

基于深度学习的腺体病理图像分割

蔡亚洁, 李畅, 杜悦, 黄道斌

(皖南医学院, 安徽 芜湖 241000)

摘要:腺体病变引起的疾病如结肠腺癌、乳腺癌等的发病率逐年增高,病理检查是临床诊断的“金标准”,从病理图像中准确分割病灶范围对疾病的诊疗至关重要,然而这是一项费时费力的工作,同时与病理医生的水平与经验有关。近年来,计算机辅助诊断系统和深度学习(Deep learning)在医学图像处理领域快速发展并得到广泛应用,为进一步减轻医生的工作负担,采用经典神经网络对腺体病理图像进行区域分割,并使其能够适用于更加广泛的腺体分割,在腺体病理图像中取得较好的分割效果,为辅助早期诊断及减小误诊概率提供可能。

关键词:腺体;病理图像;计算机辅助诊断;深度学习;卷积神经网络

中图分类号:TP183 文献标识码:A

文章编号:1009-3044(2021)23-0089-03

开放科学(资源服务)标识码(OSID):



Glandular Pathological Image Segmentation Based on Deep Learning

CAI Ya-jie, LI Chang, DU Yue, HUANG Dao-bin

(Wannan Medical College, Wuhu 241000, China)

Abstract: The incidence of diseases caused by glandular lesions such as colon adenocarcinoma and breast cancer is increasing year by year. Pathological examination is the 'gold standard' for clinical diagnosis. Accurate segmentation of lesion range from pathological images is essential for the diagnosis and treatment of diseases. However, this is a time-consuming and laborious work, which is related to the level and experience of pathologists. In recent years, computer aided diagnosis system and deep learning (Deep learning) has been developed rapidly and widely used in the field of medical image processing. In order to further reduce the heavy work of doctors, the classical neural network is used to segment the gland pathological images, and it can be applied to more extensive gland segmentation. Effective segmentation results are achieved in the gland pathological images, which provide the possibility for early diagnosis and reducing the probability of misdiagnosis.

Key words: glands; pathological images; computer aided diagnosis; deep learning; convolutional neural network

1 引言

腺体包含管腔、细胞质和上皮层,是病理图像中的关键组织结构,H&E染色的病理切片中,部分呈椭圆形,由黑色细胞核环绕腺体基质以及透明空腔构成。传统上,病理学家常采用手动分割腺体,然而这种方法主要通过人眼观察显微镜系统下的组织进行判识,其处理质量与可靠程度可能会受到不同医生的影响,不仅效率较低,主观因素引起的误差大^[1]。在恶性情况下,腺体的结构被破坏,尽管良性病例通常具有圆形结构,但恶

性腺却显示出不规则的形状,因此,与良性结构相比,恶性病例的自动分割具有挑战性。

乳腺癌是来自乳腺终末导管小叶单位的上皮性恶性肿瘤,自从20世纪70年代末,乳腺癌的发病率缓慢上升,并且有年轻化的趋势,我国每年女性乳腺癌发病人数达16.9万,占全球总发病数的12.25%,仅次于美国。由中国肿瘤登记中心统计,乳腺癌是城市女性最常见的癌症,是农村女性第四大常见癌症^[2-3]。胃癌在我国各种恶性肿瘤中发病率第一^[4],胃腺癌的发生率占胃恶性肿瘤的95%。前列腺癌在男性常见的恶性肿瘤

收稿日期:2021-04-20

基金项目:安徽省大学生创新创业训练项目(S201910368041);安徽省大学生创新创业训练项目(S201910368121);皖南医学院校级精品开放课程“医学数据挖掘”(2018KFKC08);皖南医学院校级教学研究项目“大数据背景下医学数据挖掘课程实践教学研究”(2018JYXM10);安徽省质量工程教学研究项目“以学科竞赛为驱动的应用型医学信息人才培养研究”(2019JYXM0260)

作者简介:蔡亚洁(1999—),女,安徽合肥人,本科在读,研究方向主要为人工智能与医学影像的结合研究,包括医学图像的分类、分割、检测等;李畅(2000—),男,安徽淮南人,本科在读,研究方向主要为人工智能与医学影像的结合研究;杜悦(1999—),女,安徽阜阳人,研究方向主要为人工智能与医学影像的结合研究;黄道斌(1981—),男,讲师,博士在读,研究方向为人工智能相关领域。

中排名第二,仅次于肺癌^[5],我国前列腺癌呈明显高发趋势。

现在对腺体病变的诊断主要依赖病理、超声、钼靶、核磁共振等,通常先使用钼靶进行筛查,若无法定性,则进行超声和核磁共振检查,出现疑似情况,将进行病理穿刺检查,并对是否癌变、恶性程度以及分子分型做出判断^[6]。组织病理学检查作为腺体病变诊断的“金标准”,对于早期诊断至关重要。然而这项工作并不是十分容易,必须面对腺体在形状、大小、位置、纹理等方面的多变性带来的影响,因此,探寻合适的腺体病理图像处理办法,对于降低死亡率、减轻医生负担以及提高患者生活质量等方面具有重要意义。

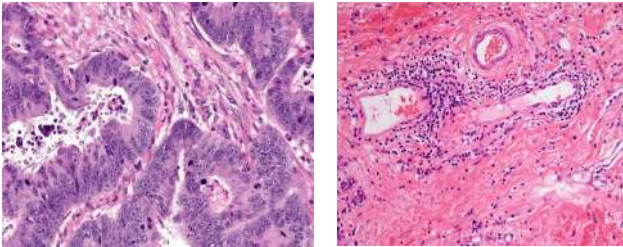


图 1 腺体病理图像

近些年随着技术的进步,数字化病理系统的出现使得病理资源数字化、网络化。医生可以通过电脑完成阅片工作,同时便于更好地分析与交流,打破时间和空间的限制,数字化病理为疾病的精准医疗以及计算机辅助诊断提供了良好的基础。

2 方法

2.1 深度学习与卷积神经网络

深度学习的概念由 Hinton 等人于 2006 年首次提出^[7],是机器学习算法中的一项新兴技术,通过构建多层网络,以及高层次特征表示数据的抽象语义信息,获得良好的特征鲁棒性。传统的医学图像分割算法包含区域的分割方法、边缘的分割方法、图论的分割方法和泛函的图像分割方法等^[8]。由于计算机视觉相关领域发展迅速,医学图像分割与深度学习的结合性研究取得显著的成果^[9],深度学习正在成为通用成像和计算机视觉领域的领先机器学习工具^[10]。然而,多层感知器引入了更大的参数量,导致神经网络难以正常完成训练,为解决此问题,Krizhevsky 等人^[11]提出了卷积神经网络(Convolutional Neural Networks, CNN),采用局部连接和权值共享的方式进行网络输入,避免传统算法中繁琐的特征提取和数据重建过程^[12-13],减少了权值的数量使得网络易于优化、模型更加简便,同时也减少了过拟合发生的几率,使其在医学图像处理过程中有很大的优势^[14]。

2.2 U-net

研究人员在卷积神经网络的基础上进一步探寻更加优化的分割网络。2015 年加州大学的 Long Jonathan 等人^[15]提出全卷积网络(Fully Convolutional Networks, FCN),能够进行图像的像素级分类,同年,Olaf Ronneberger 等人吸取 FCN 的优点,提出 U-net 神经网络结构^[16]。该网络呈现为 U 型结构,将下采样层与上采样层相连,使得下采样层提取到的特征可以直接传递至上采样层。U-net 在医学图像分割中有良好的表现,一方面由于直接编码端的特征于解码端,便于目标细节重建^[17],另一方面由于改良的 U-net 将低分辨率信息和高分辨率信息相结合,不仅提供了物体识别的依据,也提供了精准分割定位的依据。

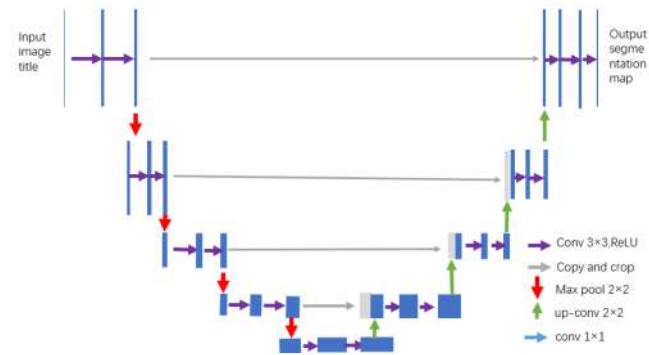


图 2 U-net 网络结构简图

3 实验

3.1 数据采集和预处理

训练数据的数量与质量水平,对于深度学习中网络模型有很大影响。数据集 Warwick_qu_dataset 是来自 2015 年 MICCAI 腺体分割挑战(GLAS)的公开数据集,训练数据集由 85 个(良性/恶性= 37/48)的 20*(0.62005 微米/像素)的 BMP 图像,经过苏木精-伊红染色的病理切片,由病理学专家提供真实注释。测试数据包含两部分:轮廓评估的 A 部分(60 幅图像)和部分 B(20 幅图像)的现场评估。神经网络需要大量样本进行训练才能达到较好的泛化,丰富数据集样本库,获得不同情况、不同形态、不同个体尽可能多的数据,对于研究有着重要意义。为了提高模型性能,对原始数据集进行以下预处理:

- 1)为防止过拟合,对原始图像进行旋转、翻转、随机裁剪来扩充增强数据集;
- 2)病理图像的采集和制作过程中造成图像的差异,采用颜色标准化的方法,将所有图片都转化成标准图片颜色(RGB)分布,进行病理图像标准化。



图 3 总体框架图

3.2 训练

3.2.1 参数设置

整体训练过程是在 Ubuntu16.04 操作系统上,使用 Nvidia GeForce 1080Ti GPU 计算卡进行。在 Tensorflow 后台的 Keras 环境完成 U-net 网络模型的构建,选用优化效率最高的自适应 Adam 优化器,初始学习率设置默认,送入训练的图片批次量(batch-size)为 12, epoch 为 40,分割训练的过程中选择 Dice Loss 损失函数。

3.2.2 评价指标

Dice 系数是一种集合相似性度量函数,通常用于计算两个样本的相似度,是评价分割结果好坏的常用指标,取值范围在 0 到 1 之间,DiceLoss 是医学图像分割中的常用损失函数。本实验中 X 代表医生标注的腺体区域,Y 代表模型分割的腺体区域。

$$DICE = \frac{2|X \cap Y|}{|X| + |Y|}$$

$$DICE\ LOSS = 1 - \frac{2|X \cap Y|}{|X| + |Y|}$$

4 结果

在腺体病理图像数据集上进行分割模型训练操作,训练过程中,若学习率设置过小,则收敛过程缓慢;若学习率过大,会导致梯度在最小值附近来回震荡,难以进行收敛。根据多次实验观察以及调试经验,将学习率 lr 设为0.00001较合适。绘制loss曲线图可看出曲线存在一定范围的上下摆动,最终train loss以及test loss都趋于平稳,说明学习过程遇到瓶颈,训练存在细节问题,可逐步减小学习率或批量数目,进一步进行模型优化。

由分割效果图可看出,较为正常腺体的分割效果较好,能够完成相关区域与其他区域分割的任务,并且与医生标注图像对比差距较小,分割后的腺体形态较完整,边缘较为光滑清晰,而病变程度大的腺体分割准确度相对较低,存在组织之间相互牵连、分割差异较大等问题。

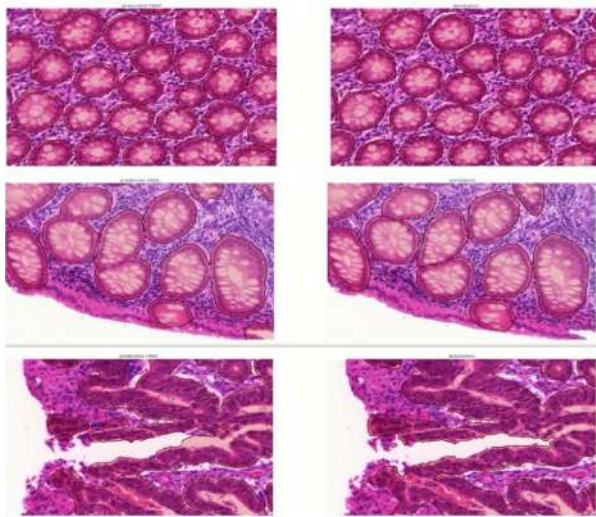


图4 分割效果图(左为腺体分割区域,右为医生标注区域)

目前的研究成果尚有不足和提升空间,分割过程中遇到的困难和难点会影响最终准确率,具体干扰因素如下:

- 1)图像的采集和特征提取存在个体差异;
- 2)切片的组织制备过程可能导致组织形变、伪影和组织外观不一致;
- 3)切片染色过程中可能引起染色的变化,染色前后差异对鲁棒性的要求更高,对神经网络模型的构建造成更大的困难;
- 4)在中、低分化腺癌等恶性肿瘤中,腺体正常结构退化明显;
- 5)腺体肿块分割易受边缘及周围腺体干扰,造成相互牵连等问题;
- 6)U-net用于腺体病理图像分割可能存在模型训练过度或不充分的情况,容易造成误差等。

5 结论

伴随着时代的进步,医学图像逐渐成为人工智能的热门研究领域之一,深度学习在医学图像处理上取得了不俗的成绩。文章以腺体病理图像作为研究对象,从经典深度学习网络入手,能够大致完成目标区域分割,尤其对于较为正常的腺体有较好的分割效果,对于病变程度较大的腺体还存在相关问题,

这也是未来的研究工作中需要进行优化的地方,弥补其中的不足之处,使之能够真正减轻临床的工作负担。

参考文献:

- [1] 宁旭,罗晓晖,唐建民.病理切片图像分割技术的研究[J].中国医学物理学杂志,2005,22(5):648-650.
- [2] WANG S, LIU J, YAO-YAO B I, et al. Automatic Recognition of Breast Gland Based on Two-step Clustering and Random Forest [J]. Computer Science, 2018,
- [3] 郑莹,吴春晓,张敏璐.乳腺癌在中国的流行状况和疾病特征[J].中国癌症杂志,2013,23(8):561-569.
- [4] 田雅婷.中国胃癌发病率居各类肿瘤之首[J].老年健康,2011(3):5.
- [5] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA: A Cancer Journal for Clinicians, 2018, 68(6):394-424.
- [6] Robertson S, Azizpour H, Smith K, et al. Digital image analysis in breast pathology—from image processing techniques to artificial intelligence[J]. Translational Research, 2018, 194:19-35.
- [7] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553):436-444.
- [8] 林瑶,田捷.医学图像分割方法综述[J].模式识别与人工智能, 2002, 15(2):192-204.
- [9] Xu Y, Li Y, Wang Y P, et al. Gland instance segmentation using deep multichannel neural networks[J]. IEEE Transactions on Biomedical Engineering, 2017, 64(12):2901-2912.
- [10] Schmidhuber J. Deep learning in neural networks: an overview [J]. Neural Networks, 2015, 61:85-117.
- [11] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6):84-90.
- [12] Kim Y. Convolutional neural networks for sentence classification[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar. Stroudsburg, PA, USA: Association for Computational Linguistics, 2014.
- [13] 庞丝丝,黄呈铖.基于卷积神经网络的图像分类研究[J].现代计算机,2019(23):40-44.
- [14] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6):84-90.
- [15] 王帅,刘娟,毕姚姚,等.基于两步聚类 and 随机森林的乳腺腺管自动识别方法[J].计算机科学,2018,45(3):249-254.
- [16] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015:234-241.
- [17] 柴志忠.基于深度卷积神经网络的病理影像研究[D].厦门:厦门大学,2019.

【通联编辑:唐一东】