

# 图像处理期末作业

题目：医学图像分割算法综述



学 生 姓 名 : 李闯

学 号 : 2253214

专 业 班 级 : 计算机科学与技术 2 班

指 导 教 师 : 何良华

2025 年 06 月 22 日

## 目录

|                            |    |
|----------------------------|----|
| 一、 引言 .....                | 1  |
| 1. 1. 医学影像分割的研究意义 .....    | 1  |
| 1. 2. 医学影像分割面临的主要挑战 .....  | 1  |
| 二、 传统医学影像分割算法 .....        | 2  |
| 2. 1. 基于阈值的分割方法 .....      | 2  |
| 2. 2. 基于区域的分割方法 .....      | 3  |
| 2. 3. 基于边缘检测的分割方法 .....    | 3  |
| 2. 4. 基于图论的分割方法 .....      | 4  |
| 2. 5. 基于活动轮廓的分割方法 .....    | 5  |
| 2. 6. 基于聚类的分割方法 .....      | 5  |
| 三、 基于深度学习的医学影像分割算法 .....   | 6  |
| 3. 1. 全卷积网络 (FCN) .....    | 6  |
| 3. 2. U-Net 架构及其变种 .....   | 7  |
| 3. 3. DeepLab 系列 .....     | 9  |
| 3. 4. Transformer .....    | 10 |
| 3. 5. 生成对抗网络 (GAN) .....   | 12 |
| 四、 未来发展方向 .....            | 12 |
| 4. 1. 跨模态通用模型与统一表示 .....   | 13 |
| 4. 2. 自监督学习与模型可解释性 .....   | 13 |
| 4. 3. 模型轻量化设计与临床应用优化 ..... | 13 |
| 4. 4. 泛化能力增强与实用性提升 .....   | 13 |
| 五、 总结 .....                | 13 |
| 参考文献 .....                 | 15 |

## 一、引言

### 1.1. 医学影像分割的研究意义

随着精准医疗和智能辅助诊断技术的快速发展，医学影像分割作为医学图像分析的核心技术环节，其研究价值与应用潜力日益凸显。在当今医疗环境中，每日产生的医学影像数据呈现爆炸式增长。面对如此庞大的影像数据规模，传统依赖医生人工判读的方式已难以满足临床需求，亟需高效、自动化的分析手段。

除此之外，医学影像具有高维度、模态多样、结构复杂、差异性强等典型特征。不同成像技术（如 CT、MRI、PET、超声等）所呈现的图像在分辨率、对比度、组织边界表现等方面差异显著，而同一模态下，不同患者、不同解剖部位的图像特征也存在巨大变异。医生在面对海量、高异质性的图像数据时，难以依靠肉眼逐一精确分析，尤其是在边界模糊、病灶微小、伪影干扰明显的情况下，人工解读不仅耗时费力，而且易受主观经验影响而产生误差。

医学影像分割技术正是在此背景下应运而生。通过图像处理、机器学习与深度神经网络等方法，分割模型能够自动、准确地提取器官结构、病灶区域和功能区划等语义区域，为图像配准、特征提取、三维重建和临床决策等后续任务提供标准化输入。尤其是在深度学习的推动下，现代分割模型已逐渐具备端到端建模能力，能够适应多模态、多尺度图像数据，显著提升了分割的精度与适应性。

更重要的是，医学影像分割作为交叉领域研究的典型代表，融合了图像处理、计算机视觉、机器学习、医学知识等多个方向的技术成果，不仅推动了人工智能在医疗领域的深度渗透，也带动了多学科协同创新的加速发展。

综上所述，医学影像分割在提升医学图像解读效率、促进临床决策智能化、支撑医学大数据分析以及推动医疗科技创新等方面均发挥着核心作用，是当前医学图像分析领域最具活力与挑战性的研究方向之一，具有重要的科学意义与应用价值。

### 1.2. 医学影像分割面临的主要挑战

医学影像分割技术的发展虽然取得了显著进展，但在实际应用中仍面临诸多关键性挑战，这些挑战主要来自数据、算法和临床应用三个维度。

在数据方面，高质量医学影像标注的高度专业化要求构成了首要障碍。精确的医学图像分割标注必须由经验丰富的临床医师完成，这一过程不仅耗时费力，而且对专业知识要求极高。同时，医疗数据严格的隐私保护政策使得跨机构数据共享面临诸多限制，数据孤岛现象严重制约了大规模高质量数据集的构建。

从算法设计的角度来看，医学影像的特殊性带来了独特的技术难题。不同成像设备产生的多模态数据在特征表现上存在显著差异，这对算法的泛化能力提出了更高要求。此外，医学图像中微小病灶的精确分割、低对比度组织的边界识别以及三维体数据的高效处理等问题，都需要开发更先进的算法来解决。

临床应用场景对分割技术提出了最为严格的要求。医疗诊断对分割精度有着近乎苛刻的标准，特别是在手术规划和治疗评估等关键环节。实时性要求是另一个重要挑战，许多临床应用场景需要算法在极短时间内完成处理。此外，医疗设备的多样性导致的性能差异问题也不容忽视，同一算法在不同设备上的表现可能存在显著波动。

这些挑战的存在，促使研究者不断探索创新方法和技术路线，持续推动医学影像分割领域的技术进步。

## 二、传统医学影像分割算法

在深度学习方法尚未普及之前，传统图像分割技术在医学图像分析中发挥了重要作用，形成了多个具有代表性的分割策略。这些方法各具原理，覆盖了从图像灰度特性、区域一致性、边缘信息到图结构建模和能量优化等多个角度，构建了医学图像分割方法论的早期基础。以下将依次介绍几类经典的传统医学影像分割方法，包括基于阈值、区域、边缘、图论、活动轮廓和聚类的分割技术，以全面理解其原理特点、应用场景及存在的局限性。

### 2.1. 基于阈值的分割方法

基于阈值的分割方法是医学图像分割中最基础的技术之一，其核心思想是通过设定一个或多个灰度阈值，将图像像素分为不同的类别。这种方法假设不同的组织结构在图像灰度分布上具有可区分的特征，通过选择合适的阈值即可实现目标区域与背景的分离。

该方法的关键在于如何确定最优的阈值。理想的阈值应当能够最大程度地区分不同组织结构的灰度特征。在医学图像处理中，常用的阈值确定方法主要包括：

1. 最大类间方差法（大津法）：这是一种基于灰度直方图的自适应阈值选择方法。它通过最大化类间方差来确定最佳分割阈值[2]，特别适用于具有双峰直方图的医学图像。例如在骨骼 CT 图像中，可以有效地分离骨组织与软组织。
2. 最大熵阈值法：该方法基于信息熵理论，通过最大化分割后图像的信息量来确定阈值[3]。这种方法在低对比度医学图像（如乳腺 X 光片）的分割中表现良好。
3. 基于模糊熵的阈值法：这是一种结合模糊集理论的改进方法。Cheng 等人[4]提出的模糊熵阈值法通过引入隶属度函数来处理医学图像中的灰度不确定性，特别适用于边界模糊的软组织分割。该方法首先定义像素属于前景和背景的模糊隶属度，然后通过最大化模糊熵来确定最优阈值。

尽管基于阈值的分割方法在医学图像处理中具有计算简单、实现高效等优势，但其在实际应用中仍存在明显的局限性。首先，这类方法对图像质量要求较高，当医学图像中存在严重的噪声干扰、伪影或强度不均匀性时，阈值选择的准确性会显著下降。其次，对于灰度分布重叠严重的组织结构，基于全局灰度特征的阈值方法往往难以获得理想的分割效果。此外，传统的阈值分割方法通常无法充分利用医学图像中的空间上下文信息，导致在复杂解剖结构分割时容易产生不连续的分割边界。最后，多数阈值方法需要预先确定分割类别数目，这在

面对未知的医学图像分析任务时可能带来额外挑战。

## 2.2. 基于区域的分割方法

区域生长算法是一种基于像素相似性的迭代分割方法。该方法首先需要选取一个或多个种子点作为起始区域，然后根据预先定义的相似性准则（如灰度值、纹理特征等）逐步将相邻像素合并到当前区域中[5]。在医学图像分割中，这种方法的有效性很大程度上依赖于三个关键因素：种子点的准确定位、合适的相似性阈值设定以及合理的终止条件。在医学图像分析中，区域生长算法被成功应用于多种组织的分割任务，如肺部结节检测和肝脏分割等。然而，该算法对与初始条件较为敏感，初始种子点的微小偏差可能导致完全不同的分割结果，且在处理弱边界和异质区域时容易产生渗漏现象。

分裂合并算法采用了一种完全不同的分割策略。该算法首先将整个图像视为一个初始区域，然后通过递归地将异质区域分裂为更小的同质子区域，最后再根据相似性准则合并相邻的相似区域。这种方法不需要预先指定种子点，能够自动处理图像中的多个目标区域，因此在脑部MRI等多组织分割任务中表现出色。分裂合并算法通过建立区域邻接图来管理分割过程，其中区域相似性的计算通常基于统计特征或纹理描述子。然而，这类算法在确定最优的分裂和合并准则方面仍面临挑战，特别是在处理医学图像中常见的噪声和部分容积效应时。

在实际应用中，常将这两种方法与其他技术相结合以提高分割性能。例如，刘莹等人[6]将区域生长与水平集方法相结合，实现了宫颈癌病变区域的准确分割，该方法不仅保持了病变区域的完整性，还显著提升了边界定位的准确性。范群贞等人[7]则在肺CT图像分割中采用区域生长与数学形态学相结合的策略，通过形态学操作有效修正了初始分割结果中的误差，获得了更加精确的分割效果。

## 2.3. 基于边缘检测的分割方法

边缘检测是医学图像分割中一类重要的方法，其核心思想是通过识别图像中灰度的突变位置来确定目标边界。这类方法基于微分运算原理，通过计算图像的一阶或二阶导数来检测边缘位置。Sobel算子作为典型的边缘检测方法，采用 $3 \times 3$ 的卷积核计算图像梯度，具有实现简单、计算效率高的特点，常被用于医学图像的初步边缘提取。

然而，医学图像的特殊性给边缘检测带来了显著挑战。医学影像通常包含各种噪声和伪影，这会导致传统边缘检测方法产生大量虚假边缘。同时，由于部分容积效应的影响，不同组织间的边界往往呈现渐进式过渡，而非理想的阶跃边缘。此外，成像参数的变化和患者个体差异也会影响边缘检测的稳定性。

针对这些问题，Canny边缘检测器通过多阶段处理流程有效提升了边缘检测质量。该方法首先使用高斯滤波器对图像进行平滑处理以抑制噪声，随后计算梯度幅值和方向信息。在边缘细化阶段，非极大值抑制技术被用于消除冗余响应，保留真正的边缘点。最后，通过双阈值策略和边缘连接处理，得到完整连续的边缘轮廓。这种系统化的处理方式使其在脑部MRI等医学图像分析中展现出优越性能。

另一类重要的边缘检测方法是基于二阶导数的拉普拉斯算子。该方法通过寻找图像二阶导数的过零点来定位边缘，能够提供更精确的边缘位置信息。然而，拉普拉斯算子对噪声极为敏感，因此通常需要与高斯平滑相结合，形成 LoG (Laplacian of Gaussian) 算子。这种改进方法在胸部 CT 图像的肺结节检测中表现出良好的性能。

#### 2.4. 基于图论的分割方法

基于图论的医学图像分割方法通过将图像抽象为图结构来实现分割，将像素或超像素作为图的顶点，相邻关系作为边，边的权重反映像素间的相似性，从而将图像分割问题转化为图的划分问题。这类方法主要包括 Graph Cut (最小割/最大流)、归一化切割算法 (Normalized Cut) 以及随机游走算法等，各具特色且在不同医学图像分析场景中展现出独特优势。

Graph Cut 算法 [9] 是最早应用于医学图像分割的图论方法之一。该算法的核心思想是将图像建模为一个带权图，其中每个像素对应一个普通顶点，同时引入两个特殊终端节点：源节点 S (代表前景) 和汇节点 T (代表背景)。所有普通顶点均与 S 和 T 相连，形成 t-links (终端连接)，而相邻像素顶点之间通过 n-links (邻域连接) 建立边。这种图结构的关键在于边的权重设计：t-links 的权重由区域项决定，反映像素属于前景或背景的概率；n-links 的权重则由边界项控制，通常基于像素间的灰度或特征相似性采用高斯核函数计算。

通过这样的图结构转换，图像分割任务就被转化为一个最小割问题。具体来说，我们需要在图结构中找到一个割集，使得该割集将图分成包含 S 节点的前景区域和包含 T 节点的背景区域，同时使得割集中所有边的权重之和（即割的代价）最小。这个最小割问题可以通过最大流算法来求解，其中 Boykov-Kolmogorov 最大流算法因其高效性而被广泛采用。该算法通过迭代地寻找增广路径并更新流量，最终能够找到一个最优或近似最优的分割方案，从而实现对医学图像的准确分割。这种方法在交互式医学图像分割中表现优异，特别适用于脑部 MRI 和肿瘤分割等应用场景。

与 Graph Cut 不同，归一化切割算法 [11] 通过计算归一化割值来划分图像，避免产生过小的分割区域。算法首先建立像素间的相似度矩阵，然后求解对应的特征值问题，最后对特征向量进行聚类得到分割结果。这种方法不需要人工指定种子点，能够自动完成图像分割。在医学图像分析中，NCut 算法经过优化后可以应用于各类组织的分割任务。

随机游走 (Random Walker) 算法 [10] 是另一类重要的基于图论的方法，采用完全不同的概率框架。该方法将图像分割视为一个标记传播问题，通过计算每个未标记像素到达不同标记种子点的概率来确定其类别。相比于图割方法，随机游走算法对种子点位置的选择不敏感，且在弱边界处理上表现更优，特别适用于心脏 MRI 等具有复杂结构的器官分割。

尽管基于图论的方法在医学图像分割中取得了显著成功，但仍面临计算复杂度高、权重设计复杂等挑战。特别是在处理高分辨率三维医学图像时，图结构的构建和优化需要大量计算资源。此外，传统图论方法通常需要人工提供种子点，这在一定程度上限制了其自动化程度。未来研究需要进一步优化算法效率，探索自适应权重设计方法，以提高其在复杂医学图

像分析中的适用性。

## 2.5. 基于活动轮廓的分割方法

基于活动轮廓 (Active Contour) 的分割方法是医学图像分析中一类重要的边界检测技术。这类方法的核心思想是通过定义一条可变形曲线 (称为轮廓)，在能量最小化原理的驱动下，使曲线逐渐收敛到目标结构的边界位置。其数学本质是将分割问题转化为一个能量泛函的优化问题，根据曲线表示方式的不同，主要可分为参数活动轮廓模型 (Parametric Active Contour) 和几何活动轮廓模型 (Geometric Active Contour) 两大类别。

参数活动轮廓模型 (Parametric Active Contour Model) 由 Kass 等人在 1987 年提出 [13]，又称 Snake 模型。该方法采用参数方程显示表示曲线，通过最小化包含内部能量和外部能量的目标函数实现分割。其中，内部能量用于保证曲线的光滑性和连续性，通常由弹性和弯曲力构成；外部能量则引导曲线向图像特征 (如边缘) 移动。该模型通过梯度下降法迭代求解能量最小化问题，具有计算效率高的优点，但对初始轮廓位置敏感，且难以处理拓扑结构变化 (如轮廓分裂或合并)。

几何活动轮廓模型采用水平集 (Level Set) 方法隐式表示曲线 [14]，通过求解偏微分方程实现曲线演化。其优势在于能自动处理拓扑变化 (如分裂、合并)，适用于复杂结构的分割。根据驱动机制的不同，主要分为两类：

1. 基于边缘的模型：主要利用图像梯度信息构造速度函数，驱动轮廓向目标边界演化。Caselles 等人 [15] 提出的测地线活动轮廓 (Geodesic Active Contour, GAC) 是这种方法的典型代表，该模型通过梯度阈值来控制曲线的演化过程。

2. 基于区域的模型：通过分析图像区域统计特征 (如强度分布、纹理) 来实现分割。Chan & Vese [16] 提出的 Chan-Vese 模型 (无边缘活动轮廓) 是最经典的区域方法，通过最小化区域内方差实现分割，对弱边缘或均匀区域目标具有更好的鲁棒性。

## 2.6. 基于聚类的分割方法

基于聚类的图像分割方法是一类重要的无监督分割技术，其核心思想是将图像中具有相似特征的像素或区域自动归类到同一类别中。这类方法不依赖于先验的形状或边界信息，而是通过分析像素的特征相似性来实现分割，在医学图像处理中具有广泛的应用。

其中，K-means 聚类是最经典的算法，由 MacQueen 于 1967 年提出 [17]，该方法通过迭代优化将像素分配到最近的聚类中心，直到中心点稳定。虽然计算高效，但需要预先指定聚类数目且对初始中心敏感。模糊 C 均值聚类 (FCM) 由 Dunn [18] 和 Bezdek [19] 在 1973–1981 年间发展完善，通过引入模糊隶属度的概念，允许像素以不同概率属于多个类别，特别适合处理医学图像中常见的边界模糊问题。Fukunaga 和 Hostetler [20] 于 1975 年提出的均值漂移聚类采用非参数的概率密度估计方法，通过寻找特征空间中密度函数的局部最大值来自动确定类别数量，对噪声具有较好的鲁棒性。在具体实现上，这些算法都需要定义适当的特征空间和相似性度量。K-means 通常使用像素强度作为特征，采用欧氏距离度量相似性；FCM 则在目标

函数中引入隶属度权重，通过交替优化隶属度矩阵和聚类中心来实现分割；均值漂移聚类则通过核密度估计和梯度上升来寻找模态点。

这些基于聚类的方法在医学图像分析中展现出独特优势，但也面临一些挑战。虽然它们不需要训练数据且适用于多种模态，但对噪声敏感、可能产生过分割等问题仍需解决。近年来，研究者通过优化特征表示、改进距离度量等方式不断提升算法性能。例如，Zhang 等[21]在 2008 年提出的核化 FCM 算法提高了对复杂组织的分割能力；Cai 等人[22]在 2013 年发展的快速鲁棒聚类算法显著提升了计算效率。这些进展使得基于聚类的方法在医学图像分割领域持续发挥重要作用，为后续结合深度学习方法奠定了基础。

### 三、基于深度学习的医学影像分割算法

随着深度学习技术的迅猛发展，医学图像分割领域迎来了革命性的变革。相较于传统分割方法，深度学习算法展现出多方面的显著优势：首先，深度学习模型具备强大的特征自学能力，能够自动提取图像的多层次特征，克服了传统方法依赖人工设计特征的局限性；其次，通过端到端的训练方式，深度学习可以直接优化分割目标，避免了传统方法中多步骤处理带来的误差累积问题；再者，深度学习模型对图像噪声和伪影具有更强的鲁棒性，在处理低对比度医学图像时表现尤为突出。

本节将系统梳理当前主流的深度学习分割网络结构及其关键技术，全面介绍其设计思路与实现方法。

#### 3.1. 全卷积网络（FCN）

全卷积网络（Fully Convolutional Network, FCN）是第一个实现端到端像素级分割的深度学习模型，由 Long 等人于 2015 年首次提出[28]。其核心思想是将传统图像分类网络中的全连接层替换为卷积层，使得网络可以接受任意尺寸的输入图像，并生成与输入大小一致的密集预测输出，从而实现像素级的语义分割。

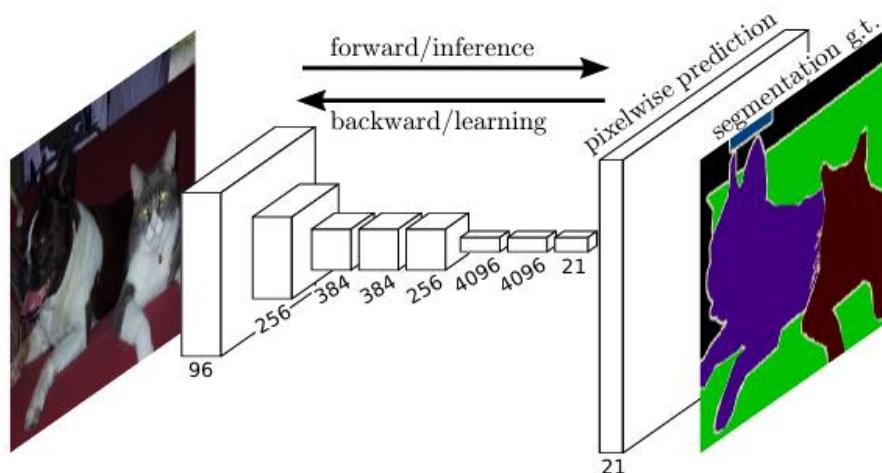


图 1 全卷积网络架构示意图

FCN 的网络结构主要包括两个阶段：特征提取阶段和上采样阶段。在特征提取阶段，FCN 通常借助预训练的分类网络（如 VGG、ResNet 等）提取深层特征。在上采样阶段，FCN 通过反卷积（转置卷积）操作对低分辨率特征图进行逐步恢复，同时融合来自浅层的特征图以增强空间定位能力。这种“粗到细”的上采样方式使得 FCN 具备了一定程度的空间分辨能力，适用于图像中的结构化预测任务。

在医学图像分割任务中，FCN 被广泛用于器官轮廓提取、病灶检测等场景，尤其在结构相对清晰的影像（如胸部 X 光、肝脏 CT）中表现良好。例如，Shelhamer 等人将 FCN 应用于脑部 MRI 图像的肿瘤区域分割，取得了比传统方法更高的准确率[29]。由于 FCN 具备端到端训练、无需手工特征设计的优点，其提出一度成为后续分割模型（如 U-Net、SegNet、DeepLab 等）的基础。

然而，FCN 在医学图像分割中的应用也存在一定局限性。首先，由于其上采样过程相对简单，往往难以准确恢复边界信息，导致分割结果模糊。其次，FCN 缺乏显式的结构对齐机制，在处理小目标或边界复杂区域时表现不稳定。此外，由于语义特征主要来源于深层卷积，容易忽视浅层的细粒度信息，这在处理医学图像中小尺度、低对比度结构时尤其明显。

尽管如此，FCN 的提出仍具有里程碑意义，它首次将卷积神经网络拓展到像素级分割任务，开启了语义分割研究的新阶段。其架构理念被大量后续模型继承和扩展，成为现代医学图像分割算法的重要基础。

### 3.2. U-Net 架构及其变种

U-Net 是最早专为医学图像分割任务设计的深度卷积神经网络之一，由 Ronneberger 等人于 2015 年提出[24]，并在细胞图像分割挑战赛中取得了优异成绩。U-Net 的提出标志着深度学习方法在医学图像分割领域的广泛应用起点。该网络采用“编码器-解码器”结构，通过对称的网络设计和跳跃连接机制，有效整合了局部细节与全局语义信息，实现了高精度的像素级分割。

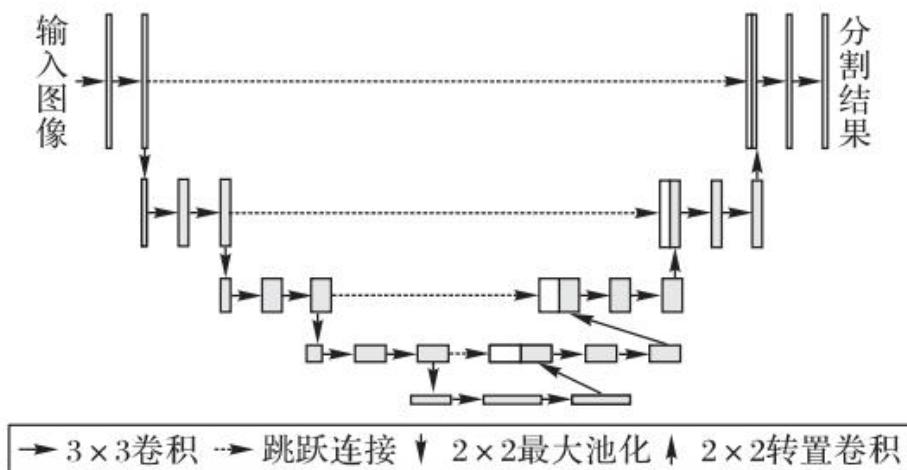


图 2 U-Net 模型网络架构图

U-Net 的结构由两部分组成：编码器部分负责提取图像的深层语义特征，通常由多层卷积与池化操作组成；解码器部分则逐步上采样恢复空间分辨率，并结合编码器的对应层特征（通过跳跃连接）来增强边界和细节信息的还原能力。这种结构使得 U-Net 在处理高分辨率图像、细粒度目标以及边界模糊的区域时具有显著优势。

为了进一步提升 U-Net 在复杂医学图像场景中的性能，研究者基于其原始结构提出了多种改进型架构：

1. Attention U-Net：Oktay 等人[25]引入注意力机制，使得网络能够自动关注于目标区域，对小病灶或组织边界模糊的情况尤为有效。注意力门控机制可以在跳跃连接中动态调整特征通道的权重，过滤掉不相关的信息，从而增强感兴趣区域的特征表达能力。

2. Residual U-Net：该变种结合了 ResNet 的残差模块，通过引入残差连接缓解了深层网络训练中的梯度消失问题[26]。残差结构的加入不仅提高了训练稳定性，也使网络具备更强的特征学习能力，尤其适合复杂解剖结构的多级表示。

3. UNet++：Zhou 等人提出的 U-Net++在编码器和解码器之间引入了密集跳跃连接，并采用深度监督策略[27]，增强了特征图之间的语义一致性。该模型通过逐层重构融合路径，改善了原始 U-Net 中直接拼接可能引起的语义鸿沟问题，特别适合于需要高精度分割的场景，如肿瘤边界提取。

4. 三维度扩展变体：针对体数据处理的 3D U-Net[33]和 V-Net[34]开创了三维分割新范式。3D U-Net 采用全三维卷积操作，在 MICCAI 胰腺分割挑战赛夺冠；V-Net 则创新性地直接优化 Dice 损失函数，为前列腺 MRI 分割设立新基准。这些变体推动医学图像分析从二维切片进入真正的三维时代。

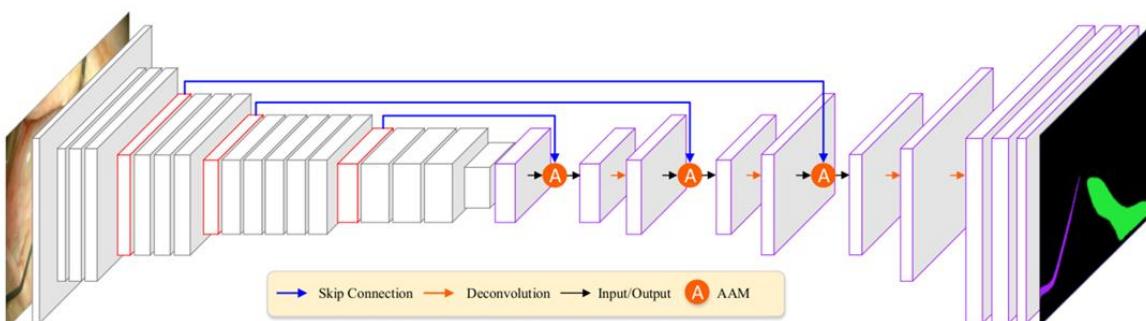


图 3 Residual U-Net 模型网络架构图

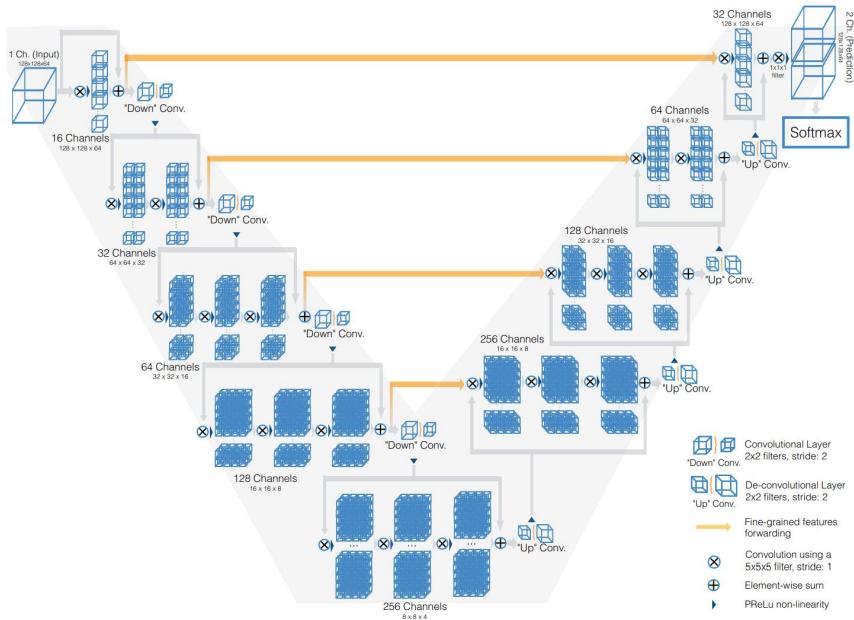


图 4 V-Net 模型网络架构图

这些 U-Net 变种在多个医学图像分割任务中均展现出优越的性能。例如，在肺结节分割、脑肿瘤定位、视网膜血管提取等任务中，改进型 U-Net 模型通常能够取得更高的 Dice 系数和 IoU 指标，尤其在处理边界模糊、目标小而密集、组织形态多变的医学图像中具有良好适应性。总的来看，U-Net 及其变种模型已成为医学图像分割的“标准架构”，其灵活的结构设计和强大的可扩展性使其成为构建新型分割模型的重要基础。

### 3.3. DeepLab 系列

DeepLab 系列是 Google 提出的一系列用于图像语义分割的深度神经网络架构，其在保持较高分割精度的同时兼顾了计算效率，对医学图像中的复杂边界和多尺度结构表现尤为出色。自最初的 DeepLabv1 提出以来，整个系列不断演化，至今已发展至 DeepLabv3+，成为语义分割领域的经典代表之一。

DeepLab 系列的核心技术之一是空洞卷积（Atrous Convolution），又称为扩张卷积。该技术通过在卷积核中插入“空洞”，即在元素之间加入间隔，有效地扩大了感受野而不增加参数量或计算复杂度[30]。这种方式能够捕捉更大范围的上下文信息，非常适合医学图像中组织结构尺度差异大的场景，如同时检测大器官与小病灶。

在 DeepLabv1 中，空洞卷积被用于替代传统的卷积层，从而在不降低图像分辨率的前提下提取更丰富的语义特征。随后，DeepLabv2 进一步结合空洞空间金字塔池化（ASPP）模块，通过并行多个不同空洞率的卷积操作，从多个尺度提取上下文信息，提高了模型对目标尺度变化的鲁棒性[31]。

在早期版本中，DeepLab 还引入了全连接条件随机场（DenseCRF）作为后处理模块，用于细化分割边界。由于卷积特征存在空间信息丢失的问题，直接输出的分割图往往边界模糊，

尤其是在医学图像中肿瘤边界、器官轮廓等细节敏感的任务中。CRF 通过考虑像素之间的颜色和位置相似性，实现对分割结果的边缘对齐，显著提升分割的细节保真度。

在 DeepLabv3 中，模型进一步强化了 ASPP 模块，通过串联多个不同空洞率的卷积层，并引入全局平均池化增强全局感知能力。该版本不再使用 CRF，而是将上下文建模完全交由网络完成，简化了模型结构，提升了端到端训练能力[32]。

DeepLabv3+[35]在此基础上引入编码器-解码器结构，结合 U-Net 式的低级特征与高级语义特征，改善了上采样过程中对空间细节的恢复效果。通过引入浅层的边界信息和深层的语义特征，DeepLabv3+在分割精度和效率之间取得了更优平衡，广泛应用于脑部 MRI、肺 CT、视网膜 OCT 等医学图像的分割任务。

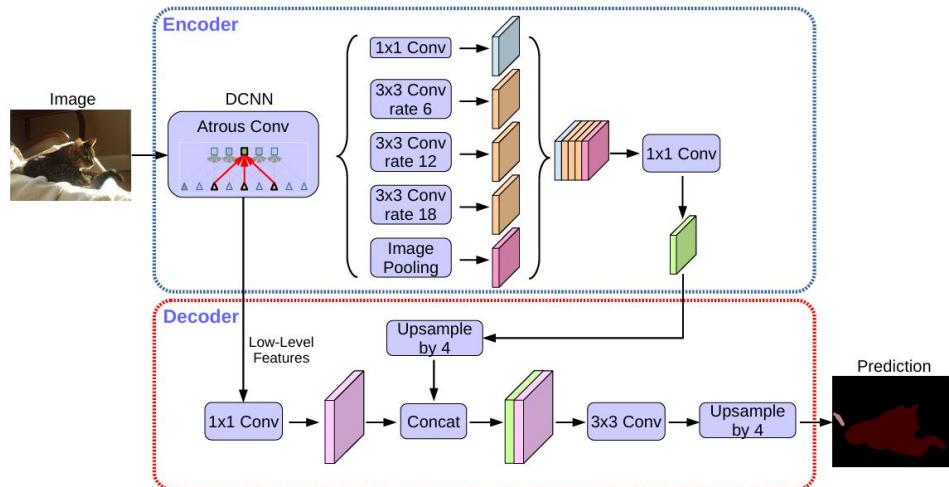


图 5 DeepLabv3+模型架构图

总体而言，DeepLab 系列通过多尺度感知、边界优化与语义增强等手段，在医学图像分割中展现出强大潜力，是当前深度学习分割框架中的重要基准模型。

### 3.4. Transformer

Transformer 最初由 Vaswani 等人于 2017 年提出 [36]，广泛应用于自然语言处理任务，凭借其强大的全局建模能力迅速在计算机视觉领域引起关注。随着 Vision Transformer (ViT) [37] 及其变种的提出，研究者逐渐将 Transformer 结构引入医学图像分割任务，解决卷积神经网络 (CNN) 在建模长程依赖和上下文信息整合方面的局限性。

Transformer 通过自注意力机制 (Self-Attention) 在全局范围内计算特征间的相关性，具备更强的上下文建模能力，尤其适合处理形状不规则、结构复杂的医学图像目标。例如在脑部 MRI 或肿瘤区域分割任务中，Transformer 能够捕捉远距离像素间的联系，弥补 CNN 局部感受野的不足。

当前主流的医学分割 Transformer 模型主要采用三种技术路线：1) 纯 Transformer 架构 (如 Swin-UNet [38]) 完全摒弃卷积操作，通过图像分块和位置编码实现序列建模；2) 混合

架构（如 TransUNet[39]）在 CNN 特征提取基础上引入 Transformer 编码器，兼顾局部特征与全局上下文；3) 轻量化设计（如 PVT[40]）则通过空间缩减注意力等创新大幅降低计算复杂度。

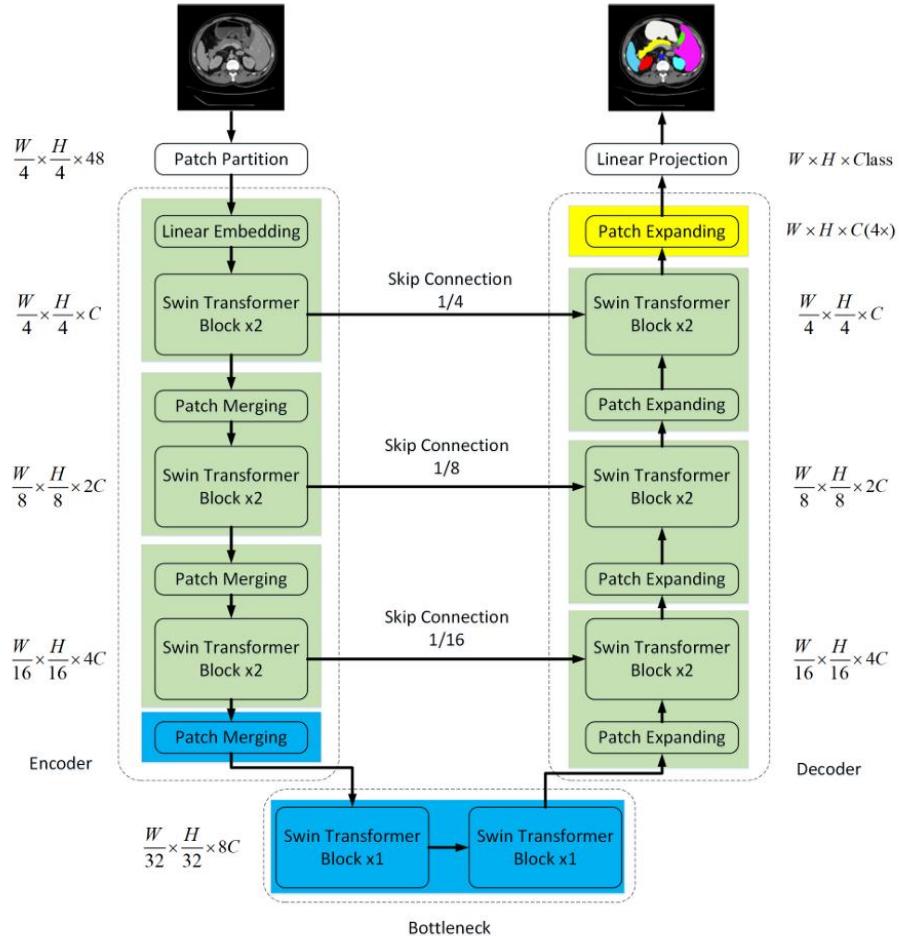


图 6 Swin-Unet 模型架构图

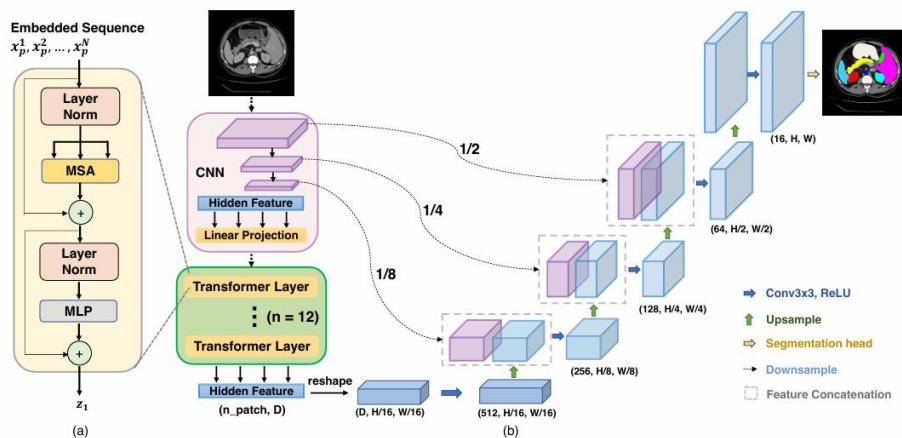


图 7 TransUNet 模型架构图

Transformer 架构在医学图像分割领域的应用虽然展现出巨大潜力，但仍面临诸多关键性挑战。首先，数据需求方面的矛盾尤为突出，这类模型通常需要海量训练样本才能达到理想性能，而医学领域高质量标注数据的获取却极为困难且成本高昂。其次，计算资源消耗问题不容忽视，处理高分辨率医学影像时会产生巨大的计算负担，严重制约了模型的临床应用可行性。再者，模型可解释性的不足也构成重要障碍，传统注意力机制缺乏直观的医学语义关联，难以获得临床医生的充分信任。此外，医学图像特有的模态多样性、解剖结构复杂性以及病理表现异质性等问题，也对模型的泛化能力提出了更高要求。最后，隐私保护和伦理规范等现实因素，也在一定程度上限制了大规模医学数据的获取与共享。这些挑战共同构成了 Transformer 在医学领域深入应用需要突破的关键瓶颈。

### 3.5. 生成对抗网络（GAN）

生成对抗网络（GAN）在医学图像分割领域展现出独特价值，其核心在于生成器与判别器的对抗训练机制[41]。通过二者的动态博弈，GAN 能够有效提升医学图像分割的精度，特别是在处理边界模糊或数据稀缺的情况下。早期研究如 Xue 等人提出的 SegAN 模型[42]首次将对抗训练引入医学图像分割，采用 U-Net 结构的生成器配合多尺度判别器，显著提升了脑肿瘤分割的边缘精度。随后的条件 GAN 研究进一步拓展了应用范围，Isola 团队开发的 pix2pix 框架[43]通过将原始影像作为条件输入，在视网膜血管分割等任务中取得了突破性进展[2]。这些开创性工作为后续研究奠定了重要基础。

针对医疗领域标注数据稀缺这一核心挑战，研究者们开发了多种创新解决方案。Mondal 等人提出的半监督 GAN 方法通过同时利用标注和未标注数据进行训练，在仅使用 20% 标注数据的胸部 X 光分割任务中取得了接近全监督模型的性能[44]。Zhu 团队开发的 CycleGAN 框架则开创性地实现了跨模态的无监督域适应，为不同成像设备获取的医学影像分割提供了新的解决思路[45]。这些方法有效降低了模型训练对标注数据的依赖，极大提升了模型在真实医疗场景中的适用性。

在三维医学影像处理方面，GAN 技术也取得了显著突破。Fidon 等人开发的 Vox2Vox 模型采用 3D U-Net 生成器架构，在脑肿瘤分割任务中展现出优于传统 2D 方法的性能[46]，这为处理具有复杂空间结构的器官分割问题提供了新的工具。然而，该领域仍面临评估标准不统一、计算成本高昂等挑战，未来的研究需要建立更专业的医疗影像评估体系，并优化模型的计算效率。

## 四、未来发展方向

医学图像分割技术正处于快速演进阶段，其发展路径也从早期的传统图像处理方法逐步迈向以深度学习为核心的智能分割模型。尽管深度学习已在多个分割任务中取得显著进展，但仍存在一定的局限性，如对大量标注数据的依赖、泛化能力不足、缺乏可解释性等。面向未来，医学图像分割将进一步迈向多模态融合、智能化、轻量化和实用化，主要包括以下几个方向：

## 4.1. 跨模态通用模型与统一表示

医学成像技术呈现出多样性和复杂性，不同模态（如 CT、MRI、PET、超声）各自具有独特的图像特征和临床价值。传统分割方法多为单模态设计，无法有效整合多源信息。未来的研究将重点聚焦于构建多模态协同分析模型，通过统一的特征编码器、模态对齐机制和跨模态知识迁移策略，实现不同模态间的深度融合和互补优势的发挥。特别是基于 Transformer 的结构因其卓越的全局建模能力，预计将在这一方向成为核心技术框架。

## 4.2. 自监督学习与模型可解释性

数据标注的高成本一直是医学 AI 发展的瓶颈，自监督、弱监督等智能学习方法为缓解该问题提供了新思路。未来分割模型将进一步融合自监督预训练、对比学习、图文联合建模等机制，以最大程度挖掘未标注数据的潜在结构信息，减少对人工标注的依赖。同时，模型可解释性将成为临床采纳 AI 系统的重要保障。通过注意力热图、显著区域可视化、逻辑推理路径追踪等方法，构建可理解、可追溯的决策过程，增强医生对 AI 模型的信任与接受度。

## 4.3. 模型轻量化设计与临床应用优化

目前的大规模分割模型通常计算复杂、参数众多，难以直接应用于资源受限的设备。未来将加快向轻量化、低功耗、高效能方向转型。研究将重点关注网络剪枝、权重量化、知识蒸馏等模型压缩技术，同时探索新型高效架构在医学图像分割中的应用，推动模型在边缘计算设备、移动终端、床旁设备等场景下的实际部署，满足实时性与可移植性的临床需求。

## 4.4. 泛化能力增强与实用性提升

医学影像在设备、医院、区域等层面存在显著异质性，模型的泛化能力成为影响实用化的重要瓶颈。未来的研究将致力于构建跨域鲁棒的分割模型，采用领域自适应、数据增强、跨中心对比学习等手段，缓解模型对特定数据分布的过拟合问题。同时，通过引入联邦学习、隐私计算等机制，促进多中心协同训练，在保障数据安全的前提下提升模型泛化性能。最终目标是构建具有普适性和可持续迭代能力的分割系统，推动其在大规模临床场景下稳定运行与持续优化。

## 五、总结

医学图像分割技术作为医学影像分析的核心环节，经历了从传统图像处理方法到深度学习模型的革命性发展。传统方法如基于阈值、区域、边缘、图论、活动轮廓和聚类的分割算法，为早期医学图像分析奠定了重要基础，但在处理复杂医学影像时仍存在诸多局限性。随着深度学习技术的兴起，全卷积网络、U-Net 及其变种、DeepLab 系列、Transformer 以及生成对抗网络等先进模型相继涌现，显著提升了分割精度和鲁棒性，推动了医学图像分割技术的快速发展。

当前，医学图像分割技术已广泛应用于肿瘤检测、器官分割、血管提取等临床场景，为精准医疗和智能辅助诊断提供了有力支持。然而，该领域仍面临数据标注成本高、模型泛化

能力不足、计算复杂度高、可解释性欠缺等挑战。未来的研究将聚焦于跨模态通用模型构建、自监督学习、模型轻量化设计以及临床应用优化等方向，通过技术创新解决现有瓶颈问题。

展望未来，医学图像分割技术将继续向着智能化、精准化、实用化的方向迈进。随着多学科交叉融合的不断深入，新一代分割算法将更加注重与临床需求的紧密结合，在提升技术性能的同时，兼顾模型的可靠性、可解释性和实用性。通过算法研究者与临床医生的紧密协作，医学图像分割技术有望在疾病早期诊断、手术规划、疗效评估等更多医疗场景中发挥关键作用，最终为提升医疗服务质量、改善患者预后做出重要贡献。

## 参考文献

- [1] 孙淑婷,刘铖枨,周广茵,等.图像分割算法在医学图像中的应用综述[J].现代仪器与医疗,2024,30(02):59-68.
- [2] Otsu N. A threshold selection method from gray-level histograms[J]. Automatica, 1975, 11(285-296): 23-27.
- [3] Kapur J N, Sahoo P K, Wong A K C. A new method for gray-level picture thresholding using the entropy of the histogram[J]. Computer vision, graphics, and image processing, 1985, 29(3): 273-285.
- [4] Cheng H D, Chen J R, Li J. Threshold selection based on fuzzy c-partition entropy approach[J]. Pattern recognition, 1998, 31(7): 857-870.
- [5] Adams R, Bischof L. Seeded region growing[J]. IEEE Transactions on pattern analysis and machine intelligence, 1994, 16(6): 641-647.
- [6] 刘莹,李筠,杨海马,等.结合区域生长与水平集算法的宫颈癌图像分割[J].电子测量与仪器学报,2020,34(09):146-152.DOI:10.13382/j.jemi.B1902879.
- [7] 范群贞,吴浩,林真.基于形态学和区域生长法的医学图像分割[J].机电技术,2019,(03):26-29.DOI:10.19508/j.cnki.1672-4801.2019.03.008.
- [8] ZHANG Yong,张勇,YUAN Jia-zheng,等.基于图论的图像分割方法综述[C]//中国计算机用户协会网络应用分会 2016 年第二十届网络新技术与应用年会. ;中国计算机用户协会网络应用分会;:, 2016.
- [9] Boykov Y Y, Jolly M P. Interactive graph cuts for optimal boundary & region segmentation of objects in ND images[C]//Proceedings eighth IEEE international conference on computer vision. ICCV 2001. IEEE, 2001, 1: 105-112.
- [10] Grady L. Random walks for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2006, 28(11): 1768-1783.
- [11] Shi J, Malik J. Normalized cuts and image segmentation[J]. IEEE Transactions on pattern analysis and machine intelligence, 2000, 22(8): 888-905.
- [12] Rother C, Kolmogorov V, Blake A. "GrabCut" interactive foreground extraction using iterated graph cuts[J]. ACM transactions on graphics (TOG), 2004, 23(3): 309-314.
- [13] Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models[J]. International journal of computer vision, 1988, 1(4): 321-331.
- [14] Osher S, Sethian J A. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations[J]. Journal of computational physics, 1988, 79(1): 12-49.
- [15] Caselles V, Kimmel R, Sapiro G. Geodesic active contours[J]. International journal of computer vision, 1997, 22: 61-79.
- [16] Chan T F, Vese L A. Active contours without edges[J]. IEEE Transactions on image processing, 2001, 10(2): 266-277.
- [17] MacQueen J. Some methods for classification and analysis of multivariate observations[C]//Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics. University of California press, 1967, 5: 281-298.

- [18] Dunn J C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters[J]. 1973.
- [19] Bezdek J C. Pattern recognition with fuzzy objective function algorithms[M]. Springer Science & Business Media, 2013.
- [20] Fukunaga K, Hostetler L. The estimation of the gradient of a density function, with applications in pattern recognition[J]. IEEE Transactions on information theory, 1975, 21(1): 32-40.
- [21] Zhang D Q, Chen S C. A novel kernelized fuzzy c-means algorithm with application in medical image segmentation[J]. Artificial intelligence in medicine, 2004, 32(1): 37-50.
- [22] Cai W, Chen S, Zhang D. Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation[J]. Pattern recognition, 2007, 40(3): 825-838.
- [23] 邢素霞,李珂娴,方俊泽,等.深度学习下的医学图像分割综述[J].计算机工程与应用,2025,61(07):25-41.
- [24] 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.
- [25] 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.
- [26] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [27] 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.
- [28] 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.
- [29] Pereira S, Pinto A, Alves V, et al. Brain tumor segmentation using convolutional neural networks in MRI images[J]. IEEE transactions on medical imaging, 2016, 35(5): 1240-1251.
- [30] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[J]. arXiv preprint arXiv:1511.07122, 2015.
- [31] Chen L C, Papandreou G, Kokkinos I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(4): 834-848.
- [32] 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). 2018: 801-818.
- [33] Çiçek Ö, Abdulkadir A, Lienkamp S S, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]//Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International

Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19. Springer International Publishing, 2016: 424-432.

- [34] Milletari F, Navab N, Ahmadi S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]//2016 fourth international conference on 3D vision (3DV). Ieee, 2016: 565-571.
- [35] 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). 2018: 801-818.
- [36] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [37] 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.
- [38] Cao H, Wang Y, Chen J, et al. Swin-unet: Unet-like pure transformer for medical image segmentation[C]//European conference on computer vision. Cham: Springer Nature Switzerland, 2022: 205-218.
- [39] Chen J, Lu Y, Yu Q, et al. Transunet: Transformers make strong encoders for medical image segmentation[J]. arXiv preprint arXiv:2102.04306, 2021.
- [40] Wang W, Xie E, Li X, et al. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 568-578.
- [41] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [42] Xue Y, Xu T, Zhang H, et al. Segan: Adversarial network with multi-scale 1 1 loss for medical image segmentation[J]. Neuroinformatics, 2018, 16: 383-392.
- [43] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1125-1134.
- [44] Bortsova G, Dubost F, Hogeweg L, et al. Semi-supervised medical image segmentation via learning consistency under transformations[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22. Springer International Publishing, 2019: 810-818.
- [45] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2223-2232.
- [46] Cirillo M D, Abramian D, Eklund A. Vox2Vox: 3D-GAN for brain tumour segmentation[C]//Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I 6. Springer International Publishing, 2021: 274-284.