

引用格式: DING Lihao, GAO Zhishan, ZHU Dan, et al. Classification Method of Breast Tissue OCT Images Based on a Double Filtering Residual Network[J]. Acta Photonica Sinica, 2023, 52(12):1210001

丁李昊,高志山,朱丹,等. 基于双重过滤残差网络的乳腺组织 OCT 图像分类方法[J]. 光子学报, 2023, 52(12):1210001

# 基于双重过滤残差网络的乳腺组织 OCT 图像分类方法

丁李昊,高志山,朱丹,袁群,郭珍艳

(南京理工大学 电子工程与光电技术学院, 南京 210094)

**摘要:**提出一种基于残差网络改进的卷积神经网络模型,应用于乳腺组织光学相干层析图像的辅助诊断。该方法采用残差网络作为基础框架,利用级联的小卷积层替换大卷积层,使用八度卷积替换部分卷积层,并且加入注意力机制,构造出一个具有双重过滤结构的残差网络模型。双重过滤结构不仅能够减少模型整体的参数计算量,还能使模型在特征提取时侧重于含有丰富结构信息的高频分量,降低低频分量占比,减少信息冗余,从而提高模型对于近似结构图像的分类识别能力。对比实验结果表明,提出的双重过滤残差网络模型对纤维脂肪、基质和肿瘤 3 类乳腺组织光学相干层析图像实现 96.88% 的分类精确度,且分类性能比目前经典的卷积神经网络模型更加优异,在乳腺癌切缘术实时诊断领域具有潜力。

**关键词:**光学相干层析术;图像分类;高频信息;八度卷积;注意力机制;深度学习

中图分类号:R737.9;TP18

文献标识码:A

doi:10.3788/gzxb20235212.1210001

## 0 引言

近年来,乳腺癌已经成为发病率最高的肿瘤,严重影响着女性的生理健康<sup>[1-2]</sup>。由于乳腺癌没有特别好的预防方法,所以在目前的医疗领域中,保乳手术是早期(I或II)乳腺癌的主要治疗方法之一<sup>[3]</sup>。然而,乳腺癌组织的切缘诊断较为困难,导致手术的再切除率较高(高达 15%~25%)<sup>[4-5]</sup>。目前主流的切片诊断技术存在三种缺陷:1)定位取材困难;2)乳腺的脂肪组织较多,导致制片困难;3)乳腺中高度散射的组织通常看起来极其相似,人为诊断具有主观性,容易发生误诊<sup>[6-8]</sup>。

针对上述三种缺陷,研究人员尝试将光学相干层析术(Optical Coherence Tomography, OCT)与计算机辅助诊断技术相结合,以实现乳腺组织 OCT 图像的自动分类。2014 年,伊利诺伊大学的 SOUTH F A 等<sup>[9]</sup>在 ERICKSON-BHATT S J<sup>[10]</sup>的研究基础上研制出了对生物组织双折射敏感的偏振敏感型 OCT (Polarization-Sensitive OCT, PS-OCT),能够根据正常组织和浸润性导管癌(Invasive Ductal Carcinoma, IDC)之间的胶原含量差异,提供更明显的显微结构信息,并且具有较好的实时性。2018 年,同是伊利诺伊大学的 WANG Jianfeng 等<sup>[11]</sup>利用 PS-OCT 的光强信息和偏振信息,结合统计学方法对纤维脂肪、基质和 IDC 图像进行分类,达到了 89.4% 的分类精度。2020 年,BUTOLA A 等<sup>[12]</sup>提出了一种新型卷积网络模型“LightOCT”,整体网络中只有两个卷积层和一个全连接层。模型在分类 44 张正常和 44 张 IDC 乳腺组织 OCT 图像时,达到 98.9% 的准确度;在分类超过 100 000 幅图像的大型公共数据集时,达到了 96% 的测试准确度。2021 年,ZHU Dan 等<sup>[13]</sup>利用机器学习和 PS-OCT 技术对人类乳腺组织中肿瘤、纤维脂肪和基质图像自动分类,模型总体准确率达到 93.5%。

基金项目:国家自然科学基金(Nos. 62175107,62205148),国家重点研发计划(No. 2022YFF0706302)

第一作者:丁李昊,dinglihao1999@163.com

通讯作者:朱丹,danzhu@njust.edu.cn

收稿日期:2023-05-10;录用日期:2023-07-26

<http://www.photon.ac.cn>

在乳腺组织 OCT 图像中,基质和肿瘤两类组织在整体结构上具有极高的相似性。而先前对乳腺组织 OCT 图像分类的研究中,大多使用传统的机器学习或是层次结构较浅的卷积神经网络(Convolutional Neural Networks, CNN)模型<sup>[14-16]</sup>,按照指定顺序提取图像特征,对高低频分量没有侧重,导致模型可能提取较多的低频分量,造成信息冗余,从而降低模型整体的分类精度。虽然可以通过增加 CNN 模型的结构层深度来解决特征提取不够充分的问题,但是一味地加深结构层不仅会增加整体模型的参数计算量,还会造成梯度消失或梯度爆炸现象<sup>[17]</sup>。因此,本文将图像中含有丰富结构信息的高频分量作为关键特征信息,以残差网络(Residual Network, ResNet)<sup>[18]</sup>为整体框架,使用八度卷积(Octave Convolution, OctConv)<sup>[19]</sup>和轻量的注意力模块(Convolutional Block Attention Module, CBAM)<sup>[20]</sup>相互配合提取特征信息,构建了一种改进的“双重过滤(Double Filtering)”残差网络(DF-ResNet)模型,该模型在特征提取时更侧重于高频分量,对于结构近似的基质和肿瘤两类 OCT 图像表现出更好的分类识别能力。

## 1 网络架构组成

构建的 DF-ResNet 模型结构和工作流程如图 1 所示。首先,DF-ResNet 模型整体采用 ResNet-18 作为结构框架,以免发生梯度消失或者梯度爆炸现象;其次,在保证感受野不变的情况下,用 3 个级联的  $3 \times 3$  小卷积层堆叠代替初始网络结构中  $7 \times 7$  的大卷积层,减少参数计算量并起到隐式正则化的作用;接着,分别在级联的小卷积层和最后一层卷积层之后加入通道注意力模块与空间注意力模块相结合的 CBAM,实现图像特征的自适应细化;同时,利用 OctConv 替换原始模型中大小为  $3 \times 3$  的卷积层,使网络在提取图像特征时具备调整高低频分量占比的能力;最后,利用全局平均池化层(Global Average Pooling, GAP)<sup>[21]</sup>代替全连接层,减少网络参数计算量,并对网络结构做正则化处理,防止模型发生过拟合。

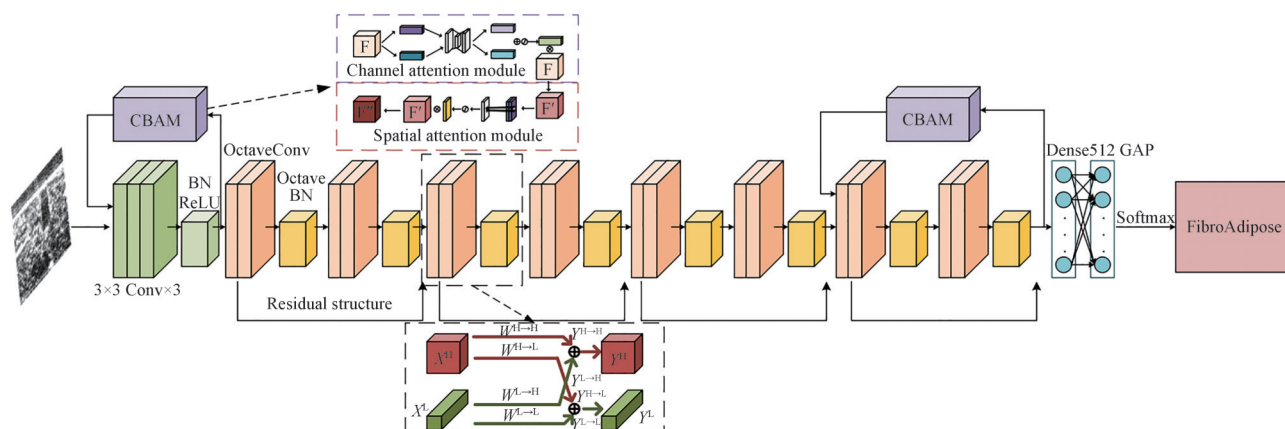


图 1 DF-ResNet 模型工作流程  
Fig. 1 DF-ResNet model workflow

### 1.1 初始层结构

DF-ResNet 模型中使用 3 个级联的  $3 \times 3$  小卷积层堆叠代替初始层中  $7 \times 7$  的大卷积层,目的是在保证感受野大小不变的情况下,利用小卷积层堆叠把原先大卷积层进行非线性分解,不仅减少了模型参数计算量,还起到了隐式正则化的作用。接着,在级联的小卷积层之后连接批量归一化层(Batch Normalization, BN)<sup>[22]</sup>,并使用非线性激活函数 ReLU<sup>[23]</sup>进行激活,最后使用最大池化层进行下采样降维,去除冗余信息,提高模型整体的收敛速度。图 1 中绿色组合块代表经过修改后的初始层结构。

### 1.2 CBAM 工作原理

DF-ResNet 模型分别在级联的  $3 \times 3$  小卷积层和最后一层卷积层之后加入 CBAM 过滤特征信息。CBAM 是一种用于前馈 CNN 中的简单而有效的注意力模块,相比于挤压激励网络(Squeeze-and-Excitation Networks, SE-Net)<sup>[24]</sup>和高效通道注意力网络(Efficient Channel Attention, ECA-Net)<sup>[25]</sup>,CBAM 不仅实现了通道和空间注意力的双机制,而且模块采用最大池化相加和平均池化堆叠代替单一的最大池化和平均池

化。对于一张输入的特征图  $F$ , CBAM 机制先对其进行通道维度计算, 得到一个与原始特征图大小相同的通道注意力权重矩阵, 将通道注意力权重矩阵与原始特征图  $F$  相乘得到权重修正后的特征图  $F'$ ; 然后对特征图  $F'$  进行空间维度计算, 得到一个与特征图  $F'$  大小相同的空间注意力权重矩阵, 将空间权重矩阵与经过通道注意力机制修正后的特征图  $F'$  相乘以进行自适应特征细化, 最终得到经过通道和空间注意力机制双重修正后的特征图  $F''$ 。特征图  $F'$  和  $F''$  的计算公式为

$$M_c(F') = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) = \sigma(W_1(W_0(F_{\text{arg}}^c)) + W_1(W_0(F_{\text{max}}^c))) \quad (1)$$

式中,  $M_c$  表示通道注意力图,  $F' \in R^{c \times w \times h}$ ,  $R$  表示实数空间,  $c, w, h$  分别表示输入特征的长度、宽度和高度,  $\sigma$  表示 sigmoid 函数<sup>[24]</sup>, MLP 表示一个多层感知器, AvgPool 表示平均池化操作, MaxPool 表示最大池化操作,  $W_0$  和  $W_1$  分别为 MLP 第一层和第二层的权重参数,  $W_0 \in R^{c/r \times c}$ ,  $W_1 \in R^{c \times c/r}$ ,  $r$  表示缩减率,  $F_{\text{arg}}^c$  表示通道注意力中的平均池化特征,  $F_{\text{max}}^c$  表示通道注意力中的最大池化特征。

$$M_s(F'') = \sigma(f^{7 \times 7}([\text{AvgPool}(F''); \text{MaxPool}(F'')])) = \sigma(f^{7 \times 7}([F_{\text{avg}}^s; F_{\text{max}}^s])) \quad (2)$$

式中,  $M_s$  表示空间注意力图,  $f^{7 \times 7}$  表示  $7 \times 7$  大小的卷积核,  $F_{\text{avg}}^s$  表示空间注意力中的平均池化特征,  $F_{\text{max}}^s$  表示空间注意力中的最大池化特征。

### 1.3 OctConv 工作原理

一张图像的中低频分量代表图像强度变换平缓的区域, 通常是连续渐变处; 高频分量代表图像强度变换剧烈的区域, 通常是含有大量结构信息的细节轮廓处。显然特征图也属于图像, 可以认为经 CNN 模型特征提取后得到的特征图同样具有高低频分量信息。

CHEN Y 等<sup>[19]</sup>在 2019 年以 ResNet-18 为基础框架, 利用 OctConv 替换部分卷积层, 构造出改进的残差模型 Oct\_ResNet-26 用于提取图像中的高频分量。OctConv 是一种新型卷积结构, 通过降低低频分量占比, 使卷积核更专注于提取特征图中的高频分量, 减少了特征图在空间维度上的信息冗余。OctConv 在进行卷积操作之前, 会先将特征图分为高频分量  $X^H$  和低频分量  $X^L$ , 再使低频分量  $X^L$  的分辨率降低至原来的二分之一。OctConv 的卷积核可根据其结构中的低频维度占比参数  $\alpha$  分成 4 个部分: 高频到高频  $O^{H \rightarrow H}$ , 高频到低频  $O^{H \rightarrow L}$ , 低频到高频  $O^{L \rightarrow H}$ , 低频到低频  $O^{L \rightarrow L}$ 。其中  $\alpha$  取值为  $[0, 1]$ , 并且应为 0.125 的整数倍。

OctConv 将输入的高低频分量经过对应的卷积核, 通过调整各自分辨率和点加等操作最终可以得到低频输出特征图  $Y^L$  和高频输出特征图  $Y^H$ 。

### 1.4 残差单元

残差单元是残差神经网络最基本的结构块, 残差神经网络令残差单元不直接学习目标映射, 而是学习残差  $I(x) = E(x) - x$ ,  $E(x)$  表示观测值,  $x$  表示上一层输出的特征映射, 即为估计值。所以原始的残差映射由一个线性映射  $x \rightarrow x$  和一个非线性映射  $I(x)$  两部分组成, 最终变更为  $I(x) + x$ 。残差单元一般是由卷积层、批量归一化层和激活函数 ReLU 层构成。

假设第  $l$  个残差单元的输入为  $x_l$ , 那么该层的输出  $x_{l+1}$  可以表示为

$$x_{l+1} = f_R(x_l + I(x_l, W_l)) \quad (3)$$

式中,  $f_R$  表示激活函数 ReLU,  $I(x_l, W_l)$  表示残差函数;  $W_l$  表示该残差函数对应的权重参数。式(3)是  $x_l$  和  $I(x_l, W_l)$  维度相同的情况, 若是两者维度不同, 则可以通过连接一个线性映射  $W_s$  来使两者维度匹配, 即

$$x_{l+1} = f_R(W_s x_l + I(x_l, W_l)) \quad (4)$$

## 2 分类实验

### 2.1 数据集简介

本次分类任务的图像数据来源于伊利诺伊大学香槟分校, 包含 41 位接受了保乳术、缩乳术或者乳房全切术的受试者。所有受试者在术前均签署了知情同意书, 允许对其组织进行研究使用。术中切除的组织样本浸泡在生理盐水中, 并放置于低温保温箱内保存, 在术后 24 h 内利用自建的 OCT 系统<sup>[11]</sup>完成对其成像。利用墨水对已经成像完全的样本的成像区域进行标记, 然后将样本浸泡在 10% 福尔马林溶液中并固定 72 h 以上。固定完成后, 由专业人员按组织学检验标准流程对样本进行切片以及苏木精和伊红染色 (Hematoxylin and Eosin stains, H&E stains)。对获得的组织切片图像与 OCT 层析图以组织切片的病理检



验结果作为黄金标准进行相关性检验。

## 2.2 数据集制作和预处理

本文挑选了3类成像效果较好的样本,每类样本选取20个主要包含纤维脂肪、基质和肿瘤的兴趣区域进行成像,获得多组 $512 \times 512 \times 2048$ (帧数,宽度 $\times$ 深度)的图像数据集。由于前后帧图像极为相似,为了保证图像数据的独立性,从图像数据集中每隔32帧截取一帧图像,并选择合适的深度方向区域进行裁剪,这样对于一个成像区域可以获得16张尺寸为 $512 \times 256$ 的图像,因此一共获得了960张乳腺组织OCT图像,将其作为初始数据集。其中包含320张纤维脂肪(FibroAdipose)图像,320张基质(Stroma)图像和320张肿瘤(Tumor)图像,具体结果如图2所示,图中比例尺代表 $400 \mu\text{m}$ 。

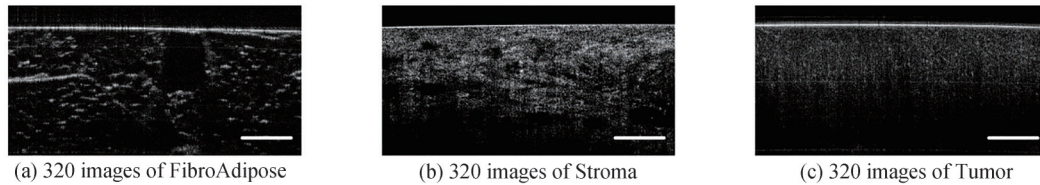


图2 三类原始数据图像  
Fig. 2 Three types of raw data images

在进行分类实验之前,为了满足CNN模型训练对于数据量的要求,减少过拟合,对初始数据集中960张乳腺组织OCT图像进行数据预处理<sup>[26-27]</sup>。首先,从每一张OCT图像中随机截取大小为 $224 \times 224$ 的区域作为新数据集中的一幅图像;接着,对新图像添加高斯噪声和椒盐噪声;最后,随机改变新图像的亮度、对比度和饱和度。对于原始数据集中的每一张OCT图像重复10次上述数据增强操作<sup>[28]</sup>,使原始数据集数量扩充10倍,最终得到9600张乳腺组织OCT图像用于模型训练、验证和测试。

## 2.3 训练策略

考虑到数据集中每一类图像以16张为一组来自相同的成像区域,而同一成像区域中的图像结构具有相似性,模型在训练时会重复学习到相同的特征,容易出现过拟合现象,最终导致测试集的分类精确度虚高,因此不能随机将数据集按比例分为训练集、验证集和测试集。将从每类样本的一块成像区域中获得的图像分成一组,一共得到20组图像数据,因此使用二十折交叉验证法训练模型,按顺序每次选择18组图像数据作为训练集,剩下的2组图像数据分别作为验证集和训练集。

针对可能出现的过拟合问题,所提的DF-ResNet模型首先在训练前通过数据增强,使得数据集扩充10倍。其次,选择ReLU作为激活函数,ReLU函数的特点会使网络中部分神经元的输出为0,通过减少参数间的相互依存,增加网络的稀疏性。另外,在模型中的卷积层之后加入BN层以此对模型参数进行归一化,从而提高网络泛化能力。最后,在训练过程中每经过5个周期就将验证集代入模型中验证并保存模型参数,若本次验证集的分类精确度高于上一次,便保存本次模型的参数;若低于上一次,便舍弃。这样即使周期选的很大,模型也能保存最合适的参数,以此来缓解过拟合问题。

## 2.4 实验环境

本次分类实验代码均采用Python语言编写,选择Pytorch(科学计算框架Torch在Python上的衍生)作为深度学习框架,Pytorch可以看作是加入了GPU支持的Numpy,因此PyTorch允许实验人员利用GPU的算力来实现神经网络的加速。总体的实验环境如表1所示。

表1 系统详细参数  
Table 1 Detailed system parameters

Name	Indicator
System	Windows 11
Memory	32 GB
Central processing unit	AMD Ryzen 7 5800H(3.20 GHz)
Graphics processing unit	NVIDIA RTX3070 GPU

## 2.5 评价指标

为了能够更加客观且直观地评价整体模型训练的性能,选择敏感度(sen)、特异度(spe)、阳性预测值(PPV)、阴性预测值(NPV)四个指标对所训练的模型整体性能进行评估。

敏感度表示在黄金标准判断阳性样本中,检测出真阳性的概率,衡量对阳性样本的识别能力,公式为

$$\text{sen} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

特异度表示在黄金标准判断阴性样本中,检测出真阴性的概率,衡量对阴性样本的识别能力,公式为

$$\text{spe} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (6)$$

阳性预测值表示在黄金标准判断阳性样本中,真正是阳性的样本比例,公式为

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (7)$$

阴性预测值表示在黄金标准判断阴性样本中,真正是阴性的样本比例,公式为

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad (8)$$

式(5)~(8)中,TP表示为被正确分类为阳性的样本个数;TN表示为被正确分类为阴性的样本个数;EP表示为被错误分类为阳性的样本个数;FN表示为被错误分类为阴性的样本个数。

## 2.6 热力图可视化

目前,对于常见的CNN模型,其整个训练及预测过程被认为是个黑盒,可解释性并不强。为了探究其内在联系,更清晰地展现出网络所关注的特征信息,通过梯度加权类激活映射(Gradient-weighted Class Activation Mapping, Grad-CAM)<sup>[29]</sup>对输入图像生成类激活的热力图,可以分析网络对于某个类别的关注区域;反之,根据网络关注的区域同样可以分析网络是否学习到了正确的特征信息。Grad-CAM首先会获取模型最后一层卷积层的输出梯度,然后表征出最后卷积层激活输出各点位对于模型决策分类的重要程度,并进行预处理,得到类激活图;接着,将调节后的卷积激活输出渲染为热力效果;最后,将原始图像与渲染后的热力效果叠加后得到热力图 $L_{\text{Grad-CAM}}^d$ 。Grad-CAM具体的原理公式为

$$L_{\text{Grad-CAM}}^d = f_R \left( \sum_k \mathbf{a}_k^d A^k \right) \quad (9)$$

式中, $f_R$ 表示激活函数ReLU, $\mathbf{a}_k^d$ 表示针对 $A^k$ 的权重, $d$ 表示类别 $d$ , $A$ 表示网络中最后一个卷积层输出的特征层, $k$ 表示特征层 $A$ 中第 $k$ 个通道, $A^k$ 表示特征层 $A$ 中通道 $k$ 的数据。

关于 $\mathbf{a}_k^d$ 的计算公式为

$$\mathbf{a}_k^d = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^d}{\partial A_{ij}^k} \quad (10)$$

式中, $Z$ 表示特征图的像素个数, $y^d$ 表示网络针对类别 $d$ 预测的分数,但没有通过softmax激活, $A_{ij}^k$ 表示特征层 $A$ 在通道 $k$ 中坐标为 $(i, j)$ 位置处的像素值。

## 3 实验结果与分析

首先分析低频维度占比参数 $\alpha$ 对于模型性能的影响,为整体模型中八度卷积结构确定合适的 $\alpha$ 参数值;然后改进CHEN Y提出的Oct\_ResNet-26模型,利用3个级联的 $3 \times 3$ 卷积层替换Oct\_ResNet-26模型中第一层 $7 \times 7$ 的卷积层,得到改进的Oct\_ResNet-28模型;另外,将经典的ResNet-34模型结构层深度减少至28层,得到新模型ResNet-28。这两种模型与DF-ResNet具有相似的架构特点,均为28层的深度结构并且采用了残差网络设计。选择这两种模型作为对比对象,有利于更直观地比较各模型结构在整体参数计算量方面的差异;同时,也有助于验证采用OctConv和CBAM进行双重特征信息过滤的有效性。

### 3.1 八度卷积中 $\alpha$ 参数对于模型性能的影响

八度卷积核的结构特点表明,参数 $\alpha$ 表示低频维度的比例,其取值范围为 $[0, 1]$ ,并且应是0.125的整数倍。当 $\alpha$ 取0时,意味着低频成分占比为0,图像仅提取高频分量;而当 $\alpha$ 取1时,表示低频成分占比为1,图像

仅提取低频分量。这两种 $\alpha$ 参数值都会导致特征图中低频或高频成分的完全缺失。因此,仅讨论 $\alpha$ 参数取值为0.125、0.25、0.375、0.5、0.625、0.75、0.875时,对Oct\_ResNet-28模型整体性能的影响。相应的实验结果如图3所示。

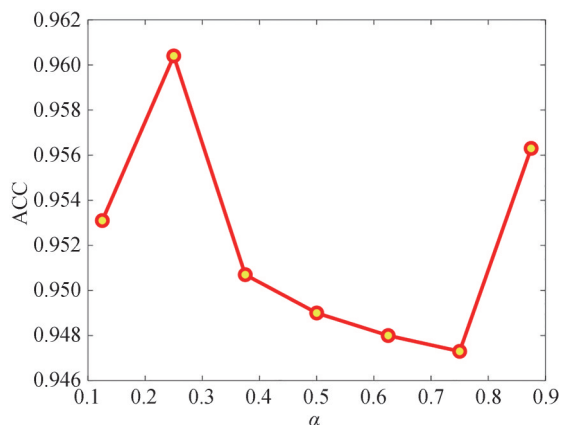


图3  $\alpha$ 参数对于模型性能的影响

Fig. 3 Effect of  $\alpha$  parameters on model performance

从图3中可以看出, $\alpha=0.25$ 时Oct\_ResNet-28模型的分类精确度最高,达到了96.05%。说明低频维度占比为0.25时,高低频分量的比例是最合适的,通过OctConv能提取到更多的有效特征,减少特征空间冗余。在接下来的模型性能分析中,模型都以 $\alpha=0.25$ 为标准进行调参实验。

### 3.2 Oct\_ResNet-28模型性能分析

以 $\alpha=0.25$ 为标准,设置批量大小batch为16,学习率lr为0.005,训练周期epochs为100,并选择带动量的随机梯度下降算法(Stochastic Gradient Descent, SGD)<sup>[30]</sup>作为优化器,在本数据集上对Oct\_ResNet-28模型进行训练。同时,选择结构层深度相同的ResNet-28模型以相同的条件在本数据集上训练,用双纵坐标表示两种模型的损失函数值和分类精度,结果如图4所示。

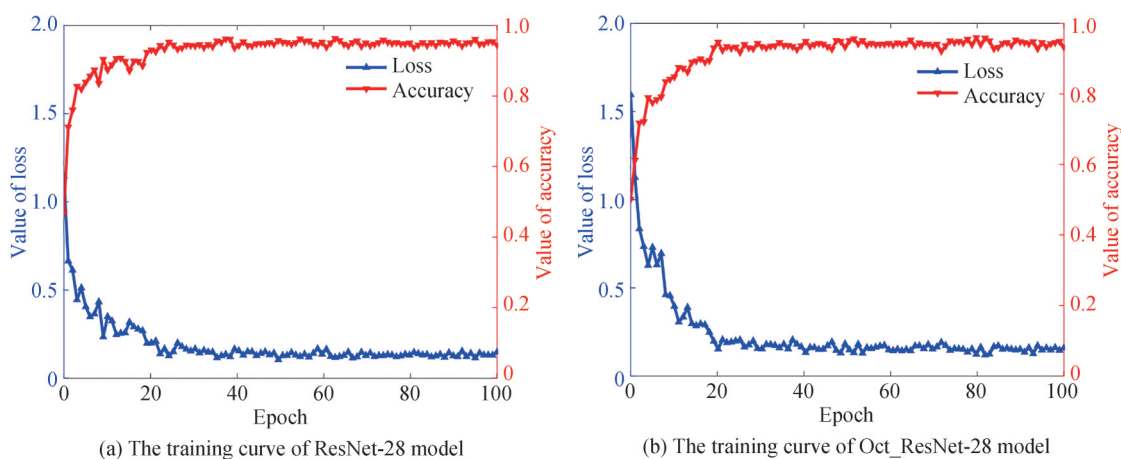


图4 模型性能训练曲线

Fig. 4 Training curve for model performance

从图4中可以发现,Oct\_ResNet-28模型初始的损失值较大,但两种模型损失值几乎都在训练了30个周期后趋向于收敛,波动幅度相似。说明在残差网络中利用OctConv代替普通卷积核这一操作不会影响整体模型的收敛速度,只是初始损失值较大,原因是OctConv工作时会将低频信息的分辨率降低至原来的二分之一,导致模型训练初期识别单个训练样本与真实值之间的误差较大。两者模型性能对比如表2所示,表中 $F_B$ 表示纤维脂肪,S表示基质,T表示肿瘤。

表 2 三种模型性能对比  
Table 2 Comparison of performance among three models

Method	Images	Sen	Spe	PPV	NNP	Accuracy
ResNet-28	F <sub>B</sub>	0.987 5	0.995 3	0.990 6	0.993 8	0.938 5
	S	0.878 1	0.971 9	0.939 8	0.941 0	
	T	0.950 0	0.940 6	0.888 9	0.974 1	
Oct_ResNet-28	F <sub>B</sub>	0.987 5	0.987 5	0.975 3	0.993 7	0.954 2
	S	0.915 6	0.976 6	0.951 3	0.958 6	
	T	0.959 4	0.967 2	0.936 0	0.979 4	
DF-ResNet	F <sub>B</sub>	<b>0.990 6</b>	0.992 2	0.984 5	<b>0.995 3</b>	<b>0.968 8</b>
	S	<b>0.946 9</b>	<b>0.981 3</b>	<b>0.961 9</b>	<b>0.973 6</b>	
	T	<b>0.968 8</b>	<b>0.979 7</b>	<b>0.959 8</b>	<b>0.984 3</b>	

从表 2 中可以看出, Oct\_ResNet-28 模型的总体分类精确度达到了 95.42%, 相比 ResNet-28 模型总体分类性能提升了 1.57%, 其识别纤维脂肪图像的能力与 ResNet-28 模型相当, 对于基质和肿瘤图像的分类识别能力明显优于 ResNet-28 模型。分析原因, ResNet-28 模型在提取图像特征时没有侧重, 特别是对于结构十分相似的基质与肿瘤图像, 容易产生信息冗余, 这些冗余信息会干扰 ResNet-28 模型最终对图像的识别分类; 而 Oct\_ResNet-28 模型在特征提取时会侧重含有丰富结构信息的高频分量, 这对于模型最终分类结构近似的图像将起到关键作用。实验结果表明: 利用 OctConv 替换部分卷积层这一操作的确能够使模型在提取图像特征时侧重于高频分量, 增强模型对于结构近似的图像的分类识别能力, 提高模型整体的分类精度, 且使用 OctConv 不会影响模型整体的收敛速度。

### 3.3 DF-ResNet 模型性能分析

以  $\alpha = 0.25$  为标准, 设置批量大小 batch 为 16, 学习率 lr 为 0.005, 训练周期 epochs 为 100, 选择 SGD 作为优化器, 在本文数据集上对 DF-ResNet 模型进行训练, 并做性能评测。最后将 DF-ResNet 模型对三类 OCT 图像的分类识别能力与 Oct\_ResNet-28 和 ResNet-28 进行比对。

从表 2 中可以看出, DF-ResNet 模型的总体分类精确度达到了 96.88%, 相较于 Oct\_ResNet-28 模型总体分类性能提升了 1.46%, 不仅在结构近似的基质和肿瘤图像分类实验中表现出更优异的识别能力, 而且对于纤维脂肪的分类精度也高于前两种对比模型。实验结果表明: DF-ResNet 模型在首层和末层加入 CBAM 并结合 OctConv 的“双重过滤”结构能够进一步过滤低频信息, 提升模型对于高频特征的提取能力, 减少信息冗余, 提高模型的分类精度。图 5 是 DF-ResNet 模型对三种图像分类的工作特性曲线 (Receiver Operating Characteristic, ROC)。

当以纤维脂肪、基质和肿瘤为阳性等级时, 曲线下面积 (Area Under the Curve, AUC) 分别为 0.990, 0.965 和 0.951。由此可以发现, DF-ResNet 模型对于纤维脂肪这一类具有独特“蜂窝状”结构的 OCT 图像几乎可以做到完全正确的识别分类, 对于结构相似性较高的基质和肿瘤图像也表现出较好的识别性能。为了进一步具现化模型学习到的特征信息, 通过 Grad-CAM 绘制出 ResNet-28 和 DF-ResNet 两种模型的热力图, 并使用 Matlab 设置阈值为 70, 对每幅热图中大于该阈值的像素点进行统计, 将其除以总像素点数可以得到提取特征面积占总图像面积的比例, 从而定量分析模型在特征提取方面的是否产生冗余信息, 具体结果如图 6 所示。图中展示了对 ResNet-28 和 DF-ResNet 两种模型在最后一层提取特征分量的比较结果, 其中比例尺代表 400  $\mu\text{m}$ 。括号内的数据反映了两种模型在最后一层提取到的特征面积所占原图面积的比例。

从图中可以看出, ResNet-28 和 DF-ResNet 两种模型在提取特征时关注的图像特征范围所占比例存在较大差异。在 ResNet-28 模型中, 对于 3 类乳腺组织 OCT 图像提取到的特征范围占比均超过 75%, 其中包括大量无用的背景信息和盖玻片阴影等。这表明 ResNet-28 模型在特征提取时没有侧重, 反而提取了较多的冗余信息, 对最终的分类识别造成了干扰。相比之下, DF-ResNet 模型对于具有独特的“蜂窝状”结构纤维脂肪图像所提取的特征范围与 ResNet-28 相近; 但对于没有明显结构特征且结构近似度高的基质和肿瘤两类图像, 该模型所提取的特征范围分别仅为 59.37% 和 67.67%, 远小于 ResNet-28。说明 DF-ResNet 模型



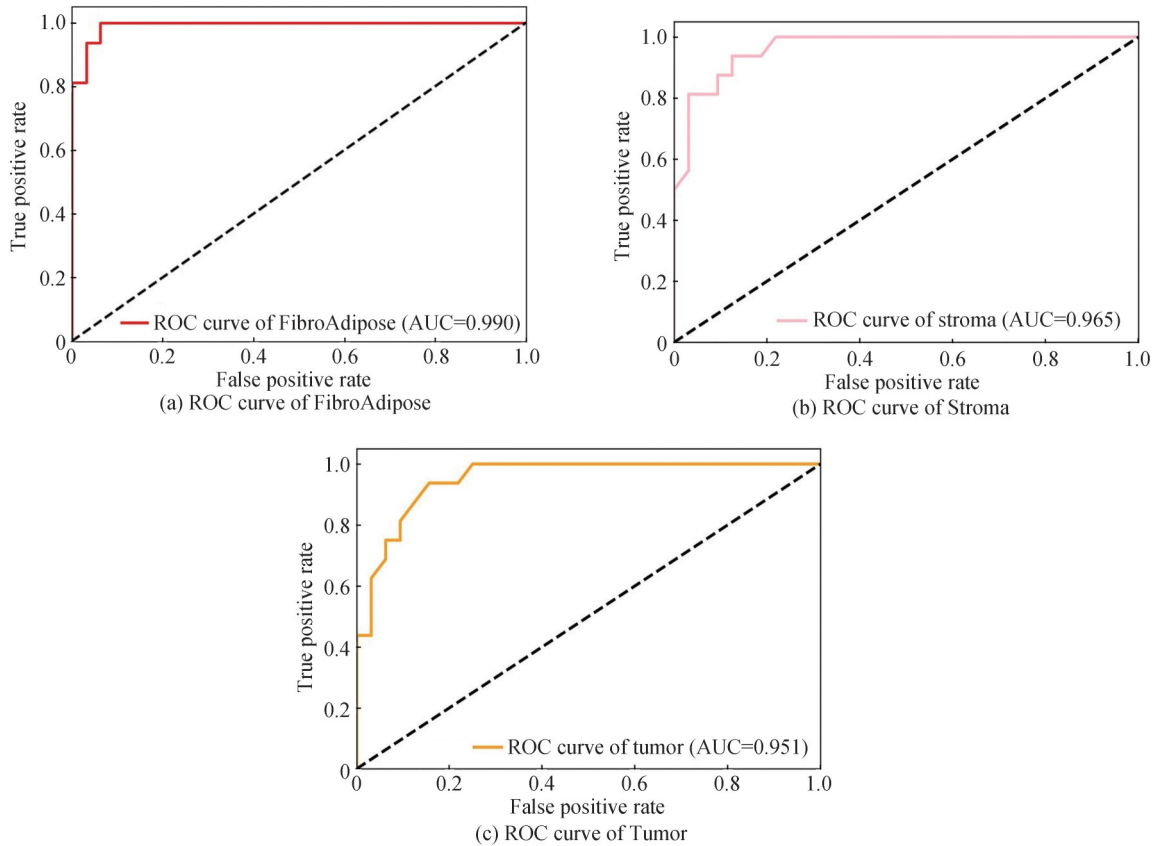


图5 DF-ResNet模型ROC曲线  
Fig. 5 ROC curve of DF-ResNet model

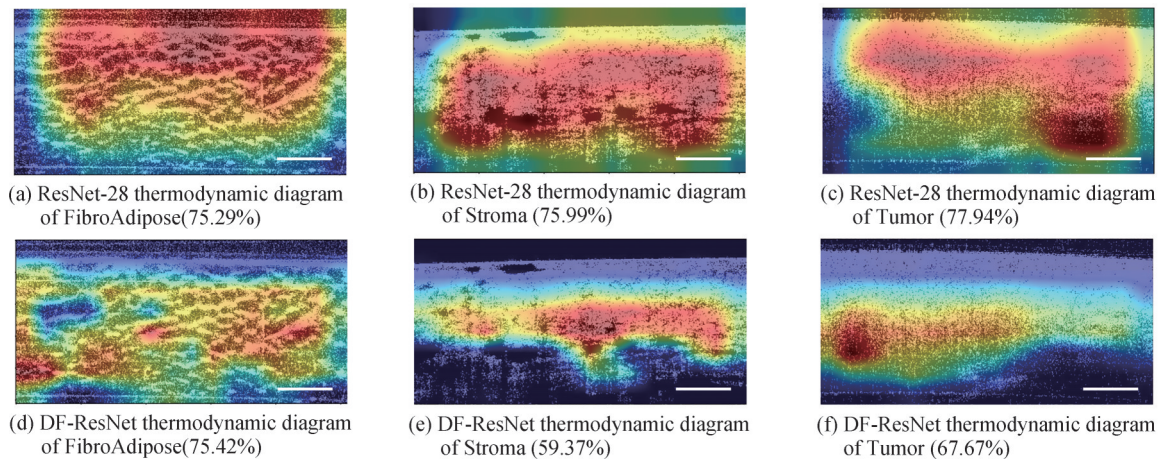


图6 两种模型热力图  
Fig. 6 The thermodynamic diagram of the two models

在特征提取方面表现出更强的针对性和精细性,侧重于对高频分量的提取,关注图像中含有丰富结构信息的轮廓边缘和细节处。以此进一步证明了DF-ResNet的“双重过滤”结构能够使模型在特征提取时更侧重于高频分量,过滤掉较多无用的低频分量,减少信息冗余。

最后,对DF-ResNet模型的泛用能力进行分析。使用预训练的DF-ResNet模型对新的乳腺组织OCT图像数据集进行识别和分类。新的数据集共有153张图像,其中包括51张纤维脂肪组织图像、51张基质图像和51张肿瘤图像。表3展示了DF-ResNet模型对新的乳腺组织OCT图像数据集的分类结果。

根据表3所示数据,预训练的DF-ResNet模型在识别纤维脂肪组织图像方面展现出卓越性能,其敏感性



表 3 DF-ResNet 对新数据集的识别分类性能  
Table 3 The recognition and classification performance of DF-ResNet on a new dataset

Method	Images	Sen	Spe	PPV	NNP	Accuracy
DF-ResNet	F <sub>B</sub>	1.000	1.000	1.000	1.000	
	S	0.882 0	0.971 0	0.938 0	0.943 0	0.941 2
	T	0.941 0	0.941 0	0.889 0	0.970 0	

(sen)、特异性(spe)、阳性预测值(PPV)和阴性预测值(NPV)均达到了 100%,意味着模型成功实现了对此类图像的完全准确分类。同时,该模型在识别结构相似的基质和肿瘤图像上也表现出良好的性能。最终模型针对新的乳腺组织 OCT 图像数据集总体分类精度达到了 94.12%,这表明预训练的 DF-ResNet 模型不仅在本文使用的数据集上具有较好效果,还在新的乳腺组织 OCT 图像数据集上展现出优秀的识别性能。这一结果证明了 DF-ResNet 模型具备良好的泛化能力,因而可以适用于术中切缘诊断场景。

### 3.4 对比 DF-ResNet 与其他网络模型性能差异

将 DF-ResNet 模型与当前流行的经典 CNN 网络,如 DenseNet-169<sup>[31]</sup>、VGG-19<sup>[32]</sup>、和 EfficientNet<sup>[33]</sup> 等进行参数计算量及性能的对比较分析。在相同的训练条件下,将这些经典模型应用于本数据集进行训练。表 4 汇总了各模型具体分类精度和模型参数计算量,为了突出对比效果,将结果绘制成柱状图如图 7,这种比较方法有助于全面评估 DF-ResNet 模型在实际应用中的优势和局限。

表 4 各模型参数量及分类精确度  
Table 4 model parameter quantity and classification accuracy

Method	Accuracy	Parameter quantity/MiB
VGG-19	0.793 9	143.7
EfficientNet	0.916 7	5.290
ResNet-28	0.938 5	19.14
DenseNet-169	0.957 2	14.15
Oct_ResNet-28	0.954 2	15.99
DF-ResNet	0.968 8	16.52

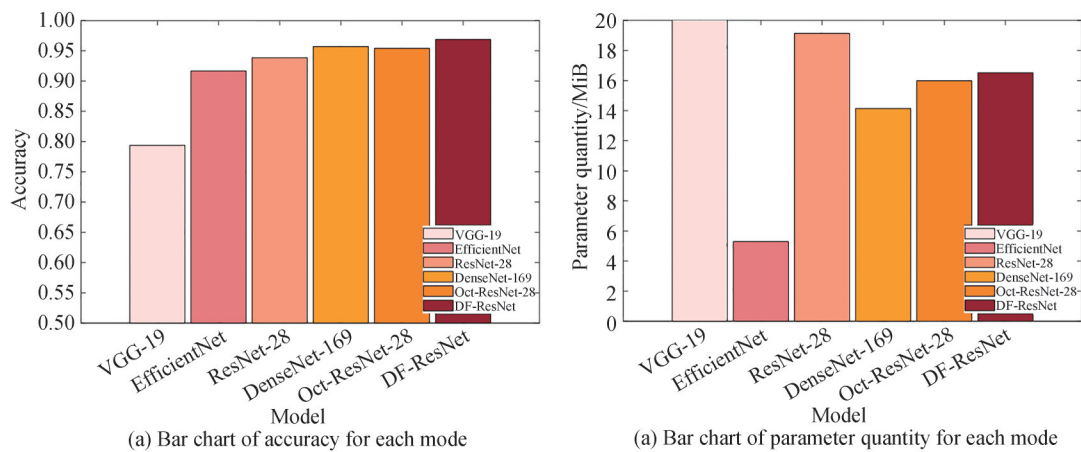


图 7 各模型参数量及分类精确度柱状图

Fig. 7 Bar chart of model parameter quantity and classification accuracy

根据表 4 中的数据,模型最终的准确率梯度为:DF-ResNet(96.88%)>DenseNet-169(95.72%)>Oct\_ResNet-28(95.42%)>ResNet-28(93.85%)>EfficientNet(91.67%)>VGG-19(79.39%)。同时,模型整体的参数计算量梯度为:VGG-19(143.67 MiB)>ResNet-28(19.14 MiB)>DF-ResNet(16.52 MiB)>Oct\_ResNet-28(15.99 MiB)>DenseNet-169(14.15 MiB)>EfficientNet(5.29 MiB)。尽管轻量化 EfficientNet 模型具有最少的参数计算量,但其图像分类识别能力相对较弱。DF-ResNet 与 ResNet-28 模型在结构层深度上相同,然而 DF-ResNet 的参数计算量比 ResNet-28 减少了 2.62 MiB,并且分类精度显著高于 ResNet 模

型。说明采用“双重过滤”结构对模型进行改进,既能降低整体参数计算量,同时还能提升近似结构图像的分类识别能力,使模型准确率得到显著提升。

此外,与未加入CBAM机制的Oct\_ResNet-28相比,DF-ResNet模型的参数计算量仅增加了0.53 MiB,但分类准确率提高了1.46%。这表明加入CBAM注意力机制虽然会略微增加模型的复杂度,但可以较大幅度提升准确率,这种权衡是可以接受的。

### 3.5 分析与讨论

考虑到OCT一般对于组织成像深度在2~3 mm,但针对切缘诊断,这种技术仍然是一种有效的工具。首先,OCT可以用于大面积组织的快速扫描,寻找可疑的肿瘤生长区域,跳过了对乳腺组织进行广泛组织学处理的繁琐过程。第二,OCT光束可以通过细针进行最小程度的侵入性输送,如细针穿刺或核心针生物检查。这些技术运用于图像引导,以在活检前将医生引导至可疑区域,从而减少了采样误差。第三,手持OCT探头是一种有效的实时无创性成像技术,通过结合使用深度学习算法,如DF-ResNet算法模型,可以自动识别和分析OCT图像中的特征,进一步提高手持OCT探头在乳腺癌术中切缘诊断中的准确性和效率,从而辅助医生在手术过程中对乳腺癌切缘进行更精确的评估。在训练阶段,将大量带有标签的OCT图像输入到DF-ResNet模型中对模型进行训练,使其能够识别正常组织与肿瘤组织之间的差异。虽然DF-ResNet模型的训练过程相对较长,但在预训练完成后,该模型识别单幅尺寸为224×224的乳腺组织OCT图像仅需6.67 ms,其识别速度足以满足术中切缘诊断的实时性要求。在应用阶段,将手持OCT探头实时采集的图像输入到预训练的DF-ResNet模型中,以快速确定探头处乳腺组织的类别。结合手持OCT探头和深度学习的术中切缘诊断方法可以提高肿瘤完全切除的成功率,降低复发风险和减少二次手术的需求。此外,该方法还可以减轻医生在术中判断切缘状态的压力,使他们更专注于手术操作本身,从而提高手术效果和患者满意度。

因此,使用手持OCT探头并结合深度学习技术进行乳腺癌术中切缘诊断具有潜力。这种方法有助于提高手术效果,确保患者获得更精确、安全的治疗方案。

## 4 结论

本文提出了一种基于残差网络改进的具有“双重过滤”结构的CNN模型DF-ResNet,实现对乳腺组织OCT图像的自动分类。DF-ResNet模型以残差网络为整体框架,利用OctConv替换部分卷积层并加入CBAM,使模型在特征提取时能够更关注于高频特征,减少低频特征占比,实现“双重过滤”结构,在很大程度上减少了信息冗余,提高了模型整体性能。利用DF-ResNet模型对乳腺组织OCT图像数据集中的3类样本数据进行训练学习、分类测试及多次调参,并使用数据增强和批量归一化等优化方式,最终分类精确度达到96.88%。对DF-ResNet模型的泛化能力进行验证,该模型在分类识别新的153张乳腺组织OCT图像实验中达到了94.12%的总分类精度。最后将DF-ResNet模型与DenseNet-169、VGG-19等经典CNN模型进行性能比对,结果表明,DF-ResNet模型在乳腺组织OCT图像分类实验中表现出更优异的性能,识别一张乳腺组织OCT图像仅需6.67 ms,足以满足术中切缘诊断的实时性要求,具有较好的鲁棒性和实时性。在未来应用阶段,可以配合手持OCT探头,将实时采集的图像输入到预训练的DF-ResNet模型中,以快速确定探头处乳腺组织的类别,提高肿瘤完全切除的成功率,为乳腺癌切缘实时诊断提供技术支持。

### 参考文献

- [1] LIMA S M, KEHM R D, TERRY M B. Global breast cancer incidence and mortality trends by region, age-groups, and fertility patterns[J]. *eClinicalMedicine*, 2021, 38: 100985.
- [2] HU K, DING P, WU Y, et al. Global patterns and trends in the breast cancer incidence and mortality according to sociodemographic indices: an observational study based on the global burden of diseases[J]. *BMJ Open*, 2019, 9(10): e028461.
- [3] ZHU Wei. Clinical value of breast-conserving surgery for early breast cancer[J]. *Henan Journal of Surgery*, 27(6): 82-84. 竹伟. 早期乳腺癌保乳手术的临床价值研究[J]. *河南外科学杂志*, 2021, 27(6): 82-84.
- [4] WILKE L G, CZECHURA T, WANG C, et al. Repeat surgery after breast conservation for the treatment of stage 0 to II breast carcinoma: a report from the National Cancer Data Base, 2004-2010[J]. *JAMA Surgery*, 2014, 149(12): 1296-1305.

- [5] JEEVAN R, CROMWELL D A, TRIVELLA M, et al. Reoperation rates after breast conserving surgery for breast cancer among women in England: retrospective study of hospital episode statistics[J]. *Bmj British Medical Journal*, 2012, 345: e4505.
- [6] HU Jintao, LAI Meina, CHEN Jingwen, et al. Application value of frozen pathological section technique in assessing the status of margins in breast-conserving surgery for breast cancer[J]. *Medical Innovation of China*, 2021, 18(23): 123-127.  
胡锦涛,赖美娜,陈敬文,等.冰冻病理切片技术在评估乳腺癌保乳手术中切缘状态的应用价值[J]. *中国医学创新*, 2021, 18(23): 123-127.
- [7] LIU Jianying, BU Hong. Pathological evaluation of breast-conserving margin[J]. *Chinese Journal of Bases and Clinics in General Surgery*, 2018, 25(2): 134-137.  
柳剑英,步宏.保乳切缘的病理学评估[J]. *中国普外基础与临床杂志*, 2018, 25(2): 134-137.
- [8] TANG Chaoyi, ZENG Jian. Research progress of breast-conserving surgery for invasive breast cancer[J]. *Guangxi Medical Journal*, 2018, 40(7): 833-838.  
唐朝易,曾健.浸润性乳腺癌保乳手术切缘的研究进展[J]. *广西医学*, 2018, 40(7): 833-838.
- [9] SOUTH F A, CHANEY E J, MARJANOVIC M, et al. Differentiation of ex vivo human breast tissue using polarization-sensitive optical coherence tomography[J]. *Biomedical Optics Express*, 2014, 5(10): 3417-3426.
- [10] ERICKSON-BHATT S J, NOLAN R, SHEMONSKI N D, et al. In vivo intra-operative breast tumor margin detection using a portable OCT system with a handheld surgical imaging probe[C]. *SPIE*, 2014, 8935: 197-202.
- [11] WANG J, XU Y, MESA K J, et al. Complementary use of polarization-sensitive and standard OCT metrics for enhanced intraoperative differentiation of breast cancer[J]. *Biomedical Optics Express*, 2018, 9(12): 6519-6528.
- [12] BUTOLA A, PRASAD D K, AHMAD A, et al. Deep learning architecture “LightOCT” for diagnostic decision support using optical coherence tomography images of biological samples[J]. *Biomedical Optics Express*, 2020, 11(9): 5017-5031.
- [13] ZHU D, WANG J, MARJANOVIC M, et al. Differentiation of breast tissue types for surgical margin assessment using machine learning and polarization-sensitive optical coherence tomography[J]. *Biomedical Optics Express*, 2021, 12(5): 3021-3036.
- [14] ZHOU Feiyan, JIN Linpeng, DONG Jun. Review of convolutional neural network[J]. *Chinese Journal of Computers*, 2017, 40(6): 1229-1251.  
周飞燕,金林鹏,董军.卷积神经网络研究综述[J]. *计算机学报*, 2017, 40(6): 1229-1251.
- [15] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [16] GU Y, LU X, YANG L, et al. Automatic lung nodule detection using a 3D deep convolutional neural network combined with a multi-scale prediction strategy in chest CTs[J]. *Computers in Biology and Medicine*, 2018, 103: 220-231.
- [17] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[C]. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings*, 2010: 249-256.
- [18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770-778.
- [19] CHEN Y, FAN H, XU B, et al. Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution[C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 3435-3444.
- [20] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [C]. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 3-19.
- [21] LIN M, CHEN Q, YAN S. Network in network[J/OL]. *arXiv preprint arXiv:1312.4400*, 2013.
- [22] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]. *International Conference on Machine Learning, PMLR*, 2015: 448-456.
- [23] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks [C]. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings*, 2011: 315-323.
- [24] HAN J, MORAGA C. The influence of the sigmoid function parameters on the speed of backpropagation learning [C]. *From Natural to Artificial Neural Computation: International Workshop on Artificial Neural Networks Malaga-Torremolinos*, 1995: 195-201.
- [25] Ganymede™ 系列 SD-OCT 系统 [EB/OL]. [2023-05-10]. [https://www.thorlabs.com/newgrouppage9.cfm?objectgroup\\_id=8214#12499](https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=8214#12499).
- [26] GU Y, LU X, ZHANG B, et al. Automatic lung nodule detection using multi-scale dot nodule-enhancement filter and weighted support vector machines in chest computed tomography[J]. *PLoS One*, 2019, 14(1): e0210551.
- [27] KHAN S U, ISLAM N, JAN Z, et al. A novel deep learning based framework for the detection and classification of



- breast cancer using transfer learning[J]. *Pattern Recognition Letters*, 2019, 125: 1–6.
- [28] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. *Journal of Big Data*, 2019, 6(1): 1–48.
- [29] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: visual explanations from deep networks via gradient-based localization[C]. *Proceedings of the IEEE International Conference on Computer Vision*, 2017: 618–626.
- [30] RUDER S. An overview of gradient descent optimization algorithms[J/OL]. arXiv preprint arXiv:1609.04747, 2016.
- [31] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 4700–4708.
- [32] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J/OL]. arXiv preprint arXiv:1409.1556, 2014.
- [33] TAN M, LE Q. Efficientnet: rethinking model scaling for convolutional neural networks[C]. *International Conference on Machine Learning*, PMLR, 2019: 6105–6114.

## Classification Method of Breast Tissue OCT Images Based on a Double Filtering Residual Network

DING Lihao, GAO Zhishan, ZHU Dan, YUAN Qun, GUO Zhenyan

(*School of Electronic Engineering and Optoelectronic Technology, Nanjing University of Science and Technology, Nanjing 210094, China*)

**Abstract:** With the emergence of Optical Coherence Tomography (OCT) technology and the rapid development of computer hardware, researchers have been attempting to utilize computer-aided to identify breast tissue OCT images. This study proposes a Convolutional Neural Network (CNN) model based on residual network enhancement for auxiliary diagnosis of OCT breast tissue images. The proposed method employs ResNet-18 framework as the basis to prevent gradient disappearance or explosion. In addition, to enhance the efficiency and regularization of the model, the original  $7 \times 7$  convolutional layer was substituted with a series of three cascading layers, each having a dimension of  $3 \times 3$ . This design choice allows for a nonlinear decomposition of the original  $7 \times 7$  layer while preserving the same receptive field size. As a result, it accomplishes two advantages: a reduction in computational cost associated with model parameters and acting as an implicit regularization technique. Subsequently, a Convolutional Block Attention Module (CBAM) was introduced after each set of cascaded small convolutional layers and the final convolutional layer. This module integrates a spatial attention module, which focuses on capturing spatial dependencies, and a channel attention module, which emphasizes informative channels, thereby serving as the initial stage of filtering and enhancing the discriminative capabilities of the network. At the same time, Octave Convolution (OctConv) is employed to substitute the  $3 \times 3$  convolutional layers in the original model. The convolutional kernel of OctConv has the capability to partition the input image sample data into four parts based on the low-frequency dimension ratio parameter within its structure. This functionality allows the network to dynamically balance high and low-frequency components during the process of extracting image features as the secondary filtering stage. After this, the Global Average Pooling (GAP) layer is used instead of the fully connected layer to reduce the computation of network parameters, and the structure of the network is regularized to prevent overfitting of the models. Ultimately, a residual network model “Double Filtering” ResNet (DF-ResNet) is constructed. The “double filtering” structure can not only reduce the overall parameter computation of the model, but also focus on high-frequency components with rich structural information during feature extraction. By decreasing the ratio of low-frequency constituents, the reduction of informative duplication is attained, resulting in an enhancement of the model’s proficiency towards categorization and identification of images manifesting proportional configurations. The proposed DF-ResNet model is employed to train and classify the OCT image dataset of three breast tissue types. It conducts multiple tuning tests and optimization techniques such as data augmentation and batch normalization, achieving an overall classification accuracy of 96.88%. After conducting three comparative experiments, the performance of the DF-ResNet model has been validated. In the first experiment, the DF-ResNet model was compared with the ResNet-28 model with an equivalent number of layers. The

experimental results showed that the replacement of some convolutional layers with OctConv allows the model to focus on high-frequency components during the image feature extraction process. This led to an improvement in the model's ability to classify and recognize images with similar structures and ultimately culminating in an overall increase in classification accuracy. Additionally, it is important to note that the use of OctConv did not negatively impact the overall convergence speed of the model. In the second experiment, the DF-ResNet model was compared with the Oct\_ResNet-28 model, which incorporated OctConv as a modification to improve its performance. The experimental results validated that the DF-ResNet model effectively filtered out low-frequency information. In the third experiment, the performance of the DF-ResNet model was assessed against established CNN models such as DenseNet-169 and VGG-19. The DF-ResNet model not only boasts a reduced parameter count but also demonstrated superior classification accuracy compared to several traditional CNN models. For the classification of OCT images of breast tissue, the DF-ResNet model displayed exceptional performance, robustness, and real-time processing capabilities. As a result, it is well-suited for providing technical support for real-time margin diagnosis in the clinical applications of breast cancer.

**Key words:** Optical coherence tomography; Image classification; High-frequency information; Octave convolution; Attention mechanism; Deep learning

**OCIS Codes:** 100.3008; 100.6950; 110.4500; 200.3050; 200.4260