



(12) 发明专利申请

(10) 申请公布号 CN 112101451 A

(43) 申请公布日 2020.12.18

(21) 申请号 202010960168.2

(22) 申请日 2020.09.14

(71) 申请人 北京联合大学

地址 100101 北京市朝阳区北四环东路97号

(72) 发明人 杨萍 满芮 季程雨 芦博
李欣桐

(74) 专利代理机构 北京驰纳智财知识产权代理
事务所(普通合伙) 11367

代理人 吴建国

(51) Int.Cl.

G06K 9/62 (2006.01)

G06N 3/04 (2006.01)

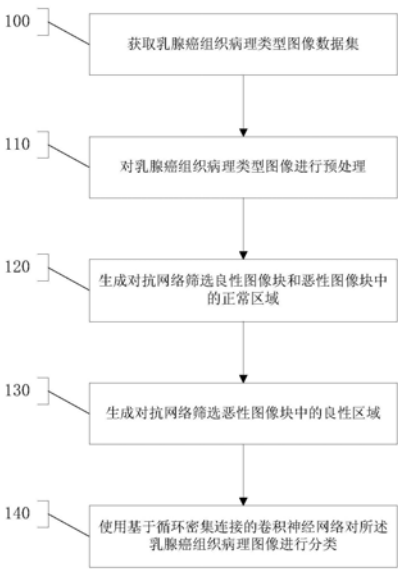
权利要求书2页 说明书13页 附图3页

(54) 发明名称

一种基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法

(57) 摘要

本发明提供一种基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法,包括获取乳腺癌组织病理类型图像数据集,还包括以下步骤:对乳腺癌组织病理类型图像进行预处理;生成对抗网络筛选良性图像块和恶性图像块中的正常区域;生成对抗网络筛选恶性图像块中的良性区域;使用基于循环密集连接的卷积神经网络对所述乳腺癌组织病理图像进行分类。本发明采用改进的无监督的生成对抗网络分别学习正常病理和良性肿瘤病理图像的数据分布,从而可以筛选恶性肿瘤病理图像中存在良性肿瘤的区域和正常区域,良性肿瘤病理图像中存在的正常区域,最大程度辅助医生更精确、更快速地诊断病情提供可能。



1. 一种基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法,包括获取乳腺癌组织病理类型图像数据集,其特征在于,还包括以下步骤:

步骤1:对乳腺癌组织病理类型图像进行预处理;

步骤2:生成对抗网络筛选良性图像块和恶性图像块中的正常区域;

步骤3:生成对抗网络筛选恶性图像块中的良性区域;

步骤4:使用基于循环密集连接的卷积神经网络对所述乳腺癌组织病理图像进行分类。

2. 如权利要求1所述的基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法,其特征在于:所述步骤1包括以下子步骤:

步骤11:对所述乳腺癌组织病理类型图像数据集进行染色标准化处理;

步骤12:将所述乳腺癌组织病理类型图像数据集划分为训练集、验证集和测试集;将所述训练集和所述验证集中正常图像进行随机图像块采样策略,得到正常图像块;

步骤13:使用所述正常图像块训练无监督的生成对抗网络。

3. 如权利要求2所述的基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法,其特征在于:所述生成对抗网络由两个对抗的模块,生成网络G和判别网络D组成。

4. 如权利要求3所述的基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法,其特征在于:所述生成对抗网络在训练时的目标函数为:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim P_{data}} [\log D(x)] + \mathbb{E}_{z \sim P_z(x)} [\log (1 - D(G(z)))]$$

其中,x为真实样本, P_{data} 为真实数据分布, $P_z(x)$ 为先验分布., $\mathbb{E}_{x \sim P_{data}}$ 为x来自真实数据分布 P_{data} 的计算期望,z为随机变量, $\mathbb{E}_{z \sim P_z(x)}$ 为z来自先验分布 $P_z(x)$ 的计算期望,D(x)为真实样本x在判别网络D上的表现,G(z)为生成样本。

5. 如权利要求4所述的基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法,其特征在于:所述步骤2包括使用所述训练集和所述验证集中的所述良性图像进行随机图像块采样策略,使用所述生成对抗网络对所述良性图像块进行筛选,大于第一阈值的图像块为错误标签的图像块,小于第一阈值的图像块为正确标签的图像块,筛选出良性肿瘤中存在的正常区域。

6. 如权利要求5所述的基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法,其特征在于:所述步骤2还包括所述生成对抗网络自动学习良性图像块样本数据分布,生成带有良性属性的生成对抗网络。

7. 如权利要求6所述的基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法,其特征在于:所述步骤2还包括使用所述训练集和所述验证集中的所述恶性图像进行随机图像块采样策略,使用所述生成对抗网络对恶性图像块进行初次筛选,大于第一阈值的图像块为错误标签的图像块,小于第一阈值的图像块为正确标签的图像块,筛选出恶性肿瘤中存在的正常区域。

8. 如权利要求7所述的基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法,其特征在于:所述步骤3包括使用所述带有良性属性的生成对抗网络对的恶性图像块进行二次筛选,大于第二阈值的图像块为错误标签的图像块,小于第二阈值的图像块为正确标签的图像块,筛选出恶性肿瘤中存在的良性区域。

9. 如权利要求8所述的基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法,其特征在于:所述步骤3还包括对于给定测试图像块 x' ,在潜在空间中找到一个点 z' ,该点与图像 $G(z)$ 对应,该图像在视觉上与异常图像 x' 最相似,并且位于良性图像分布 X 上。

10. 如权利要求9所述的基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法,其特征在于:所述点 z' 的获取方法为:

步骤31:定义一个损失函数,代表潜在空间向量映射到图像的损失;

步骤32:在潜在空间分布 Z 中随机采样 z'_1 并将其馈送到训练有素的生成器中以获得生成的图像 $G(z'_1)$,代入损失函数计算损失;

步骤33:计算损失函数关于的梯度,利用梯度下降法不断更新迭代,迭代过程中通过 $\delta = 1, 2, \dots, \Gamma$ 来优化 z 在潜空间 Z 中的位置,直到最相似的图像 $G(z'_\Gamma)$,其中, δ 为迭代次数, Γ 为迭代总次数。

一种基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法

技术领域

[0001] 本发明涉及图像特征描述的技术领域,特别是一种基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法。

背景技术

[0002] 乳腺癌是全球女性中最常见的癌症,每年约有210万名女性罹患乳腺癌。乳腺癌是一种十分严重的疾病,癌细胞在体内不受限制地生长,超出其细胞界限以侵入相邻部位或扩散到其他器官。根据美国癌症协会提供的最近数据,在美国,到2020年预计将诊断出276480例新发浸润性乳腺癌女性病例,以及48530例新发非浸润性乳腺癌病例。预计到2020年底,美国约有42170名妇女死于乳腺癌。

[0003] 由于乳腺癌的高死亡率,建议女性通过乳房X线照片和计算机断层扫描(CT)进行定期检查。但是上述两种检查仅属于定位性质的检查,并不能根据观察到的异常判断患者罹患乳腺癌。乳腺癌组织病理学检查是乳腺癌诊断的“金标准”。由于缺乏经验丰富的病理学专家进行疾病诊断,并且病理学专家的过度疲劳往往会导致误诊,因此迫切需要开发基于计算机辅助诊断乳腺癌病理类型的系统。针对目前的研究方法,大致可以分为两种:基于传统机器学习的乳腺癌组织病理类型分类和基于深度学习的乳腺癌组织病理类型分类。其中,基于传统机器学习算法的乳腺癌组织病理类型分类方法需要通过人工特征提取图像中感兴趣区域的图像特征,再将提取的特征送入常见的机器学习分类器,如支持向量机、贝叶斯分类器和决策树。在基于深度学习的算法中,常见的分类网络AlexNet、VGG16、VGG19、GoogLeNet等卷积神经网络被应用于乳腺癌组织病理类型的分类,与基于传统机器学习算法的乳腺癌组织病理类型分类方法相比,基于深度学习的算法可以最大程度辅助医生更精确、更快速地诊断病情。由于深度学习理论的快速发展,目前的研究已经表明乳腺癌组织病理学图像的分类采用深度学习的方法可以极大的提高分类的准确性,从而帮助医生进行诊断,更好地为病人进行治疗。然而,由于乳腺癌组织病理学图像的分辨率极高,采用基于图像块的乳腺癌组织病理图像深度学习分类方法非常具有挑战性。这是由于在基于图像块的乳腺癌组织病理图像深度学习分类方法中,往往将全切片数字病理图像的标签分配给图像块,但是在恶性全切片数字病理组织学图像中存在良性区域和正常区域,良性全切片数字病理组织图像中存在正常区域,因此只有部分提取的图像块被正确标记。

[0004] 2020年8月的IEEE Access上刊登了Rui Man, Ping Yang, Bowen Xu的文章《Classification of Breast Cancer Histopathological Images Using Discriminative Patches Screen by Generative Adversarial Networks》,该方法采用无监督异常检测生成对抗网络学习良性肿瘤图像的数据分布,从而可以筛选恶性肿瘤中存在良性肿瘤的区域,同时采用密集连接网络作为分类网络,实现多层特征融合进一步提高分类的准确性。其缺点是无法筛选良性图像和恶性图像中的正常区域,影响训练的效果;并且密集连接网络会存在参数过多不易训练的问题。

发明内容

[0005] 为了解决上述的技术问题,本发明提出一种基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法,采用改进的无监督的生成对抗网络分别学习正常病理和良性肿瘤病理图像的数据分布,从而可以筛选恶性肿瘤病理图像中存在良性肿瘤的区域和正常区域,良性肿瘤病理图像中存在的正常区域;同时采用基于循环密集连接的卷积神经网络作为分类网络,一方面密集连接可实现多层特征复用和融合,另一方面通过循环结构可以在实际训练过程中节省存储空间,有利于进一步提高分类的准确性和效率。为最大程度辅助医生更精确、更快速地诊断病情提供可能。

[0006] 本发明提供一种基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法,包括获取乳腺癌组织病理类型图像数据集,还包括以下步骤:

[0007] 步骤1:对乳腺癌组织病理类型图像进行预处理;

[0008] 步骤2:生成对抗网络筛选良性图像块和恶性图像块中的正常区域;

[0009] 步骤3:生成对抗网络筛选恶性图像块中的良性区域;

[0010] 步骤4:使用基于循环密集连接的卷积神经网络对所述乳腺癌组织病理图像进行分类。

[0011] 优选的是,所述步骤1包括以下子步骤:

[0012] 步骤11:对所述乳腺癌组织病理类型图像数据集进行染色标准化处理;

[0013] 步骤12:将所述乳腺癌组织病理类型图像数据集划分为训练集、验证集和测试集;将所述训练集和所述验证集中正常图像进行随机图像块采样策略,得到正常图像块;

[0014] 步骤13:使用所述正常图像块训练无监督的生成对抗网络。

[0015] 在上述任一方案中优选的是,所述生成对抗网络由两个对抗的模块,生成网络G和判别网络D组成。

[0016] 在上述任一方案中优选的是,所述生成对抗网络在训练时的目标函数为:

$$[0017] \min_G \max_D V(D, G) = \mathbb{E}_{x \sim P_{data}} [\log D(x)] + \mathbb{E}_{z \sim P_z(x)} [\log (1 - D(G(z)))]$$

[0018] 其中,x为真实样本, P_{data} 为真实数据分布, $P_z(x)$ 为先验分布., $\mathbb{E}_{x \sim P_{data}}$ 为x来自真实数据分布 P_{data} 的计算期望,z为随机变量, $\mathbb{E}_{z \sim P_z(x)}$ 为z来自先验分布 $P_z(x)$ 的计算期望,D(x)为真实样本x在判别网络D上的表现,G(z)为生成样本。

[0019] 在上述任一方案中优选的是,所述步骤2包括使用所述训练集和所述验证集中的所述良性图像进行随机图像块采样策略,使用所述生成对抗网络对所述良性图像块进行筛选,大于第一阈值的图像块为错误标签的图像块,小于第一阈值的图像块为正确标签的图像块,筛选出良性肿瘤中存在的正常区域。

[0020] 在上述任一方案中优选的是,:所述步骤2还包括所述生成对抗网络自动学习良性图像块样本数据分布,生成带有良性属性的生成对抗网络。

[0021] 在上述任一方案中优选的是,所述步骤2还包括使用所述训练集和所述验证集中的所述恶性图像进行随机图像块采样策略,使用所述生成对抗网络对恶性图像块进行初次筛选,大于第一阈值的图像块为错误标签的图像块,小于第一阈值的图像块为正确标签的图像块,筛选出恶性肿瘤中存在的正常区域。

[0022] 在上述任一方案中优选的是,所述步骤3包括使用所述带有良性属性的生成对抗

网络对的恶性图像块进行二次筛选,大于第二阈值的图像块为错误标签的图像块,小于第二阈值的图像块为正确标签的图像块,筛选出恶性肿瘤中存在的良性区域。

[0023] 在上述任一方案中优选的是,所述步骤3还包括对于给定测试图像块 x' ,在潜在空间中找到一个点 z' ,该点与图像 $G(z)$ 对应,该图像在视觉上与异常图像 x' 最相似,并且位于良性图像分布 X 上。

[0024] 在上述任一方案中优选的是,所述点 z' 的获取方法为:

[0025] 步骤31:定义一个损失函数,代表潜在空间向量映射到图像的损失;

[0026] 步骤32:在潜在空间分布 Z 中随机采样 z'_1 并将其馈送到训练有素的生成器中以获得生成的图像 $G(z'_1)$,代入损失函数计算损失;

[0027] 步骤33:计算损失函数关于的梯度,利用梯度下降法不断更新迭代,迭代过程中通过 $\delta=1,2,\dots,\Gamma$ 来优化 z 在潜空间 Z 中的位置,直到最相似的图像 $G(z'_\Gamma)$,其中, δ 为迭代次数, Γ 为迭代总次数。

[0028] 在上述任一方案中优选的是,所述损失函数包括残差损失和鉴别损失,所述残差损失用于度量生成的图像 $G(z'_\Gamma)$ 和查询图像块 x' 之间的相似程度,公式为:

[0029] $L_{\text{Residual}}(z'_\Gamma) = \sum |x' - G(z'_\Gamma)|$;

[0030] 所述鉴别损失用于反映鉴别器对两张图像提取的特征之间的差异,公式为:

[0031] $L_{\text{Discriminator}}(z'_\Gamma) = \sum |f(x') - f(G(z'_\Gamma))|$;

[0032] 总损失定义为所述残差损失和所述鉴别损失的加权和,公式为:

[0033] $L(z'_\Gamma) = (1-\lambda)L_{\text{Residual}}(z'_\Gamma) + \lambda \cdot L_{\text{Discriminator}}(z'_\Gamma)$

[0034] 其中, λ 为损失函数的权重, z'_Γ 为随机采样, f 为鉴别器中间某一层的输出。

[0035] 在上述任一方案中优选的是,所述步骤4包括以下子步骤:

[0036] 步骤41:输入经过所述生成对抗网络筛选的最具有判别特征的乳腺癌组织病理图像块;

[0037] 步骤42:使用循环密集连接块模块循环提取特征;

[0038] 步骤43:使用Softmax分类器进行分类。

[0039] 在上述任一方案中优选的是,所述循环密集块模块包含密集连接网络中的密集块和过渡层,每一个基本单元包含两个卷积核大小不同的批量归一化-ReLU激活-卷积操作,过渡层包含一个批量归一化-ReLU激活-卷积操作和池化操作,过渡层的输出将作为下一个循环密集连接块的输入。

[0040] 在上述任一方案中优选的是,每层特征图都要经过批量归一化、ReLU激活和卷积操作,所述循环密集连接卷积网络的第1层输出为:

[0041] $x_1 = H_1([x_0, x_1, \dots, x_{l-1}])$

[0042] 其中, l 表示层的维度, x_1 为 l 层网络的输出, H_1 为非线性变换, x_{l-1} 为 $l-1$ 层中产生的特征映射, $[x_0, x_1, \dots, x_{l-1}]$ 为 $0, 1, \dots, l-1$ 层中产生的特征映射的串联。

[0043] 在上述任一方案中优选的是,所述Softmax的损失函数的优化公式为:

[0044] $loss = \frac{1}{N} \sum_i -\log \left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}} \right)$

[0045] 其中, N 为样本总数, y_i 为第 i 个输入图像块 x_i 的标签, f 为类别分数, f_{y_i} 为为标签

为 y_i 的类别分数, j 为类别个数, f_j 为第 j 个类别的类别分。

[0046] 本发明提出了一种基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法,在实际训练过程中节省存储空间,消除特征之间的冗余,进一步提高乳腺癌组织病理图像的分类准确率和效率,解决高分辨率乳腺癌组织病理图像分类中显存占用过高的问题,有效缓解GPU显存不足的情况。

附图说明

[0047] 图1为按照本发明的基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法的一优选实施例的流程图。

[0048] 图2为按照本发明的基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法的另一优选实施例的总体技术路线图。

[0049] 图3为按照本发明的基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法的染色标准化的另一实施例的前后对照图。

[0050] 图4为按照本发明的基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法的循环密集连接的卷积神经网络的另一实施例的结构图。

[0051] 图5为按照本发明的基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法的另一优选实施例的具体实例工作流程图。

具体实施方式

[0052] 下面结合附图和具体的实施例对本发明做进一步的阐述。

[0053] 实施例一

[0054] 如图1所示,执行步骤100,获取乳腺癌组织病理类型图像数据集。执行步骤110,对乳腺癌组织病理类型图像进行预处理。本步骤包括以下子步骤:

[0055] 步骤11:对所述乳腺癌组织病理类型图像数据集进行染色标准化处理;

[0056] 步骤12:将所述乳腺癌组织病理类型图像数据集划分为训练集、验证集和测试集;将所述训练集和所述验证集中正常图像进行随机图像块采样策略,得到正常图像块;

[0057] 步骤13:使用所述正常图像块训练无监督的生成对抗网络,所述生成对抗网络由两个对抗的模块,生成网络 G 和判别网络 D 组成,所述生成对抗网络在训练时的目标函数为:

$$[0058] \quad \min_G \max_D V(D, G) = \mathbb{E}_{x \sim P_{data}} [\log D(x)] + \mathbb{E}_{z \sim P_z(x)} [\log (1 - D(G(z)))]$$

[0059] 其中, x 为真实样本, P_{data} 为真实数据分布, $P_z(x)$ 为先验分布., $\mathbb{E}_{x \sim P_{data}}$ 为 x 来自真实数据分布 P_{data} 的计算期望, z 为随机变量, $\mathbb{E}_{z \sim P_z(x)}$ 为 z 来自先验分布 $P_z(x)$ 的计算期望, $D(x)$ 为真实样本 x 在判别网络 D 上的表现, $G(z)$ 为生成样本。

[0060] 执行步骤120,生成对抗网络筛选良性图像块和恶性图像块中的正常区域。使用所述训练集和所述验证集中的所述良性图像进行随机图像块采样策略,使用所述生成对抗网络对所述良性图像块进行筛选,大于第一阈值的图像块为错误标签的图像块,小于第一阈值的图像块为正确标签的图像块,筛选出良性肿瘤中存在的正常区域。生成对抗网络自动学习良性图像块样本数据分布,生成带有良性属性的生成对抗网络。使用所述训练集和所述验证集中的所述恶性图像进行随机图像块采样策略,使用所述生成对抗网络对恶性图像

块进行初次筛选,大于第一阈值的图像块为错误标签的图像块,小于第一阈值的图像块为正确标签的图像块,筛选出恶性肿瘤中存在的正常区域。

[0061] 执行步骤130,生成对抗网络筛选恶性图像块中的良性区域。使用所述带有良性属性的生成对抗网络对的恶性图像块进行二次筛选,大于第二阈值的图像块为错误标签的图像块,小于第二阈值的图像块为正确标签的图像块,筛选出恶性肿瘤中存在的良性区域。对于给定测试图像块 x' ,在潜在空间中找到一个点 z' ,该点与图像 $G(z)$ 对应,该图像在视觉上与异常图像 x' 最相似,并且位于良性图像分布 X 上。点 z' 的获取方法为:

[0062] 步骤31:定义一个损失函数,代表潜在空间向量映射到图像的损失;

[0063] 步骤32:在潜在空间分布 Z 中随机采样 z'_r 并将其馈送到训练有素的生成器中以获得生成的图像 $G(z'_r)$,代入损失函数计算损失;

[0064] 步骤33:计算损失函数关于的梯度,利用梯度下降法不断更新迭代,迭代过程中通过 $\delta=1,2,\dots,\Gamma$ 来优化 z 在潜空间 Z 中的位置,直到最相似的图像 $G(z'_r)$,其中, δ 为迭代次数, Γ 为迭代总次数。损失函数包括残差损失和鉴别损失,所述残差损失用于度量生成的图像 $G(z'_r)$ 和查询图像块 x' 之间的相似程度,公式为:

[0065] $L_{\text{Residual}}(z'_r) = \sum |x' - G(z'_r)|$;

[0066] 所述鉴别损失用于反映鉴别器对两张图像提取的特征之间的差异,公式为:

[0067] $L_{\text{Discriminator}}(z'_r) = \sum |f(x') - f(G(z'_r))|$;

[0068] 总损失定义为所述残差损失和所述鉴别损失的加权和,公式为:

[0069] $L(z'_r) = (1-\lambda)L_{\text{Residual}}(z'_r) + \lambda \cdot L_{\text{Discriminator}}(z'_r)$

[0070] 其中, λ 为损失函数的权重, z'_r 为随机采样, f 为鉴别器中间某一层的输出。

[0071] 执行步骤140,使用基于循环密集连接的卷积神经网络对所述乳腺癌组织病理图像进行分类。本步骤包括以下子步骤:

[0072] 步骤41:输入经过所述生成对抗网络筛选的最具有判别特征的乳腺癌组织病理图像块。

[0073] 步骤42:使用循环密集连接块模块循环提取特征。环密集块模块包含密集连接网络中的密集块和过渡层,每一个基本单元包含两个卷积核大小不同的批量归一化-ReLU激活-卷积操作,过渡层包含一个批量归一化-ReLU激活-卷积操作和池化操作,过渡层的输出将作为下一个循环密集连接块的输入。每层特征图都要经过批量归一化、ReLU激活和卷积操作,所述循环密集连接卷积网络的第1层输出为:

[0074] $x_l = H_l([x_0, x_1, \dots, x_{l-1}])$

[0075] 其中, l 表示层的维度, x_l 为 l 层网络的输出, H_l 为非线性变换, x_{l-1} 为 $l-1$ 层中产生的特征映射, $[x_0, x_1, \dots, x_{l-1}]$ 为 $0, 1, \dots, l-1$ 层中产生的特征映射的串联

[0076] 步骤43:使用Softmax分类器进行分类。所述Softmax的损失函数的优化公式为:

[0077] $loss = \frac{1}{N} \sum_i -\log \left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}} \right)$

[0078] 其中, N 为样本总数, y_i 为第 i 个输入图像块 x_i 的标签, f 为类别分数, f_{y_i} 为为标签为 y_i 的类别分数, j 为类别个数, f_j 为第 j 个类别的类别分。

[0079] 实施例二

[0080] 针对高分辨率乳腺癌组织病理学图像,现有传统机器学习方法和用于直接分析全切片数字病理组织学图像的神经网络模型将导致非常复杂的体系结构问题,在过去的几年中,基于图像块的乳腺癌组织病理图像分类方法在乳腺癌组织病理学数据集中取得了可喜的成果。然而,采用基于图像块的乳腺癌组织病理图像分类方法非常具有挑战性,由于恶性全切片数字病理组织学图像中存在良性区域和正常区域,良性全切片数字病理组织图像中存在正常区域,因此只有部分提取的图像块被正确标记。为了解决这个错误标签的图像块问题并进一步提高分类的准确性。我们提出了一种基于生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法,采用改进的无监督的生成对抗网络分别学习正常病理和良性肿瘤病理图像的数据分布,从而可以筛选恶性肿瘤病理图像中存在良性肿瘤的区域和正常区域,良性肿瘤病理图像中存在的正常区域。同时采用基于循环密集连接的卷积神经网络作为分类网络,一方面密集连接可实现多层特征复用和融合,另一方面通过循环结构可以在实际训练过程中节省存储空间,有利于进一步提高分类的准确性和效率。为最大程度辅助医生更精确、更快速地诊断病情提供可能。

[0081] 为了实现上述目的,本发明采用如下方案:

[0082] 一种基于无监督异常检测生成对抗网络筛选图像块的乳腺癌组织病理类型分类方法,包括以下步骤:

[0083] 1) 获取乳腺癌组织病理类型图像数据集,对其进行染色标准化处理,并将其划分为训练集、验证集和测试集三部分,训练集和验证集用于本发明方法的模型训练,测试集用于本发明方法的模型测试;

[0084] 2) 将步骤1)中训练集和验证集中正常图像进行随机图像块采样策略,并进行数据增强处理,包括一系列仿射变换操作;

[0085] 3) 使用步骤2)得到的正常图像块训练无监督的生成对抗网络,通过生成对抗网络学习可以得到正常图像块的数据分布,而这种学习到正常图像块样本数据分布的生成对抗网络,在用于学习良性和恶性样本的数据分布时会有较大的异常分数。

[0086] 4) 使用步骤1)中训练集和验证集中良性图像进行随机图像块采样策略,使用步骤3)中的生成对抗网络对良性图像块进行筛选。若大于某一阈值则系统判定为错误标签的图像块,若小于某一阈值则系统判定为正确标签的图像块,进而筛选良性肿瘤中存在的正常区域;

[0087] 5) 使用步骤1)中训练集和验证集中恶性图像进行随机图像块采样策略,使用步骤3)中的生成对抗网络对恶性图像块进行筛选。若大于某一阈值则系统判定为错误标签的图像块,若小于某一阈值则系统判定为正确标签的图像块,进而筛选恶性肿瘤中存在的正常区域;

[0088] 6) 使用步骤4)中筛选出的良性图像块进行步骤2)的数据增强处理,包括一系列的仿射变换操作;

[0089] 7) 使用步骤6)得到的良性图像块训练无监督的生成对抗网络,通过生成对抗网络学习可以得到良性图像块的数据分布,而这种学习到良性图像块样本数据分布的生成对抗网络,在用于学习恶性样本的数据分布时会有较大的异常分数;

[0090] 8) 使用步骤7)中的生成对抗网络对步骤5)得到的恶性图像块进一步筛选。若大于某一阈值则系统判定为错误标签的图像块,若小于某一阈值则系统判定为正确标签的图

像块;进而筛选出恶性肿瘤中存在的良性区域,为近一步提高分类网络的训练性能提供了可能;

[0091] 9) 使用步骤8)中筛选的恶性图像块进行步骤2)的数据增强处理,包括一系列的仿射变换操作,以生成最终用于训练分类网络的训练集和验证集;

[0092] 10) 使用步骤4)得到的良性图像块和步骤9)得到的恶性图像块完成基于循环密集连接的卷积神经网络的训练;

[0093] 11) 对步骤1)的测试集进行步骤2)的随机图像块采样策略和数据增强操作,使用步骤10)的基于循环密集连接的卷积神经网络进行测试,通过多数投票策略组合步骤1)的测试集的分类结果,计算在乳腺癌组织病理数据集在图像级别和患者级别的准确率。

[0094] 本发明的有益效果:

[0095] (1) 本发明重在针对采用基于图像块分类方法存在错误标签的问题,通过生成对抗网络学习正常图像和良性图像样本的数据分布,筛选恶性肿瘤病理图像中存在良性肿瘤的区域和正常区域,良性肿瘤病理图像中存在的正常区域。有助于提高后续分类网络的性能;

[0096] (2) 本发明使用基于循环密集连接的卷积神经网络作为分类网络,密集连接可保证网络中层与层之间最大程度的信息传输,实现特征的复用,循环结构可以在实际训练过程中节省存储空间,消除特征之间的冗余,进一步提高乳腺癌组织病理图像的分类准确率和效率。

[0097] (3) 本发明通过采用图像块分类的方法解决高分辨率乳腺癌组织病理图像分类中显存占用过高的问题,有效缓解GPU显存不足的情况。

[0098] 实施例三

[0099] 在对现有技术的研究中发现,存在着以下缺点:无法筛选良性图像和恶性图像中的正常区域,影响训练的效果;并且密集连接网络会存在参数过多不易训练的问题。

[0100] 针对上述缺点,本发明在筛选图像块策略增加了对良性和恶性图像块中正常区域的筛选,以保证得到更精确的数据,采用网络结构更加紧凑的基于循环密集连接的卷积神经网络方面进行了修改,能够解决上述问题,从而实现更精确更高效的乳腺癌组织病理图像的分类结果。

[0101] 本发明主要由四个模块构成:

[0102] (1) 乳腺癌组织病理图像预处理

[0103] (2) 基于生成对抗网络筛选良性图像块和恶性图像块中的正常区域,解决存在错误标签的图像块的问题

[0104] (3) 基于生成对抗网络筛选恶性图像块中的良性区域,解决存在错误标签的图像块的问题

[0105] (4) 基于循环密集连接的卷积神经网络作为分类网络,对模块(2)和(3)中的得到的具有判别性的乳腺癌组织病理图像块进行分类。

[0106] 本发明的总体技术路线如图2所示:

[0107] 其中以上模块的详细介绍如下:

[0108] (1) 乳腺癌组织病理图像预处理

[0109] 染色标准化是H&E染色图像预处理中至关重要的一步。通过染色标准化可以减少

数字扫描仪的颜色响应,染色供应商的材料和制造技术,以及不同实验室不同的染色方案造成的颜色差异都会给图像自动分类带来影响。本发明采用了一种新型保留结构信息的染色标准化方案。首先,将染色分离问题转化为非负矩阵分解,并对其加入稀疏约束,即稀疏非负矩阵分解。这种方法的一个优点是,颜色基础是在无人监督的情况下确定的,不需要手动标记不同区域的纯污渍。保留结构信息的染色标准化方法的工作原理是将源图像的基准颜色替换为病理学家青睐的目标图像颜色,同时可靠地保持源图像的结构信息完整,并保持其原有的染色浓度。如图3所示,A为目标图像,B为源图像,C为经过处理的图像。

[0110] 其次,深度学习模型,尤其是卷积神经网络,需要大量的训练数据。由于乳腺癌组织病理图像的特殊性,获得大量与疾病进展和治疗监测相关的乳腺癌组织病理图像数据集十分具有挑战性。因此,我们提出一种图像块采样策略和数据增强方法生成足够数量的数据样本,为后续的网络的训练提供了保障。同时,本专利提出了一种基于提取图像块的数据增强算法来训练密集连接卷积神经网络。值得一提的是,我们避免使用尺寸为 32×32 或 64×64 的较小图像块,这是因为在乳腺癌组织病理图像数据集中,标签被指定为整个输入的乳腺癌组织病理学图像, 32×32 或 64×64 尺寸的图像块并不能保证将携带足够的诊断信息。因此,我们将大小为 700×460 的乳腺癌组织病理图像分割成大小为 224×224 的图像块。与较小尺寸的图像块相比, 224×224 尺寸的图像块可以携带更多的具有辨别特征的局部信息。

[0111] 最后,对于乳腺组织病理学图像,我们可以使用翻转和旋转等一系列仿射变换来增加数据量,同时提高计算机辅助诊断系统的预测精度,而不改变图像的组织形态和细胞结构。

[0112] (2) 基于生成对抗网络筛选良性图像块和恶性图像块中的正常区域

[0113] 生成对抗网络由两个对抗的模块,生成网络G和判别网络D组成。从潜在空间Z采样的服从某一分布(例如均匀分布)的随机向量z通过生成网络G后,可以得到一个生成样本G(z)。生成样本G(z)和真实样本x一起作为判别网络的输入。判别网络估计G(z)来自于x的概率。在训练过程中,生成网络通过判别网络的判别结果进行优化,以提高生成能力,使G(z)不断接近x以“欺骗”判别网络;判别网络则通过误判概率,对自身进行优化,提高对生成样本的辨别能力。通过生成网络和判别网络的相互博弈,生成网络最终能够输出接近于真实样本分布的生成样本,且判别网络最终无法对生成样本进行辨别。

[0114] 对于生成网络,生成样本G(z)应尽可能接近于真实样本x;对于判别网络,它可以给出输入样本来自于真实样本的可信度,即:当判别网络输入x时, $D(x)$ 接近于1,当输入G(z)时, $D(G(z))$ 接近于0。因此生成对抗网络在训练时的目标函数可以表示为:

$$[0115] \quad \min_G \max_D V(D, G) = \mathbb{E}_{x \sim P_{data}} [\log D(x)] + \mathbb{E}_{z \sim P_z(x)} [\log (1 - D(G(z)))]$$

[0116] 当对抗训练完成时,生成器已经学习了从潜在空间表示z到正常图像块x的映射 $G(z) = z \rightarrow x$ 。但是GAN不会自动无代价地生成由测试图像x到潜在空间z的逆映射 $\mu(x) = x \rightarrow z$,需要迭代地寻z。潜在空间具有平滑的过渡,换句话说,从潜在空间中距离相近的两个点生成的图像非常相似。给定测试图像块x,我们旨在在潜在空间中找到一个点z,该点与图像G(z)对应,该图像在视觉上与异常图像x最相似,并且位于正常图像分布X上。为了找到最佳的z,受特征匹配技术的启发,采用如下的方式:

[0117] (a): 定义一个损失函数, 代表潜在空间向量映射到图像的损失;

[0118] (b): 在潜在空间分布 Z 中随机采样 z_1 并将其馈送到训练有素的生成器中以获得生成的图像 $G(z_1)$ 。代入损失函数计算损失;

[0119] (c): 计算损失函数关于 z_1 的梯度, 利用梯度下降法不断更新迭代, 迭代过程中通过 $\gamma = 1, 2, \dots, \Gamma$ 来优化 z 在潜空间 Z 中的位置。直到最相似的图像 $G(z_\Gamma)$ 。

[0120] 上述(a)定义了一个将异常图像块映射到潜在空间的损失函数, 该函数包括两个分量: 残差损失和鉴别损失。

[0121] 残差损失用于度量生成的图像 $G(z_\Gamma)$ 和查询图像块 x 之间的相似程度, 残差损失为:

$$[0122] \quad L_{\text{Residual}}(z_\Gamma) = \sum |x - G(z_\Gamma)|$$

[0123] 受特征匹配技术的启发, 将鉴别器看成是一个特征提取器, 选择鉴别器中间某一层的输出作为函数 f , 鉴别损失反映了鉴别器对两张图像提取的特征之间的差异:

$$[0124] \quad L_{\text{Discriminator}}(z_\Gamma) = \sum |f(x) - f(G(z_\Gamma))|$$

[0125] 为了映射到潜在空间, 我们将总损失定义为两个分量的加权和:

$$[0126] \quad L(z_\Gamma) = (1-\lambda) L_{\text{Residual}}(z_\Gamma) + \lambda \cdot L_{\text{Discriminator}}(z_\Gamma)$$

[0127] 我们将上述总损失定义为异常分数, 该分数表示测试图像 x 与正常图像的拟合度。通过生成对抗网络学习可以得到正常图像块的数据分布, 而这种学习到正常图像块样本数据分布的生成对抗网络, 在用于学习良性和恶性样本的数据分布时会有较大的异常分数。从而筛选出良性和恶性图像块中的被错误标记的图像块。并使用翻转和旋转等一系列仿射变换来增加筛选得到的良性图像块的数据量, 为下一模块提供足够的训练数据。

[0128] (3) 基于生成对抗网络筛选恶性图像块中的良性区域

[0129] 使用筛选得到的良性图像块进行生成对抗网络的训练。当对抗训练完成时, 生成器已经学习了从潜在空间表示 z' 到良性图像块的映射。给定测试图像块 x' , 我们旨在潜在空间中找到一个点 z' , 该点与图像 $G(z)$ 对应, 该图像在视觉上与异常图像 x' 最相似, 并且位于良性图像分布 X 上。为了找到最佳的 z' , 受特征匹配技术的启发, 采用如下的方式:

[0130] (a): 定义一个损失函数, 代表潜在空间向量映射到图像的损失;

[0131] (b): 在潜在空间分布 Z 中随机采样 z'_1 并将其馈送到训练有素的生成器中以获得生成的图像 $G(z'_1)$ 。代入损失函数计算损失;

[0132] (c): 计算损失函数关于的梯度, 利用梯度下降法不断更新迭代, 迭代过程中通过 $\delta = 1, 2, \dots, \Gamma$ 来优化 z 在潜空间 Z 中的位置, 直到最相似的图像 $G(z'_\Gamma)$ 。

[0133] 上述(a)定义了一个将异常图像块映射到潜在空间的损失函数, 该函数包括两个分量: 残差损失和鉴别损失。

[0134] 残差损失用于度量生成的图像 $G(z'_\Gamma)$ 和查询图像块 x' 之间的相似程度, 残差损失为:

$$[0135] \quad L_{\text{Residual}}(z'_\Gamma) = \sum |x' - G(z'_\Gamma)|$$

[0136] 受特征匹配技术的启发, 将鉴别器看成是一个特征提取器, 选择鉴别器中间某一层的输出作为函数 f' , 鉴别损失反映了鉴别器对两张图像提取的特征之间的差异:

$$[0137] \quad L_{\text{Discriminator}}(z'_\Gamma) = \sum |f(x') - f(G(z'_\Gamma))|$$

[0138] 为了映射到潜在空间, 我们将总损失定义为两个分量的加权和:

[0139] $L(z'_{\Gamma}) = (1-\lambda)L_{\text{Residual}}(z'_{\Gamma}) + \lambda \cdot L_{\text{Discriminator}}(z'_{\Gamma})$

[0140] 我们将上述总损失定义为异常分数,该分数表示测试图像 x' 与良性图像的拟合度。通过生成对抗网络学习可以得到良性图像块的数据分布,而这种学习到良性图像块样本数据分布的生成对抗网络,在用于学习恶性样本的数据分布时会有较大的异常分数。从而筛选出恶性图像块中的被错误标记为恶性的图像块。并使用翻转和旋转等一系列仿射变换来增加筛选得到的恶性图像块的数据量,为后续分类网络提供足够的训练数据。

[0141] (3) 基于循环密集连接的卷积神经网络对乳腺癌组织病理图像分类块

[0142] 密集连接卷积网络是结合了模型ResNet与Highway的优点,旨在解决深层网络中的梯度消失问题。密集连接卷积网络的思想是保证网络中层与层之间最大程度信息的传输,从而直接将所有层连接起来。本发明设计一个更紧凑的循环密集连接网络以进一步减少网络参数,同时保持与密集连接网络近似的性能,每一个循环密集连接块包含密集连接网络中的密集块和过渡层,同时将过渡层的输出作为下一个循环密集连接块的输入,通过此设计可以使得网络变得更深同时减少了参数的数量,一方面密集连接可以实现特征的复用,另一方面通过循环结构可以在实际训练过程中节省存储空间,有利于提高训练的效率。如图4所示,基于循环密集连接的卷积神经网络对乳腺癌组织病理图像分类算法由三部分组成:输入经过生成对抗网络筛选的最具有判别特征的乳腺癌组织病理图像块、使用循环密集连接块提取特征、Softmax分类器。经过上述三部分实现乳腺癌组织病理学图像的分类。

[0143] 循环密集块模块是循环密集连接卷积网络的核心部分,主要特点是其中的各层网络不仅与下一层连接,而且与后面的每一层都直接连接,每一层的输入来自前面所有层的输出。从而能够促进信息的传递,减轻了梯度消失现象,并且网络能够更好地收敛。密集连接块中每一层的输出将进行通道的拼接做为该块的输出,因此由于通道数量的不同,无法直接将其输出重新输入到循环密集块中,需要通过一个过渡层进行通道数的压缩,所以循环密集块模块包含密集连接网络中的密集块和过渡层,与密集连接网络中的密集块相同,每一个基本单元包含两个卷积核大小不同的批量归一化-ReLU激活-卷积操作,过渡层包含一个批量归一化-ReLU激活-卷积操作和池化操作。过渡层的输出将作为下一个循环密集连接块的输入,从而可以达到节省存储空间的效果。

[0144] 假设一张图像块 x_0 在卷积网络中传播,网络共有 L 层,每一层都有一个非线性变换 H ,即每层特征图都要经过批量归一化、ReLU激活和卷积操作。其中 l 表示层的维度, x_l 为 l 层网络的输出,则循环密集连接卷积网络的第 l 层输出为:

[0145] $x_l = H_l([x_0, x_1, \dots, x_{l-1}])$

[0146] 其中, x_{l-1} 为 $l-1$ 层中产生的特征映射, $[x_0, x_1, \dots, x_{l-1}]$ 为 $0, 1, \dots, l-1$ 层中产生的特征映射的串联。

[0147] 与传统的卷积神经网络相比,密集连接加强了乳腺癌组织病理图像类型特征的传递,提高了特征在各个层间的流动,大大提高了特征利用率,使得网络能够自动学习乳腺癌组织病理图像中鲁棒性强且具有区分度的特征,提高分类的精确度;循环结构使得网络更加紧凑,通过循环结构可以在实际训练过程中节省存储空间,有利于提高训练的效率。。

[0148] 定义第 i 个输入图像块 x_i ,标签为 y_i ,通过Softmax损失函数不断优化:

$$[0149] \quad loss = \frac{1}{N} \sum_i -\log \left(\frac{e^{f_{y_i}}}{\sum_j e^{f_{j_i}}} \right)$$

[0150] 实施例四

[0151] 本发明采用BreKHis乳腺癌组织病理图像数据集作为实例数据集,BreKHis数据集是以患者为单位发布的乳腺组织病理图像数据集,它由巴西巴拉那的P&D实验室病理解剖学和细胞病理学实验室合作建立的。该公共数据集是由82名患者的7909张苏木精和伊红染色的乳腺肿瘤组织显微图像组成的,其中每组图像都由不同放大倍数(40X、100X、200X和400X)的图像组成,图片格式是PNG格式,包含3个RGB通道,单通道是8比特位宽,每张图片的分辨率为700*460。数据集分为良性肿瘤和恶性肿瘤两大类,其中良性有2480张,恶性有5429张。

[0152] 如图5所示,本发明解决其技术问题的主要步骤如下:

[0153] 1、采集BreKHis数据集,对其进行稀疏非负矩阵分解的染色标准化处理,按照7:3划分训练数据,测试集,进一步按照8:2将训练数据划分为训练集和验证集。训练集和验证集用于本发明方法的模型训练,测试集用于本发明方法的模型测试;

[0154] 2、对乳腺癌组织病理正常图像进行随机图像块采样,并进行旋转45度、90度、135度、180度、225度、270度、315度操作、水平翻转和竖直翻转操作,使用大小为224×224乳腺癌组织病理正常图像训练生成对抗网络,通过生成对抗网络学习可以得到正常图像块的样本的数据分布;

[0155] 3、以随机采样64张为基数,按照良性图像和恶性图像之间不平衡比率进行随机图像块采样,对训练集和验证集中良性恶性图像进行随机图像块采样策略,并通过正常图像块训练得到的生成对抗网络对恶性图像块进行筛选。若大于某一阈值则系统判定为错误标签的图像块,若小于某一阈值则系统判定为正确标签的图像块;

[0156] 4、对筛选正常区域得到的良性图像进行旋转45度、90度、135度、180度、225度、270度、315度操作、水平翻转和竖直翻转操作,在40X、100X、200X和400X分别得到63750张、66332张、64169张和58212张224×224尺寸的良性图像块;

[0157] 5、使用40X、100X、200X和400X分别得到63750张、66332张、64169张和58212张224×224尺寸的良性图像块训练无监督的生成对抗网络,通过生成对抗网络学习可以得到良性图像块的样本的数据分布;

[0158] 6、通过良性图像块训练得到的生成对抗网络对恶性图像块进行进一步地筛选。若大于某一阈值则系统判定为错误标签的图像块,若小于某一阈值则系统判定为正确标签的图像块;

[0159] 7、对具有正确标签的恶性图像块进行旋转45度、90度、135度、180度、225度、270度、315度操作、水平翻转和竖直翻转操作,在40X、100X、200X和400X分别得到63020张、66102张、63940张和57904张224×224尺寸的恶性图像块,从而生成最终用于训练分类网络的训练集和验证集;

[0160] 8、使用在40X、100X、200X和400X分别得到63750张、66332张、64169张和58212张224×224尺寸的良性图像块和在40X、100X、200X和400X分别得到63020张、66102张、63940张和57904张224×224尺寸的恶性图像块训练基于循环密集连接的卷积网络;

[0161] 本发明使用训练批次大小为64的Adam优化器来训练分类模型。学习率设置为0.001。算法基于Python语言实现,使用Pytorch深度学习框架,在具有三块NVIDIA GeForce GTX 1080iT GPU上运行,显卡内存为24GB。

[0162] 9、对测试集进行随机图像块采用策略和数据增强操作,使用基于循环密集连接的卷积网络进行测试,通过多数投票策略组合步骤1)的测试集的分类结果,计算在乳腺癌组织病理数据集在图像级别和患者级别的准确率。

[0163] 采用多数投票算法将不同图像块的分类结果融合,令 $\gamma \in [0, 1, 2 \dots n]$ 对应于n个不同的组织病理类型, $F_i^{(k)}$ 为每个类型得分函数,其中 $k \in [0, 1, 2 \dots n]$,如下面公式所示:

$$[0164] \quad F_i^{(k)} = \begin{cases} 1, & \text{if } \gamma = k \\ 0, & \text{if } \gamma \neq k \end{cases}$$

[0165] 令 C^k 为每个类别的总投票数,则可以表示为:

$$[0166] \quad C^k = \sum_{i=1}^n F_i^{(k)}$$

[0167] 假设T是测试图像I的预测标签,将图像I中最多数图像块所属的组织病理类型的标签赋予图像I:

$$[0168] \quad T = \operatorname{argmax}(C^k)$$

[0169] 在对比不同算法模型时,需要使用统一的衡量算法能力的评价标准,即性能度量方法,不同的任务需求有各自适用的性能度量方法。对于医学图像分类任务,采用公认的两个方面评价模型的分类性能:图像级别准确率和患者级别准确率。

[0170] 1) 图像级别准确率

[0171] 令 N_{all} 代表验证集和测试集中病理图像的数量, N_r 代表其中被正确分类的图像数量,则图像级别的准确率可表示为:

$$[0172] \quad \text{Image Recognition Rate} = \frac{N_r}{N_{all}}.$$

[0173] 2) 患者级别准确率

[0174] 对于乳腺癌组织病理学图像的分类,可以从两个方面评价模型的分类性能:患者级别和图像级别。从患者的层面计算准确率,设 N_{np} 为每位患者病理图像的数量, N_{rp} 为每位患者被正确分类的图像数量, N_p 为患者总数量,则有:

$$[0175] \quad \text{Patient Score} = \frac{N_{rp}}{N_{np}}$$

[0176] 则患者级别的准确率可表示为:

$$[0177] \quad \text{Patient Recognition Rate} = \frac{\sum \text{Patient Score}}{N_p}$$

[0178] 通过采用多数投票算法将图像块的分类结果融合,得出高分辨率乳腺癌组织病理学图像二分类图像级别和患者级别的准确率,为了验证本发明的有效性,与基于深度学习的医学图像分类算法中常见的分类网络AlexNet、VGG16卷积神经网络和仅使用基于循环密集连接的卷积网络(未使用生成对抗网络进行图像块筛选)的方法进行了对比,实验结果如

表1所示。

[0179] 表1 BreakHis数据集图像级别和患者级别的准确率

评价指标	方法	不同放大倍数下的准确率			
		40X	100X	200X	400X
[0180] 图像级别准确率	AlexNet	85.6±4.8	83.5±3.9	83.1±1.9	80.8±3.0
	VGG-16	77.81±3.6	72.02±2.9	81.28±4.1	83.24±2.5
	基于循环密集连接的卷积网络	96.25±2.73	94.43±1.52	86.57±3.53	86.57±3.53
	发明方法	98.75±0.89	97.59±0.43	91.77±0.68	87.34±0.82
患者级别准确率	AlexNet	90.0±6.7	88.4±4.8	84.6±4.2	86.1±6.2
	VGG-16	83.8±4.1	82.1±4.9	84.2±4.1	82.0±5.1
	基于循环密集连接的卷积网络	93.74±2.29	92.09±1.59	84.65±2.97	82.98±2.54
	发明方法	95.22±1.77	94.71±1.32	87.39±1.63	85.11±2.41

[0181] 综上所述,所有方法均在40X放大倍数下获得最佳的性能,这是由于随着放大倍数的增大,图像中的细胞数目变少并且细胞面积变大,导致携带的信息减少。本发明提出的基于生成对抗网络筛选错误标签图像块的乳腺癌组织病理类型分类方法可以有效提升分类的准确率。首先,通过生成对抗网络学习良性图像样本的数据分布,可以解决基于图像块分类方法存在错误标签的问题,筛选出最具有判别性的乳腺癌组织病理图像块,有助于提高后续分类网络的性能;其次,与使用AlexNet、VGG16卷积神经网络的分类算法和仅使用基于循环密集连接的卷积神经网络算法相比,本发明提出基于生成对抗网络筛选错误标签图像块的乳腺癌组织病理类型分类方法既解决了错误标签图像块的问题,又保证网络中层与层之间最大程度的信息传输,实现特征的复用,进一步提高了分类的准确率。另外循环结构可以在实际训练过程中节省存储空间,有利于进一步提高分类的准确性和效率。从而为本发明方法应用于临床诊断奠定了基础。

[0182] 为了更好地理解本发明,以上结合本发明的具体实施例做了详细描述,但并非是对本发明的限制。凡是依据本发明的技术实质对以上实施例所做的任何简单修改,均仍属于本发明技术方案的范围。本说明书中每个实施例重点说明的都是与其它实施例的不同之处,各个实施例之间相同或相似的部分相互参见即可。对于系统实施例而言,由于其与方法实施例基本对应,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

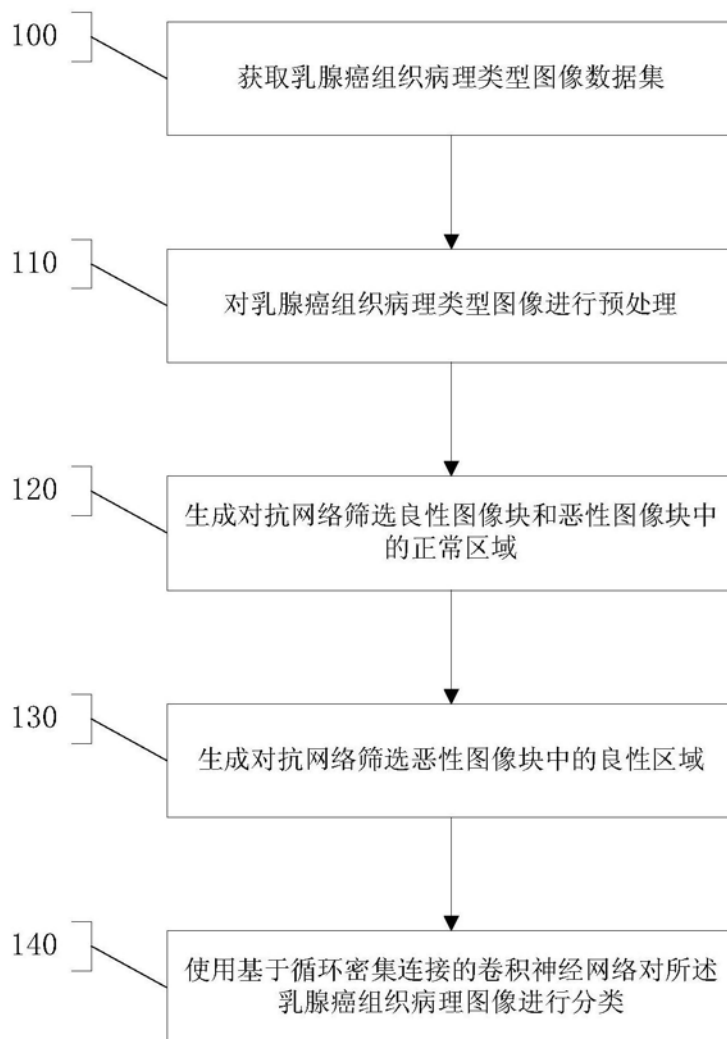


图1

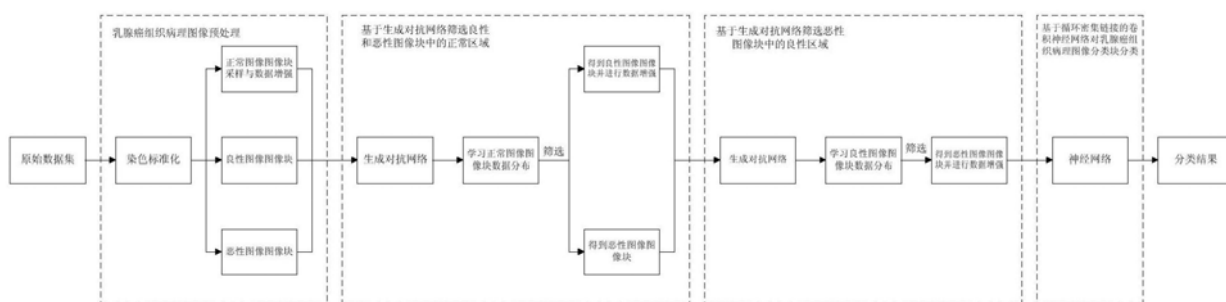


图2

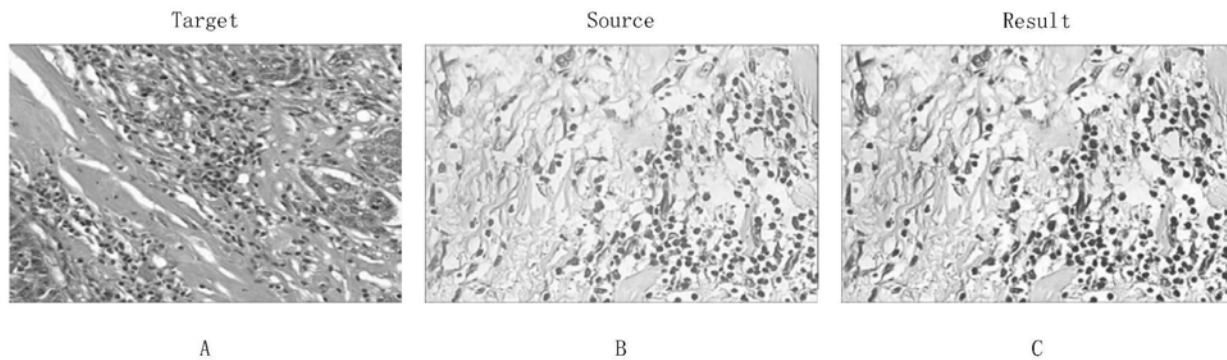


图3

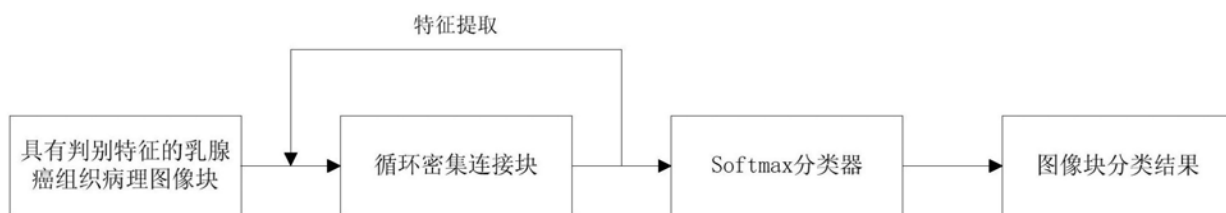


图4



图5