

基于多特征描述的乳腺癌肿瘤病理自动分级

龚磊¹, 徐军^{1*}, 王冠皓¹, 吴建中², 唐金海²

(1. 南京信息工程大学 江苏省大数据分析技术重点实验室, 南京 210044; 2. 江苏省肿瘤医院, 南京 210000)

(* 通信作者电子邮箱 jxu@nuist.edu.cn)

摘要: 为了辅助病理医生快速高效诊断乳腺癌并提供乳腺癌预后信息, 提出一种计算机辅助乳腺癌肿瘤病理自动分级方法。该方法使用深度卷积神经网络和滑动窗口自动检测病理图像中的细胞; 随后综合运用基于稀疏非负矩阵分解的颜色分离、前景标记的分水岭算法以及椭圆拟合得到每个细胞的轮廓。基于检测到的细胞和拟合出的细胞轮廓, 提取出肿瘤的组织结构特征和上皮细胞的纹理形状特征等共 203 维的特征, 运用这些特征训练支持向量机分类器(SVM), 实现对病理组织图像自动分级。17 位患者的 49 张 H&E 染色的乳腺癌病理组织图像自动分级的 100 次十折交叉检验评估结果表明: 基于病理图像的细胞形状特征与组织的空间结构特征对病理图像的高、中、低分化等级分类整体准确率为 90.20%; 同时对高、中、低各分化等级的区分准确率分别为 92.87%、82.88%、93.61%。相比使用单一结构特征或者纹理特征的方法, 所提方法具有更高的准确率, 能准确地对病理组织图像中肿瘤的高级和低级分化程度自动分级, 且各分级之间的准确率差异较小。

关键词: 乳腺癌; 组织病理图像; 自动病理分级; 计算机辅助预后分析

中图分类号: TP391.41; TP182 **文献标志码:** A

Multi-feature based descriptions for automated grading on breast histopathology

GONG Lei¹, XU Jun^{1*}, WANG Guanhao¹, WU Jianzhong², TANG Jinhai²

(1. Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science and Technology, Nanjing Jiangsu 210044, China;

2. Jiangsu Cancer Hospital, Nanjing Jiangsu 210000, China)

Abstract: In order to assist in the fast and efficient diagnosis of breast cancer and provide the prognosis information for pathologists, a computer-aided diagnosis approach for automatically grading breast pathological images was proposed. In the proposed algorithm, cells of pathological images were first automatically detected by deep convolutional neural network and sliding window. Then, the algorithms of color separation based on sparse non-negative matrix factorization, marker controlled watershed, and ellipse fitting were integrated to get the boundary of each cell. A total of 203-dimensional image-derived features, including architectural features of tumor, texture and shape features of epithelial cells were extracted from the pathological images based on the detected cells and the fitted boundary. A Support Vector Machine (SVM) classifier was trained by using the extracted features to realize the automated grading of pathological images. In order to verify the proposed algorithm, a total of 49 Hematoxylin & Eosin (H&E)-stained breast pathological images obtained from 17 patients were considered. The experimental results show that, for 100 ten-fold cross-validation trials, the features with the cell shape and the spatial structure of organization of pathological image set successfully distinguish test samples of low, intermediate and high grades with classification accuracy of 90.20%. Moreover, the proposed algorithm is able to distinguish high grade, intermediate grade, and low grade patients with accuracy of 92.87%, 82.88% and 93.61%, respectively. Compared with the methods only using texture feature or architectural feature, the proposed algorithm has a higher accuracy. The proposed algorithm can accurately distinguish the grade of tumor for pathological images and the difference of accuracy between grades is small.

Key words: breast cancer; pathological image; automated pathological grading; computer-aided prognosis

0 引言

美国癌症协会 2015 年发布的数据显示, 预计 2015 年美国女性乳腺癌新增病例仍居女性所有癌症病例的首位, 所占

有新增癌症病例的 29%^[1]。目前我国乳腺癌发病率位居女性恶性肿瘤的第一位, 但相比肺癌、胃癌和肝癌, 其预后效果最好。因此如果乳腺癌能够早发现加上恰当的治疗方案, 患者的 5 年存活率可以达到 90% 以上^[1]。在乳腺癌诊断过程

收稿日期: 2015-05-18; 修回日期: 2015-08-01。基金项目: 国家自然科学基金资助项目(61273259); 江苏省“六大人才高峰”高层次人才项目(2013-XXRJ-019); 江苏省自然科学基金资助项目(BK20141482)。

作者简介: 龚磊(1991-) 男, 江苏扬州人, 硕士研究生, 主要研究方向: 病理图像分析、癌症的计算机辅助诊断与预后; 徐军(1972-) 男, 江西乐平人, 教授, 博士生导师, CCF 会员, 主要研究方向: 计算机视觉、机器学习、癌症的计算机辅助检测、诊断与预后; 王冠皓(1989-) 男, 江苏徐州人, 硕士研究生, 主要研究方向: 深度学习、压缩感知、稀疏表示; 吴建中(1965-) 男, 江苏南通人, 研究员, 主要研究方向: 肿瘤流行病学、肿瘤遗传学; 唐金海(1961-) 男, 江苏连云港人, 教授, 博士生导师, 博士, 主要研究方向: 乳腺肿瘤的临床诊治与研究。

中,采用病理组织学分级系统来判断肿瘤的恶性程度,目前通用的评价标准是世界卫生组织推荐的诺丁汉系统,通过综合评价染色切片的腺管构成、核异型性和有丝分裂次数这三个指标来预估患者临床预后结果,辅助医生治疗。

目前临床上的病理组织学分级基本上都是依赖于病理医生人工分析。但人工分级具有明显的局限性,大量的研究^[2-3]表明,不同的病理医生由于客观原因在人工分级上具有较大的不一致性,导致这种不一致性的主要原因是人工分析方法具有较强的主观性,容易受人为和环境因素的影响,此外这个过程比较耗时费力^[4]。这种不一致性对临床诊断是非常不利的,因此迫切需要定量的分级方法来解决人工分析的缺陷。

近年来,由于数字扫描技术的快速发展,“数字病理”开始受到广泛的关注^[5]。研究计算机辅助系统(Computer-Aided System, CAD)的目标不是为了完全代替医生,而是为了向医生提供更加准确的客观建议从而提高医生的工作效率,得到更加准确的诊断结果。虽然计算机辅助诊断具有诸多优势,但由于病理组织图像呈现出高度的复杂性使得病理组织图像的自动分析仍然是一个极具挑战性的研究领域^[6]。

尽管如此,近年来很多学者在乳腺癌数字病理图像分析方面取得了瞩目的进展^[3,7-10]。病理自动分级的前提^[11]是从病理图像中准确快速地检测和分割病理标志物(如细胞)。目前细胞检测方法通常与分割算法联系在一起,这些分割方法包括自适应阈值法^[9]、区域增长法^[11]、颜色聚类法^[12]、形态学基础的方法^[13],但是这些传统的方法对于高度异质性的乳腺细胞的检测不是非常有效。Xu 等^[14]采用了深度学习的方法检测细胞,由于深度学习是采用数据驱动的方法,因此相比以上提到的传统方法具有较高的鲁棒性和准确性。基于细胞检测结果,可以结合图特征的描述^[9-10]构建病理等级的自动分级系统。Cruz-Roa 等^[15]讨论了基于图特征的病理分级的稳定性和辨别能力。基于图特征的分级方法已经被用于淋巴瘤细胞浸润程度^[11]、肿瘤类型^[16]、肿瘤等级^[17]等研究中^[18]。基于细胞纹理的分级方法同样有着广泛的研究。文献[9]发现了细胞特殊的外观表现与乳腺癌等级之间存在关联。Doyle 等^[17,19]使用了 Gray-level、Gabor 和 Haralick 特征对乳腺癌^[17]和前列腺癌^[19]进行分级。

本文提出的病理分级方法采用了卷积神经网络加两种尺寸的滑动窗口,该方法在检测高度异质性的肿瘤细胞时,与传统的细胞检测方法相比具有更高的准确率。同时与运用单一特征的病理分级方法不同,本文方法除了提取细胞个体的纹理和形状特征外,还试图运用图描述的方法定量地描述细胞间的空间关系,多种特征的组合明显提升了乳腺癌病理自动分级系统的准确率。

1 乳腺肿瘤自动病理分级算法

本文提出的乳腺癌肿瘤病理图像自动病理分级方法包含四个模块:1) 预处理;2) 检测与分割;3) 特征提取;4) 自动分级。该方法的流程如图 1 所示。在图 1(a) 预处理模块中,首先对所有输入的 H&E(Hematoxylin & Eosin)染色的乳腺癌病理切片图像运用基于参考图像颜色的颜色标准化算法进行标准化,目的是消除由于图像采集、成像、染色等因素导致的图像颜色不一致的问题。在图 1(b) 检测与分割模块图中,使用深度卷积神经网络与滑动窗口算法自动快速地检测病理图像中的细胞;同时使用基于稀疏非负矩阵分解的颜色分离^[20]和自适应阈值算法得到粗略的细胞区域;接下来根据细胞检测

模块中得到的检测结果,运用基于前景标记的分水岭算法分割重叠的细胞团得到细胞边界,最后运用椭圆拟合得到每个细胞更加精确的轮廓。在图 1(c) 特征提取模块中,首先根据图 1(b) 检测分割模块中细胞检测得到的细胞质心构建细胞间的空间关系图^[15],提取病理组织的结构特征。同时根据细胞分割结果得到的单个细胞轮廓提取出细胞的纹理和形状特征。最后在图 1(d) 自动病理分级模块中,根据图 1(c) 模块中提取出的组织结构、上皮细胞的纹理及形态特征训练支持向量机分类器进行病理自动分级。

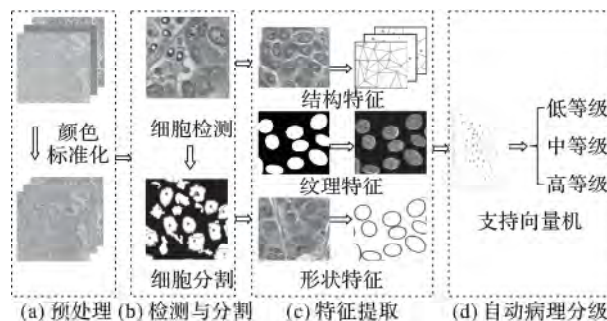


图 1 乳腺癌病理自动分级系统流程

1.1 病理图像预处理模块: 颜色标准化

在病理切片数字化制作过程中由于染色不均匀、数字成像、数字扫描设备不同等因素,会导致获得的数字化病理图像颜色差异的问题。这种差异会对随后的细胞自动检测、分割、特征提取带来不利的影响,因此使用文献[21]中提出的线性变换的方法在 Lab 颜色空间对数据库中所有的图像进行颜色标准化处理。从数据库中随机选取一张图像作为参考图像,然后对剩余的图像都运用标准化算法。该方法首先将 RGB 颜色空间的待处理 H&E 染色病理图像转化到特定的 CIELab (Lab) 颜色空间,对 Lab 各通道中的每一个像素进行一个如式(1)的线性转化:

$$l_{\text{转换图像}} = \frac{l_{\text{原始图像}} - \bar{l}_{\text{原始图像}}}{l_{\text{原始图像}}} \hat{l}_{\text{参考图像}} + \bar{l}_{\text{参考图像}} \quad (1)$$

其中 \hat{l} 和 \bar{l} 分别表示 L 通道所有像素的方差和均值。

将线性转换后得到的 Lab 颜色空间还原为 RGB 空间,即可得到与参考图像颜色分布一致的病理图像。

1.2 细胞的自动检测与分割模块

乳腺癌病理图像中主要有上皮和淋巴两种类型的细胞,80% 的癌变发生在上皮细胞区域。虽然上皮细胞和淋巴细胞都能反映出癌症病理等级信息,但由于细胞检测本身是一项极具挑战性的任务,同时检测两种类型完全不同的细胞是极为困难的,所以前人的研究一般不区分细胞类型。本文只关注上皮细胞的组织结构、形状和纹理特征。为了排除淋巴细胞的影响以及为今后研究淋巴细胞对肿瘤微环境的影响打下基础,本文采用分别训练两个卷积神经网络结合两种尺寸的滑动窗口分别检测上皮细胞和淋巴细胞。卷积神经网络能够从训练数据中隐式学习而不需要人工设计特征,面对形状高度异质性的肿瘤区域细胞具有明显的优势。本文使用的卷积神经网络共有 6 层结构:2 个卷积层、2 个池化层和 2 个全连接层,2 个卷积层和池化层交替连接,最后附加 2 个全连接层。卷积层与池化层能够保证在网络参数较少的情况下提取到对分类有效的显著特征,为了防止数据过拟合,本文在每个卷积层后都附加一个抑制率 50% 的 dropout 层。训练的两个卷积神经网络由于不同种类细胞的尺寸不同,所以两个网络第一个全连接层的参数有所不同,而前四层都是 5×5 的滤波

器加 2×2 的池化层。

本文统计了数据库中 40 倍放大的乳腺癌病理图像中大量细胞的尺寸: 上皮细胞和淋巴细胞的直径范围分别为 $25 \sim 40$ 像素和 $15 \sim 20$ 像素, 因此分别使用大小为 39×39 和 21×21 像素的滑动窗口检测上皮细胞和淋巴细胞。49 幅图像中约存在 36500 个上皮细胞和 6100 个淋巴细胞, 本文从各等级病理图像中挑选出 20 幅病理图像作为训练图像。剩下的 29 幅图像作为测试图像。在构建训练集时以专家标注的点为中心, 取滑动窗口大小的块作为训练样本。从 20 幅训练图像中随机选取出约 10000 个上皮细胞, 3000 个淋巴细胞以及 10000 个非细胞区域。

训练好卷积神经网络之后, 利用滑动框, 从病理图像的左上方开始, 从左往右、从上到下依次滑动, 每滑动一次都对窗口内的图像块进行判断, 是细胞则标记为细胞, 不是则跳过检测下一个窗口。对上皮细胞和淋巴细胞分别以 10 像素和 6 像素作为滑动步长, 能够在不降低准确率的同时缩短检测时长。检测的过程中由于步长小于窗口大小, 所以同一细胞可能会被多次标记。本文采用非最大值抑制的方法^[22], 即根据在一定区域内的检测响应值, 比较其大小, 取最大的保留。这样图像中的细胞位置就不会重复标记。

1.3 重叠细胞分割与轮廓拟合

计算细胞的形状纹理特征, 需从病理图像中分割出细胞区域。本文使用基于稀疏非负矩阵分解的颜色分离和自适应阈值方法得到细胞区域。在 H&E 染色的病理图像中: H 染色使得细胞核呈现深灰色, 而 E 染色使得非细胞区域呈现浅灰色。基于这样一个先验知识, 利用稀疏非负矩阵分解算法^[20]自动分解病理图像得到 H 通道(图 2(j)) 和 E 通道图像, 将 H 通道图像灰度化之后采用自适应阈值得到二值图像。此时得到的二值图像中基本为细胞区域。随后使用腐蚀、开操作、去除小块等形态学操作得到更为准确的细胞区域。

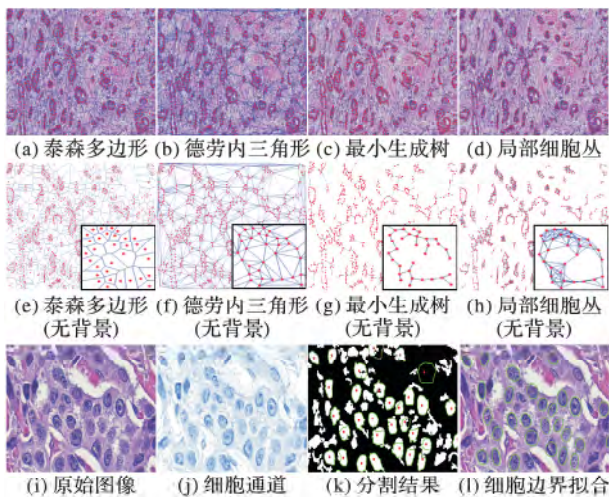


图 2 特征提取

由于细胞组织具有高密度的结构, 病理图像中存在大量的细胞重叠现象。在二值图像中获得的细胞区域存在细胞重叠的问题, 使用形态学操作能够一定程度上将细胞分开, 但对于椭圆拟合获得单个细胞边界还是不够的。根据图 1(b) 检测模块获得的细胞质心和分割算法得到的细胞区域, 将检测到的上皮细胞作为前景标记, 然后使用前景标记的分水岭算法自动分离重叠的细胞(图 2(k))。因为上皮细胞多为椭圆形, 在分水岭算法分割出单个细胞边界的基础上运用椭圆拟合的方法拟合更加精确的细胞轮廓(图 2(l))。对于可能存

在的错误分割边界, 可通过计算轮廓长短轴比进行一定程度的排除。如果椭圆轮廓的长轴不在 $25 \sim 45$ 或者长/短轴比大于 2, 则认为不是上皮细胞。

1.4 特征提取模块

根据细胞自动检测和分割模块得到的结果, 共提取出三种类型的特征: 组织结构特征、单个细胞的纹理和形状特征, 共计 203 维。具体的特征及其描述如表 1 和表 2 所示。本文分别运用全局图特征(global graph, gg)和局部细胞丛特征(cell cluster graph, ccg)这两类图特征定量描述肿瘤区域细胞间的空间结构信息。全局图特征包括: 泰森多边形(图 2(a)、(e)), 德劳内三角形(图 2(b)、(f)), 最小生成树(图 2(c)、(g))及细胞密度信息这四种图特征。局部细胞丛特征(图 2(d)、(h))包括节点数目、边长、聚类信息、半/直径等相关统计信息。在图 2 第一行展示了以病理图像为背景的全局图特征和局部细胞丛特征, 在图 2 第二行展示第一行除去背景的图特征, 并对部分细节放大展示。在构建图时以检测到的上皮细胞为质心构建空间结构图。除了空间结构图特征外, 本文还根据单个细胞分割的结果提取能定量描述细胞核多形性相关的特征: 纹理(texture)和形状特征(shape)。其中纹理特征采用了 Haralick 特征, 该特征已被广泛应用于病理组织图像的自动分级^[17-23]中。形状特征包括整张病理图像中细胞的面积、周长、长轴与短轴比、平滑度等形状统计信息。因为一般情况下淋巴细胞大小一致, 纹理均匀, 在肿瘤和非肿瘤区域中没有明显统计差异性, 所以仅提取了上皮细胞的形状和纹理特征。本文所提取特征及其相关的描述见表 1~2。

表 1 组织结构特征的具体描述

特征类型	描述	维数	具体细节
全局图特征	Voronoi	12	Area, Perimeter, Chord
	Delaunay	8	Side Length, Triangle Area
	MST	4	MST Edge Length
	Nuclear	27	Density of Nuclei, Distance, Number of Nuclei
局部细胞丛特征	Nodes	7	End Nodes, Central Nodes, Isolated Nodes
	Edges	5	Number, Edge Length
	Clustering	6	Clustering Coefficients, Connected Components
	Others	8	Radius, Diameter, Path Length, Degree

表 2 纹理特征和形状特征的具体描述

特征类型	描述	维数	具体细节
纹理特征	Contrast	10	Mean, Inverse Moment, Energy, Entropy, Variance
	Intensity	6	Mean Value, Entropy, Variance
	Energy	4	Energy, Entropy
	Others	6	Information Measure
形状特征	Area	4	Area Ratio
	Distance	16	Distance Ratio, Variance, Standard Deviation
	Perimeter	4	Perimeter Ratio
	Shape	76	Invariant Moment, Fractal DIM, Fourier Descriptor

1.5 自动病理分级模块

本文共提取 4 类特征共计 203 维, 如表 1~2 中所示。基

于这些特征以及相应的根据 Oncotype DX 评分得到的病理分化等级来训练支持向量机 (Support Vector Machine, SVM) 分类器。由于提取的特征维数较高, 为了避免“维数灾难”和分类器的“过拟合”问题, 在分类器训练之前, 采用了谱嵌入 (Graph Embedding, GE) 对高维特征进行降维。GE 是一种非线性降维方法, 能够将高维的图像特征降低到低维的同时保持原始特征空间图像之间的相关距离。实验中对降低的维数进行了多次测试, 选择效果最好的作为实验的结果。

由于病理图像的高度复杂性, 因此需要多个步骤才能达到本文设定的目标。本文主要可以分为预处理、检测与分割、特征提取和自动分级四大模块, 因此每一个环节都很重要, 如果某一个环节处理得不好, 都会影响整个结果。在预处理模块中, 只要使用相同的目标图像, 标准化后的图像都将具有相同的颜色分布, 所以对于不同的数据都是适用的。因为 H&E 染色和免疫组化染色具有较大的差别, 故本文只关注 H&E 染色的病理图像。对于检测与分割模块, 本文卷积神经网络的训练集为 40 倍放大的乳腺肿瘤细胞, 在相似的其他 40 倍放大病理图像数据库中具有较高的准确性, 已经通过实验在另外的数据库上进行了证实。在非 40 倍放大倍率或者分化等级高且存在组织坏死的病理图像中则检测准确率有所下降。基于稀疏非负矩阵分解的颜色分离在面对其他的数据时, 可能需要调整参数来达到最佳的效果。同样自适应阈值和形态学操作也可能需要调整参数来达到最佳的效果。特征提取模块中, 特征提取依赖于前面所检测到的细胞位置、轮廓信息。在自动分级模块中, 选择不同的降维维数会对准确率造成一定影响。

本文预处理和细胞检测具有较好的鲁棒性, 能够适应不同的数据库。细胞检测是病理分级重要的前提, 对后续实验有着极为重要的影响。卷积神经网络能够提供准确的细胞检测结果。同时病理图像中存在巨量的上皮细胞, 极少量的误检漏检不会对最终结果造成较大的影响, 故本文方法具有较好的鲁棒性。

2 实验设计及结果分析

为了得到可靠的准确率, 本文所有的实验过程都采用了经典的 100 次 10 折交叉验证。每一次 10 折交叉验证是将原始数据随机分成 10 组, 将每个子集数据分别作一次测试集, 其余的 9 组子集数据作为训练集, 这样得到的 10 个分类准确率的平均数作为这次 10 折交叉验证下分类器的分类准确率。对这种 10 折交叉验证迭代 100 次后得到的 100 个分类准确率作为最终分类准确率。

2.1 实验数据

实验使用的数据为从美国新泽西癌症研究所获取的来自 17 位乳腺癌患者的 49 张 H&E 染色 Aperio 光学扫描仪 40 倍放大成像的乳腺癌病理图像。图像的尺寸约为 2512×1792 。所有的图像均采用 Oncotype DX 方法评分^[24], 该评分方法根据分值的大小把病理组织切片分为高、中、低三个分化等级。这个等级作为本文确定病理等级的基准。

2.2 实验环境

实验使用的硬件设备如下: 处理器为 Intel Core i7-3770 CPU @ 3.40 GHz; 内存 (RAM) 为 16.0 GB; 系统类型为 64 位

Window7 操作系统; 开发工具为 Matlab R2012a。

2.3 基于上皮细胞的不同类型特征自动分级能力的评估

本实验中利用检测与分割模块中得到的上皮细胞的质心和边界分别提取组织的空间结构特征 (gg + ccg) 和上皮细胞的形状特征和纹理特征。为了能够分别评估这三类特征病理分级的能力, 本文分别对三种类型特征的自动分级能力进行了对比实验, 实验结果如表 3 所示。图 3(a) ~ (c) 显示了上皮细胞三类特征的自动分级准确率的受试者工作特征曲线 (Receiver Operating Characteristic, ROC) 曲线。不同的 ROC 曲线分别代表不同类型的特征, 曲线越靠近左上角, 则曲线下的面积 (Area Under Curve, AUC) 越大, 说明该类型特征的分级准确性就越高。

表 3 上皮细胞不同类型的特征分级准确率

特征类型	准确率/%
结构特征	71.83
形状特征	77.14
纹理特征	75.17

2.4 不同特征组合自动分级能力的评估

本文提取的特征有: 细胞组织的空间结构特征、上皮细胞的纹理特征和形状特征。为了评估不同类型特征组合对分级能力的影响, 分别对三种类型特征进行两两组合 (6 种) 和三三组合 (4 种)。图 3(d) ~ (i) 展示了不同类型特征组合 (两两组合、三三组合) 自动分级性能评估的 ROC 曲线。表 4 显示了不同类型特征组合的分类准确率。其中: gg 表示全局图特征; ccg 表示局部细胞丛特征; shape 表示形状特征; texture 表示纹理特征。

表 4 上皮细胞不同类型特征组合分级准确率

组合类型	特征	准确率/%
特征 两两 组合	gg + ccg	71.83
	gg + shape	88.53
	gg + texture	75.72
	ccg + shape	90.20
	ccg + texture	74.95
	shape + texture	71.46
特征 三三 组合	gg + ccg + shape	86.84
	gg + ccg + texture	76.38
	gg + shape + texture	86.16
	shape + ccg + texture	87.65

2.5 本文方法与现有方法自动分级能力的评估与对比

为了说明该方法的有效性及其优越性, 本文与近年经典的三种方法作了对比实验: 第一种方法为淋巴细胞的组织结构特征加上支持向量机的方法^[11]; 第二种方法为上皮细胞的组织结构特征结合全局纹理特征 (Gabor 特征 + 一阶统计特征 + 二阶统计特征) 加上支持向量机的方法^[17]; 第三种为结合上皮细胞组织结构特征和纹理特征的多视角多分类器 (multi-Field-of-View, multi_FOV) 的方法^[18]。考虑到不同的数据库中细胞的高度异质性, 这些对比方法的细胞检测算法不一定具有较高的鲁棒性。而细胞检测是病理分级的重要前提, 因此对比实验中将专家标注的细胞点作为这些方法特征提取的基准。而本文的方法依旧使用检测分割模块检测出的细胞点。由于本文所用的病理图像分辨率和尺寸的差异, 本文在进行第三种 multi_FOV 方法时, 并未采用 3×3 、 6×6 和

9×9的多视角模式,而是采用了2×2,3×3和4×4的多视角模式。表5中,展示了各个方法病理三分级的准确率。图

3(j)~(l)中展示了本文方法与三种对比方法自动分级性能的ROC曲线。

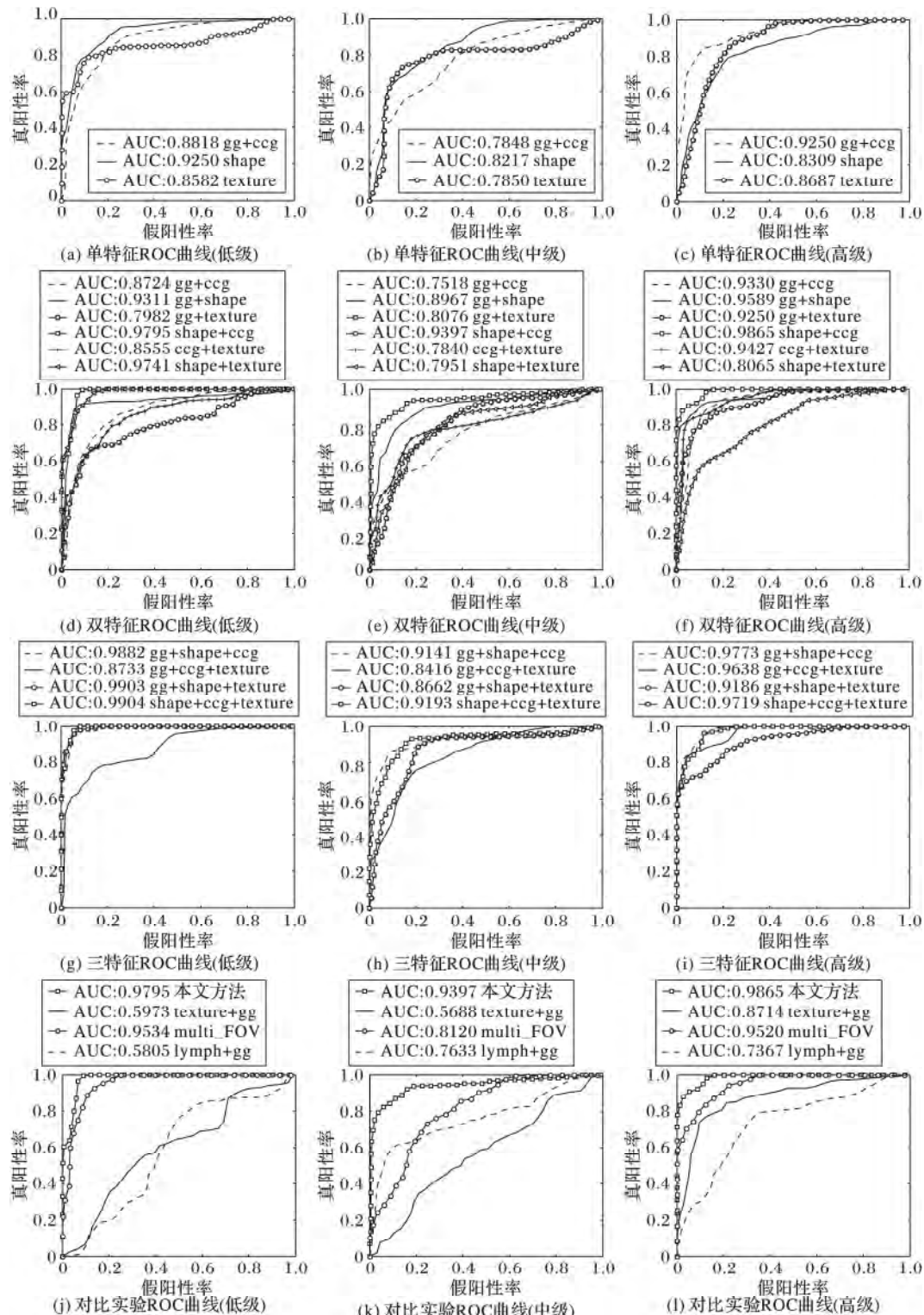


图3 自动分级性能评估 ROC 曲线

表5 不同实验方法分级准确率 %

算法	低级	中级	高级	整体
lymph + gg ^[11]	37.53	66.50	92.88	79.35
texture + gg ^[17]	30.51	54.07	91.75	75.23
multi_FOV ^[18]	56.46	66.97	74.47	67.87
本文方法	93.61	82.88	92.87	90.20

3 讨论

在进行100次10折交叉检验之后,本文得到了基于单一特征及多种特征组合的病理自动分级的评估结果。从表3可以看出,单一类型特征中形状特征的分级准确率最高,纹理特征次之,但这三种类型特征之间分级能力区别不大,均为

70% 以上。从表 4 可以看出多种特征组合实验中, ccg 和 shape 特征是最优的特征组合, 并达到了实验的最高准确率 90.20%, 可以看到 ccg 和 shape 特征组合的 ROC 曲线基本位于其他特征组合之上(图 3(d) ~ (f))。这一实验结果也验证了诺丁汉评分系统中的分级指标: 腺管形成和核异型性是两项重要指标。腺管形成即组织结构随着癌变分化程度的提高, 癌变细胞组织结构特征与正常组织结构之间的差异也就越大。对于组织结构的空间结构特征, 人的视觉是无法定量描述的。本文中运用的组织结构特征能够对病理图像中上皮细胞的空间分布作出一个量化的描述。核异型性表现为肿瘤细胞核与正常细胞核在尺寸、形状和染色等方面的异常性, 病理分化等级较高的 H&E 染色的肿瘤组织切片中, 通常细胞尺寸较大, 颜色较深, 形状也变得不规则。对于核异型性评价指标, 本文实验中形状特征相比纹理特征病理分级的区分度更高。这也与医生的诊断经验相符, 形状的变化相比纹理更加直观和更易被观察。纹理特征分类能力不及形状特征的另一个可能原因在于虽然经过颜色标准化, 但某些病理图像制作过程中由于染色过度, 很多的细胞均被染成深紫色的颜色块, 丢失了部分纹理信息。此外三种特征组合的准确率相对两种特征组合没有提高, 说明了实验数据数量一定的情况下, 特征并非越多越好, 因为特征之间可能存在相关性, 或者某些维的特征分级能力较弱。

对于最优的 ccg 和 shape 特征组合, 它们对高、中、低三个分化等级的分级区分度之间存在一定的差异性: 本文的自动分级系统对高、中、低各分化等级的区分准确率分别为 92.87%、82.88%、93.61%, 在区分高、低分化等级的病理图像上面表现出来的区分准确率较高。低级(图 3(d))与高级(图 3(f))具有较高的准确率, 它们的 AUC 比中级(图 3(e))的 AUC 值高。从图 3(d) ~ (f) 可以看出, 对于多数的特征组合均有这样的趋势: 对于低级和高级的病理图片有较好的分级准确率, 中级的图片较难以分级。这个结论也与临床中病理医生的病理分级的难易程度一致。这是因为与中分化等级相比, 高分化、低分化病理图像的细胞形状与空间结构特征在图像中表现出来的差异性较为明显, 支持向量机可以找到一个较好的分类超平面, 而介于两者之间的中分化等级的病理图像的特征与低级、高级分化等级之间的差异并不是很明显, 所以分类器较难区分。

在与其他病理分级方法对比的实验中, 由表 5 可以看出本文方法的整体准确率均高于其他 3 种方法, 且高、中、低三个病理等级分类准确率相差不大, 其他的三种方法在分类时均具有一定的偏向性。所以实验结果充分表明了本文方法的优越性。

4 结语

本文提出一种基于乳腺病理组织分析的自动分级方法。该方法能够自动检测病理图像中细胞位置并拟合细胞轮廓, 提取出包括细胞的形状特征、纹理特征和组织空间结构特征共计 203 维的特征来训练支持向量机分类器进行病理自动分级。病理图像自动分级过程中的每一步: 细胞检测、轮廓分割、特征描述均是具有挑战性的研究领域, 任何一个步骤都会对病理分级的最终准确率造成影响。由于目前样本数据库数量较少, 本文未能进行多个数据库的实验, 从而验证该方法的鲁棒性。本文针对乳腺癌病理图像自动分级的初步探索, 仍然存在一些不足: 虽然检测了淋巴细胞, 但未能将淋巴细胞的相关特征运用到自动分级中; 细胞轮廓拟合准确率还有待提高。

参考文献:

- [1] American Cancer Society. Cancer Facts & Figures 2015 [R]. Atlanta: American Cancer Society, 2015: 4-6.
- [2] ROBBINS P, PINDER S, de KLERK N, *et al.* Histological grading of breast carcinomas: a study of interobserver agreement [J]. Human pathology, 1995, 26(8): 873-879.
- [3] DALTON L W, PINDER S E, ELSTON C E, *et al.* Histologic grading of breast cancer: linkage of patient outcome with level of pathologist agreement [J]. Modern Pathology, 2000, 13(7): 730-735.
- [4] XIANG L, XU J. Nuclei detection of breast histopathology based on HOG feature and sliding window [J]. Journal of Shandong University: Engineering Science, 2015, 45(1): 37-44. (项磊, 徐军. 基于 HOG 特征和滑动窗口的乳腺病理图像细胞检测 [J]. 山东大学学报: 工学版, 2015, 45(1): 37-44.)
- [5] MAY M. A better lens on disease [J]. Scientific American, 2010, 302(5): 74-77.
- [6] BOURZAC K. Software: the computer will see you now [J]. Nature, 2013, 502(7473): S92-S94.
- [7] CHEN J, QU A, WANG L, *et al.* New breast cancer prognostic factors identified by computer-aided image analysis of HE stained histopathology images [J]. Scientific Reports, 2015(5): 10690.
- [8] WANG L, QU A, YUAN J, *et al.* Relationship between mathematical parameters of breast invasive ductal carcinoma nests and prognosis of patients based on tumor image processing [J]. Acta Biophysica Sinica, 2013, 29(5): 360-369. (王林伟, 屈爱平, 袁静萍, 等. 图像处理与分析技术研究乳腺浸润性导管癌巢数理参数与患者预后的关系 [J]. 生物物理学报, 2013, 29(5): 360-369.)
- [9] PETUSHI S, GARCIA F U, HABER M M, *et al.* Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer [J]. BMC Medical Imaging, 2006, 6(1): 14.
- [10] HALL B H, IANOSI-IRIMIE M, JAVIDIAN P, *et al.* Computer-assisted assessment of the human epidermal growth factor receptor 2 immunohistochemical assay in imaged histologic sections using a membrane isolation algorithm and quantitative analysis of positive controls [J]. BMC Medical Imaging, 2008, 8(1): 11.
- [11] BASAVANHALLY A N, GANESAN S, AGNER S, *et al.* Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology [J]. IEEE Transactions on Biomedical Engineering, 2010, 57(3): 642-653.
- [12] Di CATALDO S, FICARRA E, ACQUAVIVA A, *et al.* Automated segmentation of tissue images for computerized IHC analysis [J]. Computer Methods and Programs in Biomedicine, 2010, 100(1): 1-15.
- [13] Di CATALDO S, FICARRA E, ACQUAVIVA A, *et al.* Achieving the way for automated segmentation of nuclei in cancer tissue images through morphology-based approach: a quantitative evaluation [J]. Computerized Medical Imaging and Graphics, 2010, 34(6): 453-461.
- [14] XU J, XIANG L, LIU Q, *et al.* Stacked Sparse AutoEncoder (SSAE) for nuclei detection on breast cancer histopathology images [J]. IEEE Transactions on Medical Imaging, 2015, PP(99): 1.
- [15] CRUZ-ROA A, XU J, MADABHUSHI A. A note on the stability and discriminability of graph-based features for classification problems in digital pathology [C]// Proceedings of the 10th International Symposium on Medical Information Processing and Analysis, SPIE 9287. Bellingham: SPIE, 2015: 928703.
- [16] LOMÉNE N, RACOCEANU D. Point set morphological filtering and semantic spatial configuration modeling: Application to microscopic image and bio-structure analysis [J]. Pattern Recognition, 2012, 45(8): 2894-2911.

(下转第 3580 页)

息、大气光强、偏振态等信息进行优化,对薄雾天气条件下的图像能够很好地复原,但是对于离视点较远的浓雾天气下的图像,其去雾效果不理想。基于先验原理的光学去雾算法能较好地去除浓雾天气下的雾霾,但是光学成像系统在浓雾条件下的穿透能力差,对远处场景的识别效果不够好。因此本文提出了一种颜色空间转化的思想把偏振信息转化到彩色空间,并对转化后的三通道彩色图像用暗原色先验原理去除雾霾。实验结果表明该方法在浓雾天气条件下能取得良好的去雾效果,尤其能很好地识别离视点较远场景中的物体。但是对于薄雾条件下场景的细节增强效果不够理想,我们下一步的研究重点是优化大气光强、偏振态等信息,做到既能识别远程场景中的模糊物体又能对近处的景物细节进行增强。

参考文献:

- [1] HE K, SUN J, TANG X. Single image haze removal using dark channel prior [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(12): 2341 – 2353.
- [2] TAREL J-P, HAUTIERE N. Fast visibility restoration from a single color or gray level image [C]// *Proceedings of the 2009 IEEE 12th Conference on Computer Vision*. Piscataway: IEEE, 2009: 2201 – 2208.
- [3] PEI S-P, LEE T-Y. Effective image haze removal using dark channel prior and post-processing [C]// *Proceedings of the 2012 IEEE International Symposium on Circuits and Systems*. Piscataway: IEEE, 2012: 2777 – 2780.
- [4] NARASIMHAN S, NAYAR S. Contrast restoration of weather degraded images [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(6): 713 – 724.
- [5] NARASIMHAN S G, NAYAR S K. Chromatic framework for vision in bad weather [C]// *Proceedings of the 2000 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2000: 598 – 605.
- [6] DUAN J, FU Q, MO C, *et al.* Review of polarization imaging technology for international military application I [J]. *Infrared Technology*, 2014, 36(3): 190 – 195. (段锦,付强,莫春和,等.国外偏振成像军事应用的研究进展(上) [J]. *红外技术*, 2014, 36(3): 190 – 195.)
- [7] SCHECHNER Y, NARASIMHAN S, NAYAR S. Polarization based vision through haze [J]. *Applied Optics*, 2003, 42(3): 511 – 525.
- [8] SHWARTZ S, NAMER E, SCHECHNER Y. Blind haze separation [C]// *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2006: 1984 – 1991.
- [9] DAISUKE M, DAISUKE A, MASASHI B, *et al.* Polarization-based dehazing using two reference objects [C]// *Proceedings of the 2013 IEEE International Conference on Computer Vision Workshops*. Piscataway: IEEE, 2013: 852 – 859.
- [10] ZHOU P, XUE M, ZHANG H, *et al.* Automatic image dehaze using polarization filtering [J]. *Journal of Image and Graphics*, 2011, 16(7): 1183 – 1187. (周蒲城,薛默根,张洪坤,等.利用偏振滤波的自动图像去雾[J]. *中国图象图形学报*, 2011, 16(7): 1183 – 1187.)
- [11] ZHANG D, ZHANG J, WU X, *et al.* Fusion of polarization image based on Directionlets transform [J]. *Journal of Electronics and Information Technology*, 2011, 33(12): 2795 – 2800. (张德祥,张晶晶,吴小培,等.基于 Directionlets 变换的偏振图像融合[J]. *电子与信息学报*, 2011, 33(12): 2795 – 2800.)
- [12] YE S, TANG W, SUN X, *et al.* Characterization of the polarized remote sensing images using IHS color system [J]. *Theoretical Study*, 2006, 84(2): 11 – 14. (叶松,唐伟平,孙晓兵,等.一种采用 IHS 空间表征偏振遥感图像的方法[J]. *理论研究*, 2006, 84(2): 11 – 14.)
- [13] HE K, SUN J, TANG X. Fast matting using large kernel matting Laplacian matrices [C]// *Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2010: 2165 – 2172.
- [14] LEVIN A, LISCHINSKI D, WEISS Y. A closed-form solution to natural image matting [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(2): 228 – 242.
- [15] WANG J, ZHANG Y, WANG Z, *et al.* Single image defogging algorithm based on HIS color space [J]. *Journal of Computer Applications*, 2014, 34(10): 2990 – 2995. (王新建,张有会,王志巍,等.基于 HIS 颜色空间的单幅图像去雾算法[J]. *计算机应用*, 2014, 34(10): 2990 – 2995.)
- [17] DOYLE S, AGNER S, MADABHUSHI A, *et al.* Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features [C]// *ISBI 2008: Proceedings of the 5th IEEE International Symposium on Biomedical Imaging: from Nano to Macro*. Piscataway: IEEE, 2008: 496 – 499.
- [18] BASAVANHALLY A, GANESAN S, FELDMAN M D, *et al.* Multi-field-of-view framework for distinguishing tumor grade in ER + breast cancer from entire histopathology slides [J]. *IEEE Transactions on Biomedical Engineering*, 2013, 60(8): 2089 – 2099.
- [19] DOYLE S, HWANG M, SHAH K, *et al.* Automated grading of prostate cancer using architectural and textural image features [C]// *ISBI 2007: Proceedings of the 4th IEEE International Symposium on Biomedical Imaging: from Nano to Macro*. Piscataway: IEEE, 2007: 1284 – 1287.
- [20] XU J, XIANG L, WANG G, *et al.* Sparse Non-negative Matrix Factorization (SNMF) based color unmixing for breast histopathological image analysis [EB/OL]. [2015-04-02]. <http://dx.doi.org/10.1016/j.compmedimag>.
- [21] KHAN A M, RAJPOOT N, TREANOR D, *et al.* A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution [J]. *IEEE Transactions on Biomedical Engineering*, 2014, 61(6): 1729 – 1738.
- [22] BARINOVA O, LEMPITSKY V, KHOLI P. On detection of multiple object instances using Hough transforms [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(9): 1773 – 1784.
- [23] WANG W, OZOLEK J A, ROHDE G K. Detection and classification of thyroid follicular lesions based on nuclear structure from histopathology images [J]. *Cytometry Part A*, 2010, 77(5): 485 – 494.
- [24] BASAVANHALLY A, XU J, MADABHUSHI A, *et al.* Computer-aided prognosis of ER + breast cancer histopathology, and correlating survival outcome with oncotype DX assay [C]// *ISBI09: Proceedings of the 6th IEEE International Symposium on Biomedical Imaging: from Nano to Macro*. Piscataway: IEEE, 2009: 851 – 854.

(上接第3575页)