****

本科毕业设计（论文）

**基于残差学习与注意力机制的甲状腺**

**超声图像分类器**

|  |  |
| --- | --- |
| **学院**  **专业**  **学生姓名**  **学生学号**  **指导教师**  **提交日期** | **软件学院** |
| **软件工程** |
| **付懿轩** |
| **201430611051** |
| **陈俊颖** |
| **2018年6月1日** |

# 摘 要

本文首先对甲状腺超声图像分类器的应用场景进行了详细的介绍，并说明了作者具体的研究方向即甲状腺超声图像中结节良恶性分类。经过调研后本文详细回顾了解决此问题的各类方案，发现大多数文献所提出的思路是：先从甲状腺超声图像中分割出结节，之后归纳结节的形状或纹理特征，最后将这些特征用于分类器的训练并得到结果。但是将结节及周围区域的超声图像直接用于良恶性分类的研究鲜有人涉及。本文最终选择第二种思路来解决问题，虽然这有一定的挑战但最终还是得到了令人较为满意的成果。文章的创新点主要有两个。1.结合现有文献改进了超声图像预处理方法，如在原有算法中加入自适应直方图均衡化操使得甲状腺超声图像中结节区域与周围正常组织的对比度提高。2.通过结合深度学习领域的最新研究成果如：残差学习、注意力机制，构建了用于甲状腺超声图像中结节良恶性分类的卷积神经网络AttentionNeXt-18。

在实验部分，本文从本地及公开的甲状腺超声图像数据库中选取了58个病例用作研究分析，最终使用928张像素为224224的图片进行分类器的训练及测试。用于执行分类任务的网络完成训练后在测试集上的准确率为90.3%。在实验的最后，作者还将此网络模型与其他文献中提出的方法的分类效果进行了比较并对该网络模型进行了客观评价。

从实验结果来看本模型初步取得了较好的分类效果，这说明深度学习在医学超声领域有较大的应用价值。但想要尽早将其应用于辅助诊断系统仍需要进行更深入的研究。

关键词：甲状腺超声图像；深度学习；图像分类

# Abstract

This article first introduced the application scenario of the thyroid ultrasound image classifier in detail, and explained the specific research direction, ie the classification of benign and malignant nodules in thyroid ultrasound images. After the investigation, the author reviewed in detail the various solutions to this problem and found that most of the ideas proposed in the literature are: first segment the nodules from the thyroid ultrasound image, and then sum up the shape or texture features of the nodules, and finally, combine these features. The use of the classifier for training and obtaining results and the use of ultrasound images of nodules and surrounding areas directly for benign and malignant classifications is rarely involved. This article finally chose the second way to solve the problem. Although this project has some challenges, it eventually achieved satisfactory results. There are two main innovations in the article: 1. Using more effective ultrasound image preprocessing methods 2. By combining the latest research achievements in the field of deep learning, such as: residual learning, attention mechanism, etc., we proposed AttentionNeXt-18 for the classification of thyroid ultrasound images containing nodules.

In the experimental part, this article selected 58 cases from the local and public thyroid ultrasound image database for research analysis, and finally used 928 pixels of 224×224 pictures for classifier training and testing. The accuracy of the test set on the network used to perform the classification task was 90.3% after the training was completed. At the end of the experiment, the author also compares this network model with the classification effect of the methods proposed in other literatures and makes an objective evaluation of the network model.

From the experimental results, the model has achieved a good classification effect, indicating that deep learning has a great application value in the field of medical ultrasound. However, further research is needed to apply it to auxiliary diagnostic systems as soon as possible.

**Keywords**: Thyroid ultrasound image; Deep learning; Image classification

# 目 录

[摘 要 I](#_Toc516863707)

[Abstract II](#_Toc516863708)

[目 录 III](#_Toc516863709)

[第一章 绪论 1](#_Toc516863710)

[1.1研究背景及意义 1](#_Toc516863711)

[1.1.1研究背景 1](#_Toc516863712)

[1.1.2研究意义 2](#_Toc516863713)

[1.2文献综述 2](#_Toc516863714)

[1.3论文结构 5](#_Toc516863715)

[第二章 甲状腺超声图像结节良恶性分类算法实现 6](#_Toc516863716)

[2.1图像采集及其预处理 6](#_Toc516863717)

[2.1.1实验图像采集 6](#_Toc516863718)

[2.1.2图像预处理 6](#_Toc516863719)

[2.2 深度卷积神经网络与注意力机制 9](#_Toc516863720)

[2.2.1深度卷积神经网络VGG-16 9](#_Toc516863721)

[2.2.2残差学习网络RexNet 11](#_Toc516863722)

[2.2.3 Attention机制 12](#_Toc516863723)

[2.3 AttentionNeXt-18网络的建立 13](#_Toc516863724)

[2.3.1 ResNeXt模块 13](#_Toc516863725)

[2.3.2 Attention模块 14](#_Toc516863726)

[2.3.3 AttentionNeXt-18网络 15](#_Toc516863727)

[2.4网络训练 16](#_Toc516863728)

[2.4.1实验平台 17](#_Toc516863729)

[2.4.2实验数据集制作 17](#_Toc516863730)

[2.4.3 AttentionNeXt-18在caffe中的实现 17](#_Toc516863731)

[2.5本章小结 19](#_Toc516863732)

[第三章 实验设计与分析 20](#_Toc516863733)

[3.1 AttentionNeXt-18分类性能分析 20](#_Toc516863734)

[3.2综合对比分析 23](#_Toc516863735)

[3.2.1 AttentionNeXt-18与ResNeXt-18对比 23](#_Toc516863736)

[3.2.2 AttentionNeXt-18与经典深度卷积网络结构对比 24](#_Toc516863737)

[3.2.3 AttentionNeXt-18与迁移学习方法对比 27](#_Toc516863738)

[3.2.4 AttentionNeXt-18与传统机器学习方法对比 28](#_Toc516863739)

[3.2.5 对比分析小结 29](#_Toc516863740)

[第四章 总结与展望 30](#_Toc516863741)

[4.1论文总结 30](#_Toc516863742)

[4.2 展望 30](#_Toc516863743)

[参考文献 32](#_Toc516863744)

[致谢 35](#_Toc516863745)

# 第一章 绪论

## 1.1研究背景及意义

### 1.1.1研究背景

甲状腺是位于喉结下方具有分泌激素功能的蝴蝶状腺体。虽然它形状微小，但对人体起着极其重要的作用。它通过分泌合成甲状腺激素来控制人体代谢，而体内甲状腺激素过多或缺乏都会引起新陈代谢代异常，造成甲亢、甲减、甲状腺结节等甲状腺疾病。在这些甲状腺疾病中甲状腺结节对患者的健康影响最大，恶性甲状腺结节更有较高的致死率。根据我国卫生部的统计数据，近10年由于现代人快节奏、高压力的生活和不规律的饮食，甲状腺结节在正常群体中的病发率逐年提高。医院相应科室的就诊人数激增，医生相关工作量明显增加。

甲状腺结节是甲状腺内的小凸起、小肿块。甲状腺结节的检测有许多方法如血清学检测、核素扫描、细针穿刺细胞学检查、超声诊断、颈部X线检查等。其中超声成像技术诊断以其成本低、采集时间短、无创伤、检测灵活、辐射量低等特性被现代医学广泛用于甲状腺疾病的初期检测。

随着超声检测技术的成熟，放射科医师已经将甲状腺超声图像中的某些纹理特征作为初步判断甲状腺结良恶性的指标如：结节周围回升的高低、是否出现钙化、周围血流情况以及结节区域横纵比等。

为基于这些特点对甲状腺结节进行分类并对其可能为恶性的风险进行分层，Kwak J Y等人[1]设计了一个专门用于甲状腺成像报告及数据分级系统TI-RADS。TI-RADS得分2 , 3 , 4a , 4b , 4c 和5分别表示为“正常”、“可能是良性”、“一个可疑恶性特征”、“两个可疑恶性特征”、“三个或更多可疑恶性特征”和“可能是恶性肿瘤“。通过引入该系统可以降低医生漏判患者患有恶性肿瘤的风险同时避免了不同医生因判断标准的不同导致对患者的诊断产生较大偏差。但是该系统也有一些缺陷如：使用TI-RADS较为耗时，系统稳健性有待提高等。

近年来我国也有许多科研工作者致力于开发一套高效的甲状腺结节超声辅助诊断系统。相关工作主要涉及：从超声图像中分割出甲状腺区域、正常甲状腺与含有结节甲状腺的分类识别、甲状腺结节的分割、甲状腺结节良恶的分类识别等。本文主要关注甲状腺结节良恶的分类识别，希望通过引入深度学习领域的最新成果来建立甲状腺超声图像分类器从而较为准确地区分出良性甲状腺结节和恶性甲状腺结节。

### 1.1.2研究意义

通过建立甲状腺超声图像分类器对甲状腺结节良恶性进行分类的意义主要体现在两个方面。对医生而言该分类器可以降低医生的工作量，医生在患者完成超声检测后通过标定可疑结节区域就可以对患者的病情有一定的预判进而给出下一阶段的治疗方案。其次作为辅助诊断工具，配合医生的诊断可以大大降低此类疾病的误诊率。对患者而言，它可以降低由病情不确定性带来的焦虑，同时为他们省去不必要的检查费用。

训练分类器需要大量的甲状腺超声图像，在这个过程中收集并建立某个地区的甲状腺超声图像数据库也有着十分重要的意义。该数据库可以为后续与甲状腺疾病相关的研究打下基础，同时为医学领域深度学习提供数据支持。

## 1.2文献综述

超声成像的固有缺陷是存在斑点噪声。散斑是由包含许多次分辨率散射体的介质发生相干辐射后在图像中随机形成的确定性干涉图案。斑点噪声会对超声成像有负面影响，因为它会降低图像的有效分辨率和对比度，引起一定程度的不确定性，进而可能会干扰医学诊断。Bamber和Daft [2]已经通过实验证明：由于斑点噪声的存在，甲状腺病变的可检测性大大降低。

在这种情形下，Dimitris K 等人[3]为了获得甲状腺超声图像中噪声的不确定性感知表示，他们通过结合图像局部二进制模式的模糊分布（称为模糊LBP（FLBP）特征）和由模糊灰度直方图表示的超声回声，提出了超声图像中纹理和回声的抗噪声表示。之后用所提出的方法直接提取超声图像中甲状腺结节的特征信息用于SVM进行分类。该方法的准确率较高但是实行起来较为繁琐，并且因为使用不同设备拍摄超声图像的回声分布不尽相同，这使得该方法的泛化能力较弱。

现阶段更多的研究者并没有选择和Dimitris K 等人相同的思路去处理超声图像中的噪声，因为这使原本的问题更加复杂。故而面对噪声他们选择对图像进行预处理以减少噪声的干扰。在此基础上近些年主要有两种思路来实现甲状腺结节的良恶性分类。

第一种思路是先从甲状腺超声图像中分割出结节区域，再利用从该区域中提取的结节的图像特征训练现有的机器学习分类器进行监督学习，最后得到结节良恶性的判别结果。王昊等人[4]首先通过使用NCut和Snakes图像分割技术从甲状腺超声图像中提取甲状腺结节ROI区域，接着对ROI区域进行DT-CWT和Gabor变换得到纹理图像，之后计算纹理图像的均值、方差等特征以获得超声图像中甲状腺结节的特征向量。最后分别使用了支持SVM、BP神经网络、Fisher线性判别等分类方法完成甲状腺结节良恶得判别。其中SVM的分类效果最好，在训练集选择比较好的情况下准确率能够达到84%左右。

Rahmawaty M等人[5]首先使用中值滤波器、形态学操作、直方图均衡化一系列操作对超声图像进行预处理，之后采用ACWE算法从超声图像中分割出结节。接着利用灰度共生矩阵基于图像的纹理提取出八个最具代表性的特征用于区分甲状腺结节的良恶，最后使用多层感知器(MPL)结合输入特征进行二分类。该方法的准确率可以达到89.74%。Nugroho H A等人[6]在对超声图像经过预处理之后，结合形态学操作和活动轮廓模型分割出甲状腺结节。接着他们从分割的甲状腺结节图像中提取了七个几何特征和十四个矩特征。值得注意的是他们并没有直接使用这些特征用于后续的分类，而是通过使用基于相关性的特征选择（CFS）算法从这些特征中再次筛选了3个用于SVM进行甲状腺结节良恶性分类的特征。该方法的准确率可以达到91.52%。可惜的是以上文章均只使用了少量甲状腺结节超声图像（不超过300张图像）进行训练和测试，数据集较小使得模型泛化能力较弱。

熊伟等人[7]将提取的局部纹理特征与多示列学习算法相结合以实现超声甲状腺结节分类。他们通过对结节区域内自定义分类点的特征值进行分类进而实现对结节良恶性的分类。影响该方法性能的因素主要为分类点的选取以及如何表征分类点。他们选择分类点邻域内灰度共生矩阵的属性值表征分类点，但对于分类点的选择方法文章没有具体阐述。

第二种思路是直接使用深度卷积网络从甲状腺结节超声图像中提取特征，之后使用softmax、支持向量机(SVM)等分类算法进行良恶性二分类。Hijazi S等人[8]发现与传统图像特征提取算法相比，深度神经网络DCNN具有两个优点：使用DCNN进行图像特征检测时在面对由于相机镜头长短不同、光照条件不同、部分遮挡、水平和垂直偏移等造成的图像形变时有较强的鲁棒性；其次使用DCNN进行特征提取的计算代价较低。他们认为由于在非医学图像识别领域的良好表现，DCNN也应被用于医学图像分类及检测。

孔德兴等人[9]提出了基于深度卷积神经网络自动识别甲状腺结节良恶性的方法。他们首先利用高斯滤波和直方图均衡化对原始超声图进行预处理，得到预处理后的增强图像10000张。接着手动截取这些图像中的结节部分用于训练卷积神经网，待网络训练好后即可利用它对新输入的甲状腺结节图像进行良恶性判别。文章中的卷积神经网络是由13个卷积层、2个下采样层组成的网络。网络不足之处在于使用了较多尺寸较大的卷积核用于卷积运算，这使得网络需要训练的参数大幅增加。文章的另一个不足之处在于没有使用明确的量化指标对模型的最终输出结果进行评判。

Chi J等人[10]在使用与孔德兴等人类似的图像预处理方法后，为弥补实验样本数量的不足他们采用样本增加技术使得实验样本数量从原先的592增加到2754。之后他们从Deep Learning Caffe library[11]中下载了为参加ImageNet比赛而训练的模型—GoogLeNet用作他们实验的预训练模型。接着使用制作好的2754张甲状腺结节图像样本作为GoogLeNet新的输入并修改原网络部分参数来微调该网络以获得最终的甲状腺结节良恶性分类网络。微调后的网络性能良好准确率超过92%。作者使用微调方法在一定程度上缓解了训练数据量不足带来的问题，但作者选择的GoogLeNet在训练时使用的图像均不是医学超声图像。因此作者通过微调此模型以获得甲状腺结节超声图像良恶性分类网络的科学性及合理性有待进一步证明。

Micha Byra等人[12]使用自己建立的五层卷积神经网络对结节的良恶性进行分类识别，准确率在91%左右。文章中使用了较为新颖的预处理方法，他们直接对原始超声RF数据进行预处理以获得Nakagami map 用于卷积神经网络的训练。虽然Nakagami map可以较好的表现结节区域的纹理特征但是却受限于网络结构，较浅的卷积神经网络只能提取并学习相当有限的特征用于分类。

刘天娇等人[13]在特征融合的基础上引入迁移学习的思想完成超声甲状腺结节良恶性分类。与其他方法不同的是该方法先从经预处理后的图像中提取底层特征，再利用已通过ImageNet数据集训练后的VGG-F模型提取超声图像在全连接层的输出向量用作高层语意特征。接着将以上两个已提取特征合并，根据良恶性甲状腺结节的区分度进行特征筛选，得到最终的显著特征描述向量。最后将医生已标记好的甲状腺结节超声图像作为训练集，利用支持向量机(SVM）以及显著特征描述向量进行训练，得到针对超声甲状腺结节良恶的分类器。该方法的准确率较高但是获取显著特征描述向量十分费时，另外该方法和Chi J等人提出的方法有类似问题即用于模型预训练的数据集并非医学超声图像。

总的来讲在以上两类思路中，目前大多数用于区分甲状腺超声图像中结节良恶的方法均采用思路一。但随着医疗信息化水平的提高，面对海量的医学数据，思路一无法高效且全面地考虑用于区分甲状腺超声图像中结节良恶的所有特征。因此本文将采用思路二完成甲状腺超声图像中结节良恶性的分类识别。

## 1.3论文结构

本文主要分为四个章节。第一章是绪论，本文在绪论部分详细阐述了甲状腺超声图像中结节分类问题的的研究背景、研究意义以及相关文献中解决此类问题的方法，并在最后介绍了本论文的结构安排。第二章是甲状腺超声图像结节良恶性分类算法实现，它包含五个部分。分别是图像采集及其预处理、深度卷积神经网络与注意力机制、AttentionNeXt-18网络的建立、网络训练以及本章小结。第三章是实验结果，本章首先通过引入具体的评价指标对实验结果进行量化评判，其次通过和解决此问题的其他方法进行对比得出本文所提方法的优缺点并给出了可能的优化思路。故而本章包括模型评估和对比讨论两个部分。最后一章是总结与展望，这部分回顾了AttentionNeXt-18网络建立及训练的全过程并探讨了接下来的研究方向。

# 第二章 甲状腺超声图像结节良恶性分类算法实现

本章详细讲述了甲状腺超声图像中结节良恶性分类算法实现的全过程并展示了其中涉及的所有技术细节。本章主要包含五个小节分别是：样本采样及预处理、深度卷积网络与注意力机制、AttentionNeXt-18的建立、网络训练以及本章小结。

## 2.1图像采集及其预处理

### 2.1.1实验图像采集

用于进行实验的超声图像数据来自以下两个数据库：

数据库1是一个可公开访问的甲状腺超声图像数据库，该数据库由Pedraza等人[14]建立。数据库中包含428张尺寸为560\*360的甲状腺超声图像，这些图像是从使用TOSHIBA Nemio 30和TOSHIBA Nemio MX超声设备捕获的甲状腺超声视频序列中提取的。每张图片在经放射科专家组评估后均被勾勒出了相应的病灶区。其中有357张图像含有恶性甲状腺结节，其余71张图像中含有良性甲状腺结节。

数据库2是由实验室建立的本地数据库，甲状腺超声图像由中山大学附属医院提供。数据库中包含尺寸为500\*400的正常甲状腺超声图像、含有恶性结节的甲状腺超声图像以及含有良性结节的甲状腺超声图像共903张。这些图像中只有极少数含有医生的标注。

然而我们并不能直接使用数据库中的图像用于模型的训练，原因在于：1.由于超声图像采集于不同的医疗机构，两个数据库中有些超声图像的纹理尺度不相同。2.部分超声图像中含有放射科医师添加的病灶区域标记及注释，这些图像在被用于深度学习训练时会影响分类器的学习效果。因此在使用这些图像用于分类器训练之前采用适当的图像预处理方法以提高图像质量并适当增强图像样本的特征是十分必要的。

### 2.1.2图像预处理

图像预处共分为三个阶段，所有阶段的操作全部依托MATLAB 2017b完成。本文使用的图像预处理算法是结合其他相关文献后提出的加强版算法。主要算法流程见算法表格2-1。

表2-1 图像预处理算法

|  |  |
| --- | --- |
| 步骤 | 具体实现 |
| 输入 | 灰度图像 |
|  | 第一阶段 |
| 1. | 将变换成为二值图像 |
| 2. | 在上使用连通分量算法，得到个连通区域 |
| 3. | 统计个连通区域的面积，获得最大连通区域对应像素点的集合，将中对应像 |
|  | 素点的值重置为0，其余像素点的值重置为1。得到图像 |
| 4． | 通过点乘运算 得到可能含有注释的区域， |
| 5． | 统计区域 中灰度值在[201,255]范围内的像素点数量，获得数量排在前三的对 |
|  | 应灰度值 |
| 6. | 去掉图像中灰度值为的点得到图像 |
| 7． | 对图像使用自适应中值滤波算法[15]得到 |
|  | 第二阶段 |
| 8. | 对图像使用对比度有限的自适应直方图均衡化处理得到图像 |
|  | 第三阶段 |
| 9． | 对进行剪裁得到最终的输出 |
| 输出 | 灰度图像 |

图像预处理的第一阶段是去掉部分超声图像中由医师添加的用于标记感兴趣区域或指明其他信息的标记。如图2-1中在红色矩形框中的箭头即是需要去掉的部分。

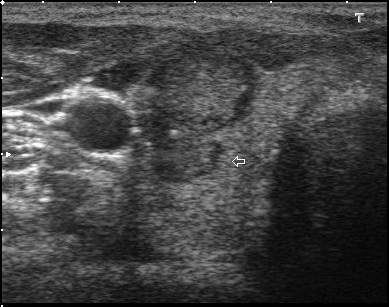


图2-1带有医生标记的原始超声图像

经过前6步处理之后，图中的注释信息基本被去除。此时去掉注释后的相关区域如图2-2所示。对去掉注释后的图像进行自适应中值滤波处理后得到降噪后的图像。

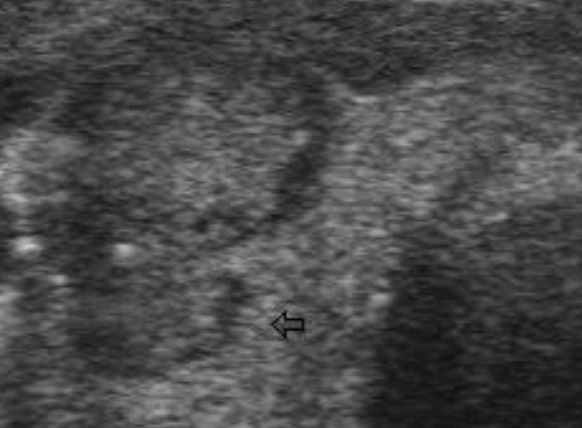


图2-2.去掉注释后的相关区域

至此图像预处理第一阶段全部完成。经过图像预处理第一阶段处理后的图像如图2-3所示。



图2-3 经图像预处理第一阶段处理之后的图像

图像预处理的第二阶段是将去掉注释的甲状腺超声图像以及不含注释的甲状腺超声图像进行对比度有限的自适应直方图均衡化处理(CLAHE) 。该算法的思想是首先根据图像的最大尺寸计算网格大小，之后指定滑窗大小。如果未指定窗口大小，则选择网格大小作为默认窗口大小。在从个左上角确定图像上网格点之后，对每个网格点计算得到其周围区域的灰度直方图，该区域是以网格点为中点以窗口大小为其面积的区域。如果指定了剪切级别，则将上面计算出的直方图剪切到该级别，然后使用新得到的直方图计算它的累积分布函数CDF。最后对于每个像素找出围绕该像素的四个最接近的相邻网格点后使用像素的灰度值作为索引，根据它们的CDF在四个网格点上找到它的映射。在这些值之间进行插值以获得当前像素位置处的映射。之后将此灰度映射到最大值和最小值范围之间并将其放入最终输出图像中

此方法在提高图像对比度的同时最大限度的抑制了局部噪声的产生。相应的实现已被MATLAB封装成函数adapthisteq() 供研发人员使用。经过该方法处理后的图像如图2-4所示：



图2-4 经过预处理两个阶段后的图像

图像预处理的最后一个阶段是从已完成前两阶段处理的超声图像中剪裁出尺寸为224224的甲状腺结节区域用于后续分类器的训练。

## 2.2 深度卷积神经网络与注意力机制

众所周知应用不同网络结构提取到的特征图在用于分类时得到的分类效果有时相差很大。因此建立与待解决问题相匹配的网络结构至关重要。在建立AttentionNeXt-18之前作者首先分析了经典的深度卷积网络结构，并从中提取深度卷积神经网络设计的基本理念用于后续模型的建立。同时作者将注意力机制运用于超声图像分类问题。

### 2.2.1深度卷积神经网络VGG-16

VGG-16网络是由牛津大学的Karen Simonyan和Andrew Zisserman提出[16]。他们在文章中指出了卷积神经网络的深度对于网络提取高层次图像特征的重要性。论文中VGG-16网络的结构如下表2-2所示：

表2-2 VGG-16网络结构表

|  |
| --- |
| input(224224 RGB image) |
| conv3 - 64  conv3 - 64 |
| maxpool |
| conv3 - 128  conv3 - 128 |
| maxpool |
| conv3 - 256  conv3 - 256  conv3 - 256 |
| maxpool |
| conv3 - 512  conv3 - 512  conv3 - 512 |
| maxpool |
| conv3 - 512  conv3 - 512  conv3 - 512 |
| maxpool |
| FC - 4096 |
| FC - 4096 |
| FC -1000 |
| soft-max |

与之前研究者提出的浅层卷积神经网络相比，该网络仅仅使用卷积核尺寸为33的滤波器。通过计算我们可以发现连续两个卷积核尺寸为33的滤波器组合之后得到的感受野大小与单独使用一个卷积核尺寸为55的滤波器的感受野相同；同理连续三个卷积核尺寸为33的滤波器组合之后得到的感受野大小大小与单独使用一个卷积核尺寸为77的滤波器的感受野相同。但使用连续尺寸较小的卷积核可以降低整个网络的训练参数并使网络加深，这对与提升模型的训练速度和训练准确度至关重要。从表格中我们不难发现，滤器数量在每个maxpool层之后都会增加一倍。这强化了缩小空间维度但增加深度的想法。总的而言这篇文章想要提示我们在设计网络结构时应尽量选择尺寸较小的卷积核用于卷积运算并在计算资源允许的范围内加深网络结构。

### 2.2.2残差学习网络RexNet

He K等人[17]发现如果简单沿用VGG网络结构，当网络加深到一定层数后梯度消失问题会使得网络很难训练。由于梯度传播回前一层，因此重复的乘法可能会使梯度变得无限。 结果是，随着网络层数量的加深，其性能趋于饱和，甚至开始迅速下降，导致网络在进行图像分类和识别时效果退化。具体来讲，He K等人在同一数据集下使用VGG-20和VGG-56这两种网络结构进行训练和测试时发现，在相同迭代次数下VGG-56网络对应的训练错误率和测试错误率均高于VGG-20网络。为解决该问题他们提出了残差学习单元如图2-1。

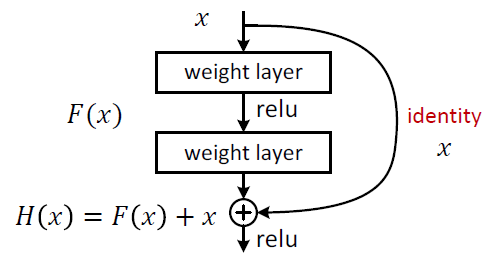


图2-1 残差学习单元

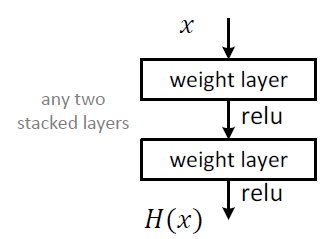


图2-2 部分VGG网络

在图2-1和图2-2中x 均表示输入图像，H(x) 表示最终输出，identity表示恒等映射，F(x)表示求和前卷积网络相关参数映射。通过对比图1和图2，我们不难发现残差学习单元在最后的输出通过短路连接保留了输入图像的信息，这使得输出变化对F(x)中参数权重的调整作用更大。引入残差学习单元的优势在于:(1)在保持网络结构简洁的同时使得网络更容易优化； (2)该结构缓解了梯度消失问题，使得网络可以更容易地加深从而提高网络在训练和测试阶段的准确率。RexNet的诞生告诉我们在设计极深网络时应当使用残差学习单元提高网络的整体性能，在较深网络中使用残差学习单元可以加速网络的最优拟合过程。

### 2.2.3 Attention机制

来自人类感知过程的证据[18]表明了注意机制的重要性，该机制使用顶层信息来引导自下而上的特征匹配前馈过程。许多研究已经尝试将注意力机制引入到深度神经网络，采用“自下上升的非监督学习+自顶向下的监督学习”： 自下上升阶段即权重的初始化阶段，在这一阶段输入大量无标签数据，分别得到每一隐含层上的权重及其输出；自顶向下阶段即权重精调阶段，在这一阶段输入有标签数据，首先获得当前网络的累计总误差，然后计算每一隐含层上每个神经元引起的的误差，接着针对每一隐含层的权重进行独立的迭代调整，而是不是像多层感知机那样对网络中所有权重进行统一迭代调整。非监督学习的目的是学习大量数据的结构，使每一层隐含层的初始权重更满足全局最优，避免随机初始化距离最优解太远而陷入局部最优；监督学习则根据有标签数据对网络权重进行精细化调整，由于每一隐含层输出都是对其输入的更精细的特征表示，因此每层权重的每次调整都仅仅企图最小化本层网络的输出误差，而不是全部网络的输出误差，这种方法同样有效遏制了反向传播通过每层连续乘法引起的梯度爆炸/消失问题。

在卷积神经网络领域，这种思想正开始慢慢活跃起来。Newell A等人[19]提出的stacked hourglass network 网络通过融合多个尺度的特征信息可以很好的预测人体姿势。该网络采用跳级的思想搭建了先降采样再升采样的全卷积结构，具体来讲就是：每次降采样之前网络分出上半支路保留原输入图像尺度信息；每次升采样之后，和来自上半支路保留的图像原尺度数据相加；两次降采样之间，使用三个残差学习单元提取特征；两次相加运算之间，使用一个残差学习单元提取特征。

Wang F等人[20]受stacked hourglass network 网络结构的启发也设计出用于图像分类的带注意力机制的网络。文章提出的Attention Module包含快速前馈扫描和自顶向下反馈机制。前馈扫描可以快速收集整个图像的全局信息，后者作将全局信息与原始特征图相结合。在卷积神经网络中，这两个机制可以通过自下而上的自顶向下全卷积结构来实现。

总的来讲用于现阶段图像分类领域的注意力机制主要是靠 bottom-up与top-down feedforward这一网络结构实现：首先自下而上的前馈结构产生具有强烈语义信息的低分辨率特征映射。之后自顶向下的网络结构发挥类似特征选择器的作用，增强特征图中较好的特征并抑制噪声。

## 2.3 AttentionNeXt-18网络的建立

本小节主要介绍用于提取甲状腺超声图像中结节特征并用于分类的卷积神经网络AttentionNeXt-18的设计思路以及具体实现细节。

AttentionNeXt-18是按照模块化的思想建立，主要由ResNeXt模块和Attention模块构成。

### 2.3.1 ResNeXt模块

ResNeXt模块的构建是在ResNet残差学习单元的基础上引入cardinality概念[30]完成的,这样可以降低网络的计算量。该模块详细信息如图2-3：

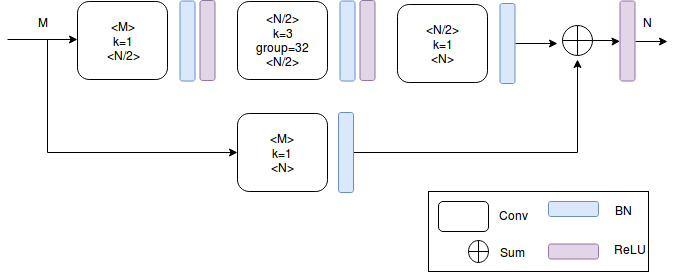


图2-3 ResNeXt模块

图中Ｍ代表输入通道数，Ｎ代表输出通道数，ｋ代表进行卷积运算时卷积核大小，group是caffe中卷积层的参数；group为32表示进行卷积运算时先采用32个输入输出通道数均为４的卷积核进行计算，最后在输出时将通道合并。

每次卷积运算完成后进行批处理归一化(BN)操作[21]可以减少隐层协方差变化。另外批处理归一化操作允许网络的每一层独立学习，而不依赖于其他层。ResNeXt模块另一个亮点在于:所有卷积层进行运算时均采用stride=1;卷积核尺寸为33的卷积层采用pading=1;卷积核尺寸为11的卷积层采用pading=0，这样确保特征图的尺寸在通过该结构时始终保持不变。故而ResNeXt模块仅仅改变输出的通道数，这使得模型整体上变得简洁也更容易进行学习。结构中BN层、ReLU函数以及卷积层设置的顺序参考了Ren S等人[22]的研究。

### 2.3.2 Attention模块

Attention 模块也是由ResNeXt模块构成，它可以逐步完善特征图所表达的语意从而提高图像分类器的识别准确率。Attention 模块的结构示意图如下图2-4：

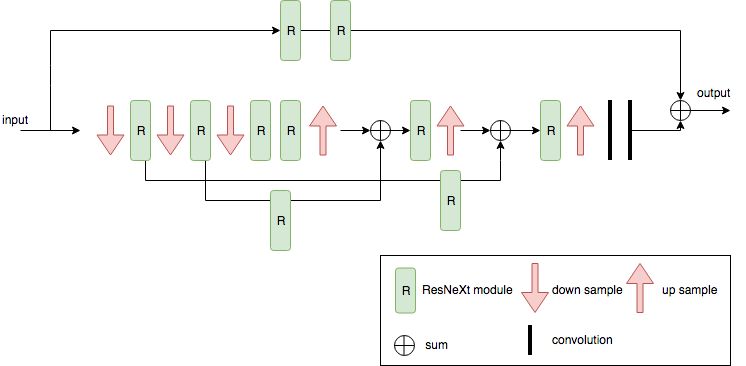


图2-4 Attention 模块

Attention 模块包含两路。上半路由两个ResNeXt模块构成，它们在输入图像的原尺度上逐步提取更深层次的特征。

在下半路我们通过下采样操作配合较小数量的ResNeXt模块可以迅速增加感受野的范围，其中下采样操作使用max pooling实现。在经过三次下采样操作后图像分辨率达到较低水平，在此之后经先前下采样操作获得的全局信息将通过对称的自顶向下结构进行扩展，更有目的的指导训练每个位置的输入特征。为确保下半路输出特征图和输入时尺寸一致，在ResNeXt模块后进行三次上采样操作。

在Attention模块中，自底向下结构对应的是三次下采样以及每次下采样后紧接着的ResNeXt模块；自顶向上结构对应的是三次上采样以及每次上采样后紧挨着的ResNeXt模块；两个对称的结构可以逐步优化最终得到的特征图。从图中可以看到在自顶向下结构和自底向上结构之间加入了跳跃连接，通过加入跳跃连接来捕获不同尺度的特征信息可以使训练得到的特征更具有代表性。下路最后两层使用卷积核尺寸为11的滤波器进行卷积运算是为了实现多个特征图的线性组合。

### 2.3.3 AttentionNeXt-18网络

用于甲状腺超声图像结节分类的网络AttentionNeXt-18最终的结构如图2-5所示，

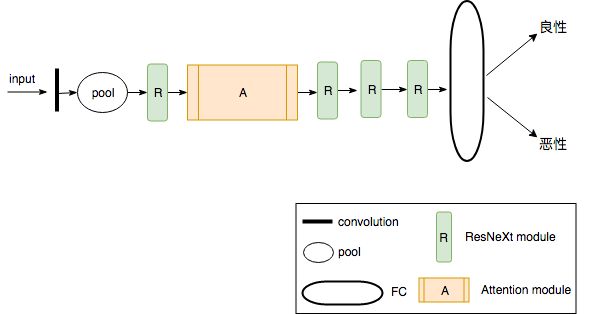


图2-5 AttentionNeXt-18网络结构图

它使用了ResNeXt模块和Attention 模块，网络最后一层是全连接层并用softmax函数进行二分类。考虑到医学超声图像的特点以及网络的分类功能并没有将网络设计的过深，网络可以学习到的特征图的最小尺度为77 像素大小。AttentionNeXt-18网络结构的细节如表2-3所示：

AttentionNeXt-18所用的注意力机制与Wang F等人[20] 提出的方法最大的不同在于AttentionNeXt-18考虑到了超声图像存在斑点噪声，以及本文所解决的问题是二分类问题等特点，重新构建了网络结构。包括网络中Attention模块放置的位置以及Attention模块内部进行上采样与下采样的顺序及次数等。

表2-3 AttentionNeXt-18结构细节

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Unit | Output Size | AttentionNeXt - 18 | | | | | | | |
| Conv1 | 112112 | kernel size | | the number of filter | | stride | | | padding |
| 77 | | 64 | | 2 | | | 3 |
| Max pooling | 5656 | kernel size | | | stride | | padding | | |
| 33 | | | 2 | | 0 | | |
| ResNeXt Module | 5656 | kernel size | the number of filter | | | stride | | padding | |
| 11 | 128 | | | 1 | | 0 | |
| 33 | 128 | | | 1 | | 1 | |
| 11 | 256 | | | 1 | | 0 | |
| Attention Module | 56 56 | Attention1 | | | | | | | |
| ResNeXt Module | 28 28 | kernel size | | the number of filter | | stride | | padding | |
| 11 | | 256 | | 1 | | 0 | |
| 33 | | 256 | | 2 | | 1 | |
| 11 | | 512 | | 1 | | 0 | |
| ResNeXt Module | 2828 | kernel size | | the number of filter | | stride | | padding | |
| 11 | | 256 | | 1 | | 0 | |
| 33 | | 256 | | 1 | | 1 | |
| 11 | | 512 | | 1 | | 0 | |
| ResNeXt Module | 2828 | kernel size | | the number of filter | | stride | | padding | |
| 11 | | 256 | | 1 | | 0 | |
| 33 | | 256 | | 1 | | 1 | |
| 11 | | 512 | | 1 | | 0 | |
| FC,Softmax | 2 | | | | | | | | |

表2-2中每个单元对应图2-5中相应的元素，Output Size为该层最终输出特征图大小。卷积层各个参数细节也相应地列在表中。

输入网络的图像大小为224224 像素。从表中可以看出该网络最终输出的特征图大小为2828像素。

## 2.4网络训练

本小节主要讲述了如何将上一小节提出的网络模型用代码实现并结合预处理后的数据完成相应的训练。

### 2.4.1实验平台

整个卷积神经网络的编写以及训练均在caffe[23]中完成，caffe是一款用C++编写含有python 接口的深度学习框架。选择caffe作为实验平台有以下几点原因：1.该框架富有极强的表达能力，定义模型和完成相关优化均无需硬编码；2.可以自由选择使用GPU或CPU来训练模型，且使用者能够十分便捷地通过更改代码中相应标记字段完成GPU和CPU之间的切换；3.图像处理的速度快、效率高且易于进行加速优化；4.提供了大量开箱即用的功能层和模型库。

### 2.4.2实验数据集制作

用预处理后的超声图像进行实验数据集的制作。用于实验进行训练和测试的数据分配如表2-3所示，

表2-4 测试集与训练集数据分布

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 总数 | | 甲状腺良性结节 | | 甲状腺恶性结节 | |
| Cases | Samples | Cases | Samples | Cases | Samples |
| 训练集 | 40 | 160 | 15 | 60 | 25 | 100 |
| 测试集 | 18 | 72 | 8 | 32 | 10 | 40 |

数据集中的每一个case对应一张含有结节的甲状腺超声图像，之后从一个case中选取四张包含结节及部分周边区域的图像作为sample。之后是制作测试集和训练集相应的lmdb文件，lmdb是caffe数据层可以接收的一种数据格。完成lmdb格式文件的制作最重要的是需要分别为良性甲状腺结节和恶性甲状腺结节的超声图像贴相应的标签。本实验中将甲状腺超声图像中良性结节的图像标记为0，将恶性结节的图像表记为1.

### 2.4.3 AttentionNeXt-18在caffe中的实现

caffe中层是网络模型的组成要素和计算的基本单位，用于实现网络结构的常用层有数据层、视觉层、激活层、softmax层、accuracy层等。

数据层是每个模型的最底层，是模型的入口，不仅提供数据的输入，也提供数据从Blobs转换成别的格式进行保存并输出。数据的预处理也可以通过该层实现。数据来源可以来自高效率的数据库(如LMDB等)，也可以直接来自于内存。

卷积层是视觉层的核心。该层可以定义个类卷积运算、网络的学习率、滤波器数量、用于初始化权值的方法、是否开启偏置项等。池化层（pooling）也是视觉层的重要组成部分，该层是为了减少运算量和数据维度而设置的一种层。常用的池化方法有MAX和AVE。

在激活层中，网络对数据进行激活操作（实际上就是一种函数变换），这种操作是逐元素进行运算的。从网络底层得到一个Blob数据输入，运算后从网络顶层输出一个Blob数据。激活层没有改变输入数据的大小。激活层函数的类型有很多种如sigmoid函数、tanh函数、relu函数等，在AttentionNeXt-18中使用的激活函数为relu，因为其收敛速度快。

模型中用到的层还有softmax层和Inner Product层。softmax层相当一个分类器，计算的是输出类别的概率属于逻辑回归的一种推广。Inner Product层又名全连接层(简写为FC)，该层的输入输出均为向量。最终的输出相当于将把Blobs的长和宽均置为1 。全连接层实际上也是一种卷积层，只是它的卷积核大小和原数据大小一致。

在对网络各个层的功能有一定了解后，就可以在caffe中编写相应的prototxt文件。prototxt是一种数据格式，网络模型的运行主要依靠两个prototxt格式的文件。第一个是train\_test.prototxt 文件，它通过对caffe中各层进行定义从而确定网络结构并指定用于该网络训练和测试的数据集。第二个是solver.prototxt 文件，它相当于模型的求解器。它定义了网络进行梯度下降的方式，并控制网络按此方式有效的进行参数训练。由于caffe原本的框架中并没有用来完成双线性插值操作的层，因此需要按照官方文档的要求编写并编译实现该功能的Interp 层。

在完成以上工作后通过编写训练的脚本语句就可以使用训练集对模型进行训练并用测试集对完成训练后的模型进行测试。

模型进行训练及测试的系统环境为Ubuntu 16.04，核心处理器为Intel i5 2.7GHz ,内存大小为8GB, 1600MHZ, DDR3。训练时采用CPU模式，在此模式下完成训练及测试任务共用182分钟。

## 2.5本章小结

第二章是本文的核心章节，该章节汇总了本文的两大创新工作即：结合已有文献提出了加强版的超声图像预处理算法；将Attention机制应用于解决具体的医学超声图像分类问题。此外本章节还对AttentionNeXt-18网络在caffe中的实现细节进行了相关说明，并对网络的训练集和测试集的数据来源及分布进行了详细的介绍。

# 第三章 实验设计与分析

本章通过设计一系列的实验对AttentionNeXt-18网络的分类效果以及网络的整体性能进行检测，并针对实验结果给出相应的分析。

## 3.1 AttentionNeXt-18分类性能分析

本文选择准确性(accuracy),敏感度(sensitivity),特异度(specificity),受试者工作特性曲线(ROC)作为检测模型分类效果的指标。对于甲状腺超声图像结节良恶性分类这个二分类问题，表3-1展示了测试集对应的混淆矩阵。

表3-1 测试集混淆矩阵

|  |  |  |  |
| --- | --- | --- | --- |
|  | | AttentionNeXt-18模型预测结果 | |
| 良性 | 恶性 |
| 真实结果 | 良性甲状腺结节 | 120 | 8 |
| 恶性甲状腺结节 | 20 | 140 |

从测试集混淆矩阵中我们可以发现在所有良性的甲状腺结节超声图像样本中，本文提出的网络AttentionNeXt-18错误的预测结果仅有8例，在所有恶性的甲状腺超声图像样本中本文提出的网络AttentionNeXt-18错误的预测结果有20例。我们推测在网络AttentionNeXt-18学习的过程中良性甲状腺结节超声图像的特点更容易学习，恶性甲状腺结节超声图像的特点因为样本数量以及恶性结节较为复杂的生理医学特性不容易进行学习。这样实验结果与现实情况基本一致：良性甲状腺结节在超声图像中特征明显容易进行分辨，恶性甲状腺结节在超声图像中特征种类较多且类型复杂有时不容易进行判断。

在机器学习分类问题中，如果模型将一个正例预测为正例则认为最后结果中产生了一个真正例；如果模型将一个反例正确预测为反例，则认为最后结果中产生一个真反例(FP).相应的如果模型将一个正例预测为反例则认为最终结果中产生了一个伪反例(FN);如果模型将一个反例预测为正例则认为最终结果产生一个伪正例。根据表格3-1我们可以确定在该测试集上：TP = 120, FN = 8, FP = 20, TN = 140。利用以上数据我们根据公式可以进行计算出：

Accuracy = =90.28%. (3-1)

Sensitivity = =93.75%. (3-2)

Specificity = =87.50%. (3-3)

使用MATLAB绘制测试集测试结果相应的ROC曲线如图3-1：

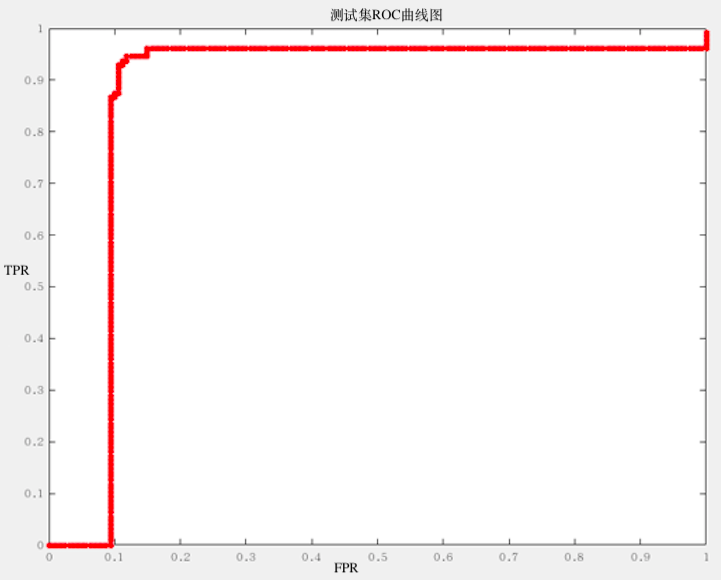


图3-1 测试集ROC曲线

ROC曲线可以用于评价一个分类器。曲线由两个变量1-specificity 和 Sensitivity绘制。1-specificity=FPR，即负正类率。Sensitivity即是真正类率，TPR(True positive rate),反映了正类覆盖程度。图中线条描述的就是AttentionNeXt-18网络在测试集上每一个预测结果的负正类率以及真正类率的变化情况，图像整体靠近左上角1的位置。这说明文章中构建的AttentionNeXt-18网络有较好的预测能力。文章还使用了具体的量化指标对分类器的分类性能进行评定，如下文。

我们通常以ROC曲线下方面积(AUC)的大小作为模型优劣的指标。简单来讲AUC值越大的分类器，正确率越高。从表3-2中我们可以看到AttentionNeXt-18的AUC值为0.88，这说明本文构建的AttentionNeXt-18网络可以用于解决甲状腺超声图像中结节良恶性分类问题，并且AttentionNeXt-18网络表现出较好的分类效果。将以上评价信息汇总如下表3-2所示。

表3-2 模型分类效果指标

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Accuracy | Sensitivity | Specificity | AUC |
| 所提出模型 | 90.28% | 93.75% | 87.50% | 0.88 |

从上述实验结果可以看出本文提出的AttentionNeXt-18模型可以有效的提取甲状腺超声图像中的高层次信息并能够较为准确地对图像中的结节进行良性和恶性的分类判断。实验表明在图像样本量还不充足的情况下，对超声图像进行预处理可以有效改善网络的性能并提高测试时的准确率。具体表现在1.统一输入网络的超声图像的尺寸可以使得网络在反向传播时参数变化不会十分剧烈。2.去除超声图像中由放射科医生手动标记的注释并在降噪的同时提升图像对比度，可以避免网络学习到一些不必要的特征，从而提高网络在分类时的准确率。当然这个问题可能会随着用于网络训练的图像样本的增加得到解决，因为在以往的深度学习分类问题研究中，用于训练网络的图像数据常常是数以万计的。大量数据可以防止不同尺度的少量噪声破坏神经网络的训练过程。

## 3.2综合对比分析

我们提出的分类网络模型是将深度学习前沿理论应用于超声图像中甲状腺结节良恶性评估的探索性工作，因此对比分析十分重要。本小节将从三个方面出发对本文建立的网络模型与其他模型方法进行对比。

### 3.2.1 AttentionNeXt-18与ResNeXt-18对比

为测试AttentionNeXt-18网络中Attention模块的有效性，本文在该部分通过实验对比了AttentionNeXt-18网络与仅仅由ResNeXt模块构成的ResNeXt-18网络在甲状腺超声图像中结节良恶性的分类效果。实验所用的训练集与测试集均为上一章所建立的数据集。以上两个网络在测试集上的预测准确率如表3-3所示。

表3-3 AttentionNeXt-18与ResNeXt-18分类效果对比

|  |  |
| --- | --- |
| 用于分类的深度学习网络 | 准确率 |
| ResNeXt-18 | 82.36% |
| AttentionNeXt-18 | 90.28% |

从表3-3中我们不难发现，仅使用ResNeXt模块构建的网络ResNeXt-18使用相同的数据集经过训练之后在测试集上的准确率仅有82.36%。而将仅有ResNeXt模块构成的网络中连续堆叠的十个ResNeXt模块替换成Attention结构，即采用本文构建的网络AttentionNeXt-18之后在测试集上的准确率达到了90.28%。在使用Attention模块之后网络整体在测试集上的预测准确率有了明显的提高，提高了近7%。这个项比实验也充分说明了注意力机制在甲状腺超声图像中结节良恶性分类问题上发挥了重要作用，同时也验证了本文所提出的第二个创新点在解决此类问题上的有效性。

综合表3-3中的实验结果以及本文在第二章中有关AttentionNeXt-18网络在实现方面的技术细节，我们可以得到以下结论：AttentionNeXt-18网络因为Attention模块的存在在进行甲状腺超声图像结节良恶性预测时拥有更高的预测准确率。本文在构建AttentionNeXt-18网络时使用的Attention机制在超声图像分类任务中具有较高的应用价值，它可以逐步完善特征图所表达的语意从而提高图像分类的识别准确率。

### 3.2.2 AttentionNeXt-18与经典深度卷积网络结构对比

本文在第二章提到了两类经典的深度卷积神经网络即VGG-16网络以及ResNet。为对比它们和AttentionNeXt-18在解决超声图像分类问题上的优劣，在本小节我们使用论文中构建的数据集对VGG-16网络、ResNet-18网络、ResNet-50网络以及本文提出的AttentionNeXt-18网络进行甲状腺超声图像中结节良恶性的预测实验。首先通过实验获得以上四个网络Training Loss与迭代次数的关系如图3-2所示。

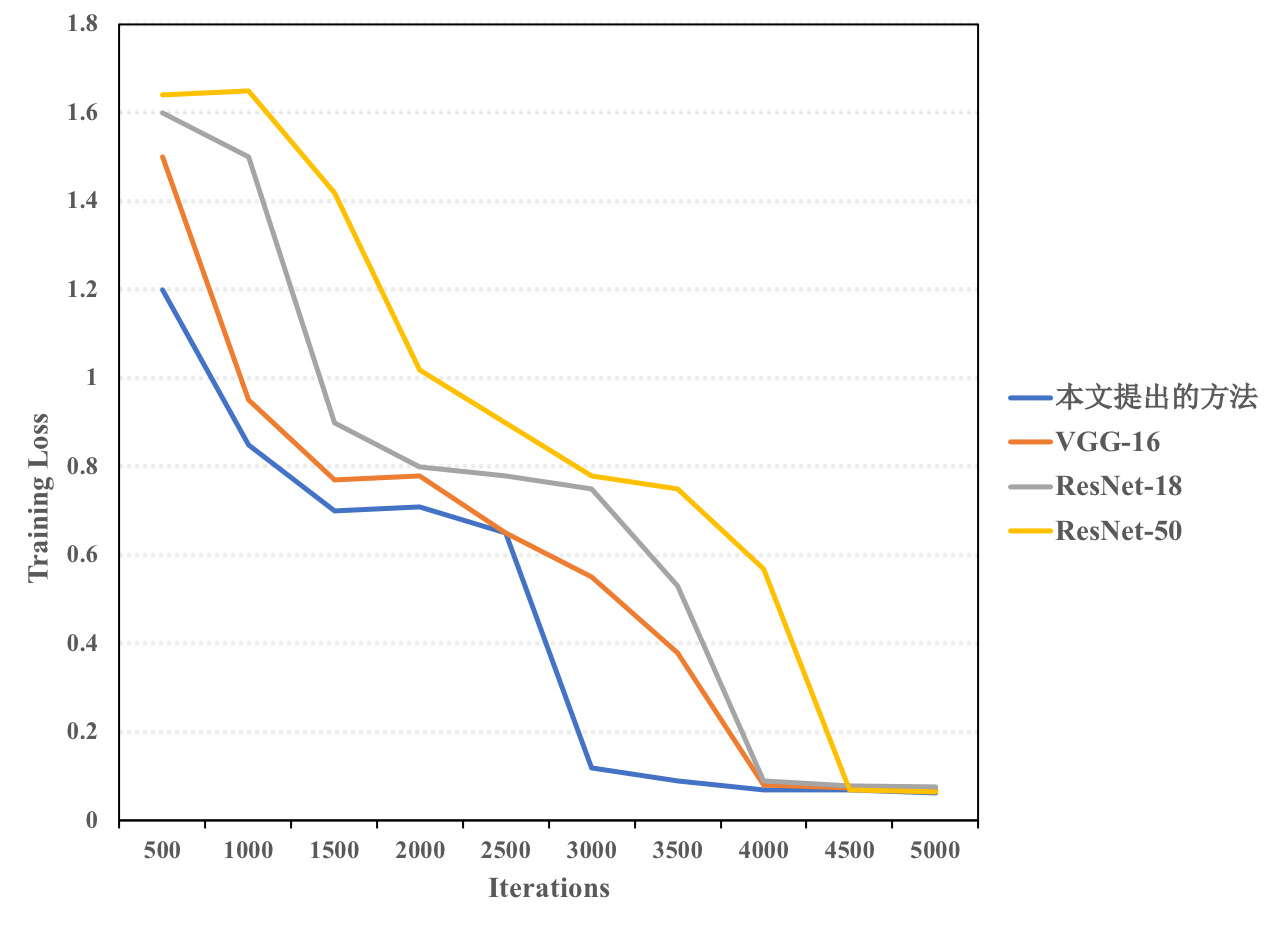


图3-2 不同网络训练迭代次数与训练误差关系图

图3-2记录了四种网络结构在训练集上进行训练时每500次输出的Training Loss值。在训练开始时由于网络的结构差异我们不难发现各个不同网络结构对应的初始的Training Loss值均不相同，大小在1.4到1.7之间。四种网络在经历了5000次迭代训练之后Training Loss值均稳定在了0.005左右，这说明四种网络在经过5000次迭代训练后均趋于收敛，四种网络在训练完成后可以用于后续测试集上的测试。值得注意的是四种网络在训练过程中Training Loss的下降方式以及模型整体的收敛速度各有不同。AttentionNeXt-18网络在训练过程中较其他网络最先达到收敛状态，在经过将近3000次迭代后Training Loss值大致稳定在0.005。其他三种网络结构几乎都是在经过4000次迭代后才逐渐趋于收敛。从图中还可以发现具有相同网络结构的网络，网络最终的收敛情况会和网络的深度有一定关系。在小规模数据集上解决甲状腺超声图像结节分类问题时较深的网络往往需要经过更多的迭代次数才能达到收敛。综合图3-2中的信息我们不难发现，本文所构建的AttentionNeXt-18较其他网络结构在Training Loss方面最先达到收敛状态。这说明本文提出的网络模型在解决超声图像分类问题时更容易训练，从而证明了网络设计在结构上的合理性。

在对四种网络在进行训练时因结构不同而导致在训练过程中相关重要参数的变化进行分析后，文章在该小节又针对以上四种网络在测试集上的获得的预测准确率进行阐述与分析。表3-4展示了本文构建的AttentionNeXt-18网络与其他三种经典深度卷积网络在测试集上的准确率。

表3-4 与不同深度学习网络进行对比

|  |  |
| --- | --- |
| 用于分类的深度学习网络 | 准确率 |
| VGG-16 | 81.26% |
| ResNet-18 | 81.28% |
| ResNet-50 | 85.30% |
| AttentionNeXt-18 | 90.28% |

从表3-4中我们可以发现本文构建的AttentionNeXt-18网络较其他三种经典的深度卷积神经网络在测试集上拥有最高的准确率，准确率可以达到90.28%。在三种经典的深度卷积网络结构中，ResNet-50拥有最高的预测准确率85.30%。ResNet-50在经典的深度卷积网络中的高准确率应该与其较深的结构有关，我们知道在一定范围内增加卷积网络的卷积层数可以有效提高整个卷积网络的分类效果。但从实验最终结果来看单纯增加卷积网络的深度在测试集上所取得的预测准确率并没有在使用了注意力机制后的准确率高。这提示我们在解决甲状腺超声图像结节良恶性分类问题时如果选择构建深度卷积网络来解决此问题，更应当注重网络结构的构建，单纯增加网络的深度不一定会取得满意的结构。VGG-16网络与ResNet-18网络在测试集上的准确率分别为81.26%和81.28%。从预测准确率的角度出发，这两种网络结构凭借较为简单且容易实现的结构可以在测试集上达到80%以上的准确率实属不易。但和本文构建的网络AttentionNeXt-18相比它们在预测准确率方面的表现就有明显差距。上文通过实验也说明了AttentionNeXt-18结构的复杂性并不会对整个网络在训练过程中的收敛速度产生较大影响。

综合图3-2和表3-4中的信息，我们可以看到本文提出的方法在解决甲状腺超声图像结节分类问题上具有一定优势。该网络较经典深度卷积神经网络不仅更容易训练而且有更高的预测准确率。

### 3.2.3 AttentionNeXt-18与迁移学习方法对比

从2017年开始逐渐有学者使用迁移学习的方法来解决超声图像中的分类问题，Jianning等人[10]率先使用迁移学习方法对甲状腺超声图像中的结节进行良恶性分类。

迁移学习简单的讲就是将别人已经训练好的的模型参数迁移到新的模型，再经过预训练来帮助构建模型。考虑到图像分类任务之间具有的相关性，所以通过迁移学习可以将先前模型已经学到的模型参数通过某种方式分享给新模型。这样做的好处是可以加快新模型的学习效率不用像大多数网络那样从零开始学习，同时加强网络的学习效果。

在该小节本文将AttentionNeXt-18与迁移学习方法进行对比。具体做法如下：在获得相同的预训练网络模型GoogLeNet之后使用本文创建的训练集对网络进行微调，再用测试集进行测试得到测试集预测准确率如表3-3所示。准确率为50次实验结果的平均值。

表3-5 与迁移学习下的GoogLeNet进行对比

|  |  |  |
| --- | --- | --- |
| 所用方案 | 用于分类的特征 | 准确率 |
| 迁移学习下的GoogLeNet | 由经过预训练的GoogLeNet提取的特征图 | 87.34% |
| AttentionNeXt-18 | AttentionNeXt-18提取的特征图 | 90.28% |

表3-5中清楚的展示了迁移学习下GoogLeNet与本文构建的网络AttentionNeXt-18在测试集上的预测准确率，分别是87.34%和90.28%。从预测准确率上讲，使用网络AttentionNeXt-18提取的特征图在用于甲状腺超声图像中结节良恶性预测时可以达到更好的实验效果。

综合表3-5中的实验数据我们可以看到在使用本文数据集进行实验时，使用迁移学习得到的结果并没有直接使用AttentionNeXt-18直接进行训练并分类的效果好。原因可能是用于微调训练的样本数量太少或是用于GoogLeNet进行预训练的数据集与超声图像领域的匹配度不够高。利用迁移学习的理论我们还可以将很多在其他领域具有较好分类效果的模型通过改进或融合之后应用在超声图像分类领域。

### 3.2.4 AttentionNeXt-18与传统机器学习方法对比

此小节将本文提出的方法与近十年内较为有代表意义的用于甲状腺超声图像中结节良恶性分类的传统方案在使用方法上、总样本数量上、准确性上、进行了比较。相应的比较结果如表3-6所示。

表3-6 甲状腺超声图像结节良恶性分类识别方案对比

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 使用方案 | 用于分类的特征 | 所用分类方法 | 总样本 | 准确率 |
| Lim et al [24]提出 | 大小、边缘信息、纹理回声、囊变特性 | 人工神经网络 | 190 个甲状腺超声图像样本 | 93.78 % |
| Savelonas et al [25]提出 | 各类形状特征 | 最邻近k-NN | 173个甲状腺超声图像样本 | 93.00 % |
| Iakovidis et al [26]提出 | 模糊强度直方图 | 支持向量机SVM | 250个甲状腺超声图像样本 | 97.50 % |
| Legakis et al [27]提出 | 纹理特征、形状特征 | 支持向量机SVM | 142个甲状腺超声图像样本 | 93.20 % |
| Luo et al [28]提出 | 应变率波形的功率谱 | 线性判别分析 | 98个甲状腺超声图像样本 | 87.50 % |
| AttentionNeXt-18 | AttentionNeXt-18提取的特征图 | 统计学方法 | 928个甲状腺超声图像样本 | 90.28 % |

表3-6将本文使用的超声图像分类网络的分类效果与其他传统机器学习方法的分类效果进行了比较，传统机器学习方法首先提取超声图像中结节的纹理特征、结节边缘特征、形状特征或是利用超声图像中的灰度特性作为特征，之后整合这些特征并将这些特征作为分类器的输入用作分类。从表中的数据可以看出在小样本下，传统机器学习方法在测试集上所取得的准确率可以高达97.5%,高于本文所用的方法。这也说明传统的机器学习方法中仍有我们值得借鉴的地方。

综合表3-6中的数据，我们发现传统机器学习方法在小样本上取得了较好的成果。原因可能是在小样本情况下手动提取的特征在一定范围内更能够表征结节的良恶性特点。但当它们在面对情况更加复杂的大样本的时，相应的性能是否依旧还会和在小样本下的情况拥有相同的表现还需进一步考证。

### 3.2.5 对比分析小结

通过对比分析我们发现在样本量较少时如果选择了有效的分类特征，采用传统方法的准确率较高。在使用深度学习网络解决此问题时本文建立的网络较传统网络有明显优势。总的来讲深度学习理论对于甲状腺超声图像分类问题的解决有十分重要的意义。在对比的过程中我们也发现在该准确率下模型还有许多可以完善的地方：1.网络结构层面，因为现阶段还没有理论证明在超声图像领域网络的最佳深度在什么范围故而可以探索不同深度对网络性能的影响。2.数据层面，坚持进行相关图像的收集工作以增加可用的有效训练样本数量。3.分类角度层面，我们的研究仅限于将甲状腺结节分为两类，可能是恶性的，或是良性的。我们还没有足够的样本按照六个不同的TI-RADS分数来训练网络，建立预测更精细粒度的模型。4.用于分类的特征层面，考虑是否通过加入较为底层的特征用于先验以提高识别准确率。

# 第四章 总结与展望

## 4.1论文总结

在第一章绪论部分，文章首先对甲状腺超声图像结节良恶性识别的应用背景及应用意义进行了详细的阐述，接着详细回顾了国内外近年来用于甲状腺超声图像中结节良恶性分类识别的各种方法并确定了本文解决此问题所用的思路。

论文在第二章详细描述了作者提出的甲状腺超声图像中结节良恶性分类识别算法。该算法与其他算法相比不同之处在于，首先作者使用改进了的超声图像预处理方法，其次作者使用自己建立的深度卷积网络AttentionNeXt-18用于图像中特征的提取及分类。建立深度卷积神经网络所用到的设计思想以及相关技术细节也在第二章中进行了详细的阐述。在第二章的最后一节作者用表格的形式展示了用于实验的训练集和测试集相应的图像数量分布。

论文在第三章通过使用Accuracy、Sensitivity、Specificity、AUC这四个指标对该分类器在测试集上的表现进行量化评估。实验结果表明该分类器可以较为准确地分辨甲状腺超声图像中结节的良恶性。在第三章的最后一节作者又通过三组对比实验更加全面的对分类器进行评估。

总的来讲本文创新点较为明确，它体现在使用新的预处理方法处理原始超声图像并构建了新深度卷积神经网络用来解决超声图像分类问题。最后作者通过对比分析也发现了一些可以使分类器得到进一步优化的方案如：考虑加入有效的底层特征进行先验、进一步增加训练集中图像的数量等。

## 4.2 展望

在甲状腺超声图像结节良恶性分类问题中传统的分类识别方法依然有极大的研究价值因为它们在小数据集中有近乎完美的表现。这提示我们可以通过研究传统分类方法中那些能够良好表征结节良恶性特征的指标进一步辅助深度卷积网络在大样本下的训练或是减少深度卷积网络的训练时间。

甲状腺超声诊断辅助软件是研究的最终目标，通过本文我们可以看到深度学习网络在其中可以发挥的巨大潜力。因此我们还需要结合超声成像领域的特点继续深入研究如何更加科学的使用深度学习算法。此外建立相应的甲状腺超声图像数据库也是一项十分重要的工作。

# 参考文献

1. Kwak J Y, Han K H, Yoon J H, et al. Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk[J]. Radiology, 2011, 260(3): 892-899.
2. Bamber J C, Daft C. Adaptive filtering for reduction of speckle in ultrasonic pulse-echo images[J]. Ultrasonics, 1986, 24(1): 41-44.
3. Iakovidis, Dimitris K., Eystratios G. Keramidas, and Dimitris Maroulis. Fuzzy local binary patterns for ultrasound texture characterization[C]. Proceedings of International conference image analysis and recognition. 2008: 1348-1359.
4. 王昊.甲状腺结节超声图像纹理特征提取及半监督分级方法研究[D].西南交通大学, 2017.
5. Nugroho H A, Rahmawaty M, Triyani Y, et al. Texture analysis for classification of thyroid ultrasound images[C]. Proceedings of Electronics Symposium (IES). 2016: 476-480.
6. Nugroho H A, Nugroho A, Frannita E L, et al. Classification of thyroid ultrasound images based on shape features analysis[C]. Proceedings of Biomedical Engineering International Conference (BMEiCON). 2017: 1-5.
7. 熊伟，龚勋，罗俊.基于局部纹理特征的超声甲状腺结节良恶性识别[J]. 数据采集与处理, 2015, 12(1):186 – 191.
8. Hijazi S, Kumar R, Rowen C. Using convolutional neural networks for image recognition[J]. Cadence Design Systems Inc, 2015, 53(5):13-19.
9. 孔德兴,吴法,马金连. 基于深度卷积神经网络自动识别甲状腺结节良恶性[P].中国专利:106056595A, 2016-10-25.
10. Chi J, 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.
11. Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell, Caffe: convolutional architecture for fast feature embed- ding[C]. Proceedings of the 22nd ACM International Conference on Multimedia. 2014: 675–678.
12. Byra M, Piotrzkowska-Wróblewska H, Dobruch-Sobczak K, et al. Combining Nakagami imaging and convolutional neural network for breast lesion classification[C]. Proceedings of Ultrasonics Symposium (IUS). 2017: 1-4.
13. 刘天娇, 孙卫东, 牛丽娟. 一种基于迁移学习与特征融合的超声甲状腺结节良恶性分类方法[P]. 清华大学中国医学科学院肿瘤医院. 106780448A. 2017-05-31.
14. Lina Pedraza, Carlos Vargas, Fabián Narváez, Oscar Durán, Emma Muñoz, and Eduardo Romero. An open access thyroid ultrasound image database[C]. Proceedings of International Symposium on Medical Information Processing and Analysis. 2015: 9287-9297.
15. D.Selvathi,V.S.Shnitha. Thyroid Classication and Segmentation in Ultrasound Images Using Machine Learning Algorithms[C]. Proceedings of International Conference on Signal Processing, Communication, Computing and Networking Technologies. 2011:34-46.
16. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]. Proceedings of IEEE International Conference on Computer Vision. IEEE Computer Society, 2015:2740-2748.
17. 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.
18. Mnih V, Heess N, Graves A. Recurrent models of visual attention[C]. Proceedings of Advances in neural. information processing systems. 2014: 2204-2212.
19. Newell A, Yang K, Deng J. Stacked hourglass networks for human pose estimation[C]. Proceedings of European Conference on Computer Vision. 2016: 483-499.
20. Wang F, Jiang M, Qian C, et al. Residual attention network for image classification[C]. Proceedings of Computer Vision and Pattern Recognition. IEEE, 2017:6450-6458.
21. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. arXiv. preprint arXiv. 2015:1502.03167.
22. He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks[C]. Proceedings of European Conference on Computer Vision. 2016: 630-645.
23. Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding[C]. Proceedings of the 22nd ACM international conference on Multimedia. 2014: 675-678.
24. Lim KJ, Choi CS, Yoon DY, Chang SK, Kim KK, Han H, Kim SS, Lee J, Jeon YH. Computer-aided diagnosis for the differentiation of malignant from benign thyroid nodules on ultrasonography[J]. Acad Radiol, 2008, 15(7):853–858.
25. Michalis A. Savelonas, Dimitris E. Maroulis, Dimitris K. Iakovidis, Nikos Dimitropoulos. Computer-aided malignancy risk assessment of nodules in thyroid US images utilizing boundary descriptors[C]. Proceedings of Panhellenic Conference on. IEEE. 2008:34-45.
26. Iakovidis DK, Keramidas EG, Maroulis D. Fusion of fuzzy statistical distributions for classification of thyroid ultrasound patterns[J]. Artif Intell Med, 2010, 50(1): 33–41.
27. Legakis I, Savelonas MA, Maroulis D, Iakovidis DK.Computer- based nodule malignancy risk assessment in thyroid ultrasound images[J]. Int J Comput Appl, 2011 33(1):29–35.
28. Luo S, Kim EH, Dighe M, Kim Y. Thyroid nodule classification using ultrasound elastography via linear discriminant analysis[J]. Ultrasonics, 2011, 51(4):425–431.
29. Rajendra Acharya U, Vinitha Sree S, Swapna G, Gupta S, Filippo M, Garberoglio R, Witkowska A, Suri JS. Effect of complex wavelet transform filter on thyroid tumor classification in three-dimensional ultrasound[J]. Proc Inst Mech Eng H J Eng Med, 2013, 227(3):284–292.
30. Xie S, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]. Proceedings of Computer Vision and Pattern Recognition (CVPR). 2017: 5987-5995.

# 致谢

本论文是在陈俊颖老师的指导下完成的，陈老师在论文选题、数据选取、卷积神经网络模型建立、论文结构等诸多方面给出了很多十分有益且中肯的建议。在和老师讨论的过程中被老师严谨的治学风格和认真负责的态度所影响，同时也学到了很多专业知识，提升了专业素养。在此向导师致以崇高的敬意和诚挚的感谢。

作为一名软件学院的学生还要感谢学院老师的付出，有了他们的陪伴才使得我在面对困难时坚持走到了最后。

同时还要感谢和我一起参加国创项目训练的同学们，他们在我两年来在超声图像处理领域的学习中给了我巨大的鼓励和帮助，这份友情将永驻我心。

最后我要感谢在物质和精神上给予我全力支持的家人，是你们的鼓励让我不畏艰难的不断前行最终完成四年的学业。