摘要

健康是人类终生的话题，一个人从孩童到暮年的各个阶段都伴随着医疗资源的消耗。截至 2021 年 5 月 11 日，根据中国第七次人口普查结果显示，我国人口总人口约 144350 万人，且同年数据显示我国约 102.3 万个医疗卫生机构，床位约 944.8 万张，医疗人员约 1066 万人。海量的人口与匮乏的医疗资源的矛盾日益凸显，在这样的背景下，人工智能（Artificial Intelligence，AI）技术以其高效的决策分析的优势全方面赋能医疗行业的各个领域，将医疗机构的诊断水平和效率提升到新的高度，并且促进了医疗资源的合理配置。

在儿童先天性肾盂输尿管连接部梗阻性积水（Ureteropelvic Junction Obstruction Connection, UPJO）领域内，俗称肾积水，其中 B 超是该疾病最常用的筛查手段，且主要治疗手段仍以手术为主。诊断阶段通过观测超声图像中肾盂经线、肾盏扩张程度、肾实质厚度、皮质回声等指标判断病理等级（SFU 0-4），这一阶段需要经验丰富的影像科医师通过细致的检测做出判断从而确定手术时机。而目前该疾病的首选高效诊断技术在国内发展仍很落后，国内分布的相关医学影像人才也不平衡、不充分，因此在诊断阶段需投入大量的医疗时间成本和人力资源成本，从而需要计算机图像处理技术来降本增效。传统的计算机图像处理方对超声图像的处理准确度不高，而超声图像本身为灰度图像，具备边界模糊性、视角多样性，且各病理等级对应的数据集不平衡，对本文所研究的智能诊断分类方法提出了一定的挑战。为解决上述问题，本文研究基于深度学习的计算机视觉技术，提出一种基于肾脏超声图像的智能诊断方案，具体所做工作如下：

(1) 构建基于肾脏超声图像的语义分割模型，针对超声图像的背景区域和器官区域边界模糊的问题，分割网络的注意力需要集中在器官和病变区域的问题，研究提出与注意力模块（Triplet Attention Module，NAM）结合的基础特提取网络并在高级特征提取阶段采用金字塔池化模块（Pyramid Pooling Module , PPM）的图像语义分割网络作为解决方法。在包含 1850 张带注释的超声图像的肾积水数据集上进行实验，包括注意力模块的排列，参数量的计算以及性能。

(2) 在语义分割模型提取的分割特征图并结合原图的数据集下，构建疾病严重程度的图像分类模型，为尽可能提高分类准确率，并针对特征信息丢失最严重的池化层，研究提出一种具有小波池化增益层的分类神经网络，在利用上一步分割模型预测的语义特征图和原图结合下，补充预测并构造的共 3289 张图像带标注的数据集上（SFU 0-4）进行实验，最终结果表面我们的诊断模型在分割任务和分类任务上均具备优越性与有效性。

关键词：肾脏超声图像；深度学习；注意力机制；金字塔场景解析；小波变换

Abstract

Health is a topic of human life, and a person is accompanied by the consumption of medical resources at various stages from childhood to old age. As of May 11, 2021, according to the results of the seventh census of China, the total population of my country is about 1,443.5 million, and the data of the same year shows that there are about 1.023 million medical and health institutions in my country, 9.448 million medical beds, and about 10.66 million medical personnel. The contradiction between the massive population and the lack of medical resources is becoming increasingly prominent. Against this background, artificial intelligence (AI) technology, with its advantages in efficient decision-making analysis, empowers all fields of the medical industry in all aspects, and integrates medical institutions The diagnostic level and efficiency have been raised to a new level, and the rational allocation of medical resources has been promoted.

In the field of congenital ureteropelvic junction obstruction (UPJO) in children, B-ultrasound is the most commonly used screening method for this disease, and the main treatment method is still surgery. In the diagnosis stage, the pathological grade (SFU 0-4) is judged by observing the renal pelvis meridian, the expansion degree of the renal calyces, the thickness of the renal parenchyma, and cortical echoes in the ultrasound image. At this stage, experienced radiologists need to make judgments through careful detection, so as to determine the timing of surgery. At present, the development of the preferred efficient diagnostic technology for this disease is still very backward in China, and the distribution of relevant medical imaging talents in China is unbalanced and insufficient, and a large amount of medical time and human resource costs need to be invested in the diagnosis stage. Therefore, it is necessary to use medical images processing to achieve cost reduction and efficiency improvement. The accuracy of ultrasonic image processing is not high enough under the traditional computer vision technology, and the ultrasonic image itself is a grayscale image with fuzzy boundaries and diverse viewing angles, and the data sets corresponding to each pathological grade are unbalanced. The above problems have caused certain challenges to the intelligent diagnosis classification method studied in this paper. In order to solve the above problems, this paper studies the computer vision technology based on deep learning, and proposes an intelligent diagnosis scheme based on kidney ultrasound images. The specific work is as follows:

(1) We built a semantic segmentation model based on kidney ultrasound images. Aiming at the problem of blurred boundaries between the background area and the organ area of the ultrasound image, the attention of the segmentation network needs to focus on the problem of organs and lesion areas. The research proposal and attention The basic feature extraction network combined with the Triplet Attention Module

(NAM) and the image semantic segmentation network of the Pyramid Pooling Module (PPM) are used as a solution in the advanced feature extraction stage. Experiments are performed on the hydronephrosis dataset containing 1850 annotated ultrasound images, including the arrangement of attention modules, the calculation of parameter quantities, and the performance.

(2) Based on the segmentation feature map extracted by the semantic segmentation model and the data set of the original image, an image classification model of disease severity is constructed. In order to improve the classification accuracy as much as possible, and for the pooling layer with the most serious loss of feature information. This research proposes a classification neural network with wavelet pooling gain layers. Combined with the semantic feature map predicted by the previous segmentation model and the original image, the predicted and constructed data of a total of 3289 images with annotations are supplemented. Experiments are carried out on the set (SFU 0-4), and the final results show that our diagnostic model is superior and effective in both segmentation and classification tasks.

Key words: ultrasound image, deep learning, attention mechanism, feature pyramid networks, discrete wavelet transform

目录

第 1 章 绪论

随着人工智能时代的到来，在国内医疗资源匮乏、不平衡的现状下，基于大数据医疗的 AI 方法变得越来越重要。但是医疗领域中的应用场景十分丰富，各类医疗背景下的 AI 处理方案并不能适用于所有类别，各类背景下与之对应的处理方法研究成果尚不成熟，因此最近几年，医疗人工智能的应用越来越多的需求产生，大量的研究人员对其产生了研究兴趣，许多成果以及技术方法也都被应用在各种基于大数据 AI 的医疗任务上，并且从成果数量上看已经是百花齐放，在各个医疗处理场景下都有相当的优秀产出，大大节省了各地的医疗资源的利用。随着医疗 AI 获得越来越多的关注度，其研究意义和应用场景也会进一步增多。本文研究在儿童群体中的先天性肾积水的医疗场景下展开，应对当前国内医疗背景和该疾病严重性的综合问题，试图探索通过 AI 等智能手段提供新时代的先天性肾积水的临床诊断方案。

本章首先对医疗 AI 与先天性肾积水的相关知识以及在当前的研究价值进行了细致的阐述。然后对国内外该场景下的研究现状进行了简述，并通过调研国内外现有方法的相关内容，得到了现有方法在各自医疗场景下的缺点与不足。最后，再详细介绍本文的研究内容和组织架构。

1.1 研究背景和意义

先天性 UPJO 是最为常见的小儿泌尿系畸形之一，是导致儿童患有梗阻性肾积水的主要原因，该疾病会使得尿液流出受阻导致系统内压力长期持续增大，肾脏的血流量持续性减少，间接导致肾脏持续性缺血性损伤，并发不可逆的病理改变，如果不给予及时的治疗将导致肾脏功能永久性损害，或者限制儿童肾脏发育最终潜能，为了尽量减少长期后遗症，患者需要进行离断型肾盂输尿管成形手术。然而，由于儿童的肾脏多有相当的代偿功能，因此对积水巨大的肾脏要谨慎处理从而不宜草率做出决定，手术时机的判断成为关键一步。B 型超声波检查可协助尿路梗阻的定位、肾积水病理等级的判别，是医疗机构对于该疾病最常见的诊断方式之一，主要的检查手段是产前超声检测和产后超声检查。超声技术依赖于通过压电元件发射和接收超高频声波脉冲的换能器，具有低成本、便携性的优点，安全阈值内的电离辐射相对安全并适合儿童群体。然而，尽管它有优势，但也有一些缺点:机械和热机制的潜在生物效应可能会限制扫描时间，检查的质量和准确性取决于操作人员的培训水平。UPJO 在超声检查的图像上主要表现为器官内部出现一定面积的阴影区域，人工的分辨方式会根据超声图像中的几何、纹理、

阴影分布等特征来判断疾病的病理等级，然而，由于超声波图像是高噪声的，即使是训练有素的医学影像医师也要获得足够的信息[1]，这种人工的特征提取处理方式使得医学影像医师的处理速度不高。因此为解决该问题，随着人工智能(AI)技术的发展，科研工作者找到了可以利用 AI 结合超声图像帮助预测梗阻性肾积水的可能性，通过 AI 领域的计算机视觉技术对超声图像做大数据 AI 分析，快速有效地提出一个初步的诊断结果，从而加快医疗进程。

早期 AI 模型的一些局限性阻碍了医学的广泛接受和应用。在 21 世纪初，深度学习的出现克服了许多这些限制。现在的 AI 能够分析复杂的算法和自我学习，通过风险评估模型应用于临床实践，提高诊断准确性和提高工作流程效率，基于数据采集与分析，应用于健康监测、临床护理等医疗领域，以互联网技术为依托，凭借基础设施与专业采集设备对各类医疗数据的收集，实现了 AI 与大数据服务在医疗行业的广泛应用，帮助提高了医疗行业的诊断效率以及医疗质量。但是，我国的医疗资源分配存在严重不均的问题，精确优良的医疗设备和经验丰富的医护资源大多集中在较为发达的城市与局部发达地区，这使得大量的发展中地区甚至贫困地区的患者群体在本地得不到可靠的医疗服务，向发达地区的大型医院集中从而造成了更为复杂难解的医疗环境。

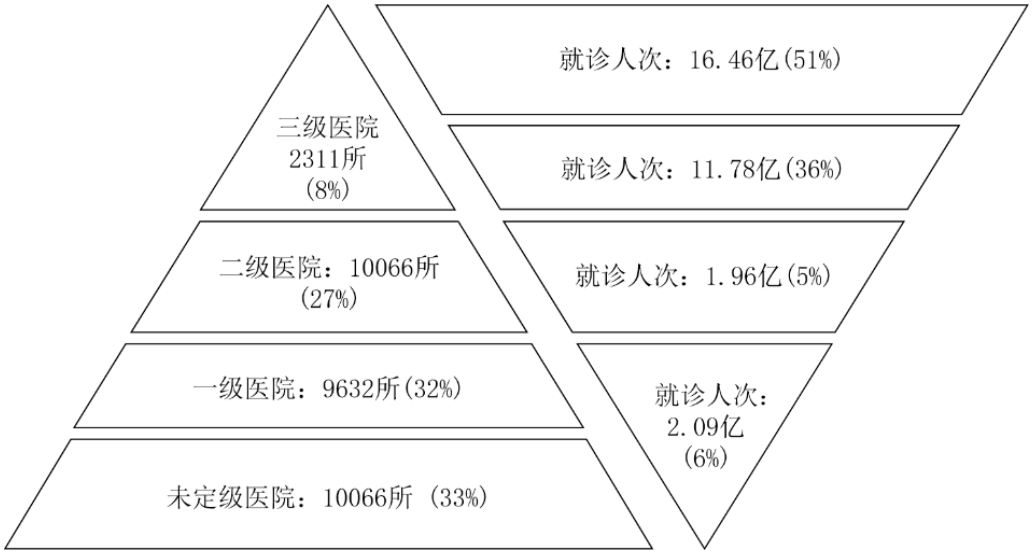

图 1-1 医疗资源比例图和对应的就诊人次比例图

Fig. 1-1 Medical resource ratio chart and corresponding visit ratio chart

如图 1-1 所示，截止 2020 年底由国家卫计委统计发布的权威数据显示，国内建设存在有大约 32476 所正规医院，其中仅有约 2498 家三级医院，约总额的 8%，然而在三级医院就诊人数却达到 16.46 亿人次约占全国就诊人次总额的 50%，由此可见医疗资源供需现状显著不匹配，医疗服务失衡的现象十分严峻。目前 AI 医疗的主要应用在于医学影像的诊断环节，主要解决的需求如表 1-1 所

示：

表 1-1 AI 医疗的主要应用现状

Tab. 1-1 Main applications status of AI in medicine

需求	描述
病灶识别标注	针对医学影像进行图像分割、特征提取、定量分析
靶区自动勾画	针对肿瘤放疗环节的影像进行处理
影像三维重建	针对手术环节的应用进行 AI 识别并完成三维重建

在医学图像处理中，大部份工作是与超声图像相结合的科学应用，因为超声具有无电离辐射、无痛无创、轻便简洁、即时成像、高复用性等等优势优点，现已广泛应用于肾、肝、心血管等内脏器官，以及浅表结构的初筛与医疗诊断[1]中，结合 AI 技术并用于超声图像的分割、特征的提取和分类等方面。例如通过超声图像分割算法将数据分割成不同的组织器官区域，从而帮助医疗工作者更准确地定位病变区域。同时也可以通过匹配算法从时间线维度，与的不同成像方式的医学图像进行配准，从而比较病变区域的变化情况。此外，AI 还可以通过特征提取算法，提取医学图像中的特征信息并加以既定的形态学处理然后得出需要的评价参数，从而辅助医生进行诊断。例如对于肾积水的诊断，AI 可以通过特征提取算法提取医学图像中的病变区域，从而帮助医生更准确地判断病变类型和程度，也可以通过分类算法对不同程度病理等级的肾积水超声图像进行分类，从而实现自动化诊断。在基于超声图像的病灶识别与标注领域，AI 的发展在不断地促进这一领域蓬勃向上的趋势，也不断有创新性的研究成果在各医学影像的应用场景下孵出，解放医学影像人才的劳动力，节省大量的超声医疗资源，帮助医疗机构与患者节省财力，并且能够辅助医生对患者的手术时机做出及时有效的判断，因此具备极大的社会价值。现有的研究现状中也有基于对比实验的研究表明 AI 与超声图像的联合可以极大地简化医师人工诊断的操作步骤，甚至是一定程度上降低了主观差异性、节约了资深医师的珍稀资源并缩短了医疗诊断报告的时间以提高诊断的效率。

1.2 国内外研究现状

深度学习技术可以实现对肾积水超声图像数据纹理、像素、空间分布等特征信息的学习，自动找出实现分类等需求的重要特征并得出结论。由先天性 UPJO 引发的儿童肾积水的手术时机以及病理原理在医学界仍处于研究阶段，尤其是手术时机的判断存在一定的争议[54]，结合相关研究相对其它医疗 AI 应用领域较少，然而针对超声图像的相关的研究在随着 AI 医疗的发展，越来越多的结合 AI 技术的跨域研究成果出现，并应用于众多地区，在医学图像处理应用场景下，它可以

极大地帮助医生的诊断能力，同时缓解欠发达地区的医疗资源失衡和短缺[2-7]。如果我们在节省了经验型医师的人工操作的情况下下，实现超声检查阶段自动区分病变区域，不仅可以支持泌尿外科医生的进一步诊断，还可以节省大量的医疗资源、人力、金钱，帮助负担过重的患者，这一技术不仅很大程度上缓解了医疗短缺资源和不同地区之间的不平衡，并且促进了医生的诊断能力[8-11]。

早期大部分 AI 医疗相关的研究都以各参数为数据的机器学习为主，之后深度学习的发展使其重心都倾向于超声图像的病灶识别与标注[12]，长期以来该技术一直应用于医学图像处理等各个领域，极大地促进了医生的诊断能力。Zhu[13]等人开发了一种使用深度卷积神经网络对超声图像中的甲状腺和乳腺病变进行分类的自动方案，其结论表明其模型在对乳腺癌和甲状腺癌进行分类时相比机器学习具有更深潜力，比放射科医生的准确性更高，除此之外，AI 在肝脏超声方面也有相关的应用，主要为肝脏脂肪检测及评估肝纤维化等[14]。Biswas[15] 等人运用深度学习方法评估脂肪肝，其结论表明使用深度学习比机器学习能更好地判别脂肪肝的病灶特征。Burgos 等人开发了一种 AI 算法，通过分析胎儿大脑信息自动估计胎龄，结论表明使用超声胎儿平面图像的自动化识别方法在孕龄估计方面产生了更低的误差。Shah 等人[16]总回顾了 47 篇文章，报告了 AI 在泌尿系统癌症中的应用，在所有良性条件下，AI 都被用来预测手术的结果，影像组学在肾脏肿块的分类和核分级、膀胱癌的膀胱镜诊断、格里森评分预测以及磁共振成像与前列腺癌的计算机辅助诊断中都有 AI 的相关应用，结论表面临床范式即将转变，AI 应用将在该领域中找到自己的位置，并彻底改变决策过程。关于肾脏区域的病理研究，肾脏病理等级是肾脏疾病诊断的金标准，例如肾积水的 SFU_0-4 级来区分病理严重程度以提供手术时机的判定。迄今为止已有多项研究显示 AI 技术能够在各项疾病诊断金标准中得到了相比人工效率更高且具有临床价值的成熟应用，比如说肾小管、肾小球、肾脏间质等基本的器官结构和病变的病灶识别、预测肾脏病理分级等，从而实现自动化的、量化的特征识别、病理分级以及临床诊断，Gallego 等人[17]使用深度学习框架从数字化肾脏切片中自动进行肾小球分类和检测，结论表明，该技术适用于在全幻灯片图像（WSI）中正确检测肾小球，显示出鲁棒性并降低了假阳性和假阴性的检测率。对于 UPJO 诊断，早期已有试图从机器学习领域寻求效率突破的研究，Lorenzo 等人[18]结合机器学习与云计算技术，提出了一个通过数据分析解决当代挑战的机会，报告了一个超越当前标准的创造性解决方案，探索利用云机器学习平台预测产前婴儿 UPJO 手术时机的潜在价值。Blum 等人[6]引入了一种动态解决方案，使用自动信号分析和机器学习来分析利尿肾图的引流曲线，研究结果表明，机器学习方法具有潜在的临床实用性，可以更早地发现 UPJO 病例，但以上研究可靠性不足，仅仅证明机器

学习方法的潜在实用性价值。

深度学习的兴起，激发了医疗 AI 领域关于先天性 UPJO 引发的儿童肾积水疾病的研究，迄今已有国内外学者完成过探索并发表了相关研究，然而已有的可靠研究成果并不丰富，大部分还停留在理论阶段，截至 2021 年，Hameed 等人[19]在该领域收集大量资料，其综述结论表明 AI 在该领域的未来是光明的，但在提供可靠的结果方面，仍然有很大的改进和增长空间以积极影响更多的患者。Adree 等人[20]研究结论表明使用定量指标对肾积水引发的膀胱输尿管反流（VUR）进行分级是可能的，未来可以应用机器学习方法对 VUR 进行客观评分。之前提到过肾积水的现阶段主要治疗手段为肾盂成形手术，Drysdale 等人[21]提出一种模型预测肾积水的肾盂成形手术后再次干预的风险和时间，结论表面模型表现良好，其方法的实施在该科领域是新颖的，可能有助于对接受肾盂成形手术的患者进行个性化的风险分层，但需要进一步的真实世界验证。Smail 等人[2]探讨了针对肾积水病理 SFU 分级的问题，提出一种平均准确率为 71%的深度学习模型并首次证明了 CNN 方法对肾积水超声图像的适用性，Tabrizi 等人[22]提出了一种基于深度学习的方法来预测肾积水的严重程度，该方法由肾脏和输尿管交界处是否存在阻塞或阻塞来定义，首先对肾脏进行半自动分割，以分析其表征梗阻的外观。然后开发了一个基于深度学习的模型，使用超声图像中的每个切片来预测阻塞，将平均准确率提高到 78%。Lin 等人[23]提出一种结合注意力的模型将肾脏和扩张的盆腔系统与液体分开，结果表明模型检测扩张盆腔系统的敏感性和特异性分别为 99%和 83%，检测肾积水的敏感性和特异性分别为 90%和 80%，将检测肾积水的准确度进一步提高。Lien 等人研究结论表面 Res-UNet 算法对 SFU_1-3 级和 SFU_4 级的肾积水分类具有显著的准确度，但数据集来源患者年龄跨度太大，并不适用于儿童先天性 UPJO 的临床诊断，也缺乏对正常和轻症患者的科学评估。综上现有方法对先天性儿童 UPJO 超声图像的准确度以及可靠性，均未达到该领域临床诊断的要求。

1.3 本文的主要工作

本文针对先天性 UPJO 引发的儿童肾积水疾病的诊断，为缓解当前医疗资源与医疗需求的不平衡的矛盾，节省医学影像科医师人力、医疗资源、金钱和帮助受苦的病人，提出了一种基于肾积水超声图像的结合计算机视觉的智能诊断技术。伴随着医疗器械的发展，如何利用超声图像来辅助医生完成最优的医疗诊断、治理过程也变得尤为关键，本文研究项目来源于国家重点研究项目，与首都医科大学附属儿童医院协同合作，并将先天性儿童 UPJO 在医学诊断阶段的智能分级

作为主要的研究方向，突破性地提出了一种从语义分割模型再到图像分类模型的分段诊断方法，模型会利用超声图像所展示的器官特征、病变特征、纹理特征等信息，分别对不同类别的患者的肾脏病理等级做出高效率、高准确度的初步诊断。本研究的模型将在基于现有研究的基础上，进一步提高诊断的准确率，为该领域医疗 AI 处理落地临床实用的最终目标贡献科研成果。以下是本文的主要研究内容：

(1) 分析数据并对现有医疗图像方法的调研与分析

首先对医院提供的超声检查的视频进行了图像帧的周期性抽取，然后对图像数据集中的脏数据进行做清理与过滤，分析思考其中运用计算机视觉技术的难点与痛点。接着针对性地调研相关该领域内前人的可靠的研究成果，探索创新性的解决方法，分析现有研究成果的模型与其优劣性，然后研究合适的优化算法，并提出完整的方法论。

(2) 提出一种基于肾积水超声图像的图像语义分割模型

在图像语义分割阶段，针对器官区域、病变区域和非相关背景区域的位置关系和边界模糊的问题，我们提出了一种结合注意力机制和金字塔场景解析结构的语义分割模型，帮助神经网络关注标注的器官与病变区域、更好地学习全局信息以区分各区域，最终提高图像语义分割的精度，然后将分割结果按一定比例与原图结合生成新的数据集。

(3) 提出一种融合小波池化增益层的图像分类模型

在图像分类阶段，创新性地引入离散小波替换池化层，从频域解析出三个低频分量与一个高频分量，并行学习，在神经网络中构造分支完成参数传播并完成异化特征组合的学习，其中在归一化处理模块、激活函数模块、神经网络剪枝优化模块方面也均调研了相关技术并应用其中，综合提高了最终分类模型对于UPJO超声影像的诊断准确率。

尽管在 UPJO 病理研究领域内，结合深度学习方法的研究比较少，对于 UPJO 引发的肾积水的诊断仍具有挑战性，因为其表现各不相同且非特异性，超声具有易于接近、无辐射暴露和重复评估的特点，成为肾积水的补充诊断工具，但经过耗时的培训后，影像科医师之间的主观差异仍然存在，也不断有学者提出最新研究成果并评估深度学习算法通过超声图像检测肾积水的可行性，表明深度学习算法对肾积水的检测具有显著的召回率、特异性、精密度和准确度，减少了超声医师之间的变异性并提高临床条件下的效率，但是这些研究忽略了轻症甚至正常的超声图像的诊断识别，直接对图像分类的诊断过程也欠缺鲁棒性。相对而言，本研究的主要贡献有三方面:

(1) 我们提出了一种分割-分类的诊断框架，并将其应用于肾积水超声图像

的智能诊断，取得了优异的综合性能。

(2) 在分割阶段不破坏骨干网络结构的前提下满足迁移学习条件并轻量级地嵌入注意力机制提高图像分割精度，结合金字塔场景解析模块针对性地解决了肾脏超声图像的上下文特征难题。

(3) 在分类阶段思考图像频率域的运用可能性，并创新型地加入小波池化增益层辅助神经网络模型学习到更多维度的、更加鲁棒的特征信息，并提高最终的分类性能。

1.4 本文的主要组织结构

本文的组织结构如下：

第一章：绪论。

本章主要从研究的相关背景和各研究意义的角度出发分别介绍了 AI 与医疗领域的结合，细分到 AI 在医学影像领域内的研究现状，再然后到 UPJO 应用场景下的现有研究，分析各研究的贡献与不足。之后引出本文研究所解决的问题和以研究内容的应用价值，以及本研究所提出的解决方案以及主要工作，最后阐述了本文的组织结构。

第二章：相关知识

本章首先对深度学习在肾脏超声领域内的相关研究分析，提出不足并引出本文研究，并基于本文对肾脏超声图像的诊断模型所用的技术进行了分解，大体分为语义分割和图像分类两个阶段，并分别介绍了两个阶段的相关技术发展、相关的评价指标和计算方式，然后背景知识内分别介绍了各个阶段内的更细粒度的技术模块，包括注意力模块、金字塔场景解析模块和各种优化算法，并简单介绍了这些模块下对应领域内的相关研究，以及对本次研究的结合的最优方案。

第三章：基于肾积水超声图像的语义分割模型

本章节首先会大致勾勒并介绍分割网络模型的大致结构，并按输入到输出的顺序一一进行阐述与理论分析，对注意力机制的算法和结合网络的方式进行公式化的介绍，具象化地介绍了注意力机制所解决的问题，然后详细介绍了在高级特征处理阶段的金字塔场景解析过程，阐述其算法原理与解决的问题，最后再详细地介绍这一阶段对整个诊断过程所做出的贡献、其作用与意义。

第四章：基于离散小波池化增益的图像分类模型

本章节首先大致介绍了分类网络模型的大致结构，同样地以输入到输出的顺序进行解析，在对离散小波变换的运用介绍中，由连续小波变换循序渐进到图像中的离散小变换，再由一维的计算公式推导出二维图像下的计算公式，由浅入深、

由简到繁的分析其理论原理与神经网络的结合。然后分别介绍了其中所用的优化方法，包括剪枝算法、归一化算法等，并从理论原理角度分析其算法过程，以及与本文研究方向为何、如何相结合的细致介绍。

第 2 章 相关知识

上文主要基于先天性 UPJO 引发的儿童肾积水疾病的研究背景和意义、国内外研究现状进行了详细的研究。在本章，我们将对肾脏超声图像的智能诊断方案的相关知识包括深度学习与医疗图像处理的相关技术，我们方法的整体框架，实验结果的评价标准，以及注意力机制的相关阐述和数字图像处理中的频域图像处理的方法，最后对模型算法的优化算法和原理进行了介绍。

2.1 医疗图像处理与深度学习

2.1.1 医学图像分割

从仿生学角度模仿人类大脑的低级视觉处理方法，通过深度卷积神经网络（DCNN）结构拟人化地检测并提取出医学超声图像中的可用特征，如图 1-1 所示，这些特征的具体表现形式可以是形态学上贴近直线的线性特征，即肾脏的肾皮质和肾柱的形态学特征，或者是逼近圆形的特征，如图中肾窦和肾锥体的形态，然后是更高阶的特征，例如局部的和全局的空间区域和纹理特征提取，上下文特征信息，例如器官与器官内组织的空间关系，比如肾窦和肾锥体呈扇形，肾锥体之间由肾柱相连。

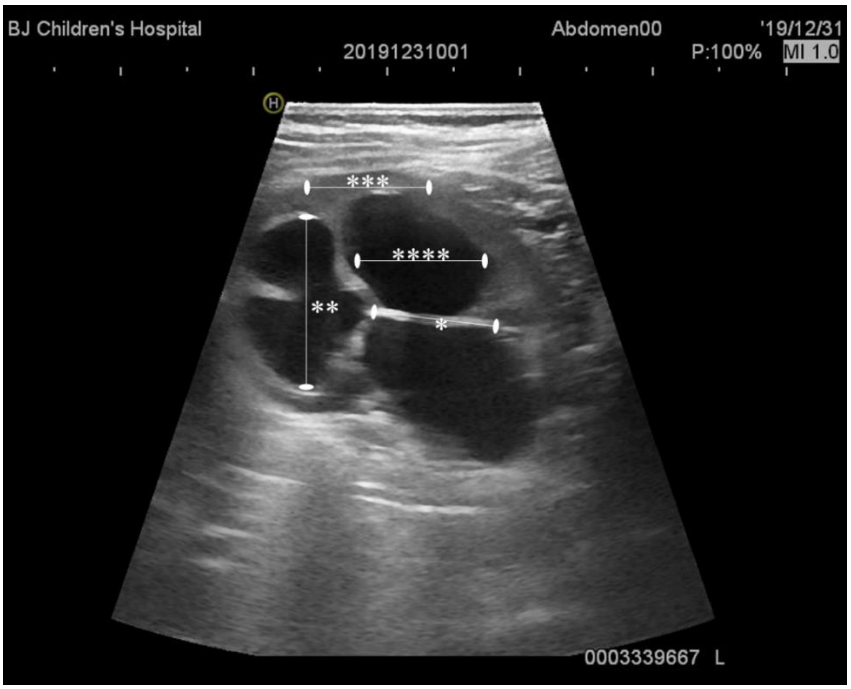

图 2-1 图示为先天肾积水患者超声影像切片图像，‘*’表示肾柱，‘**’表示病变的肾窦，‘***’表示肾皮质，‘****’表示病变的肾锥体

Fig. 2-1 The picture shows the ultrasound slice images of patients with congenital hydronephrosis, '*' indicates the renal column, '**' indicates the renal sinus, '***' indicates the renal cortex, and '****' indicates the renal cone body

图像的语义分割使计算机视觉领域内的经典问题，其具体实施例是把原始的图像数据作为输入送进语义分割神经网络模型，然后将输出转换为具有显著性标识的感兴趣区域的掩膜图像，其中输入数据的图像中的每一个单个像素会根据其所识别的类型而被分配以类别的标签[24]，具体过程实施例如图 2-2 所示。

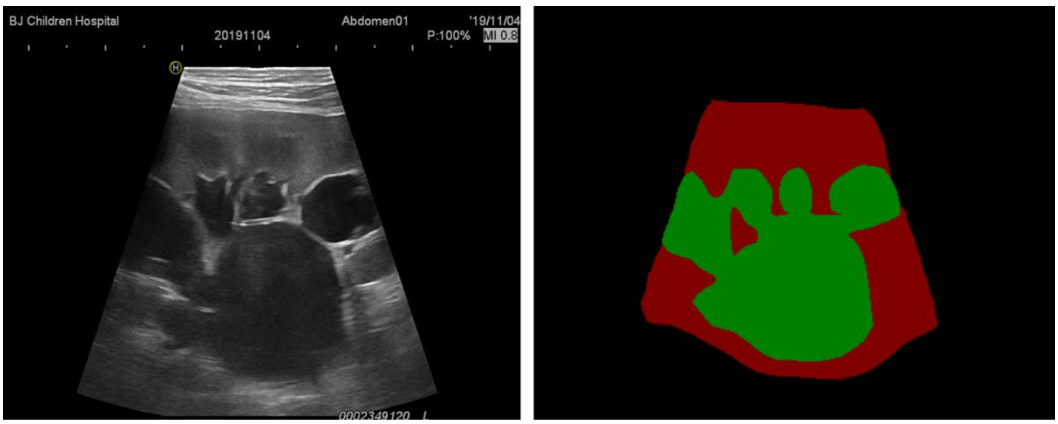

图 2-2 图示为图像语义分割的过程，将输入图像 A 输出转换成掩模图像 B

Fig. 2-2 The picture shows the process of image semantic segmentation, converting the input image A output into a mask image B

对于医学图影像数据，准确分割医学图像是放射治疗计划期间轮廓的关键步骤，并且早在 2010 年就有权威研究论述了 AI 技术在医疗图像处理中的应用[25]，其中医疗图像的语义分割同样也是依据不同区域间的区分度大小，将原始图像按不同语义分割出若干区域并显著性掩模化的过程[26-28]。目前，该领域的应用主要以各类别细胞、器官组织的医学灰度图像作为处理对象,而传统的图像语义分割技术分别有以区域为基准的和以边界为基准的分割方法，前者主要依赖于图像的局部空间特征，例如之前提到过的灰度、纹理、空间分布及像素统计特性的均匀性等等，而后者主要是利用形态学的梯度信息来模糊确定对象目标的边界特征。通常来讲，医学图像的语义分割可以用集合论模型来描述：即预定义一组医疗图像 $M.I.$，然后预定义一组相似性约束 $Cons._i\,(i=1,2,\cdots)$,$M.I.$的语义分割即是得到该数据的所有不同子集区域的划分结果，即：

$$\bigcup_{x=1}^{N} R_x = M.I., R_x \cap R_y = \emptyset, \forall x \neq y, x, y \in [1, N] \qquad (2-1)$$

其中 R_x 满足信息相似约束 $Cons._i\,(i=1,2,\cdots)$ 下所有像素的集合。对于 R_y 同理，x, y 用来区分不同区域，N 为不小于 2 的正整数，表示区域划分后的序号。

对于早期的图像语义分割算法的手段，其主要是以灰度值得聚类分割、条件随机场算法等一些相对传统的语义分割算法[29]：

(1) 最原始的语义分割形式涉及分配区域必须满足的硬编码规则或属性[30]，以便为其分配特定标签。规则可以根据像素的属性来构建。通过拆分与合并的算法递归地将输入图像按不同属性规则分割出各类子集区域，一直持续到可以分配标签为止，然后通过合并算法将相邻的子区域与相同的标签进行组合。这种方法的问题是规则必须实现硬编码。此外，仅用灰色级别的信息来表示复杂的类是极其困难的，在面对超声图像这类复杂的图像场景运用下，这种方式显然已经因为其滞后性与低效性，使得无法应用于医疗图像处理领域。因此，需要特征提取和优化技术来正确地学习这些复杂类所需的表示。

(2) 随着机器学习的兴起，科研工作者考虑并尝试通过训练出一个参数模型为每个像素完成识别与分类的任务并以此分割图像的语义。然而当模型存在缺陷的情况下，分割出来的语义图像往往是噪声密度较大的分割结果。这种问题可以通过补充考虑各像素之间的先验关系来规避，对于当前预测的像素，其附近的像素往往与其具有一致的类别标签。研究者们为了模拟这样的局部相似性关系，提出条件随机场、即一种用于结构化预测的统计建模方法来完成，该算法与离散分类器的不同在于前者可以在预测对象的类别之前考虑该对象的相邻上下文的像素类别——即像素之间的关系，这一技术要点使得该算法成为语义分割的理想候选，迄今仍有结合机器学习完成医学图像分割的研究成果[31, 32]

(3) 从深度学习逐渐取缔机器学习在语义分割任务中的地位之后，首个基于深度学习范式的图像语义分割模型全卷积网络[33]（Fully Convolutional Networks，FCN）被提出并不断地迭代创新衍生更具优越性的语义分割模型:

1） FCN 作为深度学习下语义分割的最初模型，其主要贡献有 3 点：a、将端到端的卷积网络推广到语义分割中;b、重新将预先训练的 ImageNet 网络用于分割问题中;c、使用反卷积层进行上采样并提出了跳跃连接来改善上采样的粗糙程度。该模型的关键在于网络中的全连接层可以看作是使用卷积核遍历整个输入区域的卷积操作。该模型用卷积运算实现全连接层结构，并且将预先训练的网络模型的全连接层卷积化。由于 CNN 网络中的池化操作，得到的特征图谱仍需进行上采样，其中反卷积层在进行上采样时，不是使用简单的双线性插值，而是通过参数学习实现插值操作，因此该网络层也被称为上卷积、完全卷积、转置卷积或是分形卷积层。然而，由于在池化操作中丢失部分信息，使得即使加上反卷积层的上采样操作也会产生粗糙的分割图，因此该算法还从高分辨率特性图谱中引入了跳跃连接方式，但仍然这类分割算法对于噪声密集、细节信息与边缘信息模糊的肾积水超声图像的运用场景下效果不佳。

2） 针对 FCN 在语义分割过程中网络感受野大小固定和分割对象的细节容

易忽视或被算法平滑的难题,SegNet[34]应用而生，它创新了一种编码-解码结构，编码部分主要由 VGG[35]网络的前端 13 个卷积层、5 个池化层构造组成，解码部分由相同的 13 个卷积层以及 5 个上采样层构造组成，最后的解码器输出的高级特征再通过参数可学习的分类器中，对每个独立的像素完成分类操作并分配标签。SegNet 还特定地通过池化指数来保留特性数据的轮廓特征并且降低了网络参数的数量，但该算法应用于医学影像数据的领域时，在性能上仍然有所欠缺。

3） 出于医学影像自身敏感性等特点,提供开源渠道并能使用的训练数据数量相对其它图像领域较少，因此 U-Net[36]模型算法应运而生，它高效地利用少量数据完成训练并实现高性能检测的效果，提出了处理大尺寸图像的有效方法。其主要贡献也有三点：a、以重叠-拼接的策略能够实现对任意大的图像进行无缝分割，同时每个图像块也获得了相应的上下文信息。b、随机弹性变形进行数据增强搭配图像分块在数据量较少的情况下起到了扩充数据量的作用。更重要的是，这种策略不需要对原图进行缩放，每个位置的像素大小与原图保持一致，不会因为缩放而带来误差。c、引入加权损失函数以解决数据不平衡的问题。该模型继承了 FCN 的学术思想，但相比 FCN 的结构，U-Net 是完全对称的，并且模型对解码器进行了卷积加深的处理，而 FCN 仅进行了上采样。两者算法都用了跳跃连接的结构，这种结构最明显的好处即可以联合低层的细粒度表层信息和高层的语义信息，因此很好的满足了图像分割对这两方面信息的需求。由于该算法的优越性，在医学图像分割领域的很多场景都有应用，但对于 UPJO 引发的肾积水的超声图像的应用仍未有成熟的研究成果。

4） 在处理高级语义信息的算法中还有 DeepLab[37-39]系列，DeepLabv1 由于深度卷积网络在不断的最大值池化和下采样之后会丢失关键的细节特征信息，因此改用扩张卷积算法以成倍地增加网络感受野然后可以获得更多的上下文特征信息。考虑到 DCNN 结构在图像识别中的空间低敏性，这一缺点极大限制了模型识别的精度，于是有学者提出以全连接方式构造条件随机场来提高模型更准确地捕捉细节特征的能力。DeepLabv2 语义分割模型补充空间金字塔池化结构，通过多尺度、多采样率的扩张卷积来提取图像特征，再将特征融合以捕获不同大小的上下文信息。DeepLabv3 语义分割模型进一步在空间金字塔池化层中补充了全局平均池化层，与此同时，在平行的扩张卷积处理之后补充批量归一化网络层，从而赋能捕获全局语义特征信息，后续研究还参考补充了编码-解码模块以恢复原始的像素信息，使得分割的细节信息能够更好的保留，同时编码丰富的上下文信息。尽管该系列优化的算法先

进，模型改进成熟，在很多图像分割领域都有应用，但在医疗图像分割领域仍然表现不佳，因此更少有学者在肾积水超声数据上产出过可靠的研究。

5） 聚焦全局上下文信息的网络同样还有本次研究最为相关且使用的 PSPNet[40]（Pyramid Scene Parsing Network，PSP）金字塔场景解析网络语义分割模型首先结合预训练网络和扩张网络来提取图像的特征，得到原图像八分之一大小的特征图，之后通过 PPM 实现池化与上采样操作，最后与最初输入特征图进行连接后经过卷积层输出最后的预测分割图像。其主要特点包括：PSPNet 是建立在 FCN 之上的像素级的分类网络，将大小不同的内核集中在一起激活特征图的不同区域创建空间池金字塔。特性映射来自网络被转换成不同分辨率的激活，并经过多尺度处理池层，稍后向上采样并与原始层连接进行分割的特征图。学习的过程利用辅助分类器进一步优化了像残差网络这样的深度网络模型，不同类型的池模块侧重于激活的不同区域特征图。该算法的空间处理特性很好的满足了 UPJO 图像下，肾脏器官组织的空间分布特性，因而比较适用于肾积水超声图像运用场景下的语义分割任务。

以上的内容主要对深度学习时代下图像语义分割算法的发展路程和相关知识进行了详细阐述，对相对传统的语义分割算法如何迭代到当前的深度学习模型的算法进行了综合性的介绍，对基于深度学习的语义分割模型结合肾积水超声图像以及其他医学图像的现有研究和有效性进行了评估，发现现有的图像语义分割方法依赖于各种不同于医学图像，也阐述了适用于本文研究所涉肾积水超声图像的分割方法。

2.1.2 医疗图像分类

有关对肾积水超声图像的直接病理性分类的研究现阶段未有成熟的理论成果，考虑到肾积水超声图像的场景复杂性，直接分类的效果大都不够理想。深度学习中发展最成熟的技术运用即是分类算法，在医疗图像病灶识别和人脸识别等场景都有成熟的应用。其中 DCNN 是相对广泛使用的结构，图像分类任务从传统的方法到基于 AI 的方法的演变历经了多年的发展，接下来的描述重点关注 AI 的进展并讲述图像分类的发展的几个重要的节点，进而拓展阐述医学图像的分类。

MNIST 是首个具有通用学术意义的基准数据，其成为 AI 分类算法的发展中首个手写数字的分类标准数据集并包含 60000 例训练集数据和 10000 例测试集数据，所有图像例均为像素大小为 28 的正方形灰度图。在上个世纪末的早期 AI 发展阶段，最为前沿且广泛应用于分类实验的是支持向量机和 K 最邻近算法为代表的机器学习算法，其中支持向量机实现了 MNIST 分类错误率为 0.56%的优

异效果，而当时的神经网络代表算法为 LeNet 系列网络，其实验效果仍不如机器学习算法，即使是迭代到 1998 年的 LeNet5，其对于 MINIST 的识别错误率仍停留于 0.7%的较低水平。随着计算机算力的提高、硬件的进化更新促进了神经网络结构的发展，神经网络算法逐渐表现出巨大潜力以及优越性，以识别错误率为 0.23%实现了最佳的分类效果。到了本世纪初期，信息量井喷式增长带来的庞大规模的数据集，受限于硬件的发展和算法的局限性，神经网络的训练和优化仍然是艰难重重。MNIST 和 CIFAR 数据集从规模上、数据分布上均属于语义简单的分类数据，其衍生出的算法若是在医疗图像领域中落地，面临更加复杂的图像分类任务，其准确度和召回率等指标表示算法有效性是远远不够的。直到 Krizhevsky 等人[41]在 2012 年提出了基于深度学习模型的 AlexNet 并在当年的 ImageNet 图像分类竞赛中夺得冠军，自此深度学习开始了爆发式发展，作为第一个深度卷积神经网络，它的网络层数相比 LeNet5 增加了 3 层，网络的参数量增加、数据维度也从 28 提高到 224，相比 LeNet5 补充了更好的激活函数和更快的网络模型的收敛速度，与此同时还解决了 Sigmoid 激活函数在网络层数较高的情况下引发的梯度消失或弥散的严重问题。之后随之不断地更新优化与迭代，先后补充了随机剪枝层 DropOut 来抑制过拟合的现象，补充标准归一化网络层以解决局部的网络神经元之间劣性竞争活动创建的消极机制并抑制了反馈幅度小的神经元放大反馈大的神经元而造成的网络偏移现象，从而有效地增强了网络模型的泛化能力，随后的迭代发展从数据增强的层面补充了裁剪翻转等预处理算法来进一步增强网络模型的泛化能力。最终成长为约 240M 的参数量的模型，基于此模型的成功经验不断拓展了大量的分类应用场景，例如人脸识别。并随着图像处理器(GPU)的横向拓展应用，促进深度学习进入新的范式。自此不断涌现学者从医疗 AI 的角度进行研究并提出用深度学习的方式处理医学图像分类任务的研究方向，虽大部分研究仍处于理论阶段且不具临床应用价值，尤其是肾积水超声图像场景下的模型内的诸多问题仍有待解决，但医疗 AI 仍然是 AI 发展至今相对成熟的领域之一。

2013 年，Lin 等人提出使用全局平均池化来降低过度拟合的风险[42]。基于此研究成果相继衍生出 VGGNet 和 GoogleNet 两类网络模型，前者提出了通过在已有网络架构的基础上选择性地补充和提高网络层进而提高深度即实现网络的效果提高的理论原理，简单却十分有效，因此迄今仍被很多的图像分类应用研究的实验部分选为基准模型以完成对比分析。后者由 Szegedy 等人[43]提出，其核心算法是采用并行的方式，组织 4 种不同尺寸的卷积层并且对各组织成分的运算结果进行通道上的联合，以达到融合不同尺度的图像信息的需求从而可以得到更好的特征信息。之后 He 等人提出的深度残差网络[44]（Deep residual network, ResNet）

的是基于 CNN 结果的基准网络模型的里程碑成果，其残差连接以消除网络梯度消失的思想迄今的大部分深度学习领域的新兴研究都有所引用，也渐渐成为了医学图像分类中主要使用的模型。直到 2020 年 Liu 等人[45]提出的 ConvNeXt 将 CNN 结构推进至新的高度，ConvNeXt 并没有特别复杂或者创新的结构，它是集成了各类已有研究成果下的优化的算法和网络结构，并达到了 ImageNet 中 Top-1 的准确率，也有望在医学图像领域内超越深度残差网络。

时至今天，以上所介绍图像分类算法在不同的时期下，均在医疗图像领域中提供研究意义并发挥着作用，在肾积水诊断应用下，临床医生需要借助超声图像来辅助判断与识别肾脏区域的病灶现象，并结合图像对病灶的轻重程度按照金标准 SFU 进行量化分级，因此自动识别肾积水超声图像中的病灶区域和正常组织器官是该场景下的医学图像分析的基本任务。在早期该疾病结合计算机技术的研究下，定向梯度直方图特征[46]或局部二值模式（LBP）特征[47]已经主导了图像分析领域中复杂的图像特征提取或手工判别特征的方法，但最近出现的深度学习算法开创了传统方式向非手工工程方式的转变，实现了图像分析的自动化，近两年的新冠疫情爆发期间，就有通过全玻片图像中的计算机辅助组织检测以及根据胸部超声图像诊断 COVID-19 肺炎[48]的成功案例。在一些医学图像数据分类的挑战赛中，所有排名靠前的团队都使用了 CNN 结构，在 CAMELYON17 挑战中，排名前十的解决方案几乎都使用了 CNN 对整个乳腺癌转移图像中的进行自动检测和分类，也证明了从深度学习提取的特征超过了 Shi 等人的手工方法[49]，其中评判深度学习模型算法的优越性以一系列评价指标为准，其中就有准确率，召回率等评估指标。

2.1.3 评估标准

医学图像分类同传统图像分类一致都以模式识别，的方式将不同类别的图像打上标识性的标签，如果研究过程需要统计并分析不同类别之间的分类错误的所有情况，并输出统计学评估结果，或者是检查是否存在某特定的类别之间出现相互混淆的情况，就需要根据预测结果统计混淆矩阵然后根据其输出分类实验的详细预测结果。对于三个及以上类别的多分类任务，混淆矩阵能够很清晰明确地反映出了各个类别之间分类错误的统计学概率，具体实施例是将实验的预测结果与真实结果按不同的类别统筹划分到了同一个既定的预测结果表里，具体如表 2-1 所示，其中我们可以清楚看到每个类别正确识别的数量和错误识别的数量,根据此表可以统计并计算出四个基本的评价指标。基于如表 2-1 所示的四个指标，对于医学图像分类的评价指标有：基于预测标签和真实标签的准确率指标(Accuracy),只针对正样本的真实标签和预测标签对比的精确率指标（Precision），

所有正样本中被正确识别的召回率指标（Recall）；除此之外还有进阶的基于召回率和精确率的综合指标 F1-score，以及混淆矩阵、ROC 曲线和 AUC，接下来对这些指标做出详细的介绍。

表 2-1 预测结果表

Tab. 2-1 Prediction Result

	P	N
T	正样本且分类为正（TP）	负样本但分类为正（TN）
F	正样本但分类为负（FP）	负样本且分类为负（FN）

以下介绍均结合表 2-1 中的统计结果进行阐述，在所有的数据样本集中，准确率的计算方式为：$\frac{TP+TN}{TP+FN+FP+TN}$，精确率的计算方式为:$\frac{TP}{TP+FP}$ ；正样本中被正确识别出来的概率即召回率的计算方式为：$\frac{TP}{TP+FN}$，其值越高意味着精确率越低。例如对于概率阈值 T 的影响，其中概率阈值 T 是一个设定的在[0,1]范围的内的浮点数值，通过 T 值分析算法模型预测的结果并划分到正类或负类，T 值降低会导致召回率升高以及分类精确度的下降，T 值增大会导致精确度升高以及召回率的降低，当样本类别的数量在三个及以上甚至更多，对于某项预测结果其最终的类别就是取预测概率值最大的类别。基于精确率和召回率综合分析得出的曲线是 PR 曲线，该曲线反映了精准率与召回率的关系，在通常情况下将横坐标设定为召回率以及纵坐标设为精确率，且每一条 PR 曲线要对应一个阈值，然后通过选择合适的 T 值对数据样本进行统计学上合理地统筹划分，对于概率大于 T 值的样本为正例，小于 T 的样本为负例，曲线逼近右上角则意味着该模型的表现性能越优异，除此之外还可以根据曲线和坐标轴包围的面积做微积分计算来完成进阶的定量评估方法。基于精确率和召回率的指标衍生出的评价指标同样还有 F1-score，其计算方式为：$\frac{2\times Precision\times Recall}{Precision+Recall}$，在众多模型对数据进行学习后，当某一个神经网络模型的 PR 曲线与另一个模型的 PR 曲线呈包含关系的时候，则可根据这种特点侧面证实包含模型的表现性能具有优越性，除此之外假如两个 PR 曲线存在交点甚至是多出交叉，则需要根据曲线下方的面积计算面积并根据大小进行比较，但相比 PR 曲线 F1-score 的结果更加有效，其值相对大就可以认证该模型的性能更好。

以上指标都是针对医学图像分类模型在既定参数下的单一数值指标，如果研究过程需要想观测构造的分类模型在不同的参数下的表现性能的话，则需要用到 ROC 曲线，其能够评估医学图像分类模型在不同阈值下的性能表现。对于 ROC 曲线，其横坐标统计假阳率（False Positive Rate，FPR），其计算方式为：$\frac{FP}{FP+TN}$，

纵坐标统计真阳率（True Positive Rate），其计算方式为：$\frac{TP}{TP+FN}$。ROC 曲线上的点与某个阈值唯一对应，最大阈值与原点对应，最小阈值与对应于坐标（1，1）的点，随着阈值不断地增大，TP 和 FP 都相应地降低，从而导致 ROC 曲线上的点向曲线图的左下方移动。同样的，ROC 曲线也可以用来评判多个医学图像分类模型的性能，即当曲线呈包含关系的时候则从统计学上证明了包含曲线所属模型的性能具备优越性。

因为医疗图像语义分割实质上是像素级别的分类任务，因此对于医疗图像语义分割的评估标准，像素级的微观程度上的图像分类的评估指标同样适用。除此之外，统筹针对每个像素点评估指标还有平均像素准确性（Mean Pixel Accuracy，MPA），该指标通过计算模型预测结果的类别正确的数量占总预测像素数量的比例，然后再统计其类别的所有预测结果中真实属于该类别的像素准的准确率，即类别像素准确率（Class Pixel Accuracy CPA），由于医疗图像的二维特性，对于其二维图像内的所有像素点，其计算方式为：$\frac{TP}{FN}$，最后分别计算每个类别的 CPA 后进行累加求和并计算平均结果，计算方式为：$\frac{\sum_i^N CPA_i}{N}$。同理，针对语义分割的标注数据的区域性原理，平均交并比（Mean Intersection over Union，MIoU）也可作为语义分割系统性能的评价指标，其计算的是所有类别的交并比的均值结果，其中交并比的计算方式为$\frac{TP}{FN+TP+FN}$。这种评价标准相比 MPA 而言，直观上如果某一类物体体积较小，在图像上所占的像素比例就小，如果这类物体预测全对或者全错，即交并比为 0 或者 1，对最终的平均交并比的结果影响就较大，但是 MPA 是统计的每一个像素的预测准确率，这种极端情况的影响就较小。因此一般结合多种评价指标综合分析分割模型的有效性。

以上是医疗图像处理任务结合深度学习相关的实验与结论的评估标准，分别针对深度学习下的语义分割分支和图像分类分支的评估参数的计算方式、原理以及意义做了详细的阐述，接下来围绕基于肾脏超声图像的智能诊断方法的技术框架做一个详细的背景知识的介绍。

2.2 背景知识

2.2.1 深度学习

针对肾积水超声图像的分析应用场景，深度学习的技术能够通过多个计算处理层组成的深度神经网络模型实现学习肾积水超声图像的多维特征表示。其通用的深度学习方法在很多领域极大地改善了其作业效率，例如语音信号的识别、视觉对象的识别与检测以及若干领域内的最新技术。深度学习技术通过反向传播算

法实现对计算机以一组指令或指示来更改模型内部的可学习的参数，这些可学习的参数用于从前一层的表示中更改用于计算每层中的表示，从而发现大型数据集中的复杂结构[50]。而深度学习的研究起源与认知心理学、计算神经科学领域，时间倒退到 1943 年，该年 Warren McCulloch 和 Walter Pitts 研究发明了基于数学和阈值算法的逻辑严谨的神经网络计算模型，也称为 MCP 神经元数学模型并且为之后的神经网络的发展奠定了基础。一直发展到 1958 年由 Rosenblatt 创造了感知机并将其成功运用于简单的模式识别任务中，将神经网络的研究推向了第一次的高潮。一些学者对神经网络的发展过度乐观，甚至有研究结论表明 20 年内机器可以替代完成所有的人工作业。再然后直到 1969 年 Minsky 和 Papert 直接指出目前的神经网络存在两个关键的缺陷：a、无法解决异或这类简单的线性不可分问题；b、计算机没有足够的计算力来求解大型的网络。这些问题在当时直接决定了神经网络的不可行性，神经网络的研究进入了第一次寒冬，在这期间也有部分学者取得了一定的成就，例如 1959 年 Hubel 与 Wiesel 通过观察研究猫的视觉皮层发现，哺乳动物的视网膜上同时存在着两种细胞即视锥细胞和视杆细胞，其分别对颜色和明暗敏感。视网膜接收光信号后将其转化为电信号并分两路向视觉皮层传播。视神经传来的信号主要经过初级视觉皮层（V1）、二级视觉皮层（V2）、三级视觉皮层（V3）等层层处理抽象，最终在我们大脑中形成影像。该研究于 1981 年获得诺贝尔生理学或医学奖，并直接启发了日本学者 Fukushima。1980 年 Fukushima 根据猫的视觉皮层中视觉感受野以及视神经信号层层传播处理的思想发明了感知机。该网络可以视为 CNN 最早期的雏形，但当时该网络的训练为自组织的方式，其并未采用误差反向传播，当数字的位置或形态稍有变化时其识别效果并不理想，然而这一工作却为日后 LeCun 研究的重要基石。

加州理工学院的物理学家 Hopfield 于 1983 年利用神经网络并通过电路模拟仿真的方法求解了旅行商 NP 难题，此举在学术界引起的较大的轰动，这也推动了神经网络二次快速发展。直至 1986 年 Hinton 等人发明了 BP 误差反向传播算法并为后来神经网络的发展奠定了基础，随后的 George 首次提出了 Sigmoid 激活函数的万能逼近定理，2 年后 Hornik 指出万能逼近定理并不依赖于特别的激活函数，而是由多层前馈网络结构所决定的，比如针对任意的布尔函数我们只需使用一层感知层来拟合该函数，同时对于任意的泊松分布下的分类问题我们也只需使用一层的感知层，增加感知层的层数后每一层神经元的数目将会减少同样意味着宽度减少，对于网络模型的构造一般倾向于增加深度而非宽度，主要原因是深度的增加对精度的提高更加有效。之后的图灵奖获得者 LeCun 正式提出第五代标准的 CNN 网络 LeNet，其首次引用了卷积与池化的网络结构，并结合双曲正切激活函数和极大似然估计损失函数与 BP 反向传播算法进行模型的训练。该

网络在手写数字识别上获得成功，并应用于美国的邮政系统中。然而受限于当时计算机的发展、数据的匮乏以及网络优化的困难包括梯度消失问题，神经网络的发展再次受挫。

2006 年图灵奖获得者 Hinton 在 Science 上发文指出利用 RBM 编码的预训练深度神经网络与 PCA 相比在高维特征抽取方面有更佳的性能，即深度网络拥有强大的特征提取能力。同时还指出深层的网络训练可以通过逐层训练的方式实现，这也使得更深网络的训练成为了可能。虽然这篇文章现在看来并没有在理论上做出较大的创新，尤其是逐层训练的方式早已被弃用，但是该文章却使得深度学习重新回到学者们的视野。与此同时随着计算机技术的进步以及互联网的普及，其为深度学习技术提供了井喷式发展的催化剂。真正使深度学习得到广泛关注的是在 2012 年 ImageNet 比赛中，当时 Hinton 队伍凭借 AlexNet 深度神经网络以领先第二名 10.8 个百分点的优势一举多得比赛冠军，此后便进入了深度神经网络的时代。深度学习尚处于起步阶段，但它们在分类任务上已经优于被动视觉系统，且在医学图像处理领域已有相当多的研究应用。

2.2.2 注意力机制

注意力机制（Attention Mechanism）是研究者归纳人体眼睛的注意力聚焦特性跃迁到深度学习网络模型中的一种特殊的具备可学习参数的网络层，用来自动化地计算并学习输入的数据对输出结果的贡献权重大小。本小节从人体的视觉直觉、文本翻译常识以及特征工程等多种角度对注意力机制的意义和原理进行了比较详细的阐述。注意力机制的起源是由上个世纪末的生命科学以及神经类相关的科学家在研究人类视觉的实验过程中偶然发现的一种光信号处理的特殊机制。此后 AI 领域的科研工作者们在了解这种机制后以仿生学的模式将这种特殊机制补充到深度学习神经网络模型里并取得了一定的成果。迄今为止注意力机制已经成为深度学习领域，尤其是自然语言处理领域中应用场景最为广泛的网络模型中间件之一。在翻译任务中常用的 Transformer 以及其衍生模型 GPT 等自然语言模型中都采用了这种基于人体仿生学的注意力机制。科研工作者在深度学习模型在文本数据下的情感分析任务上通常都会补充特征工程的相关研究，即将初始的文本数据通过转换算法输出向量数据，其中注意力机制就是在数据科学与工程领域里的常用工具之一，它可以帮助神经网络模型聚焦高效且规模化的有用特征，进而促使网络模型高效地完成既定的识别任务。例如通过逐步回归分析方法对原始数据的相关特征集进行严谨的筛选并得到一个价值更高的特征子数据集，从而可以让下游模型聚焦于和任务关系更为紧密相关的信号或者数据。有学者从理论的角度试图使网络模型通过自学习来掌握如何分配自身参数的注意力也就是参数权

重，即为输入数据的加权。他们用注意力机制的直接目的是为输入的各个维度赋予重要程度的分数，然后按分数对特征加以权重值以突出重要特征对下游网络模型或者特征模块的影响程度，以上便是注意力机制的基本思维。在深度学习领域，模型往往需要接收和处理大量的数据，然而在特定的某个时刻，往往只有少部分的某些数据是重要的，这种情况就非常适合注意力机制发光发热。在深度学习的多国文本数据翻译任务中，以中译英为例，机器翻译是将一串中文语句翻译为对应的英文语句，其中主要的功能模块即编码器、注意力模块和解码器，其中编码器用于将中文语句进行编码，这些编码后续将提供给解码，解码器再根据编码器的输入数据进行解码操作。然后在中间过程中将某个时间点的前一个时刻的状态和编码器输出进行注意力机制的计算，得到一个当前时刻的翻译内容，其中注意力算法会有个打分函数，其大小描述了当前的时刻在编码器的结果上的关注程度，在此基础上的下一个时刻会更加关注源中文语句中的相应内容。而针对图像数据，尤其是医疗图像中的超声图像数据，对于由像素点构成的二维的平面图像来讲，其注意力机制的表现形式即抽象为像素分布簇的权重分布，通过生成一层与输入数据同样尺寸的注意力权重层，来影响网络模型中对输入数据的参数化的学习过程进而促使网络模型从仿生的角度更加关注其构造者所期望的图像区域。

注意力机制的研究成果也存在分支，大致为分硬性注意力机制、键值对注意力机制、多头注意力机制以及自注意力机制。其中软注意力机制是通过统计学分布来加权求和并融合各个向量化的输入数据，而硬性注意力机制则不是采用这种方式，它是根据注意力的实际分布中选择向量化输入中的某单个向量组作为输出，这里有两种选择方式：选择注意力分布中，分数最大的那一项对应的输入向量作为注意力机制的输出。根据注意力分布进行随机采样，采样结果作为注意力机制的输出。这两种方式会使得最终的损失函数与注意力分布之间的函数关系不可导，导致无法使用反向传播算法训练模型，硬性注意力通常需要使用强化学习来进行训练。因此，一般深度学习算法会使用软性注意力的方式进行计算；另外是键值对注意力机制，该种机制对输出的要求是更为一般的键值对（K,V），并需要相关的查询向量 Q，在这种模式下，使用查询向量 Q 和相应的键值 K，进行计算注意力权值 A，当计算出在输入数据上的注意力分布后，通过注意力和 V 进行加权融合计算出结果，当 K 与 V 相等时，该注意力方式就退化成了普通的经典注意力机制；还有就是多头注意力机制，其利用了多个查询向量，并行地输入键值对信息（K,V），在查询过程中，每个查询向量将会关注输入信息的不同部分，即从不同角度上去分析当前的输入信息，最终向所有查询向量的结果进行拼接作为最终的结果；最后在自注意力机制中，其查询向量可以通过输入信息生成，而不用选择上一个任务相关的查询向量，即当模型读到输入信息后，根据输

入信息本身决定当前最重要的信息。

2.2.3 频域图像

传统的图像处理方式是通过对计算机编程实现对数字图像的噪声去除、原画增强、旧画复原、图像分割以及提取特征等处理操作，大部分技术随着 AI 的发展已被更具优越性的计算机视觉技术所取代。通常情况下对数字图像进行处理、加工与分析的目的与动机分三个方面：

（1）从人体视觉感官上提高图像的质量，例如增加图像的亮度值、模糊增强或通过算法抑制图像中指定的某种成分，以及通过翻转平移等操作对图像进行形态学上的几何变换的图像处理操作并以此提升图像的视感质量。

（2）提取图像中所富含的为某种既定任务下完成计算机的分析提供数据的相关特征或特殊像素群信息。通常作用于计算机视觉的图像预处理操作，其提取的特征的维度也很多样化，例如频域特征、值域的灰度特征或色彩特征、细节纹理特征、对象的几何边界特征、空间区域特征、以及各对象间的拓扑特征和结构等等。

（3）对信息量较大的图像数据进行编码实现信息量的压缩变化，便于终端的存储和通信传输。

对于医疗影像中的超声图像数据，可以通过傅里叶变换、小波变换等空域-频域转换算法，可以将图像特征的分析域从空间特性转换到频率特性，图像的频率是表征图像中灰度变化剧烈程度的指标，是灰度在平面空间上的梯度，传统的对于图像上频域的处理主要应用在降噪上，通过分析图像的频域信息，结合高通滤波器或者低通滤波器以达到图像滤波的目的。对于超声图像，主要的目的便是提取超声图中所包含的特征与特殊信息，以帮助神经网络学习到更多维度的特征知识。深度学习模型中，主要还是通过卷积学习低级特征，比如边缘、纹理与幅值，但对于模糊性强、噪声高的超声图像，仅仅是低级特征不足以支撑模型最后的拟合任务与分类效果。需要让神经网络学习到频域内的高级特征，在超声图像处理的应用场景下，频率域的特征也包含了丰富的信息量，具体可在后续章节的实验中可视化，因此在某些特定的应用场景下，深度学习的特征维度也需要拓展到频率域。

2.3.4 优化算法

本文在之前的医疗图像分类的相关知识中提到了一种基于 CNN 结构的神经网络 ConvNeXt，其相对于 ResNet 主要是以网络结构和优化算法上的宏观调整，

达到了 ImageNet 的 Top-1 准确率。下面将依次介绍常见的优化算法中的归一化算法、正则化算法以及标准化算法。

在深度学习的深度卷积神经网络中，其网络中的某一中间层的输入是其前一个神经层的输出，所以前层神经网络的输入参数的改变会引发下一层网络中输入的数据分布产生剧烈的差异化，在通过随机梯度下降算法在网络层之间不断更新模型参数的时候，每一次网络层参数的更新迭代都会致使之后神经网络模型隐藏层的输入的数据分布产生变化，并随着网络层数的加深而变化得更加显著，这种训练过程中的变化现象被称作为内部协变量偏移(Internal Covariate Shift)。为了解决该训练过程中的问题，就需要每层神经网络的输入的数据分布在训练的迭代过程中保持一致，从而促使深度学习中特有的几个归一化算法的研究产出，最先得到广泛使用的是批处理归一化（BN）算法。我们定义网络中第 l 层的净输入为 i^l，神经元的净输出为 o^l，结合参数w, b的传播函数如下，

$$i^l = f(w \cdot i^{l-1} + b) \tag{2-2}$$

为了减少内部协变量偏移问题，就要使得净输入 i^l 的分布一致，根据当前参数下，对于每一维 i^l在整个数据中的期望$E(i^l)$和方差$Var(i^l)$，将输入数据的概率分布都归一化为标准正态分布 $\tilde{i}^l = \frac{i^l + E(i^l)}{\sqrt{Var(i^l) + \varepsilon}}$，这种方法将输入数据的值集中在零点附近，并在数据通过激活函数构造的网络激活层时，使数据刚好接近线性变换的区间从而削弱了神经网络的非线性特性，进而更好地实现模型的凸优化过程。除此之外，为避免归一化计算对网络的拟合表示能力产生消极作用，BN 算法通过一个附加的缩放参数 γ 和平移变换参数 β 以改变数据的取值区间，其中 $\gamma = \sqrt{\sigma_l^2}$，$\beta = \mu_l$。BN 层可以看作是一个特殊的神经网络层，补充在每一层非线性激活函数网络层之前，即在公式 2-2 的基础上，净输入处理输出为

$$i^l = f\left(BN_{\gamma,\beta}(w \cdot i^{l-1})\right) \tag{2-3}$$

其中用整个数据集上的均值和方差来别代替每次小批量样本的均值和方差，这种方法可以简化并优化使得非常深的网络能够收敛。但是 BN 却很受批量数据的超参数大小的影响，当 BN 需要一个足够大的批量时，小的批量大小会导致对该批统计数据的准确率降低。对于图像分割这种显存占用较大的任务，Group Normalization (GN)相比 BN 更优[51]，GN 计算均值和标准差时，把每一个样本数据的特征的通道分量 C 分成 G 组，每组将有$\frac{C}{G}$组通道分量，然后将这些通道中的元素求均值和标准差，各组通道用其对应的归一化参数结合公式 2-3 独立地计算归一化结果。

然后是优化算法中的正则化参数，相比机器学习中的复杂的正则化算法，深度学习中的常用正则化手段为 DropOut，该方法下的所有模型共享参数，每个模

型继承父神经网络参数的不同子集。在单个步骤中，训练小部分的子网络的时候，参数共享会使得其余的子网络也能有好的参数设定，在前向传播到指定层时，层中每个单元乘以相应的掩码，从而决定是否被去掉这个单元，然后继续向前传播，后续更新参数等，然后更新完一次参数，恢复所有未更新单元，之后重复这个过程。这种算法只有在极少的训练样本时才可能失效。但是这种算法将每个隐藏的激活层乘以一个独立的伯努利随机变量，而当与批处理归一化结合使用时，DropOut 会失去有效性[52]，现有研究表明更适合与 BN 以及 GN 归一化算法结合使用的正则化算法为 DropPath[53]， 相比前者一个是作用于特征层，一个是作用于网络分支，都是通过对网络结构的选择来达到局部层结构失效的目的来达到正则化效果，而 DropPath 不受批处理操作的影响，只针对网络分支处理并弥补了 DropOut 的缺点。

2.4 本章小结

本章第一节结合本文研究所提出的分割-分类框架，对其深度学习在医疗图像领域的发展与应用做了一个概括性的简要概述，包括医学图像分割和分类的发展历程，对前人的研究做了综述性归纳与总结。第二部分介绍了后续章节关于方法论描述所需要的相关背景知识，以帮助读者更好地了解后续方法的理论内容，其中包括深度学习理论、注意力机制原理、图像的频域处理的意义以及神经网络中的优化算法。

第 3 章 基于肾积水超声图像的语义分割模型

3.1 模型提出的动机

之前章节有提到过对于儿童先天性 UPJO 的病理检测，在诊断阶段以超声为主要手段，需要医生根据超声图像做出科学权威的分析，给出其对应的病理情况，决定手术与否与手术时机，然后国内面临医疗资源紧缺的情况，医学影像科的权威人才更是少之又少，在这一阶段引入大数据与人工智能技术下的计算机视觉以帮助，甚至在大部分医师短缺的情况下替代其工作职能，能够大大节省医疗资源与时间成本，帮助实现医疗工作的降本增效。其中的挑战就有超声图像本身的一些难点痛点，如图 3-1 所示，每一帧关于肾脏的超声图像会包括四类区域：病变区域（a）、器官区域（b）、其他组织区域（c）、非相关超声背景区域（d）。不难发现，对于疾病诊断最重要的 a、b、c 区域之间的边界存在模糊性，且图像噪声点多且密集，传统的图像处理方法或深度学习分类方法缺乏足够的相关信息、以及边缘信息，从而降低最终分类结果的准确度。基于以上问题，本研究提出一种先将肾脏超声图像通过语义分割算法，然后融合原图以增强边缘等信息，然后再通过分类算法输出诊断结果的智能诊断方法。

首先介绍诊断方法中图像语义分割阶段的深度学习网络（Attention & Pyramid Pooling Network，APPNet）模型，经过实验与分析，本研究决定先通过专业的且经验丰富的医学影像科医生对超声图像数据集做人为的边界标注，然后基于这些标注的数据集训练一个语义分割神经网络，得到分割结果后，通过数据清洗与图像处理算法创建新的分类网络数据集，再完成之后的分类工作。针对如同 3-1 的四个类别的区域，在专业医师的帮助下，制作语义分割数据集，下一步就是选择什么样的语义分割算法，上一章节提到过各类经典的语义分割算法即从图像处理到机器学习到深度学习的一个发展过程，也列举了各算法在超声图像应用领域内的缺点。其实仔细观察图 3-1 还可以发现，随着病理严重程度增加，器官区域 b 内的病变区域 a 会越来越饱满，整张图像的其他无关区域比如 c、d 区域也会造成干扰，边缘信息更加难于识别，边界上下文更难以区分，需要分割算法具备一种注意力机制使得模型关注局部区域，同时还需要提供一种额外的算法分支能够让模型学习到多尺度的上下文信息，即学习不同尺度下、各感受野范围内不同类别之间关系的信息。因此，在基础特征提取阶段，融合注意力模块到特征提取网络中去，生成注意力权重图并使得在基础特征提取阶段的神经网络聚焦标注的病变、器官区域以及边缘信息；然后在高级特征处理阶段以金字塔场景解析方式做空间金字塔池化处理实现多尺度上下文的信息解析。

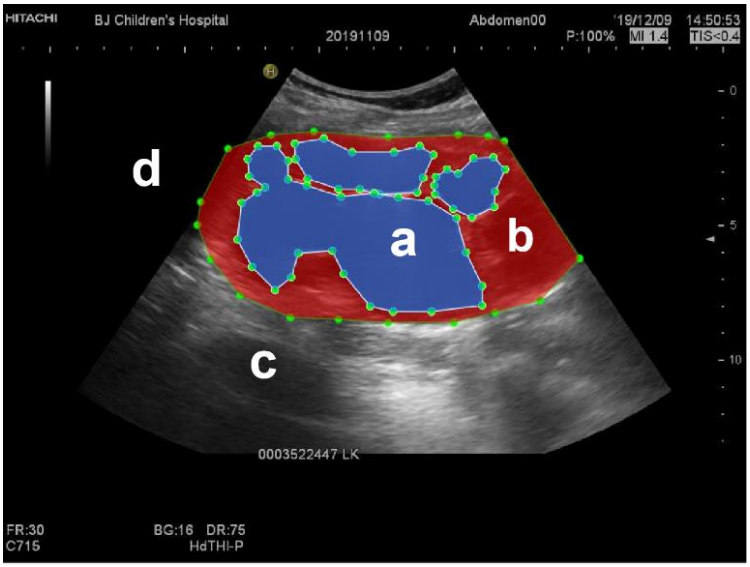

图 3-1 图示为超声图像视频中样本帧，并用标注工具完成注释后的样例，成分包括 a、b、c、d 四类区域

Fig.3-1 The picture shows the sample frame in the ultrasound image video, and the sample is annotated with the annotation tool, and the components include four types of regions a, b, c, and d

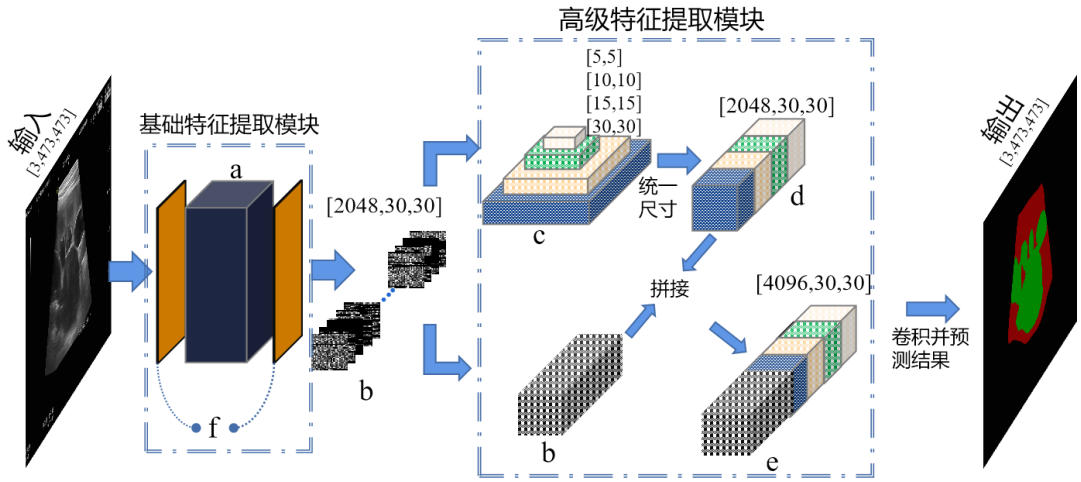

图 3-2 APPNet 的整体架构，输入为原始超声图像，输出为红绿黑颜色区分的分割结果。a. 该模块为 ResNet32 网络并在前后端嵌入两个 NAM 模块，即 f；b. 基础特征阶段所提取的特征结果；c. 金字塔池化处理的高级特征；d. 对 c 进行统一维度运算后的结果；e. d 和 b 集成后的最终特征图，然后经上采样并预测之后输出结果

Fig. 3-2 The figure shows the overall architecture of APPNet, the input is the original ultrasound image, and the output is the segmentation result of red, green and black colors. a. The module is a ResNet32 network and embeds two NAM attention modules at the front and back ends, namely f; b. The feature result extracted in the basic feature stage; c. The advanced feature of the pyramid pooling process; d. The result of the unified dimension operation on c; e. The final feature map after the integration of d and b, which is up-sampled and predicted, then output the result

因此对于肾脏超声图像分割，本研究提出了一个融合注意力机制和金字塔池化模块的新型金字塔场景解析网络 APPNet，它能满足既定应用场景下、专注于

关键部分，聚合上下文信息的能力，并很好地完成语义分割任务。其核心是一个融合了注意力模块的基本的特征提取网络和高级特征处理阶段的金字塔池化模块。整体结构如图 3-2 所示，详细解释如下跟随整体网络下面本文将详细地从数据输入到输出即端到端的的方式介绍该网络的各个模块。

3.2 基础特征提取模块

对于语义分割任务，由于其分类级别是基于像素的，低级别的图像信息往往不够，例如传统图像处理中，为了得到某些信息（例如边缘信息），需要对图像进行滤波处理，滤除不是边缘的内容，从而得到边缘特征。例如还进一步地需要纹理信息，通过卷积神经网络将这些低级特征提取出来。而像素级分类是需要高级特征的，即假设有两张图片分别为猫和狗，只用上述低级的纹理、边缘特征是无法区分其中的耳朵、眼睛、鼻子等特征区域是否属于猫类别还是狗类别，因此需要通过环境信息等高级特征去结合低级特征并给出一个综合分析的结果。

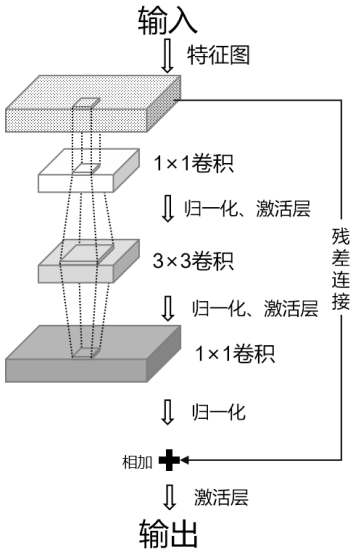

图 3-3 ResNet 的基本构成单位，主要由卷积、归一化、激活层和残差连接组成

Fig. 3-3 The basic constituent unit of ResNet, mainly composed of convolution, normalization, activation layer and residual connection

目前人工智能在欠缺可解释性的大背景下，由实验验证网络的深度对模型的性能至关重要，当增加网络层数后，网络可以进行更加复杂的特征模式的提取，所以当模型更深时理论上可以取得更好的结果，网络深度增加时，网络准确度出现饱和。在本研究的基础特征提取阶段，医学超声图像受本身成像原理的限制，致使其图像的分辨率不高而且噪声污染严重，在深度网络的传播过程中随着维度的降低，饱和率增大、易出现过拟合、网络退化等问题，而 ResNet 可以在保留

深度网络均值的同时避免网络退化的问题。关于 ResNet 的结构，其基本构成单元为瓶颈模块如图 3-3 所示,它还具有低复杂性，相比传统的基础特征提取网络具有更少的参数量，这意味着处理速度更快。随着深度网络的增加而产生的梯度退化问题意味着深层的神经网络难以训练，这里提出一个假设检验，对于一个层数不多的浅层网络，我们试图通过线性堆叠新的网络层来构造一个深层的神经网络，极端糟糕的情况下是所有累加的网络层无法学习到任何有用的参数而仅仅是对浅层网络的特征的简单复制，对于这种情况下新增的网络层相等于完成$(y = x)$的恒等映射。在以上假设成立的情况下，理论上堆叠而成的深层网络的表现性能相比于浅层神经网络不应该出现网络性能退化的现象。然而不断的实验事实证明发现了在训练的过程中仍然发生了梯度消失而导致网络退化的问题。ResNet 针对该问题提出了用残差连接来解决网络退化的问题，图 3-3 所示，对于由该基本构成单位组成的网络模型，当输入为 x 时，将传输过程中可学习的特征记为$H(x)$, 而 ResNet 可以满足该网络完成残差$F(x) = H(x) - x$的参数化学习，从而原始的学习特征为$F(\mathcal{X}) + \mathcal{X}$。这种创新的连接方式的是残差学习相比之前假设的深层神经网络更容易在避免退化的前提下完成特征学习，接下来从数学的角度分析为何要针对深层网络模型补充残差学习，首先将残差单元定义为:

$$y_n = h(x_n) + F(x_n, W_n)，\ x_{n+1} = f(y_n) \tag{3-1}$$

其中x_n和x_{n+1}分别表示的是第 n 个残差单元和输入和输出，F为残差函数即学习到的残差，$h(x_n) = x_l$表示恒等映射，f表示激活函数，基于上式可得到从浅层神经网络 n 到深层神经网络 N 的学习特征为：

$$x_N = x_n + \sum_{i=n}^{N-1} F(x_i, W_i) \tag{3-2}$$

通过链式求导法则，定义损失函数到达深层神经网络 N时的梯度为$\frac{\partial loss}{\partial x_N}$，结合公式 3-2 计算偏导，那么反向传播过程中的梯度为：

$$\frac{\partial loss}{\partial x_n} = \frac{\partial loss}{\partial x_N} \cdot \frac{\partial x_N}{\partial x_n} = \frac{\partial loss}{\partial x_N} \cdot \left(\frac{\partial \sum_{i=n}^{N-1} F\left(F(x_i, W_i)\right)}{\partial x_n} + 1 \right) \tag{3-3}$$

由公式 3-2 可发现残差连接的短路机制可以避免损失的情况下完成梯度的传播，即括号内的 1，即使主干的残差梯度即$\frac{\partial \sum_{i=n}^{N-1} F(F(x_i, W_i))}{\partial x_n}$比较小，有 1 的存在也不会导致梯度消失。对于超声图像，由于分辨率有限，我们在任何时候都会接收到分辨率单元内的大量分布的散射体的散射信号，这些散射信号相干相加，或者说它们根据每个散射波形的相对相位进行相加叠加和相消叠加。图像中就出现亮点和暗点不规则相间分布的信号。这样的散斑噪声带来的负面影响就是在局部区域随着网络深度的不断加大，梯度消失的概率会逐渐增加，因此采用 ResNet 网

络作为基础特征提取模块，利用其残差连接的优点弥补超声图像自带的噪声所带来的缺陷。本研究在其基础上融合了注意力机制，另外考虑到在本研究中肾脏超声图像的分割任务中，需要更多地识别一些不显著的特征，通过使用基于规范化的注意力模块(NAM),它抑制了较少显著性的权值，对注意力模块应用一个权重稀疏惩罚，相比其它注意力机制缺乏对权重的影响因素的考虑并抑制了不显著的像素，它通过利用预训练模型权重的方差度量来突出显著特征，其整体结构如图 3-3 所示。

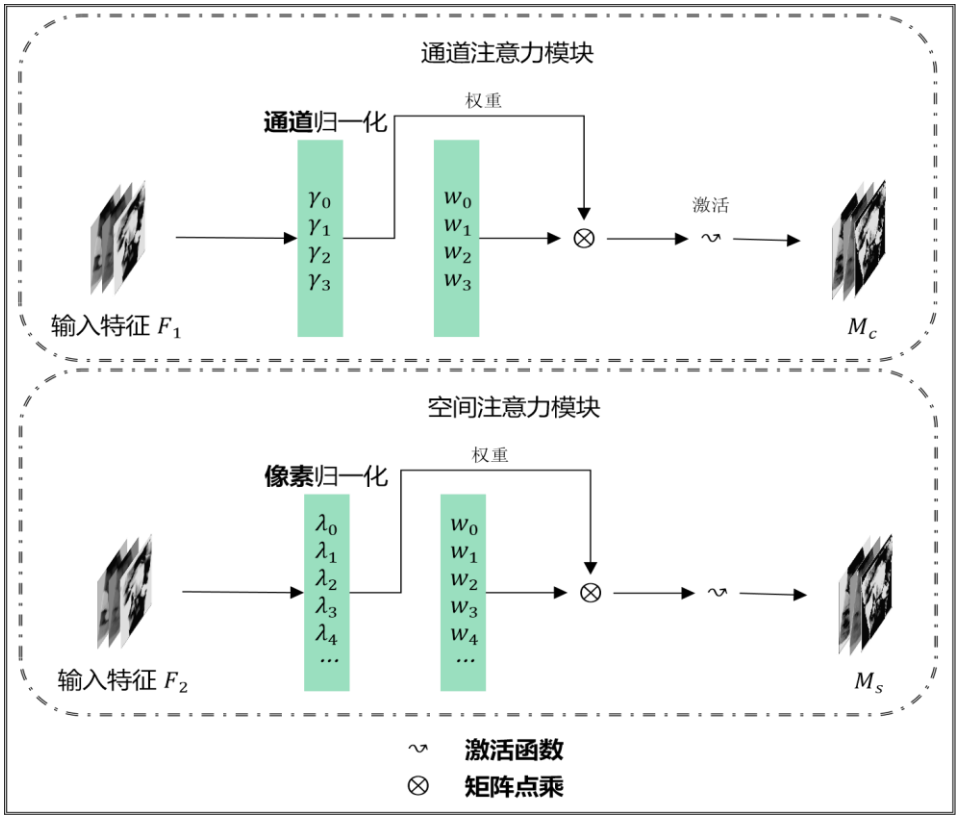

图 3-4 NAM 的结构图，分别有通道注意力与空间注意力两种模式

Fig.3-4 The structural diagram of NAM, which has two modes: channel attention and spatial attention

对于通道注意力模块，以F_1作为输入特征、M_c为输出特征，权重值W_γ等于$\frac{\lambda}{\sum_{j=0}\gamma_j}$使用批归一化（BN）中的缩放因子，如公式 3-4，

$$A_{out} = \gamma \frac{A_{in} - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \qquad (3-4)$$

其中μ为均值，σ为标准差，其平方根表方差，γ和β表示可训练的仿射变换参数，对应尺度和位移的变换，利用 BN 的缩放因子反映出所有通道的各自变化的幅度，也同样代表了当前通道的重要程度指标，缩放因子即 BN 中的方差，且值越大表示该通道变化得越快，那么该通道中包含的信息会越丰富，重要性也越大，

而那些变化不大的通道信息单一且重要性小，通过这种方式对各个通道的重要性给出分值作为注意力层的输出M_c。而对于空间注意力模块，以F_2作为输入特征、M_s为输出特征，权重值W_λ等于$\frac{\lambda}{\sum_{j=0}\lambda_j}$,将 BN 的比例因子应用于空间维度来衡量像素的重要性，并称之为像素归一化，即给各个像素的重要性给出分值并作为注意力层的输出M_s，具体计算方式分别如下：

$$M_c = ReLU\left(W_\gamma\left(BN(F_1)\right)\right); M_s = ReLU\left(W_\lambda\left(BN(F_2)\right)\right) \quad (3-5)$$

其中 $ReLU$ 表示激活函数。对于以上描述的注意力模块，其与 ResNet 的具体融合方式不同于传统的方式，即在每个卷积层之后加入注意力层，因为这种方式有两个缺点，一是会过高地提高网络的参数量，二是会破坏原有网络的结构而不利于预训练参数的初始化，于是研究并思考了一种新的在不破坏 ResNet 原有结构的情况下添加 NAM 的巧妙方法，即在主干网络的整体输入输出两端分别前后加入通道和空间注意力单元，由于 NAM 是一个轻量级的通用模块，这种嵌入方式几乎无需任何额外的计算开销，经实验验证两种嵌入方式的最终的准确率、召回率等指标几乎一致，为了更好地介绍 ResNet 与注意力层的结合原理，对于给定的特征图$\mathrm{F} \in \mathcal{R}^{\mathrm{C}\times\mathrm{H}\times\mathrm{W}}$作为输入，对于通道注意单元$\mathrm{M_c} \in \mathcal{R}^{\mathrm{C}\times 1\times 1}$和空间注意力单元$\mathrm{M_s} \in \mathcal{R}^{1\times\mathrm{H}\times\mathrm{W}}$，其整体的注意力特征处理过程可以概括为：

$$M(F_{in}) = \{M_c; M_s\} \odot F_{in} \quad (3-6)$$

其中⊙表示逐像素乘法，分号表示串联。同样的，对于输入特征$F \in \mathcal{R}^{C\times H\times W}$，将 ResNet 主干网络的特征处理过程定义为$\mathrm{Y(F_{in})}$，结合公式 3-5，其最终的基础特征提取模块处理过程可以概括为：

$$F_{out} = M_{end}\left(Y\left(M_{start}(F_{in})\right)\right) \quad (3-7)$$

其中$\mathrm{M_{end}}$和$\mathrm{M_{start}}$分别表示 NAM 在 ResNet 中的嵌入顺序，并且都由通道注意力单元和空间注意力单元串联组成。以上即基础特征提取模块，其作用是初步提取低级的超声图像的特征如纹理、边缘等特征信息，其主干网络为 ResNet32，以残差连接的方式消除了网络退化的问题，并其基础上嵌入了 NAM 注意力层，赋予主干网络在特征提取阶段更多地关注不够显著的特征的能力，并将提取的初步特征为下一步高级特征提取模块做输入准备。

3.3 高级特征提取模块

上一节提到过对于肾积水超声图像的语义分割这样的像素分类任务，低级的纹理等特征信息是不够的，还需要知道像素间的关系等高级特征信息，以此来提高像素分类的准确性。而完成这一目标，需要通过高级特征信息解决数据中的三个难点，一是环境下的不匹配导致的错误识别；二是难以识别的不明显的类别，

例如超过模型感受野极限的物体或者体积小以至于难以区分识别的物体；三是不同对象之间相近的类别难以统一而混淆。如图 3-5 所示，图中的 a 所表示的对比图即从人体正面超声检测和侧面超声检查的不同肾脏图像，同样是肾脏器官需要全局环境信息去区分；图中 b 所表示的对比图表示积水情况严重的和轻微的病灶超声图像，同样的语义但是左边红线区域的病变特征与右边的区分度太低而导致网络难以识别；图中 c 所展示的即相同的类别易混淆、难统一的问题，积水区从灰度值上与最终任务无关的蓝线背景区域相似，但是由红线构成的病变区域一定是包含在绿线肾脏器官区域内，若需要对其从计算机视觉的视角去区分的话，同样也需要利用不同的感受野结合上下文信息去学习其易混淆的相关特征。

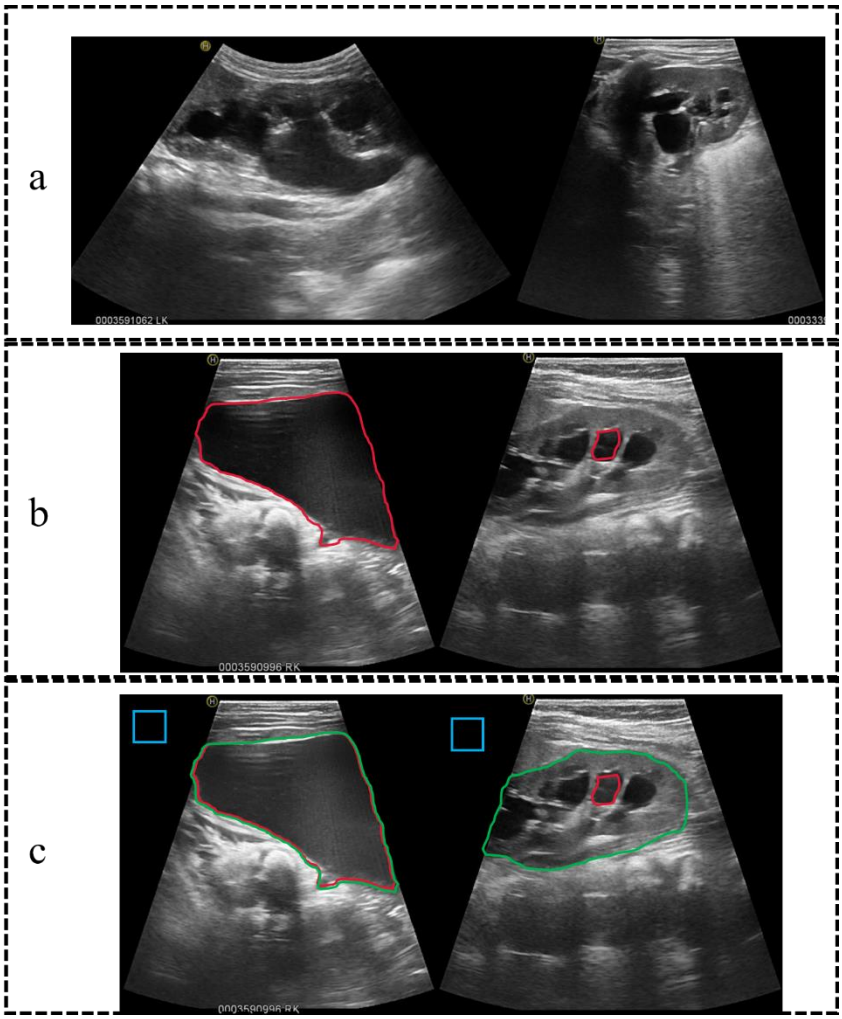

图 3-5 图示为肾积水超声图像语义分割中的难点样例图，a. 不同扫描角度下的肾脏器官；b. 小尺度积水区和大尺度积水区；c. 积水区与器官的空间分布特点

Fig. 3-5 The picture shows an example of difficulties in the semantic segmentation of hydronephrosis ultrasound images, a. Kidney organs under different scanning angles; b. Small-scale hydronephrosis areas and large-scale hydronephrosis areas; c. Spatial distribution of hydronephrosis areas and organs features

综合以上所描述的语义分割中的问题，一个能够提取全局信息的深度神经网络模型可以极大提高复杂场景解析的能力从而规避以上问题。学术界内将感受野定义为使用上下文信息的大小，通过不同大小的感受野引入更多的上下文信息进行解决，当分割层有更多全局信息时，出现误分割的概率就会低一些。因此本研究在基础特征提取结束过后，对特征图进一步处理，并使用金字塔池化结构作为高级特征处理模块的主干结构，具体过程如图 3-6 所示。

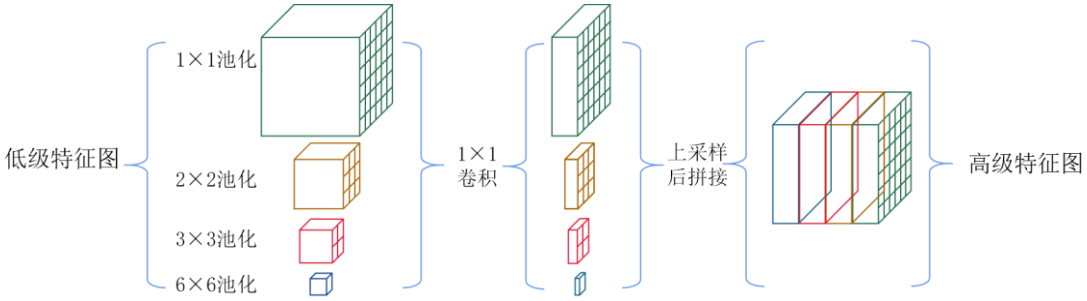

图 3-6 高级特征提取的过程，先经过不同尺寸的池化，然后通过卷积调整通道数量，最后上采样并拼接结果输出高级特征图

Fig.3-6 The process of advanced feature extraction. After pooling layers of different sizes, then adjust the number of channels through convolution, and finally up-sample and concatenate the results to output advanced feature maps

从图 3-2 中可以看到，在基础特征提取模块处理过完原图过后生成的低级特征图 b 过后，会产生分支，金字塔池化模块融合了四种不同尺度下的特征。。其中对于图 3-2 中的 c 模块，即金字塔池化模块，就是一种充分利用全局信息的方式，结合图 3-6 的细节过程展示，首先对特征分别池化到不同大小的特征块儿，且池化大小以此为 1×1、2×2、3×3、6×6、如图 3-6 所示例如由 2048 个通道的、30×30 像素大小构成的特征经过 4 层金字塔模块池化后分别生成 30×30、15×15、10×10、5×5 的结果，因为这里相当于产生了 4 个分支，相比于原通道数 2048 多了 4 倍的特征块，为了压缩通道数，对结果分别进行 1×1 卷积将通道数减少到 2048 的$\frac{1}{N}$，这里的 N 大小为 4 即金字塔模块池化层的数量。之后为了统一像素尺寸大小，将四个分支的高级特征通过双线性插值的图像处理算法将四个高级特征采样到同样大小的像素尺寸，最后如图 3-2 所示，将高级特征 c 与低级特征 d 拼接生成融合细节特征与全局特征的输出，最后的通过一层卷积生成最终的预测结果。

以上小结较为详细地介绍了基于肾脏超声的智能诊断方案的图像语义分割阶段过程，我们从原始的超声图像输入到最终语义分割结果的输出进行详细地过程描述，首先利用融合了注意力机制的基础特征提取模块提取超声图像的低级特征，然后利用金字塔池化算法在低级特征的基础上提取高级特征，通过一些图像

处理的算法统一尺寸并与低级特征组合成最终预测特征图，然后输出预测结果。

3.4 实验与结论

本小节对融合注意力机制和金字塔池化的图像语义分割网络进行实验评估，实验建立在首都医科大学附属儿童医院提供的肾脏超声图像数据集上，首先简单的描述我们的数据集，以及实验的设置与参数配置，之后根据所设计的实验来评价本研究所提出图像语义分割算法，最后分析实验结果并评估该模块的作用。

3.4.1 数据集与实验设置

我们所有的实验都是在同一个 UPJO 数据集上进行的，该数据集包含来自 17 名患者的 1850 张超声图像。原始数据由北京儿童医院于 2019 年 9 月 10 日至 2020 年 3 月 10 日收集，由专业超声医师在肾脏冠状面和横切面（标准位置）截取的 174 张截图组成，另外还有 120 扫描仪存储的视频，记录超声波扫描期间超声波技师的操作。 将所有源视频以大小为 5 的帧速率切成图像，并且通过数据清洗，去掉脏数据、包括模糊而无法区分的图像。 然后经验丰富的医生使用图像标注软件—Labelme 在每张图像上标注病变区域和肾脏区域。 最后，经过这样的预处理得到了 1850 张带注释的 UPJO 超声图像，目前胎儿肾积水的诊断标准使用最多的是胎儿泌尿外科协会(Society of Fetal Urology)制定的先天性肾积水分级系统，其中病理等级分为了 的共 5 级(SFU，0~4 级)共 5 级。本文对模型和算法进行了对比试验，并且在注意力机制的融合设置消融实验，均使用了相同的实验环境，表 3-1 列出了本次实验环境的具体配置，包括实验室用的计算机配置。实验的模型训练的开发与调试工作主要在 Windows 操作系统下进行，代码主要基于 Python 的深度学习框架 PyTorch。

表 3-1 实验环境

Tab. 3-1 Experimental environment

环境	参数	具体配置
硬件环境	CPU	Intel(R) Core(TM) i7-10700F 八核 2.90GHz
	GPU	NVIDIA GeForce RTX 3070
	内存	32GB
	显存	8GB
软件环境	操作系统	Windows 10 专业版
	开发环境	JetBrains PyCharm

我们将标注好的 1850 张图像数据分为比率为 9：1 的训练集与测试集，采用批量训练并将批量规模设置为 8，训练世代为 200，训练过程加入 Adam 优化器，

初始学习率设为 0.005，并设置学习下降的倍率为 0.9，频率为 1 世代的方案更新学习率，以平均交并比作为模型结果的评价指标，输入分辨率为 473×473，因此，在数据加载的阶段，我们将重置图像的大小。由于图像的长度和宽度不相等，为了避免失真，我们通过添加灰度边框将图像填充成一个正方形，然后重置大小。然后，基于迁移学习的思想，利用预先训练好的模型文件的权重数据进行优化，从 ImageNet 上的公共数据集生成的学习过程。此外，还对其他基于 ResNet 的特征提取网络模型进行了比较训练，并保持了相同的学习速率、优化器和训练代数。我们所有的实验都是使用 Intel CPU 和 NVIDIA GPU 进行的。由于我们的 GPU 内存大小是 8GB，因此我们使用批处理训练，批处理大小设置为 8。然后，我们使用自适应矩估计（Adam）优化器（一种常见的优化算法）来训练我们的模型直到收敛。损失函数上结合了 Dice Loss 和 Cross Entropy Loss 作为组合损失函数，补充定义真实值为G_T，预测值为S_P，损失值为L_s，其计算公式为：

$$L_s = 1 - \sum G_T \log S_P - \frac{2\sum G_T S_P}{\sum G_T^2 + S_P^2} \tag{3-8}$$

为了完整地描述一个特定的深度学习模型的性能，除了准确性外，还应该考虑模型的复杂性，如参数的数量和计算量。为了提高其在精度和速度上的性能，我们在分割模型的中分别在顶部和底部位置分别集成了独立的注意力模块。为了找出基础特征提取模块种注意力的效能，以及其子结构即空间域与通道域注意力单元的位置对整体结果的影响，我们分别对两个单元和两个串联组合进行了比较实验。结果如表 3-2 所示，以经过 50 个世代训练的最终模型（所有实验中其余参数均一致）的 MIoU 和 MPA 作为评价指标。此外，为了确认我们提出的基于注意力的 ResNet 的网络结构比原始的 ResNet 轻，我们比较了参数的数量，并计算 Giga 级的定点每秒乘的累加运算量（GMACs）来测量表 3-3 中每种方法的计算复杂度。

3.3.2 消融实验与分析

我们进行了消融实验来比较 NAM 中四种注意力单元的排列，通过删除部分结构和改变顺序来研究网络的性能从而得到最优的网络结构，并得到结果如下：

表 3-2 注意力单元排列组合的对比结果

Tab 3-2 Comparison results of attention unit permutations

实验序号	描述	MIoU	MPA
1	ResNet50 + 空间注意力	88.17	91.03
2	ResNet50 + 通道注意力	88.16	91.05
3	ResNet50 + 空间 - 通道注意力	88.39	92.78
4	ResNet50 + 通道 - 空间注意力	88.53	93.63

如表 3-2 所示，不同的模型采用了不同组合的混合域的注意力单元，我们对各模型有效性做了 MIoU 和 MPA 两个维度上的评估。通过比较发现每种方法的性能差距并不明显，其中 MIoU 的最大差值仅为 0.27。相对而言，实验 4 的最终模型效果最好。单独集成空间域或通道域中的某一个注意力单元的情况下，表现均劣于混合域组合的情况，且各自对模型的影响基本一致。混合域集成注意力模块的情况下，实验 4 在平均交并比上略优于实验 3，仅相差 0.14 个百分点。在像素准确性上，实验 4 明显优于实验 3 且相差 0.85 个百分点。综上可得出实验 4 的方案使得最终模型的有效性最高，之后小节的对比试验也是在实验 4 方案基础上，与各个图像语义分割模型完成多个指标维度上的比较分析。

3.3.3 对比实验结果与分析

这一节我们将从准确度、复杂度两个维度去对评估我们的最终模型 APPNet，对比一些经典语义分割模型的计算性能。对比最终分割模型与基础特征提取模块，并且评估了各经典基础特征提取模块的可训练的参数量和计算量，包括评估了基于表 3-2 最优结果的基础特征提取模块的可训练的参数量和计算量。首先，我们选择了几个在基础特征提取网络模型上同样采用 ResNet 的图像语义分割网络，以及几个现有的流行的图像语义分割网络。并对所有深度学习网络进行了参数一致、数据集一致、训练过程一致的对比实验。选择最优的分割模型与我们的 NAM 模块相结合，作为最终的图像语义分割网络并再次进行训练。

表 3-3 模型有效性的比较

Tab. 3-3 Comparison of Model Effectiveness

实验序号	方法	MIoU	MPA
1	PSPNet	87.61	92.01
2	Deeplab_v3	85.47	91.58
3	DANet	80.54	83.14
4	**APPNet**	**88.93**	**93.52**
5	FCNs	81.66	82.46
6	RefineNet	84.79	88.68

结果如表 3-3 所示。对比实验 2、3、4，结果显示，在基于 ResNet 作为基础特征提取网络的前提上，PSPNet 的有效性最佳，因此将其与两个轻量级注意单元相结合，对其进行改进然后构建了 APPNet，并进行对比实验来彻底评估最终模型的有效性。我们验证了融合 NAM 的 ResNet 的性能优于其它没有融合 NAM 模块的语义分割网络，即对比实验 1 和实验 4，结果显示，与原始的 PSPNet 相比，其性能提高了 1.32 个百分点。对比实验 4、5、6，可以发现与其它的不同种类的语义分割算法相比，在肾脏超声数据的场景下，我们的模型仍然具有优越性。

除此之外，我们还分别评估了空间注意单元和通道注意单元的计算参数，以论证它们的轻量性。将我们构造的基础特征提取网络作为基线，对比各经典基础特征提取网络模型和基线的复杂度。

表 3-4 复杂度的比较

Tab. 3-4 Comparison of Model Complexity

实验序号	描述	Params(M)	MACs(G)
1	ResNet50	25.55703	18.84624
2	通道注意力	0.00020	0.00004
3	空间注意力	0.52480	0.00378
4	**Ours**	**26.08203**	**18.85144**
5	ResNet101	45.07416	35.96785
6	ResNet152	60.7178	53.10086
7	AlexNet	61.10084	13.09258
8	VGGNet	138.35754	68.2407

我们计算了空间注意单元和通道注意单元的复杂度，即实验 2 和实验 3，实验结果如表 3-3 所示，与基线相比，空间注意单元小了 5 个数量级的参数量，通道注意单元相差了 2 个数量级的参数量，分别相差了 6 和 4 个数量级的计算量，由此分别证明并展示了他们的轻量性。总体上，对比实验 1 和实验 4，基线模型和原始结构相比，参数量增加了约 2%、计算量增加约了 0.03%。对比实验 4、5、6，即 ResNet 同族模型见比较，无论是可训练参数量，还是计算复杂度都比融合了注意力模块的 ResNet50 高。对比 AlexNet，实验 7 表明其参数量约为的基线的 2.4 倍，但计算量略小于基线。对比 VGGNet, 实验 8 表明其参数量约为基线的 5.3 倍，计算量约为基线的 3.5 倍。因此从计算复杂度的角度去分析，融合了注意力机制的基线模型同样具备轻量级特性，此外，结合表 3-3 实验 1 和表 3-4 实验 1，以及表 3-3 实验 4 和表 3-4 实验 4，对比结果分析，模型增加了 2%参数量，和 0.03%的计算量，最终训练模型结果的 MIoU 提高了 1.32 个百分点，MPA 提高了 1.51 个百分点，可以发现前后模型在增加了极少的复杂度的情况下，有效性明显提高了。

综上所示，我们评估了模型的复杂度和有效性，最终验证了在极小的开销下我们的模型的性能有了一定的提升，也印证了注意力机制在肾脏超声图像的应用场景下，能够有效地提高预测准确率，接下来将从抽象层面到可视化层面进一步分析注意力机制在超声图像应用场景下的实际效果。

3.3.4 注意力可视化分析

从可解释性的角度阐述，为了更加抽象地理解注意力机制的作用，分别从超

声扫描过程中的横截面与冠状面的数据中，挑选了左、右肾的图像数据通结果，并可视化了注意力层之后在数据集中的权重热力图，具体如图 3-所示，颜色从小（蓝色）到大（红色），波长越长即表示权重越大。

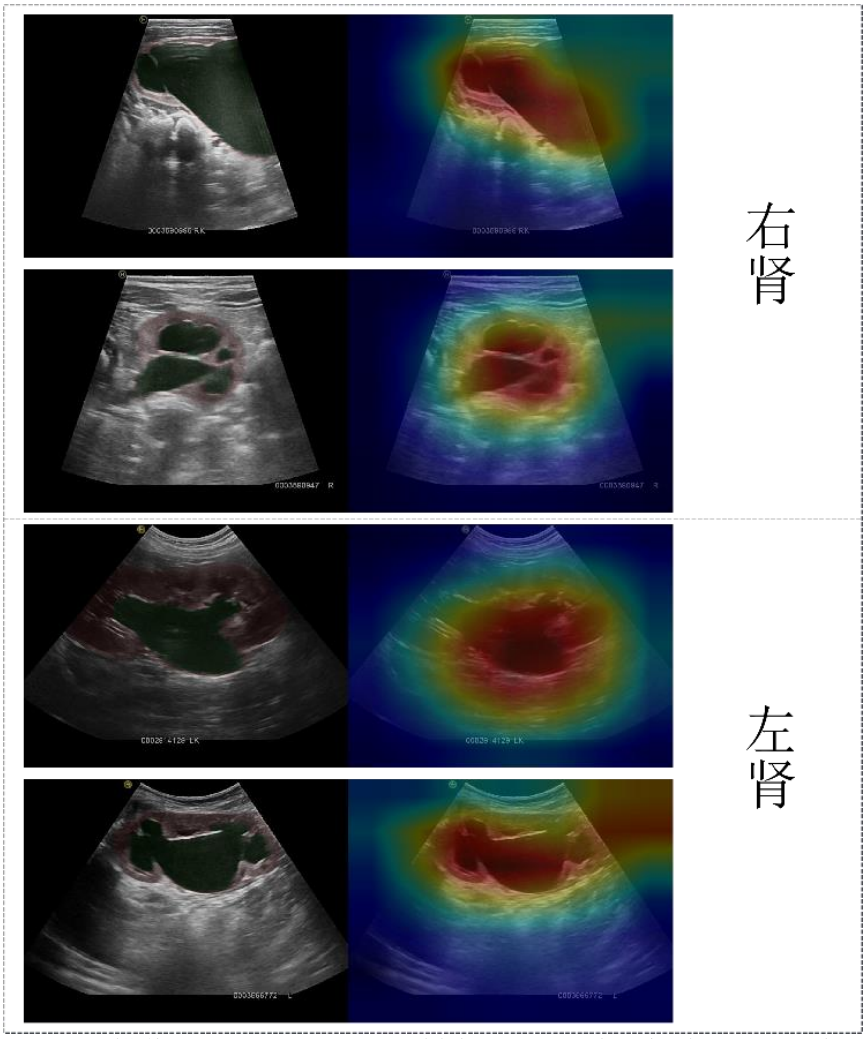

图 3-7 左肾右肾的横截面和冠状面，共四张样本图，左侧为分割结果，右侧为权重热力图

Fig. 3-7 The cross-section and coronal plane of the left and right kidneys, a total of four sample images, the left is the segmentation result, and the right is the weight heat map

理论上注意力机制得到的权重乘以输入特征图可以帮助网络进行自适应特征细化，权重数值表征了特征图该点的重要程度，通过感受野反推至原图像，即表示了该区域的重要程度，观察图 3-7 可以发现，网络的权重主要集中在病变区域，其次是肾脏区域，再其次就是其它的无关区域，包括超声图像自带的黑色背景的权重都很低，由此可证明，注意力机制能够让网络自适应关注需要关注的地方，从而使网络按照我们的期望，学习到更加重要的特征信息，并尽可能忽略掉无关区域的特征分量。

3.5 本章小结

本章节详细叙述了肾脏超声图像诊断框架中的分割的部分，我们希望以很小的代价提高模型的分割精度，因此，我们提出了一种基于注意的残差网络结构，并将其应用于金字塔场景解析网络模型。给定一个输入，得到初始特征图，模型沿通道和空间维度依次推断注意图，再乘以输入特征图，输入到剩余网络结构中，在输出端重复注意过程，然后作为下一个模块的输入。由于注意模块的轻量化，可以忽略其开销，有效提高模型的精度。我们通过从北京儿童医院获得并标记的数据集进行实验来验证我们的模型。我们的实验表明，与添加注意力模块前的模型相比，该模型的计算量和参数数量增加较少，性能有所提高，通过可视化注意力神经网络层以具象分析其作用，最后印证了该模型的轻便性和有效性。

第 4 章 基于离散小波池化增益的图像分类模型

4.1 模型提出的动机

之前章节有提到在诊断阶段分为语义分割阶段和图像分类阶段，前一阶段是为了增强肾脏超声图像的边缘信息，为最后的分类任务提供特征信息更丰富的数据集，经过上一章节所介绍的语义分割算法过后，我们获得了一个可以将原始超声图像分割成背景区域、器官区域与病变区域的分割模型，并且在图像分类阶段，我们通过语义分割模型补充分割了新的数据集，将分割结果与原图按比例叠加后作为图像分类神经网络的新的数据集。这一阶段的数据集仍然具有超声图像自带的白斑噪声，且器官的纹理特征等信息也与白板噪声高度耦合。从通信工程中的数字信息处理中可以思考这种白斑噪声从频域上处理，因此在图像本身是一种二维的数字信号载体的前提下，我们试图从数字图像频域处理的角度去挖掘不同维度中的特征信息。

图像中的频域处理算法主要有傅里叶变换和小波变换，其中傅里叶变换是居于全局图像的频域变换，其变换函数的每个点的幅值都对变换后的数据有相当程度的影响。而小波变换所使用方式的是局部基，对于指定基的系数，函数对该系数是否有影响取决于其是否在当前基所支撑的点上。这就导致小波不仅包含频率信息，即由基的频率对应反映，也包含时间信息，即由基所在的时间轴位置对应反映，在图像中级即表现为位置信息。小波的尺度收缩特性使得小波变换具有了分形的特性。这种特性能够小波捕获了信号和数据的局部相似性。这种能力可以用来压缩和特征提取，这是傅立叶所没有的，小波也是有缺点的，在对信号频域分析的时候，由于小波变换使用了下采样，从信息论的知识上解释即该操作造成了频谱的混叠，因此我们仅仅在网络中对不同的小波变换下采样结果中的在通道维度上叠加并进行神经网络深度学习的过程而不做小波变换的逆变换过程。也正是这里小波变换带来的下采样效果，使之可以完全替代一般模型前前端的池化操作，这里传统的池化操作有最大值池化、均值池化等算法，会丢失部分图像特征信息，对于一般图像，这种操作对分类结果几乎没什么影响甚至能节省网络计算复杂度，但对于超声图像这种显著性低的特征密度较大的图像来说，池化操作会大大降低最终的分类精度，因此我们尝试在目前计算机视觉上已有的表现优异的网络模型中尝试这种替换方式来达到更好的分类效果，并构建了结合小波池化增益层的网络模型，命名为 WaveConvNeXt。

4.2 神经网络中的离散小波变换

正如之前所说，为了对超声图像做病理分级判断，我们在智能诊断系统中设计添加了分类器，在分类网络中池化层以 2 倍下采样的方式分解图像，其目的是为了减少总体计算量，但是同样带来的还有信息量的损失，既需要降低复杂度，又不想损失信息量，所以我们在分类器中加入了一种数字信号处理的方法来取代传统的池化层以折中之前所述之难题。从数学角度来讲，图像特征是具有局部变化统计特性的亮度值的二维阵列，如边缘和反差鲜明的同质区域，以多个分辨率来表示图像的结构（也称图像金字塔）非常有效且概念简单，与其相关的重要技术就是子带编码，将图像分解为一组频带受限的分量，也称子带，子带可以重组用以复原图像，在特征维度上也可以表示图像中人眼难以捕捉的信息。在深度学习模型中，我们不需要使用子带以复原图像，但对于计算机而言，这种信息的捕捉与学习十分容易。如前所述，我们添加了一个分类器来获取 UPJO 病理分级。在传统的分类网络中，池化层采用下采样来分解图像，减少总的计算量，但会导致信息丢失。 由于 DWT 及其逆运算是可逆的，从而保证了在降低复杂度的同时不丢失信息。 DWT 被认为是分解图像并获得所需频域信息的有效算子，因此被应用于分类器以取代传统的池化层。为了方便理解替换过程的原理，我将从一维的角度去讲解，首先定义两个基函数，一个Ψ(x)和尺度函数φ(x)，对于一维的离散序列$f(n)$，其正向的离散小波级数展开系数为

$$W_{\varphi}(j_0,k)=\frac{1}{\sqrt{M}}\sum_{n}f(n)\,\varphi_{j_0,k}(n) \tag{4-1}$$

$$\begin{aligned} W_{\Psi}(j,k)&=\frac{1}{\sqrt{M}}\sum_{n}f(n)\,\Psi_{j,k}(n),\ j\geq j_0\\ j&=0,1,\cdots,J-1\\ k&=0,1,\cdots,2^{j}-1\\ n&=0,1,\cdots,M-1 \end{aligned} \tag{4-2}$$

在以上等式中，$\varphi_{j_0,k}(n)$ 和 $\Psi_{j,k}(n)$ 是在 J 尺度下的基函数的支撑区上使用 M 等间隔的取样，通常$j_0=0$，变换本身由 M 个系数组成，最小尺度为 0，最大尺度为J，这里再补充定义

$$\begin{aligned} \varphi_{j,k}(x)&=2^{j/2}\varphi\left(2^{j}x-k\right)\\ \Psi_{j,k}(x)&=2^{j/2}\Psi\left(2^{j}x-k\right) \end{aligned} \tag{4-3}$$

k 决定既定方向上的位置，j 决定取样的宽度，j_0 为限定的任意开始尺度，则得到的展开集合 $\{\varphi_{j_0,k}(n)\}$ 是 $\{\varphi_{j,k}(n)\}$ 的一个子集。

与一维 DWT 类似，二维 DWT 使用二维尺度函数和小波函数。 我们先对二维数组的行进行一维变换，然后对上一步结果的列进行一维变换。结果，然后再取结果的列的一维变换，但在二维情况下会得到水平、垂直和对角细节系数，对于大小为 $L \times D$ 图像，其 DWT 是

$$W_{\varphi}(j_0,l,d)=\frac{1}{\sqrt{LD}}\sum_{x=0}^{L-1}\sum_{y=0}^{D-1}f(x,y)\,\varphi_{j_0,l,d}(x,y) \tag{4-4}$$

$$W_{\Psi}^{i}(j,l,d)=\frac{1}{\sqrt{LD}}\sum_{x=0}^{L-1}\sum_{y=0}^{D-1}f(x,y)\,\Psi_{j_0,l,d}^{i}(x,y),i\ =\{H,V,D\} \tag{4-5}$$

最终得到 4 个 1/4 大小的输出子图：W_{φ}，W_{Ψ}^{H}，W_{Ψ}^{V}，W_{Ψ}^{D}，分别表示原图近似分量和水平、垂直和对角上的细节分量。我们将这四个分量构成的网络层命名为小波增益池化层，相比传统池化层的效率换时间的特性，这种小波增益池化层缓和了直接池化带来的信息损失，从特征工程的角度理解，通过这些人工设计的细节分量把较低级别的特征表示成更加抽象的特征，提高了表示学习下，输入数据的转换层次，利用深度神经网络模型的自动提取特征的能力，使模型从频率域中学习到更加抽象的特征。

4.3 图像分类网络-WaveConvNeXt

该分类模型在 ConvNeXt 基础网络上，添加了小波增益池化层，其中 ConvNeXt 参考了 ResNet 的多阶段设计思想，对于其核心宏观结构，以为 ConvNeXtBlock 为单元，在每个各个阶段具有不同的特征分辨率，各阶段且计算比率 调整为(3:3:9:3),。考虑到图像的信息冗余性，需要对输入图像下采样并聚合到适当的特征尺寸，其中 Sampleblock 起到至关重要的降采样操作，通过归一化用来保持模型的稳定性，并以步长与尺寸相同的卷积核实现无覆盖的卷积运算从而实现降采样过程。对于 ConvNeXtBlock, 相比于 ResNet 使用具有更大感受野的卷积核、更少的激活函数 ，并且使用 LayerNorm 以替换表现欠佳的 BatchNorm。

如图 4-1 所示，我们在首次下采样阶段用替换为小波池化增益层，并对高频分量做一个通道剪枝，以达到后续的维度一致性。模块内使用 Layer Scale 辅助优化，其是一种为了保持各层值和梯度量级的一致性的标准化方法，通过加入一个可学习的缩放参数矩阵以解决残差连接带来的方差增大的问题并提高训练效率。以 M_{weight} 表示可学习的缩放参数矩阵，L_{out} 表输出， X 表输入，其计算公式如下：

$$L_{out} = M_{weight} \times \frac{X - \bar{X}}{var(X) + \varepsilon} + b \qquad (4-6)$$

其中$\bar{X}$表均值，$var(X)$表样本方差，ε为一个无穷小参数以防分母为零，b为偏移量参数。

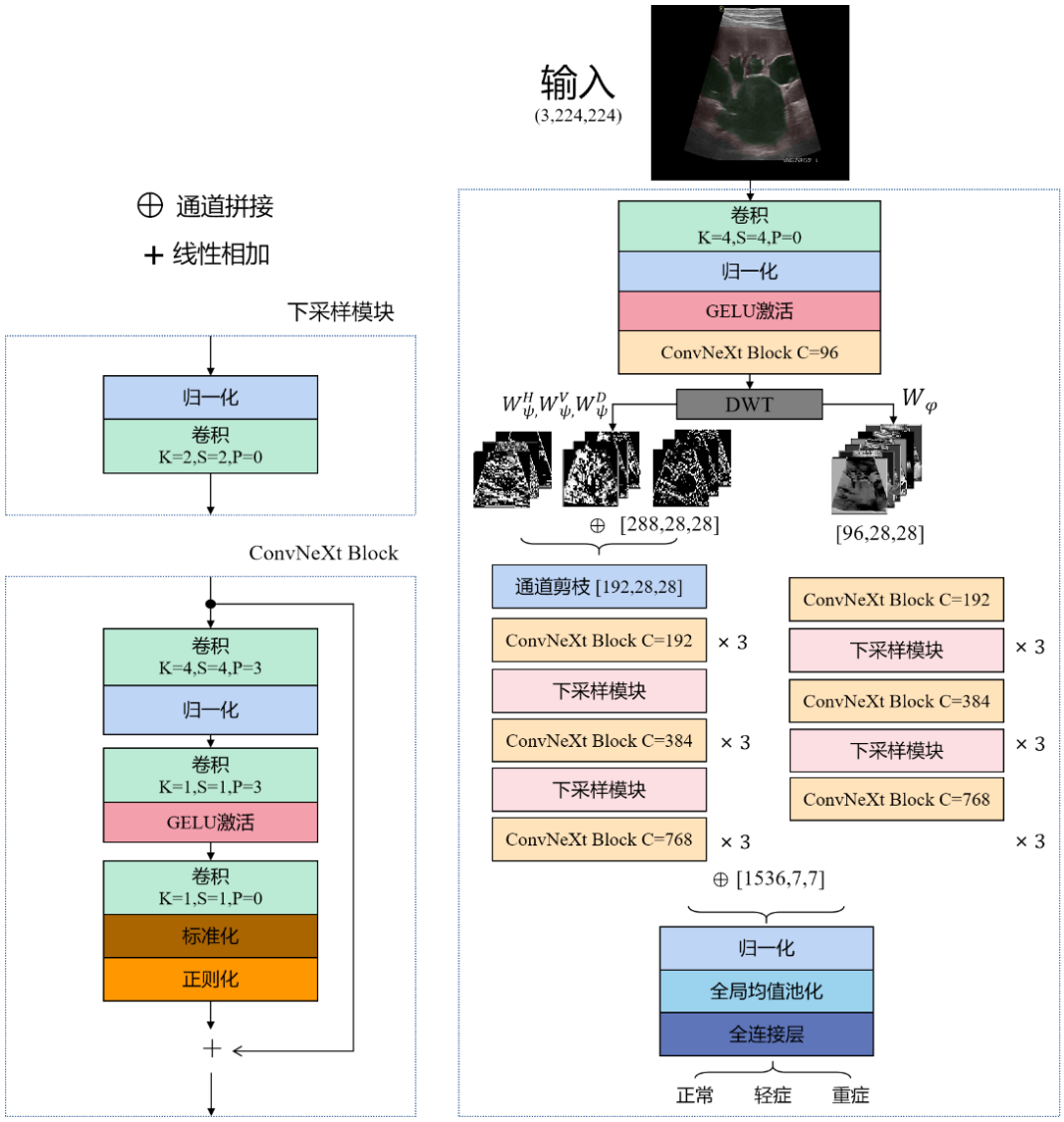

图 4-1 WaveConvNeXt 模型的框架构成，其卷积运算中的 K 表卷积核大小，S 表卷积核滑动步长，P 为填充大小，ConvNeXt Block 中的 C 为输出通道数，激活函数为 GELU，归一化算法为 LayerNorm，标准化算法为 Layer Scale，正则化算法为 Drop Path，最终输出正常、轻症、重症三类结果

Fig. 4-2 The frame structure of the WaveConvNeXt model, K in the convolution operation represents the size of the convolution kernel, S represents the sliding step size of the convolution kernel, P is the padding size, C in the ConvNeXt Block is the number of output channels, and the activation function is GELU, the normalization algorithm is LayerNorm, the normalization algorithm is Layer Scale, the regularization algorithm is DropPath, and the final output is normal, mild, and severe.

模块内还引入 DropPath 正则化方法。DropPath 通过加入可学习的超参数，即存活率，以自动化寻找 Dropout 模式。以ρ表存活率，通过一个简单的线性衰减规则，从$\rho_0 = 1$表示对于最初输入的存活率开始，对于最后一个 ConvNeXt Block 的存活率ρ_{Last}，其过程中的存活率计算如下，

$$\rho_i = \frac{i \times (1 - \rho_{Last})}{L} \tag{4-7}$$

其中 i 表计算所至的次序，L 为计算总次数。通过该方法从而可以提高网络鲁棒性和泛化能力，并提高训练速度。

4.4 实验与结论

4.4.1 数据集与实验设置

在实验环境上，因为我们所有的实验主要分为分割阶段与分类阶段，因此实验环境与第三章的 3-1 表格所示的实验环境一致，且都是在 UPJO 数据集上按处理顺序进行的。上一章节实验完成了整个智能诊断系统的分割阶段，最终产出了一个肾脏器官、病变区域、无关区域三类分割的图像语义分割模型，所有所用数据集包含了来自 17 名门诊病人的 3289 张超声波图像，并由具备专业医学影响技术的医生采用标注工具对病变区域和肾脏通过打点环绕的方式标记并注释了 1850 张图像用于分割阶段，由经验丰富的医生对所有共 3289 超声波图像进行病理分级，其中分级标准采用了胎儿泌尿外科协会制定的先天性肾积水分级系统。其中在训练出具备良好准确度的分割网络之后，加入之前未标注的所有剩余图像（1439 张），经过训练好的图像语义分割模型处理并输出，现定义图像语义分割结果为 S、原图为 F，然后定义输出的融合图像结果为 $F.I.$，合并算法如下：

$$F.I. = F \times (1 - \alpha) + S \times \alpha \tag{4-8}$$

并将病理分级完毕的 3289 图像结合医生的需求，再分为 SFU 0-2、SFU 3 和 SFU 4 共三类，即正常、轻症合重症。将分类好的 3289 张数据经过分割器，以尽量保留器官区域的原则裁剪不必要区域，结果大小为 810×608 为并分成比率为 9：1 的训练集与测试集，这一阶段加入了对比实验以证明我们的分类网络的优越性，并以准确度（Accuracy）作为评价标准，对于分类模型 f 与大小为 N 的测试集 D，指标值 $\mathrm{A}(f; D)$ 的计算方法为：

$$\mathrm{A}(f; D) = \frac{\sum_i^N \{f(x_i) = lable(x_i)\}}{N}, x_i \in D \tag{4-9}$$

所有对比试验除初始学习率为以外，其它参数一致，使用批量规模为 32 的批量训练，训练世代 50，加入 Adam 优化器并设置衰减系数为 0.9、频率为 1 世代的方案更新学习率，对于最终的分类结果：正常、轻症与重症，数据集中各个分类的数据量并不平衡,针对数据集的不均衡问题，以各分类的数据量间的比值作为权重系数加入到损失函数的计算中去，并以此算法作为整个训练过程的损失函数。整体训练网络的输入是大小为 224×224 的图像，而对于我们目标网络中的小波变换，采用哈尔基小波变换，并采用标准分解方法，首先使用一维小波对图像每一行的做变换处理，再同理对每一列进行变换，进而生成产生近似分量与细节分量，其中近似分量很好的替换了池化层的输出，为了印证这一效果，我们做了对比实验，即在其它网络上以小波变换的近似分量替代池化层输出并训练，训练世代次数设置为 30，其它参数一致。

4.4.2 对比实验与分析

在上一章节，我们进行了实验并选取了 MIoU 最高的分割模型作为我们分割阶段的图像语义分割器，将分割结果与原图合并然后裁剪作为分类阶段的输入。参考公式 4-8，我们在实际实验过程中取 α 等于 0.1，总共将 3289 张图像通过语义分割算法并完成图像融合，具体过程如图 4-2 所示。

在分类器中，我们在不改变特征维度的情况下结合了离散小波变换，获取了与 2 倍下采样效果一致的近似分量，和具备高频信息的细节分量，增强神经网络模型能力，帮助神经网络从频域学习到更抽象的纹理等特征信息，其中维度由最初的 64×56×56 增加到 256×56×56，原池化过程后的特征图如图 4-3 所示。将池化层更改为由小波池化增益层之后，将原始池化层的输出替换为近似分量，然后并行添加细节分量的网络路径，具体特征分别如图 4−4 所示。图 4-3 和图 4-4.a 相对比，离散小波变换的近似分量是池化操作的有效替代分量，从图 4-4.b 可以看出，新的特征分量包含密集的高频噪声信息和轮廓信息。

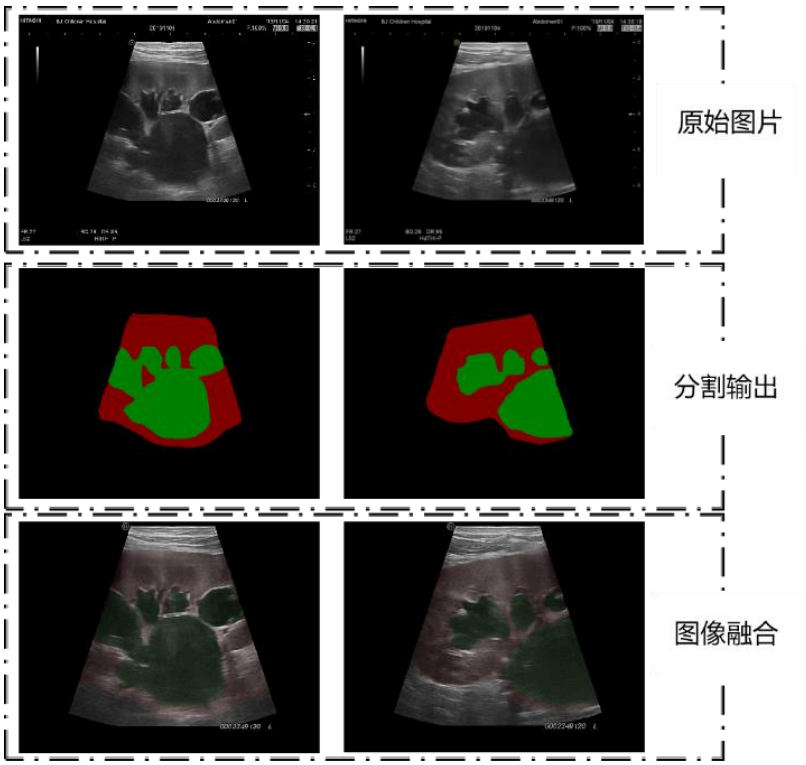

图 4-3 图示为分割阶段的处理过程，包括语义分割过程和融合图像过程

Fig. 4-4 The figure shows the processing process of the segmentation stage, including the semantic segmentation process and the fusion image process

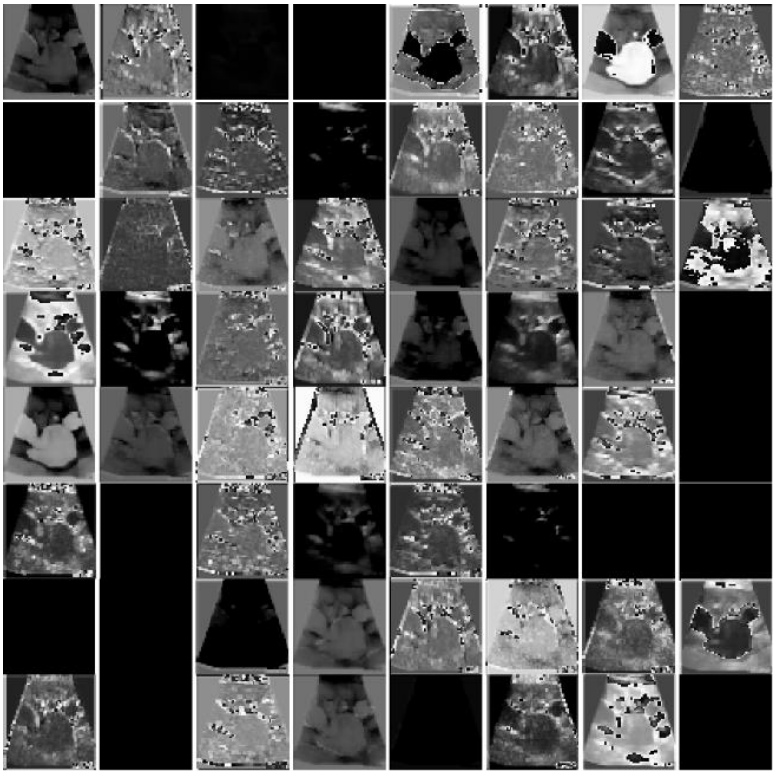

图 4-3 图示为原始网络中在经过池化层后的数据集中随机选取的超声图像的特征图，维度为 64×56×56

Fig. 4-3 The picture shows the feature map of the ultrasound image randomly selected in the data set after the pooling layer in the original network, with a dimension of 64×56×56

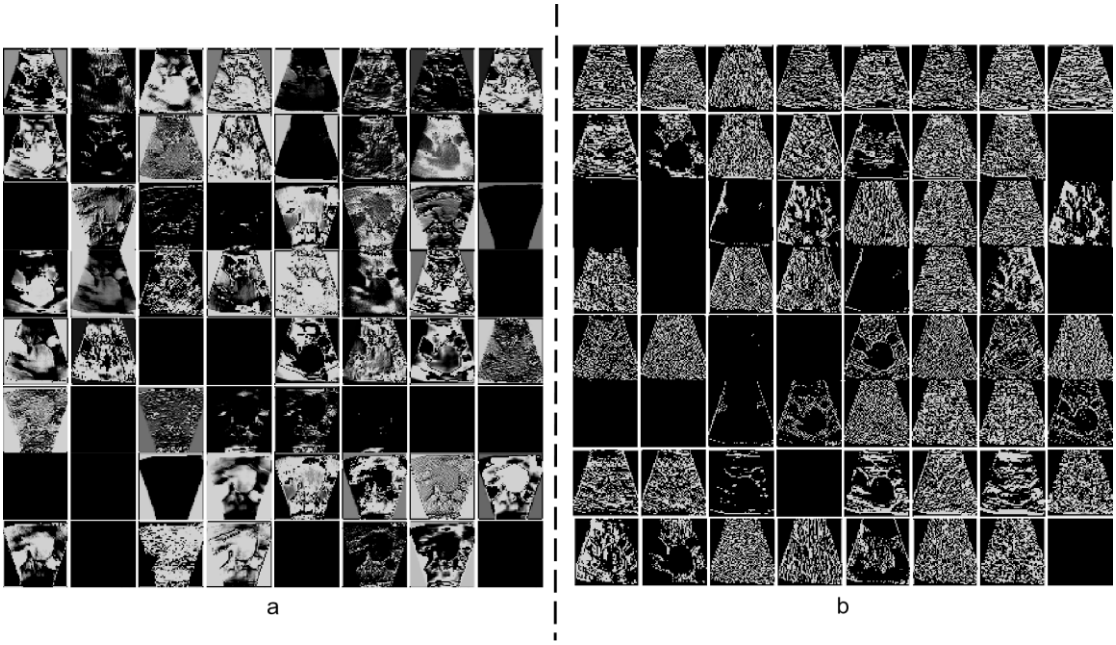

图 4-4 a.小波池化增益层后近似分量的特征图，单张维度为 56×56； b. 从水平、垂直和对角线三个细节分量中随机选择的特征图，其总数为 192。

Fig. 4-4 a. The feature map of the approximate component after the wavelet pooling gain layer, the single dimension is 56×56; b. The feature map randomly selected from the three detail components of horizontal, vertical and diagonal, the total number is 192.

除此之外，我们还对 6 组分类网络的性能，在数据集以及参数与超参数的设置均一致的情况下，进行了对比实验。具体结果如表 4-1 所示。表中给出了不同网络训练过程中的最优精度模型和训练迭代阈值。由此我们观察到，ConvNeXt 在四个基准网络中具有更好的准确性，我们提出的 WaveConvNeXt 模型更是优于所有模型，这证明了我们选择 ConvNeXt 的正确性。

表 4-1 各分量网络分类的测试的结果和我们的模型

Tab. 4-1 The results of the testing of each component network classification and our model

模型	准确度	迭代阈值
AlexNet	84.45	35
VGG	86.36	27
GoogleNet	84.75	38
ResNet	89.94	38
ConvNeXt	91.65	49
Ours	**93.47**	50

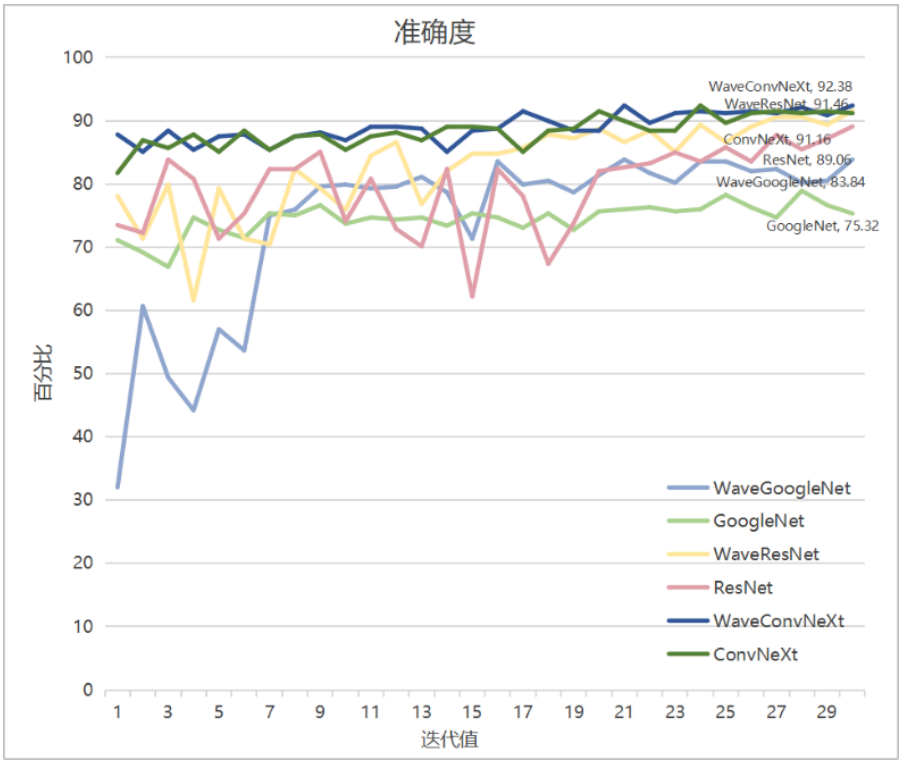

图 4-5 在经典的基础分类网络上，用小波池化增益层替换池化层后的性能

Fig. 4-5 Performance after replacing pooling layers with wavelet pooling gain layers on a classic base classification network

另外，我们还尝试在基于表 4-1 的各分量网络上做了同样的小波池化增益层替换传统池化层的操作。由于过多的池化层会大大增加网络的复杂度，因此我们只选择和测试了几个池化层在重要位置上的网络，从而剔除了 VGG 网络。以 30 迭代阈值训练其它基础分类网络在采用小波池化增益层前后的模型，然后收集训练过程中准确度曲线变化。如图 4-5 所示，用小波池化增益层替换池化层，增加特征提取的维度与数量，可以提高所有网络的性能，其中仍然以 WaveConvNeXt 为最佳。进一步证实了我们所提出的方法的优越性与有效性。

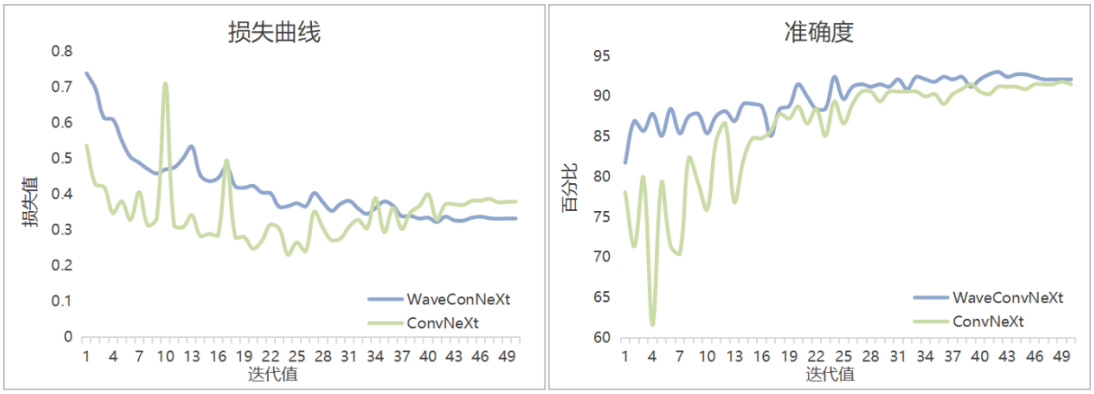

图 4-6 ConvNeXt 模型与我们提出的模型之间的训练曲线对比图

Fig. 4-6 Comparison of training curves between the ConvNeXt model and our proposed model

由于我们提出的分类网络的方法是基于 ConvNeXt 的，因此我们将以相同实验条件下的训练过程的曲线图为切入点，进一步挖掘小波池化增益层的其它优越性。如图 4-6 所示，从曲线的振动振幅和频率来看，在约迭代次数从 1-30 范围内，原网络的震荡幅度明显高于我们的模型，可以证明我们的模型相比原模型，具有更加鲁棒的训练过程和更高的性能。再观察模型训练的后半段，原模型的损失曲线出现了回升的情况，且准确度曲线趋于平滑，意味着训练过程陷入了局部最优解，甚至是过拟合，相比之下，我们的模型损失曲线稳定下降并趋于平滑，证明模型接近饱和状态。以上分析进一步证实了我们的网络结构的优越性。基于以上结果，用小波池化增益层合理地替换池化层，将从频率维度上增加特征的数量与特性，使训练过程更加稳定、模型学习的特征更加鲁棒，模型的性能更加优秀。

4.5 本章小结

本章节详细叙述了肾脏超声图像诊断框架中的分类的部分，我们希望在这一阶段尽可能准确地预测每一张超声图像。为了消除池化操作带来的信息损失的负面影响，且由于超声图像的频率信息丰富的特性，通过离散小波变换算法构造的小波池化增益层替换原始网络结构的池化层，使得神经网络模型在小波池化增益层的作用下，学习到图像的频率特征，根据算法的低频分量和高频分量分别并行构造分支网络。在充分的实验证明下，最终相比原始网络结构下，准确率提升显著，印证了我们所提出方法的有效性与优越性，并且通过分析了训练过程的各指标的曲线图，以印证我们根据医疗图像领域中，构造的超声图像分类模型的鲁棒性。

结论

本文详细阐述了我们提出的基于肾脏超声图像的智能诊断方法，研究过程大致分为前期工作阶段、方法论研究阶段和最终的实验阶段。在前期工作中，我们主要配合首都医科大学附属儿童医院完成了数据收集、清洗、分析的工作，提炼出数据与临床应用下的需求与问题，针对各类问题调研并提出解决方法，其中包括：肾脏超声数据的无关区域干扰器官区域、病变区域识别的问题，病变区域包含在器官区域从而难以区分边界、纹理等信息的问题，以及各病理等级下数据分布不均匀的问题。

结合以上问题在方法论研究阶段，对于儿童肾脏超声图像，我们创新性地提出了分割-分类的诊断框架。在分割算法中，采用注意力机制和金字塔池化模块分别强化了神经网络的识别能力与分割精度，注意力机制通过构造注意力权重图赋能网络忽略无关部分而更多地关注感兴趣部分，并且创新性地将注意力机制轻量地嵌入到骨干网络的两端，而不是嵌入到每一次卷积层，既保证了骨干网络结构不受破坏而得以使用预训练数据下的迁移学习，又使得网络以几乎可以忽略的计算开销提高分割准确率，金字塔池化模块通过多尺度感受野的解析能力帮助网络学习到不同尺度下局部特征信息和全局特征信息，以应对器官区域与病变区域呈包含关系而引发的识别难题。在分类算法在，我们首先将上一阶段的分割结果按既定算法与原图进行不超过像素最大值的融合，以正常、轻症、重症划分数据并总共构造了 3289 张图像数据用于最后三分类模型，我们创新性地通过数字图像处理中的离散小波算法构造了小波池化增益层，替换掉传统骨干网络中的池化层结构，并且在网络优化算法中根据模型与数据，针对性地选择了归一化、正则化算法。

之后在分割实验阶段，通过消融实验印证了注意力模块与骨干网络的最佳结合方式，通过对比试验印证了我们分割算法的有效性与优越性，并以最终的 94.52 的平均像素准确率完成了诊断框架中的分割任务，我们还可视化了注意层得出的权重图在原图上的热力效果，分析了注意力机制的作用与意义，间接证实了其有有效性。在分类试验阶段，通过可视化小波池化增益层来阐述我们算法的原理，通过对比可证明相比原池化层，我们的方法能够完美地替代其结构并衍生出三个高频分量的分支以补充更多的特征信息，帮助模型从频域上学习从而获得更加鲁棒的分类能力，并且我们分析了小波池化增益层的思想在不同骨干网络下的作用，相比原结构均能提高分类准确度，从而印证了我们方法的有效性与优越性。我们还通过分析模型训练过程中指标曲线变化，对比突出了我们方法的鲁棒性，

使得网络模型拟合能力、拟合速度、准确度均得到了提升。实验表明最终以 93.47 的准确率完成了最终的分类任务。

在儿童 UPJO 诊断应用中，本研究提出了一个表现优异的方法，在该领域贡献了新的框架与思路，在实际应用将解放影像科医师的劳动力，提高诊断效率降低医疗成本，缓解不同地区医疗水平不平衡的问题，间接的为我国的医疗难题贡献了理论知识与科研成果，在未来的工作中，将会把重心放到数据安全问题上，考虑到医疗数据的敏感性、人民群众广为专注的隐私问题、安全问题等，考虑将基于 AI 的模型诊断模型结合区块链等安全相关的技术，建立了一个信息安全、过程保密的医疗图像诊断框架。未来还可以包括更多的机构来扩展这种合作，可以连接便携式和超便携超声扫描仪等终端设备，方便家庭诊断，减轻患者负担，最后，还可以为其他疾病可能的联合学习和广义诊断提供技术支持。

参考文献

[1] AKKUS Z, CAI J, BOONROD A, et al. A survey of deep-learning applications in ultrasound: Artificial intelligence–powered ultrasound for improving clinical workflow[J]. Journal of the American College of Radiology, 2019, 16 (9): 1318-1328.

[2] SMAIL L C, DHINDSA K, BRAGA L H, et al. Using deep learning algorithms to grade hydronephrosis severity: toward a clinical adjunct[J]. Frontiers in pediatrics, 2020, 8: 1.

[3] TURCO S, FRINKING P, WILDEBOER R, et al. Contrast-enhanced ultrasound quantification: from kinetic modeling to machine learning[J]. Ultrasound in medicine & biology, 2020, 46 (3): 518-543.

[4] SHOKOOHI H, LESAUX M A, ROOHANI Y H, et al. Enhanced point-of-care ultrasound applications by integrating automated feature-learning systems using deep learning[J]. Journal of Ultrasound in Medicine, 2019, 38 (7): 1887-1897.

[5] DHINDSA K, SMAIL L C, MCGRATH M, et al. Grading prenatal hydronephrosis from ultrasound imaging using deep convolutional neural networks[C]//2018 15th Conference on Computer and Robot Vision (CRV). [S.l.]: IEEE, 2018: 80-87.

[6] BLUM E S, PORRAS A R, BIGGS E, et al. Early detection of ureteropelvic junction obstruction using signal analysis and machine learning: a dynamic solution to a dynamic problem[J]. The Journal of Urology, 2018, 199 (3): 847-852.

[7] HE J, BAXTER S L, XU J, et al. The practical implementation of artificial intelligence technologies in medicine[J]. Nature medicine, 2019, 25 (1): 30-36.

[8] GULSHAN V, PENG L, CORAM M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs[J]. Jama, 2016, 316 (22): 2402-2410.

[9] JUSSUPOW E, SPOHRER K, HEINZL A, et al. Augmenting medical diagnosis decisions? an investigation into physicians' decision-making process with artificial intelligence[J]. Information Systems Research, 2021, 32 (3): 713-735.

[10] KUMAR Y, KOUL A, SINGLA R, et al. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda[J]. Journal of Ambient Intelligence and Humanized Computing, 2022: 1-28.

[11] LEE D, YOON S N. Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges[J]. International Journal of Environmental Research and Public Health, 2021, 18 (1): 271.

[12] ROTEMBERG V, KURTANSKY N, BETZ-STABLEIN B, et al. A patient-centric dataset of images and metadata for identifying melanomas using clinical context[J]. Scientific data, 2021, 8 (1): 34.

[13] ZHU Y C, ALZOUBI A, JASSIM S, et al. A generic deep learning framework to classify

thyroid and breast lesions in ultrasound images[J]. Ultrasonics, 2021, 110: 106300.

[14] LASSAU N, BOUSAID I, CHOUZENOUX E, et al. Three artificial intelligence data challenges based on ct and ultrasound[J]. Diagnostic and Interventional Imaging, 2021, 102 (11): 669-674.

[15] BISWAS M, KUPPILI V, EDLA D R, et al. Symtosis: A liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm[J]. Computer methods and programs in biomedicine, 2018, 155: 165-177.

[16] SHAH M, NAIK N, SOMANI B K, et al. Artificial intelligence (ai) in urology-current use and future directions: An itrue study[J]. Turkish Journal of Urology, 2020, 46 (Suppl 1): S27.

[17] GALLEGO J, PEDRAZA A, LOPEZ S, et al. Glomerulus classification and detection based on convolutional neural networks[J]. Journal of Imaging, 2018, 4 (1): 20.

[18] LORENZO A J, RICKARD M, BRAGA L H, et al. Predictive analytics and modeling employing machine learning technology: the next step in data sharing, analysis, and individualized counseling explored with a large, prospective prenatal hydronephrosis database[J]. Urology, 2019, 123: 204-209.

[19] HAMEED B Z, S. DHAVILESWARAPU A V, RAZA S Z, et al. Artificial intelligence and its impact on urological diseases and management: A comprehensive review of the literature[J]. Journal of Clinical Medicine, 2021, 10 (9): 1864.

[20] KHONDKER A, KWONG J C, RICKARD M, et al. A machine learning-based approach for quantitative grading of vesicoureteral reflux from voiding cystourethrograms: Methods and proof of concept[J]. Journal of Pediatric Urology, 2022, 18 (1): 78-e1.

[21] DRYSDALE E, KHONDKER A, KIM J K, et al. Personalized application of machine learning algorithms to identify pediatric patients at risk for recurrent ureteropelvic junction obstruction after dismembered pyeloplasty[J]. World Journal of Urology, 2021: 1-7.

[22] TABRIZI P R, ZEMBER J, SPRAGUE B M, et al. Pediatric hydronephrosis severity assessment using convolutional neural networks with standardized ultrasound images[C]//2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). [S.l.]: IEEE, 2021: 1803-1806.

[23] LIN Y, KHONG P L, ZOU Z, et al. Evaluation of pediatric hydronephrosis using deep learning quantification of fluid-to-kidney-area ratio by ultrasonography[J]. Abdominal Radiology, 2021, 46 (11): 5229-5239.

[24] LIU X, DENG Z, YANG Y. Recent progress in semantic image segmentation[J]. Artificial Intelligence Review, 2019, 52: 1089-1106.

[25] SHARMA N, AGGARWAL L M, et al. Automated medical image segmentation techniques[J]. Journal of medical physics, 2010, 35 (1): 3.

[26] CASTLEMAN K R. Digital image processing[M]. [S.l.]: Prentice Hall Press, 1996.

[27] PRATT W K. Digital image processing: Piks scientific inside: volume 4[M]. [S.l.]: Wiley Online Library, 2007.

[28] SHOTTON J, WINN J, ROTHER C, et al. Textonboost: Joint appearance, shape and context modeling for mulit-class object recognition and segmentation[C]//European conference on computer vision (ECCV). [S.l.: s.n.], 2006.

[29] YANOWITZ S D, BRUCKSTEIN A M. A new method for image segmentation[J]. Computer Vision, Graphics, and Image Processing, 1989, 46 (1): 82-95.

[30] EDEN M, KOCHER M, EURASIP M. On the performance of a contour coding algorithm in the context of image coding part i: Contour segment coding[J]. Signal processing, 1985, 8 (4): 381-386.

[31] SEO H, BADIEI KHUZANI M, VASUDEVAN V, et al. Machine learning techniques for biomedical image segmentation: an overview of technical aspects and introduction to state-of-art applications[J]. Medical physics, 2020, 47 (5): e148-e167.

[32] XU Y, WANG Y, YUAN J, et al. Medical breast ultrasound image segmentation by machine learning[J]. Ultrasonics, 2019, 91: 1-9.

[33] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2015: 3431-3440.

[34] BADRINARAYANAN V, KENDALL A, CIPOLLA R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39 (12): 2481-2495.

[35] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[36] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. [S.l.]: Springer, 2015: 234-241.

[37] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. arXiv preprint arXiv:1412.7062, 2014.

[38] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv preprint arXiv:1706.05587, 2017.

[39] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European conference on computer vision (ECCV). [S.l.: s.n.], 2018: 801-818.

[40] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2017: 2881-2890.

[41] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60 (6): 84-90.

[42] LIN M, CHEN Q, YAN S. Network in network[J]. arXiv preprint arXiv:1312.4400, 2013.

[43] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the

IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2015: 1-9.

[44] 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. [S.l.: s.n.], 2016: 770-778.

[45] LIU Z, MAO H, WU C Y, et al. A convnet for the 2020s[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2022: 11976-11986.

[46] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05): volume 1. [S.l.]: Ieee, 2005: 886-893.

[47] HE D C, WANG L. Texture unit, texture spectrum, and texture analysis[J]. IEEE transactions on Geoscience and Remote Sensing, 1990, 28 (4): 509-512.

[48] SHORTEN C, KHOSHGOFTAAR T M, FURHT B. Deep learning applications for covid-19[J]. Journal of big Data, 2021, 8 (1): 1-54.

[49] SHI B, GRIMM L J, MAZUROWSKI M A, et al. Prediction of occult invasive disease in ductal carcinoma in situ using deep learning features[J]. Journal of the American College of Radiology, 2018, 15 (3): 527-534.

[50] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. nature, 2015, 521 (7553): 436-444.

[51] WU Y, HE K. Group normalization[C]//Proceedings of the European conference on computer vision (ECCV). [S.l.: s.n.], 2018: 3-19.

[52] IOFFE S. Batch renormalization: Towards reducing minibatch dependence in batch-normalized models[J]. Advances in neural information processing systems, 2017, 30.

[53] HUANG G, SUN Y, LIU Z, et al. Deep networks with stochastic depth[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14. [S.l.]: Springer, 2016: 646-661.

[54] 谢向辉, 杨吉江, 李明磊, 等. 儿童肾积水诊断的临床思维辅助导向系统研究[J]. 中国数字医学, 2021.

攻读硕士期间的主要研究成果

1.**P. Wen** et al., "A-PSPNet: A novel segmentation method of renal ultrasound image," 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Melbourne, Australia, 2021, pp. 40-45, doi: 10.1109/SMC52423.2021.9658740.

（录用 EI 源）

2.Y. Guan, **P. Wen**, J. Li and Z. Ma, "A DWT-Utilized Classifier for UPJO Diagnosis Using Ultrasound Images," 2022 IEEE International Conference on Networking, Sensing and Control (ICNSC), Shanghai, China, 2022, pp. 1-6, doi: 10.1109/ICNSC55942.2022.10004150.

（录用 EI 源）

致谢

伴随着毕业论文的撰写完成，我在北工大的学习和生活也即将告一段落，意味着漫漫二十六载，我的人生也即将告别校园。转眼间，我已经在北工大度过了三年的时光，虽然相比人生漫漫长路不值一提，但校园之路却一直走到现在，占据着目前为止我所有的时光。出身于中农民阶级家庭，从小被父母注入望子成龙的夙愿，替父母也替自己完成学业，历经了寒窗苦读也经历了高考的洗礼，本科的最后一年是早六晚十的一年，并让我考上了北京工业大学的软件工程专业。一路走来有亲情、有友情也有爱情，有欢笑也有泪水，很庆幸自己的人生观、价值观以及感情观没有走向歪路，成功塑造了满意的自己并即将投身祖国的建设。这么多年的美好的日夜，我向自己的所有亲人、朋友、老师、爱人表达最诚挚的感谢，我无法祝愿所有人一帆风顺，但我在此虔诚地祝福大家不断高悬。

首先我想要感谢我的研究生导师李建强老师。很遗憾本科阶段没有在北工大就读，没能听过李老师的课程，但是研究生阶段非常幸运的成为了李老师的学生，刚见到李老师就感觉到了李老师在科研上的丰富经历，给了我这个刚入门的菜鸟研究生一个崇高的积极榜样，只要努力就一定能有收获，很幸运能够加入 415 实验室和实验室的师兄师弟一起学习，一起努力，一起进步。三年的学习和生活，是李老师的悉心教导，分享科研方法、做人做事的注意事项并引导我入门，指引我成长，使我从一开始对科研的一头雾水，到现在能够独立的完成科研项目。同时，也要感谢李老师在我撰写论文的过程中不断的指正与建议，在个人的发展上，李老师也是非常关心，尽自己力所能及的帮助我，虽然很遗憾没能听取老师的建议继续攻读博士，但我仍然非常感谢李老师的真诚建议，我想说能够在李老师的教导下度过研究生的三年时光，我感到非常的幸运，这三年的时光将伴随我一生。

感谢北工大期间的朋友，一起熬夜科研、一起唱过歌，一起野餐游戏、一起唱过歌，同样也感谢过去学业时光里我的朋友们，与大家度过的校园时光是我人生中不能磨灭的记忆。想与所有的朋友包括女朋友说一句话，爱人是路，朋友是树，人生只有一条路，一条路上多棵树，风光的时候别迷路，低迷的时候靠靠树，幸福的时候莫忘路，休息的时候浇浇树。

最后，我想感谢我的父母，多年早出晚归，对我无私地付出，从来没有要求过回报。要说的太多，但感觉无论说了多少都无法完整表达他们的辛酸和对我关心与照顾，从今往后将独身在北京奋斗，希望有朝一日能驾着彩云回到家乡，“今当远离，临表涕零”，不知所言。最后，感谢生命中所有的相遇。让我的人生增添色彩，让我的生活不再平淡