(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2023 年 5 月 11 日 (11.05.2023)

WIPO | PCT

(10) 国际公布号
WO 2023/078041 A1

(51) 国际专利分类号:
G06K 9/62 (2022.01)

(21) 国际申请号: PCT/CN2022/124693

(22) 国际申请日: 2022 年 10 月 11 日 (11.10.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202111291695.X 2021年11月3日 (03.11.2021) CN

(71) 申请人: 浙江大学(ZHEJIANG UNIVERSITY) [CN/CN]; 中国浙江省杭州市西湖区余杭塘路866号, Zhejiang 310058 (CN)。

(72) 发明人: 李劲松(LI, Jingsong); 中国浙江省杭州市西湖区余杭塘路866号, Zhejiang 310058 (CN)。 田雨(TIAN, Yu); 中国浙江省杭州市西湖区余杭塘路866号, Zhejiang 310058 (CN)。 周天舒(ZHOU, Tianshu); 中国浙江省杭州市西湖区余杭塘路866号, Zhejiang 310058 (CN)。 叶前呈(YE, Qiancheng); 中国浙江省杭州市西湖区余杭塘路866号, Zhejiang 310058 (CN)。

(74) 代理人: 杭州求是专利事务所有限公司(HANGZHOU QIUSHI PATENT OFFICE CO., LTD.); 中国浙江省杭州市西湖区古墩路701号绿城紫金广场C座1506室 / 刘静, Zhejiang 310030 (CN)。

(81) 指定国(除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI,

WO 2023/078041 A1

(54) **Title:** TUMOR TISSUE PATHOLOGICAL CLASSIFICATION SYSTEM AND METHOD BASED ON ADAPTIVE PROPORTIONAL LEARNING

(54) 发明名称: 基于自适应比例学习的肿瘤组织病理分类系统及方法

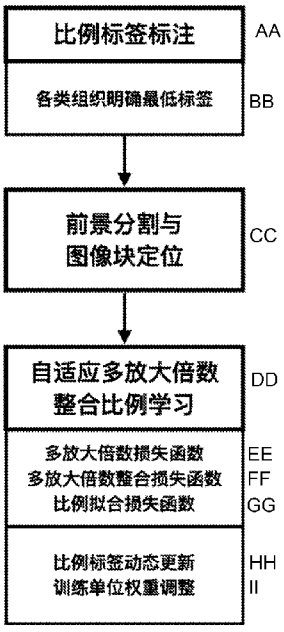


图 1

AA Make proportion labels
BB Clarify the lowest label of each class of tissue
CC Foreground separation and image block positioning
DD Adaptive multi-amplification-multiple integration proportional learning
EE Multi-amplification-multiple loss function
FF Multi-amplification-multiple integrated loss function
GG Proportional fitting loss function
HH Dynamically update proportion labels
II Adjust the weight of a training unit

(57) **Abstract:** A tumor tissue pathological classification system and method based on adaptive proportional learning. The method comprises: first acquiring several pathological sections and performing digital scanning thereon, performing manual labeling on a pathological section image, which is obtained after scanning, according to classification task target categories, and constructing a data set; then separating a tissue foreground by using difference distribution features between RGB channels and a grayscale value, so as to construct a training data set that includes image blocks of a plurality of amplification multiples; and finally, performing integration on the

[见续页]

GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布：

— 包括国际检索报告(条约第21条(3))。

plurality of amplification multiples, combining cross-entropy functions of the amplification multiples and an integrated amplification multiple to form a loss function, so as to realize multi-amplification-multiple integration learning. By means of adaptive proportional learning, global proportional labels of an image and a training weight of an image block, the proportion in which does not reach the minimum proportion, are dynamically adjusted, thereby increasing a data utilization rate, and realizing rapid convergence. In routine pathological examinations of tumor tissues, a detection rate is increased to the greatest extent while additional workloads are added as little as possible.

(57) 摘要：一种基于自适应比例学习的肿瘤组织病理分类系统及方法，首先获取若干病理切片并进行数字化扫描，将扫描后的病理图像按照分类任务目标类别进行人工标注，构建数据集；然后利用RGB通道和灰度值的差值分布特征分割组织前景，构建包含多级放大倍数的图像块的训练数据组；最后进行多级放大倍数整合，组合了各级放大倍数和整合放大倍数的交叉熵函数形成损失函数，实现多放大倍数整合学习；通过自适应比例学习，对图像全局比例标签和未达到最低比例的图像块训练权重进行动态调整，扩大数据利用率，实现快速收敛。在日常肿瘤组织的病理检查中，在尽可能低地增加额外工作负担的基础上，最大限度地提升检出率。

基于自适应比例学习的肿瘤组织病理分类系统及方法

技术领域

本发明涉及医疗图像处理及机器学习领域，尤其涉及一种基于自适应比例学习的肿瘤组织病理分类系统及方法。

背景技术

病理检查作为癌症诊疗的第一步，提供的组织病理信息对于癌症患者病情判断至关重要。但由于病理医生缺口大，较多的病理指标无法在日常病理检查中得到完整详细的调查，一些罕见的肿瘤分化如果未被病理医生识别则仅报告为中分化或低分化。胚胎发育来源接近的消化系统中，肝脏以外的器官肿瘤也有可能出现肝样分化，没有鳞状上皮分布但也可能出现鳞状细胞癌分化，以及较为常见的神经内分泌分化等等，这些类别分化的出现可能会影响患者预后。现有的病理报告中对相关类型分化的描述缺失，不利于精准的个性化医疗的开展，因此需要额外的识别方法或工具对其进行识别，有效补充病理报告。

现有方案包括两类，一类是人工阅片，一类是机器阅片。人工阅片，即由医生对病理样本制片染色后的切片图像进行检查，根据细胞分化形态将肿瘤细胞分为高、中、低等不同分化等级，以及根据细胞间的组织形态分为腺体状、巢状、实性片状等生长模式。但在病理检查采样多、病理图像尺寸大的客观条件下，病理医生无法完成不遗漏每一区域的详尽检查，容易遗漏一些罕见但倾略性较高的分化亚型。

机器阅片即利用机器学习算法（包括深度学习模型在内）根据病理图像特征与图像标签进行模型训练，当模型测试的准确性达到一定程度时可应用于病理特征识别。模型训练多采用监督学习的方式，即每一个训练最小单位都需要有类标签，为达到较高的识别准确率则需要大量的训练数据及标签，标注成本过高。目前针对是否为癌的分类模型较为常见，利用多示例学习可以有效降低标签成本，同时大规模的弱监督学习可以提升识别准确性。比例标签学习则利用数据包内类别所占比例进行机器学习，但在病理图像领域，精确的比例标签的获取成本约等同于监督学习，失去了弱监督学习低标签成本的优势。现有技术有如下缺点：

1. 人工阅片耗时耗力，且存在个体间和个体内的差异性，难以覆盖整张病理切片，无法获取较细粒度的分化差异；

2. 由于样本制备、染色标准、扫描设备的不同，数字病理切片图像的数据质量差异较大，且易存在包括笔迹、边缘伪影在内的非实质组织区域，在去除这些质量问题前不适合模型训练与直接应用，在进行自动化识别前应进行质量控制，保证数据质量；

3. 现有机器学习方法标注成本过高：监督学习需要大量像素级别或图像块级别的标注标签；多示例学习难有绝对的完全标签，且训练约束太少，模型效果不佳；全尺寸病理图像的客观比例标签获取成本同监督学习。

4. 现有模型通常仅用单一放大倍数或不规定放大倍数进行建模，而绝大多数肿瘤分化需要综合不同放大倍数的形态特征以得到准确判断。

发明内容

本发明目的在于针对现有技术的不足，提出一种基于自适应比例学习的肿瘤组织病理分类系统及方法。

本发明的目的是通过以下技术方案来实现的：一种基于自适应比例学习的肿瘤组织病理分类系统，该系统包括比例标签标注模块、前景分割与图像块定位模块和自适应多放大倍数整合比例学习模块；

所述比例标签标注模块获取若干肿瘤组织病理切片并进行数字化扫描，将扫描后的病理图像按照分类任务目标类别进行人工标注，构建数据集；

所述前景分割与图像块定位模块将病理图像在不同的放大倍数下，根据设定的目标图像块重叠率计算病理图像的缩略图尺寸并利用缩略图的 RGB 通道和灰度值的差值分布特征分割肿瘤组织前景，去除肿瘤组织背景，缩略图中每个像素点对应一个图像块；

所述自适应多放大倍数整合比例学习模块通过神经网络模型计算数据集中分割后的图像块在不同放大倍数下属于各个分类任务目标类别的概率，得到病理图像含有各个分类任务目标类别的比例，作为计算比例，对于每个分类任务目标类别，分别将计算的不同放大倍数下概率最小值与不同放大倍数下概率最大值的二分之一进行比较，取其中较大值作为图像块在该分类任务目标类别下的整合概率；将所有图像块同一类别下的整合概率进行降序排列，根据标注的比例阈值，选择训练图像块，并对每个训练图像块赋值类别标签；根据标注的比例阈值与计算比例的差值，对于病理图像中处于计算比例与比例阈值之间的训练图像块给予更高的训练权重，权重与差值呈正比，进行自适应权重赋值；使用权重调整后的训练图像块训练神经网络，损失函数为各级放大倍数损失函数项、整合损失函数项及比例拟合损失函数项之和，其中比例拟合损失函数项为若分类任务目标类别的计算比例低于比例阈值，则比例阈值与计算比例的差值作为比例拟合损失函数项，若计算比例大于等于比例阈值，则损失函数中比例拟合损失函数项为零，并根据训练结果调整比例阈值对模型迭代训练，通过训练好的神经网络模型计算病理图像含有各个分类任务目标类别的比例。

进一步地，所述比例标签标注模块共获取 n 张病理切片的全尺寸数字化扫描图像，构成数据集记为$\boldsymbol{B}=\{B_1,\ldots,B_i,\ldots,B_n\}$，其中$B_i$为第$i$张病理图像；分类任务目标类别共计 C 类，

针对第i张病理图像B_i，构建比例标签向量$\boldsymbol{R}_i$为：

$$\boldsymbol{R}_i = \left[R_i^1, \ldots, R_i^k \ldots, R_i^C\right]^T, k \in \{1, 2, \ldots, C\}, R_i^k \in [0,1]$$

其中，R_i^k为第i个病理图像中第k类的比例阈值，且满足$\sum_{k=1}^{C} R_i^k \leq 1$，$T$表示转置。

进一步地，所述前景分割与图像块定位模块计算病理图像的缩略图尺寸具体过程如下：记目标图像块尺寸为S_{tar} (pixels) * S_{tar} (pixels)，pixels为像素，目标图像块分辨率为R_{tar}，目标图像块重叠率为O_{tar}，对原始尺寸为W_{ori} (pixels) × H_{ori} (pixels)，原始分辨率为R_{ori}的病理图像B_i进行目标缩略图获取，W_{ori} (pixels)和H_{ori} (pixels)分别为原始病理图像尺寸的宽和高，则目标缩略图的宽和高尺寸分别为：

$$W_{tar} = 2 * \left\lfloor \frac{\left\lfloor \frac{W_{ori}}{2} \right\rfloor}{\left\lfloor \frac{R_{tar}}{R_{ori}} * S_{tar} * (1 - O_{tar}) \right\rfloor} \right\rfloor, H_{tar} = 2 * \left\lfloor \frac{\left\lfloor \frac{H_{ori}}{2} \right\rfloor}{\left\lfloor \frac{R_{tar}}{R_{ori}} * S_{tar} * (1 - O_{tar}) \right\rfloor} \right\rfloor$$

目标缩略图对应病理图像B_i的目标分割区域左上角点坐标为$(W_{left}, H_{up}) = (\lfloor \frac{W_{ori}}{2} \rfloor mod \left\lfloor \frac{R_{tar}}{R_{ori}} * S_{tar} * (1 - O_{tar}) \right\rfloor, \lfloor \frac{H_{ori}}{2} \rfloor mod \left\lfloor \frac{R_{tar}}{R_{ori}} * S_{tar} * (1 - O_{tar}) \right\rfloor)$；其中，$mod$表示取余数，将缩略图的目标分割区域，进行像素级分割，目标缩略图中每一个像素点对应一个的图像块，图像块的尺寸为$\left\lfloor \frac{R_{tar}}{R_{ori}} * S_{tar} \right\rfloor * \left\lfloor \frac{R_{tar}}{R_{ori}} * S_{tar} \right\rfloor$。

进一步地，所述前景分割与图像块定位模块利用缩略图的RGB通道和灰度值的差值分布特征分割肿瘤组织前景具体过程如下：提取缩略图红V_{Red}、绿V_{Green}、蓝V_{Blue}通道值以及灰度图值V_{Grey}，计算$V_{Red} - V_{Grey}$，$V_{Green} - V_{Grey}$，$V_{Blue} - V_{Grey}$，$V_{Red} - V_{Green}$，$V_{Blue} - V_{Green}$这五种差值密度分布的分割阈值；若缩略图上的一像素点超过上述五个分割阈值的数量大于等于3，则该像素点对应的图像块被视为组织前景，否则，该像素点对应的图像块作为背景剔除。

进一步地，所述损失函数中各级放大倍数损失函数项如下：

$$\text{Loss}_{op} = \begin{cases} -\frac{1}{S_i} \sum_{j=1}^{S_i} \sum_{k=1}^{C} \omega_j^k l_j^k log\left(p_{j_op}^k\right), & \sum_{j=1}^{N_i} \bar{p}_j^k < j \leq R_i^k * N_i \\ -\frac{1}{S_i} \sum_{j=1}^{S_i} \sum_{k=1}^{C} \omega^k l_j^k log\left(p_{j_op}^k\right), & j \leq \sum_{j=1}^{N_i} \bar{p}_j^k \end{cases}$$

其中，op表示不同的放大倍数，ω^k为第k类计算比例中的训练图像块的训练权重，ω_j^k为处于计算比例和比例阈值之间的训练图像块的训练权重；$S_i = \cup_{k=1}^{C} S_i^k$，$S_i^k$为第$k$类训练图像块的集合，$\bar{p}_j^k$为第$k$类中第$j$个图像块的整合概率，$l_j^k$为第$k$类中第$j$个图像块的标签；$p_{j_op}^k$为图像块在各放大倍数下归类为第$k$类的概率；$N_i$为第$i$个病理图像中前景图像块的数量。

进一步地，所述损失函数中整合损失函数项$Loss_{combine}$如下：

$$Loss_{combine}=\begin{cases}-\frac{1}{S_i}\sum_{j=1}^{S_i}\sum_{k=1}^{C}\omega_j^k l_j^k log(\vec{p}_j^k), & \sum_{j=1}^{N_i}\vec{p}_j^k < j \le R_i^k * N_i \\ -\frac{1}{S_i}\sum_{j=1}^{S_i}\sum_{k=1}^{C}\omega^k l_j^k log(\vec{p}_j^k), & j \le \sum_{j=1}^{N_i}\vec{p}_j^k\end{cases}$$

。

进一步地，所述损失函数中比例拟合损失函数项为根据去除背景图像块后的病理图像中的训练图像块数量，通过随机升采样或者随机降采样，使得一个训练批次中的训练图像块均来自同一张病理图像：比例拟合损失函数项$Loss_{ratio}^k$定义为：

$$Loss_{ratio}^k=\begin{cases}R_i^k-\frac{1}{N_i}\sum_{j=1}^{N_i}\vec{p}_j^k, & \frac{1}{N_i}\sum_{j=1}^{N_i}\vec{p}_j^k < R_i^k \\ 0, & \frac{1}{N_i}\sum_{j=1}^{N_i}\vec{p}_j^k \ge R_i^k\end{cases}$$

。

进一步地，所述自适应多放大倍数整合比例学习模块根据神经网络训练结果调整比例阈值具体公式如下：

$$R_i^{k^{epoch}}=\begin{cases}R_i^{k^{epoch-1}}, & \frac{1}{N_i}\sum_{j=1}^{N_i}\vec{p}_j^k < R_i^{k^{epoch-1}}+\alpha \\ R_i^{k^{epoch-1}}+\alpha, & \frac{1}{N_i}\sum_{j=1}^{N_i}\vec{p}_j^k \ge R_i^{k^{epoch-1}}+\alpha\end{cases}$$

其中，$R_i^{k^{epoch}}$表示调整后的比例阈值，$R_i^{k^{epoch-1}}$表示调整前的比例阈值，α为调整率。

进一步地，训练图像块的具体选择过程如下：根据分类任务目标类别中的第 k 类对应多放大倍数整合概率进行降序排列，得$\{\vec{p}_1^k,\dots,\vec{p}_j^k,\dots,\vec{p}_{N_i}^k\}$，此时$\vec{p}_1^k$对应的图像块在第 k 类中概率最高，$\vec{p}_{N_i}^k$对应的图像块在第 k 类中概率最低；对于病理图像B_i中归类为第 k 类的图像块，根据标注的比例阈值，选择在比例阈值内的图像块作为训练图像块：

$$S_i^k=\{\vec{p}_1^{k\prime},\dots,\vec{p}_j^{k\prime},\dots,\vec{p}_{N_i\times R_i^k}^{k\prime}\}$$

其中，$\vec{p}_j^{k\prime}$表示$\vec{p}_j^k$对应的训练图像块；赋值类别标签$l=\{l_j^k \mid l_j^k\in\{0,1\},\ k=1,\dots,C,\ \sum_{k=1}^{C}l_j^k=1\}$。

本发明还提供了一种基于自适应比例学习的肿瘤组织病理分类方法，该方法包括如下步骤：

步骤 1：获取若干肿瘤组织病理切片并进行数字化扫描，将扫描后的病理图像按照分类任务目标类别进行人工标注，构建数据集；

步骤 2：将病理图像在不同的放大倍数下，根据设定的目标图像块重叠率计算病理图像的缩略图尺寸并利用缩略图的 RGB 通道和灰度值的差值分布特征分割肿瘤组织前景，去除肿瘤组织背景，对缩略图进行像素级分割，每个像素点对应一个图像块；

步骤 3：通过神经网络模型计算数据集中分割后的图像块在不同放大倍数下属于各个分类任务目标类别的概率，得到病理图像含有各个分类任务目标类别的比例，作为计算比例；

步骤 4：对于每个分类任务目标类别，分别将计算的不同放大倍数下概率最小值与不同放大倍数下概率最大值的二分之一进行比较，取其中较大值作为图像块在该分类任务目标类别下的整合概率；将所有图像块同一类别下的整合概率进行降序排列，根据标注的比例阈值，选择训练图像块，并对每个训练图像块赋值类别标签；

步骤 5：根据标注的比例阈值与计算比例的差值，对于病理图像中处于计算比例与比例阈值之间的训练图像块给予更高的训练权重，权重与差值呈正比，进行自适应权重赋值；使用权重调整后的训练图像块训练神经网络，损失函数为各级放大倍数损失函数项、整合损失函数项及比例拟合损失函数项之和，其中比例拟合损失函数项为若分类任务目标类别的计算比例低于比例阈值，则比例阈值与计算比例的差值作为比例拟合损失函数项，若计算比例大于等于比例阈值，则损失函数中比例拟合损失函数项为零；

步骤 6：根据训练结果调整比例阈值对模型迭代训练，通过训练好的神经网络模型计算病理图像含有各个分类任务目标类别的比例。

本发明的有益效果：本发明通过比例标签的标定标准，以明确最低比例为标注体系进行图像比例标签的标注，降低标注成本和对标注精度的要求，扩大数据利用率；通过有效前景组织的快速提取，完成多放大倍数图像块预分割与坐标定位；最后通过自适应比例学习，提升数据利用率和训练效率；并且通过多放大倍数整合训练，提高识别准确率。本发明在日常肿瘤组织的病理检查中，在尽可能低地增加额外工作负担的基础上，最大限度地提升检出率。

附图说明

图 1 为本发明基于自适应比例学习的肿瘤组织病理分类的流程示意图；

图 2 为本发明比例标签标定示意图；

图 3 为本发明图像前景分割结果示例图。

具体实施方式

以下结合附图对本发明具体实施方式作进一步详细说明。

如图 1 所示，本发明提供的一种基于自适应比例学习的肿瘤组织病理分类系统，该系统

包括比例标签标注模块、前景分割与图像块定位模块和自适应多放大倍数整合比例学习模块；

所述比例标签标注模块获取若干肿瘤组织病理切片并进行数字化扫描，将扫描后的病理图像按照分类任务目标类别进行人工标注，构建数据集；具体过程如下

总数据集记为$\boldsymbol{B}$，共包含 n 张经苏木精—伊红染色法（hematoxylin-eosin，HE)获得的肿瘤组织病理切片的全尺寸数字化扫描图像（Whole Slide Image, WSI）B_i，即$\boldsymbol{B}=\{B_1,\dots,B_i,\dots,B_n\}$。分类任务目标类别共计 C 类，针对每张病理图像$B_i$，构建比例标签向量为

$$\boldsymbol{R}_i=\left[R_i^1,\dots,R_i^k\dots,R_i^C\right]^T, k\in\{1,2,\dots,C\}, R_i^k\in[0,1] \tag{1}$$

其中，R_i^k为第 i 个样本图像中第 k 类的比例阈值，且满足$\sum_{k=1}^{C}R_i^k\leq 1$。比例阈值指人工粗略估计时的比例范围的最低值。假设B_i中第 k 类面积占比主观估计约占 50%，不经过详细勾画计算比例无法确定是 48%或 54%或是客观确定值，但大致在 40-60%之间，则标定R_i^k为$(0,40\%]$之内的任意数值均可，设定比例阈值为 40%。

以肝样分化和普通腺癌分化的分类目标为例，C=3，分别为非癌变组织C_0、普通腺癌组织C_1、肝样腺癌组织C_2。图 2 病理诊断为肝细胞肝癌，人工评估时粗略判断虚线框内为C_2，框外为C_0，两者面积相近，则赋予$R_i^{C_0}=40\%$，$R_i^{C_2}=40\%$，$R_i^{C_1}=0$。

所述前景分割与图像块定位模块将计算病理图像的缩略图尺寸并将缩略图分割出肿瘤组织前景，去除肿瘤组织背景，缩略图中每个像素点对应一个图像块；具体过程如下：

为得到不同放大倍数（op）下的图像块坐标，以 20X 放大倍数为基础，进行不同放大倍数的图像块分割。当放大倍数为 20X 时，目标图像块分辨率$R_{tar}=0.5\ \mu m/pixel$，记目标图像块尺寸为S_{tar} (pixels) * S_{tar} (pixels)，pixels 为像素，目标图像块重叠率为O_{tar}，对原始尺寸为W_{ori} (pixels) × H_{ori} (pixels)，原始分辨率为R_{ori}的病理图像B_i进行目标缩略图获取，W_{ori} (pixels)和H_{ori} (pixels)分别为原始病理图像尺寸的宽和高，则目标缩略图的宽W_{tar}和高H_{tar}尺寸分别为：

$$W_{tar}=2*\left\lfloor\frac{\left\lfloor\frac{W_{ori}}{2}\right\rfloor}{\left\lfloor\frac{R_{tar}}{R_{ori}}*S_{tar}*(1-O_{tar})\right\rfloor}\right\rfloor, H_{tar}=2*\left\lfloor\frac{\left\lfloor\frac{H_{ori}}{2}\right\rfloor}{\left\lfloor\frac{R_{tar}}{R_{ori}}*S_{tar}*(1-O_{tar})\right\rfloor}\right\rfloor \tag{2}$$

该缩略图对应病理图像B_i的目标分割区域左上角点坐标（W_{left},H_{up}）为：

$$(W_{left},H_{up})=\left(\left\lfloor\frac{W_{ori}}{2}\right\rfloor mod\left\lfloor\frac{R_{tar}}{R_{ori}}*S_{tar}*(1-O_{tar})\right\rfloor,\left\lfloor\frac{H_{ori}}{2}\right\rfloor mod\left\lfloor\frac{R_{tar}}{R_{ori}}*S_{tar}*(1-O_{tar})\right\rfloor\right)$$

其中，*mod* 表示取余数。目标缩略图中每一个像素点对应一个 20X 放大倍数下的训练图像块，图像块的尺寸为：

$$S_{20X} \times S_{20X} = \left\lfloor \frac{R_{tar}}{R_{ori}} * S_{tar} \right\rfloor * \left\lfloor \frac{R_{tar}}{R_{ori}} * S_{tar} \right\rfloor 。$$

利用RGB通道和灰度值的差值分布特征分割组织前景，去除肿瘤组织背景，对缩略图进行像素级分割，每个像素点对应一个图像块；具体过程如下：提取该缩略图的红（V_{Red}）、绿（V_{Green}）、蓝（V_{Blue}）通道值以及灰度图值（V_{Grey}）。依据 Otsu 阈值分割算法，分别计算 $V_{Red} - V_{Grey}$, $V_{Green} - V_{Grey}$, $V_{Blue} - V_{Grey}$, $V_{Red} - V_{Green}$, $V_{Blue} - V_{Green}$这五种差值密度分布的分割阈值，若该缩略图上的第 m 个像素点(a^m, b^m)通过上述5个分割阈值的数量大于等于3，则该像素点对应的图像块$T_i^m{}_{20X}$被视为组织前景，该图像块的坐标左上角为$(W_{left} + a^m * S_{20X}, H_{up} + b^m * S_{20X})$；否则，该像素点对应的图像块作为背景剔除。如图3所示，经过前景分割后，图像上的各轮廓与无效背景得以识别，仅保留前景组织（图3中最右侧图的非白色区域）。

以 20X 图像块区域为中心，取 10X 放大倍数下图像块$T_i^m{}_{10X}$，左上角坐标为$(W_{left} + \lfloor (a^m - 0.5) * S_{20X} \rfloor, H_{up} + \lfloor (b^m - 0.5) * S_{20X} \rfloor)$，图像大小为$2 * S_{20X} \times 2 * S_{20X}$；取 5X 放大倍数下图像块$T_i^m{}_{5X}$，左上角坐标为$(W_{left} + \lfloor (a^m - 1.5) * S_{20X} \rfloor, H_{up} + \lfloor (b^m - 1.5) * S_{20X} \rfloor)$，图像大小为$4 * S_{20X} \times 4 * S_{20X}$。

最终得去除背景图像块后的总数据集$\hat{\boldsymbol{B}}$，包含 n 个样本包$\hat{B}_i$，每个样本包内含N_i个示例图像块组，记为$\hat{B}_i = \{T_i^1, \ldots, T_i^m, \ldots, T_i^{N_i}\}, T_i^m = \{T_i^m{}_{5X}, T_i^m{}_{10X}, T_i^m{}_{20X}\}^T, i \in \{1, 2, \ldots, n\}$。

所述自适应多放大倍数整合比例学习模块组合了各级放大倍数和整合放大倍数的交叉熵函数形成损失函数，实现多放大倍数整合学习；通过自适应比例学习，自适应的动态调整图像块训练权重，并对神经网络进行迭代训练；具体过程如下：

A.各级放大倍数概率推算：利用神经网络模型$f(x)$（如ResNet等）对样本包中所有图像块进行概率推理计算，得到各图像的概率矩阵组$\mathrm{P}_i = f(\hat{B}_i) = \{\mathrm{P}_{i_{5X}}, \mathrm{P}_{i_{10X}}, \mathrm{P}_{i_{20X}}\}$，各放大倍数概率矩阵可表示为：

$$\mathrm{P}_{i_{op}} = \begin{bmatrix} p_1^1 \cdots p_1^k \cdots p_1^C \\ \vdots \ddots \vdots \ddots \vdots \\ p_j^1 \cdots p_j^k \cdots p_j^C \\ \vdots \ddots \vdots \ddots \vdots \\ p_{N_i}^1 \cdots p_{N_i}^k \cdots p_{N_i}^C \end{bmatrix}_{op}$$

图像块T_i^m在各放大倍数下归类为第 k 类概率为$p_{j\,op}^k$。

B.整合概率计算：由各级放大倍数概率可计算其在第 k 类上的多放大倍数整合概率为

$$\bar{p}_j^k = \max\left(\frac{\max\left(p_{j\,5X}^k,\ p_{j\,10X}^k, p_{j\,20X}^k\right)}{2}, \min\left(p_{j\,5X}^k, p_{j\,10X}^k, p_{j\,20X}^k\right)\right)$$

C.各类概率排序与标签赋值：根据第 k 类对应多放大倍数整合概率进行降序排列，得 $\{\bar{p}_1^k, \ldots, \bar{p}_j^k, \ldots, \bar{p}_{N_i}^k\}$，此时$\bar{p}_1^k$对应的图像块在第 k 类中概率最高，$\bar{p}_{N_i}^k$对应的图像块在第 k 类中概率最低。对于B_i中应当归类为第 k 类的训练图像块：

$$S_i^k = \left\{\bar{p}_1^{k\prime}, \ldots, \bar{p}_j^{k\prime}, \ldots, \bar{p}_{N_i * R_i^k}^{k\prime}\right\} \tag{3}$$

其中，$\bar{p}_j^{k\prime}$表示$\bar{p}_j^k$对应的图像块；赋值类别标签 $l = \{l_j^k \mid l_j^k \in \{0,1\},\ k = 1, \ldots, C,\ \sum_{k=1}^{C} l_j^k = 1\}$。

D. 各级放大倍数损失函数：将所有得到赋值类别标签的训练图像块$S_i = \cup_{k=1}^{C} S_i^k$计算交叉墒，即

$$\mathrm{Loss}_{op} = -\frac{1}{S_i}\sum_{j=1}^{S_i}\sum_{k=1}^{C} \omega^k l_j^k \log\left(p_{j\,op}^k\right) \tag{4}$$

Loss_{op}为在放大倍数op下的损失函数，ω^k为第 k 类训练图像块的训练权重。

E.整合损失函数：整合损失函数项定义为：

$$Loss_{combine} = -\frac{1}{S_i}\sum_{j=1}^{S_i}\sum_{k=1}^{C} \omega^k l_j^k log\left(\bar{p}_j^k\right) \tag{5}$$

F.训练图像块权重调整：对B_i中低于比例标签阈值的差值训练图像块给予更高的训练权重，权重与差值呈正比，进行自适应权重赋值。即，若$\frac{1}{N_i}\sum_{j=1}^{N_i} p_j^k < R_i^k$，将

$$\omega_j^k = \begin{cases} \left(\dfrac{R_i^k * N_i}{\sum_{j=1}^{N_i} \bar{p}_j^k}\right)\omega^k, & \sum_{j=1}^{N_i} \bar{p}_j^k < \mathrm{j} \le R_i^k * N_i \\ \omega^k, & \mathrm{j} \le \sum_{j=1}^{N_i} \bar{p}_j^k \end{cases} \tag{6}$$

其中，ω_j^k为处于计算比例和比例阈值之间的训练图像块的训练权重，替换公式(4)、(5)中的ω^k项。

G.比例拟合损失函数：通过随机升采样（若$\hat{B}_i$中训练图像块数量小于单训练批次训练图像块数量）或者降采样（若$\hat{B}_i$中训练图像块数量大于单训练批次训练图像块数量），使得一个训练批次中的训练图像块均来自$\hat{B}_i$，第 k 的组织比例若低于比例标签，则有比例拟合损失函数项定义为：

$$Loss_{ratio}^{k}=\begin{cases}R_i^k-\frac{1}{N_i}\sum_{j=1}^{N_i}\vec{p}_j^k, & \frac{1}{N_i}\sum_{j=1}^{N_i}\vec{p}_j^k<R_i^k\\ 0, & \frac{1}{N_i}\sum_{j=1}^{N_i}\vec{p}_j^k\geq R_i^k\end{cases} \tag{7}$$

H.比例标签动态更新：由于人工标记的比例标签纪录的明确最低比例，并不等同于客观比例。因此在训练迭代模型参数前的样本赋予标签步骤中，根据前一次模型计算的各类组织比例小幅度调整比例标签，调整率α根据实验选择，本发明中设置为$\alpha=1\%$，进行自适应比例标签调整，以公式(8)中的$R_i^{k\,epoch}$更新公式(3)、(6)、(7)中的R_i^k。

$$R_i^{k\,epoch}=\begin{cases}R_i^{k\,epoch-1}, & \frac{1}{N_i}\sum_{j=1}^{N_i}\vec{p}_j^k<R_i^{k\,epoch-1}+\alpha\\ R_i^{k\,epoch-1}+\alpha, & \frac{1}{N_i}\sum_{j=1}^{N_i}\vec{p}_j^k\geq R_i^{k\,epoch-1}+\alpha\end{cases} \tag{8}$$

其中，$R_i^{k\,epoch}$表示调整后的比例阈值，$R_i^{k\,epoch-1}$表示调整前的比例阈值。

I.最终损失函数Loss_{total}为各级放大倍数损失函数项、整合损失函数项及比例拟合损失函数项之和，即

$$\mathrm{Loss}_{total}=\mathrm{Loss}_{5X}+\mathrm{Loss}_{10X}+\mathrm{Loss}_{20X}+\mathrm{Loss}_{combine}+\sum_{k=1}^{c}Loss_{ratio}^{k} \tag{9}$$

J.最佳模型参数状态选择方式：通过N^{epoch}次步骤 A-I 的循环训练迭代后，Loss_{total}将趋于稳定。对于第 epoch 次模型，其计算概率P_i与标定标签$\boldsymbol{R}_i$之间的匹配度为：

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{c}\min\left(\frac{\frac{1}{N_i}\sum_{j=1}^{N_i}\vec{p}_j^k}{R_i^k},1\right)$$

匹配度越高，模型识别效果越好。计算N^{epoch}次模型匹配度，选定具有最高匹配度的模型参数作为最佳模型参数。

K.测试/应用：对于测试数据$\vec{B}$，经过前景分割后得测试示例集$\hat{\vec{B}}$，利用最佳模型对样本包中所有图像块进行概率推理计算，$\vec{\mathrm{P}}_i=f(\hat{\vec{B}}_i)=\{\vec{\mathrm{P}}_{i\,5X},\vec{\mathrm{P}}_{i\,10X},\vec{\mathrm{P}}_{i\,20X}\}$，表示各放大倍数下的分类概率。$\vec{\vec{p}}_j^{\,k}=\min\left(\vec{p}^{\,k}_{j\,5X},\vec{p}^{\,k}_{j\,10X},\vec{p}^{\,k}_{j\,20X}\right)$表示 WSI 上图像块对于第 k 类分化的多放大倍数整合概率。

应用在肿瘤组织病理分类问题时，$\vec{\vec{p}}_j^{\,k}$高于分类阈值（一般定为 0.5）则分类至第 k 类分化，

若 WSI 中含有至少一个图像块分类为该类分化，则判断该组织具有该类分化。应用在肿瘤组织病理分类量化问题时， WSI 含有第 k 类分化的比例为$\frac{1}{N_t}\sum_{j=1}^{N_t}\tilde{p}_j^k$。

本发明还提供了一种基于自适应比例学习的肿瘤组织病理分类方法，其特征在于，该方法包括如下步骤：

步骤 1：获取若干肿瘤组织病理切片并进行数字化扫描，将扫描后的病理图像按照分类任务目标类别进行人工标注，构建数据集；该步骤具体参考比例标签标注模块的实现过程；

步骤 2：将病理图像在不同的放大倍数下，根据设定的目标图像块重叠率计算病理图像的缩略图尺寸并利用缩略图的 RGB 通道和灰度值的差值分布特征分割肿瘤组织前景，去除肿瘤组织背景，对缩略图进行像素级分割，每个像素点对应一个图像块；该步骤具体参考前景分割与图像块定位模块的实现过程；

步骤 3：通过神经网络模型计算数据集中分割后的图像块在不同放大倍数下属于各个分类任务目标类别的概率，得到病理图像含有各个分类任务目标类别的比例，作为计算比例；

步骤 4：对于每个分类任务目标类别，分别将计算的不同放大倍数下概率最小值与不同放大倍数下概率最大值的二分之一进行比较，取其中较大值作为图像块在该分类任务目标类别下的整合概率；将所有图像块同一类别下的整合概率进行降序排列，根据标注的比例阈值，选择训练图像块，并对每个训练图像块赋值类别标签；

步骤 5：根据标注的比例阈值与计算比例的差值，对于病理图像中处于计算比例与比例阈值之间的训练图像块给予更高的训练权重，权重与差值呈正比，进行自适应权重赋值；使用权重调整后的训练图像块训练神经网络，损失函数为各级放大倍数损失函数项、整合损失函数项及比例拟合损失函数项之和，其中比例拟合损失函数项为若分类任务目标类别的计算比例低于比例阈值，则比例阈值与计算比例的差值作为比例拟合损失函数项，若计算比例大于等于比例阈值，则损失函数中比例拟合损失函数项为零；

步骤 6：根据训练结果调整比例阈值对模型迭代训练，通过训练好的神经网络模型计算病理图像含有各个分类任务目标类别的比例。

步骤 3~步骤 6 具体参考自适应多放大倍数整合比例学习模块的实现过程。

本发明实施例中收集 TCGA（The Cancer Genome Atlas）数据库中的肝细胞肝癌、食管腺癌、胃腺癌、结肠腺癌、直肠腺癌、胰腺癌、宫颈腺癌、食管鳞癌、头颈鳞癌、宫颈鳞癌、肉瘤等数据集。对所有图像进行比例标签标定，以肝样分化和普通腺癌分化的分类目标为例，C=3，分别为非癌变组织C_0、普通腺癌组织C_1、肝样腺癌组织C_2。其中仅有肝细胞肝癌中的图像含有C_2组织，其余数据均不含。

经过数据预处理和模型训练后，将最佳模型应用于三家三甲医院收集的胃肠道腺癌数据

（分别有 135、95、351 例），识别其是否含有肝样分化，即肝样腺癌特征。本发明针对胃肠道肝样腺癌的识别任务，在三家三甲医院病理科的日常病理检查中，胃肠道肝样分化检出率为 0%、8.5%、56%，本方法构建模型检出率可达 100%、80%、93.8%，假阳性均在 10%左右，在尽可能低地增加额外工作负担的基础上，最大限度地提升检出率。

上述实施例用来解释说明本发明，而不是对本发明进行限制，在本发明的精神和权利要求的保护范围内，对本发明作出的任何修改和改变，都落入本发明的保护范围。

权 利 要 求 书

1、一种基于自适应比例学习的肿瘤组织病理分类系统，其特征在于，该系统包括比例标签标注模块、前景分割与图像块定位模块和自适应多放大倍数整合比例学习模块；

所述比例标签标注模块获取若干肿瘤组织病理切片并进行数字化扫描，将扫描后的病理图像按照分类任务目标类别进行人工标注，构建数据集；

所述前景分割与图像块定位模块将病理图像在不同的放大倍数下，根据设定的目标图像块重叠率计算病理图像的缩略图尺寸并利用缩略图的 RGB 通道和灰度值的差值分布特征分割肿瘤组织前景，去除肿瘤组织背景，缩略图中每个像素点对应一个图像块；

所述自适应多放大倍数整合比例学习模块通过神经网络模型计算数据集中分割后的图像块在不同放大倍数下属于各个分类任务目标类别的概率，得到病理图像含有各个分类任务目标类别的比例，作为计算比例，对于每个分类任务目标类别，分别将计算的不同放大倍数下概率最小值与不同放大倍数下概率最大值的二分之一进行比较，取其中较大值作为图像块在该分类任务目标类别下的整合概率；将所有图像块同一类别下的整合概率进行降序排列，根据标注的比例阈值，选择训练图像块，并对每个训练图像块赋值类别标签；根据标注的比例阈值与计算比例的差值，对于病理图像中处于计算比例与比例阈值之间的训练图像块给予更高的训练权重，权重与差值呈正比，进行自适应权重赋值；使用权重调整后的训练图像块训练神经网络，损失函数为各级放大倍数损失函数项、整合损失函数项及比例拟合损失函数项之和，其中比例拟合损失函数项为若分类任务目标类别的计算比例低于比例阈值，则比例阈值与计算比例的差值作为比例拟合损失函数项，若计算比例大于等于比例阈值，则损失函数中比例拟合损失函数项为零，并根据训练结果调整比例阈值对模型迭代训练，通过训练好的神经网络模型计算病理图像含有各个分类任务目标类别的比例。

2、根据权利要求 1 所述的一种基于自适应比例学习的肿瘤组织病理分类系统，其特征在于，所述比例标签标注模块共获取 n 张病理切片的全尺寸数字化扫描图像，构成数据集记为$\boldsymbol{B}=\{B_1,\ldots,B_i,\ldots,B_n\}$，其中$B_i$为第$i$张病理图像；分类任务目标类别共计 C 类，针对第$i$张病理图像$B_i$，构建比例标签向量$\boldsymbol{R}_i$为：

$$\boldsymbol{R}_i=\left[R_i^1,\ldots,R_i^k\ldots,R_i^C\right]^T, k\in\{1,2,\ldots,C\}, R_i^k\in[0,1]$$

其中，R_i^k为第 i 个病理图像中第 k 类的比例阈值，且满足$\sum_{k=1}^{C}R_i^k\leq 1$，$T$ 表示转置。

3、根据权利要求 1 所述的一种基于自适应比例学习的肿瘤组织病理分类系统，其特征在于，所述前景分割与图像块定位模块计算病理图像的缩略图尺寸具体过程如下：记目标图像块尺寸为S_{tar} (pixels) * S_{tar} (pixels)，pixels为像素，目标图像块分辨率为R_{tar}，目标图像块

重叠率为O_{tar}，对原始尺寸为W_{ori} (pixels) × H_{ori} (pixels)，原始分辨率为R_{ori}的病理图像B_i进行目标缩略图获取，W_{ori} (pixels)和H_{ori} (pixels)分别为原始病理图像尺寸的宽和高，则目标缩略图的宽和高尺寸分别为：

$$W_{tar} = 2 * \left\lfloor \frac{\left\lfloor \frac{W_{ori}}{2} \right\rfloor}{\left\lfloor \frac{R_{tar}}{R_{ori}} * S_{tar} * (1 - O_{tar}) \right\rfloor} \right\rfloor, H_{tar} = 2 * \left\lfloor \frac{\left\lfloor \frac{H_{ori}}{2} \right\rfloor}{\left\lfloor \frac{R_{tar}}{R_{ori}} * S_{tar} * (1 - O_{tar}) \right\rfloor} \right\rfloor$$

目标缩略图对应病理图像B_i的目标分割区域左上角点坐标为$(W_{left}, H_{up}) = (\lfloor \frac{W_{ori}}{2} \rfloor mod \left\lfloor \frac{R_{tar}}{R_{ori}} * S_{tar} * (1 - O_{tar}) \right\rfloor, \lfloor \frac{H_{ori}}{2} \rfloor mod \left\lfloor \frac{R_{tar}}{R_{ori}} * S_{tar} * (1 - O_{tar}) \right\rfloor)$；其中，$mod$表示取余数，将缩略图的目标分割区域，进行像素级分割，目标缩略图中每一个像素点对应一个的图像块，图像块的尺寸为$\left\lfloor \frac{R_{tar}}{R_{ori}} * S_{tar} \right\rfloor * \left\lfloor \frac{R_{tar}}{R_{ori}} * S_{tar} \right\rfloor$。

4、根据权利要求 1 所述的一种基于自适应比例学习的肿瘤组织病理分类系统，其特征在于，所述前景分割与图像块定位模块利用缩略图的 RGB 通道和灰度值的差值分布特征分割肿瘤组织前景具体过程如下：提取缩略图红V_{Red}、绿V_{Green}、蓝V_{Blue}通道值以及灰度图值V_{Grey}，计算$V_{Red} - V_{Grey}$，$V_{Green} - V_{Grey}$，$V_{Blue} - V_{Grey}$，$V_{Red} - V_{Green}$，$V_{Blue} - V_{Green}$这五种差值密度分布的分割阈值；若缩略图上的一像素点超过上述五个分割阈值的数量大于等于 3，则该像素点对应的图像块被视为组织前景，否则，该像素点对应的图像块作为背景剔除。

5、根据权利要求 2 所述的一种基于自适应比例学习的肿瘤组织病理分类系统，其特征在于，所述损失函数中各级放大倍数损失函数项如下：

$$\text{Loss}_{op} = \begin{cases} -\frac{1}{S_i} \sum_{j=1}^{S_i} \sum_{k=1}^{C} \omega_j^k l_j^k log\left(p^k_{j\ op}\right), & \sum_{j=1}^{N_i} \bar{p}_j^k < j \leq R_i^k * N_i \\ -\frac{1}{S_i} \sum_{j=1}^{S_i} \sum_{k=1}^{C} \omega^k l_j^k log\left(p^k_{j\ op}\right), & j \leq \sum_{j=1}^{N_i} \bar{p}_j^k \end{cases}$$

其中，op表示不同的放大倍数，ω^k为第 k 类计算比例中的训练图像块的训练权重，ω_j^k为处于计算比例和比例阈值之间的训练图像块的训练权重；$S_i = \cup_{k=1}^{C} S_i^k$，$S_i^k$为第 k 类训练图像块的集合，$\bar{p}_j^k$为第 k 类中第 j 个图像块的整合概率，l_j^k为第 k 类中第 j 个图像块的标签；$p^k_{j\ op}$为图像块在各放大倍数下归类为第 k 类的概率；N_i为第 i 个病理图像中前景图像块的数量。

6、根据权利要求 5 所述的一种基于自适应比例学习的肿瘤组织病理分类系统，其特征在于，所述损失函数中整合损失函数项$Loss_{combine}$如下：

$$Loss_{combine}=\begin{cases}-\frac{1}{S_i}\sum_{j=1}^{S_i}\sum_{k=1}^{C}\omega_j^k l_j^k log(\bar{p}_j^k), & \sum_{j=1}^{N_i}\bar{p}_j^k < j \le R_i^k * N_i \\ -\frac{1}{S_i}\sum_{j=1}^{S_i}\sum_{k=1}^{C}\omega^k l_j^k log(\bar{p}_j^k), & j \le \sum_{j=1}^{N_i}\bar{p}_j^k\end{cases}$$

。

7、根据权利要求 5 所述的一种基于自适应比例学习的肿瘤组织病理分类系统，其特征在于，所述损失函数中比例拟合损失函数项为根据去除背景图像块后的病理图像中的训练图像块数量，通过随机升采样或者随机降采样，使得一个训练批次中的训练图像块均来自同一张病理图像：比例拟合损失函数项$Loss_{ratio}^k$定义为：

$$Loss_{ratio}^k=\begin{cases}R_i^k-\frac{1}{N_i}\sum_{j=1}^{N_i}\bar{p}_j^k, & \frac{1}{N_i}\sum_{j=1}^{N_i}\bar{p}_j^k < R_i^k \\ 0, & \frac{1}{N_i}\sum_{j=1}^{N_i}\bar{p}_j^k \ge R_i^k\end{cases}$$

。

8、根据权利要求 5 所述的一种基于自适应比例学习的肿瘤组织病理分类系统，其特征在于，所述自适应多放大倍数整合比例学习模块根据神经网络训练结果调整比例阈值具体公式如下：

$$R_i^{k^{epoch}}=\begin{cases}R_i^{k^{epoch-1}}, & \frac{1}{N_i}\sum_{j=1}^{N_i}\bar{p}_j^k < R_i^{k^{epoch-1}}+\alpha \\ R_i^{k^{epoch-1}}+\alpha, & \frac{1}{N_i}\sum_{j=1}^{N_i}\bar{p}_j^k \ge R_i^{k^{epoch-1}}+\alpha\end{cases}$$

其中，$R_i^{k^{epoch}}$表示调整后的比例阈值，$R_i^{k^{epoch-1}}$表示调整前的比例阈值，α为调整率。

9、根据权利要求 5 所述的一种基于自适应比例学习的肿瘤组织病理分类系统，其特征在于，训练图像块的具体选择过程如下：根据分类任务目标类别中的第 k 类对应多放大倍数整合概率进行降序排列，得$\{\bar{p}_1^k,\dots,\bar{p}_j^k,\dots,\bar{p}_{N_i}^k\}$，此时$\bar{p}_1^k$对应的图像块在第 k 类中概率最高，$\bar{p}_{N_i}^k$对应的图像块在第 k 类中概率最低；对于病理图像B_i中归类为第 k 类的图像块，根据标注的比例阈值，选择在比例阈值内的图像块作为训练图像块：

$$S_i^k=\left\{\bar{p}_1^{k\prime},\dots,\bar{p}_j^{k\prime},\dots,\bar{p}_{N_i*R_i^k}^{k\prime}\right\}$$

其中，$\bar{p}_j^{k\prime}$表示$\bar{p}_j^k$对应的训练图像块；赋值类别标签

$I = \{l_j^k \mid l_j^k \in \{0,1\},\ k = 1,\ldots,C,\ \sum_{k=1}^{C} l_j^k = 1\}$。

10、一种基于自适应比例学习的肿瘤组织病理分类方法，其特征在于，该方法包括如下步骤：

步骤 1：获取若干肿瘤组织病理切片并进行数字化扫描，将扫描后的病理图像按照分类任务目标类别进行人工标注，构建数据集；

步骤 2：将病理图像在不同的放大倍数下，根据设定的目标图像块重叠率计算病理图像的缩略图尺寸并利用缩略图的 RGB 通道和灰度值的差值分布特征分割肿瘤组织前景，去除肿瘤组织背景，对缩略图进行像素级分割，每个像素点对应一个图像块；

步骤 3：通过神经网络模型计算数据集中分割后的图像块在不同放大倍数下属于各个分类任务目标类别的概率，得到病理图像含有各个分类任务目标类别的比例，作为计算比例；

步骤 4：对于每个分类任务目标类别，分别将计算的不同放大倍数下概率最小值与不同放大倍数下概率最大值的二分之一进行比较，取其中较大值作为图像块在该分类任务目标类别下的整合概率；将所有图像块同一类别下的整合概率进行降序排列，根据标注的比例阈值，选择训练图像块，并对每个训练图像块赋值类别标签；

步骤 5：根据标注的比例阈值与计算比例的差值，对于病理图像中处于计算比例与比例阈值之间的训练图像块给予更高的训练权重，权重与差值呈正比，进行自适应权重赋值；使用权重调整后的训练图像块训练神经网络，损失函数为各级放大倍数损失函数项、整合损失函数项及比例拟合损失函数项之和，其中比例拟合损失函数项为若分类任务目标类别的计算比例低于比例阈值，则比例阈值与计算比例的差值作为比例拟合损失函数项，若计算比例大于等于比例阈值，则损失函数中比例拟合损失函数项为零；

步骤 6：根据训练结果调整比例阈值对模型迭代训练，通过训练好的神经网络模型计算病理图像含有各个分类任务目标类别的比例。

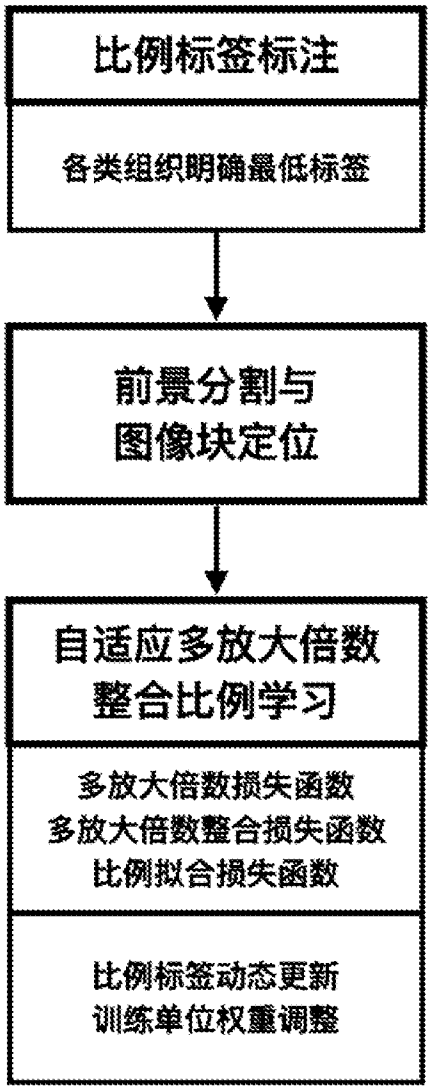

比例标签标注
各类组织明确最低标签
前景分割与
图像块定位
自适应多放大倍数
整合比例学习
多放大倍数损失函数
多放大倍数整合损失函数
比例拟合损失函数
比例标签动态更新
训练单位权重调整

图 1

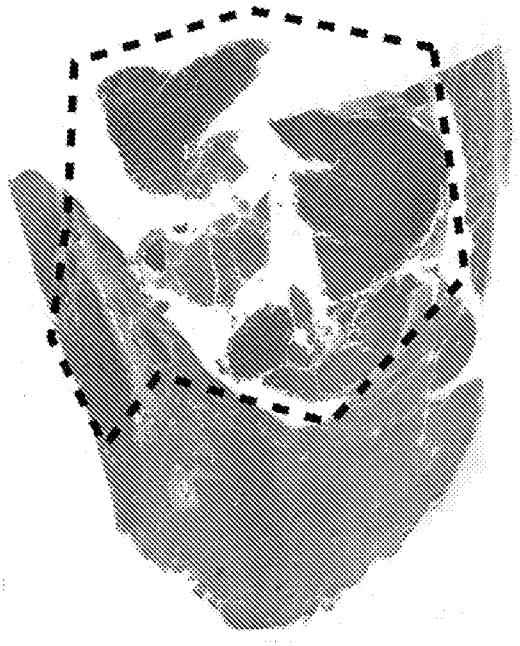


图 2

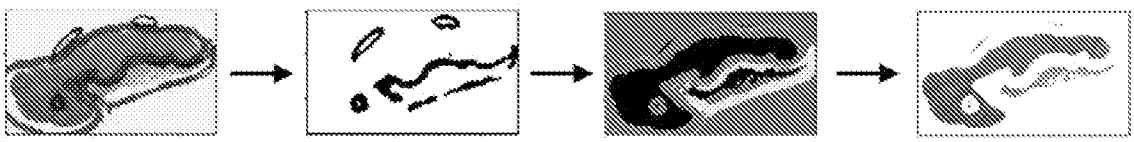


图 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2022/124693

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/62(2022.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS; CNTXT; VEN; USTXT; EPTXT; WOTXT; CNKI; GOOGLE: 自适应, 比例, 学习, 图像, 切片, 分类, 分割, 放大, 概率, 肿瘤, adaptive, proportion, learning, image, slice, class, segmentation, magnify, probability, tumor

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 113723573 A (ZHEJIANG UNIVERSITY) 30 November 2021 (2021-11-30) claims 1-10, and description, paragraphs 20-38	1-10
A	YE,Qiancheng et al. "Method of Tumor Pathological Micronecrosis Quantification Via Deep Learning From Label Fuzzy Proportions" *IEEE Journal of Biomedical and Health Informatics,* Vol. 25, No. 9, 30 September 2021 (2021-09-30), ISSN: 2168-2194, pages 3288-3298	1-10
A	CN 109528230 A (JINAN INSPUR HI-TECH INVESTMENT AND DEVELOPMENT CO., LTD.) 29 March 2019 (2019-03-29) description, paragraphs 2-30	1-10
A	US 2019156159 A1 (KOPPARAPU, K. V. K. S.) 23 May 2019 (2019-05-23) entire document	1-10

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
08 December 2022	**04 January 2023**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration (ISA/CN)** **No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088, China** Facsimile No. **(86-10)62019451**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2022/124693

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	113723573	A	30 November 2021	CN	113723573	B	14 January 2022
CN	109528230	A	29 March 2019	CN	109528230	B	17 August 2021
US	2019156159	A1	23 May 2019	US	10748040	B2	18 August 2020

国际检索报告

国际申请号
PCT/CN2022/124693

A. 主题的分类

G06K 9/62(2022.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06K

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

CNABS;CNTXT;VEN;USTXT;EPTXT;WOTXT;CNKI;GOOGLE:自适应，比例，学习，图像，切片，分类，分割，放大，概率，肿瘤，adaptive，proportion，learning，image，slice，class，segmentation，magnify，probability，tumor

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
PX	CN 113723573 A (浙江大学) 2021年11月30日 (2021-11-30) 权利要求1-10、说明书第20-38段	1-10
A	YE,Qiancheng 等. "Method of Tumor Pathological Micronecrosis Quantification Via Deep Learning From Label Fuzzy Proportions" IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, 第25卷，第9期，2021年9月30日 (2021-09-30), ISSN：2168-2194, 第3288-3298页	1-10
A	CN 109528230 A (济南浪潮高新科技投资发展有限公司) 2019年3月29日 (2019-03-29) 说明书第2-30段	1-10
A	US 2019156159 A1 (KOPPARAPU,Kavya Venkata Kota Sai) 2019年5月23日 (2019-05-23) 全文	1-10

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

"A" 认为不特别相关的表示了现有技术一般状态的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

"&" 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2022年12月8日	2023年1月4日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 薛双双 电话号码 86-(20)-28958967

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2022/124693

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	113723573	A	2021年11月30日	CN	113723573	B	2022年1月14日
CN	109528230	A	2019年3月29日	CN	109528230	B	2021年8月17日
US	2019156159	A1	2019年5月23日	US	10748040	B2	2020年8月18日