

改进型深度学习模型在乳腺肿瘤良恶性鉴别中的应用

邓竹琴¹, 俞永伟²

1. 中国人民解放军联勤保障部队第901医院妇产科, 安徽 合肥 230031; 2. 安徽省合肥市长荣医院普外科, 安徽 合肥 230001

【摘要】目的:解决传统方法在临床中对病理图像检测不足以及人工判断导致的错误判断等问题。**方法:**使用乳腺肿瘤细胞数据集,首先对数据集进行数据增强,增强后数据集为原来的2倍,将增强后数据集输入到本文提出的模型中进行训练。**结果:**经过100次迭代,训练集的准确率为97.44%,在测试集中准确率为96.4%,召回率为95.5%,与同类型文献相比都有明显提高。**结论:**本文章提出的改进型卷积神经网络具有收敛快,泛化好等特点。可以对乳腺肿瘤细胞的良恶性进行有效的辨识分类。

【关键词】卷积神经网络;乳腺癌细胞识别;图像识别

【中图分类号】R318;R377.9

【文献标志码】A

【文章编号】1005-202X(2020)11-1469-05

Application of improved deep learning model in differential diagnosis of benign and malignant breast tumors

DENG Zhuqin¹, YU Yongwei²

1. Department of Obstetrics and Gynecology, The People's Liberation Army Joint Service Support Unit No.901 Hospital, Hefei 230031, China; 2. Department of General Surgery, Changrong Hospital, Hefei 230001, China

Abstract: Objective To improve the defects and deficiency of traditional methods in clinical pathological image detection and to solve the problem of misjudgment made by human. **Methods** The data sets of breast tumor cells are from clinical date. The data sets were enhanced two fold at first and put into the proposed model for training. **Results** After 100 times of iterations, the accuracy of validation set arrives 97.44%, the test set 96.4%, and the recall rate 95.5%. They are obviously improved compared with the same type of literature. **Conclusion** The improved convolutional neural network proposed in this paper has advantages of rapid convergence and excellent generalization ability. It can identify and classify the benign and malignant breast tumor cells effectively.

Keywords: Convolutional neural network; breast tumor cell recognition; image recognition

前言

乳腺癌是临床上常见的癌症,具有发病率高、死亡率高和治愈率低等特点。据不完全数据表明,全世界乳腺癌的发病率在稳步上升,其中美国女性乳腺癌的发病率更是高达12.5%^[1-2]。虽然我国乳腺癌的发病率较低,但最近几年我国女性患乳腺癌的数量在逐年上升,成为近年来发病率最高的恶性肿瘤之一^[3]。临床中医生根据彩超、X线、核磁共振等对患者进行诊断,但病理图像具有更丰富的形态信息,

是医生确诊的重要方式。目前病理医生进行病理图像分类时主要依靠图像中细胞形状和分布进行分类。但诊断结果容易受病理医生经验和学识的影响。近年来,深度学习的兴起,使其在医学图像处理方面大放异彩^[4-5]。在乳腺癌组织病理图像分类中深度学习分类方法与传统分类方法相比准确率提高了6%^[6]。虽然识别准确率提高了6%,但识别精度依然不足。针对精度不足的情况,本文通过对 Visual Geometry Group-16(VGG-16)卷积神经网络的模型进行改进,提出基于VGG-16卷积神经网络的改进模型(VGG-Improve卷积神经网络模型)。该方法可以解决VGG-16网络过深,参数过多,收敛速度慢,训练困难等问题。另一方面采用数据增强的方法提高了模型的泛化能力和鲁棒性。

【收稿日期】2020-07-15

【作者简介】邓竹琴,妇产科主治医师,E-mail: 951340480@qq.com

【通信作者】俞永伟,普外科主治医师,E-mail: 278718482@qq.com

1 方法

1.1 VGG-Improve卷积神经网络模型

VGG-Improve卷积神经网络模型由7个卷积层和2个全连接层组成;用ReLU函数作为激活函数,在加强网络非线性映射能力的同时,也可以提高网络收敛速度^[7-8];使用正则化函数,通过对损失函数增加惩罚项,降低过拟合的风险^[9];用均值池化层,减少图像细节丢失,避免了最大池化丢失局部细节。

本文使用VGG-Improve卷积神经网络模型对乳腺癌病理图片进行训练,并与同类型文献提出的卷积神经网络进行对比。VGG-Improve卷积神经网络

模型的结构图如图1所示^[10-11]。第1层为图像输入,输入的尺寸为224×224;第2层为64×3×3的卷积核,第3层为64×3×3的卷积核;第4层为均值池化层,池化域为2×2;第5~7层与第2~4层相同;第8~10层由3个64×3×3的卷积核组成;第11层为均值池化层,池化域为2×2,每一个卷积核都使用正则化;第12~13层为两个全连接层,连接参数分别为512和256,每个全连接层后面都进行正则化。最后一层使用Softmax分类器,其中卷积层和全连接层均使用ReLU激活函数,卷积层的步长为1×1,池化层步长为2×2,全连接层的步长为2×2。

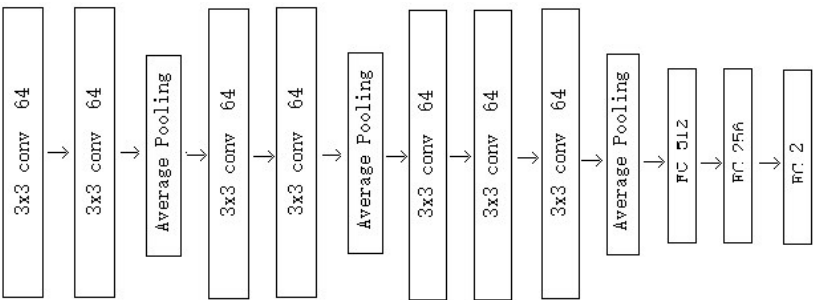


图1 改进型卷积神经网络模型
Fig.1 Improved convolutional neural network model

1.2 正则化

训练神经网络的过程中,由于数据集不足,模型选取不当等原因,容易造成过拟合。使用正则化能有效防止过拟合,使训练好的模型在训练集和测试集上都有很好的准确率。常见的正则化分为L1正则化和L2正则化。L1正则化是将权值向量 W 中各个元素的绝对值求和。L1正则化可以改变权值矩阵 W 的稀疏性,将 W 中部分元素变为零。通过改变 W 的稀疏性减少了计算量,在一定程度上可以减少模型过拟合的概率。L2正则化是将权值向量 W 中各个元素进行平方和再求平方根。使得模型的解偏向于范数较小的 W ,通过限制 W 范数的大小实现了对模型空间的限制,从而在一定程度上避免了过拟合。但L2正则化不具备稀疏化的特性,计算量并没有得到改观,但相较于L1正则化L2正则化对过拟合的抑制效果更加明显。本文采用L2正则化防止出现过拟合情况。L1正则化和L2正则化公式定义如式(1)和式(2)所示:

$$\text{loss}_{\text{new}} = \text{loss} + C\|w\|_1 \tag{1}$$

$$\text{loss}_{\text{new}} = \text{loss} + C\|w\|_2^2 \tag{2}$$

其中,loss为原始的损失损失值,C为惩罚系数,w为权重参数,loss_{new}为新生成的损失值。

1.3 数据增强

由于本研究临床数据集数据较少,无法提供大量带有标记的样本,容易在训练中造成过拟合现象。针对上述问题,本文对原有的样本进行数据增强,数据增强的方式如下:(1)旋转,将图像在0~60°范围内随机旋转。(2)图像缩放,将图像按照一定比例进行放大或缩小。(3)平移,将图像向x或y方向移动。(4)增加噪声,通过加入噪声数据抵消高频特征,抑制过拟合发生。经过数据增强后,不但增加了训练样本,还会增加模型的鲁棒性。另外本文还对图像进行归一化处理,处理后的数据能防止出现“梯度弥散”现象^[12-13]。通过数据增强,使样本扩大为原来的2倍。

2 实验

2.1 数据集的建立与实验平台

选取某医院肿瘤科采集的临床乳腺肿瘤细胞数据集作为研究对象,其中女性病例占83%,男性病例占17%,女性年龄为22~74岁,平均年龄为45岁,其中恶性乳腺肿瘤数据为2 170张,良性乳腺肿瘤数据为1 211张,乳腺肿瘤细胞图像见图2。乳腺肿瘤细胞数据集的分布情况如表1所示。数据增强后数据集扩充为原来的二倍,增强后数据集分布情况如表2

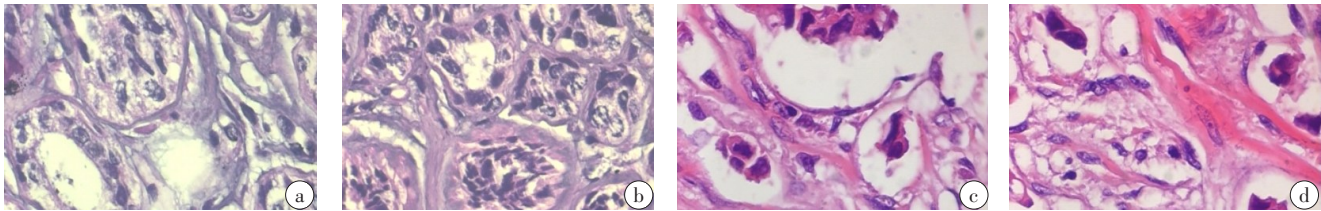


图2 乳腺肿瘤细胞

Fig.2 Breast tumor cells (a, b are benign tumors, c, d are malignant tumor cells)

图 2a 和图 2b 为良性肿瘤细胞,图 2c 和图 2d 为恶性肿瘤细胞

表 1 增强前乳腺肿瘤细胞数据集分布

Tab.1 Breast tumor cell dataset before enhancement		
类型	良性样本数	恶性样本数
训练集	1 090	2 010
测试集	121	160

表 2 增强后乳腺肿瘤细胞数据集分布

Tab.2 Breast tumor cell dataset after enhancement		
类型	良性样本数	恶性样本数
训练集	2 020	4 000
测试集	400	400

所示。本文训练和验证过程所使用的硬件平台配置为 i5-9600k/32G RAM/2T SSD/GPU GeForce RTX 2070 Ti,操作系统为 Windows 10(64bit)。本实验选取 keras 为实验框架。

训练模型时采用 Adam 优化器,训练批次为 32,学习率初始值为 4×10^{-4} ,学习率根据迭代次数动态调整。动态调整学习率公式如下:

$$Lr_{new} = Lr \times (0.95^{epoch})$$

(3)

其中,Lr为学习率,epoch为迭代次数,Lr_{new}为新生成学习率。

2.2 实验评价指标

为了分析实验数据,本文使用召回率(Recall)和准确率(Acc)对实验数据进行评价。计算定义如下:

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

(4)

$$Acc = \frac{TP + TN}{TP + FP + TP + FN} \times 100\%$$

(5)

其中,TP表示将良性肿瘤数据分类到良性类别,TN表示将恶性肿瘤分类到恶性类别,TP和TN均表示分类准确;FP表示将恶性肿瘤数据分类到良性类别,FN表示将良性肿瘤数据分类到恶性类别,FP和FN表示分类错误。召回率表示样本中正例被正确预测的比例;准确率表示判定正确的数据在总体数据中的比例。通过上

述两个指标可以反映模型的分类能力,数值越大,分类能力越强。VGG-Improve 模型评价结果表 3 所示。

表 3 同类文献网络的评价结果(%)

Tab.3 Evaluation results of similar literature networks (%)		
类型	召回率	准确率
VGG-Improve	95.50	96.40
文献[14]	91.25	93.01
文献[13]	87.25	89.87

2.3 模型改进前后的比较

本文比较了模型改进前后对肿瘤数据集分类结果的准确率和损失值。其中 VGG-Improve 卷积神经网络模型在测试集上的准确率高于 VGG-16 卷积神经网络模型,损失值小于 VGG-16 卷积神经网络模型。可以看出 VGG-Improve 相比于 VGG-16 在乳腺癌识别中有更好的表现。表 4 中为模型改进前后训练及测试时各项参数对比。由表 4 可知改进后模型测试集上的错误率和损失值均低于改进前模型。可以看出改进后模型解决了过拟合情况,并且正确率达到 96.4%,相较于为改进前拥有更低的错误率。

表 4 VGG-16 改进前后各项参数对比(性能最佳的 Epoch)

Tab.4 Comparison of various parameters before and after VGG-16 improvement (Best Epoch)				
类型	训练损失值	训练准确率	测试损失值	测试准确率
VGG-16	0.077 7	98.10%	0.453 1	84.38%
VGG-Improve	0.062 9	97.44%	0.094 5	96.40%

2.4 与同类文献对比

文献[13]中卷积神经网络模型由7个卷积层和2个全连接层组成,其中第二、三层卷积层采用两个卷积并联的方式。这种做法能够增加网络模型宽度,实现多尺度的图像特征提取,使用该模型可以有效地对宫颈癌细胞进行分类。文献[14]中卷积神经网络由4个卷积层,4个池化层和2个全连接层并联组成,其中池化层选用最大池化。

通过表5可以看出在相同的乳腺肿瘤细胞测试集下,本文提出的方法相较于文献[13]和文献[14]的模型都有很大的提升,正确率明显提升。经上述对比可以看出,本文提出的VGG-Improve卷积神经网络模型相较于同类文献提出的模型在相同数据集

的情况下拥有更好的准确率。通过表3和图3可以看出VGG-Improve卷积神经网络模型准确率和召回率都高于其他两种方法。证明了VGG-Improve卷积神经网络模型具有良好的泛化性。

表5 本文与同类相关文献准确率对比
Tab.5 Comparison of the accuracy between related literatures and the paper

类型	文献[14]	文献[13]	VGG-Improve
训练数据	8 020	8 020	8 020
测试数据	800	800	800
测试准确率	89.87%	93.01%	96.4%

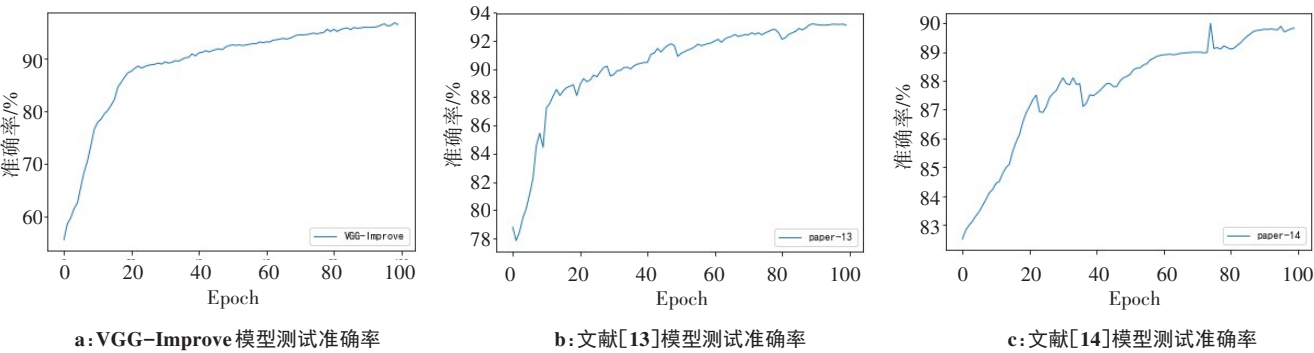


图3 测试集准确率与迭代次数的关系
Fig.3 Relationship between test set accuracy and number of iterations

3 讨论

国外在癌症分类方面,Jiang等^[16]提出一种基于卷积神经网络的新型识别系统,该系统可以对图像进行预处理并对神经祖细胞(NPC)和非NPC进行分类。Pansombut等^[17]使用多种机器学习算法对淋巴细胞进行分类,通过对比得出CNN能更好的对淋巴细胞进行分类。李正义^[18]通过对细胞边缘纹路、曲率、大小等特征的提取,使用一种改进的随机森林分类器对宫颈癌细胞进行识别。在信息化时代的背景下各种机器学习算法愈发重要,人工智能算法在医学诊断中已经广泛应用。本文提出的模型可以辅助医生对乳腺肿瘤进行良恶性分析。未来训练样本量足够时,可以代替医生进行肿瘤细胞良恶性识别。通过人工智能算法代替人工进行医学诊断可以大大缩短医生诊断时间,提高就医效率^[19-20]。现阶段人工神经网络是图像分类中重要的分支,使用人工神经网络搭建针对特定疾病的自动化医学诊断系统是未来发展的趋势。由于本文使用的数据集为乳腺肿瘤细胞,目前在乳腺肿瘤识别中有很好的效果,但尚未

对其他类型肿瘤细胞进行辨别,后续将考虑建立其他类型肿瘤细胞良恶性辨识模型。

【参考文献】

[1] 谢诗哲. 乳腺癌的危害及其防治知识推广[J]. 科技与创新, 2018(11): 6-9.
XIE S Z. The harm of breast cancer and its prevention knowledge promotion[J]. Science and Technology & Innovation, 2018(11): 6-9.

[2] TORRE LA, SIEGEL R L, WARD E M, et al. Global cancer incidence and mortality rates and trends-an update [J]. Cancer Epidemiol Biomarkers Prev, 2016, 25(1): 16-27.

[3] 郭林英, 余腾骅, 张光君, 等. 肿瘤相关成纤维细胞中 GPER 介导 HMGB1 外分泌促进乳腺癌 MCF-7 细胞自噬及增殖[J]. 肿瘤, 2017, 37(5): 448-456.
GUO L Y, YU T Q, ZHANG G J, et al. GPER-mediated HMGB1 exocrine in cancer-associated fibroblasts promotes autophagy and proliferation of breast cancer MCF-7 cells [J]. Tumor, 2017, 37(5): 448-456.

[4] LECUN Y, BENGIO Y, HINTON G, et al. (2015) DeepLearning[J]. Nature, 2015, 521(7556): 436.

[5] 王铭跃, 卢晓蕾, 黄雪兰, 等. 探讨高频超声鉴别乳腺肿瘤良恶性的临床应用价值[J]. 微量元素与健康研究, 2020, 37(1): 7-9.
WANG M Y, LU X L, HUANG X L, et al. Exploring the clinical application value of high frequency ultrasound in differential diagnosis

- of benign and malignant breast tumors[J]. *Studies of Trace Elements and Health*, 2020, 37(1): 7-9.
- [6] BAYRAMOGLU N, KANNALA J, HEIKKILA J. Deep Learning for magnification independent breast cancer histopathology image classification [C]. 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016: 2441-2446.
- [7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]. *Proceedings of the 25th International Conference on Neural Information Processing Systems*. Curran Associates Inc, 2012: 1097-1105.
- [8] 黄江珊, 王秀红. 深度学习在医学图像分析中的应用研究综述[J]. *图书情报研究*, 2019, 12(2): 92-98.
- HUANG J S, WANG X H. A review of the researches on the application of deep learning medical image analysis[J]. *Library & Information Studies*, 2019, 12(2): 92-98.
- [9] PRABHPREET K, GURVINDER S, PARMINDER K. Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deeplearning classification[J]. *Informatics in Medicine Unlocked*, 2019, 16: 100239.
- [10] KUMAR A, SINGH S K, SAXENA S, et al. Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer[J]. *Information Sciences*, 2020, 508: 405-421.
- [11] ZHAN X, ZHANG T, FENG W Y, et al. Breast cancer diagnosis from histopathological image based on deep learning[C]. 2019 31st Chinese Control and Decision Conference. Shenyang: Editorial Department of Control and Decision, 2019: 647-650.
- [12] SPANHOL F A, OLIVEIRA L S, PETITJEAN C, et al. Breastcancer histopathological image classification using Convolutional Neural Networks[C]. *International Joint Conference on Neural Networks*. IEEE, 2016.
- [13] 谢欣, 夏哲雷. 深度卷积神经网络的宫颈癌细胞图像识别[J]. *中国计量大学学报*, 2018(2): 200-203.
- XIE X, XIA Z L. Image recognition of cervical cancer cells by deep convolutional neural network[J]. *Journal of China Jiliang University*, 2018(2): 200-203.
- [14] 陶源, 王佳飞, 杜俊龙, 等. 基于卷积神经网络的细胞识别[J]. *中国医学物理学杂志*, 2017, 34(1): 53-57.
- TAO Y, WANG J F, DU J L, et al. Cell recognition based on convolutional neural network[J]. *Chinese Journal of Medical Physics*, 2017, 34(1): 53-57.
- [15] 尹宝才, 王文通, 王立春. 深度学习研究综述[J]. *北京工业大学学报*, 2015(1): 48-59.
- YIN B C, WANG W T, WANG L C. Review of deep learning research [J]. *Journal of Beijing Polytechnic University*, 2015(1): 48-59.
- [16] JIANG B, WANG X Y, LUO J Z. Convolutional neural networks in automatic recognition of trans-differentiated neural progenitor cells under bright-field microscopy [C]. 2015 5th International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC). IEEE, 2015.
- [17] PANSOMBUT T, WIKAISUKSAKUL S. Convolutional neural networks for recognition of lymphoblast cell images[J]. *Comput Intell Neurosci*, 2019: 7519603.
- [18] 李正义. 基于强特征CNN-SVM的宫颈癌细胞检测[D]. 北京: 北京交通大学, 2018.
- LI Z Y. Detection of cervical cancer cells based on strong feature CNN-SVM [D]. Beijing: Beijing Jiaotong University, 2018.
- [19] 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.
- PANG S S, HUANG C C. Research on image classification based on convolutional neural network[J]. *Modern Computer*, 2019(23): 40-44.
- [20] 胡威, 汪春梅, 张见. 基于卷积神经网络的细胞图像分割与类型判别[J]. *上海师范大学学报(自然科学版)*, 2019, 48(1): 106-112.
- HU W, WANG C M, ZHANG J. Cell image segmentation and type discrimination based on convolutional neural network [J]. *Journal of Shanghai Normal University (Natural Science Edition)*, 2019, 48(1): 106-112.

(编辑: 薛泽玲)