

# 基于超声图像的迁移学习模型在乳腺肿块良恶性鉴别诊断中的价值

余美慧 袁 泉 曾书娥 程 慧 李 楠 叶华容

**摘要** **目的** 探讨基于超声图像的迁移学习模型在乳腺肿块良恶性鉴别诊断中的应用价值。**方法** 收集我院经手术病理证实的 300 例乳腺肿块患者共计 582 张超声图像作为超声数据集(其中训练集 482 张,测试集 100 张)。采用迁移学习方法对经 ImageNet 数据集预训练的 3 种深度卷积神经网络(VGG-16、Inception-v3、ResNet-50)模型进行训练和测试。第 1 次迁移学习为 3 种模型分别对 CBIS-DDSM 数据集中良恶性乳腺肿块的 X 线图像进行识别学习,并对模型进行微调;第 2 次迁移学习为使用乳腺超声数据集中随机挑选的训练集超声图像对 3 种模型进行微调,在测试集中输出最终分类结果。绘制受试者工作特征曲线分析迁移学习后 3 种模型对乳腺良恶性肿块的鉴别诊断效能。**结果** VGG-16、Inception-v3、ResNet-50 3 种模型经过迁移学习后鉴别诊断乳腺肿块良恶性的准确率、敏感性、特异性、精准率、F1 分数、约登指数均有所提高,其中基于 ResNet-50 建立的模型具有最优的诊断效能,准确率 88.0%,敏感性 82.7%,特异性 93.8%,曲线下面积 0.915,均高于其他两种模型,差异均有统计学意义(均  $P < 0.05$ )。**结论** 基于超声图像的迁移学习模型在乳腺肿块良恶性鉴别诊断中具有较高的应用价值,其中基于 ResNet-50 构建的模型效能最佳。

**关键词** 超声检查;深度卷积神经网络;迁移学习;乳腺肿瘤,良恶性

[中图分类号]R445.1;R730.41

[文献标识码]A

## Value of transfer learning model based on ultrasound images in the differential diagnosis of benign and malignant breast masses

YU Meihui, YUAN Quan, ZENG Shue, CHENG Hui, LI Nan, YE Huarong

School of Medicine, Wuhan University of Science and Technology, Wuhan 430081, China

**ABSTRACT** **Objective** To explore the value of transfer learning model based on ultrasound images in the differential diagnosis of benign and malignant breast masses. **Methods** A total of 582 ultrasound images from 300 patients with breast masses confirmed by biopsy in our hospital were collected as an ultrasound data set (482 for training set and 100 for test set). Three deep convolutional neural network models (VGG-16, Inception-v3, ResNet-50) pre-trained on the ImageNet dataset were trained and tested by transfer learning method. For the first transfer learning, three models were used to identify and learn the X-ray images of benign and malignant breast masses in the CBIS-DDSM dataset and fine-tune the models. For the second transfer learning, three models using a randomly selected training set of ultrasound images from the ultrasound dataset were fine-tuned, and the final classification results were output in the test set. Receiver operating characteristic (ROC) curves were drawn to compare the diagnostic efficacy of three models after transfer learning for benign and malignant breast masses. **Results** The accuracy, sensitivity, specificity, precision, F1 score, and Jorden index of VGG-16, Inception-v3 and ResNet-50 models in the differential diagnosis of benign and malignant breast masses were all improved after transfer learning, and the model established based on ResNet-50 had the best efficacy with an accuracy of 88.0%, sensitivity of 82.7%, specificity of 93.8%, and AUC of 0.915, which were higher than those of the other two models, and the differences were statistically significant (all  $P < 0.05$ ).

基金项目:国家癌症中心攀登基金临床研究课题(NCC201917B04);湖北省重点研发计划项目(2020BCB022)

作者单位:430081 武汉市,武汉科技大学医学院(余美慧、程慧、李楠),计算机学院(袁泉);湖北省肿瘤医院超声医学科(曾书娥);武汉科技大学附属华润武钢总医院超声医学科(叶华容)

通讯作者:叶华容, Email: yehuarong@hotmail.com

**Conclusion** The transfer learning model based on ultrasound images has high diagnostic efficacy in the differential diagnosis of benign and malignant breast masses, among which the model established based on ResNet-50 has the best efficacy.

**KEY WORDS** Ultrasonography; Deep convolutional neural network; Transfer learning; Breast tumor, benign and malignant

2020年全球癌症统计数据显示乳腺癌已成为威胁女性健康的首位高发癌症<sup>[1]</sup>,精准诊断对改善患者预后和降低死亡率至关重要<sup>[2]</sup>。超声具有操作简便、经济、无辐射、无创等优势,是目前临床筛查乳腺癌的首选影像学检查方法<sup>[3]</sup>。然而,超声医师的工作经验、专业水平等主观因素,以及超声仪器、诊疗环境等客观因素均可能影响诊断准确性。近年来,深度学习因其在图像识别和人工智能决策中的优异表现而受到广泛关注,其可以自动对复杂的医学图像(如超声图像)特征进行定量评估,提取有诊断价值的信息,并将分析结果作为第二意见或决策支持提供给临床医师<sup>[4]</sup>,具有客观、可重复性等特点。迁移学习和卷积神经网络(convolutional neural networks, CNN)模型是目前医学领域最有效的深度学习技术,已被证实优于传统方法<sup>[5]</sup>。基于此,本研究旨在探讨基于超声图像的迁移学习模型在乳腺肿块良恶性鉴别诊断中的应用价值。

## 资料与方法

### 一、研究对象

1. 患者纳入情况:收集2018年5月至2021年3月我院收治的行超声检查、手术或穿刺活检的乳腺肿块患者,均为女性。纳入标准:①肿块经病理组织学确诊;②恶性肿瘤经病理证实为原发性乳腺癌;③可获得标准的术前超声检查图像;④术前未行相关放疗、化疗或放化疗。排除标准:①超声图像质量差,无原始格式;②术前行相关治疗。最终纳入300例患者,年龄19~79岁,平均(46.77±12.41)岁;其中良性144例(146个肿块),恶性156例(158个肿块)。本研究为回顾性研究,经武汉科技大学附属华润武钢总医院伦理委员会批准,免除了书面知情同意的要求。

2. 乳腺超声数据集情况:300例患者共582张超声图像(良性304张,恶性278张)随机分为训练集和测试集,训练集中良性患者96例(98个肿块,256张图像),恶性患者104例(106个肿块,226张图像);测试集中良性患者48例(48个肿块,48张图像),恶性患者52例(52个肿块,52张图像)。训练阶段最多提取同一肿块的4个不同截面图像和同一患者的2个肿块,随

机分配并使用训练集中200例患者204个肿块图像;测试阶段仅提取测试集同一患者一个肿块的一张超声图像。

3. 乳腺X线摄影筛查数字数据库的更新和标准化版本(CBIS-DDSM)数据集:CBIS-DDSM<sup>[6]</sup>是乳腺X线摄影筛查数字数据库的更新和标准化版本,共包含1590张图像,其中良性肿块图像845张,恶性肿瘤图像745张。将此数据集中70%的样本作为训练集,30%的样本作为测试集,以增强模型在超声数据集上的表现力,并对模型进行微调和优化。

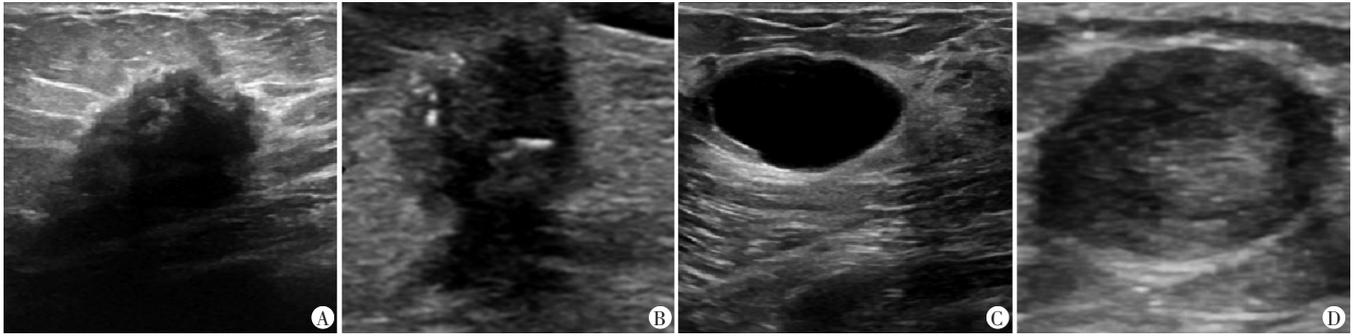
### 二、仪器与方法

1. 乳腺超声图像采集:使用Philips EPIQ 7C彩色多普勒超声诊断仪,线阵探头,频率3~12 MHz。受检者取仰卧位,双臂外展,充分暴露乳房,探头置于乳房表面,于平静呼吸状态下以乳头为中心行放射状全面扫查,二维超声重点扫查病灶区域,根据病灶位置、大小及回声适当调整深度、焦点位置和增益,存储肿块最大径超声图像。

2. 图像处理与数据增强:在完全保留乳腺病变区域的前提下,对图像进行人工裁剪预处理,具体操作如下:①从每张图像中裁剪出矩形视图,删除包含无关内容的区域;②从矩形视图的左、中、右或上、中、下正方形区域裁剪3个块,选择包括肿块在内的最合适的正方形视图;③在不改变纵横比的情况下,重新缩放裁剪后的图片尺寸为256×256(像素),见图1。由于数据集中的训练数据量有限,训练过程中对图像进行翻转、旋转、平移、对比度增强等数据增强操作,人为地将训练图像数据集扩充为其原始数量的4倍。

3. CNN模型选择:本研究选取了3种具有代表性的深度CNN模型,即VGG-16、Inception-v3和ResNet-50。

4. 迁移学习方法:使用ImageNet数据集预训练网络作为特征提取器,即从头开始学习特征,然后将训练模型参数迁移至3种CNN模型以对其进行初始化。第1次迁移学习利用CBIS-DDSM数据集中良恶性乳腺肿块X线图像对3种CNN模型进行训练优化并微调卷积层参数;第2次迁移学习使用乳腺超声数据集中随机挑选的训练集超声图像对3种CNN模型进行二次微调,得到最优的模型参数,并在测试集中输出最终分类结果。CNN模型迁移学习流程图见图2。



A, B: 恶性肿块; C, D: 良性肿块  
图 1 人工裁剪预处理后的乳腺肿块超声图像

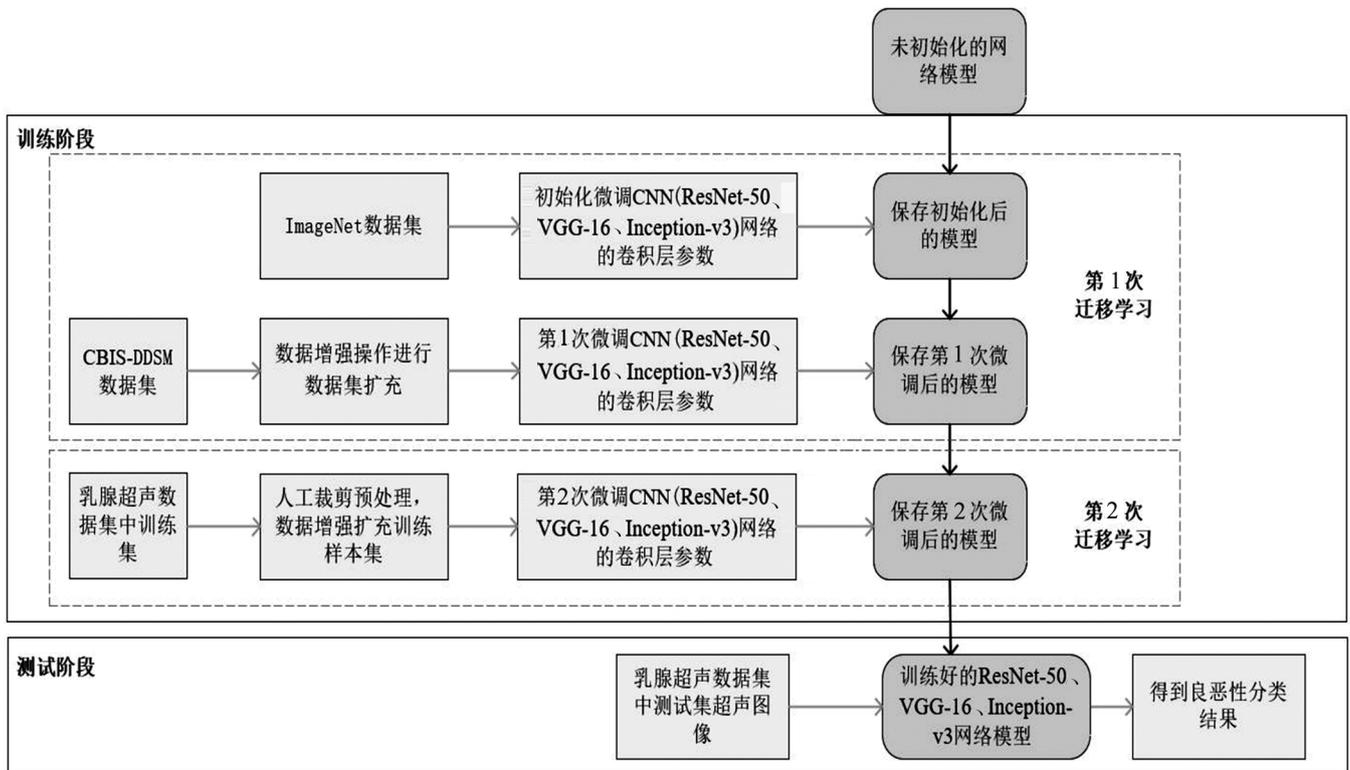


图 2 CNN 迁移学习流程图

5. 模型训练和测试: 本研究选取随机梯度下降为迁移学习的优化函数, 损失函数为交叉熵损失函数。为了训练网络, 每个 Batch 包含 16 张图像, 初始学习率设置为 0.0001。最多训练 200 个 epoch, 当 10 个 epoch 的验证误差不减少时, 训练提前结束。最后选择验证损失最低的模型用来测试。

三、统计学处理

应用 SPSS 26.0 统计软件, 计量资料以  $\bar{x} \pm s$  表示, 两组比较采用 *t* 检验; 计数资料以例或率表示, 两组比较采用  $\chi^2$  检验。以病理结果为金标准, 绘制受试者工作特征(ROC)曲线分析迁移学习后 3 种模型鉴别诊断乳腺肿块良恶性的效能, 曲线下面积(AUC)比较行 *Z* 检验。 *P* < 0.05 为差异有统计学意义。

结 果

一、临床资料比较

良、恶性乳腺肿块患者平均年龄分别为 (41.27 ± 11.64) 岁、(51.84 ± 10.86) 岁, 肿块平均最大径分别为 (2.03 ± 1.53) cm、(2.55 ± 1.22) cm, 差异均有统计学意义 (均 *P* < 0.001)。训练集和测试集患者平均年龄分别为 (46.66 ± 13.05) 岁、(46.99 ± 11.06) 岁, 肿块平均最大径分别为 (2.33 ± 1.49) cm、(2.25 ± 1.20) cm, 差异均无统计学意义。

二、未迁移与迁移学习后 3 种模型对乳腺肿块良恶性的鉴别诊断效能比较

未迁移与迁移学习后 3 种模型鉴别诊断乳腺肿块良恶性的准确率、敏感性、特异性、精准率、F1 分数、约

登指数见表1。

1. 第1次迁移学习与未迁移学习比较:3种模型第1次迁移学习后鉴别诊断乳腺肿块良恶性的准确率、敏感性、特异性、精准率、F1分数及约登指数均高于未迁移,差异均有统计学意义(均 $P<0.05$ )。

2. 第2次迁移学习与第1次迁移学习比较:3种模型第2次迁移学习后鉴别诊断乳腺肿块良恶性的准确率、敏感性、特异性、精准率、F1分数及约登指数均高于第1次迁移学习,差异均有统计学意义(均

$P<0.05$ )。

### 三、ROC曲线分析

ROC曲线分析显示,VGG-16、Inception-v3、ResNet-50第2次迁移学习后鉴别诊断乳腺肿块良恶性的AUC分别为0.841(95%可信区间0.759~0.916)、0.865(95%可信区间0.756~0.946)、0.915(95%可信区间0.869~0.949)。基于ResNet-50建立的迁移学习模型鉴别诊断乳腺肿块良恶性的AUC高于其他两种模型,差异均有统计学意义(均 $P<0.05$ )。见图3。

表1 未迁移与迁移学习后3种模型对乳腺肿块良恶性的鉴别诊断效能

模型	迁移次数	准确率(%)	敏感性(%)	特异性(%)	精准率(%)	F1分数	约登指数
VGG-16	未迁移	75.0	71.2	79.2	78.7	0.747	0.503
	1次	78.0*	75.0*	81.3*	81.3*	0.780*	0.563*
	2次	82.0#	80.8#	83.3#	84.0#	0.824#	0.641#
Inception-v3	未迁移	79.0	73.1	85.4	84.4	0.784	0.585
	1次	83.0*	78.8*	87.5*	87.2*	0.828*	0.663*
	2次	85.0#	80.8#	89.6#	89.4#	0.848#	0.704#
ResNet-50	未迁移	82.0	76.9	87.5	87.0	0.826	0.644
	1次	85.0*	78.8*	91.7*	91.1*	0.874*	0.705*
	2次	88.0#	82.7#	93.8#	93.5#	0.913#	0.764#

与同一模型未迁移比较,\* $P<0.05$ ;与同一模型迁移1次比较,# $P<0.05$

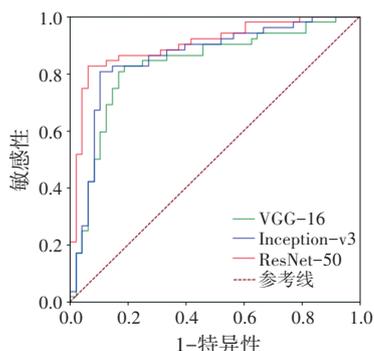


图3 第2次迁移学习后3种模型鉴别诊断乳腺肿块良恶性的ROC曲线图

## 讨论

近年来,深度学习在乳腺疾病诊断中的研究越来越多,但多是应用传统机器学习通过手动勾画感兴趣区域进行纹理分析或是基于深度学习的乳腺钼靶X线诊断,而应用两次迁移学习鉴别诊断乳腺肿块良恶性的报道较少。基于此,本研究采用深度学习的方法分别比较两次迁移学习后3种CNN模型(VGG-16、Inception-v3、ResNet-50)对乳腺肿块良恶性的鉴别诊断效能,探讨基于超声图像的迁移学习模型在乳腺肿块良恶性鉴别诊断中的价值。

在基于深度学习的图像分类问题中,数据是模型

的根本和基础<sup>[7]</sup>,虽然可对数据进行增强处理,但用于训练网络的样本仍然较少,这可能会增加出现过拟合、模型泛化能力差等问题的几率,导致在乳腺肿块良恶性鉴别诊断中无法获得良好效果。迁移学习是深度学习中一种新的学习方式,其可在一定程度上缓解数据有限的问题,解决医学图像分析中由于数据集较小导致模型不准确的问题<sup>[8]</sup>。Byra等<sup>[9]</sup>先在ImageNet数据集上对VGG19、Inception-v3及Inception ResNet-v2 3种模型进行预训练,经过迁移学习后对149例良性和82例恶性乳腺肿块的超声图像进行分类,其诊断乳腺肿块良恶性的AUC分别为0.858、0.829及0.860,表明基于超声图像的迁移学习在鉴别诊断乳腺肿块良恶性中具有较好价值。Hadad等<sup>[10]</sup>基于乳房X线摄影图像训练了一个网络模型,直接迁移至乳腺MRI数据集上并对其进行微调,分析3种CNN模型在乳腺MRI图像病变分类中的表现,结果表明在缺乏足够训练数据的情况下,乳腺X线摄影和乳腺MRI之间的跨模态迁移学习可以显著提高诊断效能。本研究通过VGG-16、Inception-v3、ResNet-50 3种模型实现乳腺X线摄影和乳腺超声之间的跨模态迁移学习,结果显示迁移学习后3种CNN模型对乳腺肿块良恶性的鉴别诊断效能均有所提高,第1次迁移学习与未迁移

学习及第 2 次迁移学习与第 1 次迁移学习比较, 准确率、敏感性、特异性均明显提高, 差异均有统计学意义(均  $P < 0.05$ ), 表明迁移学习算法能有效地在相似图像之间进行信息共享和迁移。此外, 在第 1 次迁移的基础上, 第 2 次迁移借助少量数据就能达到预期较好的诊断效果, 表明两次迁移学习较一次迁移学习效果更佳, 更适合用于小数据集的乳腺肿块超声图像良恶性鉴别诊断, 与 Hodad 等<sup>[10]</sup>研究结论一致。

本研究中基于 ResNet-50 建立的第 2 次迁移学习模型鉴别诊断乳腺肿块良恶性的准确率为 88.0%, 敏感性为 82.7%, 特异性为 93.8%, AUC 为 0.915(95% 可信区间 0.869~0.949), 较 VGG-16、Inception-v3 模型具有更优的诊断价值, 分析原因可能为 ResNet 网络创新性地引入残差网络, 改善了由于网络深度加深而产生的学习效率降低和准确率无法有效提升的问题<sup>[11]</sup>。但 Zhang 等<sup>[12]</sup>研究显示, Inception-v3 模型在测试集超声图像中鉴别诊断乳腺肿块良恶性的 AUC 为 0.905, 相较于 ResNet-50 显示出更高的诊断准确率, 这可能是由训练数据量差异引起的。

年龄是罹患乳腺癌的高危因素之一, 本研究中恶性乳腺肿块患者平均年龄为(51.84±10.86)岁, 与良性乳腺肿块患者平均年龄[(41.27±11.64)岁]比较差异有统计学意义( $P < 0.05$ ), 与以往研究<sup>[13-14]</sup>显示我国女性乳腺癌的发病高峰年龄为 45~54 岁基本一致。乳腺为多种内分泌激素作用的靶器官, 其中雌激素与乳腺癌的发病有直接关系, 分析其原因可能与年龄增长后的内分泌失调、卵巢功能紊乱及绝经后女性雌激素水平下降有关<sup>[15]</sup>。

综上所述, 基于超声图像的迁移学习模型能够提高乳腺肿块良恶性的鉴别诊断效能, 减少由于样本数据量少对基于深度卷积特征的乳腺超声图像应用产生的影响, 可为临床医师提供决策参考, 具有较高的临床应用价值。但本研究为单中心、回顾性研究, 纳入样本量有限, 图像来源于不同超声医师, 且缺乏独立的外部测试集以进一步验证模型的诊断效能, 未来仍需开展多中心、大样本研究进一步验证试验结果。

## 参考文献

- [1] Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries [J]. CA Cancer J Clin, 2021, 71(3): 209-249.
- [2] Colditz GA, Bohlke K. Priorities for the primary prevention of breast cancer [J]. CA Cancer J Clin, 2014, 64(3): 186-194.
- [3] 赵艳霞, 马兰, 连臻强, 等. 2014 年中国农村基于超声乳腺筛查多中心数据分析 [J]. 中华肿瘤防治杂志, 2020, 27(3): 172-178.
- [4] Chan HP, Samala RK, Hadjiiski LM. CAD and AI for breast cancer—recent development and challenges [J]. Br J Radiol, 2019, 93(1108): 20190580.
- [5] Shin HC, Roth HR, Gao M, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning [J]. IEEE Trans Med Imaging, 2016, 35(5): 1285-1298.
- [6] Lee RS, Gimenez F, Hoogi A, et al. A curated mammography data set for use in computer-aided detection and diagnosis research [J]. Sci Data, 2017, 4(1): 1-9.
- [7] 肖焕辉, 袁程朗, 冯仕庭, 等. 基于深度学习的癌症计算机辅助分类诊断研究进展 [J]. 国际医学放射学杂志, 2019, 42(1): 22-25.
- [8] Chan HP, Samala RK, Hadjiiski LM, et al. Deep learning in medical image analysis [J]. Adv Exp Med Biol, 2020, 1213(1): 3-21.
- [9] Byra M, Sznajder T, Korzinek D, et al. Pattern Recognition and Image Analysis [M]. Berlin: Springer-Verlag Berlin and Heidelberg GmbH & Co. KG, 2019: 41-52.
- [10] Hadad O, Bakalo R, Ben-Ari R, et al. Classification of breast lesions using cross-modal deep learning [C]// 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). IEEE, 2017.
- [11] Alaeddine H, Jihene M. Deep residual network in network [J]. Comput Intell Neurosci, 2021, 23(2): 1-9.
- [12] Zhang H, Han L, Chen K, et al. Diagnostic efficiency of the breast ultrasound computer-aided prediction model based on convolutional neural network in breast cancer [J]. J Digit Imaging, 2020, 33(5): 1218-1223.
- [13] 中国抗癌协会乳腺癌专业委员会. 中国抗癌协会乳腺癌诊治指南与规范(2021年版) [J]. 中国癌症杂志, 2021, 31(10): 954-1040.
- [14] 陈万青, 郑荣寿. 中国女性乳腺癌发病死亡和生存状况 [J]. 中国肿瘤临床, 2015, 42(13): 668-674.
- [15] Lukaszewicz S, Czeczulewski M, Forma A, et al. Breast cancer—epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—an updated review [J]. Cancers (Basel), 2021, 13(17): 4287.

(收稿日期: 2022-03-13)