中国科学院大学

University of Chinese Academy of Sciences

硕士学位论文

构建高效的肝脏超声造影连续勾画软件和迁移模型

作者姓名：

指导教师：

学位类别：工学硕士

学科专业：模式识别与智能系统

培养单位：中国科学院自动化研究所

2020 年 9 月

Construct Efficient Delineation Software and Transfer Model for Liver Contrast-enhanced Vedio Dataset

A thesis submitted to the

University of Chinese Academy of Sciences

in partial fulfillment of the requirement

for the degree of

Master of Engineering

in Pattern Recognition and Intelligent Systems

By

Institute of Automation, Chinese Academy of Sciences

Sept, 2020

中国科学院大学
学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明或致谢。本人完全意识到本声明的法律结果由本人承担。

作者签名：

日　　期：

中国科学院大学
学位论文授权使用声明

本人完全了解并同意遵守中国科学院大学有关保存和使用学位论文的规定，即中国科学院大学有权保留送交学位论文的副本，允许该论文被查阅，可以按照学术研究公开原则和保护知识产权的原则公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

涉密及延迟公开的学位论文在解密或延迟期后适用本声明。

作者签名：　　导师签名：

日　　期：　　日　　期：

摘 要

人工智能正在取代计算科学和工程学中的标准算法，超声造影已经在欧洲和亚洲用于心脏和腹部成像数十年，应用前景光明。基于这两个事实，深度学习在超声造影数据分析领域将大有可为。本研究围绕着超声造影视频数据的深度学习方法应用，在数据集清洗和视频学习框架两个话题上展开，详细研究了半自动追踪软件的开发设计和多种卓越的视频网络在超声造影数据上的迁移性能。文章主体分为四个章节：第一章介绍超声造影和超声影像组学的概念，系统回顾了基于机器学习的超声造影定量化分析和研究成果，说明了研究数据清洗和迁移学习的必要性；第二章围绕自主研发的基于点跟踪算法的半自动病灶连续勾画软件，介绍了造影视频的运动校正方法，比较不同特征点提取算法的性能，简要介绍软件的构建过程和使用；第 3 章围绕迁移学习和视频分类模型展开，在区分肝细胞癌和胆管细胞癌的分类任务下，纵向分析比较了 4 大类、16 个视频分类模型的迁移效果，并纵向比较了采样帧数、采样方法、数据清洗和全局特征对迁移效果的影响，本章节希望帮助初学者了解如何对超声造影数据集进行清洗、如何选择模型和训练参数，并在较短的时间内达到良好的分类性能；第四章总结了研究的不足，并展望了未来深度学习在超声造影和其他医学影像的的发展。虽然这些方法只在局灶性肝脏病变数据集上测试优化，但方法本身并不涉及肝脏器官图像表征的先验知识，原则上感兴趣的连续跟踪和深度学习框架可以应用到其他器官的造影序列的处理和分析中。

关键词：肝脏超声造影，图像序列处理，迁移学习，视频深度学习网络，图形界面编程，点跟踪, 数据清洗

Abstract

Artificial intelligence is replacing standard algorithms in computational science and engineering. Contrast-enhanced ultrasound(CEUS) has been used for cardiac and abdominal imaging in Europe and Asia for decades with promising future. Based on these two facts, deep learning has great potential in the field of CEUS vedio dataset analysis. This study revolves around the application of deep learning methods for CEUS video data, focusing on dataset cleaning and video learning framework choosing while introduces the development and design of semi-automatic tracking software and the transfer performance of multiple superior video networks. The main body of the article is divided into four chapters: the first chapter introduces the concepts of CEUS and radiomics, systematically reviews machine learning-based quantitative analysis researches in CEUS; illustrates the need for more research on data cleaning and transfer learning. Chapter 2 revolves around independent development point-based tracking algorithm for semi-automatic continuous outlining of liver focal lesions, introduces motion correction methods for imaging videos, and compares the performance of different feature points-based methods, briefly explains the process of constructing the software and its user manual. Chapter 3 is built around transfer learning and video classification models. We analysis and compare 4 major categories and 16 video classification models under the classification task of differentiating between hepatocellular carcinoma and cholangiocellular carcinoma, as well as compare the effects of the number of sample frames, sampling methods, data cleaning, and global features on transfer learning. This section hopes to help beginners understand how to clean CEUS dataset and know to use which model is better in CEUS. Chapter 4 summarizes the research deficiencies and looks at the future of deep learning in CEUS and other medical imaging. Although these methods only apply to focal liver lesion dataset, the methods themselves does not use any prior knowledge of the liver organ, in principle they can work for other medical problems.

Keywords: Liver Ultrasound Imaging Analysis, Image Sequence Processing, Transfer learning, Video DNN Model, Graph Guided Programming, Point-based Tracking, Data Cleaning

目 录

第 1 章 研究背景和研究意义 ····· 1
1.1 引言 ····· 1
1.2 超声造影 ····· 4
1.3 超声造影的定量分析 ····· 7
1.3.1 影像组学 ····· 7
1.3.2 动态超声造影 ····· 9
1.3.3 深度学习 ····· 10
1.3.4 研究现状 ····· 13
1.4 研究意义 ····· 17
1.4.1 数据清洗 ····· 17
1.4.2 迁移学习 ····· 18
第 2 章 基于点跟踪的半自动连续勾画软件 ····· 21
2.1 引言 ····· 21
2.2 跟踪算法 ····· 22
2.2.1 运动校正 ····· 22
2.2.2 跟踪性能测试 ····· 23
2.3 软件构建 ····· 26
2.4 软件使用 ····· 28
第 3 章 增强超声视频数据迁移学习探讨 ····· 33
3.1 引言 ····· 33
3.2 研究背景 ····· 35
3.2.1 迁移学习 ····· 35
3.2.2 数据集 ····· 37
3.2.3 图像网络 ····· 37
3.2.4 视频网络 ····· 40
3.3 实验设计 ····· 42
3.3.1 数据预处理 ····· 42
3.3.2 网络训练 ····· 43
3.4 实验结果 ····· 45
3.4.1 基本模型 ····· 45
3.4.2 采样帧数 ····· 48

3.4.3 密集采样 ····· 50
3.4.4 数据清洗 ····· 51
3.4.5 non-local 全局特征 ····· 51
第 4 章 总结展望 ····· 53
4.1 研究局限 ····· 53
4.2 展望 ····· 56
参考文献 ····· 61
作者简历及攻读学位期间发表的学术论文与研究成果 ····· 61
致谢 ····· 63

图形列表

1.1 一位 64 岁的男性，患有局灶性结节性增生。A. T1 加权磁共振图像显示肝脏右叶高血管结节性病变（箭头）。B. 在肝胆相中，病变的信号强度（箭头）高于邻近的肝实质。C. 在造影增强超声（US）的动脉期，病变表现为高血管性（箭头）。D. 在门静脉期，病变相对于邻近肝脏呈现等回声性（箭头）。经 US 引导下活检证实病变为局灶性结节性增生。 ····· 6

1.2 DCE-US 例图。（a)(b) 分别为治疗前神经内分泌转移和正常肝实质的 CEUS 图像和相应的时间-强度曲线,(c)（d）治疗 3 个月后神经内分泌转移和正常肝实质的 CEUS 图像和相应的时间-强度曲线。白色箭头表示转移病灶。曲线的起伏是由呼吸伪影引起的 ····· 11

2.1 两种 CEUS 界面展示。(a) 单幅显示，图像中只有 CEUS 一个录制窗口，(b) 双幅显示，图像的一侧为 US 视频窗口，另一侧为 CEUS 视频窗口，两个窗口显示同一成像平面的不同信号分布。····· 22

2.2 软件运行效果图。软件总体分为四个区域，菜单，跟踪控制组件，视频播放窗口和视频控制组件，图像显示了一例勾画完的视频在软件中的显示效果，图下方区域的彩色标记代表着不同跟踪任务在视频中的位置。软件界面简洁紧凑 ····· 29

3.1 网络测试结果 Heatmap 图，水平方向的刻度代表病例编号，竖直方向为网络结构名，每个位置的方块代表该网络对于此病例的是 ICC 的评分，图像的左半部分是 HCC 病例，右半部分是 ICC 病例。 ····· 47

表格列表

1.1 CEUS 机器学习方法汇总表 ····· 14
1.1 续表。 ····· 15
1.1 续表。 ····· 16

2.1 案例 1 的 CEUS 视频跟踪速度和精度 ····· 24
2.2 案例 1 的 US 视频跟踪速度和精度 ····· 25
2.3 案例 2 的 CEUS 视频跟踪速度和精度 ····· 25
2.4 案例 2 的 US 视频跟踪速度和精度 ····· 25
2.5 软件操作界面模块组成和功能表 ····· 30
2.5 续表。 ····· 31

3.1 固定输入 8×224^2 时模型迁移性能 ····· 46
3.2 不同采样帧数下模型迁移性能表 ····· 49
3.3 密集采样时模型迁移性能表 ····· 50
3.4 完整训练集上模型迁移性能表 ····· 51
3.5 non-local 模型迁移性能表 ····· 52

第 1 章　研究背景和研究意义

1.1　引言

当前世界正在发生两种模式变化：1）人工智能（Artificial Intelligence, AI）在解决问题方面正在取代人类；2）AI 也正在取代计算科学和工程学中的标准算法，这个趋势在医学领域也同样适用。超声造影剂已经在欧洲和亚洲用于心脏和腹部成像数十年，较为普遍地应用于放射学中。本研究围绕着超声造影视频数据的深度学习分析主题，在数据集清洗和视频学习框架两个话题上做探讨和分析。

本章介绍研究背景和意义，分三小节分别介绍计算机辅助医学影像分析的应用和发展，第一节介绍超声造影和超声影像组学的概念，说明计算机自动分析方法对于应用范围和数据量快速增长的超声造影十分迫切；第二节介绍了基于机器学习的超声造影定量化分析和研究成果；第三节说明超声造影数据的清洗、自动追踪软件的开发和系统研究视频网络在超声造影中迁移性能对于该领域发展的重要意义。

“人工智能” 这一术语的首次提出是在 1956 年夏，当时麦卡赛、明斯基、罗切斯特和申农等为首的一批有远见卓识的年轻科学家在一起聚会，共同研究和探讨用机器模拟智能的一系列有关问题。人工智能的核心问题包括构建能够跟人类似甚至超卓的推理、知识、规划、学习、交流、感知、移物、使用工具和操控机械的能力等。目前弱人工智能已经有初步成果，甚至在影像识别、语言分析、棋类游戏等方面的能力达到了超越人类的水平，而且人工智能的通用性代表着，能解决上述的问题的是一样的 AI 程序，无须重新开发算法就可以直接使用现有的 AI 完成任务，与人类的处理能力相同。然而，想要实现具备思考能力的强人工智能还需要时间研究。目前比较流行的方法包括统计方法，计算智能和传统意义的 AI。AI 已经取得长足的发展，成为一门广泛的交叉和前沿科学。

机器学习（Machine Learning, ML）是人工智能的一个子集，它可以在最少的人工干预下自行学习数据，以分类或预测未来或不确定的条件。由于 ML 是数据驱动的学习，所以它被归类为非符号 AI，可以对未知的数据进行预测。传统机器学习方法有支持向量机（Suport Vector Machines, SVM），决策树 (Decision Trees)、随机森林（Random Forests）、贝叶斯学习（Bayesian learning），传统机器

学习方法在图像分析领域的应用离不开各种人工设计的图像特征（如灰度共生矩阵，小波，形状描述子）和特征选择方法（如皮尔逊相关系数，L1 正则化）。人工神经网络（Artificial Neural Network, ANN）是一种受大脑启发的算法，它由具有连接节点的层组成，由输入层、输出层和隐藏层组成，在训练过程中，通过反向传播等学习算法，通过参数化权重来确定每个节点的值。ANN 有时训练的结果是局部最小值，或者只针对训练过的数据进行优化，从而导致过拟合问题。后来，研究者通过在输入层和输出层之间堆叠具有连接节点的多隐藏层，将 ANN 扩展为深度学习（Deep Learning，DL）。在分类和回归等预测任务中，DNN 一般表现出比浅层网络更好的性能, 此外，大数据和图形处理单元的出现也使得 DNN 可以高效部署。

在医学领域的主要研究问题可分为两类：二元问题（分类），如疾病是否复发，患者是否活在一定时间阈值以上等; 生存分析，即能够显示风险因素或治疗是否影响事件的时间。在医学领域流行的机器学习方法有：使用逻辑回归拟合线性变量的系数；使用 SVM 拟合非线性问题，由于 SVM 不同的核函数（例如径向基础函数）可以将数据映射到更高维度并且最大化两个类之间的边距，通过正则项容忍边界错误，具有较好的鲁棒性和泛化性，在深度学习没有大放异彩前应用非常广泛；随机森林基于决策树，可以将假设表示为连续的” 如果”，类似于人类推理，解释性好，也较多使用；CNN 网络由于强大的性能和端到端的学习方式，在使用中可以灵活的组合自定义适合数据量的小网络，也可以通过迁移学习使用训练好的复杂网络，因而近几年的相关研究非常多。

机器学习方法中常用的统计指标有：精度（Accuracies）、ROC 曲线下面积（Area Under the ROC Curve，AUC）、灵敏度（Sensitivity）、特异性（Specificity）、正预测值（Positive Predictive Values，PPV）、阴性预测值（Negative Predictive Values, NPV）、错误预测率（False Predictive Rates, FFR）和假阴率（False Negative Rates, FNRs）。常见的测试方法有“one-out”、交叉验证和随机分为训练集和测试集或者按规律分为训练集和测试集（如按时间或数据来源）。

机器学习方法研究模型能否拟合数据，通过模型的性能表明输入和输出之间是否可以通过确定的模型表达，而统计方法研究数据本身的分布能否定量或定性的描述。在现实应用中，机器学习方法如果不能在大量不同的测试数据中表现优良，是无法较为中肯的说明输入和输出之间存在明确的关联并且被模型

学习到，数据拟合的效果一方面强依赖于训练数据，另一方面受限于模型类型和参数设置；而现有的统计学习模型无法拟合复杂数据的生成过程，说明某个因素（例如是否有糖尿病）对数据生成（例如病人的肝脏造影视频）是否有重要作用，不具备高层次的抽象分析能力。目前，虽然有一些应用深度学习模型很好的应用在临床，比如医学图像分割和重建，但是对于用模型性能说明医学假设是否成立的研究，尚且存在一些严重的障碍：

1. 无法通过模型拟合效果差拒绝一个假设不成立。我们不能说明当一个模型拟合效果差是数据和问题之间无重要的因果关系，很有可能只是算法的选择不如人意，或者数据集的规模不能与模型适配，或者小规模的数据集存在一些严重的噪声，导致拟合失败。

2. 无法通过模型卓越的表现肯定假设成立并且模型能在其他数据集上表现优越。面对很多基于用深度学习或者机器学习方法的表现效果说明假设是否成立的医学研究，尤其是在单中心回顾性的小数据集上，我们不能确定模型表现良好时是否因为学习了数据集上某种无关的信息影响了模型的判断，比如亚洲人的肝癌很多是有乙肝发展到肝硬化再病变而来，肝脏实质的超声纹理感强，如果机器训练数据都是亚洲人，很可能模型学习到是肝脏背景纹理，而非病灶特征，无法迁移到脂肪肝高发的欧洲。

3. 无法确定这种关联是否重要且稳健。我们不知道标签作为影响输入数据生成的一项，它的确切的作用能力有多大，尤其是模型强大到可以学习微弱的相关性从而有较好性能的时候。比如研究是否在超声检查前喝水对肝脏超声成像影像时，可能模型可以通过图像判断是否喝水，但这是否是图像生成重要因素就有待商榷，这样的研究会使得影响数据生成的主要因素可能被忽略，反而弄巧成拙，比如随机更改图像中的一个像素值，机器模型就把原本正确识别的猫错误当成狗的经典案例。

虽然本研究探讨的所有问题都假设了超声造影视频具备解决问题所需的完备信息，即只要算法的拟合性能足够和数据集足够庞大时，通过超声造影训练的模型可以取得超人的接近金标准的结果，比如利用超声视频分类肝细胞癌和胆管细胞癌这个任务中，超声视频包含了疾病诊断的完整信息，只要分类器足够强大，就可以实现 100% 的正确分类结果。而统计学习探讨的问题可能是超声造影真的是完备的吗，包含了完整的分类任务信息吗。我完全认可，作为间接非接触

的成像技术，超声造影可能是非完备的，需要其他的临床资料辅助才可以实现100% 正确率的事实，一个很好的说明其非完备的例子就是病灶小于图像的分辨率时，图像将无能力表达这个信息，但这个例子也是一个非常极端的不适用临床的假设。

本研究并不尝试克服目前机器学习方法应用到医学数据的这三问题，但在此提出，是希望聪明的人可以去思考这样的问题，本研究也正是为了帮助人们节约时间精力，从而做更必要的研究。医学图像的存储和交换遵循即医学数字成像和通信标准（Digital Imaging and Communications in Medicine，DICOM），该标准的第一个版本于 1985 年发布，DICOM 的文件除了其他图像相关数据（例如用于捕获图像的设备和医疗背景），此格式还具有关于患者的受保护的健康信息，例如姓名，性别，年龄。本研究使用的数据均来自 DICOM 文件，在使用时通过软件转化为普通的视频图像格式，并将患者信息隐去。

1.2 超声造影

超声波是指频率高于人类听觉极限的声音。超声成像是利用目标高频声波通过组织传输，并将反射的声波用计算机转换为解剖图像。产生的图像是由组织对超声光束的反射和散射（受不同密度的组织之间的界面影响）和衰减（由超声频率和组织密度决定）的数量决定的。多普勒超声利用多普勒移位原理确定血管和分流内的血流方向和估计速度，可用于生成波形或用彩色代码显示血流的解剖图像。

US 是评估软组织（包括肝脏）的重要成像方式, 它的优点有：1）设备可广泛使用，且便于携带，可以在床边评估病人或在手术室协助手术；2）US 检查是安全的，如果在适当的参数下进行，对包括胎儿在内的敏感组织没有已知的损伤，因此 US 可以用于评估孕妇患者；3）US 通常可以快速完成，不需要特殊的准备（除了空腹评估胆囊外），可以不需要静脉注射，只需要在皮肤上轻轻施压，可用于指导诊断程序，如活检或引流积液；4）相对 CT 和 MRI 更为廉价。

尽管 US 有许多优点，但它也有重要的局限性：1）US 依赖于操作者，而且重复性有限；2）光束不容易穿过组织-气体或组织-骨的边界，因此，覆盖的肋骨或充满气体的肠道环总会在肝脏内产生盲点；3）肥胖患者的上覆脂肪可能会降低图像的质量，比如肝脏内的脂肪或纤维化可能会分散 US 光束，而脂肪可能会

减弱 US 光束，从而有可能掩盖肝脏内的病变。

对比增强超声 (CEUS) 是通过静脉注射超声造影剂的特殊超声成像方法，由于超声造影剂中的微泡成分包裹的气体与身体的软组织周围的回声性存在很大差异，使用造影剂会增强超声波的反向散射，增加了超声波图的对比度。CEUS 可以持续记录观察超声界面上血流灌注情况，能提供组织血管、组织灌注甚至内皮壁功能有关的重要定量信息。CEUS 在成人中被认为是安全的，它的安全性与 MRI 造影剂相当，并且优于对比 CT 扫描中使用的放射性对比剂，分为心脏造影成像和腹部超声造影两大类。腹部 CEUS 可评估肝、肾脏、脾、胰腺、肠膀胱、腹部血管，用于帮助诊断各种腹部/骨盆疾病，例如：常规超声检查发现肝脏病变、肝脏血流异常、肝硬化、肾脏病变、腹部外伤、脾脏病变，广泛地用于肝肿瘤的表征，广泛地在欧洲和亚洲地区应用，在部分疾病诊断中可以取代 CT 扫描或 MRI。对比增强超声可用于对器官中的血液灌注进行成像，广泛用于肝脏肿瘤的表征。1.1是一位男性的 CEUS 和 MRI 成像，摘自论文**?**，希望能帮助读者理解 CEUS 成像效果和不同 MRI 的图像风格。

超声造影在 1970 年左右出现，当时研究者在静脉注射靛青绿后观察右心室超声心动图，但由于与 X 射线方法相比，超声在没有造影剂的情况下也能很好地工作，在较长的一段时间内并没有取得真正的进展。1991 年，一家德国造影剂制造商开发并批准了造影剂 Echovist 用于不适合通过肺毛细血管的超声心动图检查，此后，CEUS 获得更多研究者的关注。2001 年，美国制造的造影剂 SonoVue 在许多欧洲国家的引入，CEUS 的应用取得了突破性进展。2004 年，欧洲生物医学超声学会联盟（European Federation of Societies for Ultrasound in Medicine and Biology, EFSUMB）发布了第一份 CEUS 指南（**?**），2015 年惰性气体六氟化硫 (SF6) 和棕榈酸外壳的微气泡 (SonoVue, Bracco Geneva, CH) 被美国食品药品监督管理局（Food and Drug Administration, FDA）批准用于成人和儿童肝脏肿瘤的诊断成像，该事件被**?** 评价为对比增强超声在美国应用的里程碑。

CEUS 在肝内的观察时长持续在 2 分钟到 6 分钟不等，从造影剂打入开始，数据记录时间一般持续到前 1-2 分钟，美国癌症学会对肝内 CEUS 的操作标准指出依据血流的灌注效果可以分为动脉期 (Arterail Phase)，门脉期 (Portal Venous Phase) 和消退期 (Lata Phase)(**?**)。通常，一次肝脏 CEUS 视频可以肉眼看到在注射造影剂后 10 至 20 秒，造影剂出现在肝脏的动脉系统中，随后，门静脉阶段

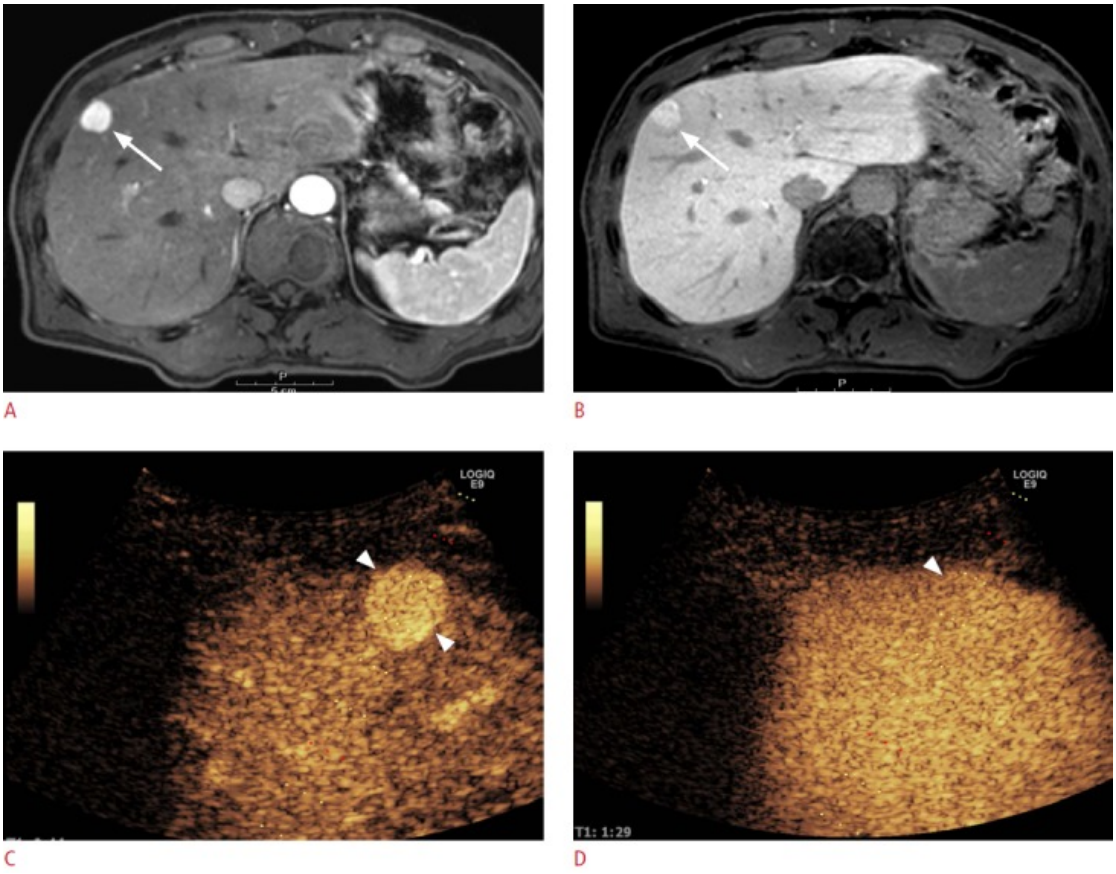


图 1.1 一位 64 岁的男性，患有局灶性结节性增生。A. T1 加权磁共振图像显示肝脏右叶高血管结节性病变（箭头）。B. 在肝胆相中，病变的信号强度（箭头）高于邻近的肝实质。C. 在造影增强超声（US）的动脉期，病变表现为高血管性（箭头）。D. 在门静脉期，病变相对于邻近肝脏呈现等回声性（箭头）。经 US 引导下活检证实病变为局灶性结节性增生。

Figure 1.1 A 64-year-old male with focal nodular hyperplasia. A .T1-weighted magnetic resonance image showed a hypervascular nodular lesion in the right lobe of the liver (arrow). B. In the hepatobiliary phase, the lesion showed a higher signal intensity (arrow) than the adjacent liver parenchyma. C. In the arterial phase of contrast-enhanced ultrasonography (US), the lesion showed hypervascularity (arrowheads). D. In the portal venous phase, the lesion showed iso-echogenicity (arrowhead) relative to the adjacent liver. The lesion was confirmed as focal nodular hyperplasia by an US-guided biopsy.

持续 30 至 120 秒，注射后约 3 至 4 分钟消退。通常 CEUS 上的事件的时间是常规记录其实际时间，以秒为单位，在扫描仪屏幕上可见的计时器显示。更好了解 CEUS 的临床应用可以从以阅读下几篇文章：**?** 回顾了 CEUS 的发展历史；**?** 回顾了自 1990 至今共三十年来超声造影剂的更迭过程，什么是理想的超声造影剂和非线性行为，市场上销售的主流造影剂和未来可能的改进；**?** 介绍了 US 成像可调节参数的含义：如振幅，频率，机器指标，增益，图像深度，焦点，帧速，图像后处理，谐波频率滤波，脉冲反转，功率调制和加速对比脉冲序列等信号调制技术；EFSUMB 非肝脏应用超声造影（CEUS）临床实践指南和建议 (**?**)；对比增强超声（CEUS）肝脏成像报告和数据系统（LI-RADS）(**??**)；最新的应用综述**?**。

目前，CEUS 的研究人员正在积极推广这项技术，有很多综合性报告和荟萃的分析都在试图总结之前零散分布的研究成果，以期促进整个学科的规范化和广泛应用。**?** 建议在数据收集之前要进行充分的测试，以达到最佳的信噪比，并减小操作带来的差异。来自 EFSUMB、WFUMB 和 CEUS LI-RADS 工作组的专家们创建了一个讨论论坛，根据已发表的证据和最佳个人经验，对 CEUS 检查技术进行标准化，对如何在临床中使用 CEUS 提供了一般性建议 (**?**)。**?** 指出 CEUS 的主要的诊断特征是：1）血管结构（在早期冲洗阶段评估）；2）与邻近组织相比，病变的对比度增强（洗入和洗出的时间过程)。**?** 在筛选 1483 篇文章后，纳入了 6 项队列研究和 10 项描述性研究，通过病理学最终诊断为良性或恶性肾脏肿块的纳入研究的汇总数据得出：CEUS 有可能成为替代是目前鉴定肾脏肿块的金标准（对比增强计算机断层扫描和对比增强磁共振成像）的有价值的选择。

1.3　超声造影的定量分析

1.3.1　影像组学

影像组学 (**?**) 是定量图像分析的一个新兴领域，其目的是将图像的大规模数据挖掘与临床和生物终点联系起来，基本思想是医学图像的信息比人眼所能分辨的要丰富得多。定量成像特征，也称为” 影像组学特征”，可以为不同的成像方式（如 MRI、CT、PET、超声等）提供更丰富的肿瘤表型的强度、形状、大小或体积、质地等信息。**?** 指出基于肿瘤活检的检测方法提供的肿瘤表征是有限的，因为提取的样本不一定能代表整个患者肿瘤的异质性，而影像组学可以通过提取相关的成像信息，全面评估肿瘤的三维（3D）景观。研究**??** 表明将众所周

知的机器学习方法应用于从医学图像中提取的影像组学特征，可以宏观地解读许多生理病理结构的表型，理论上解决了从表型推断基因型的逆向问题，提供了有价值的诊断、预后或预测信息。

影像组学的定义是仿照基因组学和蛋白质组学而来，表达了基于医学图像的个性化医学的明确意图。它的根源可以追溯到医学图像的计算机辅助检测/诊断?? 然而，随着最近医学成像采集技术的进步与处理的多样性，影像组学正在确立自己作为一个不可或缺的图像分析和理解工具的应用，超越诊断到预后和预测方法，以个性化患者的管理和治疗。与 CAD 的主要区别之一在于影像组学强调在生理病理结构的当前特征和其时间演变之间建立联系，以便确定个性化的治疗方法。影像组学特征和其他医学信息有较好的互补性，影像组学特征往往融合其他临床或实验资料可以提升分类性能。目前，影像组学已被应用于许多疾病，包括癌症和神经退行性疾病等等。在过去的几年里，发表的论文数量几乎成倍增长。

? 将影像组学的处理流程总结归纳为以下部分：(1) 影像数据的获取；(2) 肿瘤区域的标定；(3) 肿瘤区域的分割；(4) 特征的提取和量化；(5) 影像数据库的建立；(6) 分类和预测。影像学特征往往是人工设计的，用于捕获成像数据中的特征模式，包括基于形状、第一和第二阶的统计决定因素和基于模型的特征（例如分形）特征。基于功能的方法需要通过手动、半自动或自动方法对感兴趣区域（ROI）进行细分。基于形状的要素是描述 ROIS 的形状、大小和表面信息的区域的外部表示。68 典型指标包括球面性和紧凑性。3，43，69，70 一阶要素（例如均值、中位数）描述 ROIS 的总体强度和变化，而忽略空间关系。8，24 对比的二阶（纹理）特征可以提供体素之间的关系。纹理特征可以从不同的矩阵中提取，例如，灰级共发生矩阵（GLCM）、灰级运行长度矩阵（GLRLM）等 35、46、71 语义特征是可以从医学图像中提取的另一种特征。这些功能描述了放射学工作流程中通常使用的图像的定性特征。

目前，应用研究较为广泛的是基于 CT 和核磁断层图像的特征提取影像组学。这方面的影像组学的研究开展的条件较为成熟，主要体现在 CT 和核磁的图像具有三维断层图像，并且这两种成像方法对于病灶区域的边界定义较为清楚。所以，很适合使用基于特征提取的影像组学对这些图像进行分析。主要是因为 CT 和核磁图像有利于提取形状和纹理等特征，并且其数据量相对较少。由于超

声图像边界的模糊、二维的特性和操作者依赖性，使得基于特征提取的方法很难实施，特别是边界的特征定义具有很高的难度和不准确性。另一方面，由于超声数据量的优势，使得运用基于以深度学习为代表的人工智能的方法变得更有可能。

在我看来，影像组学的研究主要集中在 CT 和 MRI 数据, 很多研究成果无法直接应用在 CEUS 上。虽然在医学中，超声造影、核磁共振成像和断层成像虽然在应用方面有较多的重合，如儿童和成人的肝脏、肾脏或膀胱，单次成像数据都是体量庞大的非接触性成像序列，但是 CEUS 和 CT、MRI 序列数据含义和图像风格不同。在数据含义上，超声造影表征同一病灶切面随着时间变化造影剂的分布，CT 和 MRI 表征病灶的三维空间切片；在图像风格上，超声造影的切片相比 CT 和 MRI 图像结构性差，背景噪声大，空间分辨率低，没有清晰固定的形态。因而影像组学中提出的 3D 纹理特征、灰度特征等较低层次的局部特征不能用来的描述造影切片；影像组学中基于连通性的 CT 和 MRI 领域的自动分割算法不能用来勾勒 CEUS 病灶；在操作中由于 CEUS 在增强初期形态不明显，在动脉期病灶和背景差异小，病灶通过人工也无法精确勾画, 使得形状描述特征噪声大。目前，基于图像特征的建模方法往往通过提取不同造影时期特定的几张 2D 图像的影像组学特征实现，比如只使用开始增强图像，半达峰图像，峰值图像，峰值后 5s 图像，峰值后 10s 图像，消退图像这些特定的帧分析。

1.3.2　动态超声造影

相比超声影像组学这个概念，在 CEUS 的量化分析中，动态对比增强超声（Dynamic Contrast-Enhanced Ultrasound, DCE-US) 的知名度更广，研究应用更多。DCE-US 依靠 UCAs 的回声特性以反向散射系数为特征，将其建模为微泡总散射截面及其浓度的函数, 通过应用指示剂稀释理论（Indicator Dilution Theory，IDT）在时空两个维度定量描述血流灌注量。使用 DCE-US 量化生理指标在心脏病学和肿瘤学方面应用较为广泛，和机器学习结合使用后有望取得了更好的信息组合 (**?**)。DCE-US 中最著名的应用是提取时间-强度曲线（Time-intensity curve，TIC）,TIC 描述了 UCA 穿过被调查的器官或组织而引起的信号增强（声音强度）的时间演变，通过指示剂稀释理论和药代动力学模型可用估算和血容量，血流量，灌注，外渗、分子结合有关的定量参数 (**??**)。时间-强度曲线一般呈现出一个对数正态曲线，常用的描述特征有曲线下面积（Area under the curve，AUC），出现时

间（Appearance time, AT），峰值强度（Peak enhancement, PE），达峰时间（Time-to-peak, TTP），开始增强时间（Wash-in time，WIT），开始消退时间（Wash-out time, WOT），增强速率（Wash-in rate, WIR），消退速率（Wash-out rate, WOR)。**?** 介绍了 TIC 曲线生成的一般流程，包括感兴趣区域勾画，背景扣除，超声信号线性化等操作，并说明了量化结果在心脏，骨骼肌，皮下脂肪组织，肾脏和脑的应用。更为权威的 DCE-US 综述报告可参阅**?**。

图像1.2是非常典型的 CEUS 量化应用, 引用自论文究**?**, 放在正文中以帮助读者更好理解 DCE-US，两组图像采集在治疗前和治疗后 3 个月，记录了静脉内注射后持续 2 分钟的影像，通过在肝转移灶和邻近的正常肝实质中标记了感兴趣的区域（Region of Interest, ROI），测量随时间的变化每个 ROI 的峰值图像强度和造影达到峰值的时间。

由于超声造影视频单个数据量较为庞大，相比超声，需要花费更多的时间读片，并且，这种技术的一个关键挑战是成像结果存在显著的差异，**?** 回顾基于微气泡对比增强超声图像的组织灌注量化的潜在变异源, 指出这些来源可分为以下三类：1) 与扫描仪设置有关的因素，包括透射功率、透射焦深、动态范围、信号增益和透射频率；2) 与患者有关的因素，包括体质差异、人体与气泡的生理相互作用、在组织中的传播和衰减、组织运动；3) 与微气泡有关的因素，包括气泡的种类及其稳定性、制备和注射及剂量。

? 指出需要更多的研究来建立造影剂增强超声中主观和客观量化的可靠性和可重复性，文章解释 CEUS 成像过程中移动区域会导致伪影，可能会影响解释，尤其是在计算机辅助评估数据时；换能器产生的过大压力可使血流减少，导致血管闭塞，血管的超声信号减弱；换能器的检查方位变化会使 CEUS 在许多组织的组织灌注中显示出很大的异质性。；许多变量，如慢性和急性体力活动、空腹状态、吸烟、饮酒和使用娱乐性药物等，都可能影响 CEUS 测量结果，导致每天的变化很大；此外，超声技术依赖于操作者，这也导致研究者内部的差异。很显然，在复杂的 CEUS 分析中，量化血流灌注的目标困难重重。

1.3.3 深度学习

深度学习是一类机器学习算法，使用多层从原始输入中逐步提取更高级别的功能。例如，在图像处理中，较低的层可以标识边缘，而较高的层可以标识与人有关的概念，例如数字或字母或面部。学习可以是有监督的，半监督的或无监

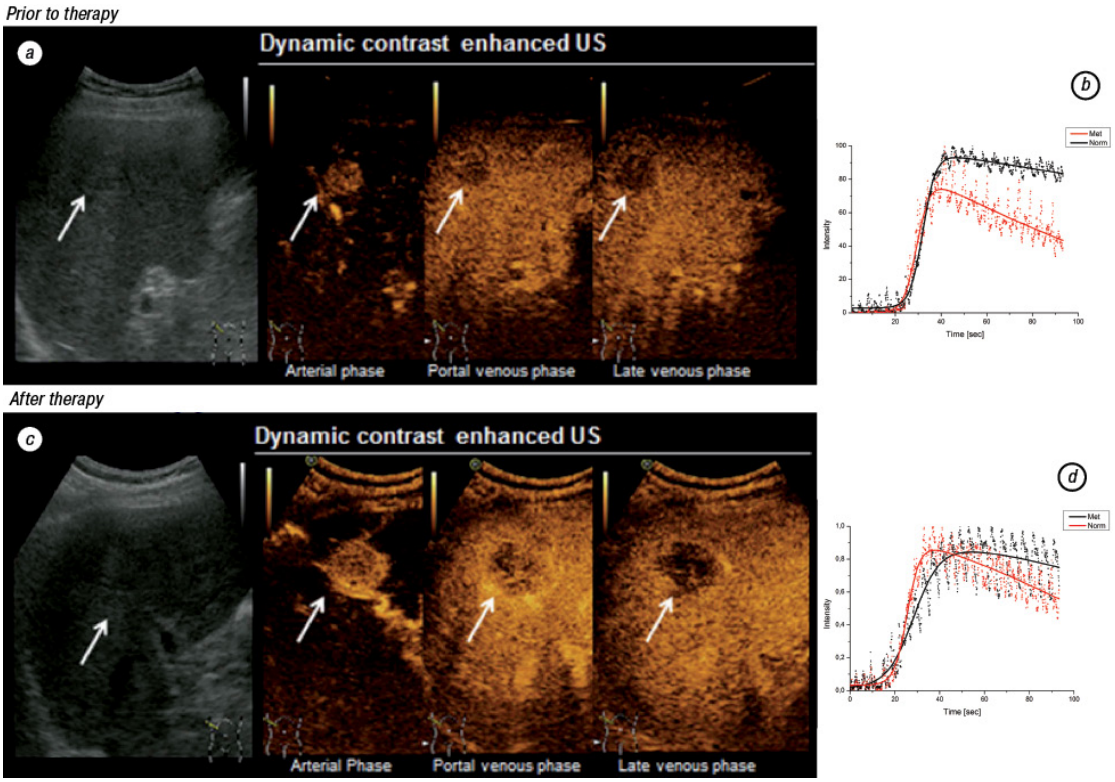


图 1.2 DCE-US 例图。（a)(b) 分别为治疗前神经内分泌转移和正常肝实质的 CEUS 图像和相应的时间-强度曲线,(c)（d）治疗 3 个月后神经内分泌转移和正常肝实质的 CEUS 图像和相应的时间-强度曲线。白色箭头表示转移病灶。曲线的起伏是由呼吸伪影引起的

Figure 1.2 An illustration for DCE-US, from. a, b. CEUS image (a) and corresponding time-intensity curve (b) of a neuroendocrine metastasis and normal liver parenchyma before therapy. c, d. CEUS image (c) and corresponding time-intensity curve (d) of a neuroendocrine metastasis and normal liver parenchyma 3 months after. White arrows indicate metastatic lesion. Undulations in curves are caused by breathing artefacts.

督的。

事实上，现在大热的浅层人工神经网络和深度学习（例如递归网络）已经存在了很多年。1967 年，Alexey Ivakhnenko 和 Lapa 发表了第一个多层感知器的通用工作学习算法 (**?**)。1971 年，论文**?** 描述了一种由八层数据通过分组数据处理方法训练的深度网络。1989 年，Yann LeCun 等人将标准反向传播算法 (**?**) 应用于深度神经网络，目的是识别邮件上的手写邮政编码。虽然该算法有效，但培训需要 3 天。在 1990 年到 2000 年间，人们大多使用特定于任务的手工特征和简单模型的研究流形（例如 Gabor 滤波器和支持向量机）。ANN 网络由于其计算成本高、解释性差和难以训练，沉寂了较长的一段时间，但该领域的先锋者没有停止探索，先后提出卷积网络（convolutional neural network, CNN）(**??**) 和长短时记忆网络（Long short-term memory, LSTM）(**?**)。2006 年,**?** 依次将每一层视为无监督的受限 Boltzmann 机器，然后使用有监督的反向传播对其进行微调，可以有效地预训练多层前馈神经网络。在 2009 年，CNN 开始应用 Nvidia 图形处理单元（Graphics Card, GPU）, 同年，Google Brain 使用 Nvidia GPU 创建了功能强大的 DNN。GPU 将训练算法的速度提高了几个数量级，将运行时间从数周缩短至数天。2012 年 10 月 ImageNet 的胜利标志着一场“深度学习革命”的开始，该革命改变了 AI 行业。2019 年 3 月，Yoshua Bengio，Geoffrey Hinton 和 Yann LeCun 因在他们在神经网络方面的研究作出了重大贡获得了图灵奖。

深度学习模型已经在广泛的诊断任务中实现了医生级的准确性，包括从黑色素瘤识别，糖尿病视网膜病变，心血管风险，以及从眼底和光学相干断层扫描图像的转诊，乳房 X 光检查中的乳房病变检测，以及磁共振成像的脊柱分析。将深度学习应用于医学图像的主要挑战是可用于构建深度模型而又不会遭受过度拟合的训练样本数量有限 (**?**)，常用折中方法有 (**?**)：1）将二维或三维的图像块而不是全尺寸图像作为输入以减少输入维数，从而减少模型参数的数量；2）通过仿射变换（即数据增强）通过人工生成样本来扩展数据集，然后使用增强的数据集从头开始训练其网络; 3）使用在计算机视觉中针对大量自然图像训练的深度模型作为“现成”特征提取器，然后使用目标任务样本训练最终分类器或输出层; 4）使用来自非医学或自然图像的预训练模型初始化模型参数，然后使用与任务相关的样本微调网络参数。更多关于 DL 在医学影像中的应用可以阅读**???**。

由于 CEUS 读片分析时间长，成像效果稳定性差的问题，人工设计的量化

方法定义的特征过于表浅，使用深度学习可能是最好辅助诊断方法。一方面它无需利用影像组学和 DCE-US 分析时的精细勾画病灶，输入只需是包含病灶的 bounding box，可以表达复杂的深层次的特征，读片速度快，另一方面，它正处于爆发阶段，在方法研究上高产，在实际应用中接口友好，无论是深度学习框架，开源项目，学习资料都可以便捷的获取使用。

1.3.4　研究现状

超声造影的定量分析目前可以概括为三个方向：1）基于影像组学特征的研究；2）基于动态超声造影的研究；3）基于深度学习的量化研究。由于造影数据的应用起步时间晚于 CT 和 MRI，在医学指南中的应用范围也较有限，在基于数据的定量分析领域的研究数远远少于 CT 和 MRI。作者在在 2020 年 7 月 3 号，通过 Web of Science 基于所有数据库检索 “CEUS” 和 “Machine Learning” 两个关键词，共计有 22 篇文章；而通过 “MRI” 和 “Machine Learning” 可以得到 6016 个检索结果，其中 2027 条与分割任务有关，2563 条与 DL 有关；通过 “CT” 和 “Machine Learning” 可以得到 3691 条检索结果，其中 1321 条与分割任务有关，2274 与 DL 有关。在 CEUS 的检索中去除一篇综述 (**?**), 两篇会议论文 (**??**) 后来在期刊发表，实质上不同的研究只有 19 篇，文本在表格 1.1中枚举了这些研究。

通过表格中枚举的 19 项研究的实施细则，我们可以看出 CEUS 的机器学习分析起源于 2012 年，主要集中在分割和分类任务中；研究的数据集中在肝脏疾病；单个研究的病例数较少，测试多 “one-out” 实现；分类任务除了基于 3D-CNN 的研究综合时间空间维度特征，其他研究都将问题简化为 2D+t 的特征提取模式：在 2D 空间提取特征构建分类任务再综合多个分类器的结果得出对序列的分类结果，或基于单个像素的 TIC 特征将时间轴序列简化为单个特征向量用来描述 2D 图像每个像素；分类算法集中在 SVM 的应用；一般，前瞻性的测试性能远优于回顾性研究，动物实验好于人体实验；数据的清洗被普遍应用，运动波动大的序列会被排除入组；部分研究设计没有将单张图像作为分类目标，忽略了来自同一个造影数据的图像高度关联的事实，即使取得很好的效果，也没有实际应用价值；深度学习方法使用的网络均为自定义的浅层网络，尚没有针对迁移学习效果的公开研究。

表 1.1 CEUS 机器学习方法汇总表

Table 1.1 Summary table of machine learning methods in CEUS

来源	研究内容	入组病例数	特征定义	分类器	数据清洗	分类器性能
?	预测 HCC 患者对跨动脉化疗栓塞的反应	单中心回顾性 36 例，“one-out” 方法测试	形态学特征，即血管数量、分叉数量、血管与组织的比例、平均血管长度、迂曲度和直径	距离加权分类器 （Distance-weighted Dis-crimination)	去除在平面外运动的图像，手动勾画第一张后使用运动校正，杂乱信号去除，多尺度增强	准确率为 86%，灵敏度和特异度分别为 89% 和 82%
?	预测 HCC 患者对跨动脉化疗栓塞的反应	单中心回顾性，130 例，三折交叉验证	网络学习的特征	基于 3D-CNN 的自定义模型	手动分割一张目标肿瘤，使用运动校正后，截取第 1 分钟连续 CEUS 图像 bounding-box 内容	AUC 为 0.93
?	区分焦点结节增生和炎症性肝细胞腺瘤	单中心回顾性，共 46 例，10 折交叉验证	利用” 光流” 方法，逐像素估计病变中的微泡传输场，定义了四个特征	综合随机森林、k-最近邻居、支持向量机和逻辑回归	去除低质量的数据，手动勾画病灶区域，分析造影剂到达后到充满病灶的过程	精确度 95.7%，AUC 为 0.97，灵敏度 93.4%，特异性 97.6%
?	前列腺癌检测	老鼠实验，共 20 例，分为训练集，验证集和测试集	网络学习的特征	自定义的 3D-CNN 网络，包含三个卷积层	每秒提取一帧，检测窗口大小 23 像素 x23 像素，步长为 5 像素，人工筛选标记每个窗口中的序列是否可用，去除过多的负样本，使正负样本数目一致	90% 以上的特异性和平均精度
?	区分 5 种肝结节	使用公开数据集，共计 95 例, “one-out” 方法测试	图像块，即小窗口中的图像为视觉词袋	支持向量机和软投票	每种病例入组 11 个病例，挑选 10 张图像（5 张动脉期，3 张静脉期，2 张后期），勾画 Bounding-box,	平均精度 64%
?	睾丸病变良恶性分类	单中心回顾性，20 例，10 折交叉验证	构建灌注模式的字典，使用非负矩阵因子提取在像素上的时间强度曲线上提取	支持向量机和软投票	依据超声图像，对造影数据进行刚性形变校正，手动勾画 ROI, 对每个像素的 TIC 使用立方平滑和时间校正	精度 100%

表 1.1 续表。

Table 1.1 Continue Table.

?	分类良性和恶性颈淋巴结	单中心回顾性，共 88 名患者，127 个淋巴结（39 个良性和 88 个恶性），五倍交叉验证	将强度平均图像向量化作为输入特征	使用点式门控博尔茨曼机进行特征选择，支持向量机分类	从 TIC 中检测到具有强度最高帧，以峰值帧为中心选择附近几秒的帧得到强度平均图像	精度、精度、灵敏度、特异性分别为 82.55%、89.58%、84.75%、77.56%
?	良恶性肝病变分类	单中心前瞻性，共 153 例，良性 76 例，恶性 77 例	CEUS 图像和 BUS 图像融合，分别采集每张图像的纹理特征	3 个串联的受限的博尔茨曼机和加权投票	人为选取三张分别在门脉期，动脉期和晚期的典型图像，人工勾画 Bounding-box	平均分类精度为 83.66±2.30%，灵敏度为 83.08±3.70%，特异性为 84.25°
?	肝脏肿瘤分类	单中心回顾性，31 例良性，67 例恶性，“one-out” 验证方法	TIC 曲线特征	使用支持向量机对多种特征组合分类得到最优模型	勾画 ROI 后使用跟踪算法构建 TIC	非线性 SVM 的灵敏度和特异性分别为 94.0% 和 71.0%
?	判别是否是类风湿关节炎	单中心回顾性，115 例，70% 训练，30% 验证	对像素级的 TIC 曲线的稳定性优化，改进了伽马变量模型和提出 singlecom-partmentrecircu-lationmodel 量化灌注动力学	支持向量机	使用运动校正并对像素级 TIC 曲线最大值规范化，去除了 16 例峰值强度和基线值小于固定阈值的病例，在人工勾画后，使用聚类方法保留有价值像素	达到 87% 的平衡精度，灵敏度为 88%，特异性为 86%
?	肝癌病变良恶性分类	单中心回顾性，共 93 例，五折交叉验证	提取单张图像统计特征描述符，以表示肿瘤的纹理信息	不同图像组合的支持向量机模型集成获得	一张 B 模式图像和三个典型 CEUS 图像，手动勾画 Bounding-box	平均精度 89.36%，特异性 89.79%
?	肝病变良恶性分类	单中心回顾性开源数据集，52 例（30 良性和 22 恶性），五倍交叉验证	使用四个灰度描述特征，在图像上以三种大小的窗口和固定步长搜索，将特征串联为向量	使用动态推理模型，寻找最合适的编码位置和编码组合，模型的结构和参数都在训练中自动优化	一张 B 模式图像和三个典型 CEUS 图像，手动勾画 Bounding-box	平均正确率为 85.8%

表 1.1 续表。

Table 1.1 Continue Table.

?	肝病变自动检测和良恶性分类	单中心回顾性开源数据集，共 186 例 HCC，109 例 HEM，59 例 FNH，“one-out” 验证方法	TIC 和形态特征，其中 TIC 使用多项式拟合平滑	通过隐马尔科夫模型确定单张图像病灶轮廓，分类支持向量机	通过医生逐帧读片，每个病例手动选择约 58 张清晰稳定的图像，医生对每张图进行精细勾画，输入数据通过小波滤波自动寻找可行的区域输入	检测重合率为 0.89±0.16， 最高分类精度为 90.3%，
?	脑肿瘤组织边界识别	单中心前瞻性，17 例，“one-out” 验证方法	视频中像素级 TIC 特征	支持向量机	移除单个患者数据中平面移动大或记录过程中系统参数的变化部分	最小平均分类误差 17.4%
?	脾脏病变二分类	基于狗的前瞻性实验，共 36 例	对 TIC 提取 1DHaar-like 特征	Adaboost	使用运动校正	正确率为 91.7%
?	五种肝脏实质病变分类	单中心回顾性，共 191 例，“one-out” 验证	提取 TIC 特征	人工神经网络	未知	正确率为 77.5%
?	图像分割	算法实验，在一个视频上演示了算法自动分割效果	提取 TIC 特征	SelfOrgnizedMap，可以自动检测和跟踪病灶，提取最优的 TICs 特征，实现无监督的图像块聚类	无	无
?	甲状腺病变良恶性分类	单中心前瞻性，20 例，基于 800 张图像（不是针对病例）的三折交叉验证，训练和测试数据不是无关的	对单张图像提取形状纹理频谱共 16 个特征	非参数满惠特尼 U 测试选择特征后使用支持向量机	去除了较大结核的病例和有不清晰的病例，每个病人获取 40 张图像，共计 800 张	100% 正确率

1.4　研究意义

1.4.1　数据清洗

在进行放射分析之前，需要对图像应用预处理步骤，旨在降低图像噪声，提高图像质量，实现可重复和可比的放射分析。

当建立用于机器学习的训练和验证的大型数据集合时，仅仅收集大量的数据是不够的，这些数据需要进过人工标记和清洗，使得数据集的质量满足预期应用的需要。以美国国家癌症研究所的癌症影像信息库（Cancer Imaging Archive, CIA）为例，虽然它包括 3.75 万多个受试者的 3000 多万张放射学图像，拥有丰富的癌症影像信息，并且通过数据描述，将数据库按肿瘤类型进行分类和组织收藏，但是由于这些数据集中分布在个别成像方法且数据分类过于细化，许多类目只有几十例数据。以 CEUS 作为关键词搜索，只能得到三个子数据集，入组病例分别为 8，21，21。很显然，目前 CIA 即使拥有巨大的数据量也不能在医学 AI 领域产生巨大收益，对比仅有 100 万张图像构成的数据集 ImageNet 数据集，数据质量的重要性不言而喻。

虽然国内肿瘤患者较多，但是具体到每家医院，肿瘤患者的数据就相对变少，而影像组学研究需要在众多的医院数据中查找严格符合入组条件的数据来保证一致性，这样做又会使数据量急剧减少 (**?**)。由于 CEUS 的大规模数据研究起步较晚，很多医院接触使用 CEUS 的时间较短，超声设备厂家型号多样标准不一，成像过程依赖医生的操作和病人的配合，而临床指南对记录的时间的要求也只是建议采集前 30s，后续的记录可以为了节约存储空间间歇记录，导致在实际应用中，尤其是回顾性分析，数据质量参差不齐。因此，为了保障数据量，需要让所有在临床标准中满足入组条件的数据尽可能不因为数据质量被排除。

虽然深度学习不像 DCE-US 那样需要对图像配准，但深度学习的性能高度依赖数据量。**?** 表明，深度学习在大数据集中具有较强的抗燥能力，在较大数据集中，增大成比例的噪声数据来扩大总体数据量比提高清洁数据的比重效果好。在很多情况下，基于深度学习方法构建数据集都希望将尽可能多的数据纳入分析中，即使数据的质量较差或与主题的吻合度较低，但这个做法在数据体量不够庞大时，可能会导致数据的错误处理或错误的结论 (**??**)。在机器学习和深度学习时代，计算机科学中定义 “垃圾进，垃圾出 (garbage in, garbage out)” 的格言在数据分析工作中仍然是非常重要的准则。对网络训练影像进行质量评估和清洗一

直以来消耗了数据科学家大量的时间。

在许多 CEUS 的研究中，虽然运动校正的使用较为广泛，但运动校正对于较明显的呼吸运动和手持探头移动效果差，对于这些病例往往只能排除入组，考虑深度学习希望能有尽可能多的质量统一的数据送入网络，深度学习实现自动定位需要人工标记的金标注训练网络，我们开发了一款灵活的用户友好的 CEUS 数据清理的应用程序，它对不标准采集数据友好，可以在对标记结果准确性高的任务上应用，利用自动校正算法（即跟踪算法）实现病灶的自动的连续勾画，允许人能够实时更改自动标注的结果，更能够允许人工在视频中任意时间，任意位置选择需要分析的时段，使得精准的数据标注工作能在人工和自动算法间高效切换，

1.4.2 迁移学习

根据表1.1可知，现有的 CEUS 中的深度学习应用都是通过自己搭建的网络，网络的框架也局限于 3D-CNN，与目前行为分析领域框架网络百花争鸣的现状相比，可以说研究方法十分局限。目前，我们需要一些数据来了解迁移学习在 CEUS 中表现，也渴望知道不同的视频分析网络在 CEUS 中的应用效果是否有差异。正如在自然数据集中，相同方法在不同数据集中的排位有差异，我们也希望知道这些框架模型迁移到 CEUS 的优劣比较。

虽然目前在其他成像方法中已经有迁移学习的应用，但不同的成像模态之间的差异明显，虽然 CEUS 源于 B 型超声，但图像网络能迁移到 US, 并不能说明视频网络可以很好应用在 CEUS, 我们尚不清楚自然视频中的时间轴上的特征编码在 CEUS 中有多少重合。一方面，迁移学习的表现和原问题和目标问题的差异有关，当源域问题和目标域的差异太大，迁移学习的效果的性能将大大降低，简单的微调训练可能效果不及直接在目标域中训练的模型 (**?**)，另一方面，迁移的效果与目标域数据集的大小有关，迁移可用的数据增多时，迁移效果会提升 (**?**)。由于 CEUS 特殊成像方式和成像效果，而且能够训练网络的数据有限，对于迁移学习的性能是否良好，现有的网络框架能否用来拟合病例有限的 CEUS 数据集，我们无法通过理论思考给出，只能通过大量实验来确定。

网络的迁移效果的好坏可以直接影响了从事 CEUS 分析量化的方向：迁移或者使用现成网络但不使用参数或者完全自定义网络结构并完整地训练网络。正如影像组学提出的寻找图像中能够在各个分类问题中发挥作用的图像描述子的

思想简化了医学图像量化的步骤，研究者只需使用通用的图像描述子就可以在自己的数据集建立较好的模型，我们希望深度学习模型中是有通用的特征描述可以直接拿来使用, 即迁移学习中涉及的底层图像特征具有良好的稳定性，迁移到 CEUS 数据中有良好的性能，并且网络在适当改进后，可以进一步提升。这样，CEUS 数据分析的将变得轻松容易，而且人们有更多时间精力探索如何改进网络，比如使用更好的正则化约束，优化方法，激活函数，损失函数，pooling 层，数据扩增方法，甚至加入新的层或者外骨架如升级为对抗网络，或者使用多个网络级联，而不用在基础的网络框架的参数设置和训练上花费大量时间。同理，如果迁移学习确实无法在 CEUS 数据集上表现优良，我们的实验也可以告诉网络设计者更多考虑使用小的自定义网络，不必在迁移学习多样的模型了解和选择花费过多精力。

综合深度学习在医学图像领域巨大的应用前景和 CEUS 在临床中多样化的应用现状，以及传统的影像特征和 DCE-US 时空特征表达能力有限和数值稳定性低，深度方法处理 CEUS 视频的研究非常有限的背景下，开展针对数据清洗和模型选择两个主题的探讨对促进深度学习应用在 CEUS 的病灶定位和辅助诊断非常必要。综上，本研究设计的针对 CEUS 数据集清洗的交互友好的半自动跟踪工具，和对比分析多种深度学习领域视频框架和模型的迁移能力研究，是具有研究价值的。

第 2 章　基于点跟踪的半自动连续勾画软件

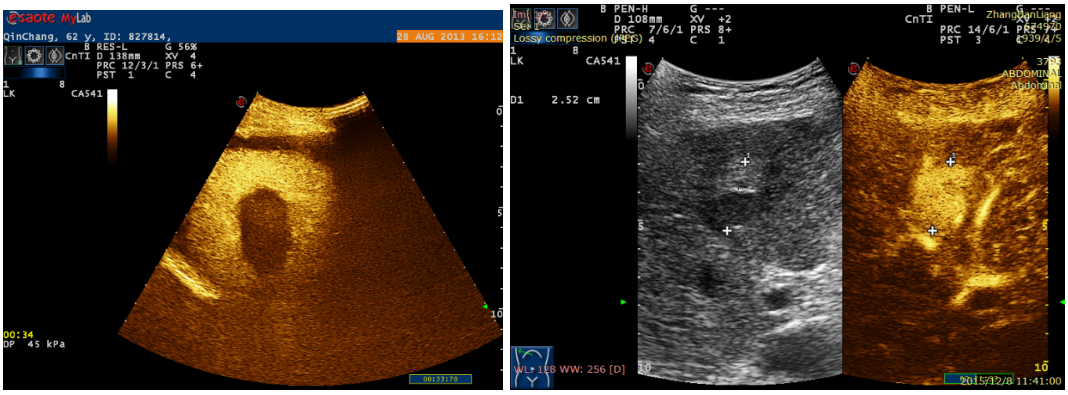
2.1　引言

本章介绍如何使用 MATLAB 2019b 构建一款用户友好的 CEUS 视频连续勾勒的半自动跟踪软件，以实现在快速精准地在病变区域绘制边界框的目标。基于深度学习在处理 CEUS 视频数据方面广阔的前景，为了加快使用 DL 对 CEUS 数据端到端地自动检测、跟踪和分类任务，设计编写了该应用程序。尽管运动的矫正在 DCE-US 中应用历史已久，但都基于自主编写的 Matlab 脚本程序，灵活性很低，不懂 Matlab 编程的医生无法使用，且当运动的矫正效果不好时，也无法及时更改。大多数情况下，CEUS 数据预处理是为每个包含的帧手动绘制草图，或以短序列手动绘制一个帧，默认情况下，病变的位置在序列中不会更改。这些研究似乎没有意识到运动补偿算法的存在。这种现象反映出，当没有实用的工具，只有理论实验时，即使是聪明的研究者也不能利用这些算法来简化或进一步改进他们的工作。

目前，除了本研究公开的软件外，还没有其他的脚本程序允许交互式的勾画操作。该软件的使用非常简单，用户只需在视频中需要跟踪的位置画一个 Bounding box，软件就可以自动跟踪，软件允许用户拖动缩放每一帧的跟踪结果，也可以批量删除一些不好跟踪结果，跟踪长度可以设置。软件除了集成数据导入，工作空间保存，勾画结果导出，视频播放等基本的数据操作接口外，还针对 CEUS 的图像特性，如双幅显示，加入了镜像功能等自动分析方法。本章分三节分别介绍运动校正和不同跟踪算法的性能测试，软件搭建，软件使用，在我看来，项目的难度在与设计和编写交互式软件，由于没有参考，需要从零开始自行构想一个好用的软件界面和功能，并让软件能够响应多种用户操作，保证数据传递无误和时序不发生混乱。

图 2.1显示了两种 CEUS 录制界面，展示了原始的未经处理的 CEUS 图像, 可以看到，CEUS 图像的录制区域具有特定的形态，通过边缘提取和形态分析可以自动识别出图像的录制模式。与此同时，本研究提出了镜像概念，即将一个录制图窗的勾画结构自动映射到另一个图窗的对应区域，帮助医生在双幅模式时利用超声清楚的边界信息来勾勒 CEUS 中未完全增强的病灶和全增强时边界无法

确定的病灶。



(a) (b)

图 2.1 两种 CEUS 界面展示。(a) 单幅显示，图像中只有 CEUS 一个录制窗口，(b) 双幅显示，图像的一侧为 US 视频窗口，另一侧为 CEUS 视频窗口，两个窗口显示同一成像平面的不同信号分布。

Figure 2.1 Two CEUS interface displays. (a) single display with only one CEUS recording window in the image, (b) dual display with US video on one side of the image window, with a CEUS video window on the other side, and two windows showing different signal distributions from the same imaging plane.

2.2 跟踪算法

2.2.1 运动校正

在 CEUS 中，有许多研究集中在通过整个序列的自动运动估计方法来补偿观察到的肝脏实质性病变（focal liver lesions, FLLs）运动，这实质上是一个连续帧上特定区域的跟踪任务。运动估计方法分为直接方法和间接方法，直接方法测量图像强度变化来估计运动矢量，对噪声和光照变化敏感 (**?**)，间接方法通过一组稀疏的图像特征进行连续帧的匹配。具体来说，基于点的跟踪技术（point-based registration techniques, PBRTs）非常适用于在 CEUS(**?**) 或 US 序列 (**?**) 中跟踪 FLLs。

PBRTs 主要由两部分组成：(1) 图像特征提取，即提取突出区域的特征 (2) 匹配和利用结果估计几何变换。经典的点描述器有 Harris 角检测器 (**?**)、利用特征值检测角点的 Kanade-Lucas-Tomasi(**??**)、HOG(**?**)、SIFT(**?**)、FAST(**?**)、SURF(**?**)、BRISK(**?**)、ORB(**?**)、CARD(**?**)。基于**?**，统计描述符的匹配策略有三种选择：基于阈值的方法、基于最近邻的方法和最近邻距离比。一般的距离指标有欧氏距

离、点积和 Hamming 距离。有时，匹配过程中可以通过估计几何变换来修正，如 M-estimator SAmple Consensus(**?**)，它利用前向和后向误差反复寻找最佳匹配和几何变换矩阵。

软件的跟踪算法的流程为：手动绘制一个边界框进行初始化；提取当前帧和下一帧的点特征；将点匹配后计算变换矩阵；通过 M-estimator SAmple Consensus 检验跟踪结果，输出更稳健的跟踪点，再重新估计变换矩阵；应用变换矩阵更新边界框的位置。该算法将进入下一帧，并进行相同的程序，直到匹配点数目少于三或达到跟踪长度。在跟踪过程中，由于不断删未匹配的点，可用的点会不断减少，因而每隔 10 张图像通过刷新点集来重新初始化需要跟踪的点，重新初始化会降低运行速度。由于 CARD 逐帧提取图像各个子区域的特征，特征向量个数由图像自身决定，不需要重新初始化。

MATLAB 2019b 图像处理工具箱提供了多种关键点检测算法，如特征值、HOG、FAST、BRISK、MSER、ORB。根据**?** 的研究，基于 HOG 的跟踪算法比较耗时，所以我们在软件中排除了这个功能。目前，在 MATLAB 中导入 SIFT 特征检测器仍然是一个需要其他特定库的重任，所以没有使用 SIFT 特征。(**?**) 也报道，虽然 SIFT 有比较高的运行速度和跟踪精度，但没有 CARD 好。由于 CARD 有官方开源，且库的安装和调用都很方便，所以我们将描述长度为 128 维和 256 维的 CARD 纳入跟踪方法中。

2.2.2　跟踪性能测试

由于 PBRTs 是一个比较成熟的系统，我们的测试主要针对该软件中不同跟踪算法的量化跟踪精度和速度, 算法的超参数都使用函数提供的默认值（通过预实验发现默认值的效果最好）。测试数据为专业医生选取的两例典型的 HCC 动脉期 CEUS 片断。其中，病例一的病灶边界不清晰，体积相对较小，呼吸运动明显病；例二的病变边界相对清晰，体积相对较大，扫描过程相对稳定。帧率均为 25 帧/秒。医师使用跟踪软件生成病灶区域，再逐帧精细调整，作为本次实验的 Ground-truth。

实验希望量化不同算法在不同跟踪长度设置下的精度和速度。跟踪长度范围为 30～270 帧，步长为 30。测试方法为，在特定的跟踪算法和跟踪长度下，用视频中的每一帧和相应人工标定的边界框来初始化多次跟踪过程。每次跟踪后，计算该跟踪序列中与 Ground-truth 的重叠比（交集面积除以并集面积），并通过

跟踪时间除以跟踪长度（frame per second，FPS）计算运行速度，再对所有跟踪结果计算平均值。我们对 CEUS 和 US 跟踪模式下的性能进行测试，分析同一算法在两种成像模式是否存在性能差异，以及跟踪精度随时间的衰减情况。测试结果见表 2.1、2.2、2.3、2.4，其中蓝色字体表示本列性能最好，红色字体表示最差。

综合表 2.1、2.2、2.3、2.4，可以发现，随着跟踪长度的增加，所有基于关键点特征的算法都呈现出精度急剧下降的现象，而 CARD 的算法性能稳定且表现较好，但速度相对较低。在研究案例和跟踪帧长度固定的情况下，在 CEUS 视频和 US 视频中，US 的跟踪精度略微好于 CEUS。对比案例 1 和案例 2，可以发现案例 2 的跟踪效果更好，但时间也更长，从一定程度上表明边界清晰的大病灶因为能检测到更多特征点，所以表现更好。综合来看，我们建议使用 CARD 128 执行跟踪任务。

表 2.1 案例 1 的 CEUS 视频跟踪速度和精度

Table 2.1 Case 1's CEUS video tracking speed and accuracy

跟踪方法		FPS	跟踪长度								
			30	60	90	120	150	180	210	240	270
Key	KTL	13.63	0.93	0.803	0.77	0.695	0.622	0.523	0.562	0.473	0.356
Points	FAST	14.15	0.894	0.804	0.771	0.695	0.625	0.524	0.562	0.475	0.364
	BRISK	11.24	0.893	0.803	0.77	0.693	0.622	0.523	0.564	0.511	0.359
	SURF	15.28	0.892	0.802	0.767	0.693	0.621	0.521	0.563	0.471	0.356
	MSER	12.11	0.892	0.801	0.768	0.692	0.621	0.522	0.559	0.471	0.356
	ORB	10.55	0.927	0.799	0.764	0.69	0.618	0.518	0.558	0.465	0.347
CARD 256		4.35	0.852	0.906	0.895	0.889	0.894	0.877	0.903	0.843	0.865
CARD 128		5.25	0.852	0.906	0.895	0.889	0.894	0.877	0.903	0.843	0.865

表 2.2 案例 1 的 US 视频跟踪速度和精度

Table 2.2 Case 1's US video tracking speed and accuracy

跟踪方法		FPS	跟踪长度								
			30	60	90	120	150	180	210	240	270
Key	KTL	10.51	0.929	0.801	0.769	0.693	0.623	0.523	0.569	0.473	0.36
Points	FAST	11.28	0.895	0.807	0.773	0.701	0.628	0.527	0.575	0.519	0.371
	BRISK	9.06	0.894	0.805	0.772	0.701	0.627	0.525	0.571	0.516	0.367
	SURF	12.31	0.891	0.8	0.766	0.692	0.623	0.522	0.566	0.47	0.354
	MSER	10.88	0.894	0.805	0.772	0.7	0.627	0.525	0.57	0.517	0.36
	ORB	9.09	0.888	0.797	0.764	0.69	0.618	0.519	0.557	0.465	0.347
CARD 256		3.81	0.919	0.923	0.911	0.892	0.892	0.894	0.873	0.901	0.851
CARD 128		4.58	0.919	0.923	0.911	0.892	0.892	0.894	0.873	0.901	0.851

表 2.3 案例 2 的 CEUS 视频跟踪速度和精度

Table 2.3 Case 1's CEUS video tracking speed and accuracy

跟踪方法		FPS	跟踪长度								
			30	60	90	120	150	180	210	240	270
Key	KTL	4.44	0.936	0.831	0.753	0.718	0.64	0.518	0.621	0.559	0.498
Points	FAST	7.87	0.881	0.829	0.752	0.718	0.641	0.519	0.624	0.558	0.5
	BRISK	6.9	0.883	0.831	0.754	0.718	0.641	0.518	0.623	0.559	0.499
	SURF	7.55	0.885	0.829	0.752	0.717	0.639	0.517	0.62	0.559	0.498
	MSER	7.75	0.883	0.828	0.75	0.714	0.637	0.516	0.619	0.556	0.497
	ORB	6.36	0.884	0.829	0.753	0.716	0.638	0.516	0.618	0.558	0.497
CARD 256		2.72	0.962	0.958	0.95	0.936	0.96	0.907	0.955	0.915	0.967
CARD 128		3.13	0.962	0.958	0.95	0.936	0.96	0.906	0.955	0.915	0.968

表 2.4 案例 2 的 US 视频跟踪速度和精度

Table 2.4 Case 1's US video tracking speed and accuracy

跟踪方法		FPS	跟踪长度								
			30	60	90	120	150	180	210	240	270
Points	FAST	8	0.878	0.825	0.749	0.716	0.64	0.517	0.623	0.556	0.5
	BRISK	6.88	0.88	0.828	0.752	0.718	0.641	0.519	0.624	0.558	0.501
	SURF	7.93	0.886	0.831	0.754	0.719	0.642	0.519	0.62	0.561	0.499
	MSER	8.01	0.884	0.83	0.753	0.718	0.64	0.519	0.622	0.561	0.499
	ORB	6.68	0.887	0.831	0.755	0.718	0.641	0.518	0.618	0.56	0.497
CARD 256		2.99	0.962	0.958	0.95	0.936	0.96	0.906	0.955	0.915	0.967
CARD 128		3.49	0.958	0.963	0.942	0.938	0.931	0.868	0.946	0.913	0.968

2.3 软件构建

由于软件的构建是一个细节比较丰富的环节，涉及很多 Matlab 使用的专业知识，而功能的介绍在下一节展开，所以本节只说明了一些重要的问题，更多细节请阅读开源代码。

- 软件的算法部分集中在图像处理和计算机视觉两个领域，主要使用 Matlab 的 Image Processing Toolbox 和 Computer Version Toolbox，视频文件的读取使用 VideoReader 函数，只需传入视频名称即可加载，可以更改 curtime 成员变量再调用成员函数 readFrame 访问视频的任意帧。由于 1÷ 帧频往往除不尽，而随意设置的时间会使算法自动插值生成对应的图像，该图像不是真实存在的，所以在加载视频后，遍历视频的 frame 并保存这些时间位置，建立字典，实际应用访问图像在视频中的次序，然后通过字典转化为时间。

- 软件搭建基于 MATLAB 图形用户界面创建应用图形用户界面（Graphical User Interface, GUIs），GUI 提供了软件应用程序的点击式控制，无需其他人学习 Matlab 语言或命令即可运行应用程序。为了对软件设计和开发执行灵活的控制，程序的主体基于编程方式创建，使用 MATLAB 函数来定义应用的布局和响应，通过函数确定每个组件在图窗中的位置。

- 作界面信息保存，交互界面的所有变量统一存放在一个字典变量中，但由于这些变量部分是系统的默认设置，部分可以通过其他变量经过标准流程计算得到，因而，为了节省存储空间，只导出字典的部分 Key 到 ".mat" 格式的文件中，包括视频文件路径，成像图窗在图像中的位置，图窗模式，造影图窗位置，当前帧，和所有 Bounding-Box 信息（跟踪序列的编号，所在帧在视频中的位置，跟踪框在图像中的位置)，Mat 文件一般只有几 KB 大小。

- 响应事件的能力基于调用交互控件的回调函数（Callback Function)，软件中交互式菜单基于 uimenu，文件载入使用 uigetfile，文件保存过使用 uisave，上下文菜单基于 uicontextmenu，使用交互控件有数字滑块、复选框、编辑字段和按钮，Bounding-Box 基于 imRect 实现，整个界面是一个 uifigure，视频播放图窗是一个 uiaxes。

- 视频播放。视频播放通过计时器 timer 实现，当计时器每隔一个间隔，就执行一次界面刷新（显示下一帧图像、Bounding-Box 等)，通过调节 time 的刷新周期可以轻易调节播放速度，比如 ×2 或者 ×0.5，timer 可以轻松的开启和暂停。

视频播放的需要注意在于刷新功能在响应时，计时器会继续工作，如果刷新操作不能再下一次 timer 启用时结束，就会导致大量任务累积未处理，容易发生系统错误，因此 timer 的最小刷新时间需要比一次刷新时间长。

- 处理关联的事件。当用户对对一个控件做出更改时，一些更改的变量会在其他控件中复用，这是这些控件也需要做出响应。

- 该软件可以自动定位录音和判断显示模式（双或单）。大多数 CEUS 机器都支持单屏成像模式和双屏成像模型，同时具有 B 模式成像功能。基于单屏 CEUS 一般为扇形的观察，基于传统的数字图像开发了自动分析算法。我们汇总了连续帧的差异图像，然后使用阈值、Sobel 边缘检测、霍夫线检测来获取记录屏幕的轮廓和形状。当轮廓为单个扇形时，成像模式将被视为单幅，否则屏幕将在水平中均匀分割。当 cine-loop 处于双显示模式时，软件将显示双边界框，用户需要通过手动选择软件哪一侧是 CEUS，软件默认为左侧。

- 默认打开文件位置自动更改。由于数据集往往在同一个大的文件夹下，为了减少用户每次加载保存文件时，文件夹路径的更改操作，软件自动保存最后一次用户加载数据和保存过数据操作的文件夹的上一层位置并且选中该文件夹，保证下次操作可以直接进入数据库目录，并且可以轻易知道上一次处理的数据是哪一个，如果操作者按顺序处理数据，那么他可以很快知道选定文件夹的下一个文件夹就是他需要处理的文件。

- 限制跟踪区域。由于超声视频的录制窗口并非铺满整个图像，跟踪结果需要限定在录制区域内。

- 灰度化和图像精度改变。由于 CEUS 和 US 都是伪彩色图像，为了降低颜色带来的不确定性和高内存占用，读入后将彩色图像转为灰度图像，并将 0-255 的像素值转为 0-1 之间的双精度数据。

- 双线程加速。由于跟踪过程以人工勾画所在位置为中心向前后两个方向跟踪，互相独立，所以使用 parfeval 执行双线程操作，再将每个线程的结果使用 fetchNext 读出，整合。需要注意，当勾画图像为第一张和最后一张时，要关闭双线程。如果电脑有显卡，也可以使用 GPU 加速。

- 处理中断。由于用户很可能密集操作，一个响应在没有结束的时候发起另一个指令，需要设置优先级，决定是按时间顺序响应，还是停止目前的操作执行新的操作，或者直接恢复到本操作之前的状态再执行下一个操作。本软件除了视

频播放允许新的中断插入，其他都按时间顺序响应。

- 处理错误。用户操作难免会发生不符合软件使用规范的操作，比如加载文件格式不支持，此时会发生错误，为了保证用户使用体验，对于错误操作自动恢复成执行前的状态，同时告知错误原因。

- 处理不稳定的造影区段。由于造影数据的录制是人为控制的，在实际操作中，难免会比较灵活或者不规范，比如扫描肝脏查找病灶，或者暂停录制但同时记录屏幕内容。由于一个造影视频的时间可能较长，而且跟踪方法也不能再剧烈运动时工作良好，所以加入光流分析接口 opticalFlowLKDoG，分析运动是否发生且是否剧烈，设置固定的高低阈值，通过对图像光流幅值求平均给出判断。由于光流的计算基于局部一小段视频得到，所以可以使用多线程将视频分为多段处理，并使用 GPU 加快计算。但光流计算对电脑配置要求较高，建议根据实际情况决定是否使用。

2.4 软件使用

本节通过图 2.2和表 2.5的方式展示软件的界面和介绍各个组件的功能。

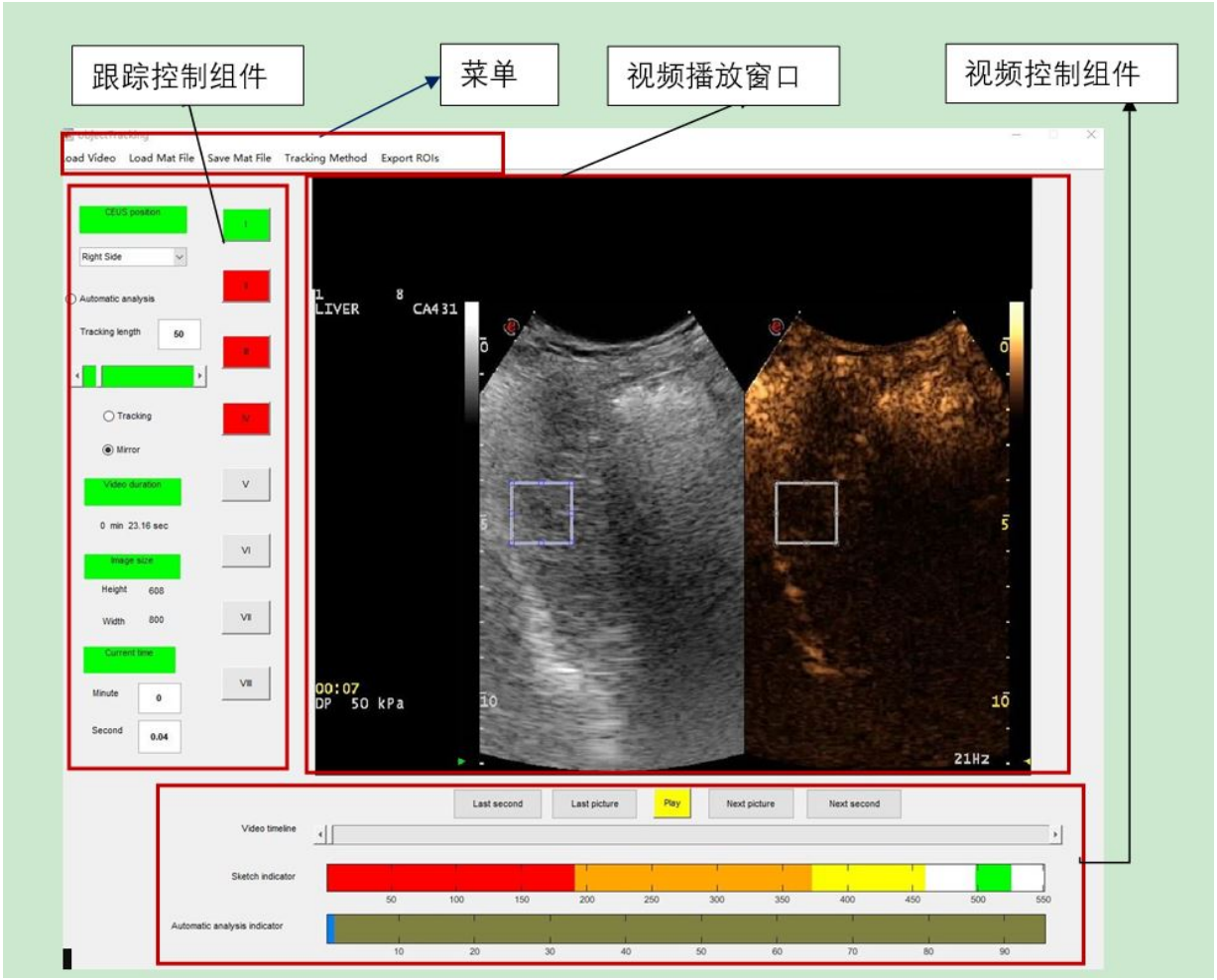


图 2.2 软件运行效果图。软件总体分为四个区域，菜单，跟踪控制组件，视频播放窗口和视频控制组件，图像显示了一例勾画完的视频在软件中的显示效果，图下方区域的彩色标记代表着不同跟踪任务在视频中的位置。软件界面简洁紧凑

Figure 2.2 A rendering of how the software works. The software can be divided into four areas: the menu, the tracking control component, the video playback window and the video control component. The image shows an example of how the finished video is displayed in soft armor, and the colored markers in the lower area of the figure represent the different tracking tasks in the Location in video. The interface is simple and compact.

表 2.5 软件操作界面模块组成和功能表

Table 2.5 Software operator interface module composition and function

所在地区	组件	功能
菜单	加载视频	通过调用 MATLAB 内置功能 VideoReader 将视频导入软件。支持格式包括 MPEG-4，H.264 编码视频（.mp4,.m4v,.mov）或其他微软媒体播放器支持的格式。DICOM 格式的文件需要使用其他软件（例如 RADIAUT DICOM 查看器）转化为可读取的视频格式。
	加载 Mat 文件	在软件保存的.mat 文件中导入数据。它可以在保存文件时恢复软件的状态。
	保存 Mat 文件	将数据保存在当前软件状态的.mat 文件中，其中包括所有变量和变量的值。在软件使用期间，不会更改原始视频。
	ROI 边界框导出	导出视频中所有勾画的边界框。
	跟踪方法	默认方法是 CARD 128。
轨道控制区域	造影所在位置选择栏	选择栏有两个值（左和右），当 CEUS 录制为双幅模式，该值有效。左值表示左侧的为 CEUS 窗口，而右侧值表示右侧为 CEUS 窗口。在处理新视频之前，应设置该值。
	自动分析	当您单击该选项时，软件将对导入视频进行自动分析。它将分析记录窗口位置和录制模式。
	跟踪长度	跟踪长度可以通过手动输入数字或拖动下面的进度条来设置一次跟踪的帧数。
	跟踪	单击该选项时，当前绘图边界框将以当前帧的 Bounding Box 为初始值，执行一次跟踪。
	镜像	该选项是自动分析后的默认设置。如果分析错误，则可以设置该选项，点击后，双幅模式下单侧的 Bounding Box 会复制到另一侧图像窗口中。
	视频基本信息文本栏	这些文本表示加载视频的时间以及帧宽度和高度，以便用户了解正在处理的视频的空间分辨率和总时长。
	当前时间	文本指示整个视频中当前帧的时间。文本框可以手动输入以快速准确地跳到特定时刻。
	跟踪控制按钮组	这 8 个按钮分别表示 8 个不同的跟踪序列。每个按钮对应的跟踪可以从录制屏幕中的任何帧和任何位置开始。它们使用灰色、绿色、红色表示没有跟踪序列，已有跟踪序列和正在处理。当按钮为灰色时，按下按钮可以绘制当前帧上的边界框。当按钮为绿色时，按下按钮可以跳转到相应的跟踪区域

表 2.5 续表。

Table 2.5 Continue Table.

所在地区	组件	功能
视频播放窗口	播放上下文菜单	单击当前帧中的图像时，将显示上下文菜单。选项包括” 下一张图片”、” 前一张图片”、” 下一张图片”、” 最后一张图片”。它们设置为检查或按帧微调边界框。
	删除上下文菜单	单击当前框架的边界框中的区域时，将显示上下文菜单。右键单击将显示” 删除当前帧的 Bounding Box”、” 删除当前帧之前的所有 Bounding Box”、” 删除当前帧之后的所有 Bounding Box” 和” 删除整个序列 Bounding Box”
	镜像上下文菜单	单击当前框架的边界框中的区域时，将显示上下文菜单。右键单击将显示” 镜像当前帧的 Bounding Box”、” 镜像当前帧之前的所有 Bounding Box”、” 镜像当前帧之后的所有 Bounding Box” 和” 镜像整个序列 Bounding Box”。. 这些操作仅影响当前跟踪序列。使用镜像意味着边界框的位置将切换到另一个图窗。
视频控制区域	视频播放控制按键组	控制视频播放的按钮，从左到右分别为” 上一秒”、” 上一帧”、” 播放/停止”、” 下一帧”、” 下一秒”，用于精细检查跟踪结果是否需要调整。
	视频播放进度条	允许手动拖动，进度条的的当前值图像在整个视频中的时间。
	勾画指示进度条	用不同颜色表示视频进度条上每个位置对应的图像是否有 Bounding Box。默认的颜色为白色，其他不同的颜色，分别代表不同的跟踪进度。
	自动分析进度条	以不同颜色指示视频进度条相应位置的图像模式。绿色表示帧处于单模式，棕色表示双幅显示模式，灰色表示未使用自动分析功能或此时间段的录制图像没有改变。

第 3 章　增强超声视频数据迁移学习探讨

3.1　引言

本章探讨自然数据集的模型在 CEUS 中的迁移性能，数据来自中山大学第一附属医院超声科，由医生回顾性收集 1999-2019 年中山大学附属第一医院胆管细胞癌（intrahepatic cholangiocellular carcinoma, ICC）患者。入组标准：1、病理确诊；2、病人进行超声造影检查；3、无肝外转移、血管侵犯；4、超声造影成像前未进行过系统或局部治疗。共 190 例胆管细胞癌患者入组，为胆管细胞癌患者按 1：1 比例匹配肝细胞癌 (（hepatocellular carcinoma，HCC）病例，所有胆管细胞癌和肝细胞癌病例组成本课题的研究病例，共计 380 人，研究采用三折的交叉验证，按肿瘤类型随机分组。入组数据为 CEUS 前 120s 的视频，使用 RandiAnt 转出为 “.bmp” 格式的图像序列，输入网络时只将需要的图片读入再合并为一个 clip，从而降低数据读取时的内存和加快读取速度，所有病例由影像医生挑选和人工勾画了以下 8 张图片的肿瘤靶区：1、病变最大切面动脉增强早期图像；2、病变最大切面动脉增强中期图像；3、病变最大切面动脉增强峰值图像；4、增强峰值后 5s 图像；5、增强峰值后 10s 图像；6、增强峰值后 20s 图像；7、增强峰值后 40s 图像；8、CEUS120s 的图像。通过对一个视频中的 8 个 Bounding-box 取最小的能够包下所有区域的 Bounding-box，并将其上下左右向外扩展 20 个像素获取单个视频统一的截取区域，从而实现病灶区域和背景分离。

肝细胞癌占据成人原发性肝癌的 90%(**?**)；胆管细胞癌是肝内第二大高发癌症，恶性程度明显高于 HCC，数据也更为稀有，病人的长期存活只能通过手术切除来实现，在临床上必须明确与肝细胞癌区分。2017 年，大规模多中心的研究**?** 表明，超声造影在 HCC 的诊断敏感度较低，对于病灶大小在 10-20mm 的病灶仅为 39.6%（CT 63.5%，MRI 70.6%），对于 20-30mm 的病灶敏感度为 52.9%（CT 71.6%, MRI 72.3%），存在和胆管细胞癌混淆的风险。2018 年，多中心研究**?** 对于 1006 个结节细化分析后表明，对于 LI-RADS LR5 期的病人，超声造影的阳性预测值可以达到 99%，虽然在这 1006 个病人中只有 519 个病人可以归类为 LR5 期病人。**?** 指出，相比 CT 和 MRI，CEUS 的消退表征对于小于 20mm 的病灶不易检测，因而敏感度较低。本次研究入住的病例没有将病灶尺寸作为入组的筛选条

件，是希望模型的通用性更好，本研究为二分类问题，HCC 标签值为 0，ICC 标签值为 1。

由于大规模数据集上训练视频网络是不能靠一台服务器和几张显卡实现的任务，本次研究使用公开发表的模型和参数，检索依据来自 Paper With Code 网站上计算机视觉-> 视频-> 视频分类任务的排名表，部分公开模型虽然有着较好的性能，但因为作者认为这些网络的计算成本太高，如 facebook 开源的视频网络模型库VMZ中的 ir-CSN-152 和 ir-CSN-152 模型，输入 32×224^2, 152 层深度，没有纳入实验中。

本研究对比了 2D CNN，2D CNN+BERT，C3D，I3D，SlowFast，R(2+1)D，TSM，GSM 后，得到结论如下：

1. GSM 模型的性能较为优越；

2. 2D CNN 可以比部分时空网络效果好；

3. BERT 结构相对 2D CNN，性能更好，且 BERT resnet50 和 BERT bninception 性能较突出；

4. Slow 通道的迁移效果最差，不适合应用在 CEUS 中；

5. TSM 网络的迁移性能较差，这可能因为 Shift 模块的模式在 CEUS 上无法迁移利用；

6. I3D 的表现较为平庸；

7. C3D 对在更多数据时，性能提升明显，实现了 0.9 的正确率和 0.867 的 ICC 敏感性，在所有实验中变现最好的；

8. R(2+1)D resnet34 相比 resnet34+BERT，性能提升更明显, 基于 resnet 50 比 resnet 34 性能好，我们推断如果使用更深的 resnet 网络，性能会提升更高，因而是最有潜力的模型；

9. 提高网络的视频采样的图像数目能够普遍地提升分类性能；

10. 虽然原模型的训练基于密集采样，但均匀采样的迁移效果更好，因为数据包含了完整的造影剂增强和消退过程；

11. 数据清洗虽然去除了 21% 的图像，但平衡了 HCC 和 ICC 的数据量，直接使用原始数据时，有 3/4 的模型对 ICC 的敏感性变差，但在 C3D 上性能和敏感度都得到了提升，所以建议在具体使用对处理过和未处理数据集都测试；

12. non-local 结构虽然可以提升原问题的分类能力，但全局特征的关联不适

合迁移到 CEUS 中；

13. 在训练过程中，2D CNN 的时间和计算消耗最少，训练是，1000 次迭代在 5 分钟左右；2D+BERT 稍微慢一点，但比 GSM 和 TSM 快；R(2+1)D 的计算量较大；运行较慢；C3D，SlowFast, Slow 基于 3D 卷积，计算量最大，运行时间最长，1000 次迭代需要 20 分钟左右。

14. 对不同网络对所有病例分类结果可视化可以发现，个别病例在绝大部分网络中都无法正确识别，这种局限可能来自数据集，也可能是迁移学习无法获得区分力强的特征。

本研究是第一个系统研究视频模型在医学三维数据集中迁移能力的报告。

3.2　研究背景

3.2.1　迁移学习

迁移学习（Transfer Learning, TL）是机器学习中的一个研究问题，着重于存储在解决一个问题时获得的知识并将其应用于另一个但相关的问题。迁移学习可以分为三类：1）正迁移，即一种情况中的学习在另一种情况下促进学习时，例如，拉小提琴技巧有助于学习弹钢琴。数学知识有助于以更好的方式学习物理；2）负迁移，指学习一项任务会使另一项任务的学习更加困难时，例如，讲泰卢固语妨碍了马拉雅拉姆语的学习；3）零迁移，即学习一项活动既不能促进也不妨碍学习另一项任务。在医学影像中的深度学习研究领域中, 有很多迁移学习的成功应用，比如**?** 使用深度卷积神经网络从乳房 X 线照片的迁移学习，开发用于数字乳房断层合成体积质量的计算机辅助检测系统，**?** 使用迁移学习对超声乳腺癌图像进行了表征，获得了强有力的结果，**?** 利用迁移技术从 X 射线图像中 Covid-19 进行自动检测，这些案例表明，网络的特征具有较好的泛化能力。从实践的角度来看，为学习新任务而重用或迁移先前学习的任务中的信息可能会显著提高强化学习代理的样本效率**?**。

最广为人知的迁移学习方法是在大数据集上从零开始训练网络，然后用需要分析的小数据集训用较小的学习率和较少的迭代次数微调（fine-tune), 该方法可以看做是把网络当成一个底层固定特征提取器，通过简单的调整高层特征的权重，获得小数据集的深层特征。与此同时，改进迁移学习效果的研究也在快速发展，这些研究往往都基于大量实验，目前还没有重大的理论进展。在我看来，

这些研究可以分为两类，1）改进训练策略，2）改进模型的泛化能力。

改进训练策略针对超参数的选择、数据集的优化和选择最优的迁移模块（网络每个子模块是固定，fine-tune 或者随机初始化）。结论具有借鉴意义的研究有：**?** 指出正确选择超参数和源数据集，可以提升 TL 的准确性。**?** 神经网络各层的学习率是重要的超参数，在 ImageNet22k 迁移到 Oxford Flowers 的任务中，选择适当选择学习率，可以将精度提高 127%；**?** 指出更多的预训练数据并不一定有帮助，通过删除不相关的数据，改进预训练数据集不同数据的比重（给弱相关的数据低权重），在细粒度分类数据集上获得了卓越的性能；**?** 对多个医学挑战中的数据集进行汇总，建立具有不同成像模态、目标器官和病理的 3DSeg-8 数据集，并通过建立名为 Med3D 的异构三维网络来共同训练多领域的 3DSeg-8，建立一系列预训练模型。相比直接从 Kinetics 数据集上的预训练模型，3DSeg-8z 上预训练的数据集在肺部分割、肝脏分割效果更好；**?** 通过比较 12 个图像分类数据集上 16 个分类网络的性能得出，当网络用作固定功能提取器或微调时，ImageNet 精度与迁移精度之间有很强的相关性，此外，在两个小型的细粒度图像分类数据集上，ImageNet 中学到的要素无法很好地迁移。

改进模型的泛化能力的研究是研究什么样的网络结构可以更好地完成迁移任务，这些研究工作集中在迁移域学习（Domain Transfer, DL)，代表性的研究有：**?** 设计了一个在结构上有多个组合的神经网络，能自动为输入在大的语义空间中选择最匹配的子特征空间。**?** 允许迁移学习在训练时和测试时的数据来自相似但不同的分布，通过自适应方法，寻找在源域上对主要学习任务具有辨别性且在不同迁移域中有较好表现的特征，以提高泛化能力。**?** 提出了一种新颖的转移学称为” 学习转移”(Learning to Transfer) 的学习框架，通过元学习，使得网络在面对不同迁移问题时能自动确定那些特征需要保留，哪些需要训练。**?** 指出虽然迁移学习通常可以通过更好的精度和更快的收敛来提升性能，但从不合适的网络中转移权重会伤害训练过程，并可能导致更低的精度，因为提出了一种正则化方法改进网络优化的方向使得网络部分参数得到较好的保留。

当然，在医学图像分析中也明确存在负迁移的现象，**?** 指出 ImageNet 预训练模型在两个大规模医学影像任务上，相比简单、轻量级的自定义模型，几乎没有性能上的提升。但研究也指出，迁移效果不好不是因为 ImageNet 上的特征不能用于医学数据上，而是由于标准模型的过度参数化，以至于在迁移问题上过度

的拟合。本次研究也发现，网络的参数越大，越容易出现过度拟合，导致泛化效果越差。

3.2.2　数据集

本研究预训练模型使用的数据集有：

- ImageNet(?),2010 年发布的 2D 视觉对象识别软件研究的大型可视化数据库，包含 1000 个类别，每个类别包含 1000 张左右的高分辨率图像，数据集基于生活图片，比如“气球”、“草莓”，对推动深度学习的发展贡献巨大。
- Kinetics(?),2017 年发布，包含 400 个人类动作类，每个动作至少有 400 个视频片段，共计 306k 个片段，每个片段持续 10 秒左右，取自不同的 YouTube 视频。这些动作以人为中心，涵盖了广泛的类别，包括人与物的互动，如演奏乐器，以及人与人的互动，如握手。Kinetics 是视频分类任务领域首个大规模的高质量数据集，是视频识别方法评估的 base-line，也是很多数据集提供预训练模型。
- IG-Kinetics(?),2019 年发布，目前规模最大的视频数据集, 数据未公开，包含 359 个类别，65M 个片段，数据来自社交媒体，没有经过人工筛选，标签较为嘈杂，通过弱监督学习初始化网络再 fine-tune 至 Kinetics 上实现了目前 Kinetics 最高正确率。
- Something-v1(?),2017 年发布的中等规模的视频片段数据集，共计 170 个类别，110k 个片段，致力于让机器学习模型对物理世界中发生的基本动作进行精细的理解，类别如 “戳破一堆罐子”,“将电缆插入充电器”, 数据以每秒 12 帧的速度从原始视频中提取，并转化为 JPG 图像保存。
- Sport-1M(?),2014 年发布，来自 YouTobe 的体育栏目视频，有 10k 个视频片段，分为 487 类，数据规模相对较小。

3.2.3　图像网络

由于视频网络是基于 2D 图像网络发展来，如 TSM 和 GSM 是在 2D 网络结构上加入时间轴上特征融合的模块实现，3D 网络也是简单的将 2D 网络的基本元素扩展为 3D（2D CNN →3D CNN, 2D Pooling →3D Pooling), 学习视频网络需要以 2D 网络为基础。而且，视频网络 Base-line 也是基于对视频每一帧使用 2D CNN 识别，最后对每一帧的输出做简单的平均得到的。在 CEUS 中，对视频数据使用 2D+t 的应用也比较常见。本节简单介绍 2D 图像分类网络的发展，当然，

由于新的改进不断涌现，大部分 SOTA(Sate of Art) 可能一夜之间被一项新的研究提升，此外改进训练策略或者数据集也是一个努力的方向，该部分必然是有局限的。

- 深度学习的复兴源于 2012 年提出的 AlexNet(**?**) 在 ImageNet 比赛中实现一骑绝尘的胜利，AlexNet 包含八层，前五层是卷积层，后三层是完全连接的层。它的激活函数——线性整流函数（Rectified Linear Unit, ReLU）使训练性能得到了改善，在 ImageNet 上最好的的 top5 Accuracy 为 71.194%。

- 2014 年之前的标准 CNN 结构是堆叠卷积层，通过最大池化，连接全连接层，但这类模型占用内存较大，容易过拟合。Inception(**?**) 网络为了减少网络尺寸，提出了 1 ×1 卷积，并在空间中使用不同规模的卷积核然后汇总，Inception-v1（GoogLeNet）使用 7 个 Inception 单元，平均池化和单层的全连接层;Inception-bn（Inception-v2)(**?**) 加入了 Batch Normlization(**?**), 去掉 5 ×5 卷积，改用两个 3 ×3 叠加的卷积, 提高了学习率和衰减系数（Weight Decay)，去除了 Dropout, 取消了 L2 正则化, 并指出 batch 越大优化效果越好；Inception-v3(**?**) 在 v2 基础上进一步缩小网络尺寸，指出特征图从输入到输出应该缓慢减小, 使用 stride 大小为 2 的池化层和卷积层，并且提升了网络的深度，将 n ×n 的卷积核分割为 n ×1 和 1 ×n 两层卷积；后来在**?** 中提出的 Inception-v4 和 Inception-v3 较为相似，Inception-Resnet 加入残差网络的连接方式，Inception-ResNet 的精度和 Inception-v4 一致，在 ImageNet 上最好的的 top5 Accuracy 为 95.3%。

- VGG(**?**) 在 AleXNet 基础上做了改进，整个网络都使用了同样大小的 3 ×3 卷积核尺寸和 2 ×2 最大池化尺寸，引入 1 ×1 的卷积核，采用了 Multi-Scale 的方法来训练和预测，网络结构简洁，虽然在 ImagNet 中的排名较靠后，但在图像风格化领域展现出难以理解的优越性，在 ImageNet 上最好的的 top5 Accuracy 为 VGG19_ BN 的 92.66%。

- ResNet(**?**) 的提出是 CNN 图像史上的一件里程碑事件，最大深度为 152 层，基本框架参考了 VGG19 网络，通过短路机制加入了残差单元，有效改善了深层网络的梯度消失和收敛慢的问题；此外，ResNet 直接使用 stride=2 的卷积做下采样，并且用全局平均层替换了全连接层, 为了保持网络不通层复杂度一致，当 feature map 大小降低一半时，feature map 的数量增加一倍。ResNetXt(**?**) 是 ResNet 的升级版，融合了 VGG 网络堆叠和 Inception 网络 “split-transform-merge”

思想，提出了 Group CNN，即将一个 CNN 层转换为多个尺度相同但通道数更少的 CNN，文章指出，该方法比增加宽度和网络深度更有效，在 ImageNet 上最好的的 top5 Accuracy 为 ResNeXt101_64 ×4d 创造的 94.252%。

- DenseNet(**?**) 吸取了 Resnet 的优点，对比于 ResNet 的 Residual Block，创新性地提出 Dense Block，在每一个 Dense Block 中，任何两层之间都有直接的连接，通过密集连接，缓解梯度消失问题，加强特征传播，鼓励特征复用，极大的减少了参数量，在 ImageNet 上最好的的 top5 Accuracy 为 DenseNet161 创造的 93.798%。

- SENet(**?**) 提出了一种全新的特征重标定策略，通过学习的方式来自动获取到每个特征通道的重要程度，然后依照这个重要程度去提升有用的特征并抑制对当前任务用处不大的特征。SENet154 在 ImageNet 上 top5 Accuracy 为 95.53%。

- SqueezeNet(**?**) 是重要的压缩网络，用比 AlexNet 少 50 倍的参数达到了和 AlexNet 相同的精度, 大量使用 1 ×1 卷积核替换 3 ×3 卷积核，延迟下采样，没有全连接层，定义了 1 ×1 卷积核的 squeeze 层和混合使用 1 ×1 和 3 ×3 卷积核的 expand 层，在 ImageNet 上最好的的 top5 Accuracy 为 80.8%。

- PolyNet() 多项式的角度推导 block 结构，提出了更丰富的结构堆叠方式，文章指出虽然增加网络的深度和宽度能提升性能，但是其收益会很快变少，从结构多样性的角度出发优化模型，也可以提升性能，在 ImageNet 上最好的的 top5 Accuracy 为 95.75。

- DualPathNet(**?**) 在 ResNeXt 的基础上引入了 DenseNet 的核心内容，使得模型对特征的利用更加充分，在 ImageNet 上最好的的 top5 Accuracy 为 DualPathNet107 实现的 94.684%。

- NASNet(**?**) 通过强化学习自动产生最好的网络结构，使用 Proximal Policy Optimization 方法优化，一次决策可以分解为输入两个并行卷积，控制器决定选择哪些 Feature Map 作为输入以及使用哪些运算来计算输入的 Feature Map，再控制器决定如何合并这两个 Feature Map，MSNet 本质上是更为复杂 Inception，在 ImageNet 上最好的的 top5 Accuracy 为 NASNet-A-Large 实现的 96.163%。

- PNASNet(**?**) 是基于 NASNet 的改进网络自动生成方法，其训练时间为 NASNet 的 0.125，采用启发式搜索的策略:Sequential model-based optimization，在缩减的空间中进行搜索降低学习难度，增加 Agent 预测模型的精度，在 ImageNet

上的 top5 Accuracy 为 PNASNet-5-Large 实现的 96.182%。

3.2.4 视频网络

1. BERT（Bidirectional Encoder Representation from Transformers）(**?**) 由 google 在 2018 年提出，提出时在 11 项自然语言处理任务中表现卓越，很快取代 LSTM 等改进的时序网络在自然语言处理中的地位。BERT 的网络结构基于**?** 提出的 transformer 结构。transformer 集成了 self-attention(确定时序信息的关联，同时可以并行计算)、multi-attention（用更复杂的结构增强模型的表达能力）、position encoding（加入位置信息，以区别对待不同位置的输入），transformer 对输入和可能的输出都进行编码以确定二者的关联，BERT 只使用了 transformer 的输入编码结构，通过对 transformer 的堆叠实现层的概念，每一层都是 seq2seq，比如输入是 “Hello, World” 两个英文单词, 输出是 “你好，新世界” 两个中文单词，时序关系任然保留。输入序列中的每一个元素除了送入对应位置的 transformer, 还输入到该层其他的 transformer 中，这样 transformer 除了看到对应位置的输入编码，也可以看到全文，帮助消除歧义（apple 可能是苹果电脑，也可能是水果，需要根据语境判断）。本研究将 2D 网络编码后的视频理解为一个” 句子”，每个” 词” 为一张图像的卷积层输出特征，词的位置关系对应图像在视频序列中的排序。

在大规模视频数据没有出现前，CNN+LSTM 是一种常用的视频分析方法，但由于时间特征和空间特征融合得太晚，且 LSTM 的表达能力有限，在大数据集上表现一般，BERT 作为一种更新的时序网络结构，它的特征表达能力更强（结构更复杂），泛化效果更好（残差网络结构）, 所以本研究将 CNN 预训练模型 + 随机初始化的 BERT 结构纳入迁移学习的范畴。同时，考虑 2t+1D 网络更直接的原因在于实验发现 2D CNN 的性能和 3D CNN 能力相当，我们研究他们之间的过度结构是希望了解 3D 网络的时空特征在 CEUS 迁移中是否是一个冗余的概念。

2. C3D(Convolutional 3D) 及其改进网络。3D CNN 的概念源于 C3D 网络 (**?**)，即使用 3 ×3 ×3 卷积核和 pooling，2D 网络的结构还以 AlexNet 为参考，只有五层卷积层接最大池化和三层全连接层，结构较为简单，输入大小为 $3 \times 16 \times 112^2$，代表 RGB 图像，clip 长度为 16，空间分辨率为 112 ×112 像素。但作者也表明，C3D 网络的参数量增加几乎没有提升网络的性能（当时还没有大规模数据集）。C3D 在 UCF-101 上正确率为 82.3%，在 Sport-1M 上正确率 85.4%。

? 指出 3D CNN 看似更适合做视频处理，但它比 2D 有更多的参数，更难训练和泛化，因此提出了使用 2D 网络预训练数据集然后将网络参数扩展至 3D 维再训练，就是将 2D 卷积核的参数在新增的维度上膨胀（repeat）原有的参数来初始化。I3D 在 HMDB-51 上达到 74.8%，在 UCF-101 上达到 95.6%，在 Kinetics 上达到了 top1 72.1%。**?** 在 I3D 的基础上，提出了 non-local 结构，基于计算机视觉中经典的非局部方法，将某一位置的响应计算为所有位置特征的加权和，以捕获长距离依赖信息，NL-I3D resnet101 在 Kinetics 上达到了 top1 77.7% 的精度，提升较为明显。**?** 在 I3D 的基础上提出了 SlowFast 网络，模型包括 1）低时间分辨率的慢速路径（Slow Pathway），以捕获空间语义，和 2）高时间帧率下分辨率的快速路径（Fast Pathway)，以捕获精细的时间分辨率的运动。快速路径可以通过降低其通道容量而变得非常轻量级，SlowFast 在视频中的动作分类和检测方面都表现出很强的性能，这些提升主要来自慢速路径，SlowFAst-NL-resnet101 在 Kinetics 上达到了 top1 79.8% 的正确率。

3. R(2+1)D 及相似的时间特征先分离再融合网络，该部分的发展基于传统的二维 CNN 计算成本低，但不能捕捉时间关系；基于 3D CNN 的方法可以获得良好的性能，但计算量很大，部署起来非常昂贵的背景。R(2+1)D 的提出源于**?** 观察到应用于视频单个帧的 2D CNNs 在动作识别中仍然表现稳定，并且 3D ResNets 相对 3D inception 可以有效避免过拟合，因而将 3D Resnet 网络的卷积滤波器分解为独立的空间和时间组件，即将 t×d ×d 的卷积核转化为 1 ×d ×d 和 t ×1 ×1 的两步卷积操作，参数量缩小的同时，更容易优化, 在 Sport-1M 上 top5 正确率为 91.2%，Kinetctic top1 为 72.0%，**?** 在使用大规模无监督预训练网络 R(2+1)D Resnet152 网络在 Kinetics 的 top1 为 79.9%，基于 Resnet34 达到 78.2%。

?? 提出了一种通用有效的时移模块：Temporal Shift Module (TSM)。TSM 沿时间维度移动部分通道，从而促进相邻帧之间的信息交换，可以插入到 2D CNN 中, 以零计算和零参数的方式实现 3D CNN 的性能，即在 1 ×d ×d 操作后将空间编码的特征在时间维度向前或向后移动（shift)。**?** 可以部署在移动段 (Jetson Nano, Galaxy Note8), 实现低延迟的在线视频识别。在超级计算机 Summit 上使用 1,536 个 GPU 的进行规模化训练，将 Kinetics 数据集上的训练时间从 49 小时 55 分钟缩短到 14 分 13 秒，达到了惊人的速度和 74.1% 的精度。**?** 指出 R(2+1)D 和 TSM 学习的都是结构化的内核，网络中的任何节点的连接不依赖输入数据，不能很

好表现时间轴上复杂的关联，因此基于 TSM 设计了一个可学习的通道分割门控（splite gate）结构，使时间轴上的特征可以选择 shift 的类型，并且使用残差网络，防止过拟合。GSM 在 Something-v1 上实现了 top1 55.16% 的突破性进步。

3.3 实验设计

3.3.1 数据预处理

1. DICOM 导出为图像序列：使用RadiAnt DICOM VIEWER 4.6.5 软件将从机器导出的 DICOM 视频数据转为图片序列，图片命名为" 肿瘤类型_视频编号_图片在序列中的排序数.bmp"，同一病例图像序列存放在一个文件夹中。

2. 添加时间标签：由于增强超声视频的时间采样频率在超声机器上可以调整，原始的 DICOM 数据可能没有记录确切的帧率，或者在视频导出过程中的处理不当导致帧率记录错误，图片的在序列中的录制时间不能通过简单的方程 (3.1):

$$\frac{frame\_index}{total\_frames} \times (end\_time - begin\_time) + begin\_time \quad \ldots (3.1)$$

由于 CEUS 录制时图像上的固定位置会显示录制的时间，如开始录制后 12s 显示为 "00:12"，一般，相同机器导出相同大小的视频时间文本位置固定，我们对每一个视频的第一张图像的时间文本手工定位后使用 Matlab 2019a 软件中的Optical Character Recognition (OCR) 工具包自动识别整个序列每张图像的时间标签，并将图像的命名方式改为 "肿瘤类型_视频编号_OCR 时间_图片该时刻的排序数"。

3. 去除噪声数据：由于计时一般在造影剂打入后开始，最初的十几秒，造影剂还没有到达肝脏，此时图像中只有背景噪声，而每个人开始增强的时间也有较大差异。此外，部分视频在 2 分钟之前病灶就消退完全，在此之后的数据也基本为背景造影，因为医生在画病灶区域时确定了开始增强时间和消退时间，通过检索图片的命名，将每个视频增强前和消退后的图片删去。该操作将 314 120 张图像缩减为 289 233 张。

4. 平衡数据分布：虽然 HCC 和 ICC 的入组病例数一致，但由于 HCC 的消退时间较晚，ICC 一般早于两分钟消退，实际医生在录制的时候，当造影剂消退了便会停止录制，整理后的数据中，HCC 的图片数大概是 ICC 的 1.3 倍，同时由于 CEUS 的帧频较高且可以调制，部分病例的图片数达到了 3000 张，而时间较短的视频只有 500 多张，为了平衡 HCC 和 ICC 的数据量和各个病例所占的比重，我们将每秒图片张数大于 7 的序列进行随机采样，只保留每秒 7 张图像的帧

率，然后计算平均每个视频的帧数，对大于平均值的视频进行随机采样至平均张数，最后，我们对整个数据集进行随机采样使得 HCC 和 ICC 的图像数一致。通过这个操作，我们分别删减了，最终 HCC 的图片数为，ICC 的图片数为。该操作将 289 233 张图像缩减为 248 262 张。

5. 数据增强：视频的增强方法使用现将图像由 RGB 转化为灰度图像，这是因为超声造影是伪彩图像，转化为灰度图像，一方面可以不同机器彩色化显示时不同的 Colormap 使图像外观有差异的影响，另一方面可以降低内存加快图像处理速度。本实验没有直接将一个完整视频直接送入网络，而是在时间轴上降采样得到多个 Clip，每个 Clip 独立送入网络分类，再将一个视频的多个 Clip 分类结果取平均。训练集的 Clip 或者图片操作流程：在空间轴上 Resize 到比网络标准输入大小多 0-32 个像素的尺寸 → 通过 Color Jittering(**?**) 对一个 Clip 或者单张图像的亮度 ±0.03、饱和度 ±0.03、对比度 ±0.03 → 发生概率为 0.5 的左右镜像操作 → 以 0.5 的概率发生的，对空间轴上某个位置随机的大小为 5000 像素的矩形区域置的像素随机生成 (Random Black)(**?**) → 对一个 Clip 或者单张图像 Random Crop 成网络标准输入的大小，图像的均一化参数与使用网络的给定值一致。测试集中的数据操作为：Resize 到网络标准输入大 16 个像素的尺寸，再通过 Center Crop 和 Normalization 转化成网络标准输入。

6. 固定测试数据：虽然采用了交叉验证，但考虑随机采样生成测试样本在性能比较时，每次测试的数据会有差异，因而对每个交叉验证组的测试集固定测试数据，一次性地随机在每个测试序列中按采样方法生成 300 个 clip 或者 500 张图片，通过灰度化，Resize 到固定大小后使用 Centercrop 去除上下左右各 8 个像素，通过减均值去方差操作后，复制到三个 RGB 通道并保存数据。

3.3.2　网络训练

所有网络训练实验均使用的 SGD 优化器，损失函数为交叉熵损失，fine-tune 过程不对任何网络结构做固定参数的操作，采用固定的学习率。由于研究的是 fine-tune 的效果，实验的学习率设置得较低，保证网络训练过程不出现发散，损失函数不断地收敛。需要注意的是，当将学习率设置较大，迭代次数设置较高，迁移学习将转变为将源数据训练的参数作为网络的初始化值，然后在目标数据集上重新学习特征，这时我们研究的问题变为各个网络能否拟合 CEUS 数据，这与本研究的初衷相违背。

实验使用的服务器配置了 4 块 Intel(R) Xeon(R) Gold 6130 CPU 和 4 块 NVIDIA TITAN Xp GPU（12 GB 显存），基于 Anaconda 环境，使用 Python 3.6.0，显卡驱动版本为 NVIDIA-SMI 440.82，CUDA 版本为 10.2，深度学习框架为 PyTorch1.4.0。具体使用的预训练模型如下：

1. 2D CNN 网络的定义的参数来自 Githb 项目Pretrained models for Pytorch，网络的预训练数据集为 ImageNet。2D CNN 方法在训练时将视频中的每一个图像作为一个独立输入，标签为该图片是否来自 HCC 或者 ICC，测试时将每个视频保存的 500 张图像一次送入网络，网络每次对单张图像预测，得到一个长度为 500 的值为 0 或者 1 的向量，当向量的和大于 250 时，判定视频为 ICC, 否则为 HCC。由于本次研究网络视频的基本框架建立在 ResNet34、ResNet50、ResNet101、BnInception、Mobilenet-v2，我们专门测试了对应 2D 网络的性能，对比改进的模块是否在 CEUS 迁移学习中表现更好。

2. 2D CNN_BERT 网络使用的前端模型和参数与 2D CNN 网络一致，使用 2D CNN 提取图像特征，再在时间轴上将这些特征打包为一个数据块送入 BERT 模块中完成分类任务。本研究使用的 BERT 参考官方代码更改为 pytorch 模型，参考 Alexnet 在卷积层后接 3 个全连接层的方式，实验设定 BERT 层数为 3，其他超参数通过在 AlexNet2D 网络上测试调整，最终确定为输入向量长度 1000，中间层向量长度为 1280，Dropout 比例为 0.2，multi-head attention 个数为 5。

3. C3D 网络参数来自个人项目c3d-pytorch。

4. I3D 和 SlowFast,Slow 网络均来自 Facebook 开源项目SLowFast。

5. R2plus1d 来自 MicroSoft 计算机视觉模型开源库VZM。

6. TSM 来自论文作者 Github 开源项目temporal-shift-module。

7. GSM 来自论文作者 Github 开源项目GSM。

为了保障迁移效果不受各个项目特殊的操作影响（如 warm_up，数据增益），我们只使用这些项目的网络结构定义文件和预训练模型，实验的数据接口等其他部分由我们自行编写，减少因训练 tricks 的不同带来的差异。研究涉及的超参数有学习率，SGD 的动量, 衰减率, 迭代次数，其中学习率, 衰减率和动量采用 grid-search 确定，学习率的搜索范围为：0.001，0.0003，0.0001，0.00003；衰减率的搜索范围为：0.005，0.001，0.0001，0；动量搜索范围为：0.9，0.99，迭代次数的设置为了节约时间，设置最大值后一次性迭代完成，每隔固定间隔，执行

一次测试，并且保存模型，最后选择最优的模型为最终性能。由于 2D CNN 网络和 2D CNN+BERT 的拟合速度快，为了节约时间，最大迭代次数设置分别为 3000 次和 6000 次，每次测试的间隔设置为 1000，其他模型的最大迭代次数设置为 12000，每次测试的间隔设置为 2000。BatchSize 的大小由网络的内存占用量决定，设置为显卡一次计算能允许的最大值。本实验在时间轴上图像采样的方法为 uniform sample，比如网络输入是 8 张图像，则将视频等分为 8 段，使用采样器随机在视频上找到一张图像，计算图像属于视频中的哪一段，在计算其在该段中的比例，再在其他分段中找到和它相同比例的图片，汇总成一个 clip。在性能测试中，除了统计单个视频基于 300 个 clip 的分类结果，我们同时也对单次输入 clip 的正确率和 AUC 进行计算。

3.4　实验结果

3.4.1　基本模型

一次系统地参数遍历后对最优模型汇总，得到表 3.1和图 3.1，这些模型都基于 $3\times8\times224^2$ 的输入尺寸，其中表格中预训练模型的正确率为在原来的数据集上的 top1 正确率，模型参数量的单位为 M(1 million)，迭代次数的单位为 k(1 thousound)，* 号表示没有对应的预训练模型或者参数未知，Slowfast 是双通道模型，输入视频帧数用 “Slowframes（Fastframes）” 表示。图 3.1为 Heatmap 图，横坐标对应不同的分类模型，纵坐标代表视频编号，每个位置的颜色块代表该模型对该视频的分类结果（[0,1] 之间的网络得分，1 为 ICC，0 为 HCC），我们可视化该结果是为了了解不同模型的错误分类结果是否有关联（是否有病例，所有模型都错误判断）图像的上半部分为 ground truth 是 HCC 的视频，下半段是 ICC 视频，理想情况下，图像的上半段颜色值都 ≤0.5，图像的下半部分都 ≥0.5, 这是代表所有病例都被正确识别。

通过图表，可以发现表 3.1可以发现，2D CNN 的性能和复杂的时序网络相比，性能不算很差，这和 Kenetics 分类任务中 2D CNN 的性能也不是很查的现象一致，这说明了 HCC 和 ICC 的外观结构属性在较多的造影图像中可以区分出来，而不必知道时间信息。2D CNN 再加入 BERT 以后，除了 bninception 模型外都出现了普遍的提升，尤其重要的是，2D moblienet+BERT 在所有网络中正确率最高，其次是 2D Bninception, Bninception 的优秀表现可能来自没有残差结构，网

表 3.1 固定输入 8×224^2 时模型迁移性能

Table 3.1 model transfer learning performance when the input size is 8×224^2

迁移模型					训练参数				测试结果						
模型	骨架	数据集	输入大小	正确率(%)	参数量（M)	学习率	衰减率	动量	迭代次数 (k)	正确率	AUC	敏感度	特异性	1-clip 正确率	1-clip AUC
2D CNN	resnet 34	Image Net	1 ×224^2	73.55	21.5	0.0003	0.005	0.99	2	0.8352	0.9077	0.7908	0.865	0.8146	0.848
	resnet 50	Image Net	1 ×224^2	76.02	25.3	0.001	0.005	0.9	1	0.8575	0.9047	0.7995	0.8673	0.8427	0.8671
	bnin -ception	Image Net	1 ×224^2	73.52	<25	0.001	0.005	0.9	1	0.8799	0.9452	0.8295	0.9062	0.8989	0.8649
	mobile - netv2	Image Net	1 ×224^2	71.8	3.47	0.0003	0.005	0.99	1	0.8687	0.9422	0.8206	0.8992	0.8933	0.8503
BERT	resnet 34	Image Net	8 ×224^2	73.55	21.5	0.0001	0	0.99	6	0.8659	0.9446	0.8372	0.9158	0.882	0.8533
	resnet 50	Image Net	8 ×224^2	76.02	25.3	0.0003	0	0.99	3	0.8743	0.9446	0.8378	0.9133	0.9045	0.8519
	bnin -ception	Image Net	8 ×224^2	73.52	<25	0.0003	0	0.99	2	0.8687	0.9365	0.834	0.9081	0.8652	0.8701
	mobilenet v2	Image Net	8 ×224^2	71.8	3.47	0.001	0	0.99	4	0.8799	0.9422	0.8386	0.9134	0.8483	0.9042
C3D	*	Sport-1T	8 ×224^2	*	*	*	*	*	*	*	*	*	*	*	*
I3D	resnet 50	Kinetics	8 ×224^2	73.5	35.3	0.0001	0	0.99	6	0.8687	0.9263	0.8292	0.8979	0.8708	0.8659
Slow fast	resnet 50	Kinetics	8(32) ×224^2	77	32.4 +0.53	0.0001	0.0001	0.9	2	0.8575	0.9104	0.7898	0.8438	0.8539	0.8588
Slow	resnet 50	Kinetics	8 ×224^2	74.8	32.4	0.0001	0.001	0.9	10	0.8184	0.8952	0.8001	0.8744	0.7528	0.8645
R(2 +1)D	resnet 34	ig65	8 ×112^2	*	33.3	0.0001	0.001	0.99	4	0.8771	0.9324	0.8546	0.9245	0.9157	0.849
TSM	resnet 50	Kinetics	8 ×224^2	74.1	25.3	0.001	0	0.9	2	0.8408	0.9182	0.8318	0.9065	0.8427	0.838
	mobile - netv2	Kinetics	8 ×224^2	69.5	3.47	0.001	0	0.9	6	0.8687	0.9253	0.8491	0.9152	0.8876	0.8541
GSM	bnin -ception	Some thing-v1	8 ×224^2	49.01	10.5	0.0003	0	0.9	8	0.8631	0.9358	0.8274	0.9045	0.8427	0.8772

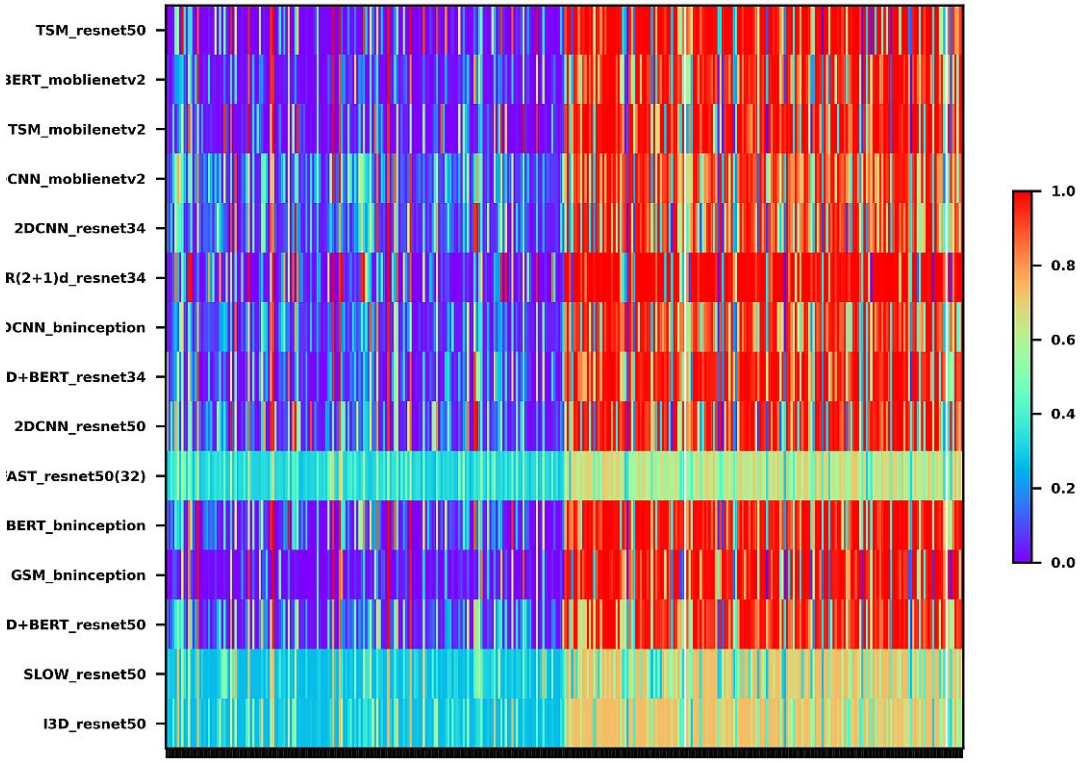


图 3.1 网络测试结果 Heatmap 图，水平方向的刻度代表病例编号，竖直方向为网络结构名，每个位置的方块代表该网络对于此病例的是 ICC 的评分，图像的左半部分是 HCC 病例，右半部分是 ICC 病例。

Figure 3.1 Heatmap of network test results, the horizontal scale represents the case number, the vertical direction is the network structure name, each position of the box represents score that the network think the vedio is ICC. The left half of the image is HCC cases, the right half of the image is ICC cases.

络的 fine-tune 逐层进行，底层的特征得到较好的保留。BERT 的性能表明了时序关系可以通过 2D CNN 转化到高级的语义空间，当成词向量理解成一个 “句子”，做一个类比，医生可以用语言描述造影视频每隔每一段时间造影图像的外观，然后让另一个医生判断这是什么疾病。BERT 的优秀表现，很可能跟造影视频的特定肿瘤的增强模式在时间轴上有规律可循有关。不同于自然场景的视频截取，一个动作可以发生在前几秒或者后几秒，而且事件的切换也比较随机，如两个人见面可以先握手再拥抱，也可以先拥抱再握手，但造影视频的达峰时间一定在开始增强后 30 秒以内，增强一定发生在衰减前。

由于 C3D 的输入尺寸要求为 16×112^2，因而不包含在本对比中，它在性能在表 3.2中。

I3D 和 SlowFast 以及 Slow 这三个网络的性能并不突出，其中 Slow 是所有模型中正确率最低的，对比 SlowFast 模型，我们认为这可能和 SLow 网络网络增加了时间轴上的参数但弱化时间的联系以强调空间有关，这个结构在功能上近似 2D CNN，但又有更多的参数。

R(2+1)D resnet34 网络是在基于视频数据预训练的网络中性能最好，这可能跟它的预训练数据集规模庞大有关，也可能和它自身的设计简洁，从局部逐渐向上的特征提取方式有关。对比 resnet34 网络，性能的提升比 2D+BERT resnet34 好。

TSM resnet50 的性能较 2D resnet50 差，与此同时 TSM mobilenetv2 的性能比 2D mobilenet 差，由于 shift 操作是零参数，对应网络的参数量一致，但新加入的功能反而起到了负面的影响，这可能跟 shift 操作在原问题上的模式不适合迁移到 CEUS 上有关。GSM 基于 bninception，虽然使用了 shift，但通过门控操作可以适应输入数据的做出改变，且通过残差网络，可以跳过不必要的 shift 操作，性能较好且计算量小。

3.4.2 采样帧数

在视频分类任务中，一般提高时间轴上的图像个数，能够提升网络的性能，不同与 8 的采样结果模型测试结果如表 3.2，对比表 3.1，可以发现增加输入量使得所有网络的性能都得到提升，C3D 网络表现较为中庸，当降低 Slow 通道的输入图像数目，SlowFast 和 Slow 网络的性能都下降了，R(2+1)D 和 GSM 的正确率在所有模型中最高，但 GSM 对 ICC 的敏感度更好。

表 3.2 不同采样帧数下模型迁移性能表

Table 3.2 model transfer learning performance under different sample frames

迁移模型					训练参数				测试结果						
模型	**骨架**	**数据集**	**输入大小**	**正确率(%)**	**参数量（M)**	**学习率**	**衰减率**	**动量**	**迭代次数 (k)**	**正确率**	**AUC**	**敏感度**	**特异性**	**1-clip 正确率**	**1-clip AUC**
BERT	**resnet 34**	Image Net	32 ×112^2	76.02	25.3	0.0003	0	0.99	1	0.8771	0.9462	0.8426	0.9171	0.8876	0.8681
	resnet 50	Image Net	16 ×224^2	76.02	25.3	0.0003	0	0.99	5	0.885	0.951	0.852	0.926	0.893	0.878
	bnin -ception	Image Net	16 ×224^2	73.52	<25	0.0003	0	0.99	3	0.863	0.928	0.826	0.893	0.882	0.849
	mobile - netv2	Image Net	16 ×224^2	71.8	3.47	0.001	0.0001	0.99	6	0.883	0.944	0.850	0.914	0.876	0.886
C3D	*	Sport - 1T	16 ×112^2	84.4	79	0.0001	0.0001	0.9	12	0.867	0.942	0.789	0.921	0.871	0.879
Slow fast	**resnet 50**	Kinetics	4(32) ×224^2	75.6	32.4 +0.53	0.0001	0.0001	0.9	10	0.8631	0.9276	0.7956	0.8507	0.8764	0.8525
Slow	**resnet 50**	Kinetics	4 ×224^2	72.7	32.4	0.0001	0.001	0.9	4	0.8073	0.8811	0.7908	0.8576	0.8146	0.8011
R (2+1) D	**resnet 34**	ig65	32 ×112^2	*	33.3	0.0001	0.001	0.99	10	0.8855	0.935	0.8288	0.877	0.9382	0.8477
TSM	**resnet 50**	Kinetics	16 ×224^2	72.6	25.3	0.001	0	0.9	8	0.8743	0.9391	0.8568	0.9345	0.8539	0.8889
GSM	**bnin -ception**	Some thing -v1	16 ×224^2	50.63	10.5	0.0003	0	0.9	8	0.8855	0.946	0.873	0.9374	0.8539	0.9102

3.4.3 密集采样

实际上，在 Kinetics 等数据的模型训练往往基于密集采样（dense sample），即每隔几张图像采集一张，或每隔一秒采集一张的方式，这样可以识别视频中精细时间尺度下的变化，对于发生速度快速的动作尤其有必要，但这个方法生成的数据在时间轴上跨度较小，由于 CEUS 的变化持续时间久，不是 Kinetics 中十几秒的短视频，理论上，均匀采样涵盖的信息更全，效果会更好。我们对 2D CNN_bninception，2D+BERT_resnet50_16，C3D_16,GSM_bninception_16 这四个在所属子类中性能较好的网络在保持最优参数不变的情况下，使用密集采用，结果如表 3.3：

表 3.3 密集采样时模型迁移性能表

Table 3.3 model transfer learning performance under dense sample

迁移模型					训练参数				测试结果						
模型	骨架	数据集	输入大小	正确率(%)	参数量（M)	学习率	衰减率	动量	迭代次数 (k)	正确率	AUC	敏感度	特异性	1-clip 正确率	1-clip AUC
2D CNN	bnin -ception	Image Net	224^2	73.52	<25	0.0003	0	0.99	2	0.832	0.885	0.779	0.847	0.837	0.828
BERT	resnet 50	Image Net	16 ×224^2	76.02	25.3	0.0003	0	0.99	3	0.858	0.921	0.816	0.880	0.876	0.843
C3D	*	Sport -1T	16 ×112^2	84.4	79	0.0001	0	0.9	12	0.902	0.958	0.853	0.924	0.899	0.904
GSM	bnin -ception	Some thing -v1	16 ×224^2	50.63	10.5	0.0003	0	0.9	8	0.866	0.938	0.838	0.910	0.876	0.857

通过表 3.3中的数据和和表 3.2中对应模型的性能，我们发现正确率都出现了明显的下降，1 clip 的真确率也都下降了，且 ICC 的敏感度变差。3.3表明短时间的造影 Clip 不能提供较多的时间轴上动态变化的信息帮助分类器做出正确的决定，均匀采样因为提供更完整的造影变化过程，迁移时，即便原模型是密集采样，仍然效果更好。

3.4.4　数据清洗

在实验中我们为了平衡 HCC 和 ICC 的数据量以及每个病例图片数量，缩减了 21% 的图片，为了探讨这样的操作对性能的影响，我们对 2D CNN_bninception，2D+BERT_resnet50_16，C3D_16,GSM_bninception_16 这四个在所属子类中性能较好的网络在保持除迭代次数外其他参数不变的情况下，使用全部的数据训练网络，结果如表 3.4:

表 3.4 完整训练集上模型迁移性能表

Table 3.4 model transfer learning performance on the full training set

迁移模型					训练参数				测试结果						
模型	骨架	数据集	输入大小	正确率 (%)	参数量（M)	学习率	衰减率	动量	迭代次数 (k)	正确率	AUC	敏感度	特异性	1-clip 正确率	1-clip AUC
2D CNN	**bnin -ception**	Image Net	224^2	73.52	<25	0.0003	0	0.99	2	0.832	0.885	0.779	0.847	0.837	0.828
BERT	**resnet 50**	Image Net	16 $\times 224^2$	76.02	25.3	0.0003	0	0.99	3	0.858	0.921	0.816	0.880	0.876	0.843
C3D	*	Sport - 1T	16 $\times 112^2$	84.4	79	0.0001	0.0001	0.9	12	0.902	0.958	0.853	0.924	0.899	0.904
GSM	**bnin -ception**	Some thing -v1	16 $\times 224^2$	50.63	10.5	0	0.0003	0.9	8	0.877	0.949	0.867	0.941	0.871	0.881

通过对比表 3.4和表 3.2中对应模型的性能，我们发现在 2D CNN_bninception，2D+BERT_resnet50_16,GSM_bninception_16 都出现正确率的下降，但在 C3D_16 中，正确率和敏感度都提升了，其他模型的 ICC 敏感度都下降，这意味着更少的 ICC 被正确判断，更严重的医疗事故将发生，这个现象出现的原因很可能是数据集中有更多的 HCC 图片，因而出现了较高的 Bias，而 C3D 性能的提升很可能跟它参数量大，需要更多数据拟合有关。

3.4.5　non-local 全局特征

在视频分类任务中，non-local 作为一个有效的模块，能够学习输入的全局关联，在 I3D 中可以有效提升性能，我们研究了 TSM_NLN_resnet50_8 和 I3D_-

NLN_resnet50_8，其中超参数的选择除了迭代次数外，与各自未加 NLN 的模型参数一致，结果如表 3.5:

表 3.5 non-local 模型迁移性能表

Table 3.5 non-local model transfer learning performance table

迁移模型					训练参数				测试结果						
模型	骨架	数据集	输入大小	正确率(%)	参数量（M)	学习率	衰减率	动量	迭代次数 (k)	正确率	AUC	敏感度	特异性	1-clip 正确率	1-clip AUC
I3D	nln resnet 50	Kinetics	0.74	35.5	8×224^2	0.0001	0.001	0.9	6	0.855	0.932	0.832	0.900	0.854	0.854
TSN	nln resnet 50	Kinetics	0.756	25.3	8×224^2	0.0001	0.001	0.9	10	0.858	0.925	0.828	0.901	0.837	0.871

通过对比表 3.1中的 I3D 模型和 TSM mobilev2，我们发现网络的性能都出现了下降。由于迁移学习是基于卷积特征能够将图像基本模块（边，角）提取出来，这些基本元素可以很好地利用在其他数据集上实现的。non-local 模块相比卷积算子，它的计算结果涉及整个输入，很容易猜想到，由于 CEUS 的视频内容不同与 Kinetics 数据集，他们的特征连接方式必然是差异较大的，正如搭建积木的材料是一样的，但最终搭建出来的东西完全不同，则自搭建方法必然是的不同的类比，使用 Non-local 这样的全局连接方法是不利于 CEUS 迁移学习的。

第 4 章　总结展望

4.1　研究局限

研究工作紧密围绕医工交叉，致力于探究如何用计算机科学辅助更好的医疗数据分析，使得更多的数据工作者可以较快高效的完成基础的造影视频数据集建立和分析网络搭建，为现有研究涉及较少但前景广泛的造影数据自动化分析提供有基础工具和参考方法。由于研究者具有扎实的数字图像处理能力和深度学习数据分析能力，通过大量的编程工作和网络训练实验，每个课题的完成度高，并开源了实验代码，使用者可以轻易的在较短时间内复现两个实验，调整跟踪算法参数或者导入自定义的跟踪方法，无需编写实际使用中需要的交互界面;或者训练适合在新的数据集上的迁移学习深度学习模型，通过多重比较，寻找最优基础模型并进一步优化改进。

研究工作的局限性总体上来自如下两点:

1. 研究工作集中在编程和数据实验，在创新上偏向应用层面，没有提出新颖的方法或者理论；

2. 研究工作围绕着特定医院的回顾性数据，没有外部实验验证模型或结论是否在其他数据上同样可靠。

跟踪软件的设计作为一个较完整的作品，它的局限性主要有以下五条：

1. 跟踪程序基于 Matlab 平台，无法在操作系统上独立运行，使用者虽然不需了解 Matlab 编程规则，但是需要自行注册和安装，如果用户所持的教育账号不支持 Matlab 免费订阅服务，则需要付费订购 Matlab 的 Computer Vision Toolbox 和 Image Processing Toolbox；

2. 程序基于跟踪算法不能自动检测出病灶的位置，需要医生给出需要跟踪的位置，这意味着医生在操作过程中还是需要通过观看连续的造影视频确定病灶的位置，没有接受专业读片训练的人无法使用该软件标记出有研究意义的跟踪区域。

3. 软件的跟踪算法基于特征点，在原则上不适合跟踪变化较大的物体，原始的造影视频数据由于较高的时间采样率使每一帧之间的物体的变化较小，因而具备较好的跟踪效果，但如果输入的造影数据进过后期加工的降采样数据，比

如一秒一帧的视频帧率，软件可能就不再适用。

4. 软件目前不能直接读取 DICOM 格式的原始数据，一方面由于不同超声设备所使用的数据结构不同，Matlab 支持的 DICOM 读取接口可能会读取错误或者降低数据质量，建议使用厂家提供的专业数据格式转化软件，另一方面直接读取压缩率低的 DICOM 数据在配置较低的笔记本电脑上极易导致内存不够用，电脑死机。通常，原始的 DICOM 数据可以经过其他专业软件如 RadiAnt 转化成 avi 等 Matlab 的 VideoReader 函数支持的视频格式，详情参见官方说明。

5. 由于超声造影记录的成像区域视野有限，病灶区域往往位于图像中心或者在图中占据较大比例。与此同时，深度学习模型也在积极集成病灶定位和分类功能，以及更进一步的弱监督学习，网络以后也许可以使用训练好的定位模型在没有人工标记的情况下清洗数据，或者直接对整个成像区域分析。当然对于采集过程规范的视频，整个录制过程中病灶的位置几乎没有改变时，这时固定位置的框也是很好的选择。当然，在目前超声造影领域的能够自动定位并且推广使用的深度学习模型还不存在的背景下，对于绝大多数采集过程不规范的回顾性数据而言软件的应用还是有必要的。与其同时，对于基于深度学习方法定位和跟踪的算法更需要这类软件来快速构建精准的训练标签。

基于深度学习的 HCC 和 ICC 分类问题的超声造影视频模型研究虽然讨论较为全面，但不足如下：

1. 没有尝试独立构建模型。医学数据分析领域很多研究使用自定义的网络结构，本研究没有这样做的理由如下：第一：研究没有外部验证，在单中心的数据集上密集调参会导致模型对数据集的针对性太强，对数据集的大小敏感，很可能在外部数据集和其他研究问题上表现的较差。通用的经典网络结构，在各种数据集和研究问题上证实了模型卓越的性能，使用时只需简单调参，当数据增加时也不用担心网络的表达能力不足，重新定义网络结构。第二：本次研究所使用的数据集在原始的迁移模型上没有产生严重的过拟合现象，所获得的结果在可以接受的性能中，因而与其把时间消耗在定义新的网络上，不如去思考如何让迁移网络性能更好。

2. 没有以较大的学习率训练网络。由于研究是关于不同视频分类模型的迁移能力探讨，当学习率过高时，模型中从自然数据集中获得的参数结构被很大程度的破坏，不再满足微调（fine-tune），反而将其变成一种常用的网络初始化方

法——将使用预训练模型的策略，将研究问题变成不同网络结构拟合造影视频能力的问题。研究使用较小的学习率和较少的迭代次数，是为了说明用自然场景的大数据训练得到网络在轻微调整后对超声造影视频的拟合能力。

3. 研究的模型有限。在视频分析领域有着大量优秀的模型和处理方法，比如在前几年非常流行的彩色图像结合光流数据的模型，随着新的方法理论的不断涌现，很多曾近的 SOTA 模型被超越，相比之下，本次研究只是选取了在激烈的竞争后目前热度最高，综合性能最好的模型框架。此外本研究基于训练好的开源模型，所以对于有些 SOTA 模型也不在比较范围中，还有如等过于庞大的模型，由于不能在单张 10G 显卡上实现一次计算，也被排除在研究外。

4. 深度模型框架带来的差异。由于本研究主要 Pytorch 框架，但部分研究如 SlowFast 是基于 Caffe 实现，实际使用中通过 OMNX 将原始的权重转化为 pytorch 可以读取的数据类型。目前，相同模型在不同框架上即使使用相同的参数，由于底层算法略微有差异，也会导致性能有轻微的改变。

5. 没有计算 p-value。原因如下：第一：本研究与基于深度学习方法，与统计方法相比，深度学习方法没有对数据分布做假设，没有分析数据本身的是否满足某种分布，而是通过正确率，AUC, 敏感性和特异性来体现模型是否可以很好拟合数据，这是两种研究文化的差异。第二：研究使用交叉验证使单个病例之间相互独立的假设被破坏，出现在第一个测试组中的数据会出现后来三组实验的训练组，基于此情况证明一次交叉验证得到的一组模型不具备统计学差异是不可行的，在实验中我发现，交叉验证不同组的 AUC 较稳定，但是 ACC 的差异很容易较大，但由于实验结果取了平均值，结果是有参考价值的。第三：数据分组没有计算训练组和测试组的临床基线，虽然分组是随机的且分析过程只使用图像数据，没有使用任何临床资料，但在不能假设图像表征和临床基线无关的情况下，没有统计确实是本研究的不足。第四：统计学的结论成立都是由有前提条件的，比如数据符合某种分布，但深度学习作为黑箱模型，不乏数据只发生轻微改变，结论发生巨大变化的怪异现象，在我看来针对高度非线性的模型，目前还没有很好的统计描述方法，而且笔者本人没有接受过系统的统计学教育，容易错误使用。第五：没有通过大量重复性实验证明结论的稳健性，一方面研究时间跨度有限，另一方面由于网络不是随机初始化的，一般而言初始化对网络的性能影响较大，不同初始化会使网络收敛到不同局部最优解，但对于迁移学习而言，网络

的初始值在每次试验中固定，其他超参数固定时，结果应该偏差较小。

6. 没有可视化，现有的可视化算法缺乏理论支持，深度学习目前还是一个黑箱模型，可视化结果在整个数据集上没有规律性，不能帮助医生诊断，但为了方便理解，论文附录部分做了 cnn+bert 的可视化实验。

4.2 展望

随着 CEUS 的应用范围扩大，CEUS 的数据量将会快速增长，关于其潜在的医学价值也将很有可能通过深度学习挖掘发现。可以预见未来一段时间，深度学习在其他学科卓越的表现会吸引很多从事 CEUS 量化分析的研究者整理出部分足够大的回顾性数据来改进现有的医学模型。这些研究会唤起人们对于数据收集的重视，当规范有目的性的 CEUS 数据收集到一定程度时，会出现在模型改进上的研究，这些改进针对 CEUS 数据的特殊性质设计，人们将不再只会照搬现有的模型框架，或者搭建简单的网络。改进的算法会因其卓越的性能大大降低 CEUS 数据分析难度, 促使人们更愿意使用深度学习这样流程简单性能强大的方法解决问题。可以说，只要 CEUS 的临床应用价值扩大，深度学习的发展不停滞，两个学科的交融和互相促进的局面就会发生。目前，已经有一批研究者成功将 DL 搬移到 CEUS 临床问题的解决中。

未来，CEUS+AI 的局面可能是这样的：

- 建立了开源的高质量的 CEUS 视频大数据集, 该数据集很可能是关于肝内各种病变的自动分割和分类，虽然 CEUS 在肝脏疾病的应用最为广泛，现有的研究往往基于结节良恶性分类，没有涉及精细分类或者分类性能差，这些研究不能帮助 AI+CEUS 成为在实际诊断中辅助医生判断的工具。建立肝内病变完备的数据集将帮助政府机构授权的计算机辅助诊断在 CEUS 领域落实。
- 研究方法将不再局限于 CNN 分类网络，生成对抗学习 (Generative Adversarial Network)，图网络 (Graph Nerual Network)，蒸馏学习 (Distill Learning)，作用域迁移学习（Domain Transfer learning），元学习 (Meta Learning) 等先进的 AI 理念将会应用在 CEUS 中，学习任务也不在局限与分类和分割。
- 物理建模和 AI 模型互相促进，参考 DCE-US 领域探讨从流体动力学和药物代谢角度建立客观公正的现实世界描述，AI 模型很可能会结合 DCE-US，帮助建立更精准稳定的 DCE-US 特征和更具解释能力的 DL 模型。

- 通过增强现实技术辅助医生执行 CEUS 操作,CEUS 的成像参数较多，成像效果差异大，通过 AI 计算可以自动化调节成像参数，提示医生寻找合适的成像切面，并通过图像后处理技术优化图像质量，提升诊断体验。
- 寻找能在不同问题上表现优良的通用特征或者网络结构，降低重复性劳动，随着 CEUS 应用在更多器官中，我们不希望开辟一个个研究分支，我们更希望一项研究能够迁移复用在其他问题上。

与此同时医疗影像 +AI 的大环境发展也能给 CEUS+AI 带来蓬勃生机。

AI 医学影像可能在未来大幅增强图像分割、特征提取、定量分析、对比分析等能力，实现常见病灶识别与标注、病灶性质判断、靶区自动勾画、影像三维重建、影像分类和检索等功能。在眼底筛查、X 线胸片阅片、脑区分割、脑疾病诊断、骨伤鉴定、骨龄分析、器官勾画、病理切片分析、皮肤病辅助诊断这些领域出现集成化的通用性强的医学辅助诊断系统。以眼科为例，目前基于眼底照的 AI 算法对于眼底疾病、视神经疾病的诊断已经接近人类医师的水准。2018 年，美国食品药品监督管理局 (FDA) 已批准了世界上首款使用人工智能检测糖尿病患者视网膜病变的二类医疗设备 IDx-DR 上市。未来，AI 影像的广泛应用能将医生从低附加值、重复性劳动中解放出来，提升诊断、放疗、手术的效率，降低医院成本、提高诊疗效果、改善病人就诊体验。

在我看来，未来的 AI 影像可能出现像手机领域那样的行业巨头，但更可能各个开发公司之间达成共识，各自针对不同器官或者图像类型寻找特色化的发展道路，研究者会在算法的鲁棒性、安全性、易用性，数据集构建，以及在算法和技术层面针对小样本、多模态、分布式样本等方向进行更多探索。目前，市场上大部分产品同质化严重，集中在糖网、肺结节等领域。目前，国内外有上百家企业具备相关的产品或算法，包括:1) 百度、Google 等科技巨头。科技巨头资金雄厚、能长期布局和投资，AI 技术和人才实力有积累，并能结合云平台提供服务，最容易形成全疾病范围、多区域覆盖、平台式的产品服务。2) 推想科技、依图医疗等 AI 创企。AI 创企对市场反应灵活，部分厂商在某些影像领域布局较早，通过与医院进行科研合作、集成进医院信息化厂商软件等形式，形成了一定数据和算法壁垒。3)GE、Phillips、联影等影像设备企业。

目前，AI 影像还没有建立标准化的监管体系和商业落地模式，虽然图玛深维、深睿医疗、雅森科技等公司的产品已获得了二类医疗器械认证，其产品应用

仅限用于病灶检出、异常征象识别，再由医生确认病症，暂不可用于诊断领域。AI 影像监管主要瓶颈在于 AI 不同于传统的计算机辅助系统，其具有自学习和快速迭代的特性，推理过程也不完全透明，需要通过有效手段明确其是否能产生一致的、稳健的、可靠的结论。此外，由于我国各地医疗信息化水平不一，影像数据标准不一、质量参差不齐，各个医院和 AI 厂商之间的数据合作推动艰难。为了加快 AI 影像的商业化落地和数据采集工作，部分 AI 厂商积极参与中国及国际标准制定，以期共同推动标准的前进。例如，2019 年初，灵医智惠作为牵头单位，向世界卫生组织和国际电信联盟成立的健康医疗人工智能焦点组递交了眼底影像和临床辅助决策系统等 2 项标准提案。

当然，AI 影像的大规模应用缺乏的不止是数据集建立，标准建立这样的应用基础，还有关键性的算法研究领域需要攻克诸多问题，我们不能因为个别成功的案例就想当然的以为方法就在那里，只是缺乏践行材料。以高维数据分析为例，在医学中广泛使用的 3 维 CT,MRI 和 CEUS, 可以借鉴视频网络中的大量网络结构框架，但事实上，即使目前最先进的视频网络在人工整理过的数据上的表现仍然无法超越人类。以 SOMETHING-V1(**?**) 数据集为例，它是大量带有密集标签的视频剪辑的集合，这些视频剪辑显示人类对日常对象执行预定义的基本动作，该数据集由大量的标记工作者创建，致力于让机器学习模型能够鉴别物理世界中发生的基本动作，比如“从上到下”这样的物体移动动作，目前，官方显示的 top-1 Accuracy 仅为 54.1%，这样的分类性能如果发生在医学诊断中，是万万不可行的。

此外，不同于现实数据集庞大的体量，很多医学问题因为发病率低无法建立大规模数据集，我们不知道，在小数据上，这些方法能否表现良好。**?** 已发现基于 AI 的医学成像系统（包括 MRI，CT 和 NMR）的缺陷。他们考虑了三个关键问题：与微小扰动或运动相关的不稳定性；关于微小结构变化的不稳定性，例如带有或不带有小肿瘤的大脑图像；以及样本数量变化的不稳定性。他们发现，某些微小的运动会导致最终图像中出现大量伪像，细节被模糊或完全去除，并且图像重建的质量会因反复进行二次采样而变差。这些错误广泛分布在不同类型的神经网络中。

深度学习方法作为端到端的黑箱模型缺乏解释性，跟经典的基于数学和物理建模方法相比，数据分析者即使是没有任何医学数据处理经验，只要给定输入

和输出数据，就可以进行大量密集的调参实验，给出一个看着像样的模型。目前，相比正统的深度学习算法研究，医学图像分析缺少数据，缺少客观公正的方法评价体系，缺乏创新性，甚至缺少有新意的研究问题，整体的研究风格偏向应用。虽然，在研究 AI 影像初期，我们可以先从简单的问题入手，比如二维影像数据或者高维数据的分割问题（不同分类，分割可以基于局部信息作为网络输入，不必将整个数据完整送入网络，一次性给出结论），但研究者终究需要面对小数据建模难以拟合的巨大挑战，独立思考医学 AI 的针对性改进方法。

面对有限数据集的拟合，除了迁移学习，元学习，笔者认为从两个方面着手，第一是降低数据集中的不确定性，比如降低分析问题的难度，细化分类任务，去除特殊的数据；第二是增加模型的确定性知识，与主流深度学习领域研究的基于光学成像直接复现事物外貌的数据不同，医学成像利用电磁场或者机械场穿透器官，间接反应人体器官的生物化学或者物理属性，灰度图像的值反映了该属性的强弱。与自然世界中存在即合理不同，我们确切的知道，正常的器官应该具备什么样的理化性质，特定的病变应该有什么图像特征，医学领域有大量成文的诊断标准和建模（如仿体实验，人工合成数据集），这些知识应该帮助我们弥补数据上的不足。

最后，正如文章在第一章引言部分提出的机器学习方法无法通过模型拟合效果差拒绝一个假设不成立；无法通过模型卓越的表现肯定假设成立并且模型能在其他数据集上表现优越；无法确定这种关联是否重要且稳健这三个问题，未来，基于数据分布的探索可能会进一步发展，人们可能更希望模型能通过无监督学习找到潜在的决定数据生成的变量，并使用这些变量来解释结果。比如现有的分类问题，通过判断病变是否是 HCC，更深层的研究会试图回答，如果正常组织病变成 HCC，它的图像是什么样的。

作者简历及攻读学位期间发表的学术论文与研究成果

作者简历

基本信息

安徽合肥

教育经历

2017 年 9 月-2020 年 9 月　研究生

中国科学院大学　人工智能技术学院　模式识别与智能系统

2013 年 9 月-2017 年 7 月本科

哈尔滨工业大学　航天学院　探测指导与控制技术专业

致　谢

致谢是在我写论文的第一天写的，因为我只有很短的时间来写这个论文，很多事情不能尽善尽美，加上自己又很喜欢抠细节，这次写作必然不能按我心意进行。但是相比写作过程不断反思自己的不争气，我选择从一开始承认自己的失败与无能，这样我可以不做无用的思想斗争，专心看看自己已经知道了什么，还有什么不知道。我从小接受克己复礼的教育，收敛自己的锋芒和克制自己的欲望，然而做人做事和做学问都是需要悉心研磨，这些年我很遗憾丢失了斗志和尊严。

如果说感谢，我感谢经验教训，希望这些能让我在以后的路走的更加坚定；我感谢我的朋友，在我低落的时候鼓励和陪伴了我；我感谢那些流淌过生命中的真善美的存在，有些人接触不深，但依然给人留下了很好的印象和深刻的影响。当然，我需要感谢很多倍感启发的书和文章；感谢计算机科学积极开源分享和创新的研究氛围，让很多问题可以通过检索得到答案或参照；感谢实验室优质的软硬件条件，可以随时调用公共资源做实验；感谢北京这座包罗万象的城市，它汇聚了各行各业的优秀人才，很多大楼的招牌就足够让人仰慕和振奋。最后，需要表达是对未来的向往，希望自己可以积极开拓人生新局面，不虚晃度日和碌碌无为。