

计算机辅助设计与图形学学报

*Journal of Computer-Aided Design & Computer Graphics*

ISSN 1003-9775, CN 11-2925/TP

## 《计算机辅助设计与图形学学报》网络首发论文

题目：集成多尺度微调卷积神经网络下的甲状腺结节良恶性识别  
作者：梁嘉炜，邱桃荣，周爱云，徐盼，谢学梅，付豪  
收稿日期：2020-02-28  
网络首发日期：2020-10-15  
引用格式：梁嘉炜，邱桃荣，周爱云，徐盼，谢学梅，付豪. 集成多尺度微调卷积神经网络下的甲状腺结节良恶性识别. 计算机辅助设计与图形学学报.  
<https://kns.cnki.net/kcms/detail/11.2925.tp.20201015.1437.004.html>



**网络首发：**在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

**出版确认：**纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

# 集成多尺度微调卷积神经网络下的甲状腺结节良恶性识别

梁嘉伟<sup>1)</sup>, 邱桃荣<sup>1)\*</sup>, 周爱云<sup>2)</sup>, 徐盼<sup>2)</sup>, 谢学梅<sup>1)</sup>, 付豪<sup>1)</sup>

<sup>1)</sup> (南昌大学信息工程学院 南昌 330031)

<sup>2)</sup> (南昌大学第一附属医院 南昌 330006)

(1318457129@qq.com)

**摘要：**针对由于训练图像样本较少与忽略多尺度的结构和纹理信息而导致分类性能不佳问题，为提升甲状腺结节良恶性诊断准确率，提出了集成多尺度微调卷积神经网络下的甲状腺结节超声图像识别算法。首先将图像转换成具有 3 种不同尺度信息作为输入数据，以便模型能够学习到图像不同尺度的特征信息，提高模型的特征提取能力。其次，通过优化 3 种预训练模型(AlexNet, VGG16 和 ResNet50)的全连接层结构和迁移学习与微调策略，构建了 3 种不同尺度的 9 个微调模型，让模型能够更好地学习源域(ImageNet)和目标域(甲状腺超声图像)上的特征差异。最后选择最优的微调模型组合并通过对模型输出类别概率的加权融合方法得到最终的集成模型，利用模型的多样性进一步提升分类性能。文中算法在真实采集的数据集上和其他算法进行对比实验，得到甲状腺结节良恶性识别的准确率为 96.0%，敏感度为 94.1%，特异度为 97.7%，AUC 为 0.98。实验结果表明，该算法在这些指标上均优于传统机器学习算法和当前甲状腺结节良恶性识别领域中的其他算法，能够有效地提取出互补的视觉特征信息，具有令人满意的分类性能。

**关键词：**集成模型；多尺度特征；卷积神经网络；迁移学习；甲状腺超声图像

中图法分类号：TP391.41 DOI: 10.3724/SP.J.1089.2020.18254

## Ensemble of Multiscale Fine-Tuning Convolutional Neural Network for Recognition of Benign and Malignant Thyroid Nodules

Liang Jiawei<sup>1)</sup>, Qiu Taorong<sup>1)\*</sup>, Zhou Aiyun<sup>2)</sup>, Xu Pan<sup>2)</sup>, Xie Xuemei<sup>1)</sup>, and Fu Hao<sup>1)</sup>

<sup>1)</sup> (School of Information Engineering, Nanchang University, Nanchang 330031)

<sup>2)</sup> (First Affiliated Hospital of Nanchang University, Nanchang 330006)

**Abstract:** Aiming at the problem of poor classification performance due to the small number of training image samples and ignoring multiscale structure and texture information, in order to improve the accuracy of diagnosis of benign and malignant thyroid nodule, this paper proposes a method for thyroid nodule ultrasound image recognition based on ensemble of multiscale fine-tuning convolutional neural networks. First, the image is converted into three different scales of information as input data, so that the model can learn the feature information of different scales of the image, and improve the feature extraction ability of the model.

收稿日期：2020-02-28；修回日期：2020-07-01。基金项目：江西省重点研发计划(20181BBG70031)。梁嘉伟(1996—)，男，硕士研究生，主要研究方向为医学图像处理、机器学习；邱桃荣(1964—)，男，博士，教授，博士生导师，CCF 高级会员，论文通讯作者，主要研究方向为数据挖掘与模式识别、智能信息处理、数据智能；周爱云(1960—)，女，博士，主任医师，主要研究方向为超声诊断；徐盼(1989—)，男，硕士，主治医师，主要研究方向为心血管超声及超声造影；谢学梅(1994—)，女，硕士研究生，主要研究方向为医学图像处理、机器学习；付豪(1997—)，男，硕士研究生，主要研究方向为医学图像处理、机器学习。

Secondly, nine fine-tuning models of three different scales were constructed by optimizing the full-connection layer structure of three kinds of pretraining models (AlexNet, VGG16 and ResNet50) and the transfer learning and fine-tuning strategy, so that the model could better learn the characteristic differences of source domain (ImageNet) and target domain (thyroid ultrasound image). Finally, the optimal fine-tuning model combination is selected and the final integration model is obtained by the weighted fusion method of model output category probability, and the classification performance is further improved by utilizing the diversity of models. The proposed algorithm was compared with other algorithms on the real data set, and the accuracy, sensitivity, specificity and area under curve (AUC) of benign and malignant thyroid nodules were 96.0%, 94.1%, 97.7% and 0.98. The experimental results show that the algorithm is superior to the traditional machine learning algorithm and other algorithms in the field of benign and malignant thyroid nodule identification, and can effectively extract complementary visual feature information with satisfactory classification performance.

**Key words:** ensemble model; multiscale feature; convolutional neural network; transfer learning; thyroid ultrasound image

近些年来,甲状腺疾病的发病率呈上升趋势,在高分辨率超声检查中,成人的发现率高达68%<sup>[1]</sup>.甲状腺结节的良恶性诊断是治疗甲状腺疾病的关键.但是当前超声影像学中的癌变结节判据尚不完善,甲状腺结节中的回声模式变化会大大地限制医生的诊断能力<sup>[2]</sup>.因此,迫切地需要一种计算机辅助医疗诊断方法,辅助医生得出更准确的诊断结果.

为了帮助医生提高临床诊断的准确性,许多学者希望通过传统机器学习算法进行甲状腺结节良恶性诊断. Iakovidis 等<sup>[3]</sup>采用一种模糊局部二值模式(fuzzy local binary patterns, FLBP)特征对甲状腺超声图像的纹理进行分类识别,其虽然相对于局部二值模式(local binary patterns, LBP)特征提升了对噪声的鲁棒性,但是准确率(accuracy, ACC)只有84%,识别效果不理想. Acharya 等<sup>[4]</sup>对甲状腺三维超声图像使用复小波变换滤波器提取特征,以分类甲状腺肿瘤,虽然在特定数据集上有较好的效果,但是复小波变换滤波器的构建过程较为复杂. Lyra 等<sup>[5]</sup>用灰度共生矩阵(gray level co-occurrence matrix, GLCM)表征甲状腺结节的纹理特征,提取不同方向的对比度、相关性、能量和均匀性等纹理特征,但其往往侧重提取图像的全局纹理特征而忽视局部纹理特征. Chen 等<sup>[6]</sup>提出了将纹理特征与

病理学特征相结合,采用支持向量机(support vector machine, SVM)对甲状腺结节进行分类,证明了通过多特征融合可以获得较好的分类结果. 传统机器学习分类算法的主要工作是对甲状腺超声图像进行纹理等特征的提取,并且结合 SVM 进行分类;但是,其提取高质量的有效特征十分困难与复杂,多依赖于人工的选择,工作量大且效率较低.

深度学习由于其易学性、通用性以及高效性而适用于计算机辅助医疗<sup>[7]</sup>. 基于卷积神经网络的深度学习算法可以通过引入局部连接、权值共享、池化操作和非线性激活等方法,允许网络从数据中自动学习特征,避免了传统机器学习中需要人工提取特征的缺陷,因此近年来深度学习成为甲状腺超声图像分类研究领域的一大热点. Ma 等<sup>[8]</sup>提出一种级联的深层卷积神经网络对甲状腺超声图像进行分类,其准确率(accuracy, ACC)为83.0%. Guan 等<sup>[9]</sup>使用深度卷积神经网络对甲状腺超声图像分类,其 ACC=90.4%. 但是,深度卷积神经网络需要大量数据样本训练,以完成目标的准确分类. 而在医学影像领域,收集大样本是一件非常困难的事情,学习样本不足会使深度学习模型学习到的特征不具有鲁棒性和代表性,从而导致模型的泛化能力较差.

为解决因数据集数量较小而导致模型训练困

难的问题, 使用迁移学习是一种新颖并且效果较好的方法. Moussa 等<sup>[10]</sup>使用经过 ImageNet 预训练的 ResNet50 进行迁移学习, 然后对甲状腺超声图像分类, 在其私有数据集上 ACC=97.3%, 敏感度 (sensitive, SENS) 为 80.7%, 特异度 (specificity, SPEC) 为 64.17%. Chi 等<sup>[11]</sup>使用经过预训练的 GoogLeNet 提取深层特征, 之后使用代价敏感随机森林分类器对甲状腺超声图像进行分类, 在其地区健康数据集上 ACC=96.3%, SENS=86%, SPEC=99%. 该算法虽然 ACC 较高, 但是 SENS 和 SPEC 的差异较大, 对良恶性图像的识别能力偏差较大; 并且单一尺度模型的识别特征能力有限, 导致模型的分类性能和泛化能力不佳.

因此, 针对上述问题与研究现状, 本文提出了集成多尺度微调卷积神经网络下的甲状腺结节超声图像识别算法, 主要在 3 个层面上进行了一定的改进. 首先将图像转换成具有 3 种不同尺度信息的图像作为输入数据, 能够让模型提取到图像不同尺度的特征信息, 提高模型的特征提取能力. 其次, 使用 3 种不同尺度的图像微调预训练模型 (AlexNet, VGG16 和 ResNet50) 得到 9 个微调模型的过程中, 重新构建了这 3 个预训练模型的全连接层结构, 并优化了迁移学习与微调的策略, 让模型能够更好地学习到源域 (ImageNet) 和目标域 (甲状腺超声图像) 上的特征差异. 最后, 在此基础上从 9 个微调模型中以微调模型的 ACC 作为集成模型的遴选依据, 选择最优的微调模型组合. 通过对模型输出类别概率的加权融合方法得到最终的集成模型, 并利用模型的多样性提取出互补的视觉特征信息, 进一步提高模型的性能. 在真实采集的数据集上, 本文算法与传统机器学习算法和当前甲状腺结节良恶性识别领域中的其他算法进行对比实验, 各项分类性能指标均优于其他算法, 具有令人满意的分类性能.

## 1 本文方法

### 1.1 甲状腺结节超声图像数据集

实验所用数据样本均为真实采集的数据, 在有经验的临床医师的指导下, 取得了如图 1 所示超声环境下甲状腺恶性结节和良性结节的医学图像. 共抽取 124 位甲状腺结节患者的甲状腺超声图像, 124 位患者均通过临床手术得出病理结果, 确保标签的准确性. 在取得的 1344 幅原始图像中, 良性结节 681 幅, 恶性结节 663 幅. 将样本的 80% 作为训练集, 共 1075 幅图像, 其中, 良性结节 545 幅, 恶性结节 530 幅; 将样本的 20% 作为测试集, 共 269 幅图像, 其中, 良性结节 136 幅, 恶性结节 133 幅. 实验数据集的统计如表 1 所示.



a. 恶性结节 b. 良性结节

图 1 数据集图像

表 1 数据集统计

数据集	良性结节	恶性结节	总数
训练集	545	530	1075
测试集	136	133	269

### 1.2 甲状腺超声图像的预处理

(1) 提取感兴趣区域 (region of interest, ROI) 并标准化. 为了提高模型对甲状腺超声图像的识别能力, 需要对输入图像进行预处理操作. 首先将甲状腺结节提取为如图 2 所示固定大小 (128×128) 的 ROI, 病变的位置和边界由影像学专家标记; 然后使用零均值处理减去图像的均值, 从而标准化图像.



图 2 提取 ROI 并标准化图像

(2) 数据增强. 深度学习模型的性能非常依赖



训练数据的数量, 当有数量较大的训练数据时, 深度学习模型表现得更好<sup>[12]</sup>. 目前有一个很好的增加数据量的方法是使用随机的几何变换来扩充训练数据. 使用幅度占图像总宽度和高度 0.2 的随机偏移、随机剪裁和随机放缩, 并且使用随机旋转 30° 和随机水平翻转对图像进行数据增强至原数据的 4 倍. 为了实现数据增强, 使用了 Keras 的图像生成器批量生成图像数据, 并实现了实时数据转换.

### 1.3 多尺度图像的提取

当前甲状腺结节识别领域中大多数算法只使用单一尺度的图像作为模型的输入图像进行训练, 在训练过程中模型提取特征的视角尺度单一, 容易忽略甲状腺结节超声图像丰富的细节信息. 为了能够丰富图像的多尺度特征信息, 本文首先在图像的中心提取出如图 3 所示, 图像尺度为  $128 \times 128$  (Scale 1),  $64 \times 64$  (Scale 2) 和  $32 \times 32$  (Scale 3) 的图像块; 然后使用这 3 种不同尺度的图像分别对 3 种不同的预训练模型 (AlexNet, VGG16 和 ResNet50) 进行迁移训练, 并微调得到 9 个能提取图像不同尺度特征的微调模型. 旨在提高模型的特征提取能力, 同时也为后续模型的集成与融合提供基础模型.

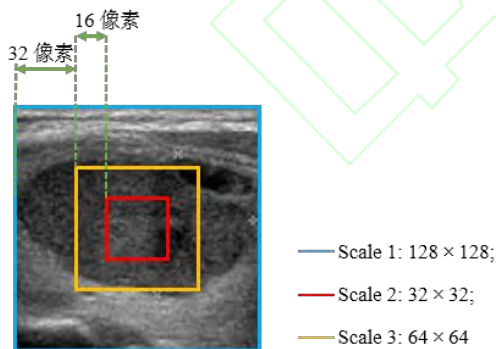


图 3 多尺度图像提取

### 1.4 多样化的深度卷积神经网络

甲状腺超声图像分类的问题之一在于当前算法使用单一结构的卷积神经网络进行图像识别, 忽略了多尺度的结构和纹理信息. 因此只凭一种卷积神经网络进行图像识别往往分类性能不佳. 而本文选取 3 种不同结构的深度卷积神经网络作为预训练模型, 以提取出多样化的特征信息, 提高

对图像的认识能力.

AlexNet, VGG16 和 ResNet50 在 ImageNet 图像分类任务上的表现优异, 而且甲状腺结节超声图像作为医疗图像具有特殊性, 难以收集到充足的训练样本数量, 并且与自然图像在特征方面的差异较大. 这种情况下预训练模型包含更多的是原数据集的特定特征, 而且训练集的数量较小, 在一个不太深的网络模型上训练一个分类器可能效果更好. 因此, 在构建集成模型过程中, 期望基础模型之间的结构差异性较大且性能表现差异性较小, 这样模型的融合能起到信息互补的作用, 表现出更好的分类性能.

AlexNet<sup>[13]</sup> 是比较经典和实用的网络之一, 在 ImageNet 自然图像库上实现了 15.3% 的 top-5 错误率. AlexNet 由 5 个卷积层和 3 个全连接层组成, 并且采用了 ReLU 激活函数代替常规的 Sigmoid 函数, 能够更快地训练并解决较深网络导致的梯度弥散问题. 采用 Dropout 随机失活来防止模型过拟合, 使用叠加的最大池化层避免平均池化层的模糊化效果, 并且步长比池化核的尺寸小, 池化输出层之间有重叠和覆盖, 进而提升了特征的丰富性. 采用局部相应归一化 (local response normalization, LRN) 对局部神经元的活动创建了竞争机制, 使得其中响应比较大的值变得相对更大, 并抑制反馈较小的神经元, 增强了模型的泛化能力.

VGG16<sup>[14]</sup> 在 ImageNet 自然图像库上实现了 9.9% 的 top-5 错误率. VGG16 由 5 个卷积块和 3 个全连接层构成, 使用 ReLU 激活函数和最大池化层. VGG16 整个网络十分规整, 都使用了同样大小的卷积核尺寸和最大池化尺寸, 并使用多个较小且尺寸相同的卷积核串联, 可以在线性变换的过程中产生更少的参数量, 让 VGG16 收敛的速度更快; 而且更小的卷积核带来隐式的正则化效果, 能够有效地减少过拟合现象.

ResNet50<sup>[15]</sup> 在 ImageNet 自然图像库上实现了 7.8% top-5 的错误率. ResNet50 总共有 16 个残差模块, 每个残差模块包含着 3 个卷积层, 共由 49 个卷

积层和 1 个全连接层组成, 使用 ReLU 激活函数与最大和平均池化层. ResNet50 的优势在于具有与其他网络不同的残差模块, 通过将前面若干层的数据输出直接跳过多层网络结构接入到后层的数据层输入部分, 这种跳跃式的结构可以显著地缓解随着网络层数加深而出现的随机梯度消失的问题, 使得这种网络体系结构在提取更加深层特征的同时, 有效地降低了过拟合现象.

与诸如 GoogLeNet, DenseNet 等结构庞大的网络相比, AlexNet, VGG16 和 ResNet50 的结构与深度非常适合训练样本较少的医学图像分类任务, 其过拟合现象较少、提取的特征质量较好、训练速度较快, 拥有更好的分类性能. 由于 AlexNet, VGG16 和 ResNet50 的结构差异较大, 非常符合选取模型的特点需求, 因此选择它们作为本文的主干模型.

使用这 3 种不同结构的深度卷积神经网络作为预训练模型, 能够提取出甲状腺超声图像多尺度的特征, 有效地解决因为只提取特征单一而影响分类性能的问题. 3 种预训练模型(AlexNet, VGG16 和 ResNet50)可以相互弥补自身的缺陷, 进一步提高图像识别能力, 它们的类激活(class activation map, CAM)热力图如图 4 所示, 其表示图像不同位置对于预训练模型决策的重要程度.

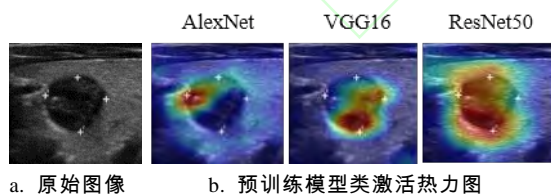


图 4 各预训练模型 CAM 热力图

从图 4 可知, 不同结构的预训练模型决策的依据也不同, 在这幅甲状腺超声图像中, AlexNet 更偏向提取甲状腺结节左上方的区域来识别其良恶性; VGG16 更偏向使用甲状腺结节右下方的区域来识别良恶性; 而 ResNet50 则更加依靠甲状腺结节左下方的区域来进行识别. 这 3 种预训练模型决策的依据不同, 能够对甲状腺结节不同尺度的特征有一个更好的把握.

## 1.5 迁移学习与微调卷积神经网络

因为医疗图像的特殊性, 训练样本数量较少, 在少量数据上训练深度学习模型是非常困难的. 处理这个问题可以通过使用预先训练好的卷积神经网络来得到改善<sup>[16]</sup>. 卷积神经网络的迁移学习首先在一个大量数据的源域(ImageNet)上训练, 然后在一个相关却不不同的目标域上训练并微调该卷积神经网络的权重来实现<sup>[17]</sup>. 研究表明, 即使在 2 个不相关的领域(自然和医学)之间, 迁移学习也是有作用的<sup>[18-19]</sup>. 使用预训练模型的参数来初始化权重并微调卷积神经网络, 可以有效地解决训练数据较少而导致的过拟合问题<sup>[20-21]</sup>.

在迁移学习与微调过程中, 本文对预训练模型结构和微调策略做了一定的改进. 在甲状腺识别方面, 目前大多使用的迁移学习是将预训练模型全连接层的输出节点个数更改为分类目标数, 然后重新训练预训练模型的全连接层, 以实现模型的迁移. 但是, 预训练模型是基于 ImageNet 数据集训练得到最终的模型, ImageNet 属于自然图像, 与甲状腺结节超声图像的差异性较大, 只有较低的特征层具有良好的泛化性能, 而较高层的特征层却相差较大, 导致迁移学习的效果不佳. 与传统算法相比, 本文首先优化改进了预训练模型的全连接层结构, 由于预训练模型的全连接层是为 ImageNet 数据集而构建的 1000 分类问题, 使用的全连接层的层数和每一层的神经元节点数较多, 在训练过程中需要计算大量的参数并且容易出现过拟合现象; 而在甲状腺结节良恶性的二分类问题上, 不需要如此复杂的神经网络作为模型的全连接层, 因此本文如图 5 所示重新优化并构建了这 3 个预训练模型(AlexNet, VGG16 和 ResNet50)的全连接层, 其中包含了 4 层具有 8 个神经元的神经网络, 并使用 Dropout 层和  $L_2$  正则化来降低模型的过拟合现象, 有效地减少了计算参数, 加快了训练的速度, 提高了模型分类性能.

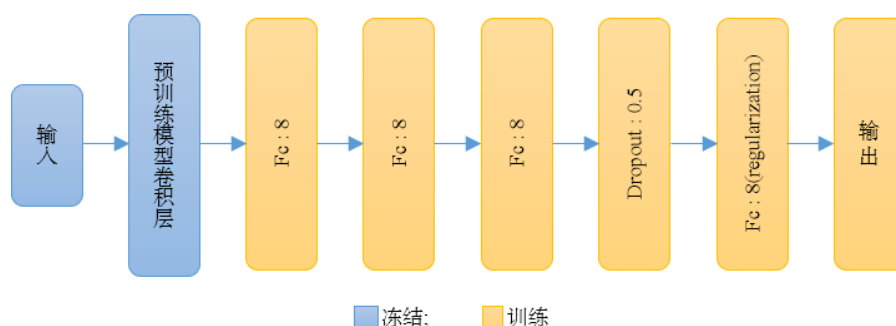


图 5 模型结构的改进

其次优化了微调的策略, 将模型的训练分为 2 个步骤完成.

Step1. 冻结预训练模型的卷积层, 只对重新构建的全连接层使用自适应学习率优化器 Adam<sup>[22]</sup>进行迁移训练, 如图 5 所示. 在进行微调之前对重新构建的全连接层进行训练是很重要的, 可以让模型在微调之前将全连接层的参数快速学习到适合模型微调的权值, 而不会破坏预训练模型卷积层所学习到的特征, 不仅可以加快训练速度, 还可以提升模型识别能力.

Step2. 对模型进行微调训练, 使用随机梯度下降(stochastic gradient descent, SGD)优化器<sup>[23]</sup>并设置较小的学习速率为  $10^{-4}$ , 动量为 0.9, 学习率

衰减为  $10^{-6}$  来微调模型的部分卷积层, 使其更具有数据特异性, 能够学习到甲状腺超声图像与 ImageNet 图像的差异特征; 并且进一步研究了不同模型最佳的微调层数, 选择合适的微调层数也是影响模型分类性能的重要因素之一. 微调层数太少会导致模型学习不到目标图像与预训练图像的差异; 微调层数太多又会导致模型破坏预先学习的特征, 使分类性能明显下降. 本文依次微调了预训练模型(AlexNet, VGG16 和 ResNet50)最后 4 个卷积层, 如图 6 所示. 通过实验分析, 得出了不同模型最佳的微调层数, 让模型能够更好地学习到源域(ImageNet)和目标域(甲状腺超声图像)上的特征差异, 进一步提高模型的性能.

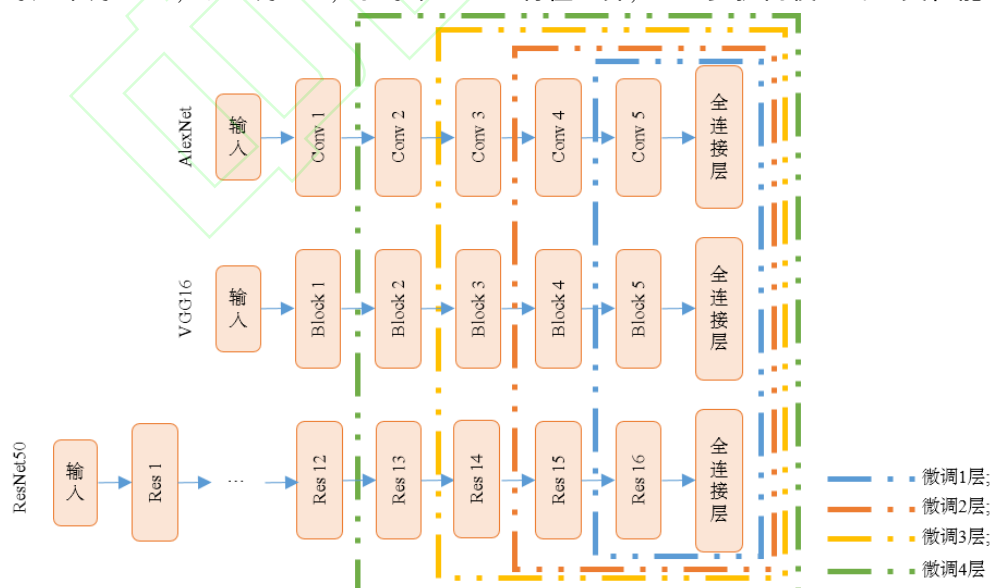


图 6 模型微调策略的改进

## 1.6 模型的集成与融合

在当前甲状腺结节识别领域中, 大多数算法只使用了某一种深度卷积神经网络, 但是单个模型提取特征的能力与考虑的因素有限, 利用的信

息不够全面, 导致分类性能不佳. 而本文提出了集成多尺度微调卷积神经网络下的甲状腺结节超声图像识别算法, 主要在以下几个方面做出了一定的改进.

本文首先提取了3种不同尺度的图像, 然后分别使用这3种不同尺度的图像迁移, 并微调训练3个不同结构的预训练模型(AlexNet, VGG16 和 ResNet50), 将得到的9个微调模型作为集成的基础模型. 在多个不同尺度和结构上微调的深度卷积神经网络能够提取出不同的视觉特征, 以表征来自不同视角的图像, 且考虑多个尺度上的结构和纹理信息, 这对甲状腺结节超声图像的分析非常重要.

其次在9个不同微调模型中选择其中一个子集组成一个集成模型的选择性很多, 选择最优子集进行组合是一个值得研究的问题. 选择子集模型的数量太多, 可能存在重复的信息, 会增大集成模型预测的时间, 降低预测的效率; 选择子集模型的数量太少, 又会降低模型的分性能. 将模型的ACC作为集成模型的遴选依据, 选出ACC最高的5个模型命名为 $A \sim E$ , 如表2所示.

表2 不同尺度下模型的ACC %

图像尺度	ALexNet	VGG16	ResNet50
Scale 1	A(85.7)	B(92.0)	C(91.8)
Scale 2	D(85.2)	E(86.6)	81.7
Scale 3	78.9	76.2	78.8

从表2可以看到,  $A \sim E$ 这5个微调模型的ACC较高, 从这5个ACC较高的微调模型中选取最优子集, 并通过模型输出类别概率的加权融合方法将它们形成一个集成模型. 在选取最优子集时, 可以发现 $B$ 和 $C$ 的ACC相比其他模型较高, 因此将 $B$ 和 $C$ 模型保留; 最终可以得出子集3种组合方式 $\{A, B, C\}$ ,  $\{D, B, C\}$ 和 $\{E, B, C\}$ . 通过选择最优的微调深度模型子集构建成集成模型, 在有效地在降低时间复杂度的同时, 具有令人满意的分类性能.

表3 不同模型组合下的ACC %

组合方式	ACC
$\{0.2A, 0.5B, 0.3C\}$	94.7
$\{0.1D, 0.5B, 0.4C\}$	93.3
$\{0.2E, 0.3B, 0.5C\}$	96.0

表3的结果表明, 使用 $\{0.2E, 0.3B, 0.5C\}$ 所集成的模型ACC最高, 达到96%; 并且这3种集成模型的ACC均要高于单独模型, 进一步说明集成模型可以综合不同模型的预测能力. 将甲状腺超声图像的粗粒度特征如甲状腺结节的轮廓、大小和细粒度特征如甲状腺结节的纹理结合起来, 从不同角度提取出视觉特征来表示图像, 可减少因为忽略多尺度的结构和纹理信息所带来的影响.

由于集成模型中的组合策略也会影响最终的预测性能, 因此本文进行了不同组合策略的对比实验, 包括对模型输出类别的投票、对模型输出类别概率的平均和对模型输出类别概率的加权的融合方法进行分析比较. 得出使用对模型输出类别概率的加权的融合方法可以得到最佳的分类性能, 更好地结合模型的多样性.

集成模型最终的预测流程如图7所示. 首先将原始图像转化为具有Scale 1和Scale 2这2种尺度的图像; 然后使用Scale 1尺度的图像对预训练模型VGG16和ResNet50进行迁移学习与微调, 得到微调模型 $B$ 和 $C$ . 使用Scale 2尺度的图像对预训练模型VGG16进行迁移学习与微调, 得到微调模型 $E$ ; 最后使用输出类别概率的加权融合方法赋予这3个微调模型不同的权重, 得到最终的集成模型 $\{0.2E, 0.3B, 0.5C\}$ .



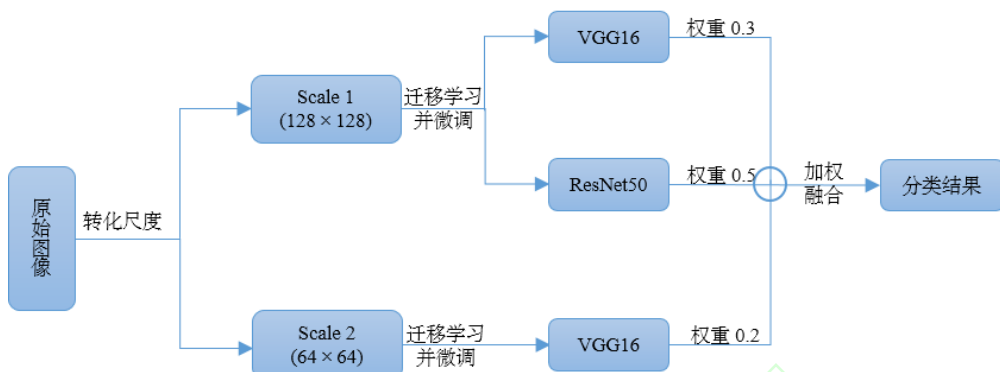


图 7 预测流程

## 2 实验结果与分析

所有实验均使用相同的测试集进行测试，测试集共有 269 幅甲状腺超声图像，其中良性结节 136 幅，恶性结节 133 幅。对于实验方法的分类性能采用 ACC, SENS 和 SPEC 进行评估，分别定义为

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$SENS = \frac{TP}{TP + FN} \quad (2)$$

$$SPEC = \frac{TN}{TN + FP} \quad (3)$$

其中，TP 是真阳性样本；TN 是真阴性样本；FP 是假阳性样本；FN 是假阴性样本。

### 2.1 非迁移学习和迁移学习

首先对非迁移学习和迁移学习的进行对比实验，以验证迁移学习的效果和可行性。对于非迁移学习，使用高斯分布随机产生一个较小的数字初始化模型权重，并训练模型的所有卷积层和全连接层。对于迁移学习，首先使用在 ImageNet 上预先训练好的权重作为模型的初始化参数，其次重新构建模型的全连接层使预训练模型适应的二分类问题(良性、恶性)，最后将模型的卷积层冻结，只训练重新构建的全连接层。非迁移学习和迁移学习均在同一训练集上训练，并在同一测试集上得到各个模型的 ACC，如图 8 所示。

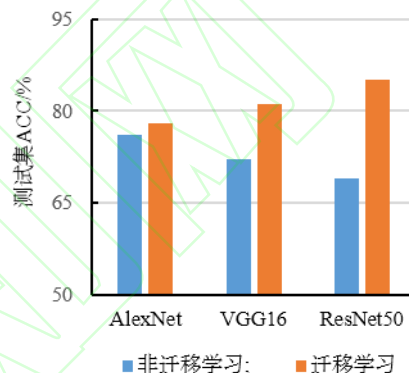


图 8 非迁移学习和迁移学习测试集 ACC

从图 8 可以看出，AlexNet, VGG16 和 ResNet50 在迁移学习下测试集的 ACC 均高于非迁移学习；并且模型层数越深，使用迁移学习的效果越明显。这是因为随着模型层数的加深，由于训练数据量的不足导致模型拟合现象更加严重，而使用 ImageNet 上预训练的权重可以让模型获得更好的初始化权重，使模型能够提取到对大多数分类任务都相似的浅层特征。实验结果证明，迁移学习可以有效地减少因为训练数据较少而带来的过拟合现象。

### 2.2 微调层数的影响

迁移学习也被称为特征提取机制和零微调策略<sup>[24]</sup>，虽然大多数分类任务的浅层特征相似，但是所需要识别的甲状腺超声图像与预训练集 ImageNet 上的自然图像的差异较大，在 ImageNet 上学习到的特征不足以很好地区分出甲状腺结节的良恶性，因此还需要在迁移学习的基础上对模型卷积层的最后几层进行微调，使模型学习到甲状腺超声图像与 ImageNet 差异较大的特征。但是，对于微调层数的选择也十分重要。为确定最佳的

微调层数, 本文冻结模型除微调层数外的卷积层, 只训练需要微调的卷积层和全连接层, 并且通过依次改变微调模型的卷积层数, 得出不同模型最佳的微调层数. 各个模型不同微调层数在测试集的实验结果如图 9 所示.

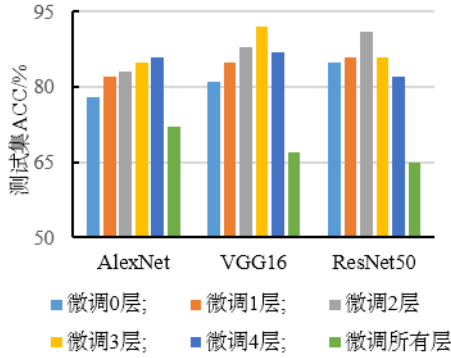


图 9 微调不同层数测试集 ACC

从图 9 中得出对于 AlexNet 最佳的微调层数是 4 层, VGG16 的最佳微调层数是 3 层, ResNet50 的最佳微调层数是 2 层. 这说明较浅的模型允许微调较多的卷积层数, 不容易发生过拟合现象, 而较深的模型可微调的卷积层数较少, 更容易发生过拟合现象. 而且在到达最佳微调层数之前, 随着微调层数的增加, 模型的 ACC 也在提高; 说明通过微调模型的最后几层卷积层, 可以让模型学习到目标图像与源域图像差异较大的特征. 但是, 当到达模型的最佳微调层数后, 再继续增加模型的微调层数就会破坏预训练模型所学习到的特征, 同时出现严重的过拟合现象, 导致 ACC 降低; 当微调所有层时, ACC 的下降更加明显.

### 2.3 单个模型和集成模型

针对单个模型忽略图像多尺度的结构与纹理特征的问题. 本文采用对模型输出类别概率的加权融合方法, 结合不同层数、不同尺度的预训练模型(AlexNet, VGG16 和 ResNet50), 以达到对图像的粗粒度特征如甲状腺结节的轮廓和大小、细粒度特征如甲状腺结节的纹理有较好的整体把握, 减少因为忽略多尺度的结构和纹理信息所带来的影响. 并且通过实验得到了最佳的权重比例, 使用最佳的权重比例对模型输出概率类别进行加权融合

得到集成模型. 分别将只具有 Scale 1 尺度的预训练模型(AlexNet, VGG16 和 ResNet50)与具有多尺度的集成模型 {0.2E, 0.3B, 0.5C} 进行对比实验, 结果如表 4 所示.

表 4 单个模型与集成模型对比实验

模型	ACC	SENS	SPEC
AlexNet <sup>[13]</sup>	85.7	85.3	85.7
VGG16 <sup>[14]</sup>	92.0	86.0	98.4
ResNet50 <sup>[15]</sup>	91.8	92.1	91.7
集成模型	96.0	94.1	97.7

从表 4 可以得出, 集成多尺度微调卷积神经网络的集成模型相对于单个模型在甲状腺结节超声图像的良恶性识别问题上的 ACC, SENS 和 SPEC 都有较大的提升, 在一定程度上降低了因单个模型忽略不同尺度的结构和纹理信息所带来的影响.

### 2.4 模型不同融合方法的对比

单个模型通常仅包含预测对象的部分信息, 通过一定规则融合各单项模型, 可通过包含更全面的预测信息来提高模型的性能. 集成模型中, 模型不同的结合策略也会影响最终的预测性能. 本文基于模型输出类别概率的加权融合方法, 在得到各个基础模型的输出类别概率值后, 通过加权平均的方法得到最终的输出类别. 由于集成模型所需要确定的加权系数只有 3 项, 所需要的计算开销不大, 并且基于学习方法确定最优的加权系数需要根据某种最优准则构造目标函数, 因此在约束条件下极小化目标函数, 以求得集成模型的加权系数. 但是, 构造目标函数和求解的过程较为复杂且具有一定的难度, 不易普及, 而且最优权重可能出现负值, 这往往与实际要求不符. 所以本文采用枚举的方法统计在不同加权系数下集成模型的性能, 使用最佳个体保留策略, 保留集成模型分类性能最佳的加权系数为最终各项的加权系数. 这样的做法计算步骤简单, 并且仍然可以得到分类性能优异的集成模型. 同时本文还对模型不同的组合策略进行了对比实验, 包括对模型输出类别的投票、对模型输出类别概率的平均和对模型输

出类别概率的加权的融合方法进行分析比较, 如表 5 所示.

表 5 不同融合方法对比实验 %

融合方法	ACC	SENS	SPEC
输出类别投票 <sup>[12]</sup>	94.4	91.9	96.9
输出类别概率平均 <sup>[14]</sup>	93.7	90.4	97.0
输出类别概率加权	96.0	94.1	97.7

通过表 5 实验结果可知, 使用对模型输出概率的加权的融合方法是更加合理的融合方法, 能够更好地结合模型的多样性, 发挥出其自身的优势. 根据实验结果表明, 这 3 种融合方法得到的集成模型分类性能均高于单个基础模型, 也在一定程度上证明了使用输出类别概率加权融合方法的可行性与优越性.

## 2.5 本文算法与传统机器学习算法

为了评估本文算法与传统机器学习算法的分类性能, 首先使用传统机器学习算法提取相应特征, 其中包括: (1) 均值局部二值模式(uniform local binary patterns, ULBP). 该特征描述了图像的局部结构, 具有灰度不变性、旋转不变性和较强的分类能力, 是图像分类中常用的纹理特征. (2) GLCM. 该特征是利用灰度空间的相关特性来分析纹理, 并计算其对比度、相异性、协同性、熵、相关性和能量这 6 个统计量作为特征. (3) Gabor 特征. 该特征与人类的视觉系统类似, 特别适合于纹理表示与判别, 也是图像分类中常用的算法. 然后使用 SVM 作为分类器, 该分类器学习的关键是找到支持向量, 支持向量之外远端样本点的价值较小, 通过设置惩罚参数, 可以较好地避免噪声支持向量对分界面的不利影响, 有利于处理样本不平衡和小样本集等问题<sup>[25]</sup>. 最后, 不同算法在相同测试集上进行测试, 分类性能的对比结果如表 6 所示.

表 6 本文算法与传统机器学习算法对比实验 %

算法	ACC	SENS	SPEC
ULBP <sup>[3]</sup>	79.2	75.0	83.5
GLCM <sup>[5]</sup>	72.3	73.5	70.6
Garbor <sup>[4]</sup>	73.8	79.4	68.4
本文	96.0	94.1	97.7

从表 6 可以看出, 本文算法在 ACC, SENS 和 SPEC 这 3 种性能指标上, 比传统机器学习算法均有较大的提升, 说明其分类性能要优于传统机器学习算法; 其且不需要人工选取和设计相应特征, 有效地降低了甲状腺结节良恶性识别的复杂性.

## 2.6 本文算法与其他文献算法

将本文提出的集成多尺度微调卷积神经网络下的甲状腺结节超声图像识别算法, 与其他深度学习算法在原理上进行对比. 例如, 文献[10]利用 ResNet50 进行迁移学习, 并微调第 1 个残差模块对甲状腺超声图进行识别; 文献[11]利用 InceptionV3 进行迁移学习, 并微调所有层后使用代价敏感随机森林对甲状腺超声图像进行识别; 文献[26]利用 VGG16 进行迁移学习, 并融合 LBP 和方向梯度的直方图(histogram of oriented gradients, HOG)对甲状腺超声图像进行识别.

本文主要从 3 个层面上做出了一定的改进. 首先, 本文使用了 3 种不同尺度的图像分别对 3 种不同的预训练模型(AlexNet, VGG16 和 ResNet50)进行迁移训练并微调, 让模型能够学习提取到图像不同尺度的特征信息, 提高了模型的特征提取能力.

其次, 在迁移学习过程中, 本文对预训练模型的全连接层结构和迁移学习的策略也做出了一定的改进. 这 3 种算法在迁移学习时只将全连接层的输出节点个数修改为分类目标数, 并没有对模型的结构进行优化. 而本文重新优化构建了这 3 个预训练模型(AlexNet, VGG16 和 ResNet50)的全连接层, 其庞大的全连接层替换为包含了 4 层只具有 8 个神经元的神经网络, 并使用 Dropout 层和  $L_2$  正则化来降低模型的过拟合现象, 有效地减少了计算参数, 加快了训练的速度; 同时优化了迁移学习策略. 在迁移学习训练过程中, 文献[26]只对模型的全连接层进行了训练, 并没有对模型的卷积层进行微调, 不能很好地学习到原始数据集和新数据集的差异特征. 文献[11,26]虽然都对模型进行了微调, 但是没有进一步地研究不同模型最佳的微调层数. 而本文将模型的迁移学习优化为 2 个步

骤完成. 首先在进行微调之前将全连接层的参数快速的训练学习到适合模型微调的权值, 在不破坏预训练模型卷积层所学习到的特征的前提下, 加快训练速度; 然后对模型进行微调训练, 并依次微调了预训练模型(AlexNet, VGG16 和 ResNet50)最后 4 个卷积层. 通过实验分析得出了不同模型最佳的微调层数, 让模型能够更好地学习到源域(ImageNet)和目标域(甲状腺超声图像)上的特征差异.

最后, 本文以多个微调模型的 ACC 作为集成模型的遴选依据, 通过选择最优的微调模型子集构建成一个集成系统, 在有效地降低时间复杂度的同时, 具有最优的分类性能. 而且使用对模型输出类别概率的加权的融合方法得到最佳的集成模型, 进一步提升集成模型的性能.

使用这 3 种其他文献算法与本文算法进行对比实验, 在同一测试集上测试如表 7 所示. 通过引入受试者工作特征曲线(receiver operating characteristic curve, ROC)及其曲线下的面积(area under curve, AUC)评估算法性能, 如图 10 所示.

表 7 本文算法与其他文献算法对比实验 %

算法	ACC	SENS	SPEC
文献[10]	86.6	77.9	95.5
文献[11]	77.0	64.7	89.5
文献[26]	89.5	87.5	91.7
本文	96.0	94.1	97.7

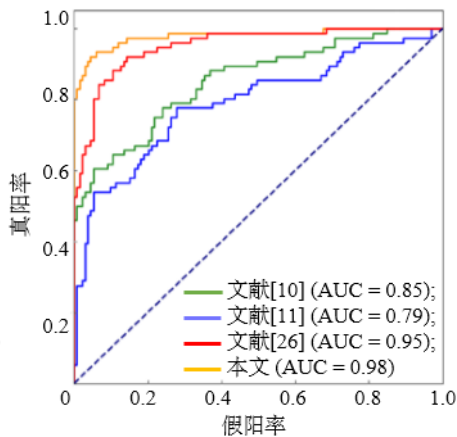


图 10 本文算法与其他文献算法 ROC 曲线

根据表 7 和图 10 进一步证明了本文算法的优越性, 不仅可以有效地减少因为训练数据量不足而导致的过拟合现象, 以及由于仅使用单一模型而忽略多尺度的结构和纹理信息的问题, 而且在甲状腺结节良恶性识别领域中的分类性能有较大的提升.

### 3 讨论

首先, 本文通过实验证实了迁移学习对初始化模型权重的影响, 使用在 ImageNet 上的预训练权重初始化模型相对于随机初始化模型有较大的提升, 能有效地避免过拟合现象; 并且随着预训练模型层数的增加, 迁移学习的提升效果也更加明显.

其次, 研究了微调层数对于模型的影响, 因为预训练数据集与甲状腺超声图像的差异较大, 需要对预训练模型的卷积层进行微调, 使模型对甲状腺超声图像更加敏感, 能够提取到甲状腺超声图像与自然数据集差异较大的特征. 实验证实对预训练模型的卷积层进行微调虽然可以很好地提升模型的性能, 但是对于不同的模型所能达到最佳性能的微调层数也不同, 对于层数较浅的预训练模型可以微调的层数较深, 不易发生过拟合现象; 而层数较深的预训练模型只能微调较少的层数, 过多的微调会破坏模型原先的特征提取能力, 并破坏预训练模型卷积层的参数而发生严重的过拟合现象, 模型的性能也会急剧降低.

最后, 通过实验证实了使用对模型输出类别概率的加权融合方法将不同尺度的微调模型相结合, 能够有效地提取甲状腺超声图像的多尺度特征. 将本文算法与传统机器学习算法和其他文献算法进行对比实验, 结果表明, 其在各项分类性能指标上均优于对比算法, 对提升甲状腺结节良恶性结节识别能力有一定的帮助. 本文提出的集成多尺度微调卷积神经网络下的甲状腺结节超声图像识别算法涉及图像尺度的转换、9 个预训练模型的微调 and 模型选择. 因此, 在训练阶段是非常耗时



的。但是在预测阶段,分类结果是通过 3 种不同尺度的微调模型正向计算得到的,速度相对较快,可以用于常规的临床工作流程。

通过分析如图 11 所示部分预测图像,可以看出良性结节和恶性结节在视觉上非常相似,都具有结节边界不清晰、内部结构为实性低回声和纹理不规则呈固态分布的特点。由于甲状腺结节的形态、边界、内部回声和钙化特征存在交叉重叠,没有具体的评价标准,因此非常难以识别。这表明,本文不仅需要进一步优化算法,使算法可以提取出更加深层次且有效的特征,还需要扩充数据集的图像数量,让算法有更多具有挑战性的学习样本来进行训练,以提升算法的识别能力。

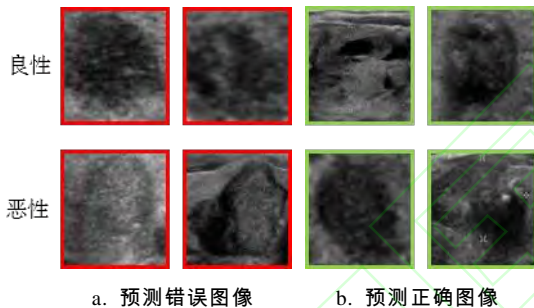


图 11 部分预测图像

## 4 结 论

综上所述,本文所提集成多尺度微调卷积神经网络下的甲状腺结节超声图像识别算法,在真实采集的数据集上得到了  $ACC=96.0\%$ ,  $SENS=94.1\%$ ,  $SPEC=97.7\%$ ,  $AUC=0.98$  的实验结果。比其他算法得到了更高的客观识别率,对识别甲状腺良恶性结节有很好的提升效果;可以有效地改善辅助医疗诊断的性能,为临床医学中甲状腺结节良恶性的诊断提供帮助。但是,本文算法比较依赖人工标注的训练数据,在制定训练数据集的过程中有一定的复杂性。

本文进一步工作为:(1) 继续持续优化模型结构和分类算法提高模型的性能。(2) 扩充数据集以训练对甲状腺超声图像更加敏感的深度卷积神经网络。(3) 进一步结合 TI-RADS 评分系统对甲状腺

结节的风险进行细分评估。(4) 研究出一种可以自动标识 ROI 的算法,避免算法对人工标注的依赖。

## 参考文献(References):

- [1] Song Yanfang, Li Rui. Progress in ultrasonic diagnosis of thyroid carcinoma [J]. Journal of Imaging Research and Medical Applications, 2019, 3(14): 2-3(in Chinese)  
(宋艳芳, 李睿. 甲状腺癌超声诊断的研究进展[J]. 影像研究与医学应用, 2019, 3(14): 2-3)
- [2] Han R L, Wang J, Zhang F J, *et al.* Ultrasound risk assessment combined with molecular markers of galectin-3, c-MET, HBME-1 and CK19 for diagnosis of malignant and benign thyroid nodules[J]. Pathology & Oncology Research, 2019, 25(3): 1075-1081
- [3] Iakovidis D K, Keramidas E G, Maroulis D. Fusion of fuzzy statistical distributions for classification of thyroid ultrasound patterns[J]. Artificial Intelligence in Medicine, 2010, 50(1): 33-41
- [4] Acharya U R, Sree S V, Swapna G, *et al.* Effect of complex wavelet transform filter on thyroid tumor classification in three-dimensional ultrasound[J]. Proceedings of the Institution of Mechanical Engineers Part H-Journal of Engineering in Medicine, 2013, 227(3): 284-292
- [5] Lyra M E, Lagopati N, Charalabistou P, *et al.* Texture characterization in ultasonograms of the thyroid gland[C] // Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine. Los Alamitos: IEEE Computer Society Press, 2010: 1-4
- [6] Chen S J, Chang C Y, Chang K Y, *et al.* Classification of the thyroid nodules based on characteristic sonographic textural feature and correlated histopathology using hierarchical support vector machines[J]. Ultrasound in Medicine & Biology, 2010, 36(12): 2018-2026
- [7] Zhang Peng, Xu Xinnan, Wang Hongwei, *et al.* Computer-aided lung cancer diagnosis approaches based on deep learning[J]. Journal of Computer-Aided Design & Computer Graphics, 2018, 30(1): 90-99(in Chinese)  
(张鹏, 徐欣楠, 王洪伟, 等. 基于深度学习的计算机辅助肺癌诊断方法[J]. 计算机辅助设计与图形学学报, 2018, 30(1): 90-99)
- [8] Ma J L, Wu F, Jiang T A, *et al.* Cascade convolutional neural networks for automatic detection of thyroid nodules in ultrasound images[J]. Medical Physics, 2017, 44(5): 1678-1691
- [9] Guan Q, Wang Y J, Du J J, *et al.* Deep learning based classification of ultrasound images for thyroid nodules: a large scale of pilot study[J]. Annals of Translational Medicine, 2019, 7(7): 137
- [10] Moussa O, Khachnaoui H, Guetari R, *et al.* Thyroid nodules classification and diagnosis in ultrasound images using fine-tuning deep convolutional neural network[J]. International Journal of Imaging Systems and Technology, 2019, 30(1): 185-195
- [11] Chi J N, Walia E, Babyn P, *et al.* Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network[J]. Journal of Digital Imaging, 2017, 30(4): 477-486

- [12] Yang H W, Hsu H C, Yang C K, *et al.* Differentiating between morphologically similar species in genus *Cinnamomum* (Lauraceae) using deep convolutional neural networks[J]. *Computers and Electronics in Agriculture*, 2019, 162: 739-748
- [13] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90
- [14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[OL]. [2020-02-28]. <https://arxiv.org/pdf/1409.1556.pdf>
- [15] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2016: 770-778
- [16] Oquab M, Bottou L, Laptev I, *et al.* Learning and transferring mid-level image representations using convolutional neural networks[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2014: 1717-1724
- [17] Bengio Y, Guyon I, Dror G, *et al.* Deep learning of representations for unsupervised and transfer learning[J]. *Proceedings of Machine Learning Research*, 2012, 27: 17-36
- [18] Chen H, Ni D, Qin J, *et al.* Standard plane localization in fetal ultrasound via domain transferred deep neural networks[J]. *IEEE Journal of Biomedical and Health Informatics*, 2015, 19(5): 1627-1636
- [19] Shin H C, Roth H R, Gao M C, *et al.* Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning[J]. *IEEE Transactions on Medical Imaging*, 2016, 35(5): 1285-1298
- [20] He H X, Khoshelham K, Fraser C. A multiclass TrAdaBoost transfer learning algorithm for the classification of mobile lidar data[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 166: 118-127
- [21] Jaderberg M, Simonyan K, Zisserman A. Spatial transformer networks[OL]. [2020-02-28]. <https://arxiv.org/abs/1506.02025>
- [22] Kingma D P, Ba J. Adam: a method for stochastic optimization[OL]. [2020-02-28]. <http://arxiv.org/abs/1412.6980>
- [23] Bengio Y. Practical recommendations for gradient-based training of deep architectures[OL]. [2020-02-28]. <https://arxiv.org/abs/1206.5533>
- [24] Chougrad H, Zouaki H, Alheyane O. Deep convolutional neural networks for breast cancer screening[J]. *Computer Methods and Programs in Biomedicine*, 2018, 157: 19-30
- [25] Wang Xuchu, Niu Yanmin, Zhao Guangjun, *et al.* Combining region proposals and deep SSAE learnt features for detecting left ventricle in cardiac MR images[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2018, 30(3): 424-435(in Chinese)  
(王旭初, 牛彦敏, 赵广军, 等. 融合候选区域提取与 SSAE 深度特征学习的心脏 MR 图像左心室检测[J]. *计算机辅助设计与图形学学报*, 2018, 30(3): 424-435)
- [26] Liu T J, Xie S N, Yu J, *et al.* Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features[C] // *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. Los Alamitos: IEEE Computer Society Press, 2017: 919-923