

单位代码	10445
学号	2020317096
分类号	TP391.9

# 山东师范大学

## 硕士专业学位论文

### 基于深度学习的甲状腺切除术前辅助诊断 方法研究

**Research on preoperative auxiliary diagnosis method  
of thyroidectomy based on deep learning**

**学位类别：电子信息硕士**

**领 域：计算机技术**

**学习方式：全日制**

**研 究 生：杨青翰**

**指导教师：张宇昂 副教授**

**提交时间：2023年6月**

# 目 录

第一章 绪论.....	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	2
1.3 本文主要研究内容及创新 .....	3
1.4 论文结构安排.....	4
第二章 颈部甲状腺超声影像及图像处理技术 .....	5
2.1 引言.....	5
2.2 超声成像原理及颈部甲状腺超声特点 .....	5
2.2.1 超声成像原理.....	5
2.2.2 超声中恶性甲状腺结节及甲状旁腺的特点 .....	6
2.3 数据处理.....	6
2.3.1 图像裁剪.....	7
2.3.2 图像标注.....	8
2.3.3 数据增强.....	8
2.4 传统图像分割方法.....	9
2.4.1 基于图像灰度值的分割方法.....	9
2.4.2 基于边缘检测的分割方法.....	9
2.4.3 基于区域的分割方法.....	10
2.5 基于深度学习的图像分割方法 .....	10
2.5.1 卷积神经网络.....	10
2.5.2 全卷积网络模型.....	14
2.6 本章小结.....	16
第三章 基于双通路镜像 U-Net 深度神经网络的恶性甲状腺结节影像 分割方法 .....	17
3.1 引言.....	17
3.2 双通路镜像 U-Net 网络(DMU-Net) .....	18
3.2.1 网络架构.....	18
3.2.2 金字塔注意力模块(PAM).....	19
3.2.3 边缘精炼模块(MRM) .....	20
3.2.4 融合模块(AM).....	21
3.3 实验与分析.....	22
3.3.1 评价指标.....	22
3.3.2 模型分割性能.....	23
3.3.3 提取边缘细节能力分析.....	25

3.3.4 消融实验.....	26
3.4 本章小结.....	27
第四章 基于“教师伙伴”策略的术前甲状腺分割的早期-晚期多阶段网络 .....	28
4.1 引言.....	28
4.2 早晚多阶段网络(EL-Net).....	30
4.2.1 早期阶段骨干网络.....	31
4.2.2 裁剪阶段.....	33
4.2.3 晚期阶段骨干网络.....	33
4.2.4 边缘感知模块(EAM) .....	34
4.2.5 还原操作 .....	35
4.3 “教师伙伴”策略 .....	35
4.4 损失函数.....	36
4.5 实验与分析.....	37
4.5.1 模型分割性能.....	37
4.5.2 消融实验.....	39
4.6 本章小结.....	42
第五章 总结与展望 .....	44
5.1 本文总结.....	44
5.2 本文展望.....	45
参考文献.....	46

## 摘要

随着甲状腺结节的发病率逐年上升，甲状腺结节已经成为了世界上最流行的疾病之一，有很多患者需要手术介入治疗。超声检查是甲状腺疾病诊断过程中关键的一步，超声检查的诊断结果决定着患者下一步治疗方式的选择和手术方案的制定。超声中恶性甲状腺结节特征不明显，容易与周围组织混淆，并且由于医生个人经验的不同，误诊和漏诊的情况时有发生。此外，对于被确诊为恶性的甲状腺结节，通常需要使用甲状腺切除术对患者进行治疗。然而，由于甲状腺周围存在体积较小，位置不固定的甲状旁腺（尤其是下甲状旁腺），使得在甲状腺切除术中经常导致甲状旁腺的受损或意外切除，造成严重的医疗事故。术前对超声中恶性甲状腺结节和甲状旁腺进行定位分割，有利于辅助医生识别结节和术中保护甲状旁腺。综上，甲状腺切除术前计算机辅助诊断方法的研究在临床上有着十分重要的意义。

本文针对颈部超声影像中不明显的恶性甲状腺结节和甲状旁腺，研究了基于深度学习的甲状腺切除术前辅助诊断技术，旨在实现甲状腺切除术前对恶性甲状腺结节和甲状旁腺的准确分割，以辅助医生诊断和治疗。本文的主要贡献和成果如下：

（1）完成了相关数据集的构建。本文构建了恶性甲状腺结节分割数据集和下甲状旁腺分割数据集，并在专业医生的支持下完成了数据集的标注工作及数据预处理操作。

（2）研究了用于术前超声中恶性甲状腺结节分割的卷积神经网络模型。本文提出的模型使用了双通路的网络结构，分别是 U 形子网和转置 U 形子网。U 形子网用于提取超声图像的上下文特征，以便定位图像中的恶性甲状腺结节；转置 U 形子网用于捕捉恶性甲状腺结节更多的细节特征，以更好地在复杂背景中分割恶性甲状腺结节。所有的实验结果都证明了所提出的模型在超声图像中恶性甲状腺结节分割任务上的优越性。

（3）研究了用于术前超声中甲状旁腺分割的神经网络模型。本文提出的模型使用了早-晚多阶段架构。其中模型早期阶段使用 Transformer 网络作为模型的主干网络，其强大的全局特征提取能力可以从复杂的超声图像中提取丰富的语义信息以分割甲状旁腺。同时，本文引入了裁切操作，从模型早期阶段的分割结果中提取目标区域的坐标信息，用于裁剪

输入的超声图像、对应的标签和早期阶段的特征图，并作为模型晚期阶段的输入。模型晚期阶段使用特征增强网络作为该阶段的主干网络，对甲状旁腺进行更精细的分割。此外，为了将早期阶段 Transformer 提取到的各像素间的结构化信息传送给晚期阶段，本文还提出了“教师-伙伴”训练策略。该方法在自建的甲状旁腺数据集上取得了优异的分割效果。

**关键词：**深度学习；医学影像分割；恶性甲状腺结节；甲状旁腺

**中图法分类：**TP391.9

## ABSTRACT

Recently, as the incidence of thyroid nodules increases, thyroid nodules have become one of the most prevalent diseases in the world. Thus, many patients require surgical intervention. Ultrasound examination is an essential step in the diagnosis of thyroid disease. The diagnostic results of ultrasonography determine the following treatment method and the formulation of the surgical plan for the patient. The characteristics of malignant thyroid nodules on the ultrasound image are not obvious, and it is prone to be confused with the surrounding normal tissues. Meanwhile, due to the differences in individual experience of clinicians, the misdiagnosis and missed diagnosis occurs frequently. In addition, patients with malignant thyroid nodules are usually treated with thyroidectomy. However, due to the presence of small and unfixed parathyroid glands (especially inferior parathyroid glands) around the thyroid gland, parathyroid glands are prone to damage or accidental remove during thyroidectomy, resulting in serious medical accidents. Preoperative localization and segmentation of malignant thyroid nodules and parathyroid glands in ultrasound can help doctors identify nodules and protect parathyroid glands during surgery. In summary, the study of computer-aided diagnosis methods before thyroidectomy has important clinical significance.

Aiming at the malignant thyroid nodules and parathyroid glands that are not obvious in neck ultrasound images, this thesis studies the preoperative auxiliary diagnosis methods based on deep learning for thyroidectomy, so as to achieve accurate segmentation for malignant thyroid nodules and parathyroid glands before thyroidectomy. The main contributions and results of this thesis are as follows:

(1) The construction of the relevant dataset is completed. In this thesis, a malignant thyroid nodule segmentation dataset and a lower parathyroid gland segmentation dataset are constructed. Meanwhile, the annotation work and data preprocessing are completed under the supervision of professional doctors.

(2) The convolutional neural network model for malignant thyroid nodule segmentation in

the preoperative ultrasound image is studied. The model proposed in this thesis uses a dual-route structure, namely the U-shaped subnet and the inversed U-shaped subnet. The U-shaped subnet is used to extract the context features of the ultrasound image to locate the malignant thyroid nodules in the ultrasound image. The inversed U-shaped subnet is used to capture more detailed features of malignant thyroid nodules to finely segment malignant thyroid nodules in the confused background. All the experimental results demonstrate the superiority of the proposed model on the task of thyroid nodule segmentation in ultrasound images.

(3) The neural network model for parathyroid gland segmentation in preoperative ultrasound image is studied. The model proposed in this thesis uses an early-late multi-stage architecture. In the early stage, the Transformer model is used as the backbone, and its powerful semantic feature extraction ability can extract rich semantic information from the confused ultrasound image to segment parathyroid glands. At the same time, the cropping operation is introduced, the coordinate information of the target region is extracted from the segmentation results of the early stage, and the input ultrasound image, corresponding labels and intermediate feature maps are cropped and transferred to the late stage. In the late stage, the FR-Subnet is used as the backbone network in this stage to accurately segment the parathyroid gland from the cropped ultrasound image. In addition, to transfer the inter-pixel structured information extracted by the Transformer backbone in the early stage to the late stage, we also propose a "teacher-mate" training strategy. The proposed method achieves excellent segmentation results on the self-built parathyroid dataset.

**Key words:** Deep Learning; Medical Image Segmentation; Malignant Thyroid Nodules; Parathyroid Gland

**Classification:** TP391.9

# 第一章 绪论

## 1.1 研究背景及意义

甲状腺结节是临床上最常见的疾病之一。据统计，甲状腺结节的发病率为 19%~67%，其中恶性甲状腺结节发病率为 5%~15%<sup>[1]</sup>。甲状腺结节通常情况下无法通过触诊进行准确诊断，大多需要超声检查来辅助临床医生判断甲状腺结节的恶性程度，进而确定治疗方式。超声图像中甲状腺结节的边缘和形状是判断甲状腺结节类型的关键特征。其中，良性结节通常边界清晰，在超声中特征较为明显。相反，超声中的恶性甲状腺结节体积相对较小，形状不规则，且边界不明显，在超声中极易与周围正常组织混淆，造成误诊和漏诊。然而，由于临床医生个体经验的不同，导致不同医生针对同一病例可能给出的诊断结果也不尽相同。所以一个能对超声中恶性甲状腺结节进行准确分割的模型，可以辅助医生提高甲状腺结节的确诊准确性，降低误诊率和漏诊率。

甲状腺癌多发源于甲状腺滤泡上皮细胞，是常见的恶性肿瘤之一，其治疗方式多为手术切除。手术造成的甲状旁腺功能减退症是甲状腺切除术中最常见且最严重的并发症之一。其中暂时性甲状旁腺功能减退症和永久性甲状旁腺功能减退症的发病率分别为 14%~60% 和 4%~11%<sup>[2]</sup>。甲状旁腺功能减退症的临床表现多为肢体麻木、手脚刺痛、焦虑、痉挛，甚至呼吸肌麻痹，严重影响患者的生活质量甚至危及患者生命，同时也是甲状腺及甲状旁腺外科医患冲突的主要原因<sup>[3]</sup>。因此，如何在手术中准确辨识并有效保护甲状旁腺，是甲状腺及甲状旁腺外科医生面临的一个严峻挑战。

然而，甲状旁腺的颜色、外观、大小、质地与周围淋巴结、脂肪颗粒、异位甲状腺结节等组织十分相似，导致甲状旁腺在甲状腺切除术中很难鉴别<sup>[4]</sup>。此外，在甲状腺切除术中，无论是单侧甲状腺叶切除术还是甲状腺全切除术，几乎无一例外地要切除淋巴结及其周围的脂肪和结缔组织，这进一步增加了血供中断、甲状旁腺损伤和意外切除的发生率。

放射性核素显像(99mTc-MIBI)和 SPECT/CT 等技术在临床上检测甲状旁腺是有用的，但目前在甲状腺切除术前仍没有可靠的方法来识别甲状旁腺<sup>[5,6]</sup>。前期研究发现超声在甲状旁腺的鉴别中具有明显优势<sup>[68]</sup>：术前超声可辨识及定位双侧下甲状旁腺的几率为 36.9%，



术前超声可辨识及定位至少一侧下甲状旁腺的几率为 81.0%，术前超声对甲状旁腺的辨识和定位，对于术中保护甲状旁腺（尤其是下甲状旁腺）极为有利。它可以帮助甲状腺外科医生预先划定甲状旁腺的关键保护区域，节省外科医生的精力和手术时间，降低手术风险。同时能有效避免术后甲状旁腺功能减退引起的相关并发症，提高患者生活质量，减少医患矛盾。虽然超声在鉴别甲状旁腺方面具有明显优势，但甲状旁腺（尤其是下甲状旁腺）由于其位置多变，外部特征不明确，在超声图像中对其进行准确识别仍存在一定困难<sup>[7]</sup>。因此，对超声影像中甲状旁腺进行准确定位和分割，有助于临床医生在甲状腺切除术前确定甲状旁腺的位置，从而在术中对甲状旁腺进行有效保护，降低甲状旁腺意外损伤和切除的概率。

## 1.2 国内外研究现状

颈部甲状腺超声检查是甲状腺结节疾病诊断的关键步骤<sup>[8]</sup>，医生可以通过超声图像中甲状腺结节的边缘和钙化情况来判断甲状腺结节的恶性程度，也可以在甲状腺切除术前对位置不固定的甲状旁腺（尤其是下甲状旁腺）在超声中进行分割定位，以免在手术中损伤和误切甲状旁腺，造成医疗事故。目前，对于术前超声中甲状腺疾病的辅助诊断方法可以分为三类，分别是甲状腺结节良恶性分类，甲状腺结节检测和甲状腺结节分割。

近年来，使用基于深度学习方法的辅助诊断方法发展迅速，对于甲状腺结节良恶性分类任务来说，其使用特征提取器从复杂的超声影像中提取底层的纹理、几何信息，通过训练分类器来对甲状腺结节进行良恶性的分类。其分类的准确性主要取决于数据中有效信息的质量。为了提高网络从有限的的数据中尽可能提取到更多的有效信息，研究者们基于深度学习提出了很多模型和算法。然而，目前图像级别的甲状腺结节分类任务仅仅只能将甲状腺结节分类为良性和恶性两个类别，这种分类方式与临床上常用的 TI-RADS（Thyroid Imaging Reporting And Data System）分类标准差距较大，无法为医生提供有效的帮助。另外，恶性程度较低的结节与恶性程度较高的结节的特征差异较为明显，将甲状腺结节简单地划分为良性和恶性两种类别对临床医生帮助有限。

甲状腺结节检测方法旨在通过向系统中输入超声影像，将超声影像中疑似甲状腺结节的区域使用检测框标记出来，以辅助医生在复杂的超声影像中快速找到甲状腺结节。该方

法可以有效降低不同医生阅片时的主观性，降低工作强度。因此该方法受到了越来越多研究者的关注，尤其是基于深度学习的甲状腺结节检测方法。比如，Li 等人<sup>[9]</sup>，受目前基于深度学习的 Faster R-CNN 启发，研究开发了一种检测超声影像中甲状腺乳头状癌的方法。Song 等人<sup>[10]</sup>，在目标检测网络 SSD 的基础上进行改进，提出了自动的甲状腺结节检测方案。甲状腺结节检测方法可以准确、快速地在超声影像中定位出甲状腺结节的位置，可以有效辅助医生对甲状腺结节的诊断。

### 1.3 本文主要研究内容及创新

本文主要研究内容是使用基于深度学习的模型实现甲状腺切除术术前对超声影像中恶性甲状腺结节的分割和甲状旁腺的分割任务，从而辅助医生诊断和手术。为了验证本文中模型在处理恶性甲状腺结节和甲状旁腺分割的有效性，首先需要解决的是相关数据集的问题。通过与山东省立医院相关医生的合作，采集了患者颈部的超声影像用于本文中数据集的构建，并在医生的指导下完成了数据集的标注过程。在自建的恶性甲状腺结节数据集和甲状旁腺数据集上，测试相关模型，最后达到辅助诊断的目的。具体内容如下：

(1) 数据集中所用到的图片的质量，图片中病灶或目标的特征以及图片的数量都影响着模型的最终性能。在本文中，用的到所有数据都是使用超声机扫描得到的颈部甲状腺超声影像。其以灰阶亮度（brightness）的不同来对目标区域进行诊断，故也称为 B 超。为了完成恶性甲状腺结节分割和甲状旁腺分割两项任务，在医院及相关医生的帮助下完成了两项私有数据集的标注，分别是针对诊断阶段不明显的恶性甲状腺结节容易与周围组织混淆的问题构建的恶性甲状腺结节数据集（Malignant Thyroid Nodule Segmentation dataset, MTNS dataset）和针对甲状腺切除术术前使用超声对手术中极不明显且位置不固定的下甲状旁腺进行预先定位的甲状旁腺数据集（Parathyroid Gland Segmentation dataset, PGS dataset）。

(2) 超声中恶性甲状腺结节分割。恶性甲状腺结节在超声中极不明显，影响医生对病人的病情进行诊断。针对超声中恶性甲状腺结节极易与周围正常人体组织混淆，本文提出了双通路镜像 U-Net 模型对超声中恶性甲状腺结节进行准确分割。通过 U 形子网提取上下文特征，转置 U 形子网提取更多有关恶性甲状腺结节的边缘细节信息，将上下文信息和提取到的边缘细节信息融合，从而获得高精度的恶性甲状腺结节分割结果。

(3) 术前超声中甲状旁腺分割。在甲状腺切除术中，由于甲状旁腺体积非常小，且位置不固定（尤其是下甲状旁腺），导致其极易被误伤或者误切，造成医疗事故，给患者的健康造成严重影响。因此，为了在术中避免意外伤害脆弱的甲状旁腺，在本文中提出了早晚多阶段网络以在术前对甲状旁腺进行分割定位，以在术中准确找到甲状旁腺，避免误伤。早晚多阶段网络由两个阶段构成，分别是早期阶段和晚期阶段。早期检测目标区域，并将相关特征映射裁剪成小图像块。然后，晚期阶段根据裁剪后的小图像块进行精确的局部分割，并将裁剪后的分割结果恢复到正常大小。

## 1.4 论文结构安排

本文各章节的主要内容如下：

第一章“绪论”，主要说明本文的研究背景和意义，结合国内外研究现状引出本文的相关研究内容，同时给出了本文各章节的具体结构安排。

第二章“颈部甲状腺超声影像及处理技术”，介绍了超声影像和颈部甲状腺超声影像特点。并介绍了处理医学影像的相关方法。

第三章“基于双通路镜像 U-Net 深度神经网络的恶性甲状腺结节影像分割方法”，针对超声中容易与周围人体组织混淆的恶性甲状腺结节，提出双通路镜像 U-Net 对恶性甲状腺结节进行分割。

第四章“基于‘教师伙伴’策略的术前甲状旁腺分割的早期-晚期多阶段网络”，针对甲状腺切除术中甲状旁腺容易误切的问题，提出了术前在超声中检测分割甲状旁腺的早晚多阶段分割网络。

第五章“总结与展望”对全文的研究内容进行了总结概括，突出所提出方法在辅助诊断上的优势，同时也给出了不足和后续的研究方向。

## 第二章 颈部甲状腺超声影像及图像处理技术

### 2.1 引言

颈部超声检查目前已经成为临床上诊断甲状腺类疾病最基本的手段。其在术前病情诊断，术中引导手术都有十分广泛的应用。本文围绕甲状腺切除术前对恶性甲状腺结节的辅助诊断以及术前超声中甲状旁腺的定位分割展开，本章将重点阐述超声影像中颈部甲状腺的特点以及相关辅助诊断方法。

### 2.2 超声成像原理及颈部甲状腺超声特点

#### 2.2.1 超声成像原理

超声检查由于其技术成熟，安全性高，开销较小等特点一直被广泛应用于各种疾病的临床诊断过程中。以甲状腺结节诊断为例，医生通过触诊检测甲状腺结节患病率的概率为 3%~7%<sup>[1]</sup>，而超声检查中甲状腺结节的检出率为 20%~76%<sup>[1]</sup>。

超声检查是利用超声设备向人体发射频率在 1M 赫兹到几十 M 赫兹不等的超声波，并收集超声波经过人体不同组织和器官时产生的反射波，之后将其显示出来的过程。目前常见的超声检查方式有：

**A 型超声 (Amplitude Mode)：**A 型超声是最早应用于医疗行业的超声检测。其接受到的回声以振幅形式显示。常用于测量组织和器官的大小等方面，目前多用于眼科领域。

**B 型超声 (Brightness Mode)：**B 型超声又简称 B 超，其将不同深度反射回来的超声转变为亮度强弱不同的灰阶，并将其通过屏幕显示出来。这是目前应用范围最广的一种超声检查。

**M 型超声 (Motion Mode)：**M 型超声是在 B 型超声的基础上加上锯齿波的一种特殊形式。此模式主要显示组织回声界面的位置在不同时间变化的情况，目前主要应用于心脏领域。

**D 型超声 (Doppler Mode)：**多普勒型超声，即目前常说的彩超。它在超声波技术的基

础上，利用多普勒现象的原理，通过声波回声频率的不同，记录其中的变化，并将其转换成彩色信号。目前常用于血管和心脏的检查。

### 2.2.2 超声中恶性甲状腺结节及甲状旁腺的特点

目前临床上对于甲状腺结节的诊断主要分为超声检查和细针穿刺活检两步。虽然超声检查可以对甲状腺结节进行有效的检查，但是由于甲状腺超声图像的复杂性，临床上经常会出现误诊和漏诊的情况。颈部甲状腺结节超声影像的特点可以概括为以下几点：

（1）边界模糊。由于人体正常组织的浸润，导致超声中甲状腺结节的边界不明显。尤其是恶性甲状腺结节，其尺寸通常较小，特征不明显，极大地增加了医生辨认病灶区域的难度，很容易造成误诊和漏诊。

（2）容易与周围组织混淆。在超声影像中，甲状腺的周围分布着特征与甲状腺结节十分相似的气管，血管等各种复杂的组织，且由于个体的差异性，导致这种干扰没有规律可循，只能依靠医生的经验对其进行辨别判断，严重增加了医生的负担。

（3）噪声。由于人体结构的复杂性，超声波进入人体后会不可避免地造成不同程度的折射、反射等光学现象，进而产生大量干扰性噪声。不可避免地噪声极大的影响了超声的成像质量，影响医生诊断。

类似地，甲状旁腺由于其体积较小（与黄豆粒大小类似），且位置不固定（尤其是下甲状旁腺），其在超声中的特征格外不明显。图 2-1 为超声中的恶性甲状腺结节和甲状旁腺。

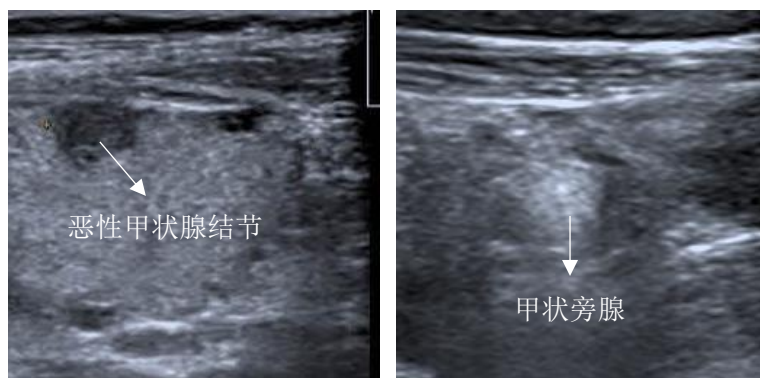


图 2-1 超声中的恶性甲状腺结节和甲状旁腺

## 2.3 数据处理

本文为了完成超声中的恶性甲状腺结节分割和甲状旁腺分割，共采集并完成了相应数

数据集的制作。为了保证数据的质量并有效对所提出模型能进行有效训练。本节将详细介绍相关数据的处理过程，包括图像裁剪，图像标注和数据增强几个方面。

### 2.3.1 图像裁剪

本文所使用的数据皆由山东第一医科大学附属山东省立医院乳腺甲状腺外科提供，其中恶性甲状腺数据集的图片来自 70 位患有恶性甲状腺结节的病人，甲状旁腺数据集的图片来自于 250 位患者。所有超声图片皆由超声探头为 5-14MHZ 的东芝 Aplio500 超声机采集。

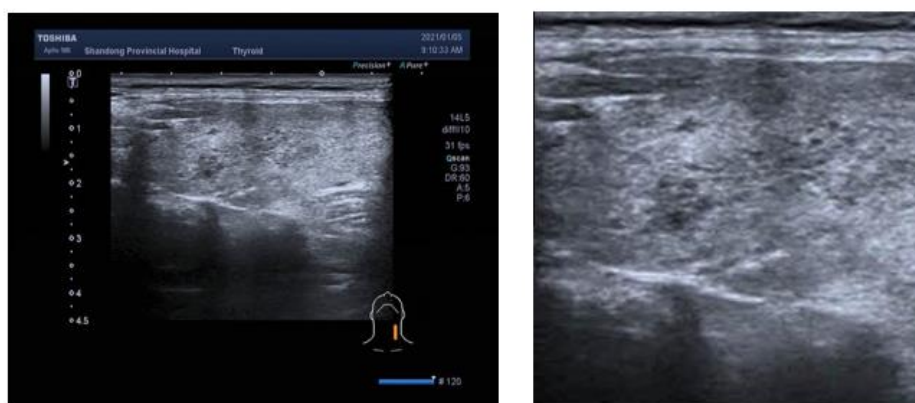


图 2-2 超声中恶性甲状腺结节和裁切后的图像

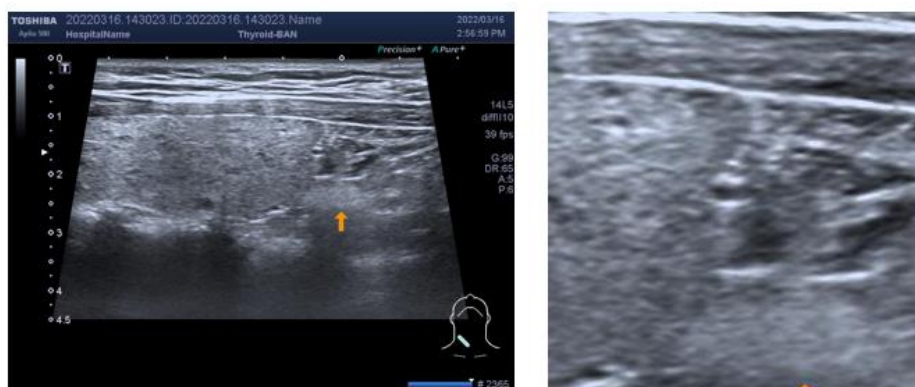


图 2-3 超声中下甲状旁腺和裁切后的图像

如图 2-2 和图 2-3 所示，超声图像四周边缘存在很多文字信息，比如检查日期，设备名称等等。为了移除这些干扰信息，本节对采集的恶性甲状腺结节和甲状旁腺超声图片进行了裁剪。同时为了将图片顺利输入到设计的模型中，本节固定了裁剪框的大小。其中恶性甲状腺结节裁剪框设置为  $256 \times 256$  分辨率，甲状旁腺裁剪框设置为  $224 \times 224$  分辨率。

### 2.3.2 图像标注

本文所提出的方法都属于监督学习，因此数据集中的每张图片都需要人工标注。由于图像分割任务本质上是对输入图像每个像素点进行分类，因此需要对目标区域（恶性甲状腺结节/甲状旁腺）的像素归为目标类，将其余区域的像素归为背景类，如图 2-4 所示。

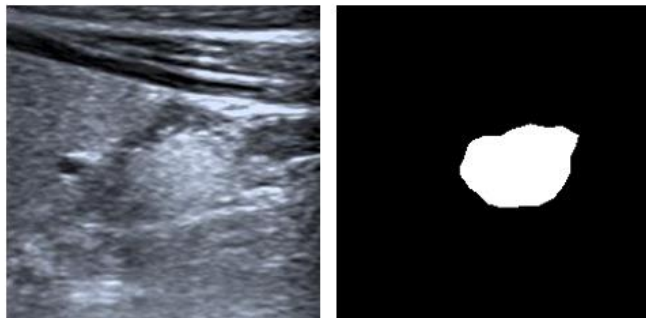


图 2-4 裁切后超声中的甲状旁腺和对应二值分割标签

不同于常见的自然图像，医学影像尤其是超声图像中，存在大量干扰性的造成和组织，同时恶性甲状腺结节和甲状旁腺边缘等特征不明显，因此，图像的标注过程必须由专业的医生帮助和指导。本文中，数据集的标注过程使用的软件为 LabelMe，标注结果中像素值为 0（黑色）的像素为背景，像素值为（255）的像素为目标类（恶性甲状腺结节/甲状旁腺）。

### 2.3.3 数据增强

为了增加恶性甲状腺结节数据集和甲状旁腺数据集中训练集的数据量，尽可能避免模型过拟合情况的发生，本节对采集的图像进行了数据增强，包括图像平移、图像旋转、图像翻转，如图 2-5 所示。其中图 2-5（a）经过平移、旋转及翻转变换得到图 2-5（b）（c）（d），图 2-5（b）在图 2-5（a）平移后的基础上旋转 90 度，图 2-5（c）在图 2-5（a）平移后的基础上旋转 180 度，图 2-5（d）在图 2-5（a）平移后又经过了水平翻转。

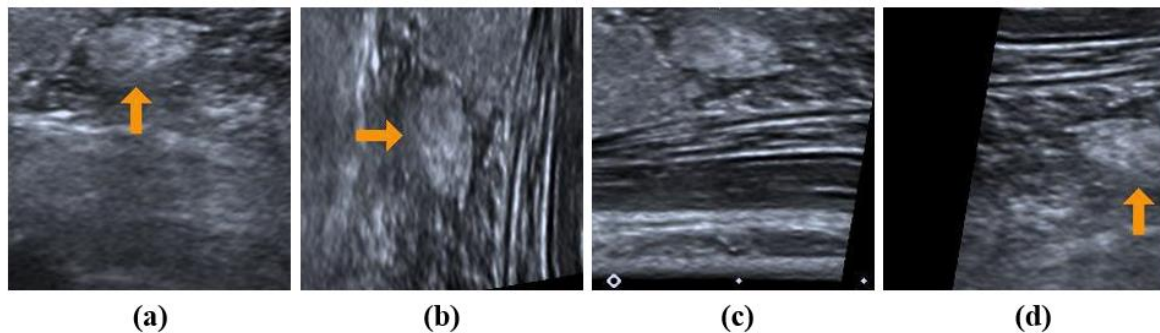


图 2-5 经过数据增强后的甲状旁腺超声图片



## 2.4 传统图像分割方法

传统医学影像分割方法又可以分为像素灰度值分割法、边缘检测分割法、统计学分割法、区域分割法、活动轮廓模型分割法等。

### 2.4.1 基于图像灰度值的分割方法

像素灰度值分割法是一种仅仅依靠像素值的大小来确定医学影像中感兴趣区域（Region of interest, ROI）的传统影像分割算法，其最具代表性的算法是阈值分割法。顾名思义，阈值分割法就是通过手动设定阈值，将医学图像中在阈值范围内的像素点标记为目标区域，剩下区域标记为背景区域。这种分割方法因其实现简单、计算量小、性能较稳定的特点，成为了最早的分割方法之一。根据阈值选择方法的不同，该方法又可以分为人工经验选择法、直方图法和最大区间方差法（OTSU）<sup>[67]</sup>。其中，人工经验选择法就是根据需要处理的图像的先验知识，对医学影像中目标区域和背景区域进行分析，选择出目标区域像素值所在区间，并通过不断实验，确定分割效果较好的阈值区间。这种方法效率较低，且需要耗费较多人力成本；直方图法是利用直方图对影像进行分析，根据直方图波峰和波谷之间的关系进而选择分割效果较好的阈值；最大区间方差法利用最大类间方差确定阈值的方法，当取最佳阈值时，两部分之间的差别应该是最明显的，前景和背景之间的类间方差如果越大，就说明构成图像的两个部分之间的差别越大，当部分目标被误分为背景或部分背景被误分为目标，都会导致两部分差别变小。尽管这种方法得到了广泛的使用，但是由于医学影像中噪声的影响，导致其分割效果受限。

### 2.4.2 基于边缘检测的分割方法

人类的视觉系统是对物体的边界信息十分敏感的，因此基于边缘检测的分割方法得以提出和发展。图像中物体边缘周围的像素的值往往存在突变，因此，该类方法主要是根据目标边缘像素值的突变来对目标区域进行分割<sup>[50]</sup>。

这种方法理论上可以快速定位目标的边缘，进而对目标进行分割，但是如果图像本身容易产生混淆的信息和噪声较多，则会对分割结果造成较大影响。由于颈部甲状腺超声中噪声和相似组织较多，因此该方法不适合超声中恶性甲状腺分割和甲状旁腺分割任务。



### 2.4.3 基于区域的分割方法

与基于边缘检测的分割方法类似，基于区域的分割方法同样利用了目标和背景的相似性来分割目标区域。以区域生长算法<sup>[66]</sup>为例，其通过种子和合适生长策略的选取来不断聚合与种子相似的像素区域，最终实现对目标区域的分割。此算法的关键是种子的选取和相似区域判断标准的设定。该方法由于需要预先选择种子，因此其分割过程需要人为干预，其分割结果在一定程度上也受人为干预的影响。此外，该方法同样对噪声敏感，造成分割不准确，因此其更适用于像素灰度值变化均匀的场景，不适用于超声中恶性甲状腺分割和甲状旁腺分割任务。

## 2.5 基于深度学习的图像分割方法

上一节中所介绍的传统医学影像分割方法大多需要人工介入，无法有效降低医生工作量，且易受噪音影响导致分割精度较低，因此亟需一类抗噪声能力强，分割精度高且能有效降低医生工作量的自动分割方法。

近年来，随着计算机硬件水平的不断提高，使用深度学习技术对复杂的医学影像进行分割受到越来越多研究者的广泛关注。深度学习逐层提取图像特征，不同层提取特征的侧重点不同，通常，深度学习模型浅层用于提取图像的纹理、边缘等细节信息，而模型深层用于提取图像的抽象特征和语义信息。这种从浅层到深层的特征提取方式使其能够很好地完成各种复杂场景下的医学图像分割任务。

### 2.5.1 卷积神经网络

卷积神经网络（Convolutional Neural Network, CNN）通常应用于自然图像及医学影像处理的相关工作，其灵感来源于视觉感知<sup>[54]</sup>，使用人工神经元来模拟生物神经元以完成各种复杂任务。CNN 是一种前馈神经网络，能够从具有卷积结构的数据中提取特征。与传统的特征提取方法<sup>[51][52][53]</sup>不同，CNN 提取特征的过程中并不需要人工介入。总体来说，CNN 与全连接网络相比，CNN 具有以下几个优势：（1）局部连接：每个神经元不再与上一层的所有神经元连接，而只与少量神经元连接，可以有效降低参数，加快收敛速度。（2）权重共享：一组连接可以共享相同的权重，进一步降低了模型参数量。（3）下采样：使用池化层对图像进行下采样，在保留有用信息的同时减少数据量。同时，下采样操作还可以通过

去除大量冗余的特征来减少参数量。正是由于 CNN 的这些特性，才使得其成为深度学习领域最具代表性的算法之一。卷积神经网络通常由卷积层，池化层，激活函数和全连接层四部分构成。

### (1) 卷积层

卷积操作是模型提取特征过程中的关键步骤，其输出可以被称为特征图 (feature map)。当使用固定大小的卷积核来提取特征时，通常会丢失边界处的图像特征，此时通常在特征图周围使用填充操作(padding)来调整特征图大小。此外，通过设置卷积操作的步长(stride)来控制卷积的密度。步长越大，代表密度越低；步长越小，代表密度越高。模型在提取特征时，不同的卷积层提取的特征也有所不同。其中，浅层卷积层擅长于提取图像中的低级特征，例如纹理和边缘等；深层卷积层则更擅长提取抽象特征和语义特征用于更准确地完成下游任务。

具体来说，卷积操作可以视为卷积核在图像上不断滑动，并不断加权求和的过程。如图 2-6 所示。通常情况下，卷积操作都是在二维图像上完成的，对于医学影像中的三维影像，可以将二维的卷积操作拓展为三维卷积以对图像进行特征提取。

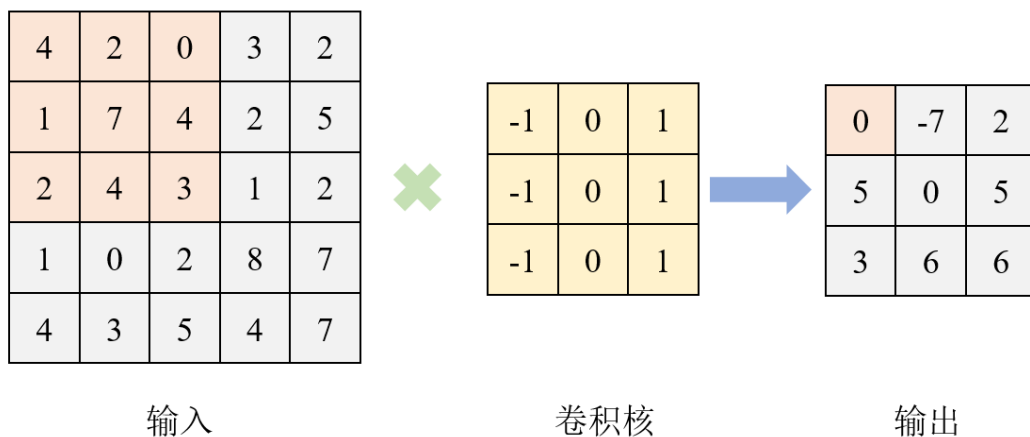


图 2-6 卷积运算示意图

### (2) 池化层

在特征提取的过程中，为了进一步提取更深层的上下文信息并去除过多冗余信息，通常在卷积层之后使用池化层对提取到的特征进行下采样。与卷积操作类似，池化操作同样是在特征图上进行“滑动”，对每一个池化覆盖区域进行降维。池化操作一般分为最大池化 (Max Pooling) 和平均池化 (Average Pooling)。其中，最大池化意味着取特征图的池化覆

盖区域的最大值，而平均池化则是取特征图池化覆盖区域的平均值，如图 2-7 所示。

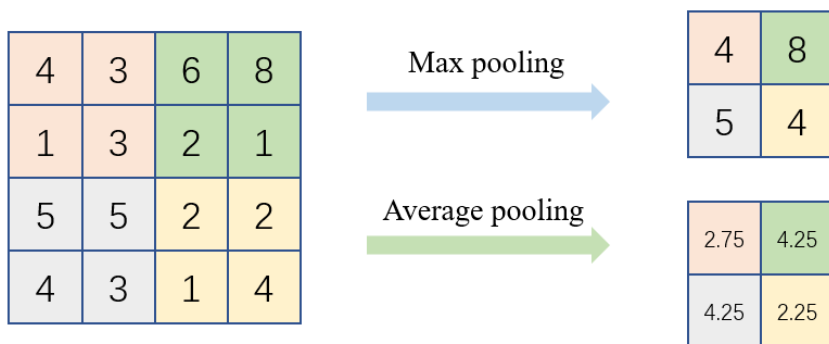


图 2-7 最大池化和平均池化示意图

### (3) 激活函数

CNN 可以利用不同的激活函数来表达复杂的特征。在多层神经网络中，两层之间存在一个函数，此函数称为激活函数。激活函数决定着卷积操作提取到的哪些特征和信息应该传递到下一个神经元。神经网络中的每个神经元接受上一层神经元的输出值作为输入，并将处理后的值传递给下一层。

如果神经网络模型没有使用激活函数，则每一层的输入将是前一层输出的线性函数。在这种情况下，He 等人<sup>[55]</sup>验证了不管神经网络有多少层，输出始终是输入的线性组合，这意味着如果没有激活函数，整个神经网络模型只相当于做了一次简单的线性变换。这也是原始的感知器<sup>[56][57]</sup>，它的学习能力有限。为此，引入非线性函数作为激活函数。理论上，具有非线性激活函数的深度神经网络可以拟合任何函数，大大提高了神经网络对数据的拟合能力。以下是卷积神经网络模型中常用的几个激活函数。

**Sigmoid 激活函数。**Sigmoid 函数整体呈 S 形，是最经典的非线性激活函数之一，如图 2-8 所示。其数学表达如公式 (2-1) 所示。当  $x$  值趋于 0 时，梯度变得更陡，随着  $x$  值增大或减小，其梯度也变得相对更平稳。Sigmoid 函数可以将一个实数映射到  $(0, 1)$ ，因此，它可以用于二分类问题。

$$Sigmoid = \frac{1}{1+e^{-x}} \quad (2-1)$$

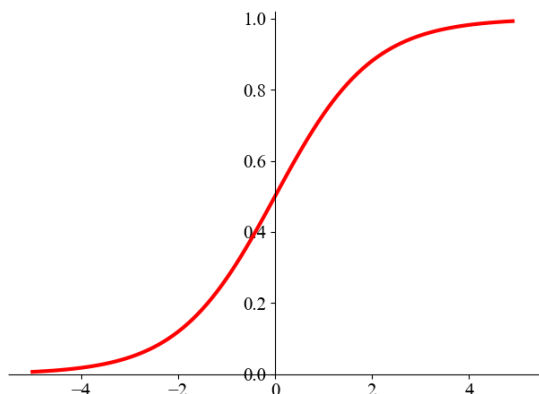


图 2-8 Sigmoid 激活函数

Tanh 激活函数。与 Sigmoid 激活函数类似，Tanh 激活函数的梯度在  $x$  轴的两端较小，而在  $x$  轴中间较大，如图 2-9 所示。与 Sigmoid 不同的是其可以将输入的  $x$  值映射到  $(-1, 1)$ ，因此 Tanh 的输出均值为 0，可以实现归一化，更便于网络学习。其数学表达如公式 (2-2) 所示。

$$\text{Tanh} = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad (2-2)$$

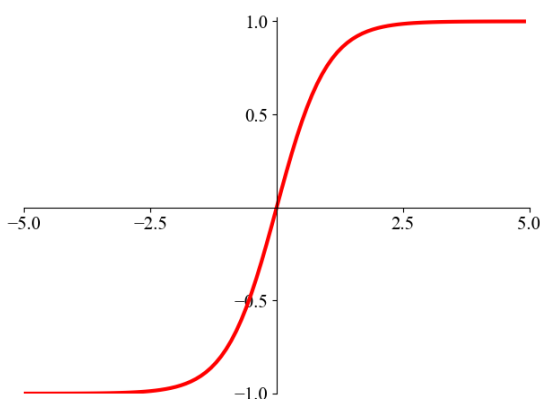


图 2-9 Tanh 激活函数

ReLU 激活函数。ReLU 是一种有效且常用的激活函数。当  $x < 0$  时，其输出值为 0；当  $x > 0$  时，其函数输出值为  $x$ ，如图 2-10 所示。其数学表达如公式 (2-3) 所示。在 Sigmoid 激活函数和 Tanh 激活函数中，在  $x$  值较大或者较小时，其梯度很小，这导致了训练时模型收敛较慢。相比之下，ReLU 激活函数最大的优势在于其梯度为常数，因此其可以有效加快网络的训练速度

$$\text{ReLU} = \max(0, x) \quad (2-3)$$

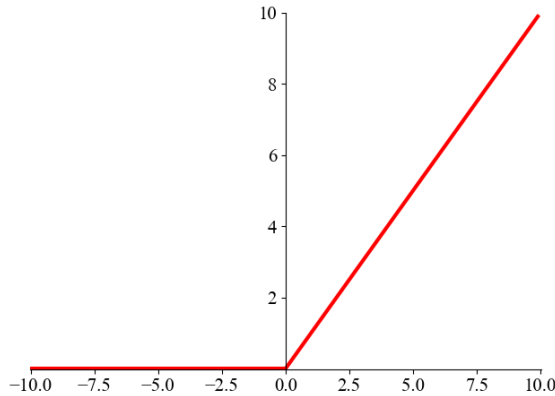


图 2-10 ReLU 激活函数

#### (4) 全连接层

全连接层通常用于卷积层、池化层和激活函数之后，对上层提取到的特征进行整合。在卷积神经网络中，全连接层通常以上层  $N$  个尺寸为  $H \times W$  的特征图作为输入，并将其转换为一维特征  $1 \times (N \times H \times W)$ 。

Softmax 函数通常放在全连接层之后，对输出的特征向量进行归一化，已完成模型后续不同的下游任务。

### 2.5.2 全卷积网络模型

卷积神经网络通常应用于图像分类任务，已经在医学影像分类任务中取得了很好的效果。但在临床上，简单的图像级分类任务已经不能满足临床医生日益增长的需求。在背景复杂的医学影像中，对病灶区域或目标区域进行精确有效的分割被认为是相关疾病确诊和分析的必要步骤之一。

#### (1) FCN 模型。

Long 等人<sup>[58]</sup>在 2015 年首次提出了全卷积神经网络模型(Fully Convolutional Networks, FCN)，不同于基于图像的分类任务，FCN 将图像级的分类任务进一步延伸到了像素级分类任务，即图像分割领域。不同于传统的卷积神经网络，将提取到的二维特征通过全连接层转换为一维特征，进而得到图像级的分类结果，FCN 为了实现像素级的分类任务，使用卷积层替代了全连接层，因此不会将提取到的二维特征转换为一维向量，而是直接输出与输入图像相同尺寸的预测特征图，其输入图像的每一个像素点进行了准确的分类。图 2-11 展示了这一过程。

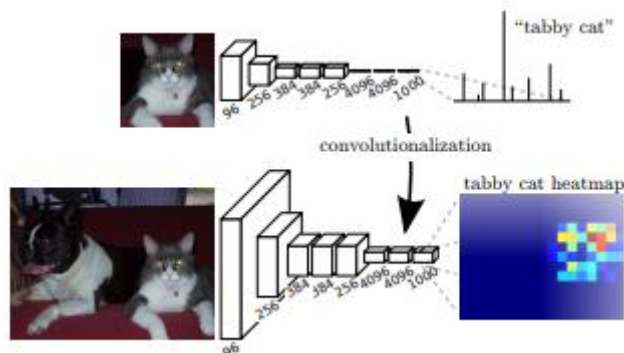


图 2-11 将卷积神经网络中的全连接层转化为卷积层的示意图

图中上半部分为用于图像分类任务的卷积神经网络模型，其中的小立方体块代表卷积操作，最后的三个线条则代表着全连接层。全连接层将提取到的特征转换为一个代表着 1000 个类别的输出向量，将其中概率最大的作为最终的输出结果。图中下半部分为 FCN 网络，其用 4096 个  $7 \times 7$  大小的卷积核替代了卷积神经网络的第一个全连接层，用两个  $1 \times 1$  卷积替代了后两个全连接操作。此外，全卷积网络还在网络的浅层和深层之间使用的跳跃连接，以得到更加精确的特征。最后将提取到的特征进行上采样恢复到与原始输入相同的尺寸，以对每一个像素都进行分类预测。

## (2) Unet 模型

在生物医学影像分割领域，Ronneberger 等人<sup>[59]</sup>基于全卷积网络提出了迄今为止医学影像分割领域乃至图像分割领域最著名的 Unet 模型，如图 2-12 所示。

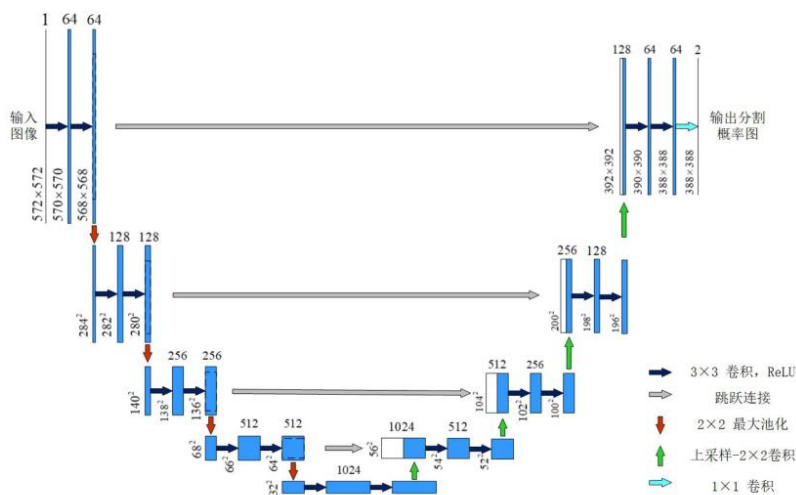


图 2-12 Unet 模型示意图

Unet 整体由 23 个卷积层组成，没有全连接层。Unet 可以分为三部分，分别是：编码器，解码器和跳跃连接。编码器部分如图 2-12 左侧所示。其由四个编码器层组成，每个编

码器层包括两个  $3\times 3$  卷积层，一个 ReLU 激活函数和一个  $2\times 2$  的最大池化层构成，用于提取输入图像的特征。以  $1\times 572\times 572$  大小的输入图像为例，当输入图像每经过一个编码器层时，其图像尺寸会减半，通道数增加一倍，最终得到  $512\times 32\times 32$  大小的特征图。在编码器的前几层，由于输入图像的尺寸较大，经过特征提取后特征图中通常会包含很多纹理和边缘信息，随着编码器深度的增加，图像尺寸变小，网络的感受野变大，此时图像会包含更多的语义信息，有利于对图像中的病灶进行准确的定位。Unet 的解码器部分如图 2-12 右侧所示，其由四个解码器层组成。其组成与编码器层类似，不同的是将编码器层的  $2\times 2$  的最大池化层替换为  $2\times 2$  的反卷积层。特征图通过每一个编码器层后，通道数减半，图像尺寸加倍，最后得到的预测特征图的尺寸与输入图像一致。Unet 的跳跃连接可以为输入解码器的特征图补充更多的细节信息，有利于最终分割结果的提高。

## 2.6 本章小结

本章对甲状腺超声影像和图像分割方法进行了简要介绍。首先，对医学上不同超声成像技术的原理进行介绍。之后，介绍了超声中恶性甲状腺结节和甲状旁腺的特点，包括超声中甲状腺结节边界模糊，周围组织容易造成混淆以及超声检查中不可避免的噪声，说明了对超声中的甲状腺结节和甲状旁腺进行分割的必要性。随后，介绍了传统的图像分割方法，对其优势和局限性进行了分析。最后，介绍了目前基于深度学习的图像分割方法，包括了卷积神经网络中使用的卷积层，池化层，全链接层和激活函数，并介绍了对医学影像分割影响最大的两个全卷积网络模型，FCN 和 Unet。

## 第三章 基于双通路镜像 U-Net 深度神经网络的恶性甲状腺结节影像分割方法

### 3.1 引言

近些年, 甲状腺癌<sup>[12][13]</sup>确诊人数逐年增加, 尤其在东亚地区的国家, 甲状腺癌的发病率在过去 30 年里从 200% 上升到 300%。大多数恶性甲状腺结节属于乳头状癌, 生长速度缓慢, 通过合理的治疗可得到有效的控制<sup>[14][15]</sup>, 手术切除是有效的治疗方式之一<sup>[17]</sup>。使用超声检查对甲状腺结节的恶性程度进行判断决定着治疗方式和手术方案的制定。超声中甲状腺结节的边缘及形态是判断甲状腺结节类型的关键特征。良性甲状腺结节通常边界清楚, 边缘光滑。相反, 恶性甲状腺结节形态不规则, 边界模糊, 周围血管丰富<sup>[18][19]</sup>。目前, 甲状腺分类和甲状腺结节分割任务已经取得了很多优秀的成果, 并能够帮助临床医生有效地识别甲状腺结节的类型。但是, 目前临床上对恶性甲状腺结节的诊断仍然存在问题。比如由于其容易与周围正常组织混淆, 导致诊断过程中造成误诊和漏诊。随着卷积神经网络 (CNN) 的快速发展, 一些基于 CNN 的方法在甲状腺结节分割任务上得到了广泛的发展。

比如, Ma 等人<sup>[11]</sup>在超声中使用深度 CNN 自动分割甲状腺结节。Ying 等人<sup>[20]</sup>设计了一种级联 CNN 网络来分割甲状腺结节。虽然此类基于深度学习的方法较传统医学影像分析方法已经有了很大的改进, 但这些方法主要针对的是界限清楚、特征明显的良性甲状腺结节, 忽视了超声中形态不规则、特征不明显的恶性甲状腺结节。为了准确定位并分割超声图像中的恶性甲状腺结节, 避免医生在诊断时将正常组织和恶性甲状腺结节混淆, 本文提出了一种基于深度学习的双路镜像 U-Net (Dual-route Mirroring U-Net, DMU-Net) 框架来对超声影像中的恶性甲状腺结节进行自动分割。DMU-Net 的主干是由 U 形子网和转置 U 形子网组成的双子网结构, 两个子网由汇聚模块 (Aggregation Module, AM) 连接。此外, 本章设计了一个金字塔注意模块 (Pyramid Attention Module, PAM), 它被添加到 U 形子网的跳连接中, 以提取更多的上下文信息。同时, 本章还设计了一个边缘精炼模块 (Margin Refinement Module, MRM) 以从特征图中获取更多关于恶性甲状腺结节的边缘细节特征。综上所述, 此方法的主要创新点如下:



(1) 提出了一种基于双通路的镜像 U 形分割网络 DMU-Net，它主要由 AM 模块连接的两个分支组成，分别是用于提取语义信息以定位病灶位置的 U 形分割子网分支和用于提取更多边缘细节信息的转置 U 形子网分支。该模型可以帮助医生更准确地在超声中诊断和辨别恶性甲状腺结节。

(2) 将 PAM 和 MRM 两个模块分别引入到两个子网中，用于从甲状腺结节超声图像中提取上下文信息和超声中甲状腺结节的边缘细节特征。

## 3.2 双通路镜像 U-Net 网络 (DMU-Net)

### 3.2.1 网络架构

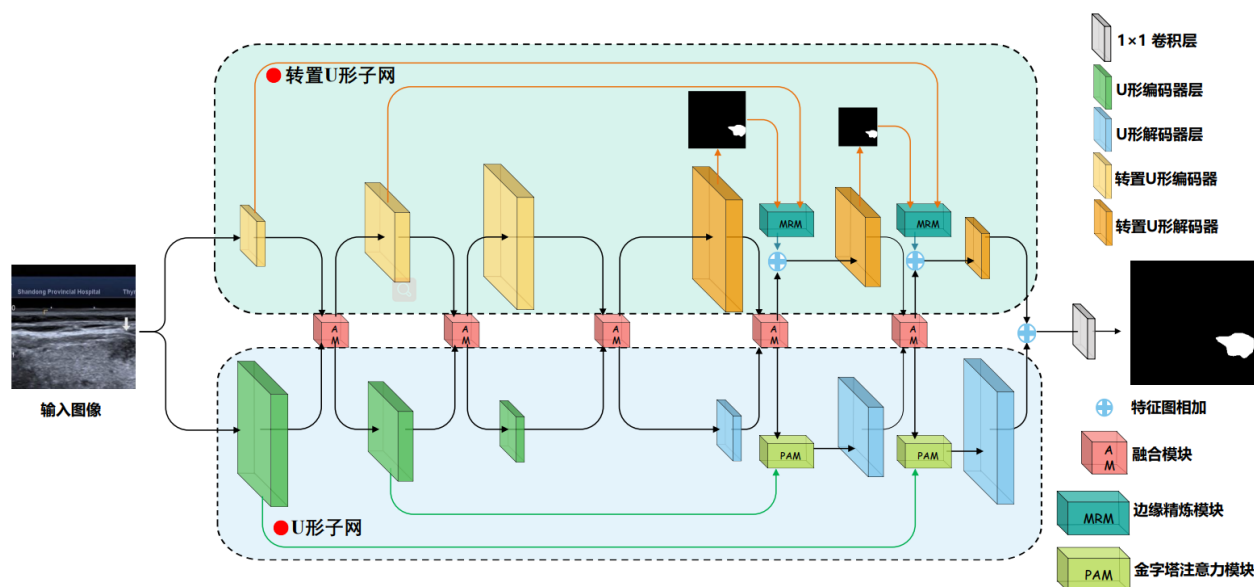


图 3-1 双通路镜像 U-Net (DMU-Net) 示意图

如图 3-1 所示，所提出的 DMU-Net 采用两个子网作为主干，分别称为 U 形子网和转置 U 形子网。前者侧重于从超声图像中提取语义信息，后者擅长于捕获病变区域的边缘特征。其中，MRM 模块和 PAM 模块用来辅助两个子网以捕获更多的语义信息和细节特征。此外，模型中还添加了 5 个 AM 模块用来融合提取到的不同特征。

U 形子网由三层编码器和三层解码器组成。编码器用于捕获特征图的语义信息，每个编码器层由  $3 \times 3$  Conv 层、最大池化层和 batch-normal 层组成，最后使用 ReLU 激活函数。解码器层的主要功能是恢复特征图的大小，每个解码器层由  $3 \times 3$  卷积层、双线性插值层和 batch-normal 层组成，然后再进行 ReLU 激活。此外，为了从 U 形子网解码器的特征图和

AM 的特征图中获得更密集的上下文信息，本文将 PAM 嵌入到 U 形子网中。与 U 形子网相反，转置 U 形子网的编码器通过逐层增加特征图的尺寸来增强超声中恶性甲状腺结节区域的边缘特征。在其解码器中，转置 U 形子网用最大池化层替换 U 形子网的双线性插值层，将增强后的特征图恢复到原始的大小。为了提取更多的边缘细节特征，在解码器恢复特征图尺寸的过程中，将 MRM 模块插入到转置 U 形子网的解码器中。具体来说，设  $X \in \mathbb{R}^{C \times H \times W}$  为输入图像， $C, H, W$  分别是输入图像的通道数，长度和宽度， $k \times k$  设为网络的初始感受野。在 U 形子网中，经过第一个 U 形编码器层后，感受野的大小增大到  $2 \times k \times 2 \times k$ ，特征图尺寸变为  $1/2 \times X$ （假设下采样系数为 2）。同样的，经过 U 形子网的第二层和第三层编码器层后，接收域尺寸增大到  $4 \times k \times 4 \times k$  和  $8 \times k \times 8 \times k$ ，特征图尺寸变为  $1/4 \times X$  和  $1/8 \times X$ 。相比之下，转置 U 形子网的编码器使用扩张系数为 2 的双线性插值层代替了最大池化层。因此，在经过转置 U 型子网的第一个编码器层后，其感受野的大小被压缩为  $1/2 \times k \times 1/2 \times k$ ，特征图的大小变为  $2 \times H \times W$ 。同样，经过第二层和第三层转置 U 形编码器层后，其感受野的大小分别缩小为  $1/4 \times k \times 1/4 \times k$  和  $1/8 \times k \times 1/8 \times k$ ，特征图的大小分别为  $4 \times H \times W$  和  $8 \times H \times W$ 。U 形子网和转置 U 形子网的解码器的主要目标都是通过其编码器的反向操作将特征图恢复到初始大小。总的来说，U 形子网具有较大的感受野，专注于从图像中提取上下文特征，而转置 U 形子网较小的感受野更擅长于捕获恶性甲状腺结节的更多细节特征。下面将对该模型涉及到的不同模块进行详细介绍。

### 3.2.2 金字塔注意力模块（PAM）

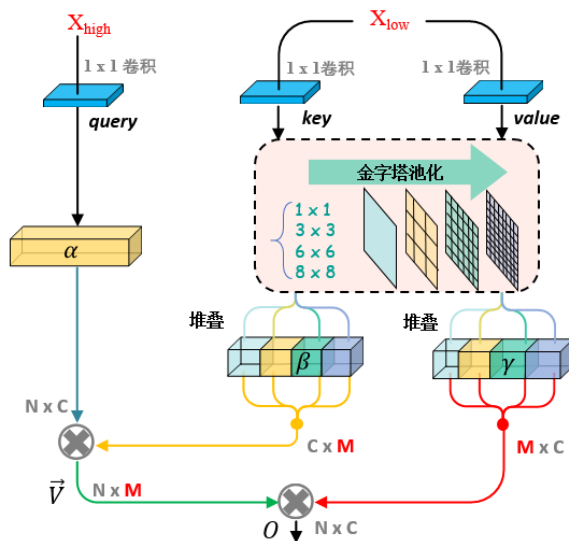


图 3-2 金字塔注意力模块（PAM）示意图

受 Non-local 注意力机制<sup>[22]</sup>和空间金字塔池化操作<sup>[23]</sup>的启发, 本章设计了金字塔注意力模块 (PAM), 以少量的计算代价获得特征图中不同像素之间的长程依赖关系, 如图 3-2 所示。具体来说, 首先将来自融合模块 (AM) 的高维度特征图  $X_{high} \in R^{C \times H \times W}$  通过  $1 \times 1$  卷积操作  $W_{query}$  得到特征向量  $query$ :

$$query = W_{query}(X_{high}) \quad (3-1)$$

之后将特征向量  $query$  拉伸到大小为  $N \times C$  维度的特征向量  $\alpha$ 。其中  $N$  是特征图中的像素点的个数,  $C$  代表通道数。相似地, 将来自 U 形编码器层的低维度特征图  $X_{low} \in R^{C \times H \times W}$  分别输入到两个  $1 \times 1$  卷积层  $W_{key}$  和  $W_{value}$  得到  $key$  和  $value$  两个特征向量:

$$key = W_{key}(X_{low}) \quad (3-2)$$

$$value = W_{value}(X_{low}) \quad (3-3)$$

然后, 将  $key$  和  $value$  两个特征向量输入到四个具有不同池化大小的金字塔池化模块中, 得到四个不同大小的特征向量  $\beta_1, \beta_2, \beta_3, \beta_4$ :

$$\beta_1, \beta_2, \beta_3, \beta_4 = P_\beta(key) \quad (3-4)$$

$$\gamma_1, \gamma_2, \gamma_3, \gamma_4 = P_\gamma(value) \quad (3-5)$$

其中,  $P_\beta$  和  $P_\gamma$  代表金字塔池化操作。将  $\beta_1, \beta_2, \beta_3, \beta_4$  四个特征向量在通道维度进行堆叠, 得到一个多尺度的特征向量  $\beta$ 。对  $\gamma_1, \gamma_2, \gamma_3, \gamma_4$  四个特征向量做同样的操作得到特征向量  $\gamma$ 。特征向量  $\beta$  和  $\gamma$  的大小都是  $M \times C$ ,  $M$  同样代表特征图中像素点的个数, 但要远远小于  $N$ 。之后对特征向量  $\alpha$  和  $\beta$  使用矩阵乘法, 生成相似度矩阵  $\vec{V}$ :

$$\vec{V} = \alpha^T \times \beta \quad (3-6)$$

最后, 将相似度矩阵  $\vec{V}$  与特征向量  $\gamma$  做矩阵乘法得到一个具有各像素间长距离依赖关系的输出  $O \in R^{N \times C}$ :

$$O = \vec{V} \times \gamma^T \quad (3-7)$$

### 3.2.3 边缘精炼模块 (MRM)

边缘精炼模块是为了从转置 U 形子网增强的特征图中提取更精细的细节特征而设计的。MRM 模块的结构如图 3-3 所示。在 MRM 模块中, 首先, 对输入特征图  $X_{pre}$  使用边缘细化操作 (公式 3-8), 得到具有精确边缘细节的注意力图  $Att$ 。

$$Att_i = 1 - \frac{|p_i - T|}{\max(T, 1 - T)}, i \in [1, 2, \dots, N] \quad (3-8)$$

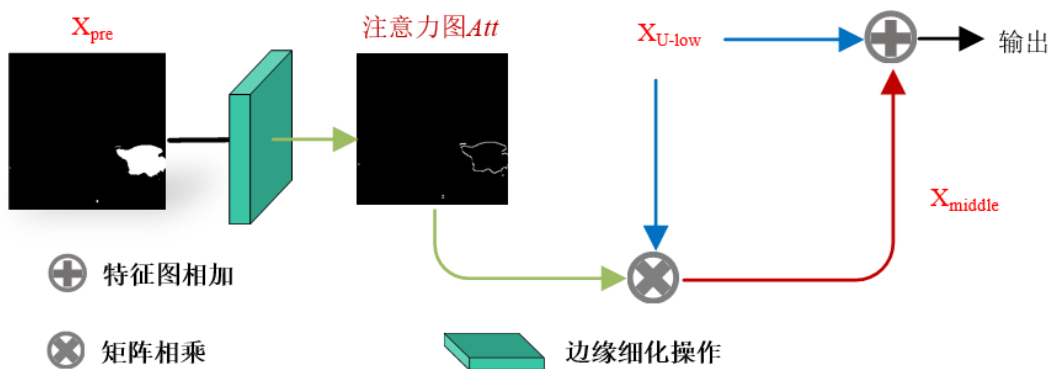


图 3-3 边缘精炼模块 (MRM) 示意图

其中， $i$  表示每个像素在特征图中的位置， $N$  表示特征图中的像素数， $P_i \in (0, 1)$  表示预测图中每个位置的预测值。 $T$  是用于判断当前像素点是否属于病灶边界位置的超参数。在本文中设定  $T=0.5$ 。 $P_i$  与  $T$  的差距越大，预测值属于甲状腺结节边界的概率越小；相反， $P_i$  与  $T$  的差距越不明显，像素越可能属于甲状腺结节边缘。然后，具有精细病变边缘的注意力图  $Att$  与具有更多低维特征的特征图  $X_{U-low}$  进行矩阵乘法，从而生成特征图  $X_{middle}$ 。最终的输出由  $X_{middle}$  和  $X_{U-low}$  进行像素级相加运算产生。

### 3.2.4 融合模块 (AM)

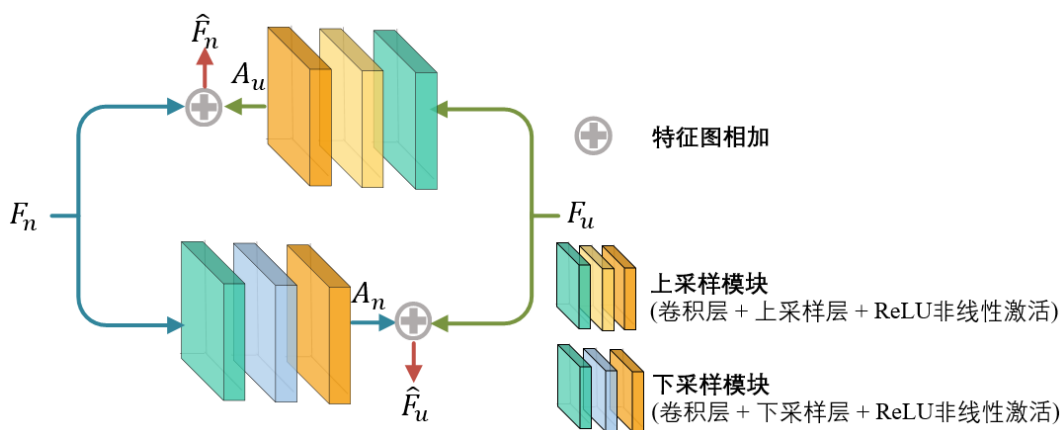


图 3-4 融合模块 (AM) 示意图

转置 U 形子网通过限制感受野增大来获取处理细节信息的能力。随着网络深度的增加，转置 U 型子网生成的特征图包含了更多的细节，这有利于转置 U 形子网提取更多细节信息。相反，U 形子网的感受野随着子网深度的增加而增加，导致 U 形子网的特征图中包含更多的上下文信息，有利于在复杂的超声影像中定位病灶的位置。然而，这也造成了两个子网在每一个阶段的特征图尺寸不匹配，两子网提取到的不同信息不能进行直接融合。

为了融合来自不同子网的特征，本章设计了融合模块（AM）来统一来自两个不同子网特征图的尺寸，并融合来自不同子网提取到的特征。在 AM 模块中，分别为 U 形子网和转置 U 形子网设计了上采样块和下采样块，上采样块由卷积层、双线性插值层和 ReLU 激活函数组成；下采样块由卷积层、最大池化层和 ReLU 激活函数组成。最大池化层和双线性插值层用来改变特征图的大小，使得来自不同子网的特征映射在融合操作之前具有相同的大小。同时，卷积层和 ReLU 激活可以尽可能地防止从图像中捕获到的上下文信息和细节信息丢失。AM 的结构如图 3-4 所示。

设  $F_u$  和  $F_n$  分别表示来自 U 形子网和转置 U 形子网的特征图。首先，将  $F_u$  和  $F_n$  输入到上采样块和下采样块中，得到与  $F_u$  和  $F_n$  大小相同的特征图  $A_u$  和  $A_n$ 。然后，通过像素级相加来聚合来自不同子网的特征。融合后的特征图  $\hat{F}_u$  和  $\hat{F}_n$  可以表示如下：

$$\hat{F}_u = F_u + A_n \quad (3-9)$$

$$\hat{F}_n = F_n + A_u \quad (3-10)$$

最后，将融合了丰富上下文信息和细节信息的特征图  $\hat{F}_u$  和  $\hat{F}_n$  分别输入到 U 形子网和转置 U 形子网的下一层中。

### 3.3 实验与分析

在这一部分，进行了对比实验来评估 DMU-Net 在恶性甲状腺结节数据集上的泛化能力。此外，本节还通过实验分析了 DMU-Net 的边缘细节提取能力。最后，通过消融实验验证了 DMU-Net 各个模块的有效性。

#### 3.3.1 评价指标

为了定量评价所提出方法的性能，使用 Dice 指数（F1-Score）和 Jaccard 相似度得分作为本模型的评价指标。

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (3-12)$$

$$Jaccard = \frac{TP}{TP + FP + FN} \quad (3-13)$$

其中， $TP$  (True Positive) 是指被正确预测的正像素， $FP$  (False Positive) 和  $FN$  (False Negative) 分别表示预测错误的正像素和负像素。需要注意的是，病灶区域的像素为正像素，而背景

区域像素为负像素。DMU-Net 在恶性甲状腺结节分割数据集上用 ADAM 优化器进行了 100 个次的监督训练，学习率设定为 0.0001。所有实验都在两块 NVIDIA RTX 3090 GPU 上验证。

### 3.3.2 模型分割性能

在这一节中，进行了 DMU-Net 与 SegNet<sup>[60]</sup>、U-Net<sup>[59]</sup>和 U-Net++<sup>[61]</sup>的分割性能进行比较。实验结果见表 3-1。具体而言，DMU-Net 在恶性甲状腺结节分割数据集上的 Dice 指数比 Unet 相比提高了 25.86%（56.91% vs 82.77%）。同样，与 Segnet 和 Unet++相比，DMU-Net 也取得了最好的结果。

表 3-1 不同医学影像分割方法在恶性甲状腺结节数据集上的性能

方法	Dice (%)	Jaccard (%)
SegNet	77.40	66.25
Unet	56.91	49.12
Unet++	50.66	38.66
DMU-Net	82.77	72.58

表 3-2 不同医学影像分割方法在恶性甲状腺结节数据集上的 FLOPs 和网络参数

方法	FLOPs (G)	Parameters (M)
SegNet	75.33	16.31
Unet	64.52	34.53
Unet++	138.07	36.63
DMU-Net	283.21	0.30

表 3-2 中是对四种网络的 FLOPs 和参数进行了比较。尽管 DMU-Net 的参数数目最少，但由于其能够提取上下文信息和更多的细节信息，因此可以获得更好的性能。转置 U 型子网随着网络深度的增加，其成倍增加的特征图尺寸导致了 DMU-Net 中 FLOPs 代价最大。

图 3-6 给出了 Unet 和 DMU-Net 在 PGS 数据集上的 Jaccard 指数曲线。从图中可以看出 DMU-Net 的 Jaccard 指数曲线开始时较低，但随着迭代次数的增加，其增长和收敛较快。相比之下，Unet 的曲线缓慢且不稳定。

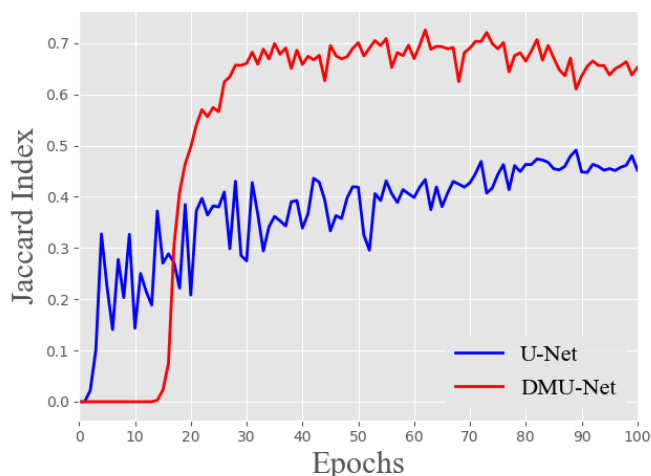


图 3-6 Unet 和 DMU-Net 在前 100 轮分割性能比较

为了定性评价 DMU-Net 对恶性甲状腺结节的分割性能，本节将 DMU-Net 的分割结果与经典的医学影像分割方法进行了比较，如图 3-7 所示。可以发现 DMU-Net 的分割结果要明显优于所对比的模型，尤其是在微小和具有复杂形状的病灶上。

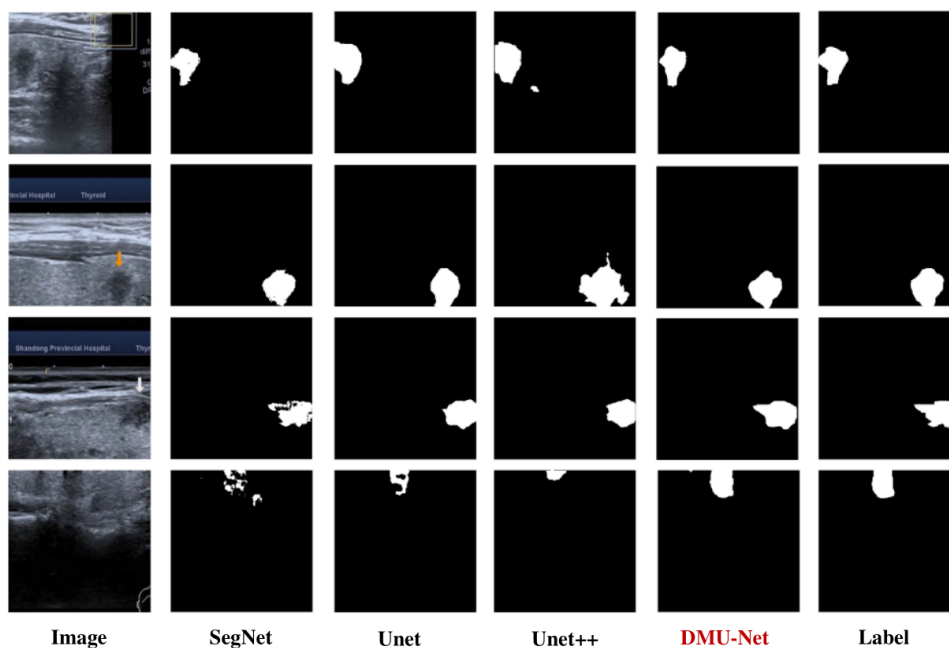


图 3-7 不同模型在恶性甲状腺结节数据集上的分割结果

从图 3-7 中可以看出，SegNet、Unet 和 Unet++在部分图片中取得了不错的分割结果，例如 SegNet 在第一张图片上很好地将甲状旁腺与复杂的背景分离开来，但是这些经典的分割模型在剩余图片上有很多误分割和分割不完全的结果，并不能有效地对超声图片中恶性甲状腺结节进行准确且有效的分割。相比之下，DMU-Net 能够从复杂的超声图片中完整、

准确地分割出恶性甲状腺结节。

### 3.3.3 提取边缘细节能力分析

为了突出 DMU-Net 在提取病灶边缘特征方面的优势，本节从恶性甲状腺结节分割数据集中选取了一些边缘细节复杂的超声影像，并对甲状腺结节边缘的分割结果进行了详细的分析。超声图像中存在许多类似于甲状腺结节区的组织区域，正确地将超声中恶性甲状腺结节和正常的人体组织区别开来对医学影像分割模型来说是一个挑战。图 3-8 定量地列举了部分实验结果来说明 DMU-Net 具有良好的边缘处理能力。

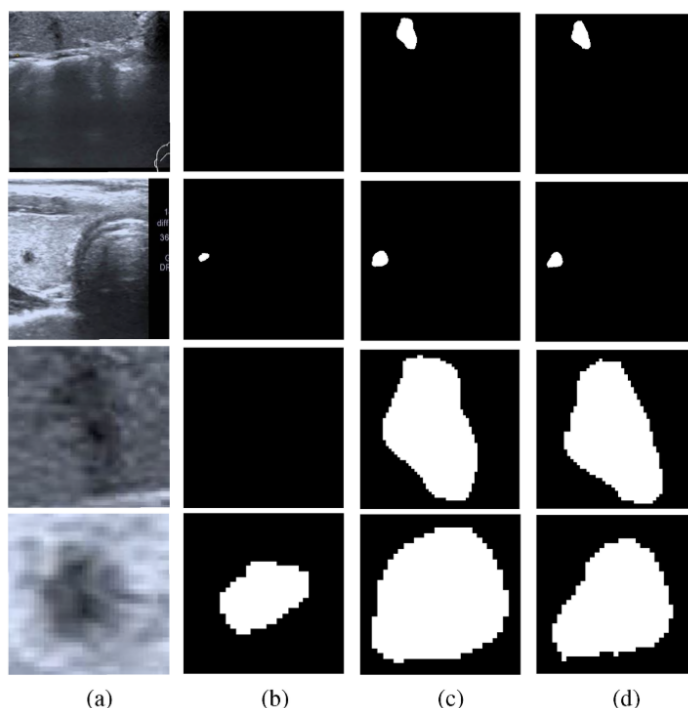


图 3-8 对于边缘不规则的恶性甲状腺结节，DMU-Net 的分割结果更为准确，而 U-Net 的分割结果则较差。底部两行图像是对顶部两行图像进行了局部放大。(a) 超声图像；(b) U-Net 分割结果；(c) DMU-Net 分割结果；(d)为标签

从图 3-8 中可以观察到恶性甲状腺结节的边界不规则，边界上有明显的毛刺。如图 3-8 (d) 所示，虽然人工分割的错误不可避免，但放射科医生已尽了最大努力将毛刺区与恶性甲状腺结节图像分离出来。在这种情况下，DMU-Net 仍能将甲状腺结节的不规则部分尽可能地分割出来。同时，本节还比较了 DMU-Net 和 U-Net 的分割结果。DMU-Net 可以精确地分割出甲状腺结节的毛刺区域。相比之下，U-Net 甚至不能完整地分割甲状腺结节。



在图 3-9 中甲状腺结节（红色虚线框）的颜色和质地与周围正常组织（黄色虚线框）相似。DMU-Net 可以准确分割甲状腺结节区域，而 U-Net 不能正确地识别甲状腺结节区域，造成了误分割。图 3-9 的第二幅图像显示甲状腺结节（红色虚线框）与组织（黄色虚线框）更相似。DMU-Net 不仅能准确定位结节区域，而且能准确分割甲状腺结节。然而，U-Net 的结果不仅对甲状腺结节的分割较差，并且还将正常人体组织错误地归类为甲状腺结节，导致了误分割。

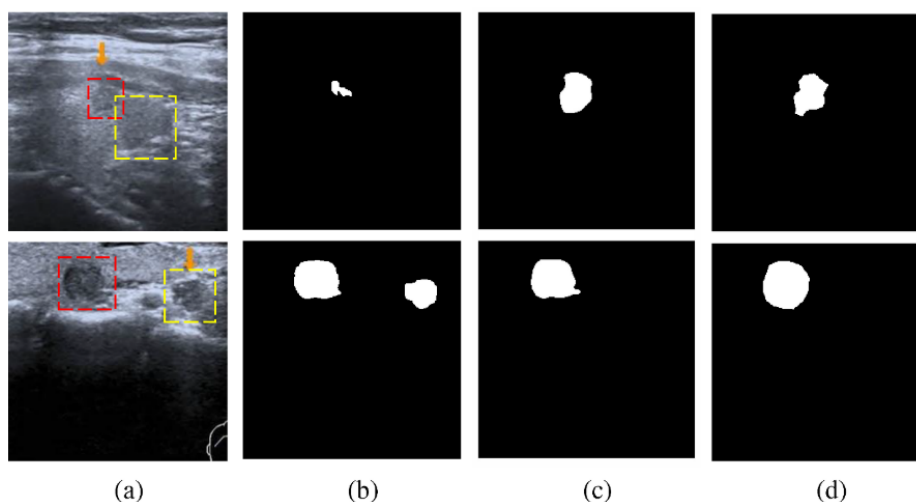


图 3-9 大多数分割模型的挑战是在图像中辨别恶性甲状腺结节和正常人体组织。(a)恶性甲状腺结节超声图像，其中黄色虚线框代表正常组织，红色虚线框是恶性甲状腺结节；(b) 为 U-Net 的分割结果；(c) 为 DMU-Net 分割结果；(d) 标签

### 3.3.4 消融实验

本节评估了 DMU-Net 每个组成部分的作用。所有消融实验均在恶性甲状腺结节分割数据集上进行测试。消融研究从转置 U 形子网（N）和 U 形子网（U）开始，然后测试由转置 U 形子网和 U 子网组成的主体结构（NU）的性能。在测试 NU 的性能时移除金字塔注意力模块（PAM）和边缘细化模块（MRM）两个模块。此外，还研究了边缘细化模块（MRM）在捕获病变区域边缘特征方面的优势。同样，通过在基于 NU 的 U 形子网的跳连接中添加 PAM 模块来研究 PAM 模块对整个网络的影响。最后，作为消融研究的一部分，本节比较了 DMU-Net 与 N，U，NU，NU + PAM，NU + MRM 几个不同组合的分割性能。所有消融研究结果如表 3-3 所示。

表 3-3 消融实验结果

方法	N	U	NU	NU + MRM	NU + PAM	DMU-Net
<b>Dice (%)</b>	0.3	68.82	77.18	79.57	81.07	<b>82.77</b>
<b>Jaccard (%)</b>	0.16	54.95	65.43	68.45	70.62	<b>72.58</b>

### 3.4 本章小结

本章提出了一种基于双通路镜像 U-Net 深度神经网络 (DMU-Net) 的恶性甲状腺结节影像分割方法, 该方法使用 U 形子网从超声图像中提取上下文信息来准确定位恶性结节区域, 并使用转置 U 形子网来捕获恶性结节的边缘细节信息。DMU-Net 优越的分割性能可以辅助临床医生对患者的病情进行诊断。

(1) 优势。DMU-Net 可以在恶性甲状腺分割数据集上取得优异的分割效果。得益于 U 形子网与转置 U 形子网结合的双通路设计和三个辅助模块对模型整体性能的提升, DMU-Net 可以在恶性甲状腺结节分割任务中提取更有价值的病变区域特征, 从而对超声中恶性甲状腺结节进行准确分割。

(2) 临床价值。首先, 对于甲状腺结节, 尤其是在超声图像中尺寸小、边界不规则, 容易与周围组织混淆的恶性结节, DMU-Net 可以帮助医生准确定位恶性甲状腺结节, 避免因其与周围组织混淆而引起的误诊和漏诊。其次, 本章所使用的恶性甲状腺结节分割数据集由山东省立医院乳腺甲状腺外科相关医生标注, 包含了临床医生丰富的诊断经验, DMU-Net 可以从临床医生那里学习到丰富的经验和知识, 并对超声中的恶性甲状腺进行分割, 以帮助缺乏经验的年轻医生在超声图像中识别特征不明显的恶性甲状腺结节。

综上所述, 本章提出了一种基于深度学习的恶性甲状腺结节分割方法, 并在医院的帮助下建立了恶性甲状腺结节分割数据集。实验结果表明, DMU-Net 的分割性能优于经典的医学影像图像分割方法。

## 第四章 基于“教师伙伴”策略的术前甲状旁腺分割的早期-晚期多阶段网络

### 4.1 引言

甲状腺癌是目前世界上增长速度最快的恶性肿瘤之一。从 1990 年到 2019 年，甲状腺癌的全球甲状腺癌发病率（ASIR）迅速增长了 40.65%<sup>[24]</sup>。手术是甲状腺癌的首选治疗方法。然而，手术诱发的甲状旁腺功能减退症是甲状腺手术最常见、最严重的并发症之一。甲状旁腺功能减退症可引起暂时性或永久性的低钙血症，严重影响患者的生活质量，甚至威胁到患者的生命。这也是甲状腺手术后造成医患冲突的一个主要原因<sup>[3]</sup>。因此，甲状旁腺的识别和保护对甲状腺和甲状旁腺外科医生来说是一个严峻的挑战。

一般来说，甲状旁腺可分为两种类型，分别是上甲状旁腺和下甲状旁腺。其中，上甲状旁腺位于甲状腺两侧的较高位置，下甲状旁腺通常位于甲状旁腺两侧较低位置。值得注意的是，由于胚胎发育起源的不同，上甲状旁腺的位置相对固定，而下甲状旁腺的位置是可变的，通常难以预测。此外，甲状旁腺的颜色、形状、大小和质地与周围淋巴结、脂肪颗粒、异位甲状腺结节等组织相似，导致甲状旁腺在甲状腺切除术中难以识别<sup>[4]</sup>。并且，在甲状腺切除术中，无论是单侧甲状腺叶切除术还是甲状腺全切除术，无一例外地都要切除淋巴结及其周围的脂肪和结缔组织。这进一步增加了手术中甲状旁腺损伤和意外切除的发生率。

超声检查是甲状腺相关疾病的首选筛查方法<sup>[25]</sup>。然而，关于正常甲状旁腺的超声特征的研究很少。超声检查可以帮助甲状腺外科医生预先确定甲状旁腺的关键区域，大大减少外科医生的工作量和手术时间，降低手术失误的风险，从而有效避免甲状旁腺功能减退相关的术后并发症，提高患者的生活质量，减少医患冲突。虽然超声检查在识别甲状旁腺方面具有显著优势，但仍有一些挑战需要克服，如甲状旁腺难以确定（尤其是下甲状旁腺），外部特征不明显<sup>[7]</sup>等。

近年来，随着人工智能技术的快速发展，基于机器学习的计算机辅助诊断（CAD）方法已成为诊断甲状腺相关疾病的新手段。Ahmad 等人<sup>[26]</sup>专注于甲状腺疾病的诊断，并提出

了一种混合决策支持系统,准确率可达到 98.5%。此外,Banu 等<sup>[27]</sup>采用线性判别分析(LNA)对甲减和甲亢分类任务进行分类,准确率为 99.62%。同样,也有许多先进的方法<sup>[28][40]</sup>用于甲状腺疾病的诊断,并取得了令人满意的结果。此外,也出现了许多针对甲状腺结节的医学图像分析方法。Chang 等<sup>[41]</sup>提出了一种决策树模型来解决超声图像中甲状腺结节的分割任务,该模型包括图像预处理阶段和图像分割阶段。Keramidas 等<sup>[42]</sup>设计了一种甲状腺结节检测系统,可以用来处理甲状腺结节图像和视频。Ma 等人<sup>[11]</sup>利用基于深度学习的方法对超声图像中的甲状腺结节进行分割,取得了较好的分割效果。与 Ma 等人<sup>[11]</sup>提出的端到端模型不同,Ying 等人<sup>[20]</sup>提出了一种用于甲状腺结节分割的级联卷积神经网络(CCNN),并取得了显著的性能。Kang 等<sup>[43]</sup>引入了多任务学习策略来探索甲状腺结节分割与分类任务之间的内在联系。以上研究均以甲状腺相关疾病为研究对象,针对甲状旁腺术前定位和分割以提高甲状旁腺在超声影像中可读性的方法较少。由于甲状旁腺在超声图像中几乎不可见,所以分割超声影像中的甲状旁腺是一个具有挑战性的任务。

此外,由于甲状旁腺在超声图像中所占的比例较小,大多数分割方法容易被超声图像中混乱复杂的背景所误导,导致模型性能下降。幸运的是,已经有一些相关的方法专注于解决此类问题。Zhou 等<sup>[44]</sup>提出了一种由粗到细的多阶段方法,该方法利用粗分割阶段检测下一个细分割阶段的目标区域。然而,这两个阶段各自工作,信息交互受到限制,影响了模型分割性能的提高。Yu 等<sup>[45]</sup>在 Zhou 等人<sup>[44]</sup>的基础上提出了显著性转换模块,该模块将前一次迭代结果转换为空间权重分享到当前迭代。该方法将粗分割阶段提取到的信息传播到细分割阶段。理论上,粗分割阶段提取到的信息可以帮助细分割阶段识别目标区域,但从非目标区域提取的无用特征会细分割阶段造成一定程度上的误导,进而导致性能下降。

本章中提出了一个基于“教师伙伴”策略的早-晚多阶段网络(Early-late multi-stage network, EL-Net)。通常,多阶段网络意味着模型在不同的阶段执行不同的特征提取任务,通过各个模型阶段之间的协作,达到传统单阶段网络所不能达到的性能上限。EL-Net 由两个阶段构成,分别是早期阶段和晚期阶段。早期阶段负责检测目标区域,并将早期阶段骨干分割网络的预测结果和相关特征图裁剪成较小的图像。然后,晚期阶段根据裁剪后的小图像进行更精确的局部分割,并将裁剪后的分割结果恢复到正常大小。本方法将 Vision Transformer(ViT)<sup>[46]</sup>作为模型早期阶段的骨干网络。ViT 可以从输入图像中提取丰富的上

下文信息用于检测甲状旁腺。此外，本章还设计了基于 CNN 的特征增强子网作为模型晚期阶段的骨干网络。其负责从裁剪后的图像和特征图中捕获更详细的细节特征。与传统的单阶段网络相比，EL-Net 网络的优点是能够定位有价值的区域，消除混乱的背景对分割甲状旁腺时的影响。此外，EL-Net 可以在不耗费更多计算资源的情况下，从较小但更有价值的目标区域中挖掘更详细的目标特征。

遗憾的是，模型晚期阶段的特殊设计决定了它无法提取丰富的上下文信息，因此仅仅将裁剪后的图像输入到模型晚期阶段会导致性能严重下降。其中，解决晚期阶段上下文信息缺乏最直接途径是将模型早期阶段提取到的丰富的上下文信息直接转移到模型晚期阶段，但这种方式忽略了像素间的关系。为此，本章提出了“教师-伙伴”策略，该策略在训练的早期阶段将模型早期阶段和晚期阶段的关系定义为“教师-伙伴”，它可以将模型早期阶段提取到的像素间的结构化知识转移到模型晚期阶段。随后，在训练后期将模型早期阶段和晚期阶段的关系转化为“伙伴-伙伴”关系，早期阶段负责从复杂的超声图像中定位并裁切有价值的区域，晚期阶段用于进一步提取甲状旁腺的边缘细节特征以得到更精细的分割结果。

本章的主要创新点可以概括为以下四个方面：

（1）提出了由早期和晚期两个阶段构成的 EL-Net。它能从复杂的超声图像中检测并裁切有价值区域，减轻了复杂的超声背景对后续晚期阶段进一步分割甲状旁腺时的影响。

（2）设计了“教师-伙伴”策略从早期阶段向晚期阶段补充像素间的结构化信息。

（3）设计了特征增强子网（Feature Reinforcement Subnet, FR-Subnet）和边缘感知模块（Edge Aware Module, EAM）作为 EL-Net 晚期阶段的骨干网络，用于捕捉超声中甲状旁腺的细节信息。

（4）本章所提出的计算机辅助诊断方法为甲状腺切除术中保护甲状旁腺提供了一种新的术前辅助诊断思路。

## 4.2 早晚多阶段网络（EL-Net）

整个网络结构可以分为早期阶段和晚期阶段两个部分，如图 4-1 所示。其中早期阶段目标是从输入的超声图像中检测出有价值区域。早期阶段又可分为两部分，分别是用于准

确地分割和定位有价值区域的主干特征提取网络和对有价值区域进行裁剪，避免过多复杂的背景噪声影响模型的分割效果的裁剪操作。

模型晚期阶段以早期阶段裁剪后的图像和特征图作为输入。这一阶段的目标是对前一阶段的预测结果进行细化，从而在提出的基于 CNN 的晚期骨干网（FR-Subnet）中提取更精细的有关目标区域细节特征，并通过还原操作将分割图片的大小恢复到正常尺寸。

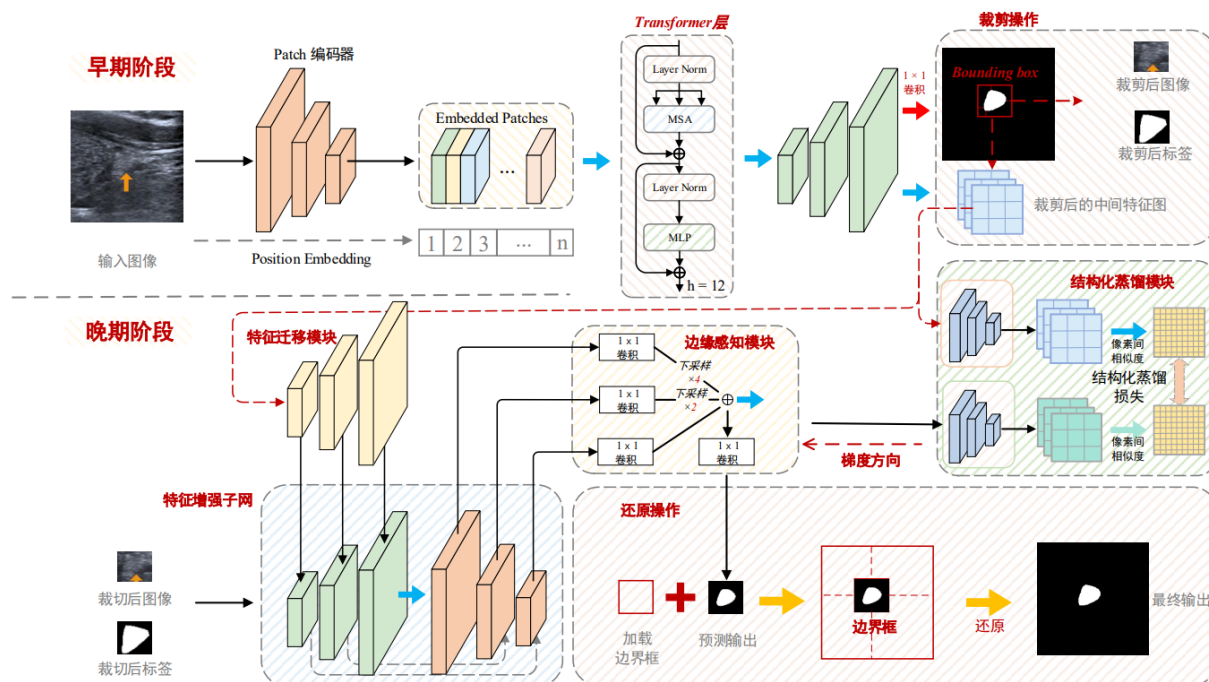


图 4-1 早晚多阶段网络（EL-Net）模型示意图

### 4.2.1 早期阶段骨干网络

考虑到 Transformer 在医学图像分割任务中提取上下文信息的优异性能，本方法使用 Vision Transformer<sup>[46]</sup>作为 EL-Net 的早期阶段的骨干网络来从输入的超声图像中提取全局上下文信息。与通过堆叠更多卷积层来获取全局上下文信息的传统 CNN 网络不同，Transformer 模型通过堆叠 Transformer 层来直接建立特征图中任意像素之间的依赖关系。

在将图片输入 Transformer 层之前，需要对输入图像需要进行若干预处理操作。首先，给定通道数为  $C$ ，空间分辨率为  $H \times W$  的输入图像  $X \in \mathbb{R}^{C \times H \times W}$ ，使用切片编码器将  $X$  分割成  $N = H/P \times W/P$  个图像块，其中  $P$  是每个图像块的大小。之后，二维图片  $X \in \mathbb{R}^{C \times H \times W}$  通过 Linear projection 操作转换为一维向量  $X_p \in \mathbb{R}^{N \times C_e}$ 。此外，可学习的位置向量被加入到

$X_p$  中来建立任意两个图像块之间的位置关系。最后，这些图像块被输入到之后的 Transformer 层中进一步处理。

Transformer 层一般由多头自注意力模块 (MSA) 和多层感知器模块 (MLP) 组成。在 MSA 中，首先使用  $X_p$  或者上一层 Transformer 的预测结果  $X_l$  计算  $query = X_l W_Q$ ,  $key = X_l W_K$ ,  $value = X_l W_V$  三个向量，其中  $W_Q \in \mathbb{R}^{C_e \times d_Q}$ ,  $W_K \in \mathbb{R}^{C_e \times d_K}$ ,  $W_V \in \mathbb{R}^{C_e \times d_V}$  代表三个可学习矩阵。之后，对  $query$  和  $key$  进行矩阵乘法运算，得到注意力向量。随后，生成的注意力向量被输入到 *Softmax* 激活函数中进行归一化处理。最后，将每个  $value$  向量与相应的注意力向量相乘，并将加权值向量叠加，得到输出向量。

通常，每个 MSA 包含  $h$  个自注意模块 (默认  $h = 12$ )。因此每个自注意力模块输入  $query$ ,  $key$ ,  $value$  三个向量的通道数为  $d_Q = d_K = d_V = C_e/h$ 。自注意力操作如下：

$$SA(X_l) = \text{Softmax}\left(\frac{query(key)^T}{\sqrt{d_k}}\right)value \quad (4-1)$$

之后，将来自不同自注意力模块的结果在通道维度进行堆叠，最终结果用公式表示如下：

$$MSA(X_l) = X_l + \text{Concat}[SA(X_l)_1, \dots, SA(X_l)_h]W_0 \quad (4-2)$$

其中， $W_0$  是一个可学习的矩阵。与直接在原始图像上计算自注意相比，加入  $W_0$  可以有效地减少计算量。此外，不同的向量具有不同的分布，MSA 可以从自注意力模块中捕捉不同结果之间的各种相关性。

Transformer 层的最终结果可以表示为：

$$X_{l+1} = \text{MLP}(MSA(X_l)) + MSA(X_l) \quad (4-3)$$

其中，MLP 包含两个线性层和一个 *Gelu* 非线性激活函数构成。MLP 通过对 MSA 的特征空间进行变换，有效地增强了 Transformer 层的建模能力。

为了便于后续的裁剪和融合操作，本节对输出向量进行了一定变换。首先，将  $X_l \in \mathbb{R}^{N \times C_e}$  变换为  $X_r \in \mathbb{R}^{C_e \times \frac{H}{h} \times \frac{W}{h}}$ ，之后使用  $1 \times 1$  卷积将其通道数变为 512，此时  $X_r \in \mathbb{R}^{512 \times \frac{H}{h} \times \frac{W}{h}}$ 。一般地，当输入图片的空间分辨率为  $224 \times 224$  时， $C_e = 768$ ,  $h = 12$ 。为了方便后续网络的晚期阶段进一步从图像中提取细节特征，EL-Net 使用连续的卷积操作将特征图  $X_r \in \mathbb{R}^{512 \times \frac{H}{h} \times \frac{W}{h}}$  上采样为  $X_f \in \mathbb{R}^{16 \times H \times W}$ ，最后使用  $1 \times 1$  卷积得到最终预测输出图  $X_{crop} \in \mathbb{R}^{2 \times H \times W}$ 。

### 4.2.2 裁剪阶段

超声图像中的背景区域存在与甲状旁腺特征相似，具有混淆性的组织。由于晚期阶段骨干网络不具备从图像中提取上下文信息的能力，因此，无法对准确地对超声中的甲状旁腺进行定位，如果将大量具有混淆性的背景区域直接输入到模型的晚期阶段，容易误导后续晚期骨干网络的判断，导致分割效果较差。早期阶段 Transformer 可以从输入图像中提取丰富的上下文信息用于确定超声中甲状旁腺的位置，但将 Transformer 提取到的全部信息传递给晚期阶段是不必要的。因此，有必要为模型的晚期阶段剔除超声图像和相关的特征图中无用的内容。在这一阶段，将 Transformer 主干网络的预测输出  $X_f \in \mathbb{R}^{16 \times H \times W}$  和  $X_{crop} \in \mathbb{R}^{2 \times H \times W}$  作为模型晚期阶段的输入。首先，需要生成目标区域位置的矩形边界框（Bounding box, bbox）。边界框的生成过程如下：

$$bbox = Location(X_{crop}, m) \quad (4-4)$$

其中 *Location* 函数表示有价值区域的检测过程，*m* 表示目标区域周围的填充距离，默认设置为 20 像素。值得注意的是，在目标区域周围保留部分区域更有利于模型晚期阶段提取特征。之后，将输入的超声影像  $X$ ，对应的二值标签图  $X_{label}$  和用于晚期阶段的中间特征图  $X_f$  进行裁剪。具体描述如下：

$$X^c, X_{label}^c, X_f^c = Crop(X, X_{label}, X_f; bbox) \quad (4-5)$$

*Crop* 函数用于裁剪输入的超声图像，对应的二值标记和晚期阶段的中间特征图。

### 4.2.3 晚期阶段骨干网络

一般情况下，CNN 网络的感受野随着特征图的减小而逐渐增大。较大的感受野可以使 CNN 网络更多地关注上下文信息，但这也使一定程度上其忽略了目标区域的有价值的细节信息。受 Valana-rasu 等人<sup>[47]</sup>的启发。本章为网络的晚期阶段设计了基于 CNN 的主干网—FR-Subnet，它可以通过逐步限制感受野的增加，从有价值的区域捕捉更多的细节信息。

FR-Subnet 由编码器和解码器组成。与 U-Net 的编码器不同，FR-Subnet 的编码器更类似于 U-Net 的解码器，FR-Subnet 的编码器包含三个 *SampConv* 块。每个 *SampConv* 块由上采样/下采样层和卷积层及激活函数组合而成，用于增大/缩小特征图的大小以缩小/增大感受野。首先，将裁切后的特征图  $X_f^c$  输入到特征迁移模块（Feature Transfer Module, FTM）。FTM 由三个堆叠的卷积层组成，这些卷积层可以将输入的特征映射上采样到三个不同大小



的特征映射，从而为 FR-Subnet 补充来自模型早期阶段的丰富上下文信息。之后，将裁剪后的图像  $X^c$  输入到 FR-Subnet 编码器中，提取更多甲状旁腺的细节特征。然后，将编码器每一层的输出特征图与  $T_i$  进行求和，以为 FR-Subnet 补充丰富的上下文信息。其中  $T_i$  表示来自 FTM 第  $i$  层的输出特征图。公式表示如下：

$$X_{i+1}^c = \text{Mean}(\text{SampConv}_i(X_i^c) + T_i), i \in [1, 2, 3] \quad (4-6)$$

其中， $i$  表示第  $i$  个 *SampConv* 块，*Mean* 代表取平均值。相似地，将增强后的特征映射  $X_i$  输入 FR-Subnet 的解码器中，以将放大后的特征图恢复到原有尺寸。其数学表达如下：

$$X_{i+1}^c = \text{SampConv}_i(X_i), i \in [4, 5, 6] \quad (4-7)$$

具体来说，对于 FR-Subnet 的编码器（前三个 *SampConv* 层），设  $k \times k$  表示 FR-Subnet 第一层的感受野，通过第一个 *SampConv* 层的处理，感受野变为  $k/2 \times k/2$ 。相似地，经过第三个 *SampConv* 层的处理后的感受野变为  $k/4 \times k/4$ 。小的感受野意味着更大的特征图。经过每一 *SampConv* 层，感受野都会变为之前的一半，但是特征图的尺寸增大了一倍，这代表着随着网络编码器深度的增加，FR-Subnet 可以捕捉图像中更多的细节信息，然而，较小的感受野意味着 FR-Subnet 从图像中提取语义信息的能力不足。因此，EL-Net 将早期阶段从 Transformer 主干网络中提取到的丰富语义信息通过 FTM 模块分层地引入到 FR-Subnet 中。FR-Subnet 的解码器（后三个 *SampConv* 层）主要用于恢复放大后的特征图的大小。与 FR-Subnet 编码器的 *SampConv* 层不同，解码器的 *SampConv* 层利用卷积运算后的池化操作来下采样被放大的特征图。然后，将特征图输入后续的边缘感知模块（EAM）中，对提取到的不同尺度的细节特征进行融合。此外，为了向 FR-Subnet 解码器补充编码器阶段提取到的细节信息，FR-Subnet 的编码器和解码器之间设置了跳跃连接。

#### 4.2.4 边缘感知模块（EAM）

尽管 FR-Subnet 在提取细节特征方面有较好的性能，但其最终预测输出忽略了来自不同尺度的信息。多尺度特征是解决医学图像分割中复杂尺度变化问题的有效方法。因此，本节引入边缘感知模块（EAM）来融合从 FR-Subnet 捕获的多尺度细节特征。

如图 4-1 所示，FR-Subnet 解码器层的输出表示为 EAM 模块的输入，即  $X_i^c, i = 5, 6, 7$ 。具体来说，首先，用  $1 \times 1$  卷积层对三个不同尺度的特征图进行通道维度匹配。然后，通过在空间上使用不同池化系数的池化操作来对齐这些特征图。具体来说，分别对  $X_5^c$  和  $X_6^c$  进行

下样，其中池化系数设置为 4 和 2。 $X_7^c$ 的大小保持不变。然后将这些特征图进行像素级的相加操作得到多尺度特征图 $X_{multi}^c$ 。

$$X_{multi}^c = Sum(X_5, X_6, X_7) \quad (4-8)$$

其中 $Sum$ 表示求和操作。最后，将处理后的特征图 $X_{multi}^c$ 作为后续结构化蒸馏模块（SDM）的输入，然后进行 $1 \times 1$ 卷积运算，生成二值预测结果。

#### 4.2.5 还原操作

模型的晚期阶段可以捕获更多的细节特征从而精确地分割超声影像中的甲状旁腺，但它严重依赖于早期裁剪后的超声影像和特征图作为晚期阶段的输入。为了对分割结果进行整体预测，在这一阶段，将裁剪后的分割结果恢复到正常尺寸。

具体来说，首先，加载早期阶段用于裁剪有价值区域的边界框（bbox）。之后，将 EAM 的输出作为模型晚期阶段的分割结果嵌入到边界框中。然后，使用 bbox 提供的坐标信息计算出裁剪后的分割结果在最终预测图中的具体位置，通过水平和垂直方向的填充操作将分割结果恢复到正常尺寸。

### 4.3 “教师伙伴”策略

知识蒸馏策略，即“师生”策略。其目的是尽可能地将复杂模型提取到的更多有用信息转移到轻量化模型中，帮助轻量化模型实现更好的性能。本文所提出的模型由早期阶段和晚期阶段两部分构成。早期阶段用于从图像中提取上下文信息以准确地找到目标区域。晚期阶段则用于捕捉从裁切后的超声影像中捕获关于甲状旁腺的更多细节特征。模型晚期阶段主干网络 FR-Subnet 的特殊设计决定了其没有从输入的超声影像中捕获全局语义信息的能力。因此，EL-Net 使用特征迁移模块将模型早期阶段从超声影像中提取到丰富语义信息传递到晚期阶段用于为其补充上下文信息。为了进一步将模型早期阶段 Transformer 骨干网络提取到的像素间结构化知识也传授给 FR-Subnet, 受 Hinton 和 Liu 等人<sup>[48][49]</sup>的启发，本章引入了“教师-伙伴”策略。在模型训练的前期，早期阶段充当“教师”，晚期阶段充当“学生”，“学生”从“教师”那里接受像素级结构化信息，此时模型早期阶段和晚期阶段两部分的关系为“教师-学生”。在模型训练阶段的后期，即模型晚期阶段在“教师”的帮助下具备了提取细节信息的能力后，模型早期和晚期的关系转化为“伙伴-伙伴”，模型

早期阶段负责定位有价值区域，晚期阶段对有价值区域进行进一步细化分割。

图像分割问题本质上是基于像素的分类任务。一般情况下，相邻像素的预测结果是连续且相互关联的，为了将这种像素间的关联信息从模型的早期阶段传递给晚期阶段，本章提出了结构化蒸馏模块（SDM）以实现“教师-伙伴”策略。

具体来说，首先使用两个由三个堆叠的卷积层组成的特征提取器对模型早期阶段和晚期阶段的特征图进行编码，使得两阶段的特征图尺寸统一。然后，计算早期阶段特征图中任意两个像素之间的相似度 $a_{ij}$ 。可以表示为：

$$a_{ij} = p_i^T p_j / (\|p_i\|_2 \|p_j\|_2) \quad (4-9)$$

其中， $p$ 表示像素点。为了将结构化知识从网络早期阶段转移到网络晚期阶段，使用平方差损失来实现结构化蒸馏损失。

$$L_{distill} = \frac{1}{(W' \times H')^2} \sum_{i \in R} \sum_{j \in R} (a_{ij}^S - a_{ij}^T)^2 \quad (4-10)$$

$a_{ij}$ 的上标 $T$ （教师）和 $S$ （学生）表示网络早期阶段和网络晚期阶段特征图中任意两像素之间的相似度。 $W'$ 和 $H'$ 表示特征图的宽和高。

## 4.4 损失函数

为了加快收敛速度并避免分割任务的过拟合，本节组合了 Dice loss 和 Binary Cross-Entropy loss 作为训练模型的损失函数。Dice loss 损失可以用如下的数学形式表示：

$$L_{Dice} = 1 - 2 \times \frac{\sum_i P_i G_i}{\sum_i P_i + \sum_i G_i} \quad (4-11)$$

此外，Binary Cross-Entropy loss 可以用数学表示为：

$$L_{BCE} = -\sum_i [(1 - G_i) \ln(1 - P_i) + G_i \ln(P_i)] \quad (4-12)$$

其中 $i$ 表示输出特征图中的第 $i$ 个像素。 $P_i$ 表示第 $i$ 个像素的预测值， $G_i$ 表示第 $i$ 个像素的标签值。

网络早期阶段和晚期阶段的损失函数可以描述为：

$$L^T = L_{Dice}^T + L_{BCE}^T \quad (4-13)$$

$$L^S = L_{Dice}^S + L_{BCE}^S \quad (4-14)$$

上标 $T$ （教师）表示网络早期阶段，上标 $S$ （学生）表示网络晚期阶段。值得注意的是，预测结果和对应标签在晚期阶段都进行了裁切。

最终模型的损失函数可以描述为:

$$L = L^T + \lambda_1 L^S + \lambda_2 L_{distill} \quad (4-15)$$

其中,  $\lambda_1$  和  $\lambda_2$  的初始值默认设置为 0, 当早期预测结果的 Dice 系数达到阈值  $\psi$  时,  $\lambda_1$  设置为 1。根据经验,  $\psi$  被设置为 0.7。此外, 当晚期阶段预测结果的 Dice 系数落后于早期阶段的 Dice 吸收时, 表示模型晚期阶段需要早期阶段的帮助, 此时  $\lambda_2$  也设置为 1。

## 4.5 实验与分析

本节进行了充分的对比实验和消融实验, 以验证 EL-Net 的整体性能和各个模块功能的有效性。

### 4.5.1 模型分割性能

为了验证模型的分割性能, 本章将提出的方法与六种用于医学图像分割任务的优秀方法进行了比较, 包括 U-Net<sup>[59]</sup>、U-Net++<sup>[61]</sup>、Attention U-Net (AttU-Net)<sup>[62]</sup>、TransUnet<sup>[63]</sup>、FAT-Net<sup>[65]</sup>和 Medical Transformer (MedT)<sup>[64]</sup>。其中, U-Net 是医学图像分割任务中公认最经典的应用模型。同时, U-Net++和 AttU-Net 是基于 U-Net 的重要改进版本。此外, TransUnet 和 FAT-Net 是结合 Transformer 与卷积神经网络两者优势的两个方法。MedT 是一种基于 Transformer 的双通路分割模型。本节接下来将从定量和定性两个方面详细验证并对比所有方法的性能。值得注意的是, 为了确保与上述方法的公平比较, 所有的方法都是在相同的硬件和软件环境下进行训练并评估的。

表 4-1 给出了六种分割方法在 Dice 系数、Jaccard 指数和 Sensitivity 上的定量比较结果。从表中可以看出, 本章的方法在甲状旁腺分割数据集上取得了最好的分割性能, Dice 系数和 Jaccard 指数分别为 78.87%和 66.42%。此外, 该方法的 Dice 系数、Jaccard 指数和 Sensitivity 分别比次优方法 (U-Net) 高出 2.26%、3.42%和 2.25%。虽然 MedT 的整体性能落后于本文提出的方法, 但考虑到其低成本的 FLOPs 和参数量, 它仍然是一种非常有竞争力的医学图像分割方法。FAT-Net 在甲状旁腺数据集上的分割结果与其他方法相比最差, Dice 和 Jaccard 的分割结果分别为 72.53%和 59.21%。其余三种方法 (U-Net++、AttU-Net 和 TransUnet) 在甲状旁腺数据集上取得了相近的分割性能。

表 4-1 不同医学影像分割方法在 PGS 数据集上的性能

方法	Parameters	FLOPs	Dice	Jaccard	Sensitivity
	(M)	(G)	(%)	(%)	(%)
U-Net	7.85	10.72	76.2	63.00	79.10
U-Net++	9.16	26.51	75.82	62.41	75.67
AttU-Net	7.98	11.03	75.07	61.59	75.01
TransUnet	105.28	24.64	74.01	60.94	74.12
FAT-Net	28.76	30.53	72.53	59.21	71.79
MedT	1.56	1.64	72.94	59.52	71.85
Ours	98.75	112.27	<b>78.87</b>	<b>66.42</b>	<b>81.62</b>

图 4-2 定性地比较了本方法与其余六种对比方法在甲状旁腺数据集上的分割结果。具体来说，从图 4-2 (b) 到图 4-2 (e)，U-Net、U-Net++、AttUnet 和 TransUnet 容易将一些非目标区域错误地归类为目标区域，而 EL-Net 可以准确地预测目标区域。此外，与 FAT-Net 和 MedT 相比，EL-Net 的分割结果包含了目标区域更多细节特征。这也证明所提出的 EL-Net 在超声图像中分割甲状旁腺的优越性。

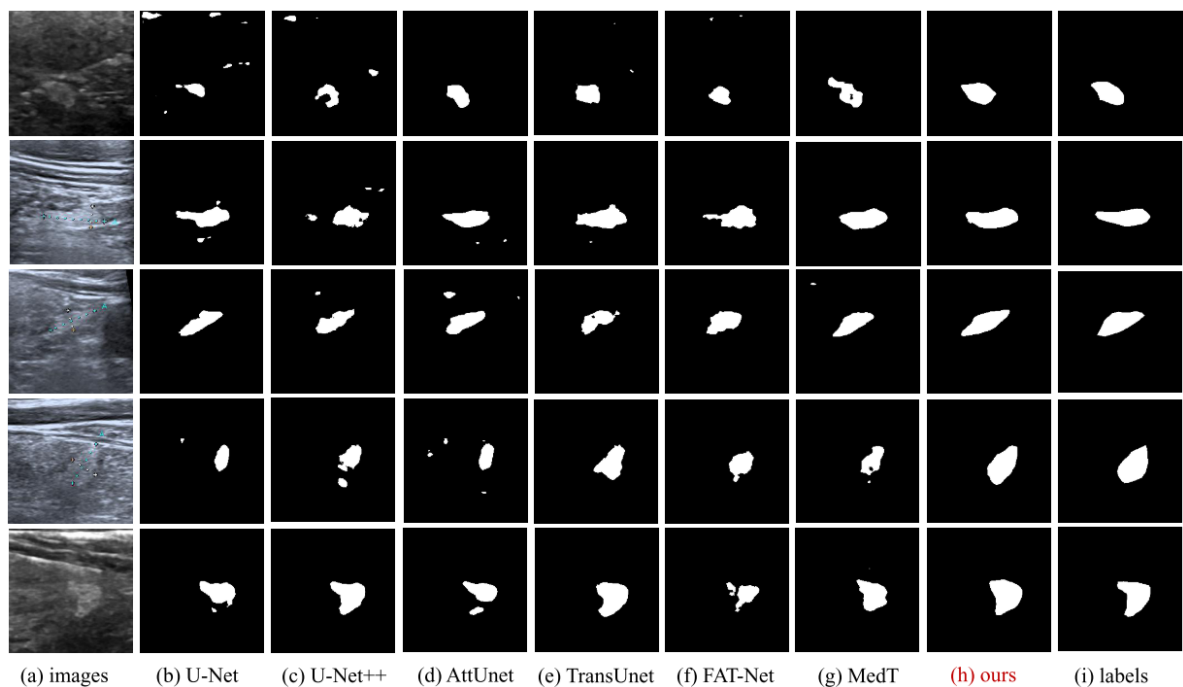


图 4-2 不同方法的甲状旁腺分割结果的定性比较。(a)为输入的甲状旁腺超声图像；(b) - (h)分别是 U-Net、U-Net++、AttU-Net、TransUnet、FAT-Net、MedT 和 EL-Net 的分割结果；(i)标签

### 4.5.2 消融实验

为了证明 EL-Net 每个组件的有效性,本节进行了大量的消融研究。首先,本节验证了模型早期阶段和晚期阶段对模型整体的作用。然后,对模型晚期阶段的各组成部分的有效性进行了评估。每种组合都在甲状腺数据集上进行了五折交叉验证,并且都在相同的软硬件环境下测试。

为了评估 EL-Net 早期和晚期的性能,首先对早期阶段和晚期阶段进行了消融研究。通过移除“教师-伙伴”策略的主体结构化蒸馏模块(SDM),测试各个阶段的性能,具体实验结果如表 4-2 所示。一方面,在甲状腺数据集上,模型使用早期-晚期多阶段与仅使用晚期阶段相比,在 Dice 系数、Jaccard 指数和 Sensitivity 方面分别提高了 28.17%、29.99% 和 33.03%;另一方面,早期-晚期多阶段在 Dice 系数、Jaccard 指数和 Sensitivity 方面比仅使用早期阶段提高了 1.28%、1.67%和 1.52%。

表 4-2 早期阶段和晚期阶段消融实验的定量分析结果

方法	Parameters (M)	FLOPs (G)	Dice (%)	Jaccard (%)	Sensitivity (%)
早期阶段	94.82	22.22	77.59	64.75	80.10
晚期阶段	3.93	90.05	50.11	35.59	47.07
早-晚多阶段	98.75	112.27	<b>78.28</b>	<b>65.58</b>	<b>80.10</b>

从表 4-2 中可以看出,模型仅使用晚期阶段与使用早期-晚期多阶段性能差距较大,其原因是模型晚期阶段善于从输入的超声图像中提取更多的细节特征,但缺乏对有价值区域的识别能力,其中超声图像中复杂混乱的背景容易误导 EL-Net 的晚期阶段。同时,模型早期阶段的移除,导致其不能使用模型早期阶段裁切后的图像作为输入,这进一步增加了模型在复杂的超声图像中分割甲状腺的难度。幸运的是,模型的早期阶段可以提取丰富的上下文信息来精确定位输入超声图像中的目标区域。虽然失去了晚期阶段对细节特征的提取,会一定程度上造成模型对目标区域分割精度的下降,但不会有分割精度的大规模下降的情况出现。因此,模型在早期阶段的支持下,配合晚期阶段对目标区域局部特征的提取,EL-Net 可以实现很好的分割性能。图 4-3 进一步对比了 EL-Net 早期阶段和晚期阶段的性能。可以发现模型晚期阶段如果没有早期阶段的介入,分割效果较差;模型早期阶段虽然可以取得不错的结果,但晚期阶段可以提取更多的细节信息,以帮助模型进一步提升模型

性能。

为了直观地观察不同阶段的分割效果，图 4-4 中分别列出了模型早期阶段、晚期阶段和早期-晚期阶段的定量分割结果。由于早期阶段的 Transformer 主干网络具有捕获上下文信息的能力，因此从图 4-4 (b) 可以看出，Transformer 主干网络可以准确地识别出甲状旁腺，但从图 4-4 (b) 中可以看到，Transformer 主干网络的分割结果与标签相比遗失了部分细节特征。图 4-4 (c) 中，由于晚期阶段的主干网络 FR-subnet 感受野有限，晚期阶段在非目标区域产生了大量的错误分割，导致了整体分割性能的严重下滑。然而，结合了早期阶段和晚期阶段的多阶段模型不仅能准确地从混乱的背景中识别出甲状旁腺，而且还能提取到更多细节的特征，如图 4-4 (d) 所示。

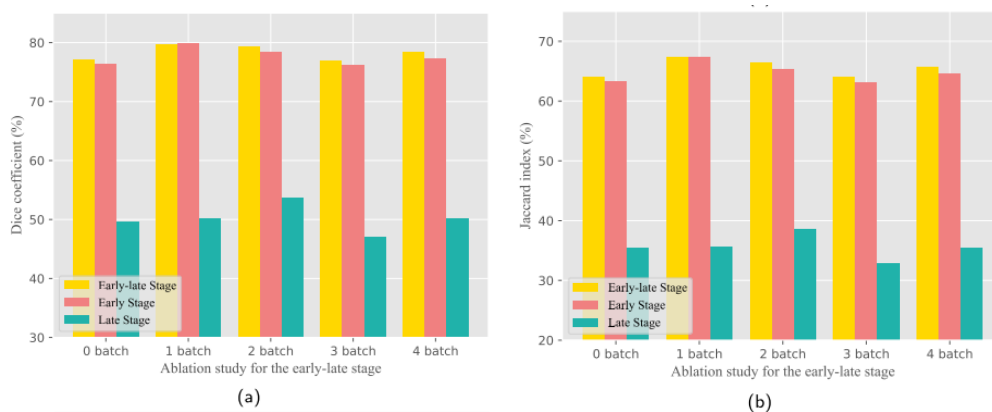


图 4-3 EL-Net 早期阶段和晚期阶段消融实验，其中 Early Stage 代表模型早期阶段，Late Stage 代表模型晚期阶段，Early-late Stage 代表两阶段同时工作。(a)和(b)分别采用 Dice 系数和 Jaccard 指数对两个阶段的性能进行评估。所有实验均在甲状旁腺数据集上进行五折交叉验证

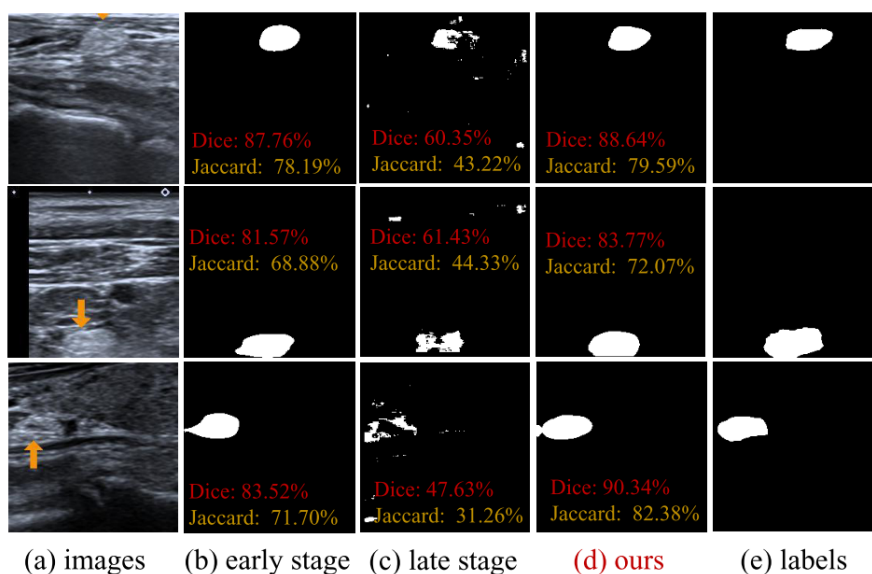


图 4-4 早期阶段和晚期阶段消融实验的定性分割结果

同时, 本节还验证了模型晚期阶段各个模块的有效性。其中, 将 FR-subnet + EAM 作为模型晚期阶段的基线 (Baseline), 不同模块与 Baseline 可以构成几种不同的模型组合。参与模型晚期阶段消融实验的模块包括裁切操作 (Crop, CP)、结构蒸馏模块 (Structured Distillation Module, SDM) 和特征转移模块 (Feature Transfer Module, FTM)。在此基础上, 可以得到五种组合: Baseline, Baseline + CP, Baseline + CP + FTM, Baseline + CP + SDM, Baseline + CP + FTM + SDM。不同组合的比较结果如表 4-3 所示。

表 4-3 在甲状腺数据集上对所提出方法的消融实验的定量结果。其中 Baseline 由 FR-subnet 和边缘感知模块 (EAM) 构成, CP 表示早期阶段的裁剪操作, FTM 表示特征转移模块, SDM 表示“教师伙伴”策略的结构化蒸馏模块。

方法	Dice (%)	Jaccard (%)	Sensitivity (%)
<b>Baseline</b>	50.11	35.59	47.07
<b>Baseline + CP</b>	69.67	54.73	68.84
<b>Baseline + CP + SDM</b>	70.10	55.21	69.76
<b>Baseline + CP + FTM</b>	78.28	65.58	80.10
<b>Baseline + CP + SDM</b>	<b>78.87</b>	<b>66.42</b>	<b>81.62</b>
<b>+ FTM (Ours)</b>			

虽然 Baseline 可以很好地从甲状腺超声图像中提取特征, 但由于超声图像背景复杂, 容易对模型造成一定程度上的误导。因此, Baseline 的 Dice 系数和 Jaccard 指数分别仅为 50.11% 和 35.59%。当将 CP 加入 Baseline 时, Dice 系数和 Jaccard 指数分别可以达到 69.67% 和 54.73%。这是因为通过裁剪操作, 超声图像中大部分有误导性的背景信息被模型的晚期阶段忽略, 降低了模型晚期阶段误分割的概率。此外, SDM 作为“教师-伙伴”策略的核心, 可以将 Transformer 主干的提取到的结构化特征转移到晚期阶段, Baseline + CP + SDM 的 Dice 系数和 Jaccard 指数分别可以达到 70.10% 和 55.21%。此外, FTM 负责将模型早期阶段提取到的上下文信息传送到晚期阶段。与 Baseline 和 Baseline + CP 相比, Baseline + CP + FTM 的 Dice 系数分别提高了 28.17% 和 8.61%。与此同时, 在 Jaccard 指数方面分别实现了 29.99% 和 10.85% 的提升。最后, 通过组合所有模块, EL-Net (Baseline + CP + SDM + FTM) 在甲状腺数据集上取得了最佳的分割性能, 得到的 Dice 系数和 Jaccard 指数分



别为 78.87%和 66.42%。

图 4-5 中比较了消融实验中不同组合对甲状旁腺的分割结果,进一步定性地说明了 EL-Net 每个模块的有效性。在图 4-5 (a) 中,可以观察到颈部甲状旁腺超声影像的背景比较混乱,对分割模型具有一定程度的误导性。因此,在图 4-5 (b) 中, Baseline 不能正确识别甲状旁腺。在图 4-5 (c) 中,由于模型早期阶段的裁剪操作可以有效降低超声图像中复杂背景对模型晚期阶段分割性能的影响, Baseline + CP 的性能相较于 Baseline 的提升明显,但 Baseline + CP 的结果仍然是混淆且不连续的。在图 4-5 (e) 中,由于 FTM 模块将早期阶段 Transformer 主干网络从超声图像中提取到的丰富语义信息输入到模型晚期阶段主干网络 FR-Subnet 中, Baseline + CP + FTM 预测结果得到有效提升。此外,相比于图 4-5 (c) 和图 4-5 (e),在图 4-5 (d) 和图 4-5 (f) 中可以看到,“教师伙伴”策略 (SDM) 的引入使得模型早期阶段分割结果的像素间的结构化信息得以转移到模型的晚期阶段,因此, Baseline + CP + SDM 和 Baseline + CP + SDM + FTM 两种组合可以输出更连续的像素级预测输出。

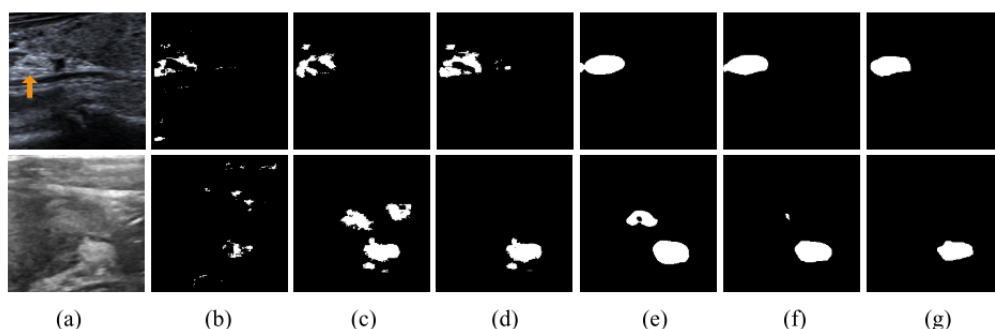


图 4-5 消融实验的定性结果。(a) 超声图像; (b) Baseline; (c) Baseline + CP; (d) Baseline + CP + SDM; (e) Baseline + CP + FTM; (f) Baseline + CP + SDM + FTM; (g)标签

## 4.6 本章小结

本章提出了一种用于术前超声图像中甲状旁腺分割的早期-晚期多阶段网络 (EL-Net)。模型早期阶段可将目标区域从超声图像的复杂背景中分离出来,晚期阶段则可进一步从裁切后的目标区域分割甲状旁腺,得到更为准确的分割结果。该方法有效缓解了超声图像中复杂的背景对模型特征提取过程中负面影响。同时,通过裁切操作,EL-Net 在提取更多细节特征的同时节省了大量计算资源。此外,本章所提出的“教师-伙伴”策略,也为 EL-Net

两个不同阶段之间的像素间结构化信息的交互架起桥梁。本章所提出的 EL-Net 在甲状旁腺数据集上进行了详细的对比实验和消融实验，所有实验结果都证明了 EL-Net 分割性能的优越性和各个模块的有效性。综上所述，该方法可以有效地帮助外科医生在超声影像中进行术前甲状旁腺的分割。

1) 优势。与乳腺肿瘤分割、皮肤病变分割、胃肠道息肉分割等医学图像分割任务一样，甲状旁腺的分割研究也具有重要的临床价值。遗憾的是，在超声图像中没有用于术前甲状旁腺分割的开源数据集。因此，本工作在山东第一医科大学附属山东省医院乳腺甲状腺外科资深医生的帮助下制作了甲状旁腺分割数据集。本章所提出的早晚多阶段网络 (EL-Net) 能够准确地从复杂的超声图像中定位和裁剪目标区域，并通过晚期阶段进一步优化获得更好的分割结果。此外，所提出的“教师-伙伴”策略将早期阶段和晚期阶段的关系在训练前期定义为“教师-学生”，在训练后期转化为“伙伴-伙伴”。该策略可以将像素级的结构化信息从模型的早期阶段传递到模型的晚期阶段，从而进一步提升模型对甲状旁腺的分割质量。

2) 临床价值。到目前为止，术中对甲状旁腺的保护大多依靠外科医生的经验，对于术中极不明显不明显和极度脆弱的甲状旁腺来说是不够的。甲状腺切除术中仍然存在甲状旁腺损伤或意外切除的巨大风险，并会导致严重的医疗事故。因此，在甲状腺切除术前对超声图像中的甲状旁腺进行分割定位是十分必要的。不幸的是，目前很少有研究关注于术前超声图像中甲状旁腺的定位。本章提出的方法能够从复杂的超声图像中准确地分割出甲状旁腺，有助于外科医生在甲状腺切除术前定位和发现甲状旁腺。同时，此方法也可以帮助初级外科医生在超声图像中辨认不明显的甲状旁腺。

## 第五章 总结与展望

### 5.1 本文总结

甲状腺是人体内最重要的内分泌器官，同时，甲状腺结节是临床上最常见的疾病之一。近年来，甲状腺相关疾病的患者不断增加，给临床医生造成了越来越大的压力。甲状腺结节通常情况下无法通过触诊对病情进行准确诊断，大多需要超声检查来辅助临床医生判断甲状腺结节的恶性程度，因此通过超声检查来准确辨别超声影像中的甲状腺结节至关重要。然而，超声影像中不可避免的噪声和特征与甲状腺结节相似的组织对甲状腺结节的诊断造成了极大的困扰，尤其是尺寸相比于良性甲状腺结节小很多的恶性甲状腺结节，其在超声中的特征相比与良性结节非常不明显。此外，对于已经确诊，并需要进行甲状腺切除术的病患，如何避免甲状腺切除术中甲状旁腺的意外损伤和切除，进而有效保护脆弱的甲状旁腺是困扰着临床医生的一大难题。术前使用超声检查对位置不固定的甲状旁腺（下甲状旁腺）进行定位，可以有效地在甲状腺切除术前帮助医生确定手术中的保护区域，从而在术中对甲状旁腺进行有效保护。然而，甲状旁腺体积极小，在超声中的特征极为不明显，这给医生在术前使用超声设备寻找甲状旁腺造成了极大的困难。因此，使用合适的图像分割技术对超声影像中特征不明显的恶性甲状腺结节和甲状旁腺进行准确分割有助于辅助医生进行相关疾病的诊断和治疗，降低误诊、漏诊以及术中误切甲状旁腺的概率。随着计算机技术和深度学习技术的不断发展，使用计算机视觉技术对超声影像中的甲状腺结节和甲状旁腺进行分割，对甲状腺疾病的辅助诊断具有重要意义。本文基于深度学习技术，设计了分别用于超声中恶性甲状腺结节分割和甲状旁腺分割的网络模型，以辅助医生诊断和治疗相关疾病。本文主要工作总结如下：

（1）在山东省立医院乳腺甲状腺外科相关医生的帮助下构建了用于超声中恶性甲状腺结节分割的数据集和甲状旁腺分割的数据集。

（2）提出了双通路镜像 U-Net 模型对超声中恶性甲状腺结节进行准确分割。模型通过 U 形子网提取图像的上下文特征，通过转置 U 形子网提取更多有关恶性甲状腺结节的边缘

细节信息，将提取到的上下文信息和边缘细节信息融合，以获得高精度的恶性甲状腺结节分割结果。同时，本文中将该方法与主流图像分割方法进行了详细的对比，证明了该方法在对恶性甲状腺分割任务上的优越性。

(3) 提出了基于“教师伙伴”策略的早晚多阶段分割网络以在甲状腺切除术前对甲状旁腺进行分割。早晚多阶段网络由两个阶段构成，分别是早期阶段和晚期阶段。早期阶段检测目标区域，并将相关特征图裁剪成小图像块。然后，晚期阶段根据裁剪后的小图像块进行精确的局部分割，并将裁剪后的分割结果恢复到正常大小。此外，该方法使用“教师伙伴”策略，将模型早期阶段提取到的像素间的结构化信息转移到模型晚期阶段，从而进一步提高网络模型的分割性能。该方法与其他主流方法进行了详细的对比实验，证明了该方法在对超声中甲状旁腺分割任务上的优越性。

## 5.2 本文展望

本文针对恶性甲状腺结节和甲状旁腺辅助诊断进行了研究，基于深度学习技术提出了适用于两种不同任务的超声图像分割网络。虽然取得了优异的分割结果，但仍有不足和需要进一步完善的地方，在后续研究中进一步改进。

(1) 模型所需计算成本较大。双通路镜像 U-Net 中的转置 U 形子网与常规 U-Net 不同，其为了提取特征图中更底层的细节信息，要将特征图不断放大，这虽然可以捕捉到更多有助于分割恶性甲状腺结节的细节信息，但也消耗了较多的计算资源。相似的，用于术前甲状旁腺分割的早晚多阶段网络中的晚期阶段网络同样存在此问题。

(2) 数据集较小。数据的采集过程中需要依靠临床医生丰富的临床经验和相关知识储备，且由于临床医生通常工作量较大，导致数据的采集周期较长且数量较少。因此，除了尽可能增加数据集的数据量之外，可以使用半监督和弱监督的方法，降低模型对数据的依赖。

(3) 所研究方法仅针对超声图片。本文研究的两个方法仅仅针对超声图片（视频帧）并没有对超声视频每一帧中的恶性甲状腺结节和甲状旁腺进行准确的实时分割。对超声视频中的恶性甲状腺结节和甲状旁腺进行实时分割将更有利于辅助医生诊断。

## 参考文献

- [1] 中华医学会内分泌学分会, 中华医学会外科学分会内分泌学组, 中国抗癌协会头颈肿瘤专业委员会, 等. 甲状腺结节和分化型甲状腺癌诊治指南[J]. 中国肿瘤临床, 2012, 39(17): 1249-1272.
- [2] Giordano D, Valcavi R, Thompson G B, et al. Complications of central neck dissection in patients with papillary thyroid carcinoma: results of a study on 1087 patients and review of the literature[J]. Thyroid, 2012, 22(9): 911-917.
- [3] Orloff L A, Wiseman S M, Bernet V J, et al. American Thyroid Association statement on postoperative hypoparathyroidism: diagnosis, prevention, and management in adults[J]. Thyroid, 2018, 28(7): 830-841.
- [4] Tjahjono R, Nguyen K, Phung D, et al. Methods of identification of parathyroid glands in thyroid surgery: A literature review[J]. ANZ Journal of Surgery, 2021, 91(9): 1711-1716.
- [5] Lorberboym M, Minski I, Macadziob S, et al. Incremental diagnostic value of preoperative <sup>99m</sup>Tc-MIBI SPECT in patients with a parathyroid adenoma[J]. Journal of Nuclear Medicine, 2003, 44(6): 904-908.
- [6] Johnson N A, Carty S E, Tublin M E. Parathyroid imaging[J]. Radiologic Clinics, 2011, 49(3): 489-509.
- [7] Ha T K, Kim D W, Jung S J. Ultrasound detection of normal parathyroid glands: a preliminary study[J]. La radiologia medica, 2017, 122(11): 866-870.
- [8] Vanderpump M P J. The epidemiology of thyroid disease[J]. British medical bulletin, 2011, 99(1).
- [9] Li H, Weng J, Shi Y, et al. An improved deep learning approach for detection of thyroid papillary cancer in ultrasound images[J]. Scientific reports, 2018, 8(1): 1-12.
- [10] Song W, Li S, Liu J, et al. Multitask cascade convolution neural networks for automatic thyroid nodule detection and recognition[J]. IEEE journal of biomedical and health

- informatics, 2018, 23(3): 1215-1224.
- [11]Ma J, Wu F, Jiang T, et al. Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks[J]. International journal of computer assisted radiology and surgery, 2017, 12(11): 1895-1910.
- [12]La Vecchia C, Malvezzi M, Bosetti C, et al. Thyroid cancer mortality and incidence: a global overview[J]. International journal of cancer, 2015, 136(9): 2187-2195.
- [13]Pellegriti G, Frasca F, Regalbuto C, et al. Worldwide increasing incidence of thyroid cancer: update on epidemiology and risk factors[J]. Journal of cancer epidemiology, 2013, 2013.
- [14]LiVolsi V A. Papillary thyroid carcinoma: an update[J]. Modern Pathology, 2011, 24(2): S1-S9.
- [15]Gyorki D E, Untch B, Tuttle R M, et al. Prophylactic central neck dissection in differentiated thyroid cancer: an assessment of the evidence[J]. Annals of Surgical Oncology, 2013, 20(7): 2285-2289.
- [16]Papini E, Pacella C M, Solbiati L A, et al. Minimally-invasive treatments for benign thyroid nodules: a Delphi-based consensus statement from the Italian minimally-invasive treatments of the thyroid (MITT) group[J]. International Journal of Hyperthermia, 2019.
- [17]Gharib H, Papini E, Paschke R, et al. American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association medical guidelines for clinical practice for the diagnosis and management of thyroid nodules: executive summary of recommendations[J]. Journal of endocrinological investigation, 2010, 33(5): 287-291.
- [18]Haugen B R, Alexander E K, Bible K C, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer[J]. Thyroid, 2016, 26(1): 1-133.
- [19]Horvath E, Majlis S, Rossi R, et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management[J]. The Journal of Clinical Endocrinology & Metabolism, 2009, 94(5): 1748-1751.

- [20] Ying X, Yu Z, Yu R, et al. Thyroid nodule segmentation in ultrasound images based on cascaded convolutional neural network[C]//International Conference on Neural Information Processing. Springer, Cham, 2018: 373-384.
- [21] Zhang Y, Xiang T, Hospedales T M, et al. Deep mutual learning[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4320-4328.
- [22] Wang X, Girshick R, Gupta A, et al. Non-local neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7794-7803.
- [23] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.
- [24] Zhang M N, Liang X Y, Li M T, et al. Current status and temporal trend of disease burden of thyroid cancer in China from 1990 to 2019[J]. Asia-Pacific Journal of Clinical Oncology, 2022.
- [25] Uchino S, Noguchi S. Examination strategy for thyroid nodules[J]. Nihon rinsho. Japanese Journal of Clinical Medicine, 2007, 65(11): 2016-2020.
- [26] Ahmad W, Ahmad A, Lu C, et al. A novel hybrid decision support system for thyroid disease forecasting[J]. Soft Computing, 2018, 22(16): 5377-5383.
- [27] Banu G R. A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease[J]. International Journal of Computer Sciences and Engineering, 2016, 4(11): 64-70.
- [28] Mishra S, Tadesse Y, Dash A, et al. Thyroid disorder analysis using random forest classifier[M]//Intelligent and cloud computing. Springer, Singapore, 2021: 385-390.
- [29] Sankar S, Potti A, Chandrika G N, et al. Thyroid Disease Prediction Using XGBoost Algorithms[J]. Journal of Mobile Multimedia, 2022, 18(3): 1-18.
- [30] Chaubey G, Bisen D, Arjaria S, et al. Thyroid disease prediction using machine learning approaches[J]. National Academy Science Letters, 2021, 44(3): 233-238.
- [31] Asif M A A R, Nishat M M, Faisal F, et al. Computer aided diagnosis of thyroid disease using machine learning algorithms[C]//2020 11th International Conference on Electrical and

- Computer Engineering (ICECE). IEEE, 2020: 222-225.
- [32] Borzouei S, Mahjub H, Sajadi N A, et al. Diagnosing thyroid disorders: Comparison of logistic regression and neural network models[J]. Journal of Family Medicine and Primary Care, 2020, 9(3): 1470.
- [33] Shankar K, Lakshmanaprabu S K, Gupta D, et al. Optimal feature-based multi-kernel SVM approach for thyroid disease classification[J]. The journal of supercomputing, 2020, 76(2): 1128-1143.
- [34] Duggal P, Shukla S. Prediction of thyroid disorders using advanced machine learning techniques[C]//2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2020: 670-675.
- [35] Begum A, Parkavi A. Prediction of thyroid disease using data mining techniques[C]//2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS). IEEE, 2019: 342-345.
- [36] Vairale V S, Shukla S. Classification of Hypothyroid Disorder using Optimized SVM Method[C]//2019 International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, 2019: 258-263.
- [37] Aswathi A K, Antony A. An intelligent system for thyroid disease classification and diagnosis[C]//2018 Second international conference on inventive communication and computational technologies (ICICCT). IEEE, 2018: 1261-1264.
- [38] Tyagi A, Mehra R, Saxena A. Interactive thyroid disease prediction system using machine learning technique[C]//2018 Fifth international conference on parallel, distributed and grid computing (PDGC). IEEE, 2018: 689-693.
- [39] Kousarrizi M R N, Seiti F, Teshnehlal M. An experimental comparative study on thyroid disease diagnosis based on feature subset selection and classification[J]. International Journal of Electrical & Computer Sciences IJECS-IJENS, 2012, 12(01): 13-20.
- [40] Saiti F, Naini A A, Shoorehdeli M A, et al. Thyroid disease diagnosis based on genetic algorithms using PNN and SVM[C]//2009 3rd International Conference on Bioinformatics



and Biomedica.

- [41]Chang C Y, Huang H C, Chen S J. Thyroid nodule segmentation and component analysis in ultrasound images[C]//Proceedings: APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference. Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference, International Organizing Committee, 2009: 910-917.
- [42]Keramidas E G, Maroulis D, Iakovidis D K. TND: a thyroid nodule detection system for analysis of ultrasound images and videos[J]. Journal of medical systems, 2012, 36(3): 1271-1281.
- [43]Kang Q, Lao Q, Li Y, et al. Thyroid nodule segmentation and classification in ultrasound images through intra-and inter-task consistent learning[J]. Medical Image Analysis, 2022, 79: 102443.
- [44]Zhou Y, Xie L, Shen W, et al. A fixed-point model for pancreas segmentation in abdominal CT scans[C]//International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2017: 693-701.
- [45]Yu Q, Xie L, Wang Y, et al. Recurrent saliency transformation network: Incorporating multi-stage visual cues for small organ segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8280-8289.
- [46]Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
- [47]Valanarasu J M J, Sindagi V A, Hacıhaliloğlu I, et al. Kiu-net: Towards accurate segmentation of biomedical images using over-complete representations[C]//International conference on medical image computing and computer-assisted intervention. Springer, Cham, 2020: 363-373.
- [48]Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network[J]. arXiv preprint arXiv:1503.02531, 2015.
- [49]Liu Y, Chen K, Liu C, et al. Structured knowledge distillation for semantic

- segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 2604-2613.
- [50]Torre V, Poggio T A. On edge detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986(2): 147-163.
- [51]Lindeberg T. Scale invariant feature transform[J]. 2012.
- [52]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). IEEE, 2005, 1: 886-893.
- [53]Ahonen T, Hadid A, Pietikainen M. Face description with local binary patterns: Application to face recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2006, 28(12): 2037-2041.
- [54]Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex[J]. The Journal of physiology, 1962, 160(1): 106.
- [55]He K, Zhang X, Ren S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1026-1034.
- [56]Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain[J]. Psychological review, 1958, 65(6): 386.
- [57]Rosenblatt F. Principles of neurodynamics. perceptrons and the theory of brain mechanisms[R]. Cornell Aeronautical Lab Inc Buffalo NY, 1961.
- [58]Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.
- [59]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. Springer International Publishing, 2015: 234-241.

- [60]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.
- [61]Zhou Z, Rahman Siddiquee M M, Tajbakhsh N, et al. Unet++: A nested u-net architecture for medical image segmentation[C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Springer International Publishing, 2018: 3-11.
- [62]Oktay O, Schlemper J, Folgoc L L, et al. Attention u-net: Learning where to look for the pancreas[J]. arXiv preprint arXiv:1804.03999, 2018.
- [63]Chen J, Lu Y, Yu Q, et al. Transunet: Transformers make strong encoders for medical image segmentation[J]. arXiv preprint arXiv:2102.04306, 2021.
- [64]Valanarasu J M J, Oza P, Hacihaliloglu I, et al. Medical transformer: Gated axial-attention for medical image segmentation[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. Springer International Publishing, 2021: 36-46.
- [65]Wu H, Chen S, Chen G, et al. FAT-Net: Feature adaptive transformers for automated skin lesion segmentation[J]. Medical image analysis, 2022, 76: 102327.
- [66]陆剑锋, 林海, 潘志庚. 自适应区域生长算法在医学图像分割中的应用[J]. 计算机辅助设计与图形学学报, 2005(10): 28-33.
- [67]Otsu N. A threshold selection method from gray-level histograms[J]. IEEE transactions on systems, man, and cybernetics, 1979, 9(1): 62-66.
- [68]Chen R, Zhang K, Liu J, et al. Preoperative ultrasound identification and localization of the inferior parathyroid glands in thyroid surgery[J]. Frontiers in Endocrinology, 2023, 14: 415.