

DOI:10.13232/j.cnki.jnju.2021.01.003

## 基于深度主动学习的甲状腺癌病理图像分类方法

张 萌<sup>1</sup>, 韩 冰<sup>1\*</sup>, 王 哲<sup>2</sup>, 尤富生<sup>3</sup>, 李浩然<sup>1</sup>

(1. 西安电子科技大学工程学院, 西安, 710071; 2. 空军军医大学基础医学院病理学教研室, 西安, 710071;  
3. 空军军医大学生物医学工程系, 西安, 710071)

**摘 要:** 甲状腺癌是内分泌系统最常见的恶性肿瘤, 甲状腺病理图像对于甲状腺癌的分级、预后和后续治疗有重要的指导作用。近年来, 深度学习在病理图像分类分级中表现出色, 然而, 为了获得良好的分类性能, 这些方法往往需要大量的标注数据。众所周知, 医学图像的手动注释非常繁琐、耗时, 并且需要领域知识的指导。为了降低标注成本, 提出一种将卷积神经网络(Convolutional Neural Networks, CNN)和主动学习相结合的分类方法, 无须标记所有数据, 仅选择少量样本进行标注。此方法利用 CNN 提取病理图像的特征, 进而使用该特征计算未标注样本的不确定性和相似性, 选择“有价值”的样本; 然后由病理学家对选定的样本进行标注, 并不断微调网络以增强模型的分类性能。在甲状腺病理图像上的实验结果表明, 该方法能够在不牺牲最终分类准确率的情况下降低标记成本。

**关键词:** 甲状腺乳头状癌病理图像, 深度学习, 主动学习, 图像分类, 哈希码

中图分类号: TP183

文献标识码: A

## Papillary thyroid carcinoma pathological image classification based on deep active learning

Zhang Meng<sup>1</sup>, Han Bing<sup>1\*</sup>, Wang Zhe<sup>2</sup>, You Fusheng<sup>3</sup>, Li Haoran<sup>1</sup>

(1. School of Electronic Engineering, Xidian University, Xi'an, 710071, China; 2. Teaching and Research Section of Pathology, School of Basic Medicine, Air Force Medical University, Xi'an, 710071, China; 3. School of Biomedical Engineering, Air Force Medical University, Xi'an, 710071, China)

**Abstract:** Thyroid cancer is the most common form of endocrine malignancy and the informative pathological images are critical for thyroid cancer risk stratification, prognosis and treatment guidance. Recent advances in deep learning have achieved promising results on pathology image classification benchmarks. However, to achieve an acceptable classification performance, most of the existing methods require a large number of labeled images. The manual annotation for the medical image is known to be tedious, time-consuming, and requires guidance from domain knowledge. To reduce the labeling cost, this paper proposes a classification method which integrates the convolutional neural network (CNN) and active learning to actively select a few samples for annotation. Specifically, we utilize CNN to extract feature of the pathological image. And then the deep feature is employed to estimate the uncertainty and representativeness of pathological image for “valuable” sample selection. Finally, the selected samples are annotated by pathologists and continuously fine-tune CNN to enhance the classification performance. The experimental results on thyroid pathological images demonstrate that the proposed method can reduce the labeling cost without sacrificing the accuracy of the classification system.

**Key words:** papillary thyroid carcinoma pathological image, deep learning, active learning, image classification, Hash code

基金项目: 国家自然科学基金(61572384)

收稿日期: 2020-09-19

\* 通讯联系人, E-mail: bhan@xidian.edu.cn

甲状腺癌是内分泌系统最常见的恶性肿瘤,也是全球发病率增长最快的恶性肿瘤<sup>[1]</sup>. 在众多的甲状腺癌检查方法中<sup>[2]</sup>,病理穿刺活检是当前敏感性和特异性最高的方法<sup>[3]</sup>. 在甲状腺癌的四中主要病理类型中,甲状腺乳头状癌约占甲状腺癌的 85% 以上,虽然甲状腺乳头状癌预后非常好,十年的存活率高达 90% 左右,但此类癌症的淋巴转移率高达 40%~50%<sup>[4]</sup>,因此甲状腺乳头状癌的早期确诊并制订合适的治疗方案对防止病情恶化、挽救病人的生命具有重要意义.

卷积神经网络(Convolutional Neural Networks, CNN)在生物医学图像分析中被广泛应用,然而大部分基于 CNN 的图像分析方法是以大量带标注的数据集为基础的. 对生物医学图像进行标注不仅繁琐、耗时,而且需要领域内的专业知识指导,这些专业知识需要系统学习,难以迅速掌握,所以很难获取大量带标签的数据. 然而,目前已经有公开的其他类型疾病的病理图像数据集,我们可以充分利用现有带标签的其他数据集,辅助挑选有价值的样本进行标注,减少标注代价.

近年来,由于数字数据的爆炸性增长,存储和检索效率极高的哈希编码在各种计算机视觉任务中已被广泛应用. Venkateswara et al<sup>[5]</sup>使用深度神经网络学习样本的代表性哈希码来解决域适配问题,在无监督域自适应方面取得良好的效果. 本文以该深度哈希网络为基础,结合主动学习算法,提出一种基于深度主动学习的甲状腺乳头状癌病理图像分类方法,将深度学习与主动学习集成到同一框架中. 该方法通过 CNN 来提取图像特征,并利用不确定性信息与代表性信息对未标注数据集中“有价值的”样本进行注释. 通过在每一轮迭代中加入新注释的样本对模型进行微调以增强模型的分类性能. 在甲状腺癌病理图像数据集上的实验结果表明,和原始的 CNN 网络相比,本文提出的基于深度主动学习的甲状腺乳头状癌病理图像分类方法将标注样本数量降低了 70% 左右.

本文的主要贡献如下:

(1)提出一个新的深度主动学习分类框架对甲状腺癌病理切片进行分类,该方法将深度学习与主动学习集成到一个框架中;

(2)设计了一种样本不确定性评估方法评估未标记样本的不确定性;

(3)设计了一种样本间相似性评估方法,从同一患者样本间相似性以及不同患者的样本间相似性两个角度对样本相似性进行评估;

(4)在甲状腺癌数据库上的实验表明该方法能够有效降低标注代价,仅需标注少量样本便能取得良好的分类效果.

## 1 相关工作

目前,解决标注问题的方法可以分为两大类:无监督方法和有监督方法. 无监督学习能够根据未知类别的训练样本解决分类等问题,然而无监督学习的效果远远比不上监督学习,尤其对于医学图像,准确的分类是十分重要的,关乎后续治疗方案的制订. 主动学习(Active Learning, AL)可以很好地平衡这两个问题,在减少标注代价的同时能保证分类准确率,因此众多研究者对主动学习进行了深入研究. 在主动学习的早期研究中,主要依赖手工特征进行信息查询与样本选择. Abe and Mamitsuka<sup>[6]</sup>提出一种查询学习策略,结合委员会查询和装袋查询学习策略,在实际应用中获得了较好的效果. Ebert et al<sup>[7]</sup>分析不同的采样标准并提出一种基于强化学习的新型反馈驱动的主动学习框架,能够根据经验在学习过程中调整采样策略. Huang et al<sup>[8]</sup>提出一种基于主动学习的最小-最大观点,为结合样本的信息性和代表性提供了一种系统的方法. Tang and Huang<sup>[9]</sup>提出一种自定步长的主动学习方法:一方面考虑了信息量和代表性,选取的实例对模型的改进具有较高的潜在价值;另一方面,利用样本的易用性,使模型能够充分利用样本潜在价值.

近年来,由于深度学习的发展,主动学习与深度学习的研究<sup>[10-14]</sup>逐渐受到越来越多的重视. Wang and Shang<sup>[10]</sup>可能是第一次将主动学习与深度学习相结合,提出基于堆叠受限玻尔兹曼机和堆叠自动编码器的主动标记方法. Li<sup>[11]</sup>也将类似的方法应用于高光谱图像分类. Stark et al<sup>[12]</sup>利用主动学习提高基于 CNNs 的验证码识别性能,Al Rahhal et al<sup>[13]</sup>利用深度学习和主动学习进行心电图分类. Zhou et al<sup>[14]</sup>利用同一样本的不同裁剪图

像块的类别及概率的选择样本进行标注,并在三个不同的生物医学图像数据集上获得了良好的效果,然而由于病理图像不同的裁剪的图像往往属于不同的类别,因此该方法不适用于甲状腺癌病理图像分类. Yang et al<sup>[15]</sup>将全卷积神经网络与主动学习进行结合,通过对最有效的注释区域进行注释来显著减少注释工作量,在2015年MICCAI腺体分割数据集上获得了较好的分割结果.

虽然上述方法在解决样本标注问题上取得了很大进展,但这些方法往往只关注低预测置信度的不确定样本,未能综合考虑样本的不确定性与代表性,同时也没有充分利用CNN训练过程中得到的特征以及现有的带标签数据库信息. 为了克服上述缺陷,结合甲状腺癌病理图像特点,本文提出一种基于深度主动学习的甲状腺乳头状癌病理图像分类方法,充分利用现有的带标签的其他病理切片数据集,能够综合多种信息,同时考虑样本的不确定性与样本的重复性,通过有限数量的标记训练实例实现最优的分类效果.

## 2 基于深度主动学习的甲状腺癌病理图像分类方法

本文提出的方法主要包括三个部分:(1)深度CNN框架;(2)样本不确定性评估与相似性评估;(3)样本选择与模型更新. 将包含 $n_u$ 个实例的未标记甲状腺癌数据集表示为 $U = \{x_j^n\}_{j=1}^{n_u}$ . 首先根

据未标记样本的不确定性信息和代表性信息从数据集 $U$ 中选出 $b$ 个样本,组成集合 $U_B = \{x_j^u\}_{j=1}^b$ ,并对 $U_B$ 进行标记;然后用这些已标记的样本微调分类模型,改善模型的分类效果.

**2.1 深度CNN框架** 本文的深度CNN框架包括两个模块,如图1所示. 绿色虚线框内模块用于计算无标签样本的信息熵以及模型参数更新,红色虚线框内模块为特征哈希码提取模块,用于学习辅助数据集中的样本和甲状腺数据集中样本的深度哈希码. 选用包含 $na$ 个带标签的乳腺癌病理图像数据集<sup>[16]</sup> $A = \{x_i^a, y_i^a\}_{i=1}^{na}$ 作为辅助数据,对应的标签为 $y_i \in Y = \{1, \dots, C\}$ .

在特征哈希码提取模块中,使用VGG-f网络作为特征提取网络,该网络包括五个卷积层、三个全连接层. 通过引入哈希编码层将网络学习到的特征转化为二进制哈希码 $h(x)$ . 将VGG-f网络输出的样本深度特征记为 $u(x)$ ,将二进制哈希码记作 $h(x)$ , $h(x) = \text{sgn}(u(x))$ . 该哈希编码层受到有标签乳腺癌病理图像数据集 $A$ 的监督哈希损失函数 $L(H_A)$ 以及无监督熵损失函数 $L(U, A)$ 驱动. 有监督的哈希损失函数 $L(H_A)$ 确保乳腺癌病理图像样本生成的特征哈希码的相似性与判别性,也就是说,如果数据集 $A$ 中的样本 $x_i$ 和样本 $x_j$ 属于同一类别,则它们的哈希码是相似的. 无监督熵损失函数 $L(U, A)$ 根据数据集 $U$ 产生的哈希

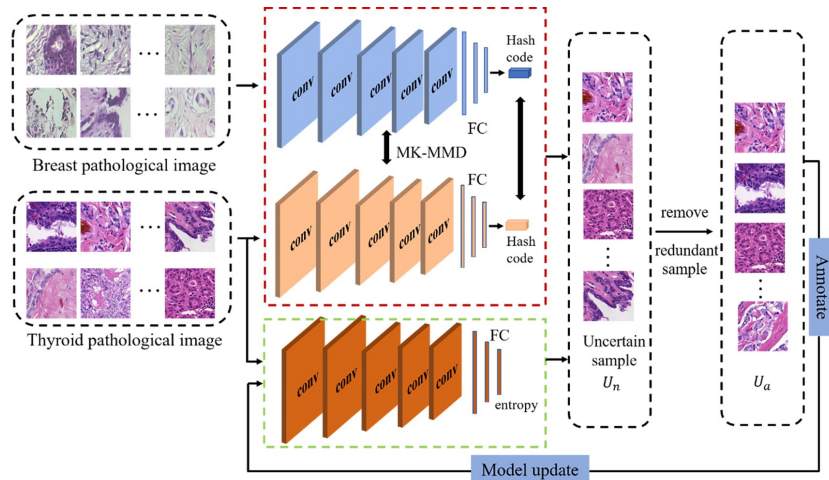


图1 分类方法整体框架图

Fig. 1 The overall framework of the proposed classification method

码和数据集  $A$  产生的哈希码的相似性对这两个数据集中的样本进行对齐,使每个甲状腺病理图像样本与乳腺癌数据集中某个类别的样本相似并且与其他类别的样本不相似.  $L(H_A)$  定义为:

$$L(H_A) = - \sum_{s_{ij}} \left( s_{ij} h(x_i^a)^T h(x_j^a) - \lg \left( 1 + \exp \left( h(x_i^a)^T h(x_j^a) \right) \right) \right) \quad (1)$$

其中,  $s_{ij} \in \{0, 1\}$  表示样本  $x_i^a$  和样本  $x_j^a$  生成的哈希码的相似性. 如果样本  $x_i^a$  和样本  $x_j^a$  属于同一类, 则  $s_{ij} = 1$ , 否则  $s_{ij} = 0$ . 无监督熵损失函数  $L(U, A)$  表示为:

$$L(U, A) = - \frac{1}{n_u} \sum_{i=1}^{n_u} \sum_{j=1}^C p_j^s \lg(p_j^s) \quad (2)$$

其中,  $p_j^s$  为甲状腺癌数据点  $x_j^a$  分配到某一类别  $s$  的概率, 通过式(3)计算:

$$p_j^s = \frac{\sum_{k=1}^K \exp \left( (x_j^u)^T x_k^{a_s} \right)}{\sum_{d=1}^2 \sum_{k=1}^K \exp \left( (x_j^u)^T x_k^{a_d} \right)} \quad (3)$$

其中,  $K$  为类别  $s$  的样本数量,  $s \in \{1, \dots, C\}$ ,  $x_k^{a_s}$  和数据集  $A$  中类别  $s$  和类别  $d$  的第  $k$  个输出值. 通过最小化  $L(U, A)$ , 甲状腺癌数据集中的每个样本产生的哈希特征码与某一类乳腺癌病理图像特征哈希码相似, 而与其他类不相似. 由于本方法是对甲状腺乳头状癌病理图像的二分类,  $C=2$ , 辅助数据集的类别与甲状腺癌病理切片数据集类别相同.

甲状腺癌病理图像与乳腺癌病理图像虽然都是病理数据, 但这两种组织取自人体不同部位, 切片制作过程不同, 且扫描时使用的扫描仪不同, 因此需要减小这两种数据的差异. 对于病理图像, 细胞核的特征是关键, 底层感受野较小的神经元能够学到核的形状及分布等特征, 网络的高层特征对分类十分重要, 因此对网络的五个卷积层与全连接层进行特征对齐, 以减少甲状腺癌病理图像特征与乳腺癌病理图像特征表示之间的域差异. 多核最大均值差异 (Multi-Kernel Maximum Mean Discrepancy, MK-MMD) 在减少域差异方面表现优异. 然而现有的减小域差异的模型大多针对自然图像数据, 需要对 MK-MMD 进行改进.

多层 MK-MMD 损失函数表示为:

$$M(U, A) = \sum_{o \in O} m(U_r^o, A_r^o) \quad (4)$$

其中,  $O$  为五个卷积层与全连接层的索引,  $U_r^o$  和  $A_r^o$  分别代表甲状腺癌病理图像数据和乳腺癌病理图像数据的在第  $o$  层的输出. 最大平均偏差  $m(U_r^o, A_r^o)$  表示为:

$$m(U_r^o, A_r^o) = \left\| \frac{1}{|U|} \sum \phi(x^u) - \frac{1}{|A|} \sum \phi(x^a) \right\|_K \quad (5)$$

其中,  $x^u$  和  $x^a$  分别表示数据集  $U$  和  $A$  的样本,  $\phi(\cdot)$  表示将数据集  $U$  和  $A$  的样本映射到一个再生核希尔伯特空间的核函数.  $K$  表示  $k$  个单一内核的组合:

$$K \doteq \left\{ k: k = \sum_{m=1}^k \beta_m k_m, \sum_{m=1}^k \beta_m = 1, \beta_m \geq 0, \forall m \right\} \quad (6)$$

网络总的损失函数为:

$$L_Q = \alpha L_{\text{cls}} + \beta L(H_A) + \lambda L(U, A) + \mu M(U, A) \quad (7)$$

其中,  $\alpha \in \{0, 1\}$  为自适应参数,  $L_{\text{cls}}$  为交叉熵分类损失函数. 主动学习过程中样本训练的不同阶段,  $\alpha$  取值不同. 在计算样本的特征哈希码阶段  $\alpha=0$ , 在样本信息熵计算阶段  $\alpha=1$ .  $\beta, \lambda$  和  $\mu$  用于平衡各个损失函数的重要性.

样本信息熵计算模块主要用于计算未标注样本信息熵并更新模型. 本文采用 VGG-f 模型来计算样本的信息熵. 为了适应甲状腺癌病理图像数据的类别, 用一个二输出的全连接层代替原有的最后一个全连接层.

**2.2 不确定性评估与相似性评估** 在主动学习中, 如何选择“有价值”的样本对提升模型的性能十分重要, 主动学习的关键在于建立一个判断候选对象“价值量”的标准. 本研究通过评估样本的不确定性与样本间的相似性来选择标注样本. 较高的不确定性表示样本有较高的信息量, 样本间相似性的评估能够辅助去除冗余样本, 以更少的标注样本获得最优的分类模型. 本方法中样本的不确定性评估准则由两部分组成: 特征哈希码的相似性以及样本信息熵. 特征哈希码的相似性定义为: 在获取甲状腺癌病理图像数据集与乳腺癌病理图像数据集的样本特征哈希码后, 甲状腺癌



数据集  $U$  中的样本  $x_j^u$  的特征哈希码应与某一类别乳腺癌病理图像生成的特征哈希码相似,而与其他类别乳腺癌病理图像特征哈希码有较大差异,即样本  $x_j^u$  的特征哈希码分配到某一类别的概率接近 1,而分配到其他类别的概率接近 0. 由于甲状腺乳头状癌病理图像分类为二分类,则分配到两个类别的概率  $\{p_j^{(s)}\}_{s=1}^2$  差值越小,表示样本的不确定性越大.

未标记样本  $x_j^u$  的信息熵  $e_j^u$  也是不确定性评价的重要指标,信息熵  $e_j^u$  表示数据集  $U$  中样本  $x_j^u$  提供的信息量,因此不确定样本的选择综合考虑了样本熵以及哈希特征码的相似性. 选择不确定性样本的指标如下:

$$Uncer(x_j^u) = \lambda_1 e_j^u - \lambda_2 |p_j^{(1)} - p_j^{(2)}|, \text{st. } \lambda_1 + \lambda_2 = 1 \quad (8)$$

其中,  $e_j^u$  的计算式为:

$$e_j^u = -(P_1 \lg P_1 + P_2 \lg P_2) \quad (9)$$

$P_1$  和  $P_2$  分别为信息熵计算模块中未标记样本  $x_j^u$  被分类为正常图像和 PTC 的概率.

直接进行图像的相似度计算十分复杂,然而通过 CNN 学到的深度特征能够很好地表征图像,因此可以将图像深度特征向量之间的相似度作为图像间的相似度. 取分类网络最后一个卷积层的多个特征图均值作为样本  $x_j^u$  的特征向量  $f_j^u$ . 在本方法中, VGG-f 网络最后一个卷积层输出特征图的尺寸为  $6 \times 6 \times 256$ , 即有 256 个特征图, 每个特征图大小为  $6 \times 6$ . 为了减小计算量, 以每个特征

图的均值代表该特征图, 最终得到一个尺寸为  $1 \times 256$  的特征向量. 最后, 使用余弦相似度来计算特征向量的相似性. 即数据集  $U$  中的两个样本  $x_i^u$  和  $x_j^u$  的相似性  $\text{sim}(x_i^u, x_j^u)$  可以表示为:

$$\text{sim}(x_i^u, x_j^u) = \cos\_sim(f_i^u, f_j^u) \quad (10)$$

**2.3 样本选择与模型更新** 注释样本的选择综合考虑样本的不确定性以及代表性. 首先从未标注甲状腺癌数据集  $U$  中选出不确定性较高的样本集合  $U_n$ , 然后评估集合  $U_n$  中样本间的相似性并去除冗余样本, 从  $U_n$  中选择有代表性的样本集合  $U_B$  并进行标注.

在待标注样本选择阶段, 目标是根据预先设定的标注数量选出同时具有不确定性和代表性的样本集合  $U_B$  并进行标记, 进一步进行模型更新, 改善模型的分类效果. 首先根据式(8)计算样本的不确定性, 选取不确定性得分最高的  $b$  个图像组成不确定性样本集合  $U_n$ ; 然后从同一患者样本相似性评估以及患者间样本相似性评估两个角度评估样本的相似性, 选出代表性样本构成集合  $U_B$ , 如图 2 所示.

在初始阶段, 由于不同数据集之间样本哈希码相似性概率的获取不需要样本标签, 因此能够直接对特征哈希码提取模块进行训练. 对于熵的计算, 随机选择少量训练样本进行标注, 然后使用标注的数据微调预训练 VGG-f 模型, 并使用微调后的模型计算样本信息熵. 主动学习是一个不断循环的学习过程, 在训练过程中, 特征哈希码提取

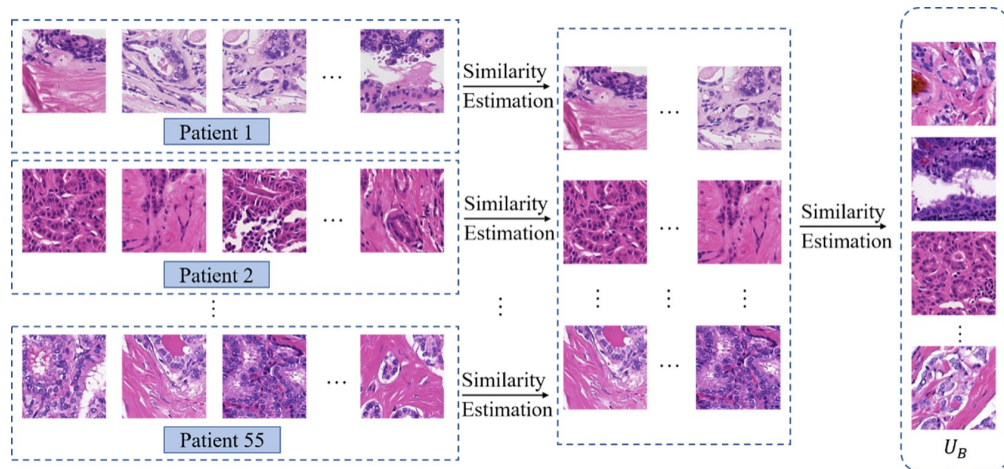


图 2 相似性样本评估示意图

Fig. 2 The framework of similarity estimation

模块和样本信息熵计算模块交替进行训练. 当特征哈希码提取模块进行训练时,  $\alpha$  设置为 0, 此时得到样本哈希码的相似性. 当样本信息熵计算模块进行训练时,  $\alpha$  设置为 1,  $\beta, \lambda, \mu$  设置为 0, 获取样本信息熵.

随着训练的进行, 新标记的样本被合并到  $U_B$  中, 并使用  $U_B$  不断微调 CNN 网络. 一些研究<sup>[5,17]</sup>已经证明, 对模型进行微调和使用标注样本从头开始训练模型相比, 可以获得更好的性能. 因此, 在我们的实验中, 当新标注的样本合并到数据集  $U_B$  时, 使用新标记的数据对 CNN 进行微调, 以不断更新模型参数, 改善模型的分类性能.

### 3 实验结果与分析

本文建立了甲状腺病理图像数据集以证明该方法的有效性. 收集西京医院 2017 年至 2018 年的 55 例甲状腺乳头癌的病理图像, 构建的数据集包括 7928 例正常病理图像和 8572 例 PTC 病理图像, 均由江丰生物信息技术有限公司的 KF-PRO-005 扫描仪扫描. 将数据集在患者级别随机分为训练、验证和测试数据集, 并使用 30% 的数据集作为测试数据集.

本文构建了三组实验并使用甲状腺癌病理图像数据库对该方法进行验证. 算法通过深度学习工具箱 MatConvNet<sup>[18]</sup>实现, 运行的软件环境为 64 位 Ubuntu14.04 系统, MatlabR2014b, Matconvnet 深度学习工具包.

**3.1 特征哈希码维度对分类效果的影响** 待标注样本选择的主要依据为样本的不确定性, 不确定性信息评估准则包括样本特征哈希码的相似性与样本信息熵. 其中, 哈希码的维度对不确定性样本的选择有一定影响, 并影响模型的分类性能. 为了验证特征哈希码维度对模型分类效果的影响, 本实验将哈希码维度设置为 16, 32, 64 和 128 bits, 实验结果如图 3 所示.

可以看出, 当哈希码维度为 16 bits 时分类准确率略低, 主要是因为当哈希码的维度较低时, 信息表征的准确性受到影响, 使分类准确率略微下降. 当哈希码的维度设置为 32 和 64 bits 时, 性能略微提升; 当哈希码维度增加为 128 bits 时, 分类准确率却没有明显变化, 说明 64 位哈希码已能较

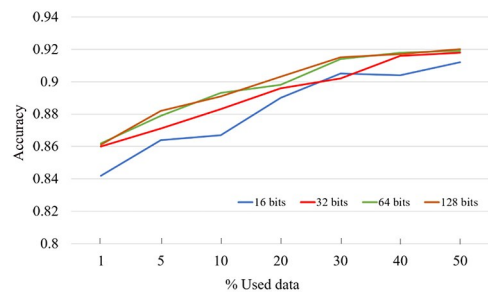


图 3 不同哈希码维度分类准确率的对比

Fig. 3 Classification accuracy with different hash length

准确地表征病理图像信息. 哈希码维度在 64 bits 的基础上再增加, 对病理图像的特征表示没有更多贡献, 因此将哈希码的维度设置为 64 bits.

**3.2  $\lambda_1$  与  $\lambda_2$  比值对分类效果的影响** 前文已经说明样本不确定性度量准则包括样本的特征哈希码相似性与样本信息熵两个部分, 为探究这两者比重对样本挑选的影响, 构建以下实验来确定不确定信息计算(式(8))中比例系数  $\lambda_1$  与  $\lambda_2$  的取值对病理图像分类准确率的影响, 实验结果如表 1 所示, 其中黑体字表示最好的分类结果. 可以看出, 当仅使用特征哈希码相似性与样本信息熵选择不确定性样本时模型分类准确率较低, 且需标注 40%~50% 的数据才能达到 90% 的分类准确率. 而两者联合进行不确定性样本选择时, 挑选出来的样本对模型的贡献率较高, 仅需标注 20%~30% 的样本便可以达到 90% 的分类准确率.

表 1 不同标注比例的分类结果 ( $\lambda_1: \lambda_2$ )

Table 1 Classification results with different ratio for  $\lambda_1: \lambda_2$

$\lambda_1: \lambda_2$	1%	5%	10%	20%	30%	40%	50%
0:10	0.822	0.834	0.887	0.894	0.905	0.904	0.912
1:9	<b>0.862</b>	<b>0.879</b>	<b>0.893</b>	<b>0.898</b>	<b>0.914</b>	<b>0.918</b>	<b>0.919</b>
2:8	0.821	0.859	0.882	0.891	0.904	0.913	0.912
3:7	0.830	0.849	0.856	0.859	0.899	0.914	0.914
4:6	0.859	0.860	0.892	0.901	0.904	0.907	0.908
5:5	0.831	0.840	0.868	0.879	0.888	0.905	0.907
6:4	0.832	0.825	0.860	0.865	0.904	0.908	0.915
7:3	0.833	0.844	0.843	0.899	0.890	0.898	0.911
8:2	0.840	0.851	0.879	0.886	0.906	0.908	0.913
9:1	0.835	0.840	0.863	0.893	0.900	0.912	0.916
10:0	0.835	0.837	0.845	0.852	0.866	0.902	0.917

率,  $\lambda_1: \lambda_2 = 1:9$  时模型的分类性能最好.

**3.3 不同标注比例分类结果对比实验** 为了验证本文算法在不同标注比例数据上的有效性, 将其与现有的几种基于主动学习的分类方法进行了对比, 共对比了 Uncertainty<sup>[19]</sup>, QBC<sup>[6]</sup>, GD<sup>[7]</sup>, QUIRE<sup>[8]</sup>, SPAL<sup>[9]</sup> 以及 Random 六种方法.

由于本文算法的目的是使用尽可能少的标注成本获取能准确进行甲状腺癌病理图像分类的模

型, 因此主要对不同比例的标注图像的分类准确率进行比较. 实验结果如表 2 所示, 其中黑体字表示最好的分类结果. Random 方法表示在未标记池中随机选取样本进行标注的方法. 为确保实验的公平性, 基于主动学习的对比方法使用的样本特征与本文方法使用的样本特征相同, 即 VGG-f 网络提取的深度特征. 同时, 初始标记样本数均设置为训练样本总数的 0.5%.

表 2 不同方法分类结果

Table 2 Classification results of different methods

Used data	Uncertainty	QBC	GD	QUIRE	SPAL	Random	Ours
1%	0.817	0.782	0.823	0.830	0.840	0.702	<b>0.862</b>
5%	0.834	0.795	0.842	0.835	0.854	0.809	<b>0.879</b>
10%	0.849	0.800	0.858	0.850	0.836	0.848	<b>0.893</b>
20%	0.865	0.802	0.867	0.855	0.857	0.860	<b>0.898</b>
30%	0.874	0.801	0.876	0.868	0.898	0.866	<b>0.914</b>
40%	0.877	0.802	0.873	0.873	0.912	0.866	<b>0.918</b>
50%	0.880	0.803	0.876	0.879	0.915	0.872	<b>0.919</b>

可以看出, 当标注数据量占总数据量的 1%, 5%, 10%, 20%, 30%, 40% 及 50% 时, 与其他方法相比, 本文方法的分类准确率最高, 且仅需标注不到 30% 的数据, 就可获得 90% 的分类准确率.

此外, 还将本文方法与原始的 VGG-f 网络的分类效果进行对比, 如图 4 所示. 可以看出, 本文方法仅需标注 30% 的样本, 便超过原始 VGG-f 网络使用全部数据时的分类准确率, 能够节约 70% 左右的标注成本.

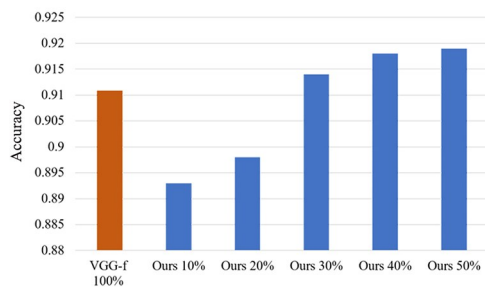


图 4 VGG-f 与本文方法分类性能对比

Fig. 4 Performance of VGG-f and our method

## 4 结 论

本文提出一种基于深度主动学习的甲状腺癌病理图像分类方法, 能有效减少标注代价. 该方法

针对病理图像标注困难的问题, 将深度学习与主动学习集成到一个框架中, 无须标注所有的训练数据, 只需选择对模型“有价值的”样本并进行标注, 使用较少的标注数据便能提高模型的性能.

所提出的方法首先利用样本特征哈希码相似性与样本信息熵挑选不确定样本, 然后进一步评估不确定性样本间的相似性, 并去除冗余样本. 最后对选择的样本进行标注并使用标注的数据更新模型参数, 提升模型的性能. 在甲状腺病理图像数据集上的实验结果表明, 所提出方法能够有效降低标注代价.

## 参 考 文 献

- [1] Global Burden of Disease Cancer Collaboration. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the global burden of disease study. JAMA Oncology, 2017, 3(4): 524—548.
- [2] 黄庆文, 陈伊, 李华贵. 甲状腺癌 CT、B 超诊断与病理诊断对照分析研究. 中外医疗, 2017, 36(28): 179—180, 183. (Huang Q W, Chen Y, Li H G, et al.

- Comparative analysis of CT diagnosis, B ultrasound diagnosis and patho-logical diagnosis of thyroid carcinoma. *China Foreign Medical Treatment*, 2017, 36 (28):179—180,183.)
- [3] 陈健. 精确病理诊断在甲状腺癌精准医疗中的意义. *中国肿瘤临床*, 2017, 44(4):181—185. (Chen J. Accurate pathological diagnosis of thyroid cancer in the era of precision medicine. *Chinese Journal of Clinical Oncology*, 2017, 44(4):181—185.)
- [4] 王伟. 甲状腺乳头状癌颈部转移淋巴结清扫研究进展. 硕士学位论文. 蚌埠: 蚌埠医学院, 2015. (Wang W. Research progress in metastatic cervical lymph node dissection of thyroid papillary carcinoma. Ph. D. Dissertation. Bengbu: Bengbu Medical College, 2015.)
- [5] Venkateswara H, Eusebio J, Chakraborty S, et al. Deep hashing network for unsupervised domain adaptation//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA: IEEE, 2017:5018—5027.
- [6] Abe N, Mamitsuka H. Query learning strategies using boosting and bagging//*Proceedings of the Fifteenth International Conference on Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998, 388:1—9.
- [7] Ebert S, Fritz M, Schiele B. RALF: a reinforced active learning formulation for object class recognition//*2012 IEEE Conference on Computer Vision and Pattern Recognition*. Providence, RI, USA: IEEE, 2012:3626—3633.
- [8] Huang S J, Jin R, Zhou Z H. Active learning by querying informative and representative examples//*Proceedings of the 23<sup>rd</sup> International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2010:892—900.
- [9] Tang Y P, Huang S J. Self-paced active learning: Query the right thing at the right time//*Proceedings of the AAAI Conference on Artificial Intelligence*, Honolulu, HI, USA: IEEE, 2019, 33:5117—5124.
- [10] Wang D, Shang Y. A new active labeling method for deep learning//*2014 International Joint Conference on Neural Networks*. Beijing, China: IEEE, 2014: 112—119.
- [11] Li J M. Notice of removal: active learning for hyperspectral image classification with a stacked autoencoders based neural network//*2016 IEEE International Conference on Image Processing*. Phoenix, AZ, USA: IEEE, 2016:1062—1065.
- [12] Stark F, Hazirbaş C, Triebel R, et al. CAPTCHA recognition with active deep learning//*German Conference on Pattern Recognition Workshop*. Springer Berlin Heidelberg, 2015:94—101.
- [13] Al Rahhal M M, Bazi Y, AlHichri H, et al. Deep learning approach for active classification of electrocardiogram signals. *Information Sciences*, 2016, 345:340—354.
- [14] Zhou Z W, Shin J, Zhang L, et al. Fine-tuning convolutional neural networks for biomedical image analysis: actively and incrementally//*2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA: IEEE, 2017: 7340—7349.
- [15] Yang L, Zhang Y Z, Chen J X, et al. Suggestive annotation: a deep active learning framework for biomedical image segmentation//*International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer Berlin Heidelberg, 2017:399—407.
- [16] Camelyon. <https://camelyon16.grand-challenge.org/Data/>. 2016.
- [17] Zhou Z W, Shin J, Feng R B, et al. Integrating active learning and transfer learning for carotid intima-media thickness video interpretation. *Journal of Digital Imaging*, 2019, 32(2):290—299.
- [18] Vedaldi A, Lenc K. MatConvNet: convolutional neural networks for MATLAB//*The 23<sup>rd</sup> ACM International Conference*. Brisbane, Australia: ACM, 2015:689—692.
- [19] Lewis D D, Gale W A. A sequential algorithm for training text classifiers//*Proceedings of the 17<sup>th</sup> ACM International Conference on Research and Development in Information Retrieval*. Springer Berlin Heidelberg, 1994:3—12.

(责任编辑 杨可盛)