

摘要

基于通道分组机制的脑部肿瘤图像分割方法研究

脑肿瘤是如今生活中常见的脑部疾病，由脑细胞异常增殖导致，轻则恶心呕吐，严重时危及生命。随着科技的发展，人类预期寿命逐渐延长，患脑肿瘤的几率也随之增加。大脑是人体最复杂的结构之一，且脑肿瘤早期不易被发现，因此准确的找到肿瘤的位置而不影响其他结构极为重要。大量的 CT 对医生来说也是一种负担，因此借助计算机把重要的器官、肿瘤病变部分从图像中提取出来就非常重要。深度学习在医学方面的应用越来越普遍，这让半自动化的分割出目标区域成为可能。在此前提下，本文研究脑肿瘤分割算法，提出了基于 U-Net 的通道分组的脑肿瘤图像分割方法。

脑肿瘤图像有多种的序列，因此对其分割不仅需要三维的空间考量，还需要多模态的联合学习。为提高模型计算效率及更关注重点分割区域，本文方法对初始数据集进行了裁剪和数据增强。在基础网络的选择上，为了提取三维切片间的联系，使用 3D 模型分割。U-Net 在医疗图像分割领域一直是佼佼者，因其跳跃连接的存在，融合浅层和深层信息，扩充了上采样阶段的信息容量，因此选用其作为基础模型。本文的创新集中在两个方面：一是降低模型计算量：在提取信息的编码阶段加入了通道分组机制，显著降低了模型的计算量，实现轻量化改进，方便信息的传递与学习；二是提高分割精度：为了实现精确分割，本文添加了 CBAM 注意力机制，忽略掉与脑肿瘤分割无关的信息，在训练中同时关注通道和空间信息。

本文对脑肿瘤疾病和图像分割方法进行了深入的研究，提出了应用 CBAM 注意力的通道机制。使用多路复用器网络降低模型复杂度，用膨胀卷积代替普通卷积扩大感受野。添加 CBAM 注意力机制，更好的提取边界信息。本文基于公开的脑肿瘤数据集，进行了消融实验，验证了加入膨胀通道分组模块和 CBAM 注意力的可行性，并且 CBAM 模块对于膨胀通道分组模块的提高有着显著效果。在肿瘤

核心上有着 3%的提升,在增强肿瘤和肿瘤整体上也分别有着 1.4%和 1.2%的提升。同时为了公平客观的比较,横向实验比较了同一领域其余模型与本文模型的差别,在三个分割目标的平均值上也体现了本文模型的优势,平均来也看增长了 1.9 个百分点,进一步验证了本文提出的方法的有效性。本文还进行了分割效果可视化,直观的感受模型预测能力。

关键词:

通道分组, 图像分割, 注意力机制, 脑肿瘤。

Abstract

Research on Brain Tumor Image Segmentation Method Based on Channel Grouping Mechanism

Brain tumors, prevalent neurological disorders caused by abnormal proliferation of brain cells, range in severity from inducing nausea and vomiting to posing life-threatening risks. As technological advancements extend human life expectancy, the incidence of brain tumors correspondingly increases. Given the brain's complexity and the difficulty in early detection of brain tumors, precise localisation of the tumor without disrupting other structures is paramount. The abundance of CT scans can significantly burden medical practitioners, underscoring the importance of computer-assisted extraction of critical organs and tumorous regions from images. The proliferation of deep learning applications in medicine has facilitated the semi-automatic segmentation of target areas. This study explores brain tumor segmentation algorithms and proposes a brain tumor image segmentation method predicated on U-Net's channel grouping mechanism.

Brain tumor images encompass multiple sequences, necessitating three-dimensional spatial considerations and multimodal joint learning for segmentation. This study implements data trimming and augmentation on the initial dataset to enhance the model's computational efficiency and focus on crucial segmentation areas. A 3D model segmentation is employed to extract interconnections between three-dimensional slices. U-Net, a leading model in medical image segmentation, is selected as

the foundational model due to its skip connections that amalgamate shallow and deep information, thereby augmenting the information capacity of the upsampling stage. This study's innovations are twofold: firstly, it reduces the model's computational load by incorporating channel grouping mechanism in the encoding stage of information extraction, significantly diminishing the model's computational load, facilitating lightweight improvements, and enabling efficient information transmission and learning. Secondly, it enhances segmentation precision by integrating a Convolutional Block Attention Module (CBAM) attention mechanism, which disregards information irrelevant to brain tumor segmentation and concurrently focuses on channel and spatial information during training.

This research delves into brain tumor diseases and image segmentation methodologies, proposing a channel grouping mechanism with CBAM attention. A multi-channel multiplexer network decreases model complexity, and dilated convolution supersedes standard convolution to broaden the receptive field. The CBAM attention mechanism is incorporated to better extract boundary information extraction. The study utilises a publicly available brain tumor dataset to conduct ablation experiments, validating the feasibility of integrating dilated channel modules and CBAM attention. The CBAM module significantly enhances the dilated channel grouping module, with a 3% improvement in the tumor core and 1.4% and 1.2% improvements in enhanced and overall tumors, respectively. Comparative horizontal experiments demonstrate the superiority of this study's model over other models in the same field, with an average increase of 1.9 percentage points, further corroborating the efficacy of the proposed method. The study also visualises the segmentation effect, providing an intuitive understanding of the model's predictive capabilities.

Keywords:

Channel grouping, image segmentation, attention mechanism, brain tumor.

目录

第 1 章 绪论.....	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	2
1.2.1 医学图像处理.....	2
1.2.2 传统图像分割算法.....	2
1.2.3 卷积神经网络在医学图像的应用.....	3
1.2.4 脑部肿瘤图像分割算法.....	3
1.3 论文工作.....	4
1.4 论文结构.....	5
第 2 章 相关技术背景简介.....	6
2.1 核磁共振影像相关背景介绍.....	6
2.1.1 核磁共振成像原理.....	6
2.1.2 核磁共振成像序列.....	7
2.2 传统图像分割算法.....	7
2.2.1 基于阈值的分割方法.....	7
2.2.2 基于区域的分割算法.....	9
2.3 人工神经网络概述.....	12
2.3.1 神经元.....	12
2.3.2 前馈神经网络.....	12
2.3.3 反向传播 (BP) 算法.....	13
2.3.4 激活函数.....	13
2.4 基于深度学习的图像分割方法.....	15
2.4.1 深度学习介绍.....	15
2.4.2 卷积神经网络的基本组成.....	16
2.4.3 编码器-解码器框架.....	19
2.4.4 注意力机制.....	19
2.5 本章小结.....	20
第 3 章 基于通道分组机制的脑肿瘤图像分割.....	21

3.1	U-Net 整体结构.....	21
3.1.1	U-Net 编码器解码器介绍.....	22
3.1.2	把 U-Net 模型推广到三维网络.....	23
3.2	膨胀通道分组机制介绍.....	24
3.2.1	多路复用器.....	26
3.2.2	膨胀卷积.....	26
3.2.3	通道分组.....	27
3.2.4	批归一化.....	29
3.3	加入 CBAM 注意力机制.....	29
3.3.1	注意力概述.....	29
3.3.2	加入 CBAM 注意力模块.....	30
3.4	使用 CBAM 通道分组模型分割脑肿瘤图像.....	33
3.4.1	三维图像分割方法分析.....	33
3.4.2	加入 CBAM 注意力模块的通道分组模型.....	33
3.5	本章小结.....	35
第 4 章	实验过程与结果分析.....	36
4.1	实验过程介绍.....	36
4.1.1	数据集介绍.....	36
4.1.2	实验环境配置.....	36
4.1.3	数据预处理与数据增强.....	37
4.1.4	可视化工具.....	38
4.2	实验方法.....	39
4.2.1	评价标准.....	39
4.2.2	损失函数.....	40
4.2.3	模型实验细节.....	40
4.3	实验结果与分析.....	41
4.3.1	本文模型 Dice 值.....	41
4.3.2	分割效果.....	42
4.3.3	模型性能纵向对比.....	44
4.3.4	模型性能横向对比.....	44

4.4 本章小结.....	45
第5章 总结与展望.....	46
5.1 总结.....	46
5.2 未来研究方向.....	46
参考文献.....	48
作者简介及在学期间所取得的科研成果.....	55
致谢.....	56

第 1 章 绪论

1.1 研究背景及意义

随着人类预期寿命的不断延长，细胞癌变的几率逐渐增加，而癌变细胞异常的增殖就形成了肿瘤。而良性肿瘤逐渐恶化，就变成了恶性肿瘤。肿瘤增大，风险也会提高，这时就会伴随着恶心呕吐等症状，影响视力和嗅觉，严重者会导致失明甚至死亡。在日常医学检验中，皮样囊肿、胶质瘤、脑膜瘤、颅内表皮样囊肿和听神经瘤等肿瘤出现频率较高。电离辐射和脑部胚胎发育中的残留组织分化生长都可以导致脑肿瘤发病，染发和外伤也可能增加患脑瘤的风险，但并没有充分证据表明两者之间的直接联系。脑肿瘤的死亡率高达百分之三，在癌症中排进了前十。脑肿瘤大小形态均有不同，即使资深专家也存在误诊的可能^[1]。幸运的是随着医疗技术和机器学习技术的发展，患者的脑肿瘤更容易被发现和治愈。因此，更准确的脑肿瘤分割将在治疗过程中有很大帮助。

目前对于脑肿瘤的治疗方式有很多，有着放疗、化疗、药物治疗等方法，但最主要的还是手术。对于良性颅内肿瘤来说，手术切除最为有效。如果切除彻底，甚至不需要其他治疗方法辅助。因此医生重点关注的脑肿瘤部分的提取分割就尤为重要，以便于医生了解手术目标的结构。

然而脑肿瘤图像的精确分割比较困难，主要难点在于：第一，肿瘤生长周期不可预测，导致形状各异，边界不规整，有些肿瘤内部也不均匀。第二，单模态图像提供的信息不足，并且受到成像设备的限制，可能出现偏移，甚至导致边界不连续。在多模态分割中，虽然上述问题并没有消失，但充分利用多模态图像互相关联，可以较好解决信息不充分的问题。但这也意味着脑肿瘤分割模型需要适应多种模态的输入并同时提取多张图片的特征。第三，在图像采集过程中可能产生噪声，这让脑肿瘤分割任务更具挑战性。

计算机视觉一直都是计算机发展的焦点，它是一门让计算机以人类思维方式进行识别的科学。具体来说，它教会了计算机模拟人眼识别物体并跟踪目标物，进一步对图像进行处理。经过两次寒冬之后，人工智能又迎来了第三波的兴起。机器学习^[2]在人工智能中占主导地位，相较于以前的程序，机器学习更

侧重于学习能力。为了完成最终的目标，借助大量数据集，在不断的迭代中修正模型的预测能力。深度学习^[3]是机器学习的一个分支，主要用于大数据的处理。在深度学习的加持下，计算机视觉又得到飞速发展，进行端到端的学习，摆脱人工设计的误差，通过对模型结构和损失函数的设计上的优化来更新模型。

在上述的背景下，本文研究了脑肿瘤的特点，设计了一种卷积神经网络，下面围绕其进行介绍。

1.2 国内外研究现状

1.2.1 医学图像处理

医学图像是反映人体内部结构的图像，目前医学图像处理任务主要包含四个部分：检测；分割；配准和融合。这四部分结合起来进行医学图像的处理^[4]：图像检测技术确定器官的种类与方位；图像分割技术提取患病区域；图像配准和图像融合技术把人体器官患病处重建成三维模型。直观的定性定量分析，大大加快临床医疗的速度，以帮助医生做出更准确的诊断和治疗计划。

在深度学习发展之前，通过核磁共振成像等方法获得三维图像后，对其进行切片形成二维图像，然后再由医生进行诊断，这就很依赖医生的经验。尤其是分割任务，需要有相关经验的医生来判断。现如今通过医学图像处理技术，计算机自动对三维医学图像进行解析，避免了二维切片传递数据不充分，还减少了医生主观臆断的影响。这样就可以全面的了解病人身体信息，对患病部位进行快速的检测、分割并处理，提高医生的效率，并且也更为可信。这种利用深度学习半自动分割病灶的方法效率更高，尤其是编码器-解码器结构的分割网络，结果也更准确。当然最理想的还是全自动分割，然而目前技术还达不到，还需要医生做出最后的诊断。

1.2.2 传统图像分割算法

传统的图像分割方法有很多，分割原理也各不相同。下面简单介绍几种：

(1) 基于阈值的分割方法：先确定若干个阈值，经过所有像素点逐一与阈值比

较，划分出不同区域，进而实现全部像素的分类完成分割；（2）基于边缘的分割方法：通过 Sobel 算子、LoG 算子等微分算子找到不同区域的分界点，把所有分界点连接起来就完成了分割；（3）基于图论的分割方法：将整张图像视为一个无向图，像素点作为节点，像素点的关系作为边，用像素点的距离和灰度值衡量相似程度，以相似程度设置边的权重。

然而传统方法有着很大的局限性，它需要有技术的人工设置，不能广泛推广，应对突发情况的能力较差，因此很长时间图像分割领域都在原地踏步。直到深度学习的出现，人们发现 CNN 适合解决此类分割任务，比起传统方法有着很多优势，于是迎来了利用 CNN 进行图像分割的浪潮。

1.2.3 卷积神经网络在医学图像的应用

近年来发表了大量 CNN 交叉医学的论文，开拓了 CNN 与医学交叉学科的广阔前景。目前主要有三大类应用：分类^[4]、检测^[5]、分割^[7]。分类是医学图像分析的基本方法，查找单一病灶是简单的二分类任务，而把病灶标上等级以及判断多种疾病，这就涉及到多分类。不仅需要判别病症的种类，还要把患病部位画框定位时，就应用到医学图像检测。而精准的描述病症的轮廓，就需要逐像素点级别的医学图像分割。这三类应用对于医生在诊断和手术中都有着巨大的帮助。

根据标签的数量可以将深度学习分为三种不同的类型。有监督学习要求所有的数据都必须有标签，模型根据标签来修正权重。半监督学习是指只有部分数据集有标签，而大部分数据没有标签的情况下进行学习，它可以通过标记的数据来指导机器对未标记的数据进行分类或者回归。无监督学习数据集中所有数据都没有标签，算法需要自行发现数据中的模式和结构，通过聚类、降维等技术进行数据分析和处理，可以用于异常检测、数据压缩、图像和语音识别等领域。虽然医学图像有着少量且难以获得的特点，但医学图像对于精度有着较高要求，所以还是采取有监督学习。

1.2.4 脑部肿瘤图像分割算法

近年来，深度学习方法在解决各种计算机视觉问题方面显示出非常有前途的性能。许多基于深度学习的方法已应用于脑肿瘤分割，并获得了令人印象深

刻的系统性能。脑肿瘤分割算法是近年来医学影像处理领域的热门研究方向之一，旨在对医学影像中的脑肿瘤进行自动化、准确的分割，实现较高的准确度和鲁棒性。通过同时考虑多个不同的医学影像模态，如 MR、CT 等，以提高分割的准确性和可靠性。准确度的提高与复杂度的降低是优化模型的主要方向。本小节对新提出的基于深度学习的脑肿瘤分割算法进行一个介绍。

基于 U-Net 的编码解码结构，具有一个不对称的更大的编码器部分来提取图像特征。在下采样的最下层加了一个 VAE 分支，为编码器分支添加额外的指导和正则化，用来重建原始的图像^[8]。TransBTS^[9]在 3 维空间中利用 Transformer 进行脑肿瘤分割的首次尝试，在 Transformer 中进行全局特征建模。解码器利用 Transformer 嵌入的功能执行渐进式的上采样来预测准确的分割图。T-BraTS^[10]通过为疾病诊断中的图像分割引入置信水平，为脑肿瘤提供有着鲁棒性的分割结果和可靠的不确定性估计。ME-Net^[11]针对三维分割问题，提出了一种基于多编码器的脑肿瘤分割模型。该结构包含四个编码器和一个解码器。四个编码器对应于脑肿瘤图像的四个模态，进行一对一的特征提取，然后将四个模态的特征映射合并到解码器中。M-Net^[12]通过将多维卷积深入到基本模块中，实现了大量表示过程的融合，使表示过程的选择更加灵活，从而自适应地平衡稀疏切片间信息和密集切片内信息表示。结合 2D 视图的高分割精度和 3D 器官轮廓的高平滑度表示的优势，从而更准确的分割。董阳^[13]等人引入小样本分割方法，并提出基于 U-net 的原型网络 PU-net 模型用以对脑肿瘤磁共振图像进行分割。

1.3 论文工作

本文进行了预处理，把数据集转成字符串格式，以 U-Net^[14]为基础网络，调整编码器结构，加入通道分组模块减少计算量实现模型结构轻量化。并且加入 CBAM 注意力让模型更关注重点区域，训练中选择适当的 loss 函数：

(1) 脑肿瘤分割方法的研究：阅读脑肿瘤分割相关论文，了解有效的医疗方法，学习 MRI 成像原理与序列特征，调整医学图像格式便于输入网络，研究提取脑肿瘤的基本流程，设立评定分割任务效果的指标。

(2) CNN 医学图像分割：学习多种分割方法，了解它们的优缺点与适用范围，在医学分割领域 U-Net 精度较高，因此选用 U-Net 作为基础网络，使用通道分组模块轻量化进行脑肿瘤图像的分割。

(3) 预处理和调参：多模态脑肿瘤图像是 nii.gz 格式，要把它们转换成网络需要的 pk1 字符串格式。医学图像数量较少，要对其数据增强。调节学习率方便模型学习权重，根据符合行业指标的评判标准选取 loss 函数，保证权重的平滑修正，确保模型连贯的优化。

(4) 在模型中加入注意力：在本文的通道分组基础上加入注意力，让模型关注图像特定区域，提取重要信息。加入前后的网络在相同条件下进行实验和对比，说明加入 CBAM 注意力是有效的。由于 CBAM 注意力是轻量级的模块，所以将其加入到网络中只增加少量额外参数，该优化取得了较好的效果。

1.4 论文结构

第一章：绪论，分析了本文的方向，介绍了医学图像领域的发展，从传统方法到 CNN 的加入。探讨了医学图像领域中 CNN 的发展前景，总结了本文所做的改进，结尾是论文总结。

第二章：介绍医学影像和 CNN 技术，主要阐述了核磁共振影像技术的发展流程，包括成像原理，和本文数据集四种核磁共振成像序列获得方式。介绍了几种传统的分割方法，重点强调了卷积神经网络的原理和发展历程。

第三章：介绍了本文模型，从经典的 U-Net 到适用于本文任务的模型，逐步介绍了本文的模型是如何改进设计的。

第四章：基于本文提出的方法进行了实验的论证，对数据进行处理，添加模块优化模型的性能。进行了大量的实验论证，定量的阐述了模型的性能，还做了模型分割结果的可视化，直观展现本文模型适用于脑肿瘤图像分割任务。

第五章：总结本文方法，展望未来的工作，分析了脑肿瘤分割模型仍存在的局限和不足，提出了未来工作的优化方向。

第2章 相关技术背景简介

2.1 核磁共振影像相关背景介绍

核磁共振成像（MRI）^[15]技术温和，在不侵入图像原本信息的同时生成三维的解剖图像。它的初衷就是在不开刀的前提下，了解患者体内信息。MRI 可以展现充足的内部组织结构，通常用于辅助疾病检测，医生根据它提供的信息给出治疗意见。MRI 技术实现层面上较复杂，氢原子可以引发核磁共振，人体内的水分子包含氢原子，根据其位置画出人体内部组织结构图。一般成像分为三个方位，MRI 视图如图 2.1 所示：

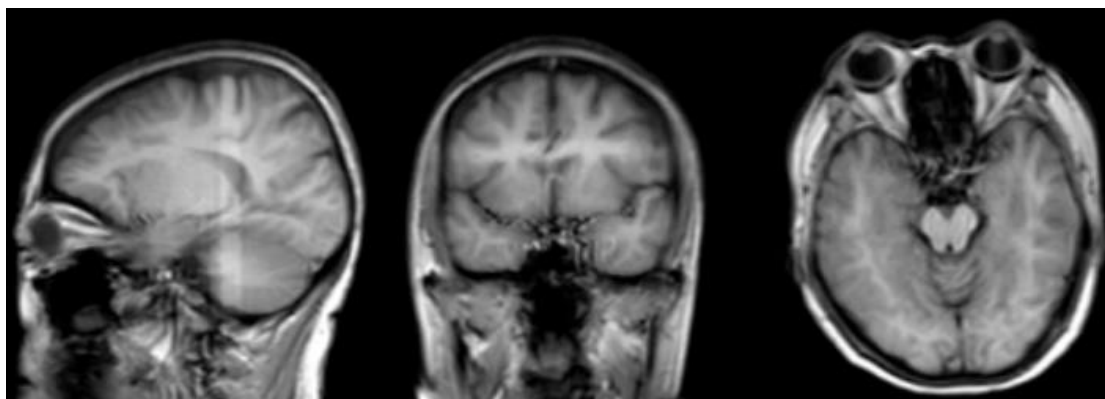


图 2.1 矢状位、冠状位、轴状位可视觉解剖图

2.1.1 核磁共振成像原理

原子核能够自旋，当它受到静磁场作用时，两极磁体取向固定不再随机，想要恢复静止态就要向外释放能量，于是开始环绕磁场运动。一个集合的原子核合成取向形成宏观磁化，对其施加一定频率的射频能量，就能发生共振。通过共振就可以从大量空间位置点收集自由感应衰减信号，然后利用傅里叶变换等方法，复合多频率信号组合成不同强度级别的振幅，并捕获对应的位置。按照他们的比例关系，依据灰度值呈现这些信息，然后把把这些点排列实现成像，这就是核磁共振成像的原理。

2.1.2 核磁共振成像序列

核磁共振成像序列：按照一定逻辑和目的采集梯度脉冲、射频脉冲、MR 信号等参数进行排序实现 MR 成像。T1 和 T2 在核磁共振成像序列中比较常见，他们之间有所区别的原因涉及到两个概念，重复时间和回波时间。重复时间就是一个射频脉冲之后直到下一周期出现一个同样射频脉冲的时间间隔，回波时间是指发出射频脉冲之后接收到回波的时间。T1 加权像重复时间大概是 500ms 左右，回波时间大概只有 20ms 左右。T2 加权像用的重复时间大概是 2500ms 左右，回波时间一般要到 500ms 左右，所以产生的影像不一样。本文使用的数据集还提供了 Flair 序列，也称为水抑制成像技术。Flair 序列就是将脑脊液水分压缩出去，而脑组织的信号却不受影响，在黑色背景下高低信号对比很容易看到脑组织的病变区域。

2.2 传统图像分割算法

图像有着颜色，纹理等方面的区别。根据这些区别，可以把一张图像分为不同的部分，这些部分互不重叠，这就是传统图像分割算法的基本原理。在深度学习出现以前，研究者们一般采取此种方法分割图像，因此称其为传统的。下面就简单介绍几种传统方法：

2.2.1 基于阈值的分割方法

阈值法是要对图像进行分区，分区的标准是根据图像特征选取的阈值，不同图像得到阈值的大小和个数也不尽相同。观察图像所有点和阈值的关系，区分出不同区域。由于该方法思路简单，计算简便，在传统方法中最为常用。

设图像像素点的灰度值为 $P(m, n)$ ，以单一阈值分割为例，阈值取 A ，就可以把图像 $P(m, n)$ 分割为两个区域。两区域的灰度值设为 x 和 y 。这里， x 和 y 都来自于原图像中的灰度值，如表达式 2.1 所示：

$$P(m, n) = \begin{cases} x, & P(m, n) < A \\ y, & P(m, n) \geq A \end{cases} \dots\dots\dots (2.1)$$

虽然该方法逻辑简单，计算量小，但需要分割的区域和相邻区域差别较大，这样才能获得较好的分割效果。而且阈值的选取直接影响预测结果的好坏，不同

阈值造成分割效果的差异较大。因此在选择的过程中就要谨慎，简单介绍几种选取办法：

(1) 人工选择法

此方法需要人工参与，需要了解相关图像的知识，由人工筛选出需要分割的目标和非目标区域。通过积累的经验，大致估计出阈值的范围，依靠一定量的实验来确定最佳的阈值。此种方法可以获得比较优异的阈值，但也存在较大的局限，依赖于人工的经验选择，判断过程中可能出现多种不确定因素的干扰。当数据量较大时，人工就远远不够，效率低下，人工的实时状态也有起伏，因此需要降低人工依赖的方法。

(2) 直方图双峰法

直方图双峰法只需要找到一个阈值，使用该方法的图像只有单一的分割目标。此方法先总结出图像的灰度值分布区域，由于分割目标和非目标特征不同，导致灰度直方图会出现两个较明显的峰值，这两个峰值之间的低谷位置就是需要的最好的分割阈值。该方法思路比较清晰，在分割目标和非目标差异较大的条件下非常好用。

如图 2.2 所示，左边是左右心室的灰度图片，右边是根据左边图片画出的灰度直方图。横轴是从 0 到 255 的灰度值，纵轴是像素点个数统计。可以看到，右图体现了左图的灰度频率统计特性，有两个非常明显的峰值在 5500 和 3000 处，它们之间的峰谷，也就是灰度值 190 左右处就是分割的最佳阈值。

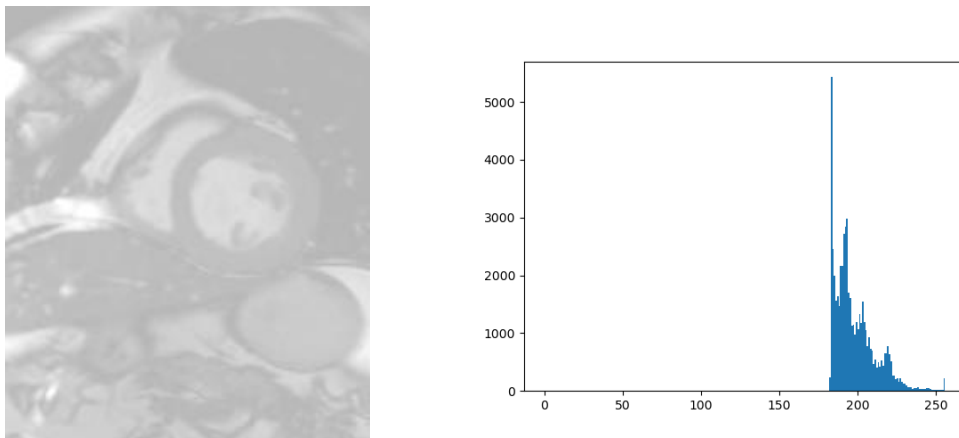


图 2.2 灰度直方图表示图例

(3) 自适应阈值法

前几种方法阈值的作用范围都是整张图片，自适应阈值法则不然，不同区域使用不同的值，减少噪声对于图像分割的影响。有很多种情况可能产生噪声，比如光线分布不均匀，影响像素点的灰度值，可能导致相邻的原本相等的像素点不再相等。而如果这些噪声使像素点灰度值改变较大，单一阈值分割的方法就会发生误判，以上几种方法就不适用了。因此阈值的自适应就很有必要，该方法是选取多个像素点作为中心，由这些像素点向外发散出多个邻域，构成多个区域。

阈值法的主要特点就是思路清晰，计算简便，然而也存在着缺点：需要图像在相对理想，即分割目标和边缘背景差异较大的条件下。因此实际工作中往往结合其他分割方法共同完成分割任务^[16]。

2.2.2 基于区域的分割算法

阈值法虽然简便明了，但对于复杂的图像精度并不高，下面介绍应对复杂图像的分割算法：基于区域的分割算法，该方法利用图像的特征，把图像划分为不同区域。简单介绍几种基于该方法的分割算法：

(1) 区域生长法

区域生长，顾名思义是由一组像素点组成的区域扩大成更大区域的方法。先选取一些种子点的集合，研究该集合的共有属性，比如灰度级，纹理颜色等信息。然后再把邻域拥有相同信息的像素点扩充进来，这种相同信息的选取便是生长的规则。当然不能进行无休止的生长，该方法第一步就要确立好生长的终止条件。

以图 2.3 为例，a 图就是要分割的原图像的灰度值矩阵。第一步在图像上选取一个点作为种子，生长规则就是每轮生长过程中，在种子的八邻域内，如果灰度值和种子相同或差值为 1，那么就将该点划分入生长区域内。下图可以看出，第一轮生长把 7 邻域内的 6 归入生长区域，第二次以生长区域内的两个新点为种子，继续生长。停止条件设置为全部的种子都无法继续生长，生长就停止。第三次生长结束后，如图 2.3 所示已经把所有满足要求的点都归入生长区域，区域生长算法停止^[17]。

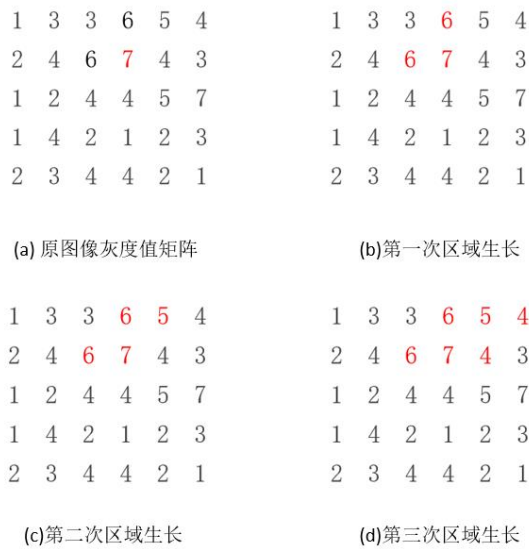


图 2.3 区域生长过程图例

(2) 区域分裂与合并

分裂合并的标准是区域分裂与合并算法的关键。区域内特征不一致时，就把该区域分裂成几个大小相等的区域，直到不能再分裂。分裂结束后开始合并，合并前提是相邻区域满足特征一致性，就把这两个相邻区域合并成一个大区域，直到不能再合并^[18]。以分裂成四份为例，四叉树可以形象的描述分裂的过程，如图 2.4 所示：左边表示图像分裂，先将一张图片分裂成 A_1 、 A_2 、 A_3 、 A_4 四个区域，然后将这四个区域都再分成四个区域， A_4 再分成 A_{41} 、 A_{42} 、 A_{43} 、 A_{44} ，右边是四叉树的形式：

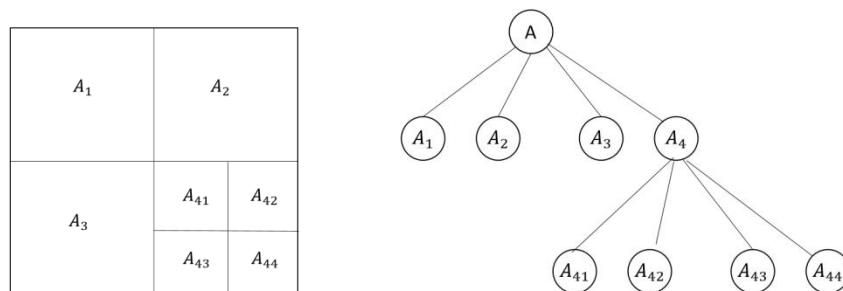


图 2.4 区域分裂示意图

分裂的依据是根据图像选定一个阈值，该阈值的计算方法如公式 2.2 所示：

D 为像素点个数, $A(m, n)$ 是灰度值, 先计算出区域的灰度均值 A :

$$A = \frac{1}{D} \sum_{m,n \in D} A(m, n) \quad \dots\dots\dots (2.2)$$

$$S^2 = \frac{1}{D} \sum_{m,n \in D} [A(m, n) - A]^2 \quad \dots\dots\dots (2.3)$$

设定阈值 S , 如果 $S^2 > S$, 区域就分裂。与此相对的合并时判断方差与阈值的关系, 如果 $S^2 < S$, 区域就合并。该方法对于阈值设定要求较高, 不合适的阈值可能会导致比较糟糕的效果, 因此该方法有着一定的局限性, 很难推广到其他领域。

(3) 分水岭分割算法

分水岭分割算法^[19]基于拓扑的形态学分割方法。该方法把图像看作地形表面, 每一个灰度等级对应一条等高线。如图 2.5 左图, 高灰度部分视为山峰, 低灰度部分视为山谷。用不同颜色的标签填充山谷, 这样做之后山谷, 也就是局部最小值就有了自己的影响区域, 这部分区域就命名为“分水岭”。如图 2.5 右图, 随着水上升, 不同颜色的水开始融合。为了避免这种情况, 在水要融合的地方及时增加屏障(图中的 Barrage)。不断进行这样的操作直到淹没所有的山峰, 根据创造的屏障得出分割结果。

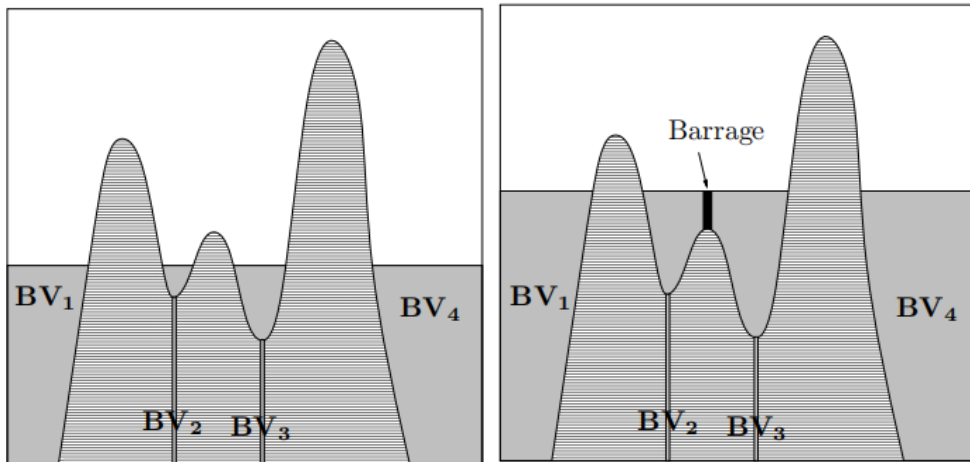


图 2.5 分水岭算法示意图

分水岭算法把极值点附近构成的屏障视为图像分割的边缘, 这样就导致分割效果很不稳定, 如果图像中有噪声之类的不规则因素干扰, 就会导致过度分割。为了解决这个问题, 又提出了基于标记的分水岭算法。该方法通过给了解的对象赋予不同的标签可以指定哪些是要合并的谷点, 哪些不是。对于不确定的对象标

签就赋值为 0，然后使用分水岭算法更新标签，对象的边界值赋为-1。更新完成之后就完成了图像分割。

2.3 人工神经网络概述

2.3.1 神经元

神经网络是对于人脑的模拟，其基本单位就是神经元，用来传递和处理信号。Frank Rosenblatt^[20]于十九世纪五十年代受到人脑启发，简化电信号刺激神经元方式，多个输入转化成一个输出。如图 2.6 所示：

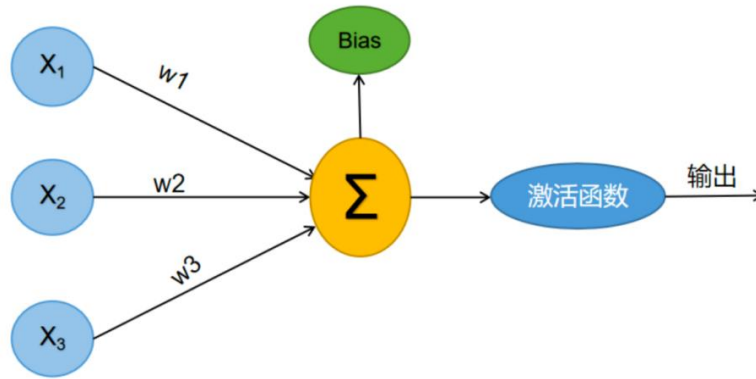


图 2.6 神经元结构示意图

其中， X_i ($i = 1, 2, \dots, n$) 表示其他相邻神经元的输入信号， w_i 表示输入权重，它表示不同输入对于输出的重要性，bias 是偏移量，输出加上偏移量后可以对结果进行更好的拟合。之后是激活函数起作用，激活函数为之前的结果附一个 0 或 1 的标签进行分类。输出值大于阈值则分到 1 类标签，小于阈值则分到 0 类标签。神经元对于固定的输入有着固定的输出值，通过监督学习逐步提升模型分类的能力，使预测值尽量接近真实值。之后神经网络逐渐发展成前馈神经网络，输出 y_i 可以表示为：

$$y_i = f(\sum_{i=1}^n w_i X_i + b) \quad \dots\dots\dots (2.4)$$

2.3.2 前馈神经网络

之所以叫前馈神经网络，是因为它是依靠后一层求导对激活函数进行调优，逐层的向前一层传播，恰恰与输入传递方向相反。如图 2.7 所示网络的简单结构：

输入层负责接收数据，中间隐藏层做数据处理，输出层输出想要的结果。同一层内神经元相互不影响，两个相邻层的神经元之间进行数据传输。信号的传递单向，整个网络有向无环。图中画了两个隐藏层，在实际实验中，隐藏层的数量可以根据实验效果设立，并不固定。

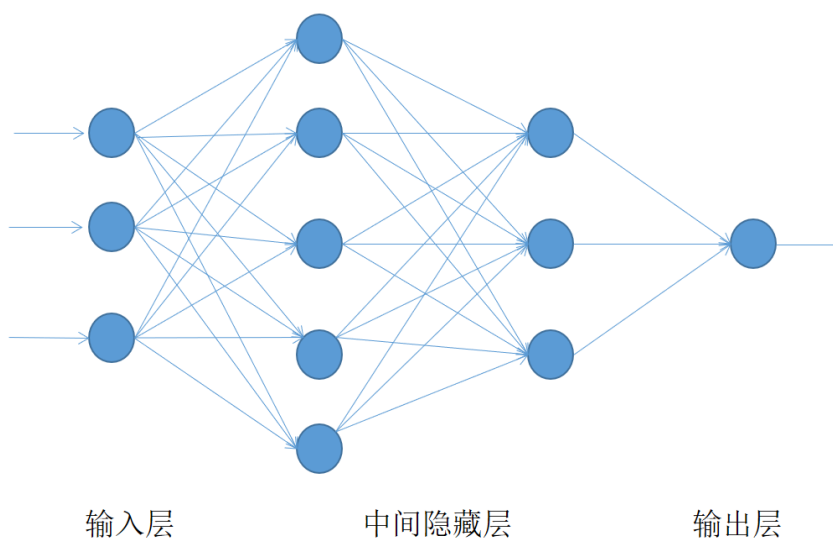


图 2.7 前馈神经网络结构示意图

2.3.3 反向传播 (BP) 算法

BP 算法是一个迭代算法^[21]，它的具体实现可分为三个步骤。第一步：数据输入到网络中，计算每一层的状态和激活值，信号正向传递。第二步：应用恰当的方法，计算每一层的误差，反向传播。第三步：修改神经元中 w 和 b 的参数值^[22]。不断迭代到整个网络学习结束，两次迭代误差足够小或者迭代次数达到所设次数，BP 算法就结束了。

2.3.4 激活函数

激活函数作用于网络的神经元，它决定了传递给下一个神经元的信号的最终形态。有了激活函数，模型就可以处理非线性问题，这使神经网络可以更好的模拟人脑，突破了模型只能解决线性问题的瓶颈。

(1) ReLU

ReLU 函数的表达式为:

$$F(a) = \text{Max}(0, a) \quad \dots\dots\dots (2.5)$$

如图 2.8 所示, ReLU 函数是线性函数, 因此它计算非常快速, 而且当输入为正数时, 梯度不会饱和, 因此可以多次使用。

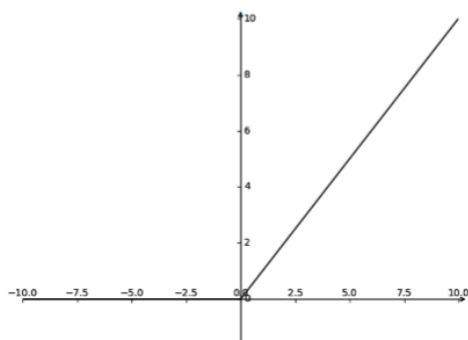


图 2.8 ReLU 函数图像

(2) Sigmoid 函数

Sigmoid 函数的表达式为:

$$F(a) = \frac{1}{1+e^{-a}} \quad \dots\dots\dots (2.6)$$

如图 2.9 所示, Sigmoid 函数适合需要输出归一化信号的神经元, 比如需要预测概率的模型就很实用。但存在着梯度不稳定的缺点, 可能在隐藏层中爆炸或消失, 而且相比 ReLU 函数, 指数型计算也比较慢。

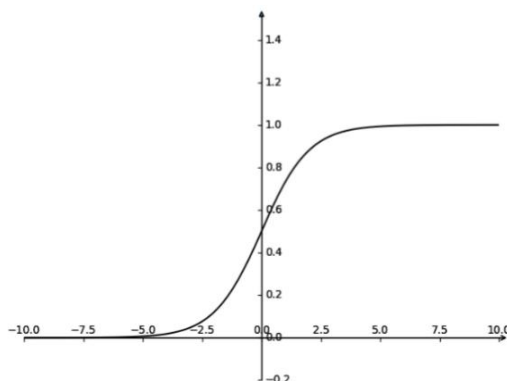


图 2.9 Sigmoid 函数图像

(3) tanh 函数

tanh 函数的表达式为:

$$F(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}} \dots\dots\dots (2.7)$$

函数图像如图 2.10 所示:

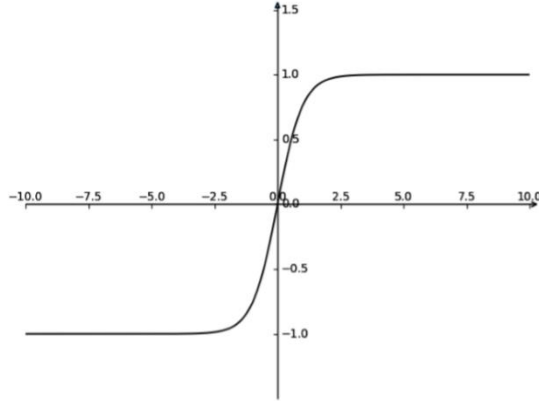


图 2.10 tanh 函数图像

tanh 函数形态类似于 Sigmoid 函数, 但相比有一些优势, 负输入可以负输出, 但输入较大或较小时输出差别不大, 不利于更新权重, 在实际实验中也很少使用。

2.4 基于深度学习的图像分割方法

2.4.1 深度学习介绍

深度学习是目前机器学习热门的研究分支之一, 比起机器学习的其他方向, 它更符合人工智能的定义。深度学习的目标是让计算机类似于人脑进行思考学习, 辅助人们驾驶汽车, 帮助医生找到病人患处所在等等, AlphaGo 甚至击败了八冠王柯洁。深度学习使用多重非线性结构模拟人脑, 学习有效的特征表示。起初人们为了获得更好的效果, 进行大量重复的训练, 更好的提取数据样本中的复杂特征。深度学习理论的横空出世, 让相关科研人员不再仅仅关注数据特征, 而是转向如何构造更深层次的网络结构, 促进了残差网络等优秀网络的诞生, 主动学习出更有效, 更高维的特征。

经过长久的发展, 根据网络结构和使用的层类型, 神经网络可以被划分为不同的类型, 包括 CNN^[23]、DNN^[24]和 RNN^[25]。目前普遍使用 CNN 解决计算机视觉任

务，卷积层提取图像特征，池化层降低图像规模，提升学习效率。而 DNN 则是由全连接层组成，每一层的所有节点都与其前一层和后一层的所有节点连接，可以更好的分离数据样本的特征。RNN 较为特殊，多用于提取时间序列的相关信息，对比卷积神经网络的卷积核，循环神经网络采用循环核。循环核具有记忆力，通过不同时刻权重共享递归学习时间序列的特征，广泛用于生成时间序列相关的文本。

2.4.2 卷积神经网络的基本组成

卷积神经网络广泛应用于计算机视觉的任务，实现便捷，不需要人工干预。休伯尔和维厄瑟尔对于猫视觉皮层的研究实验中，发现了视功能主结构。受此启发，福岛邦彦^[26]开创了卷积神经网络；Waibel^[27]加入权值，实现了不同语音的辨认；LeCun^[28]提出的 LeNet5，成功的识别了手写数字，标志着卷积神经网络理论的正式形成。之后研究者们就在网络改进上钻研，Hinton 团队提出了 AlexNet^[29]，首次采用 Dropout 策略，在 ImageNet 图像分类任务上一举夺冠；VGGNet^[30]用于大规模深层网络识别，拓宽了网络深度；GoogLeNet^[31]在其基础上降低了参数；何凯明 ResNet^[32] 获得了 2016 年 CVPR 最佳论文；DenseNet^[33]一定程度上避免了梯度消失，加强特征传递；涌现了越来越多的优秀网络。

卷积神经网络按功能可以分为数据入口的输入层，特征提取的卷积层、数据处理的正则化层和池化层、空间映射的全连接层，下面介绍一下每一层的相关操作和功能：

(1) 输入层

输入层是数据的入口，输入层对输入数据进行预处理，把数据转化为适合网络操作的形式。对数据进行洗刷：填补数据的遗漏值，删除不一致的数据，去除数据的噪声等影响。或者进行数据集成：现在数据源越来越多，就需要将多个数据源进行合并，保持规范化去重。还有降维处理：摒弃掉信息较少的维度，减小模型学习压力，加快模型学习效率。

(2) 卷积层

卷积层的作用是特征提取，提取图片中的信息，从左到右由上至下，卷积核在输入数据上按步长大小移动，图像特征逐渐抽象数据化，如图 2.11 就是卷积操作的示意图：

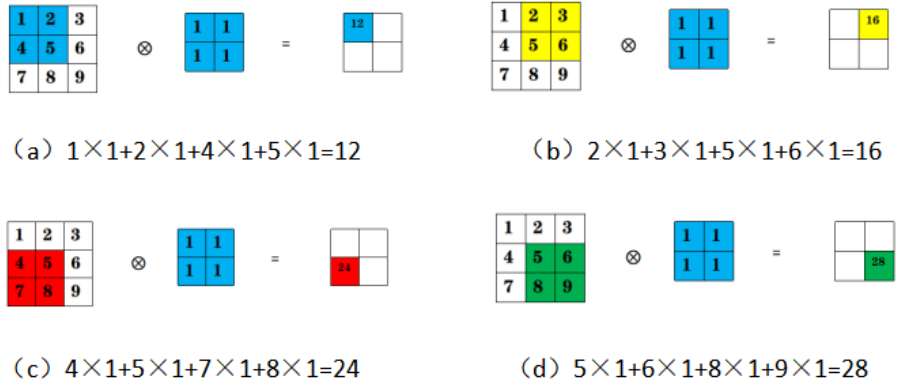


图 2.11 卷积运算过程示意图

卷积操作的具体实现是通过矩阵乘法，按步长在输入矩阵上进行滑动。在该示意图上，矩阵大小为 3×3 ，卷积核大小为 2×2 ，步长为 1。每次卷积核完成计算后就向下移动一个步长的距离，以此类推在运算全部完成后就得到卷积运算的结果。卷积核的参数并不是固定的，可以通过模型训练进行学习。随着梯度下降与反向传播，参数得到修改，预测效果也随之提高。

(3) 池化层

池化层的目的是对特征进行降采样，位于卷积层之间，将特征下采样，提取不受位置干扰的信息，减少数据与运算参数。池化方法有很多，下面简单介绍几种池化方法：

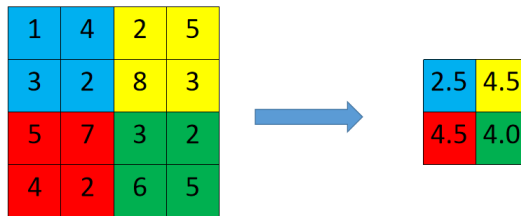


图 2.12 平均池化示意图

平均池化：将分块区域中矩阵所有值的平均值作为结果，使数据均匀的下采样，公平的提取数据原始信息。

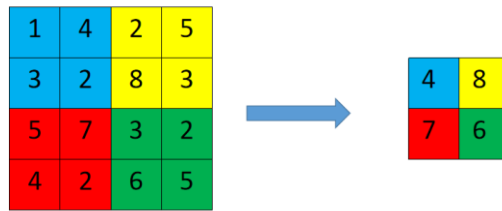


图 2.13 最大池化示意图

最大池化：选取区域内数据的最大值，该方法认为像素点的值越高所包含的信息就越丰富，更好的获得纹理特征，保留重要信息，在训练中广泛应用。

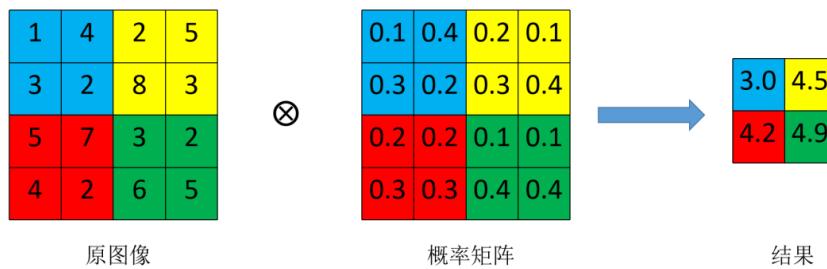


图 2.14 随机池化示意图

随机池化：为区域内每个像素点都赋一个概率值，区域内概率和为 1，概率大的像素点被选中的概率也大。这种方法选值随机，可能提取到无用信息，因此这种方法并不稳定，可能产生较坏的结果，所以使用的很少。

(4) 反卷积

反卷积又叫做转置卷积，相当于卷积操作的反操作，通过反卷积可以把输入矩阵的一个值映射到输出矩阵的多个值。这样图像的输出规模就会大于它的输入规模，可以对图像进行上采样。为了实现这个过程，需要对图像外围先用 0 做填充，增加原图像的大小，然后再进行卷积运算，最后可以得到更大的图像。如图 2.15 所示，由 2×2 上升到 3×3 ：

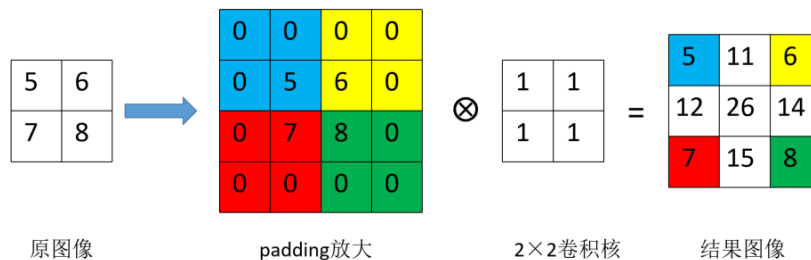


图 2.15 反卷积运算示意图

2.4.3 编码器-解码器框架

编码器-解码器结构^[34]应用广泛，起初它多用于自然语言处理中的任务，而这种任务序列长度通常都不固定。为了处理这种可变的结构，需要两个组件。第一个组件是编码器，把输入转成一个状态；第二个组件是解码器，把这种状态转换成输出序列。

对于图像来说，编码器把图像进行编码，提取特征，转化为序列形式，然后解码器把序列进行解码还原成图像形式，还原成输入的大小，还有编码器与解码器之间的跳跃连接传递信息。

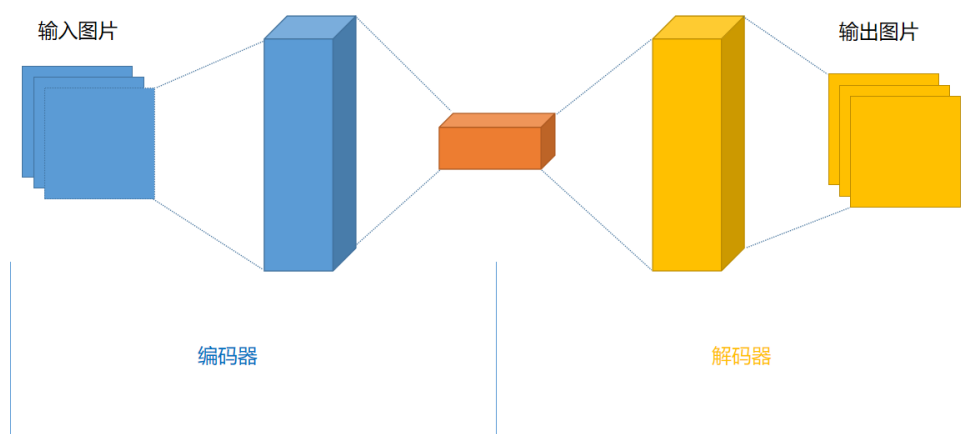


图 2.16 编码器-解码器结构示意图

2.4.4 注意力机制

注意力机制^[35]是一种常用于自然语言处理和计算机视觉等任务的技术。它的目的是帮助模型在处理输入数据时，专注于处理当前任务最重要的部分。注意力机制通过计算每个输入位置的权重，并将其特征按权重加权，从而得到当前任务的输入表示。

在机器翻译中，注意力机制可以专注于输入中与当前翻译目标最相近的词，从而生成更准确的翻译结果。在 NLP 任务中，注意力机制常常作为深度学习模型如双向循环神经网络、transformer^[36]模型的一部分来使用。同时，它也可以作为一个独立的模块，添加到其他基础模型中来增强模型的性能。注意力也可用于图像处理，可以帮助模型专注于当前任务中最重要的特征，提高模型

的准确率。例如，在目标检测任务中，注意力机制可以辅助模型集中于目标位置，减少误判。

除了 NLP 和 CV 任务之外，注意力还可以应用于生物信息学、图形学等其他方向。注意力机制两种最基本的方法是前馈注意力机制和递归注意力机制，前馈注意力机制通过在前向传播过程中计算注意力权重，而递归注意力机制则基于递归神经网络来实现。

总之，注意力机制是一种非常有效的技术，已经在许多领域得到了广泛的应用。它可以帮助模型从海量信息中抽取关键特征，从而提高模型的准确性和效率，是深度学习模型取得进展的重要因素之一。

2.5 本章小结

本章讨论了核磁共振成像原理及磁共振成像序列的获取方法，同时介绍了一些传统的图像分割方法。重点介绍了 CNN 的发展历程、基本框架、组件的作用，详细说明了儿种激活函数和池化类型。简单介绍了编码器-解码器框架和注意力机制。

第 3 章 基于通道分组机制的脑肿瘤图像分割

目前医学图像分割存在着两个主要的难点，一是三维图像带来的模型复杂度问题：现有的医学图像分割模型多采用深度卷积神经网络，如 U-Net、SegNet 和 Mask R-CNN 等。这些模型通常具有大量的卷积层，每个卷积层包含多个滤波器。同时，在模型设计时还通常会采用各种技术，如跳跃连接、池化和上采样等，以提高模型的精度和鲁棒性。这种复杂的架构通常需要大量的计算资源和时间来训练和优化，也需要大量的数据来提高精度。二是分割的精度需要：医学图像数据数量往往有限且分辨率较低，加之病理部位常被肿瘤等异常所侵蚀，病灶、器官的轮廓并不明显，使得图像的质量更低。难点是如何准确划分边界。

本文对于问题一提出了膨胀通道分组模块，在编码器阶段使用多路复用器和通道分组模块降低模型复杂度，用膨胀卷积代替普通卷积扩大感受野。降低了接近 4 倍的模型复杂度。对于问题二添加 CBAM 注意力机制，更好的提取边界信息。并且 CBAM 模块对于通道分组模块的提高有着显著效果，有效的提高了分割精度。在肿瘤核心上有着 3% 的提升，在增强肿瘤和肿瘤整体上也分别有着 1.4% 和 1.2% 的提升。

3.1 U-Net 整体结构

U-Net 在图像分割中广泛使用，尤其在医疗图像领域，采用 U-Net 作为基础模型分割十分契合，它也影响了很多后来人的工作。很多优秀的模型都是它的变种，针对各种不同的需求对其进行修改。虽然 U-Net 对于 3D 医疗图像的相邻切片之间的关联性提取的很好，但随着层数的增加，基于 3D 的 CNN 会带来极为庞大的计算量与开销，对实验设备要求极为严苛。因此本文构造了一个轻量级的网络模块，利用通道分组模块，大大的减少了计算量，降低了成本。

由于跳跃连接的存在，U-Net 可以很好的整合浅层和深层的信息。如图 3.1 所示，该网络整个呈对称状，形似 U 型，因此命名为 U-Net。编码器对应左半部分，主要负责提取图片中的特征，解码器对应右半部份，主要负责特征整合与补充。组合在一起构成分割网络中的编码器-解码器结构。

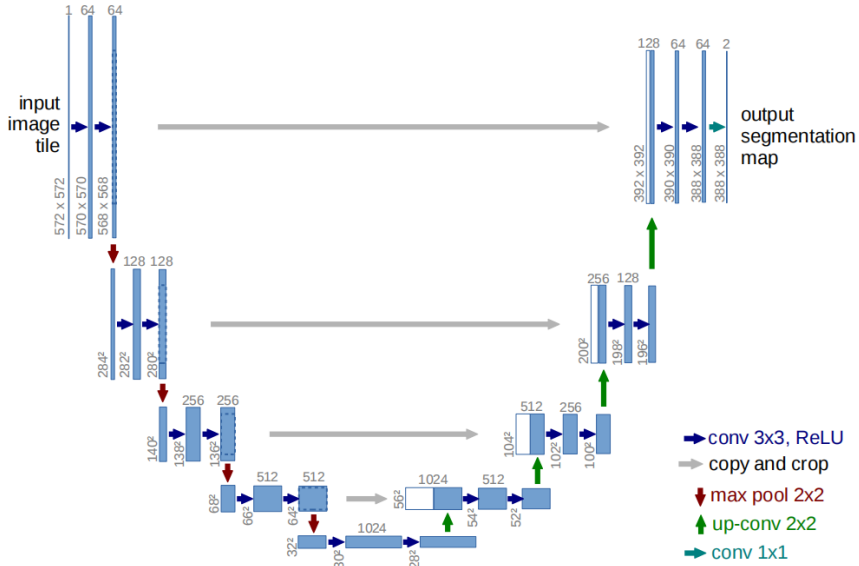


图 3.1 U-Net 网络框架

3.1.1 U-Net 编码器解码器介绍

U-Net 的网络结构由编码器部分和解码器部分组成。编码器部分负责提取特征和降采样，如图 3.2 所示：编码器每一层一般都由两个卷积层构成，每个卷积层包括 3×3 的卷积，BN 层和 ReLU 激活函数。经过卷积提取特征后的图像会流向两个方向：一方面是进入下一层编码器；另一方面是通过跳跃连接进入同层解码器和解码器的输入图片进行 concat 拼接。

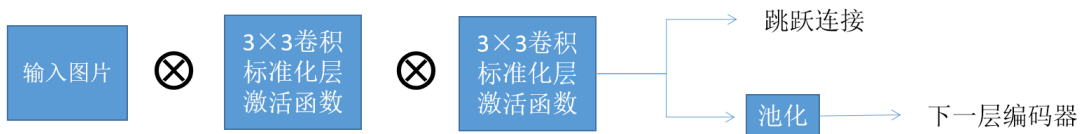


图 3.2 编码器结构示意图

如图 3.3 所示，解码器图片的来源有所差别。解码器的输入为编码器跳跃连接而来的输入图片和上一层解码器的输入图片 concat 连接组成，这种连接需要两张图片相同大小，因此上一层解码器的输入就要上采样操作。有了跳跃连接，就可以兼顾编码器层和解码器层的图像，丰富了上下文信息。因此 U-Net 取代了 FCN 网络，在医学图像分割领域广泛流行。



图 3.3 解码器结构示意图

3.1.2 把 U-Net 模型推广到三维网络

最初的 U-Net 用来解决 2D 问题，而本文的数据集是 3D 的，因此需要把网络推广到 3D。如表 3.1 所示，把二维卷积都替换为三维卷积，卷积层之后都有标准化层和激活函数，再通过最大池化减少计算量。

表 3.1 编码器 结构

	输入图像规模	操作
输入图片	$128^3, 4$	
编码器 1	$128^3, 4$	$3 \times 3 \times 3$ 卷积 $3 \times 3 \times 3$ 卷积
降采样	$128^3, 32$	$2 \times 2 \times 2$ 最大池化
编码器 2	$64^3, 32$	$3 \times 3 \times 3$ 卷积 $3 \times 3 \times 3$ 卷积
降采样	$64^3, 64$	$2 \times 2 \times 2$ 最大池化
编码器 3	$32^3, 64$	$3 \times 3 \times 3$ 卷积 $3 \times 3 \times 3$ 卷积
降采样	$32^3, 128$	$2 \times 2 \times 2$ 最大池化
编码器 4	$16^3, 128$	$3 \times 3 \times 3$ 卷积 $3 \times 3 \times 3$ 卷积
降采样	$8^3, 256$	$2 \times 2 \times 2$ 最大池化
U-Net 底层	$16^3, 256$	$3 \times 3 \times 3$ 卷积 $3 \times 3 \times 3$ 卷积

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/967044063030006046>