**技术交底书**

**发明名称： 一种基于多专家机制的前列腺癌肿瘤分期分类方法**

**本专利发明人： 倪张凯 朱俊泽 肖润宇 ­**

**技术交底书撰写人及技术联系人：\_倪张凯**

**电话： 13646019080 FAX: \_ \_ E-MAIL：\_zkni@tongji.edu.cn \_**

注意：

1.**第2部分和第4部分**，**一定要写的全面、清楚**。

2.英文缩写要有中文译文和英文原词，最好在术语解释部分给出。

3.**全文对同一事物的叫法应统一，避免出现一种东西多种叫法。**

4.应该阐述发明目的是通过什么技术方案来实现的，不能只有原理，也不能只做功能介绍。

**术语解释：**

解释一些跟技术方案有关的专业术语

1. **本发明要解决的技术问题是什么？**（务必明确一个最主要的技术问题）

前列腺癌症超声影像的遮掩**自动编码器的重建预训练和微调方法**

**2、详细介绍技术背景,并描述已有的与本发明最相近似的实现方案**（与本发明最接近的技术方案的说明（对于方法，应说明现有方法的步骤，对于装置，应当说明结构组成及其关系））

任务解释

前列腺癌是全球男性中第二常见的癌症类型，也是导致癌症相关死亡的主要原因之一。在《中国前列腺癌筛查与早诊早治指南（2022年）》中，强调了提高筛查效果和规范性。尤其是在低资源地区，低成本且高效的诊断方法尤为关键。作为现代医学诊断的重要工具，医疗影像技术广泛应用于疾病检测中，并发挥着不可替代的作用。医疗影像技术包含磁共振成像（Magnetic Resonance Imaging, MRI）、计算机断层扫描（Computed Tomography, CT）、X光成像以及超声成像技术等。这些技术提供了宝贵的视觉数据，帮助医生诊断和评估疾病。在前列腺癌的诊断中，MRI是最常用的影像学工具。MRI具有较高的特异性，可用于前列腺癌的T分期预测，但其敏感性较低，且受制于价格、禁忌症及设备普及等因素，难以普及到广泛的临床应用中。作为另一种常见的影像诊断工具，超声成像在前列腺癌检测中也有所应用。与MRI相比，超声成像具有成本低、适用人群广和对人体危害小的优势，更适用于低成本且高效的前列腺癌检测。但前列腺癌多发、散灶的特征使得单张影像的诊断价值降低，包绕整个前列腺腺体及周围组织、器官的超声视频弥补了单张影像诊断的不足。然而超声视频成像存在分辨率低的问题，导致难以准确进行诊断。此外，超声的敏感性和特异性存在较大波动，尤其在小病灶和边界模糊的情况下，诊断的准确性依赖于医生的个人经验。因此，如何有效提升超声视频的分辨率，以提供更加清晰且高保真的影像，并提高医生对小病灶及边界模糊等复杂区域的诊断准确性，已成为当前基于超声成像开展前列腺癌诊断的重要研究方向和技术发展重点之一。

在前列腺癌的诊断中，超声视频作为一种重要的成像工具，虽然具备了低成本、适用人群广泛等优点，但由于其分辨率低和诊断准确性受限，特别是在处理小病灶或边界模糊的区域时，仍然面临许多挑战。因此，如何提升超声视频的诊断能力，特别是在处理复杂病灶区域时，已经成为了前列腺癌超声诊断领域的核心问题之一。视频分类任务在这一背景下成为了提升超声视频诊断效果的重要手段之一。与传统的单张静态影像不同，超声视频通过连续的帧信息提供了时序特征，这对于捕捉前列腺癌的动态变化、病灶发展及其与周围组织的关系等具有重要意义。通过对超声视频中的多个帧进行分析，可以更全面地评估前列腺组织的结构、血流情况以及肿瘤的形态等，从而提高疾病的检测率和诊断的准确性。

近似实验方案

VideoMAE工作采用了基于遮掩自动编码(Masked Autoencoder, MAE)的自监督预训练方法，基于视频的MAE预训练流程如下图所示：

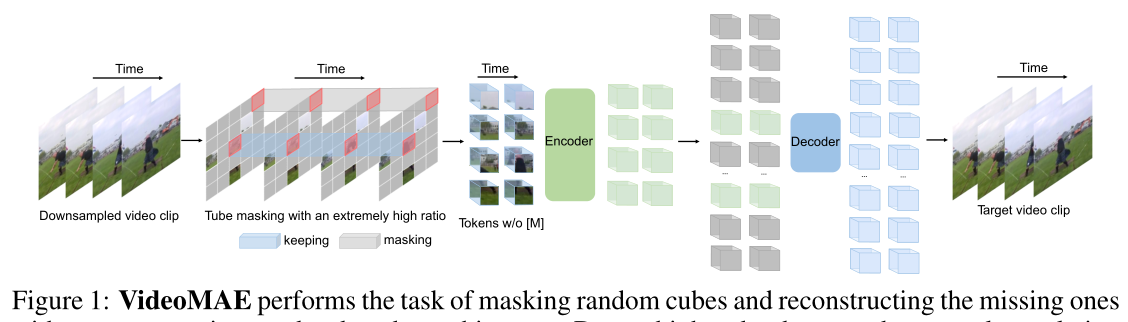


图1.VideoMAE预训练流程框架

将下采样帧作为输入，使用跨行时间采样策略来进行更有效的视频预训练。形式上，首先从原始视频V中随机采样一个由t个连续帧组成的视频剪辑。然后，使用时间采样将剪辑压缩为T帧，每个帧的分辨率为，并使用立方体嵌入来获得视频令牌。其中我们将每个大小为的立方体视为一个令牌嵌入。因此，立方体嵌入层获得个三维令牌，并将每个令牌映射到通道维度d。这种设计可以降低输入的时空维度，有助于缓解视频中的时空冗余。

对于获得的三维令牌，VideoMAE随机选择90%的信息进行遮掩。然后把没有遮掩的令牌输入到了视觉编码器(Vision Encoder)中，在VideoMAE工作中使用了视觉转换器(Vision Transformer, ViT)作为主干。最后利用解码器进行重建训练，优化重建损失。

随后针对微调过程，VideoMAE利用预训练好的编码器，将三维令牌输入之前预训练好的视觉转换器(Vision Transformer, ViT)的视觉编码器(Vision Encoder)中获得特征，取最深维度的特征作为获取的特征输入分类头，通过分类损失训练分类头和视觉编码器。

**3、以因果关系推理的方式推导出现有技术的缺点是什么？**

**针对这些缺点，说明本发明的目的。**（缺点可以有多个，但一定要分清主次。最主要的缺点原则上只能有一个。）

其中针对前列腺癌症超声影像的数据方面有如下问题：

1. ：前列腺超声影像的数据中存在大量的无关信息，尤其是无病灶的区域。随机遮掩会导致模型学习过程中对重点病灶区域的忽略，这在医学影像中尤其重要，因为局部特征往往决定了疾病的早期发现和准确诊断。MAE在遮掩了过多非关键的局部特征时，模型对真正关键的局部特征重建特征的学习就会相对的差。这也是最主要的缺陷。

本发明提出了一种基于多专家的大模型丰富基础知识指导的可泛化的前列腺癌症T0/T1期诊断系统。该方法针对MAE流程中随机遮掩的重建预训练未让模型学习到关键区域的特征问题，设计了一个基于病灶分割内容的遮掩得分网络。迫使模型学习关键区域的特征也就是关键的局部特征。在专家模型的搭建中，受制于模型的泛化性问题，其在不可见数据集上的分割结果对于其中的一部分实例存在错误。由于不同的分割模型见到的数据并不一致，因而存在不同的分割错误实例，然而三个模型同时对同一实例分割错误的可能性较低，可视化如图2所示。我们做了如下设定：我们构建模拟医生会诊的多专家会诊机制，利用三个专家模型降低失误概率。

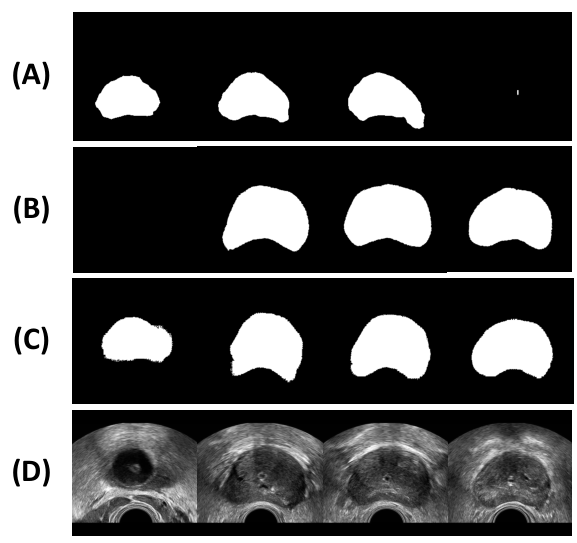
****

图2.A)MedSAM2模型分割结果；B)Deeplabv3模型分割结果；

C)nnUNet2++模型分割结果；D)超声影像原图像；

1. 现有模型未充分考虑多医生采样之间的个体差异、诊断风格和数据质量波动，导致验证结果不稳定，无法满足医工结合应用中对准确性和鲁棒性的高要求。

在下游微调时，利用在重建训练阶段预训练好的遮掩得分网络，将输入的视频中的重要区域的块作为额外提示添加给特征辅助分类头训练。来解决多医生采样之间的个体差异、诊断风格和数据质量波动带来的模型效果问题。

1. **本发明技术方案的详细阐述，应该结合流程图、原理框图、电路图、时序图进行说明**（所有英文缩写都应有中文注释；**所有附图都应该有详细的文字描述，以别人不看附图即可明白技术方案为准**；同时附图中的关键词或方框图中的注释都尽量用中文；方法专利都应该提供流程图，并提供相关的系统装置）。

技术问题

本发明涉及表征学习领域，具体涉及遮掩自动编码器的重建方法。

技术方案

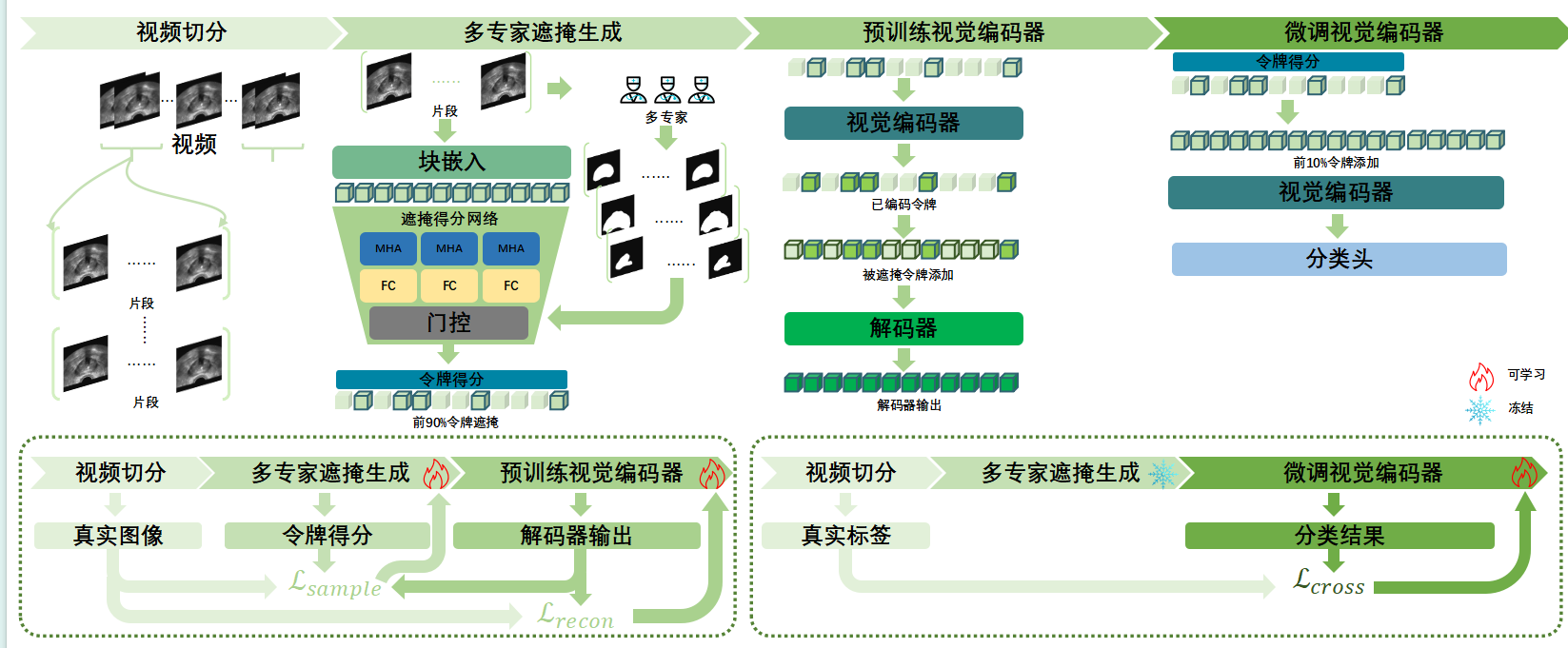


图3.本专利框架流程

一种基于多专家机制的前列腺癌肿瘤分期分类方法，主要流程图如图3所示，具体而言包含以下步骤：

步骤1 数据预处理

将视频切分为多个等长的剪辑片段，片段长度固定。将片段中的每一帧通过MedSAM2，Deeplabv3，nnUNet2++，三个预训练的分割大模型获得三个模型对帧的分割结果。

步骤2 遮掩得分生成

将整个视频通过块嵌入(Patch Embedding)获得多个三维令牌(Token)，将三维令牌输入我们设计的遮掩得分网络，该网络由三个独立的多头注意力块(Multi-Head Attention , MHA)和全连接层(Fully Connected , FC)组成，三个全连接层在输出各自对每一个令牌的得分之后，通过之前获得的三个分割图像，若某一个令牌是在分割图内的话，那么这个遮掩得分就会被加上一个值。以此来达到核心区域遮掩的目的。迫使模型学习核心病灶区域的特征。

步骤3 预训练编码器

受VideoMAE设计的启发，我们开始先预训练编码器，在这里我们选择的是视觉转换器基础版本(Vision Transformer-base, ViT-B)作为视觉转换器编码器解码器。通过步骤2，我们获得了每一个令牌的得分，我们将前90%排名的令牌遮掩，之后输入编码器，获得编码过的令牌，再将之前遮掩的令牌添加之后输入解码器，解码器重建出原图像计算重建损失训练来更新视觉转换器，通过重建内容和原内容差距和遮掩得分网络计算采样损失来更新遮掩得分网络。

步骤4 微调编码器、分类头

在经过步骤3预训练好视觉编码器和遮掩得分网络后，我们冻结遮掩得分网络，开始进行下游分类任务中编码器和分类头的微调。对于带有标签的少量数据，我们将视频分出的剪辑经过遮掩得分网络后获得每个令牌的得分，将前10%的令牌作为额外的提示(Prompt)添加给令牌序列，输入预训练好的视觉编码器之后将获得的特征输入给分类头进行分类，通过交叉熵来更新分类头和编码器。

技术效果

与现有技术相比，本发明具有以下有益效果：

1. 预训练阶段重建针对性：本发明提出的基于医学分割模型先验知识引导的遮掩重建的预训练策略对不同医生操作的不同数据具有良好的适应性。
2. 下游微调阶段额外提示针对性：本发明提出的基于预训练过程中训练的遮掩得分网络中重要信息添加作为额外提示，对不同医生操作的数据做下游分类时也能表现出较好的适应性。

附图说明

图1.遮掩自动编码器在视频中的应用VideoMAE示意

图2.三个医学分割模型的可视化

图3.本专利的框架流程

图4.本专利的遮掩得分网络框架流程

图5.本专利预训练参数更新流程

图6.本专利微调参数更新流程

具体实施方式

下面结合附图和具体实施例对本发明进行详细说明。本实施例以本发明技术方案为前提进行实施，给出了详细的实施方式和具体的操作过程，但本发明的保护范围不限于下述的实施例。

一种基于多专家机制的前列腺癌肿瘤分期分类方法包括如下具体步骤：

步骤1：数据预处理

针对有T帧的视频V，我们分割为多个等长剪辑，长度l设定为16，每个剪辑采用均匀步长的采样，具体来说我们设定采样间隔为step = ，从第1帧到第step帧。其中每一帧开始往下均匀的采样，每隔step帧就采样进入本剪辑内，最终获得多个等长剪辑片段：

（1）

此时经过预训练的分割模型MedSAM2，Deeplabv3，nnUNet2++我们获得对应的分割内容：

（2）

步骤2：遮掩得分生成

我们采用联合时空的立方嵌入(Embedding)，具体来说对于一个剪辑Clip大小为，我们在时间上采样2单位，空间上采样单位，就获得了N=()个大小为的块，将每个令牌映射成维度。这样做来缓解时空冗余。

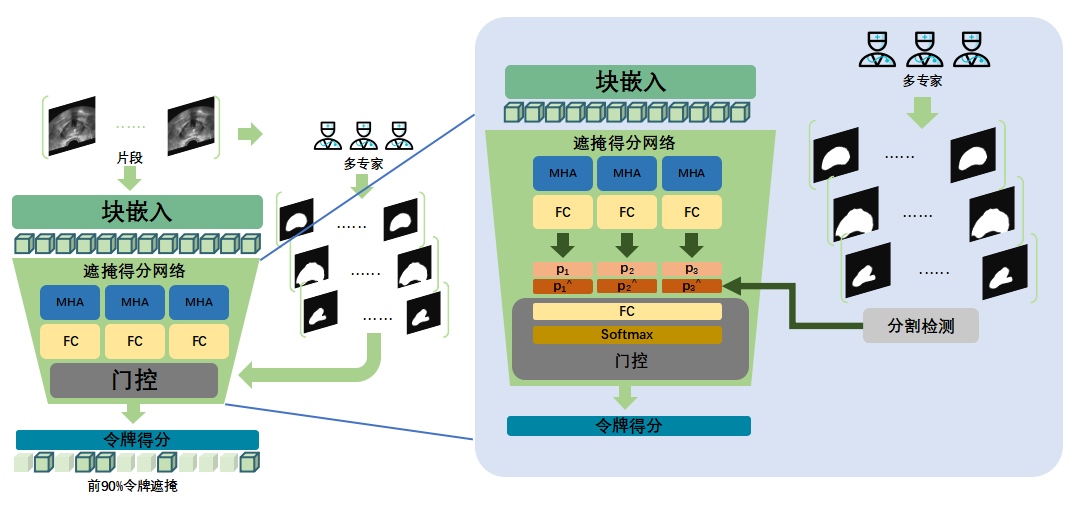


图4.遮掩得分网络

对于这些令牌，我们输入遮掩得分网络，网络架构如图4所示，遮掩得分网络分为3个独立的多头注意力块和全连接层，具体计算如下：对于已经嵌入的令牌

（3）

（4）

这里我们就获得了三个得分网络对所有三维令牌的遮掩评分

（5）

此处加入在数据预处理阶段就获得的分割内容，针对我们现在计算的这一个剪辑，我们获得也就是对应该剪辑片段的对应三个大模型分割的结果。我们做如下计算：

（6）

其中的InRoI函数表示的是N个令牌代表的位置是否在分割图像内，如果是就返还一个小值增加该令牌的得分。如果不在就返还0。具体如下：

（7）

最终我们获得了三个独立的遮掩得分评价

（8）

将这个的得分输入门控网络中

（9）

然后将W**权值广播**后和的得分相乘后相加得到最终的令牌得分。

步骤3：预训练编码器

在预训练阶段，我们的训练和参数更新图5所示：

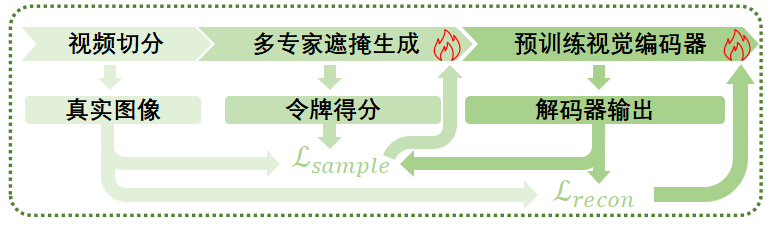


图5.预训练流程

对于得到的令牌得分，我们将前90%的令牌遮掩后将剩下的10%可见令牌输入编码器(Encoder)，进行编码之后将之前遮掩的90%令牌添加到编码过的令牌中，然后输入视觉解码器(Decoder)，进行重建，得到重建结果。此时我们计算重建损失来更新编码器解码器的权重。损失公式如下：

（10）

其中N代表所有令牌的个数，和分别代表重建输出的和原令牌的RGB（Red, Green, Blue）值。

然后我们也要训练遮掩得分网络，我们计算采样损失来更新遮掩得分网络的权重。损失公式如下：

（11）

这种公式的制定是由RL中REINFORCE算法的预期奖励最大化激发的。这里，可见令牌采样过程是动作，MAE通过给定输入数据和掩蔽部分来反馈模型的行为，就像强化学习中的环境提供反馈，而掩码令牌重构误差是返回。在图像中，不同的区域含有不同的“信息量”。例如，图像的某些区域可能包含更多的结构化信息（如物体的轮廓、颜色变化等），而其他区域可能是背景区域，包含的信息较少。在MAE中，高信息区域（例如物体、边缘等）通常比低信息背景区域对重构更为关键，它们的重构误差通常较高。因此，最大化期望重构误差将导致网络预测高信息区域的概率得分较高。

步骤4：微调编码器、分类头

在微调阶段，我们的训练和参数更新流程如图6所示。

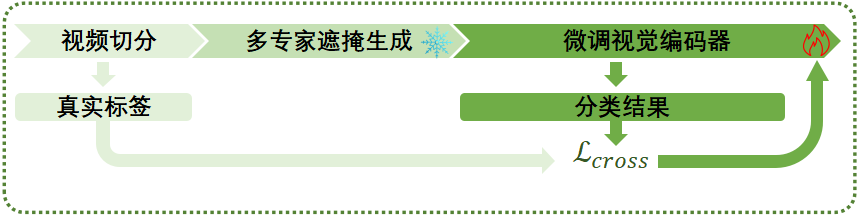


图6.微调流程

考虑到我们的任务是进行前列腺肿瘤分期检测，因此我们在概率MAE的表征输出上添加一个新的分类头。在步骤3结束后，我们预训练好了遮掩得分网络和编码器，我们冻结遮掩得分网络。对于一个带有T0或者T1类标签的视频，我们同样按照步骤1拆分出多个Clip，对于单个Clip，我们还是通过遮掩得分网络生成遮掩得分，然后我们将前10%得分的令牌作为额外提示添加到经过嵌入的令牌队伍中。再输入编码器中，输出了特征之后我们通过分类头输出分类结果。然后利用重建损失更新分类头和编码器。

**5、本发明的关键点和欲保护点是什么？**（对于上一部分给出的详细的完整技术方案，在本部分是提炼出技术方案的关键创新点，列出1、2、3...，以提醒代理人注意，便于专利代理人撰写权利要求书）

1. 遮掩得分网络设计：本发明提出了一种基于医学分割模型先验知识引导的遮掩得分网络，该网络能够自动计算出每个令牌的重要性，并且在预训练阶段通过遮掩得分机制确保模型学习到关键区域的特征。与传统的随机遮掩方法相比，使用分割模型生成的遮掩得分能够有效引导模型重点关注肿瘤病灶区域，减少对无关区域的干扰。本发明结合多个医学分割模型（MedSAM2、Deeplabv3、nnUNet2++）来处理不同医生操作和数据质量波动问题。通过多专家机制，显著降低单一模型的误差率，从而提高模型在不同数据集上的稳定性和泛化能力。
2. 下游微调过程中的额外提示机制：在微调阶段，本发明通过预训练得到的遮掩得分网络生成的遮掩得分作为额外提示添加到输入令牌序列中，增强了模型对前列腺癌肿瘤的分类能力，尤其是在面对少量标注数据时，能够实现高精度分类。

**6、用推理方式推导出本发明的优点**（务必与第3部分的现有技术的缺点相对应），可以对应3部分所要解决的技术问题或发明目的来描述。

1. 提高了模型对关键区域特征的学习能力：现有的MAE方法通过随机遮掩可能忽略重要的局部特征，而本发明通过引入基于医学分割模型的遮掩得分网络，强制模型关注前列腺癌肿瘤的关键区域，确保了肿瘤特征的学习和重建，提升了模型的诊断精度。
2. 增强了对多医生操作和不同数据质量的适应性：现有的MAE方法忽视了数据质量和医生操作之间的差异，导致模型在实际应用中的稳定性较差。本发明通过多专家机制将不同分割模型的输出进行融合，显著降低了单个模型的误差，提升了系统对不同医生诊断风格和数据质量波动的适应性。

与现有技术相比，本发明具有以下有益效果：

1. 预训练阶段重建针对性：本发明提出的基于医学分割模型先验知识引导的遮掩重建的预训练策略对不同医生操作的不同数据具有良好的适应性。
2. 下游微调阶段额外提示针对性：本发明提出的基于预训练过程中训练的遮掩得分网络中重要信息添加作为额外提示，对不同医生操作的数据做下游分类时也能表现出较好的适应性。

**7、针对4中的技术方案，是否还有别的替代方案同样能完成发明目的？**

无

**8、其他有助于专利代理人理解本技术的资料**（给代理人提供更多的信息，可以有助于代理人更好更快的完成申请文件）

无

1. **案例（实施例）**

数据集：

实验在上海第十人民医院的数据集上进行训练和验证，数据统计如下表，表1展示了我们训练集和验证集的统计：

表1.本专利数据集统计信息

|  |  |  |
| --- | --- | --- |
| 数据 | 上海第十人民医院（训练） | 上海第十人民医院（验证） |
| 病例数量 | 440 | 110 |
| 癌症初期（T0期）数量 | 230 | 58 |
| 癌症晚期（T1期）数量 | 210 | 52 |

方法应用：

本实验将训练批次大小（Batch Size）设置为 4，整个过程中在两块RTX 3090显卡上训练。对于我们的工作，我们使用AdamW优化器，在1e-3的学习率下预训练500个轮次来最小化重建损失和采样损失。随后我们在4e-4的学习率下微调300轮次最小化交叉熵损失。

效果对比

在实验效果对比上，由于医学场景和任务的特殊性，我们将效果对比在工作的消融中。针对遮掩方式进行对比实验。如下实验表2所示，我们对比了预训练视觉编码器中随机掩码和我们设计的基于三个医学分割模型的先验分割引导掩码的差距，两者同样使用了我们改进的微调方式，结果如表2所示

表2.针对掩码的消融实验

|  |  |  |
| --- | --- | --- |
| 掩码策略 | ACC | AUC |
| 随机掩码 | 0.8207 | 0.875 |
| 分割引导掩码（OURS） | **0.9136** | **0.963** |

在下游微调方法效果对比上，我们同样针对微调方法进行了对比实验，两者同样使用了我们改进的基于三个医学分割模型的先验分割引导掩码的预训练方法，对比了直接将所有令牌输入后取视觉编码器最深层特征的微调方法和我们设计的利用10%最重要的额外令牌作为视觉编码器输入特征并且取出最深层特征的微调策略。如表3结果所示

表3.针对微调方法的消融实验

|  |  |  |
| --- | --- | --- |
| 掩码策略 | ACC | AUC |
| 直接取最深层特征微调 | 0.9024 | 0.941 |
| 利用10%的额外令牌输入微调 | **0.9136** | **0.963** |

同时我们和一些现有工作进行了比较，我们把实验分为两组，基于图片的工作和基于视频的工作。我们使用统一的训练验证集划分，同时我们遵循了各个工作自己的训练验证设置。对于基于图片的工作，我们将视频中每一帧输入，将模型对所有图片的评分的平均作为对视频的评价。效果如下表4所示：

表3.和其他工作的对比

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 组别 | 方法 | 主干 | ACC | AUC |
| 基于图片 | ResNet50 | CNN | 0.726 | 0.739 |
| US\_UCL | 0.609 | 0.713 |
| RadFormer | Transformer | 0.717 | 0.758 |
| PVTv2(SOTA) | 0.783 | 0.829 |
| 基于视频 | VideoMAEv2 | Transformer | 0.904 | 0.942 |
| m2clip(SOTA) | 0.906 | 0.949 |
| (OURS) | **0.9136** | **0.963** |

上述描述仅是对本申请较佳实施例的描述，并非是对本申请范围的任何限定。任何熟悉该领域的普通技术人员根据上述揭示的技术内容做出的任何变更或修饰均应当视为等同的有效实施例，均属于本申请技术方案保护的范围