \sqrt{d_{\mathrm{k}}}\right), \tag{10}$$

$$z_2 = \boldsymbol{A}\cdot\boldsymbol{V}\,, \tag{11}$$

式中：$\boldsymbol{Q} = \boldsymbol{W}_{\mathrm{q}}\boldsymbol{H}$，$\boldsymbol{K} = \boldsymbol{W}_{\mathrm{k}}\boldsymbol{H}$，$\boldsymbol{V} = \boldsymbol{W}_{\mathrm{v}}\boldsymbol{H}$；$d_{\mathrm{k}}$代表键 $\boldsymbol{K}$ 的维度；平方根项$\sqrt{d_{\mathrm{k}}}$代表比例因子；$\boldsymbol{W}_{\mathrm{q}}$、$\boldsymbol{W}_{\mathrm{k}}$、$\boldsymbol{W}_{\mathrm{v}}$代表可学习参数矩阵；$\boldsymbol{H}$=$\{h_1, h_2, \cdots, h_n\}$代表特征矩阵，采用混合蒸馏策略融合经过 RIC 方法之后剩余示例聚合后的特征信息，形成包级表示：

$$\boldsymbol{Z} = cat(z_1 + z_2)\,, \tag{12}$$

式中：cat代表$concat(\cdot)$；$\boldsymbol{Z} \in \mathbb{R}^{2\times d}$，其中 d 代表特征向量的维度。最后使用分类器 Y 基于融合后的特征 $\boldsymbol{Z}$ 推测出最终的预测标签 X：

$$X = Y(\boldsymbol{Z})\,. \tag{13}$$

3 实　验

3.1 数据集及预处理

本文采用 Camelyon16[20] 和 BRACS[21] 两个公共数据集来验证所提方法的有效性, 这两个数据集均为经过苏木精和伊红 (H&E) 染色并附有专业注释的图像集。Camelyon16 数据集包括 400 张 WSI 图片，其中 270 张用于训练，130 张用于测试。为了确保模型的有效性和泛化能力，将训练集以 9:1 比例随机分割为训练集和验证集[22]。BRACS 数据集已经由官方进行明确的拆分，包含 395 个训练集、65 个验证集和 87 个测试集。BRACS 数据集包括三种关键的病变类型：良性、恶性和非典型。在预处理阶段，执行以下具体操作以确保数据质量：首先，利用 OTSU 分割方法裁剪掉不包含任何病理信息的背景图像，随后将数据集中的每个 WSI 裁剪成互不重叠的图像块，以便于模型训练和特征提取。这些预处理步骤有助于提高模型的效率和准确性，使其能够更专注于学习关键的病理特征。

3.2 实验环境及评价指标

本文中所有实验都是基于 PyTorch 库，为了在不同的聚合算子之间进行全面的比较，使用两组不同的特征作为输入，利用 ResNet-18 对 ImageNet 数据集进行训练，DINO[23] 预训练的 VIT-S/16 网络对 WSI 数据集[24] 进行训练。BatchSize 大小为 1，使用恒定学习率为 0.0001 的 Adam[25] 优化器在训练过程中更新模型权值。使用 Adam 优化器迭代更新 100 轮，学习率为10^{-4}，权重衰减为10^{-5}，操作系统为 Windows 10，GPU 为 NVIDIA GeForce 3080Ti GPU。

为更准确地评估模型性能，本文采用三项常用指标：准确率 (accuracy, Acc)、曲线下面积 (area under the ROC curve, AUC) 与 F1 分数 (F1-score, F_1)。准确率 (Acc) 反映模型整体预测正确的比例，定义为

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}\,, \tag{14}$$

式中：TP、TN、FP 和 FN 分别表示真阳性、真阴性、假阳性和假阴性样本数。AUC 可衡量模型区分阳性与阴性样本的综合能力，其值越接近 1 表示模型分类性能越优。AUC 对样本类别不平衡具有较好的鲁棒性，因而适用于弱监督病理任务。F1-score 综合考虑模型的精确率 (precision) 与召回率 (recall)，定义为

$$F_1 = 2\times\frac{Precision\times Recall}{Precision + Recall}\,. \tag{15}$$

本文研究任务存在类别分布不均、样本标签弱化的特点，故综合采用 AUC 与 F1-score 作为主要评价指标，以确保性能评估的科学性与全面性。

3.3 对比实验

为验证 MDM-MIL 模型在病理图像分类中的性能，选择 2 种不同的特征提取方法以及 6 种先进方法在

Camelyon16 数据集和 BRACS 数据集上进行对比分析，同时采用 F1-score 和 AUC 指标对其性能进行评估。

对比方法包括平均池化、最大池化、CLAM、DSMIL、TransMIL、DTFD 以及 ABMIL。其中，平均池化和最大池化分别通过对池化区域内的像素值取平均以及选择池化区域内的最大值来降低数据维度，但这两种方法均可能会丢失一些重要的特征信息；CLAM 使用注意力池化的方式获得包级表示，并对 Top-k 示例进行示例级聚类进一步细化特征空间；DSMIL 通过最大池化识别关键示例，进而将每个示例与关键示例的距离作为权重，最后通过加权求和的方式聚合所有示例形成包级表示；TransMIL 利用 Transformer 对长距离关系进行建模，并采用一个随机初始化的 class token 作为查询聚合示例的特征信息；DTFD 通过引入伪包增加包的数量，并在伪包的基础上构建双层多示例框架，以有效地利用内在特征；ABMIL 利用注意力机制聚合示例级别的特征表示，从而生成包级的特征表示。表 1 为采用 ResNet-18 特征提取器时，不同模型在 Camelyon16 和 BRACS 数据集上的对比实验结果。据表 1 可知，在 Camelyon16 数据集上，MDM-MIL 的 F1-score 为 0.823，比第二名的 DTFD 模型高 0.058，AUC 为 0.854，略低于 DTFD；在 BRACS 数据集上，MDM-MIL 的 F1-score 为 0.639，比第二名的 ABMIL 模型高 0.088，AUC 为0.771，与第一名的模型 CLAM 有略微差距，低 0.017。表 2 为采用 VIT-S/16 特征提取器时，不同模型在Camelyon16 和 BRACS 数据集上的对比实验结果。据表 2 可知，在 Camelyon16 数据集上，MDM-MIL 的分类性能均优于其它模型，在 F1-score、AUC 指标上分别高于第二名的 DTFD 模型 0.044，0.008；在 BRACS 数据集上，MDM-MIL 的 F1-score 和 AUC 指标大小也均排在第一，分别比第二名的 DTFD 模型高 0.127 和 0.042。

综上可知，本文所提方法在两项评价指标上明显优于其他方法。这是由于： 1) 平均池化与最大池化会导致一些重要的特征信息被忽略； 2) CLAM 和 ABMIL 都是基于独立同分布假设设计的，这限制了它们在复杂多变的数据环境中的适用性； 3) TransMIL和 DSMIL 虽然考虑到了示例的相关性，但 TransMIL在建模示例相关性时使用的是外部查询，DSMIL 虽然使用的是包内的示例作为查询，但是它在生成内部查询时依赖的是一个简单的线性分类层，导致分类结果不佳。除了本文提出的方法之外，DTFD 的性能最优，它通过引入伪包来增加包的数量，从而使每个伪包内的示例数变少，这样做可以充分利用包内部的特征和示例间的相关性，但在随机划分原始包和创建伪包的过程中，存在引入噪声标签的风险，它与本文提出的方法仍有较大的差距，这也验证了本文提出的模型在特征提取方面具有卓越的性能，特别是采用自监督学习作为预训练方法，模型的特征提取能力得到了进一步的增强，从而显著提升了模型的检测性能。

3.4 消融实验

3.4.1 模块的有效性

为了验证 MDM-MIL 模型各模块的有效性，本文分别在 Camelyon16 和 BRACS 两个数据集上进行消融实

表 1 不同模型在 Camelyon16 和 BRACS 数据集的分类结果对比 (特征提取器为 ResNet-18)

Table 1 Comparison of classification results of different models on the Camelyon16 and BRACS datasets (feature extractor: ResNet-18)

Dataset	Index	Average pooling	Max pooling	CLAM	DSMIL	TransMIL	DTFD	ABMIL	MDM-MIL (ours)
Camelyon16	F1score	0.595	0.584	0.752	0.745	0.672	0.765	0.759	**0.823**
	AUC	0.599	0.625	0.783	0.778	0.745	**0.862**	0.795	0.854
BRACS	F1score	0.495	0.499	0.526	0.523	0.468	0.472	0.551	**0.639**
	AUC	0.693	0.751	**0.788**	0.776	0.768	0.749	0.758	0.771

表 2 不同模型在 Camelyon16 和 BRACS 数据集的分类结果对比 (特征提取器为 VIT-S/16)

Table 2 Comparison of classification results of different models on the Camelyon16 and BRACS datasets (feature extractor: VIT-S/16)

Dataset	Index	Average pooling	Max pooling	CLAM	DSMIL	TansMIL	DTFD	ABMIL	MDM-MIL (ours)
Camelyon16	F1score	0.587	0.905	0.895	0.932	0.912	0.939	0.903	**0.983**
	AUC	0.694	0.932	0.943	0.952	0.934	0.976	0.925	**0.984**
BRACS	F1score	0.513	0.572	0.602	0.552	0.605	0.637	0.635	**0.764**
	AUC	0.732	0.801	0.859	0.796	0.831	0.863	0.841	**0.905**

验，并分别使用 ResNet-18 和 VIT-S/16 特征提取器提取图像特征。为了保证公平性，实验中其他的设置均保持一致，实验结果分别见表 3 和表 4。

表 3 展示了多种模型在 Camelyon16 数据集上使用不同特征提取器的实验结果。通过对比分析，可以看出模型 2 较模型 1 而言，它的 AUC 指标有明显提升，这表明随机屏蔽一部分高注意力值的示例可以提高模型对其它关键示例的关注度；模型 3 较模型 2 在 Acc 和 AUC 上均有显著提升，这表明 RIC 方法和 ADA 模块两者的结合可以使模型在更全面分析示例信息的同时，分析不同示例间的相关性。对比 1、2、3、4 四个模型，模型 4 的指标均排在第一，这表明 RIC 方法、ADA 模块和 MDM 三者的结合能更好地提升模型的检测性能，MDM 将二者学习到的示例信息进行充分融合从而更好地理解示例间的复杂关系；将模型 8 与模型 5、6、7 相比，说明将 RIC 方法、ADA 模块和 MDM 三者结合起来，它们对于模型性能的提升才更显著。最后将模型 5、6、7、8 和 1、2、3、4 进行对比，可以看出各指标均大幅提升，这主要归因于 VIT-S/16 能从大量无标签数据中学习到有用的特征表示并能同时处理图像中的多个区域，分析示例间的相关性信息，从而有效提取全局特征。

表 4 展示了不同模型在 BRACS 数据集上使用不同特征提取器的实验结果，通过对比可以发现尽管模型 2、3 分别与模型 1 相比 AUC 指标有所提升，但 Acc 并未表现出明显的提升，这表明单独使用 RIC 或 ADA 对于模型性能的提升可能不够明显。然而模型 4 较模型 1 在 Acc、AUC 指标上均有显著提升，分别提升 5.813、0.063；模型 5、6、7、8 相比较，模型 8 的指标 Acc 和 AUC 分别为 77.852 和 0.905 均排在第一。综上所述，本文提出的三个方法协同起来，能够显著提升模型的性能，而采用自监督学习作为特征提取器的预训练方法能大大提升模型的各项分类指标。

3.4.2 随机示例屏蔽中超参数的影响

T 和 p 是 RIC 方法中关键的超参数，其中，T 决定被屏蔽示例的个数，p 控制示例被屏蔽的概率。图 4(a) 和图 4(b) 分别展示了在两个数据集上，超参数 T 在不同值下对应的模型性能。可以看出不同 T 值对于模型的各项指标均有一定影响。T 值过大意味着被屏蔽的示例个数过多，这很可能导致一些关键信息的丢失。由于医学图像诊断通常需要依赖多个关键示例进行准确判断，因

表 3 在 Camelyon16 上的消融实验结果

Table 3 Results of ablation experiments on the Camelyon16 dataset

Feature extractor	Model	RIC	ADA	MDM	Acc	AUC
ResNet-18	1	—	—	—	79.070	0.758
	2	√	—	—	77.519	0.805
	3	√	√	—	81.395	0.827
	4	√	√	√	**84.496**	**0.854**
VIT-S/16	5	—	—	—	93.798	0.958
	6	√	—	—	89.147	**0.986**
	7	√	√	—	89.152	0.968
	8	√	√	√	**98.445**	0.984

表 4 在 BRACS 上的消融实验结果

Table 4 Results of ablation experiments on the BRACS dataset

Feature extractor	Model	RIC	ADA	MDM	Acc	AUC
ResNet-18	1	—	—	—	58.140	0.708
	2	√	—	—	58.138	**0.785**
	3	√	√	—	56.977	0.753
	4	√	√	√	**63.953**	0.771
VIT-S/16	5	—	—	—	67.442	0.811
	6	√	—	—	75.581	0.877
	7	√	√	—	72.093	0.884
	8	√	√	√	**77.852**	**0.905**

此过多的屏蔽会导致模型性能下降；相反，T 值过小虽然能够保留更多的示例信息，但也可能导致模型过于关注注意力分数排在前面的关键示例，而忽略了其他同样含有重要信息的示例，这种情况下，模型可能无法全面捕获图像中的有用信息，从而导致分类性能下降。综合模型的三项指标而言，当 T=9 时，模型的性能最佳。在这个超参数的设置下，模型既能够屏蔽掉一部分示例以避免模型对于部分示例的过度关注，又能够保留足够的关键信息以维持较高的诊断准确性。

图 5 展示了在两个数据集上，超参数 p 在不同值下对应的模型性能。可以看出在不同数据集上 p 的最优取值是不同的。图 5(a) 展示的是在 Camelyon16 数据集上不同 p 值下三项指标的变化情况：当 p=0.6 时，虽然 AUC 值达到最高，但 Acc 和 F1-score 的值均不理想；当 p=0.5 时，Acc 和 F1-score 两项指标均达到峰值，在这个设置下，模型既能够保留足够的关键信息，又避免了对非关键示例的过度关注。因此，在 Camelyon16 数据集上设置 p 为 0.5 时，模型的性能最优。图 5(b) 展示的是在 BRACS 数据集上不同 p 值对于模型的影响，当 p=0.7 时模型的各项指标均是最优的。

综上，在不同数据集上超参数 p 的设置对于模型的性能会有一定的影响。p 设置得过大会导致大部分关键示例被屏蔽，从而丢失大部分分类的关键信息；p 设置得过低又会导致 RIC 方法失去意义，模型还是会将大部分注意力放在注意力分数排在前面的示例上。因此在实际应用中，需要根据数据集的特点和模型的性能需求来选择合适的 p 值。

3.4.3 自适应多注意力模块中超参数 M 的影响

M 是 ADA 模块中重要的超参数，它表示自适应多注意力模块捕获的示例的分支数量。图 6(a) 显示，在 Camelyon16 数据集上，不同 M 的取值对于模型的三项指标有着显著的影响，由此可以得出结论：在 Camelyon16 数据集上，M=5 时模型的各项指标均较高，若 M 取值较大 (>5) 时，模型的 Acc 和 F1-score 则均呈现下滑趋势，

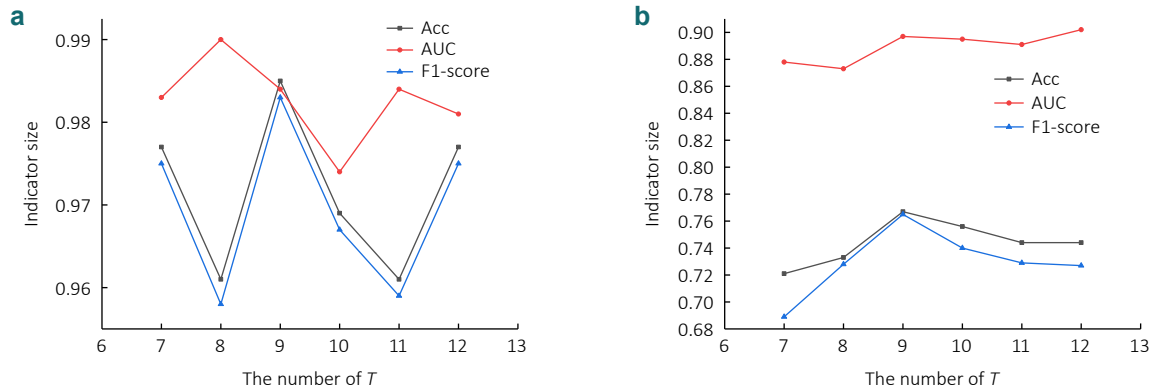


图 4 在不同数据集上超参数 T 对模型性能的影响。(a) Camelyon16；(b) BRACS

Fig. 4 The influence of hyperparameter T on model performance on different datasets. (a) Camelyon16; (b) BRACS

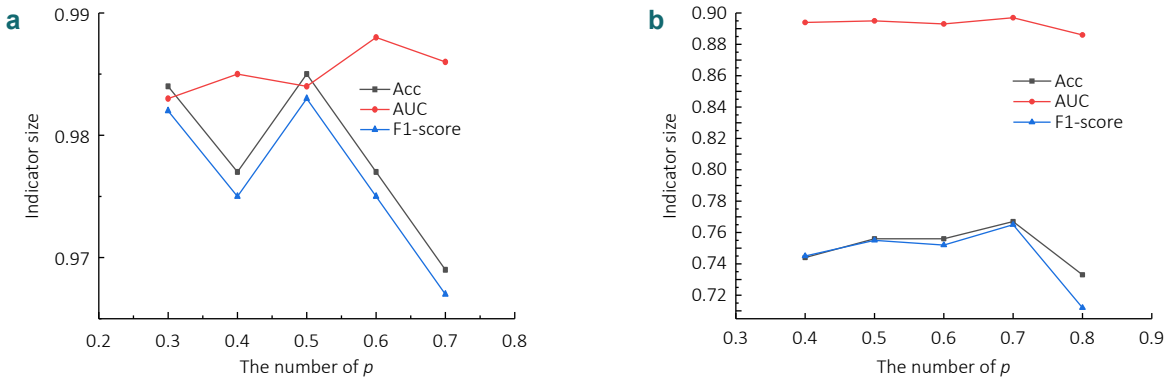


图 5 在不同数据集上超参数 p 对模型性能的影响。(a) Camelyon16；(b) BRACS

Fig. 5 The influence of hyperparameter p on model performance on different datasets. (a) Camelyon16; (b) BRACS

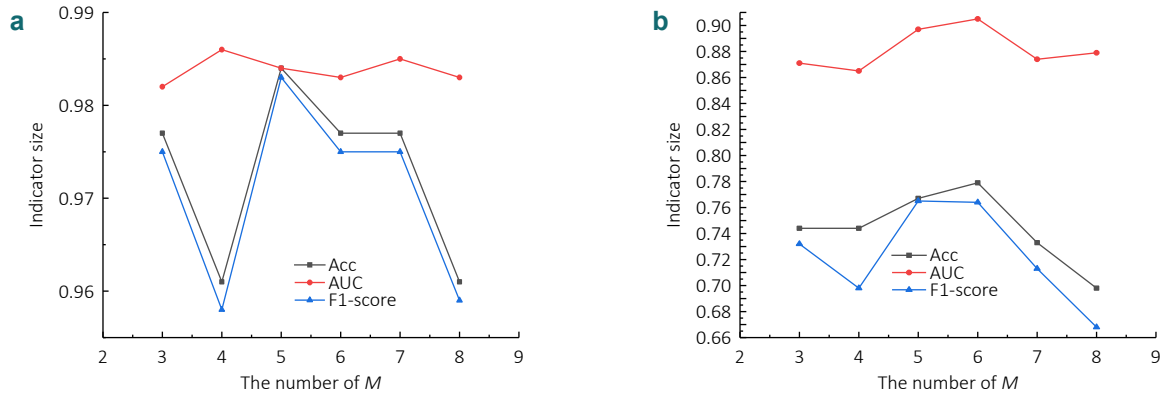


图 6 在不同数据集上超参数 *M* 对模型性能的影响。(a) Camelyon16；(b) BRACS

Fig. 6 The influence of hyperparameter *M* on model performance on different datasets. (a) Camelyon16; (b) BRACS

当 *M* 小于 5 时，虽然 AUC 的值能达到其最高值，但是 Acc 和 F1-score 均不能达到其峰值。进一步观察图 6(b)，在 BRACS 数据集上，*M*=6 时模型的性能最佳。

为提高模型的可复现性与参数选择的系统性，本文在超参数调优过程中采用了网格搜索 (grid search) 结合经验性调整的策略。具体而言，首先在参数初始区间内 ($T\in[5,15]$，$p\in[0.3,0.8]$，$M\in[3,8]$) 进行粗搜索，以获得性能上升趋势；随后在局部最优区间内进行细粒度搜索，以确定最终取值。同时，为评估模型对超参数扰动的敏感性，对最优参数附近 ±20% 区间进行了性能评估。结果表明，AUC 与 F1-score 波动均小于 2%，表明模型在参数微调下具有较高的鲁棒性与稳定性。

4 结 论

针对现有多示例学习中注意力分配过于集中和模型过拟合的问题，本文提出一种基于混合蒸馏策略的多示例学习模型。本模型结合协同训练的双路自适应注意力模型与随机示例屏蔽方法，最终利用混合蒸馏策略拼接双路的聚合结果，模型在考虑到多种不同类型示例间相关性的情况下还能充分学习到更多关键示例的特征信息，最终提高模型的性能并且缓解模型过拟合现象。在 Camelyon16 和 BRACS 数据集上的各项实验结果表明，本模型和其它先进的多示例学习模型相比，最终的分类指标都明显优于其它模型，这表明本模型在病理图像分类上更具有优势。如何将模型轻量化，以及如何将其应用到其它不同种类的医学图像分类任务中是未来工作的重要研究内容。

作者贡献

陈金令：文章撰写、定稿，确定整体技术路线；刘蓉：设计试验方案，实验验证；唐卓葳：确定研究方案，病理图像分析；柯琦，季语祝，高子清：病理图像分析。

利益冲突

所有作者声明无利益冲突

参考文献

1. Alotaibi A, Alafif T, Alkhilaiwi F et al. ViT-DeiT: an ensemble model for breast cancer histopathological images classification[C]// *2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC)*, 2023: 1–6. https://doi.org/10.1109/ICAISC56366.2023.10085467.
2. Sung H, Ferlay J, Siegel R L et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. *CA Cancer J Clin*, 2021, **71** (3): 209−249.
3. Cruz-Roa A, Basavanhally A, González F et al. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks[J]. *Proc SPIE*, 2014, **9041**: 904103.
4. Campanella G, Hanna M G, Geneslaw L et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images[J]. *Nat Med*, 2019, **25** (8): 1301−1309.
5. Amores J. Multiple instance classification: Review, taxonomy and comparative study[J]. *Artif Intell*, 2013, **201**: 81−105.
6. Chen Zenghai, Chi Zheru, Fu Hong et al. Multi-instance multi-label image classification: a neural approach[J]. *Neurocomputing*, 2013, **99**: 298−306.
7. Dietterich T G, Lathrop R H, Lozano-Pérez T. Solving the multiple instance problem with axis-parallel rectangles[J]. *Artif Intell*, 1997, **89** (1-2): 31−71.
8. Li B, Li Y, Eliceiri K W. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning[C]//*IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 14313–14323. https://doi.org/10.1109/CVPR46437.2021.01409.
9. Shao Z C, Bian H, Chen Y et al. TransMIL: transformer based

correlated multiple instance learning for whole slide image classification[C]//*Proceedings of the 35th International Conference on Neural Information Processing Systems*, 2021: 164.
10. Ilse M, Tomczak J M, Welling M. Attention-based deep multiple instance learning[C]//*Proceedings of the 35th International Conference on Machine Learning*, 2018: 2132–2141.
11. Lu M Y, Williamson D F K, Chen T Y et al. Data-efficient and weakly supervised computational pathology on whole-slide images[J]. *Nat Biomed Eng*, 2021, **5** (6): 555−570.
12. DeVries T, Taylor G W. Improved regularization of convolutional neural networks with cutout[Z]. arXiv: 1708.04552, 2017. https://doi.org/10.48550/arXiv.1708.04552.
13. Huang Z Y, Wang H H, Xing E P et al. Self-challenging improves cross-domain generalization[C]//*16th European Conference on Computer Vision*, 2020: 124–140. https://doi.org/10.1007/978-3-030-58536-5_8.
14. Dehaene O, Camara A, Moindrot O et al. Self-Supervision closes the gap between weak and strong supervision in histology[EB/OL]. (2020-12-07)[2023-05-20]. https://doi.org/10.48550/arXiv.2012.03583.
15. Holdenried-Krafft S, Somers P, Montes-Majarro I A et al. Dual-query multiple instance learning for dynamic meta-embedding based tumor classification[EB/OL]. (2023-11-17)[2024-11-15]. https://doi.org/10.48550/arXiv.2307.07482.
16. Wang X G, Yan Y L, Tang P et al. Revisiting multiple instance neural networks[J]. *Pattern Recognit*, 2018, **74**: 15−24.
17. Geirhos R, Jacobsen J H, Michaelis C et al. Shortcut learning in deep neural networks[J]. *Nat Mach Intell*, 2020, **2** (11): 665−673.
18. Geirhos R, Rubisch P, Michaelis C et al. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness[C]//*7th International Conference on Learning Representations*, 2019.
19. Liang L M, Lu B H, Long P W et al. Adaptive feature fusion cascade transformer retinal vessel segmentation algorithm[J]. *Opto-Electron Eng*, 2023, **50** (10): 230161.
梁礼明, 卢宝贺, 龙鹏威, 等. 自适应特征融合级联 Transformer 视网膜血管分割算法[J]. 光电工程, 2023, **50** (10): 230161.
20. Ehteshami Bejnordi B, Veta M, Van Diest P J et al. Diagnostic assessment of deep Learning algorithms for detection of lymph node metastases in women with breast cancer[J]. *JAMA*, 2017, **318** (22): 2199−2210.
21. Brancati N, Anniciello A M, Pati P et al. BRACS: A dataset for BReAst carcinoma subtyping in H&E histology images[J]. *Database (Oxford)*, 2022, **2022**: baac093.
22. Zhang H R, Meng Y D, Zhao Y T et al. DTFD-MIL: double-tier feature distillation multiple instance learning for histopathology whole slide image classification[C]//*Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 18780–18790. https://doi.org/10.1109/CVPR52688.2022.01824.
23. Caron M, Touvron H, Misra I et al. Emerging properties in self-supervised vision transformers[C]//*Proceedings of IEEE/CVF Conference on Computer Vision*, 2021: 9630–9640. https://doi.org/10.1109/ICCV48922.2021.00951.
24. Kang M G, Song H, Park S et al. Benchmarking self-supervised learning on diverse pathology datasets[C]//*Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 3344–3354. https://doi.org/10.1109/CVPR52729.2023.00326.
25. Kingma D P, Ba L J. Adam: a method for stochastic optimization[C]//*International Conference on Learning Representations*, 2015.

作者简介



陈金令 (1975-)，博士，教授级高级工程师，主要研究方向为图像处理和模式识别。

E-mail：chengjl2002@163.com



柯琦 (1980-)，博士，副主任医师，主要研究方向为病理诊断。

E-mail：iminfeng@126.com



刘蓉 (1996-)，硕士研究生，主要研究方向为深度学习和医学图像处理。

E-mail：409664012@qq.com



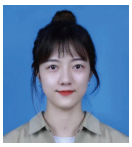
季语祝 (1982-)，硕士，副主任医师，主要研究方向为病理诊断。

E-mail：00150811@163.com



【通信作者】唐卓葳 (1978-)，硕士，副主任医师，主要研究方向为乳腺外科基础与临床研究。

E-mail: tungtung120928@163.com



高子清 (1994-)，硕士，医师，主要研究方向为病理诊断。

E-mail：793057761@qq.com

Research on whole slide image recognition based on weakly supervised learning and mixed distillation mechanism

Chen Jinling[1], Liu Rong[1], Tang Zhuowei[2*], Ke Qi[3], Ji Yuzhu[3], Gao Ziqing[3]

[1] School of Electrical Engineering and Information, Southwest Petroleum University, Chengdu, Sichuan 610500, China;
[2] Department of Breast Surgery, Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang, Sichuan 621000, China;
[3] Department of Pathology, Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang, Sichuan 621000, China

Abstract

Objective In clinical practice, pathological examination is considered the "gold standard" for diagnosing various types of cancer. In the field of pathological image intelligent diagnosis, the detection performance of deep learning models typically relies on large-scale and high-quality manually annotated datasets. Due to the high cost and expertise required in annotating medical images, large-scale annotation is often infeasible, resulting in challenges such as limited training data and significant domain variation in clinical scenarios. Therefore, achieving accurate, efficient, and generalizable medical image recognition under limited annotation conditions has become a important challenge in the field of intelligent medical image processing. Mainstream methods adopt multiple instance learning (MIL), which enables whole slide image(WSI) classification using only slide-level labels. However, MIL-based approaches still suffer from issues such as overly concentrated attention allocation and model overfitting when handling complex pathological patterns.

Methods To address these challenges, this study proposes a MIL framework based on a mixed distillation training strategy. First, Starting from publicly available datasets, the processes of image segmentation, image partitioning, and image feature extraction for digital pathology slides have been implemented. The OTSU segmentation method is used to remove the background images that do not contain any pathological information. Each WSI in the dataset is divided into non-overlapping image patchs to facilitate model training and feature extraction. These preprocessing steps help improve the efficiency and accuracy of the model, enabling it to focus more on learning the key pathological features. The self-supervised Vision Transformer-S/16(ViT-S/16) was used to train the model on unlabeled pathological image data. The ViT-S/16, which is a self-supervised learning method based on contrastive learning, can learn key feature representations from images. This process enables the model to learn transferable features from unlabeled data and provide better initialization for the downstream classification task. Then, a mixed distillation mechanism multiple instance learning (MDM-MIL) model integrating dual-path adaptive attention and random instance concealing is designed. The adaptive attention module dynamically adjusts the degree of focus on different texture features and captures semantic relationships between them to further enhance the classification accuracy. The first path adaptive gated attention submodule calculates the attention scores of all instances, and then aggregates the attention scores adjusted by random instance masking to integrate the remaining instance information. The second path uses an adaptive multi-head attention submodule to calculate the attention scores of all instances, and then aggregates the attention scores adjusted by random instance masking to integrate the remaining instance information. This adaptive method not only enhances the model's ability to recognize local features but also significantly improves its understanding of global context. Meanwhile, the random instance concealing (RIC) mechanism randomly masks instances with high attention scores during training, forcing the model to explore additional informative regions and mitigating overfitting to a few critical patches. Mixed distillation mechanism (MDM) integrates the feature information of the remaining instances from the two paths after being processed by the RIC method, thereby enhancing the learning efficiency and classification accuracy of the model. This method captures semantic relationships between different texture features of an image to further increase the classification accuracy. Finally, the classifier achieved accurate classification results on a dataset of breast cancer whole slide pathological images. This paper proposes a multi-instance learning model based on a mixed distillation strategy to address the issues of excessive attention allocation and model overfitting in existing multi instance learning. This model combines a collaborative training dual-path adaptive attention model

Foundation item: Open Funds of State Key Laboratory of Oncology in South China (HN2025-14), Sichuan Provincial Key R&D Plan (Major Science and Technology Project) (2022YFS0020), Science and Technology Project of the Sichuan Provincial Health Commission (23LCYJ020), Mianyang Central Hospital Clinical Special Fund (2025LC014), and the Special Project for City-College Science and Technology Strategic Cooperation of Nanchong City in 2022 (22SXQT0292)

*Correspondence: Tang Z W, E-mail: tungtung120928@163.com

with a random instance masking method, and ultimately uses a mixed distillation strategy to concatenate the aggregation results of the dual paths. The model can fully learn more key instance feature information while considering the correlation between various types of instances, ultimately improving the performance of the model and alleviating model overfitting.

Results and Discussions This study has the characteristics of uneven distribution of categories and weakened sample labels. Therefore, AUC and F1-score are comprehensively adopted as the main evaluation indicators to ensure the scientificity and comprehensiveness of the performance assessment. Experimental results demonstrate the superiority of the proposed method. On the Camelyon16 dataset, the model achieves an accuracy of 98.445%, an AUC of 0.984, and an F1-score of 0.983. On the BRACS dataset, the corresponding values reach 77.852%, 0.905, and 0.764, respectively, showing significant improvements over existing MIL approaches. The experimental results on the Camelyon16 and BRACS datasets show that compared with other advanced multiple instance learning models, This model has significantly better classification metrics, indicating that our model has more advantages in pathological image classification.

Conclusions This thesis proposes novel algorithm designs and conducts comparative experimental evaluations on corresponding public datasets, achieving superior results compared to the current state-of-the-art methods. This paper contributes to the advancement of automatic analysis techniques for pathological images based on deep learning, which holds significance in assisting cancer diagnosis, and improving the efficiency of clinical analysis.

Keywords: image classification; self-supervised learning; multi-instance learning; random instance concealing; adaptive attention; mixed distillation

Chen J L, Liu R, Tang Z W et al. Research on whole slide image recognition based on weakly supervised learning and mixed distillation mechanism[J]. *Opto-Electron Eng*, 2026, **53**(3): 250258; DOI: 10.12086/oee.2026.250258

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文