

清 华 大 学

# 综 合 论 文 训 练

题目：基于深度学习的靶区自动勾画

系 别：电子工程系

专 业：电子信息科学与技术

姓 名：朱海东

指导教师：吴及教授

2019年6月19日

## 关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：学校有权保留学位论文的复印件，允许该论文被查阅和借阅；学校可以公布该论文的全部或部分内 容，可以采用影印、缩印或其他复制手段保存该论文。

(涉密的学位论文在解密后应遵守此规定)

签名：朱海东 导师签名：黄从 日期：2019.6.13

## 中文摘要

近年来随着深度学习算法的迅速发展，计算机相关的技术在生活图片和医学图片等多项任务上均取得了长足的进步，在图像分类、图像分割、视频分析、物体检测、人脸识别等多个领域的任务中均有突破性的性能提升。在生物医学领域，一些需要广泛的知识储备的任务中，深度学习算法同样能够取得与人类水平相当，甚至最终效果超过人类水平的全自动化的专业水平。

在生物医学图像领域，放疗技术是治疗癌症的重要方法之一。放疗效果的好坏取决于相应靶区，尤其是 CTV 区域勾画位置的准确性。靶区勾画任务是对于医学图像中需要进行放射治疗的部位（包括病灶、潜在淋巴结等）进行勾画的操作。由于器官大小、潜在淋巴结转移区域的确定等使得靶区勾画结果严重依赖于医生的经验，通过整合医生的专业知识搭建起基于 CTV 的靶区勾画自动化系统，是具有重要意义的。

靶区勾画任务可以视作一种深度学习在语义分割算法中的应用。但是，因为其特殊的标注方式和不同医生的标注方法和习惯，靶区勾画任务相比较传统的图像分割任务有许多额外的困难，导致了标注信息存在有标注标准不统一、勾画边界不清晰等问题。这些问题在生物医学图像这个需要大量的知识储备的问题中十分常见，而这些问题与深度学习训练过程中需要清晰、统一、明确边界的标注存在矛盾。为了解决这一问题，本文首先训练得到一个可用于靶区勾画任务的深度学习模型，然后针对生物医学图像中标注标准不统一的问题设计了新的解决方法，最终将模型整合成为系统，提供给医生使用。

本文的主要贡献有：

- (1) 实现一个基于深度学习的 CTV 靶区勾画模型，实现靶区勾画任务自动化
- (2) 设计实现算法，解决深度学习算法在训练过程中训练集的标注标准不统一、标注中存在有噪声的难题。
- (3) 搭建可使用的靶区勾画系统

**关键词：**深度学习；靶区勾画；噪声标注；医学图像；语义分割

## ABSTRACT

With the development of deep learning in recent years, computers have made significant progress in many vision tasks. There are breakthrough improvements in missions such as image classification, image segmentation, video analysis, object detection, face recognition, etc. In the field of the biomedical image, deep learning algorithms can also achieve a fully automated professional level, which is comparable to or even better than human annotators in tasks that require a broad knowledge base.

In the field of biomedical imaging, radiotherapy is one of the essential ways to treat cancer. The effectiveness of radiotherapy depends on the accuracy of the corresponding target area, especially the CTV area. The target delineation task is to delineate the parts that need to be treated for radiation, such as lesions, latent lymph nodes, etc. Due to the size of the organs and the potential lymph node metastasis, the target delineation results are heavily dependent on the doctor's experience. It is of great significance to build the CTV-based target delineation automation system by integrating the doctor's expertise.

The target delineation task can be regarded as the application of deep learning in semantic segmentation algorithms but is more difficult due to its unique labeling method and different doctors' habits. Ambiguous boundaries and labels with different criterions, which are common in biomedical images segmentation tasks, contradict with the prerequisite of the deep learning models: accurate annotation with the same standards. To solve the problem, this paper first trains a deep learning model that can be used for the CTV target delineation task, then designs a new solution for training on the noisy-labeled images. Finally, we integrate the model into an system.

The main contributions of this article are:

- (1) Realize a CTV target learning delineation model based on deep learning
- (2) Design an algorithm for training deep learning models on noisy-labeled datasets
- (3) Build a usable target area delineation system

**Keywords:** Deep Learning; Target Area Delineation; Noisy Label; Medical Image; Semantic Segmentation

# 目录

<b>第 1 章 引言</b> .....	1
1.1 问题提出.....	1
1.2 研究现状.....	2
1.2.1 深度学习发展概述.....	2
1.2.2 语义分割算法发展概述.....	4
1.2.3 含噪声标注样本的训练方法概述.....	6
1.3 文章结构.....	7
<b>第 2 章 基于深度学习的靶区勾画算法研究</b> .....	8
2.1 数据转换与预处理.....	8
2.2 靶区勾画模型设计.....	9
2.2.1 基础网络.....	9
2.2.2 基础网络改进与提升.....	10
2.3 结果与性能分析.....	12
<b>第 3 章 噪声标注下图像分割模型训练</b> .....	14
3.1 问题描述.....	14
3.2 方法介绍.....	15
3.2.1 分割模块.....	15
3.2.2 质量评估模块.....	16
3.2.3 防过拟合模块.....	16
3.3 实验结果.....	17
3.3.1 和基础方法的对比.....	18
3.3.2 防过拟合模块评估.....	19
3.3.3 质量评估模块评估.....	20
<b>第 4 章 系统设计</b> .....	22
4.1 需求分析.....	22
4.2 系统用例图.....	22
4.3 系统动态建模.....	23

4.4 前端模块功能需求.....	25
4.4.1 上传功能模块.....	25
4.4.2 可视化结果显示模块.....	25
4.4.3 勾画结果下载模块.....	25
4.5 后端模块功能需求.....	26
4.5.1 预处理与数据转换模块.....	26
4.5.2 勾画结果生成模块.....	26
4.5.3 勾可视化结果生成模块实现.....	26
4.5.4 数据上传和下载管理模块.....	27
4.6 前端实现.....	27
4.6.1 上传功能模块实现.....	27
4.6.2 结果显示与对比模块实现.....	28
4.6.3 勾画结果下载模块实现.....	28
4.7 后端实现.....	29
4.7.1 勾画结果生成模块实现.....	29
4.7.2 可视化结果生成模块实现.....	29
4.7.3 数据上传和下载管理模块实现.....	30
<b>第5章 结论</b> .....	<b>31</b>
5.1 工作总结.....	31
5.2 系统架构及方法描述.....	31
5.3 问题与改进.....	32
<b>插图索引</b> .....	<b>33</b>
<b>表格索引</b> .....	<b>34</b>
<b>公式索引</b> .....	<b>35</b>
<b>参考文献</b> .....	<b>36</b>
<b>致谢</b> .....	<b>39</b>
<b>声明</b> .....	<b>40</b>
<b>附录 A 书面翻译</b> .....	<b>41</b>
A.1 U-Net: 为生物医学图像分割设计的卷积网络 .....	41

A. 1. 1 介绍.....	41
A. 1. 2 网络结构.....	44
A. 1. 3 训练 .....	44
A. 1. 4 实验.....	46
A. 1. 5 结论.....	47
A. 2 3D U-Net: 从稀疏标注中学习密集的体分割 .....	49
A. 2. 1 引言.....	49
A. 2. 2 网络结构.....	51
A. 2. 3 实施细节 .....	52
A. 2. 4 实验.....	53
A. 2. 5 结论.....	56
在学期间参与课题的研究成果 .....	58

## 主要符号表

GTV	肿瘤区
CTV	临床靶区
PTV	计划靶区
QAM	质量评估模块(Quality Awareness Module)
OCM	防过拟合模块(Overfitting Control Module)
ILSVRC	Imagenet 分类任务比赛
SE	压缩与注意力模块(Squeeze-and-Excitation)
IoU	面积/体积交并比(Intersection over Union)



# 第 1 章 引言

## 1.1 问题提出

近年来癌症已经成为全球人类的第二大死因。世界卫生组织在 2018 年曾经报道：“世界范围内，癌症是发病和死亡的主要原因，2012 年约有 1400 万新发癌症病例，在 2015 年造成了 880 万人死亡，预计今后二十年新发病例数将增加约 70%。”对于癌症的预防和治疗工作也已经成为了各国科学家和医生的关注重点。癌症主要的治疗手段一共由四种：化疗、放疗、手术和分子靶向药物。作为放疗方法的成员之一，调强放射<sup>[1]</sup>，即调强适形放射治疗，是对于体内部位进行三维放疗的一种。它需要对于特定的部位进行特定的剂量强度的调节实现放疗效果。作为一种常见的癌症治疗方法，对于肿瘤区域进行靶区勾画是放疗计划的研究基础。

靶区勾画，是放疗医生在临床放射治疗前进行癌变区域的部位进行勾画的操作，医生依据不同部位局部复发和区域淋巴结转移几率定义勾画区域。医学上的靶区通常存在三种不同的定义：肿瘤区（GTV）、临床靶区（CTV）和计划靶区（PTV），三者的位置关系如图 1.1 所示。肿瘤区是临床病灶部位，包括临床上能够诊断出的具有一定大小和形状的恶性病变部位，该部分区域的细胞已经或可能病变和发生淋巴结转移；临床靶区是在肿瘤区的基础上，增加了周围的部分可能转移的健康的细胞部分；计划靶区则是在考虑到器官的运动、成影仪器的位置偏差以及日常治疗的效果，在 CTV 的范围的基础上，对于待放射的部位进行了适当的扩充，从而保证手术中 CTV 部位得到足够剂量的治疗。靶区勾画任务本身需要极高的专业知识储备，这一高层次的知识储备需求与医院中的大需求间存在明显的冲突。随着自动化技术的推广，对于靶区勾画的自动化技术研究也在逐步得到研究和发展。自动化的靶区勾画技术一方面能够极大程度减少医生的工作量，缓解医院中对于靶区勾画的高需求；另一方面也可以将靶区信息提升到高维度进行更高层次的分析。依据靶区自动勾画任务的本身性质，该任务也被拓展为一个计算机图像语义分割任务。

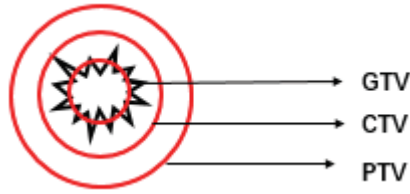


图 1.1 靶区示意图。黑色部位为肿瘤部位，红色圆圈为三个靶区范围

2012 年以来，深度学习算法在计算机视觉的多个任务上已经具有了显著的效果提升，并且在生物医学图像领域的检测、分类、分割等任务上也有了显著的性能提升。但是在靶区勾画任务上，基于深度学习方法实现靶区勾画的任务尚处于初级阶段。与其他勾画任务显著不同的是，靶区勾画任务通常不仅仅需要包括较为明显的病变区域，同样也要向周围健康细胞部分进行适当的拓展，以实现更全更广对于病变区域的照射，因此临床上 CTV 部分的勾画有很重要的意义。准确的 CTV 勾画对于放疗结果的准确性来说非常重要，决定着疾病的复发率。这一部分不但需要额外的经验和知识，也和不同的医生标注习惯有很大的关系。并且因为这一部分也属于健康细胞范围，造成靶区勾画最终的边界不清晰，给模型带来了额外的困难。靶区勾画中所可能遇到的标注噪声、标注标准不统一、医生标注习惯不同等多个问题都会对整个的深度学习方法有很大的影响，使得最终训练得到的分割模型性能降低。之前有学者曾经提出过一些分类问题中含有噪声标注样本的解决方法，但是这些方法并不能对于分割问题进行扩展和使用<sup>[17-23]</sup>。对于靶区勾画任务，本文将首先设计并实现一个深度学习网络模型，并且将会基于这个模型设计得到一个可用的基于 CTV 部位的靶区勾画系统；其次，本文将进一步研究标注不良问题，通过对于标注的噪声水平高低的衡量挑选出质量较高的标注样本进行训练，最终在含噪声标注的情况下训练得到一个稳定的模型。

本次毕业设计的贡献主要有 3 点：1) 设计得到对于宫颈癌部分进行自动勾画的靶区勾画算法；2) 提出质量评估模块 (QAM) 和防过拟合模块 (OCM)，使得在含噪声标注的样本中进行分割训练得到的模型能够同时保持有良好的性能和泛化能力；3) 基于靶区勾画算法设计实现系统。

## 1.2 研究现状

### 1.2.1 深度学习发展概述

2012 年以来，深度学习算法在计算机硬件的支持下得到了迅速的发展。深度学习中的神经网络的梯度回传概念由 Lecun<sup>[2]</sup>提出，并在对于图片的分类问题上取得了极佳的效果。LeCun 提出的 LeNet 通过卷积层和全连接层结合的方式实现对于手写数字数据集 MNIST 的分类，并且通过梯度回传的方式对于网络中的参数进行优化。因为梯度回传过程时间开销极大，当时硬件技术无法给予足够的支持，因此在很长的时间里深度学习并没有得到很有效的发展。在 2012 年的 ILSVRC<sup>[3]</sup>分类比赛上，AlexNet<sup>[4]</sup>将测试集上的计算机自动分类 top5 错误率从传统算法 SIFT+FVs<sup>[5]</sup>的 26.2%降低到了 15.3%，比先前的算法有了极大程度的提升。伴随显卡性能的提升和相应的卷积算法的优化改进，2012 年后深度学习算法也得到了快速的发展。2014 年 Simonyan 等人通过对于多重的卷积层层数和卷积核大小进行不同的组合尝试后得出了一个只包含卷积、下采样和全连接层的网络 VGG<sup>[6]</sup>，这一网络也作为之后的许多的深度学习模型的基础网络。随着硬件的计算能力和算法的提升，网络层数也慢慢加深。但是科研工作者发现随着网络的加深，网络自动的分类准确率的下降。同时梯度随着网络的层数增加会产生梯度消失或者梯度爆炸的问题。为了解决这些问题，Kaiming He 等人提出了 ResNet 网络结构<sup>[7]</sup>，其网络结构如图 1.2 所示。这一结构通过旁路的方式，将上一层的结果直接和新产生的结果进行相加的方式实现梯度的短路连接，使得最终的准确率得到了提升，并且解决了梯度消失和梯度爆炸的问题。2016 年 Gao Huang 等人提出的 DenseNet 网络结构<sup>[8]</sup>相比较于 ResNet 的效果又有了一定的提升。这些大型的网络结构给所有的计算机视觉自动化任务提供了一个网络结构基础和一个的网络骨干。在网络的性能逐渐稳定性之后，深度学习算法在图像分类、人脸识别、物体检测、图像分割等多个领域中取得了极大的进步，效果比传统算法在计算机视觉中的应用有了很大程度上的提升。

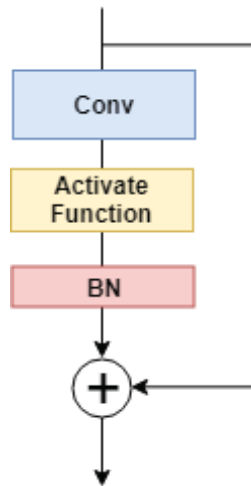


图 1.2 ResNet 中的残差模块，Conv 为卷积层结构，Activate Function 为激活函数，BN 为批正则化层

### 1.2.2 语义分割算法发展概述

语义分割任务在日常生活的图像和医学图像中均有大量的应用。在深度学习方法出现前，传统的语义分割方法，例如 Normalized cut<sup>[9]</sup>、Histogram thresholding<sup>[10]</sup>等，都是通过临近像素的关系判断相邻的像素之间的联系进行图片分割。在深度学习算法提出之后，基于 VGG 等网络骨架上产生的 FCN<sup>[11]</sup>、DeepLab<sup>[12]</sup>、DeconvNet<sup>[13]</sup>和 SegNet<sup>[14]</sup>在图像的语义分割任务中的分割效果逐步提升，并且这一类深度学习的方法在医学图像分析领域上也开始发展和推广。相比较于基础的 VGG 结构，FCN<sup>[11]</sup>将 VGG 最后的全连接层删除，并且使用卷积层进行替代，且使得输出的通道数为类别数的方法，对于图像中的每一个像素点进行分类，从而实现最终的分割。在 FCN 的基础上，DeepLab<sup>[12]</sup>通过在最后的热度图上实现了条件随机场的方法对于生成的结果进行修正，并且提出了空洞卷积的概念，如图 1.3 所示，使得网络的输出的感受野得以增加，让分割网络最终能够产生分辨率相同的分割结果图像。随后出现的 DeconvNet<sup>[13]</sup>则是将最初 FCN 的单层卷积层改成了多层的解卷积层操作，这样的好处是在相对而言更少的参数情况下，实现对于图像更为细致化的结果分割，从而仅仅使用卷积操作即可实现原分辨率的分割结果生成，并且网络中间部分的感受野相比而言更大。SegNet<sup>[14]</sup>则在 DeconvNet 的基础上实现了对于下采样和上采样的脚标的传递，通过对于相同位置的上采样操作，

网络能够将下采样的信息传输给上采样层，从而实现解码器和编码器之间的信息传递。

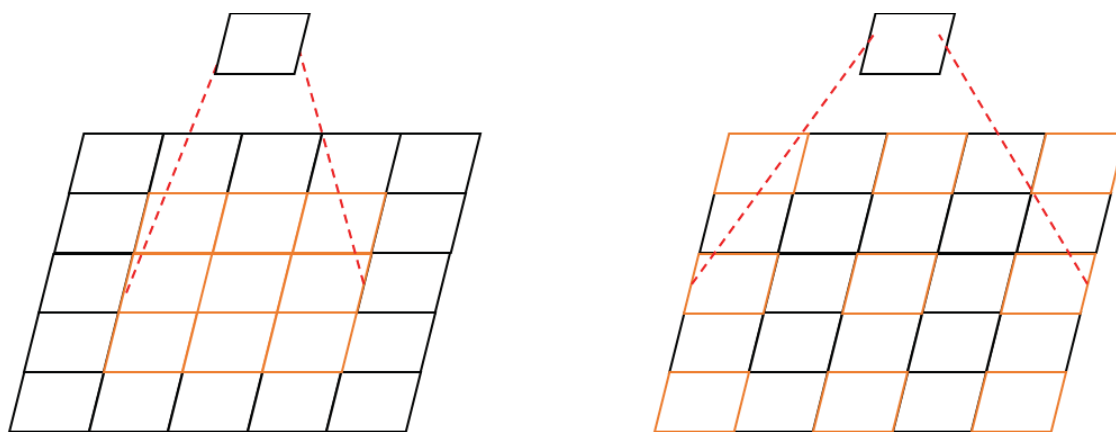


图 1.3 普通卷积和  $n=2$  的空洞卷积操作对比，橙色区域为下一层像素视野的卷积部分。双方的卷积核大小均为  $3 \times 3$ ，普通卷积中的下一层的像素感受野只能接受到相邻的八个像素和当前像素对应的区域信息，而空洞卷积感受野可以达到  $5 \times 5$  的范围大小

分割网络在日常图像中不断改进的同时，在生物医学图像分割专项任务上也有长足的进步。2015 年的 MICCAI 上，Olaf Ronneberger 等人提出了 UNet<sup>[15]</sup>这一专门给生物医学图像分割设计的算法。UNet 的网络结构如图 1.4 所示。在网络逐步下采样的过程中，感受野逐步扩大，但是最终的解卷积网络结构获取的有效信息也会逐渐减少。相比较于其他的在日常图像的分割算法，这一网络结构通过将编码器中的编码结果直接在相同分辨率的解码器中进行拼接的形式，将编码器中的信息传输给解码器，从而极大程度上增加了解码器中进行分割结果生成时的有效信息。相比较于之前的 SegNet 仅仅传输脚标的形式，UNet 结构将编解码结果进行拼接的方法传输的信息量明显更大。在 SNEMI2D 的比赛上，UNet 远超过当时的其他的分割方法。一年后的 MICCAI2016 上，Çiçek 等人在 UNet 基础上提出的 UNet3D<sup>[16]</sup>可以对于三维的数据进行分割，相比较于对于三维数据的二维分割后拼接，直接进行三维分割所获得的准确率更高，并且在 12G 显存的显卡上能够正常对于该网络进行训练，实现了对于二维和三维数据进行分割处理的基础网络结构。

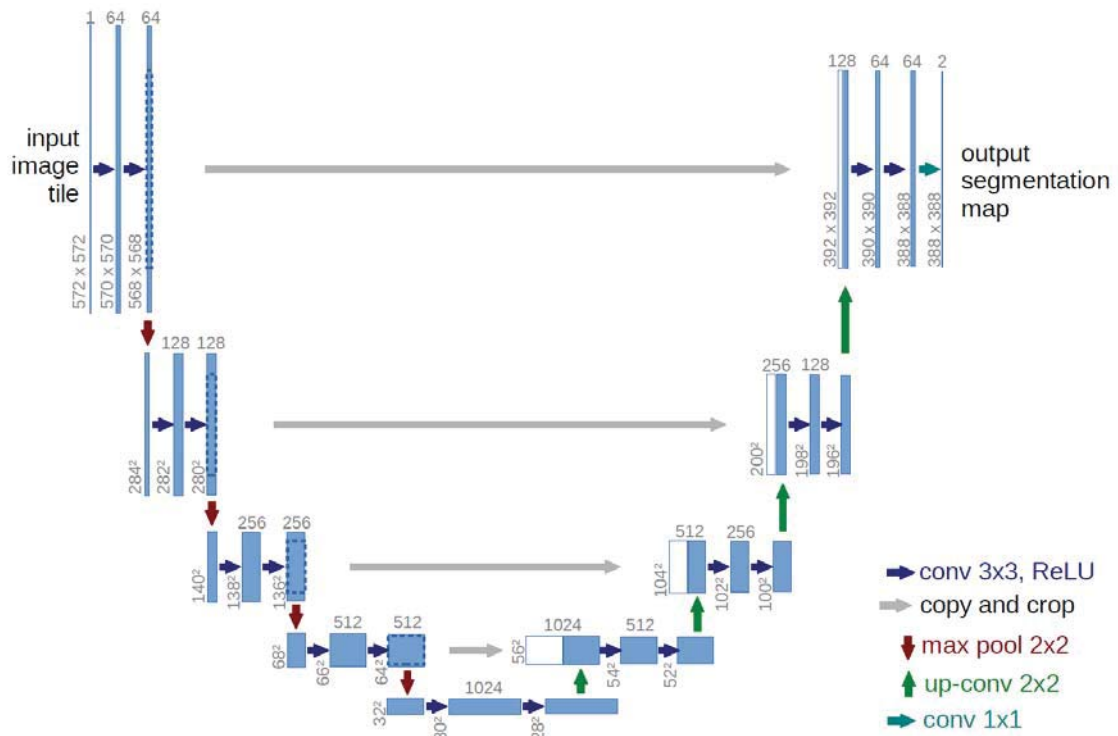


图 1.4 Rossenberger 等人提出的 UNet 网络结构

### 1.2.3 含噪声标注样本训练方法概述

标注质量较低的问题在生物医学任务中十分常见，因为标签被作为真值通常需要被直接使用在深度学习中进行训练，这一问题相比较于图像质量较低的训练集处理起来也更为困难，造成的后果也更为严重。之前有学者对于分类问题中出现的噪声标注问题进行过研究<sup>[17-23]</sup>。例如，Goldberger<sup>[17]</sup>和 Patrini<sup>[18]</sup>等人通过噪声的分布来解决标注干净和含噪声标注的数据的分布问题。通过干净数据和收到噪声污染的标注的数据的分布不同对于两类数据进行区分和处理。这一解决方法需要对于数据分布有很强的依赖性和准确的预测。Veit<sup>[20]</sup>等人则通过专注于先对于干净的数据样本进行训练后再加入噪声数据进行训练。通过人为挑选少量的干净标注的训练样本后，网络首先在这些干净的数据上进行训练，之后再逐步拓展到含有噪声的数据集上进行进一步的训练和加强。这一算法需要对于原先的数据质量进行额外的标注，在没有专家和额外的标注的情况下无法实现这一操作。Xue<sup>[23]</sup>和 Jiang<sup>[18]</sup>等人通过聚类等方法对于数据进行处理，之后对于聚类的距离关系等对于最终的训练结果进行加权，从而实现对于部分标注较好的样本予以更大权重的训练。这一方法能够在无须额外标注的情况下进行部分含噪声标注的样本

的剔除，但是这些方法并不能在对于质量评估的过程中有效结合训练的图片 and 训练的标注，使得两者的关系依旧停留在浅层，而没有进行更为深入一步的语义层次上的结合与处理。上述所有的方法都是对于分类问题的噪声标注训练中的解决办法，无法适用于图像分割问题中。

### 1.3 文章结构

本文将会主要分为 3 个部分。首先，我们会在当前的数据集上训练得到一个靶区自动勾画系统，并训练得到一个可靠的靶区自动勾画模型；然后，在图像分割问题中，考虑到训练过程中可能遇到的噪声标注的问题，我们会提出自己的解决方法；最后我们会把所有的方法和模型整合，得到一个可用的靶区自动勾画系统。基于以上的描述及功能介绍，文章分为 5 章，文章结构如下：

第 2 章将会设计基于深度学习的靶区勾画算法，实现其自动勾画功能。之后基于实验结果对于系统的性能进行进一步提升和改进，对于改进方案的性能提升以及进一步过程中遇到的新问题进行描述和分析，并得到最终的可使用的靶区勾画算法模型。

第 3 章将会在基础算法的基础上，对于算法功能进行进一步提升改造，针对靶区勾画任务中存在的噪声多、标注标准不统一等问题进行描述，提出并实现对于含有大量的噪声标注的样本数据集情况下的一个性能良好的稳定分割模型。为了进行对比，我们在公开数据集上进行了这一组实验，并且和当前最好的方法进行了效果比对。

第 4 章将会在之前所有算法的基础上，评估与陈述可实现的靶区勾画系统的需求，描述整体的实现方法，根据特定的需求进行系统的整体框架设计，并且对于不同组成模块及其所需功能进行评估描述。我们会对于整个系统中的难点进行衡量评估，设计解决方案，并设计系统的前后端。对于前端，我们会进行用户交互页面的设计，并且进行用户交互功能流程的阐述工作，对于用户交互模块各个子模块及其功能进行阐述，实现用户与服务器之间的自动化交互工作；对于后端，我们会介绍系统的核心模块的功能及算法的具体实现方法与过程，并且实现系统的模块化与功能。

第 5 章将会进行全文的综述和总结，对于之前所有实验结果进行进一步的分析，对于试验中遇到的问题进行分析和总结。

## 第 2 章 基于深度学习的靶区勾画算法研究

### 2.1 数据集与预处理

作为模型的训练集,本文采用的数据来自于安徽医科大学附属第一医院的 CT 横扫信息数据,其中可以使用的数据共 498 组,每一组均为一位病人的三维腹部区域的 CT 横扫图像的 DICOM 文件,其中存在有统一标注的一共是 340 组,其标注结果为宫颈癌部分 CTV 靶区信息,组成每组数据的二维图像张数分布在 48 张到 125 张区间范围内,每张图像大小均为  $256 \times 256$  像素,且均为灰度值图像。数据集中每张图片的像素数值最低为 0,最大为 2072。所有的图像均为对于同一部位进行的不重复单方向 CT 横扫所得结果。图 2.1 显示了其中一组数据中的一张。

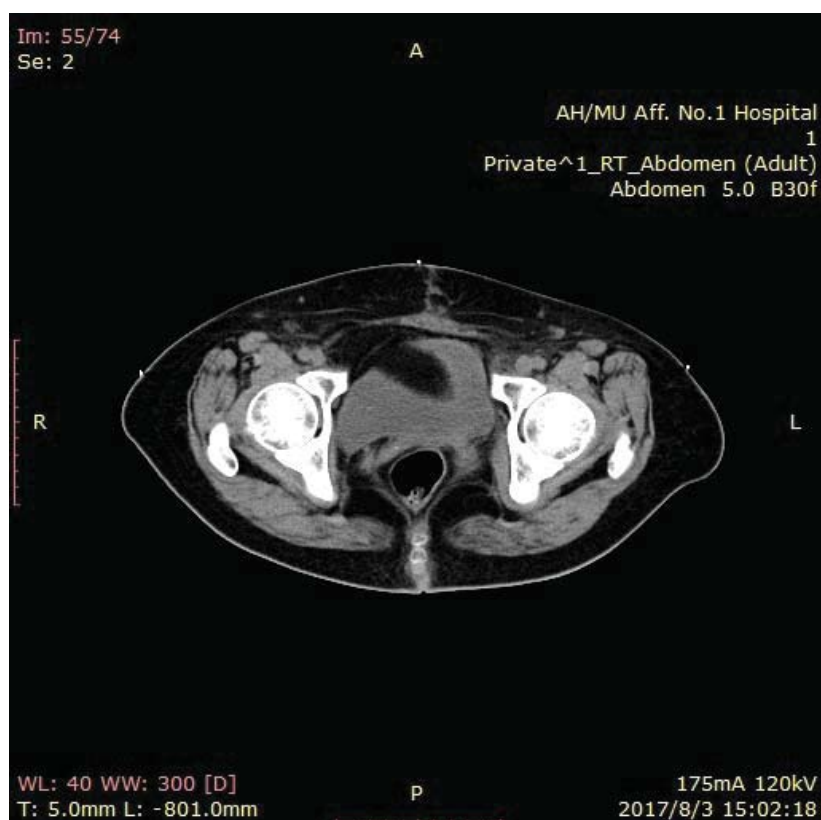


图 2.1 数据集中的样例图片可视化



深度学习网络模型对数据进行预处理时需要归一化。考虑试验中标注的稀疏性和实验的可操作性，我们统一将所有的三维图片通过双线性插值法将图像的  $z$  方向的数值固定为 72， $x$  和  $y$  轴方向的大小保持为 256 像素不变，从而将所有的横扫结果固定成为相同大小的体积。同时我们对于像素值大于 1600 的部分进行切顶，并将原图的像素值线性归一化，使得每个像素值均处于 0 到 1 区间范围内。我们将 340 组数据分为了 3 组，进行 3 折的交叉验证过程，计算其平均准确率。因为试验中数据量有限，我们在试验中进行了度数不大于  $10^\circ$  的随机旋转， $x$ 、 $y$  和  $z$  轴三个方向的翻转以及不超过 8 个像素数值的三个方向随机平移的数据增强操作，并在训练样本中加入了方差为 0.001 的噪声以提升模型的泛化能力。

## 2.2 靶区勾画模型

为了在试验中对于算法的性能进行评估衡量，获得最好的分割效果和泛化能力的模型，我们设计实验对于核心的靶区勾画的深度学习模型进行训练和测试。作为勾画模型，我们使用了 UNet3D 的网络结构作为基础，并且尝试在这一网络结构上进行多种的提升。我们采用了 Dice Ratio 这一指标对于最终的分割结果进行衡量。这一指标定义为

$$dice = \frac{2 \times Intersection(V_{pred}, V_{gt})}{V_{pred} + V_{gt}} \quad (2-1)$$

其中  $V_{pred}$  是预测为勾画结果的体素， $V_{gt}$  是标注的勾画结果的体素值。Dice Ratio 数值最小值为 0，最大值为 1，数值越大代表最终的模型的分割效果越好。

### 2.2.1 基础网络

试验中我们将 UNet3D<sup>[16]</sup> 网络作为我们实验的基础网络模型。网络结构图 2.1 所示。我们在网络中使用了 5 次的下采样和上采样操作，从而实现在更少的参数下实现原尺寸的分割结果的生成。每次下采样和上采样层之前都两个连续的卷积模块，每个卷积模块分别含有一个  $3 \times 3 \times 3$  的卷积层，一个 ELU 的激活函数层和一个批正则化层。考虑到试验中的显卡显存大小的限制，试验中的 5 次下采样后的通道数分别是 16, 32, 64, 96 和 128，并且设置每批送入的样本数为 1。在下采样层和上采样层之前有一座桥进行连接，其结构同样为两个相连的卷积模块。

在每次的下采样前，下采样的编码部分将数值传递给上采样解码网络的对应的同精度的层并与解码网络的结果进行拼接，在解码器中引入编码器的细节信息，以此来提升最终分割的细节分割处理的效果。试验中我们选择使用交叉熵作为网络的损失函数。

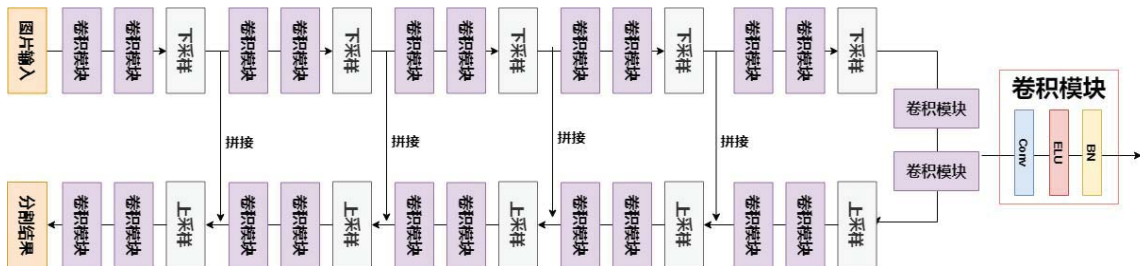


图 2.1 UNet3D 基础网络结构图。Conv 为卷积层，ELU 为激活函数，BN 为批正则化层。

## 2.2.2 基础网络改进与提升

我们在原始的 UNet3D 模型上进行了多次修改来提升基础的网络模型：

- 残差模块<sup>[7]</sup>

在基础的 UNet3D 模型的基础上，我们首先引入了 Kaiming He 等人提出的残差模块<sup>[7]</sup>。残差模块作用在卷积层的方法如图 2.2(1)所示。当网络层数增加的时候，网络中的梯度因为累积效应容易造成梯度爆炸或者梯度消失，使得梯度回传的过程中浅层的梯度值出现极端变化，导致了层数加深的网络最终的性能反而不如浅层的具有相同结构的网络。残差模块这一结构的引入可以保证在浅层的网络上梯度可以无损进行传播，一方面可以在网络更深的情况下保证训练效果的提升，另一方面可以在整个训练过程中让训练更为稳定。残差模块在 ILSVRC<sup>[3]</sup>和 COCO 2015<sup>[24]</sup>比赛中获得了五项冠军，对于网络效果提升明显。

- SE 模块<sup>[25]</sup>

SE 模块<sup>[25]</sup>是在 2018 年 CVPR 上由 Jie Hu 等人提出并且获得了 2017 年 ILSVRC<sup>[3]</sup>比赛分类任务的冠军的网络模块结构，其结构如图 2.2(2)所示。考虑到不同的通道之间的重要性可能存在部分的区分，Hu 等人在残差模块的基础上，通过对于不同的通道的信息之间进行整合后，再对于不同的通道之间的信息进行类似注意力机制的加权后，能够对于不同的通道之间信息的相对重要性程度进行一定的调整。从而对于最终的网络性能进行提升。

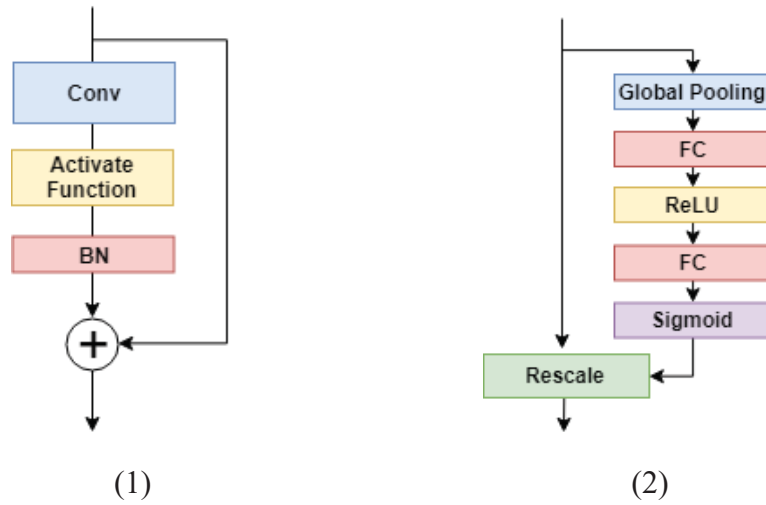


图 2.2 残差模块<sup>[7]</sup>和 SE 模块<sup>[25]</sup>

- 循环损失函数<sup>[26]</sup>

Agata Mosinska 等人在 CVPR2018 中提出了循环损失函数的思路<sup>[26]</sup>，通过网络对于分割结果的循环生成实现分割细节的进一步完善化处理。我们在网络生成分割结果的过程中借鉴了这一方法。循环生成是通过把上一次生成结果和原始图像再次送入网络中进行二次分割结果生成。在第一次结果生成前，卷积网络首先对于一个全部为 0 的分割结果和原始图像产生第一次分割的结果，之后对于这一个结果进行进一步的细化和细节的填充。这一方法可以在之前生成结果的基础上对于细节方面有更好的修改与润色，从而使得分割结果在细节方面进行结果逐步优化，对于分割细节性能上进一步提升。

- Dice 损失函数<sup>[27]</sup>

在二值交叉熵损失函数的基础上，我们引入了 Dice 损失函数<sup>[27]</sup>进行损失函数的约束和优化。Dice 损失函数的定义为

$$D_{loss} = 1 - \frac{2 \times Intersection(V_{pred}, V_{gt})}{V_{pred} + V_{gt}} \quad (2-2)$$

相比较于对于每一个体素进行分类的二值交叉熵损失函数方法，这一方法更多的对于预测的体积这一整体概念进行了约束，加强了所有的被标注的体素之间的联系。最终的损失函数 $loss$ 定义为

$$loss = Bloss + Dloss \quad (2-3)$$

其中  $Bloss$  为二值交叉熵损失函数项对应结果， $Dloss$  对应 Dice 损失函数的结果。

## 2.3 实验结果与性能分析

在这一小节里，我们会对于以上所述方法进行实验并进行结果和性能的分析。试验中我们采用 PyTorch<sup>[28]</sup> 框架，每组实验最大的迭代次数设置为 100000 次。试验中我们采用的是 Adam<sup>[29]</sup> 方式进行梯度优化，训练过程中的学习率按照经验数值设定为 0.0001。我们通过原始的 UNet3D 网络上逐一添加我们的模块进行最终分割结果的分析检验。试验中我们进行了 3 折的交叉验证，得到了如表 2.1 所示的数据，表中结果为三次实验获得的分割结果的均值。

表 2.1 勾画生成模块实验结果

方法	Dice ratio
UNet	0.747
UNet+残差模块	0.753
UNet+残差模块+SE 模块	0.755
UNet+残差模块+SE 模块+循环损失函数	0.757
UNet+残差模块+SE 模块+循环损失函数+Dice 损失函数	0.763

我们在训练过程中，每次修改都是在上一次有提升的方法基础上添加的新的模块或方法。在我们的基础模型里，Dice Ratio 值为 0.747。在添加了四种模块或者新的损失函数后，整体 Dice Ratio 数值提升到了 0.763。在试验中，每一次小的修改都有一小步的指数的提升，并且在四种模块或损失函数都使用的时候分割结果最好。在图 2.3 中我们取出了三组实验中的数据进行生成结果和标注的对比。在三个不同的测试集上，我们的 Dice Ratio 分别是 0.01, 0.64 和 0.91。三者数值相差较大，而从三张图中不同的标注我们可以看出，不同的标注者之间标注习惯存在有比较大的偏差。在 Dice Ratio 为 0.01 的标注中，医生几乎把所有的标注标在了腰椎上，这个明显和其他的标注风格有较大的不同；而 0.64 的结果中，医生更喜欢类似于条状的标注，这个和其他医生的标注结果同样存在有一定的偏差。

经过医生查看后，医生认为模型的勾画结果得到的 Dice Ratio 为 0.64 情况下，模型已经能够给整个的医学诊断提供足够的指导信息，因此这一自动勾画的核心模型可以认为是能满足医生的正常要求。

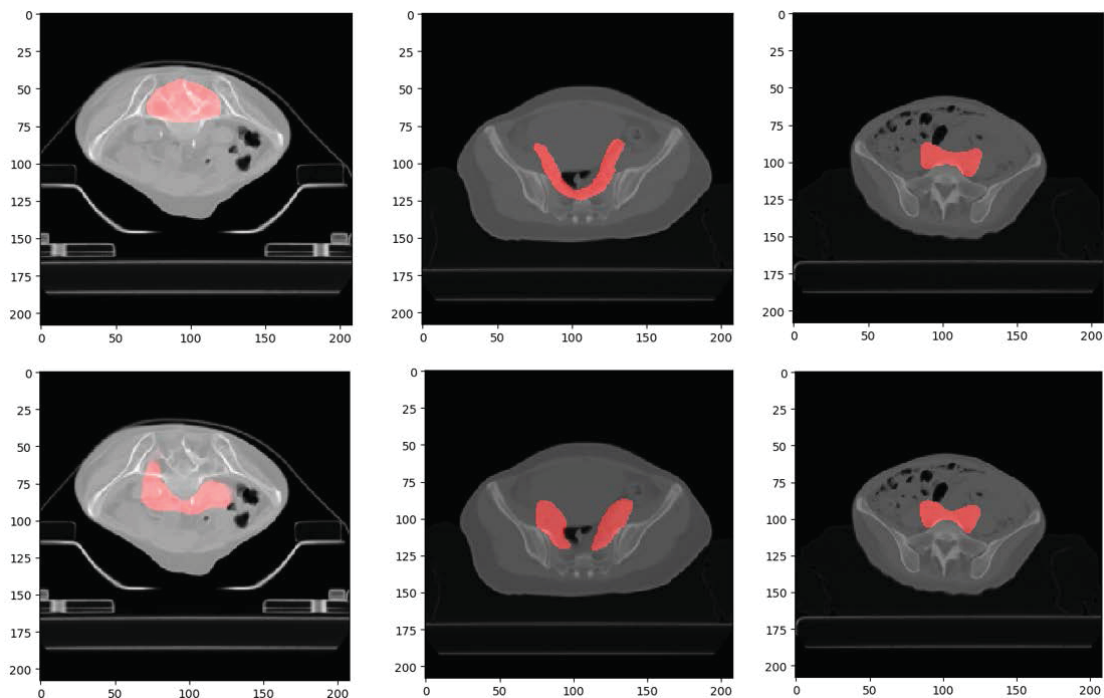


图 2.3 三组标注和预测结果。左中右 Dice Ratio 为 0.01，0.64 和 0.91

在最终的系统网络结构上，考虑到最终的实际效果，我们使用的是 UNet+残差模块+SE 模块+循环损失函数+Dice 损失函数训练得到的模型作为我们后端的勾画结果生成模块。最终的神经网络结构如图 2.4 所示。相比较于原始的 UNet3D 基础网络结构，SE 模块和残差模块在性能上有一定提升的同时，训练过程中 loss 下降更为平稳可靠。模型的参数选择上，我们选择了最终在测试集上得到的 Dice Ratio 为 0.763 的模型作为最终的后端模型的参数组成。

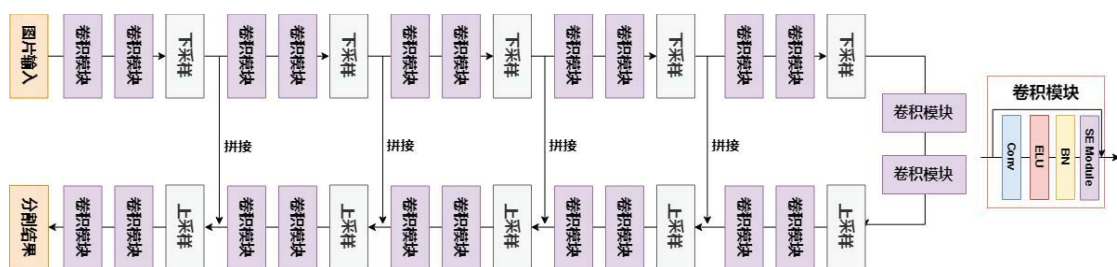


图 2.4 自动化靶区勾画最终使用模型的网络结构图

## 第 3 章 噪声标注下图像分割模型训练的研究

靶区勾画标注任务中，不同医生的标注习惯的不同，造成不同的样本之间所使用的标注标准并不是统一的。对于层次不齐的标注质量，如果直接采用原始的标注进行训练的话，所训练的模型会受到噪声样本的影响，导致最终的预测标注结果会比在没有噪声标注的数据集上训练的效果更差。如果噪声越多，这种现象会更为明显。这一问题在靶区勾画以及类似的医学图像分割问题上更为严重，因为进行类似的医学图像精确标注需要大量的专业人士进行标注工作，这个与当前的深度学习中需要大量的数据进行训练之间存在有严重的矛盾。因此如何从这些层次不齐的数据标注情况下训练得到更好的模型是一项很有意义的工作。

### 3.1 问题描述

相比较于分类问题中的标注错误通常是个例，分割问题中产生的分割的标注噪声更多的则是区域性的扩张或者缺失，如图 3.1 中所示。为了实现对于这一标注偏差问题的重现，我们通过在精确标注的分割数据集上进行 $n_i$ 个像素值的向周围扩展或者收缩，从而实现对于这一问题的仿真。同时为了满足对于现实的标注情况的仿真，变化像素值 $n_i$ 被设定为一个变动的区间范围，训练时候我们在 $n_i$ 这一范围内等概率挑选可能的像素偏差值，实现对于精确图像的加噪声处理。

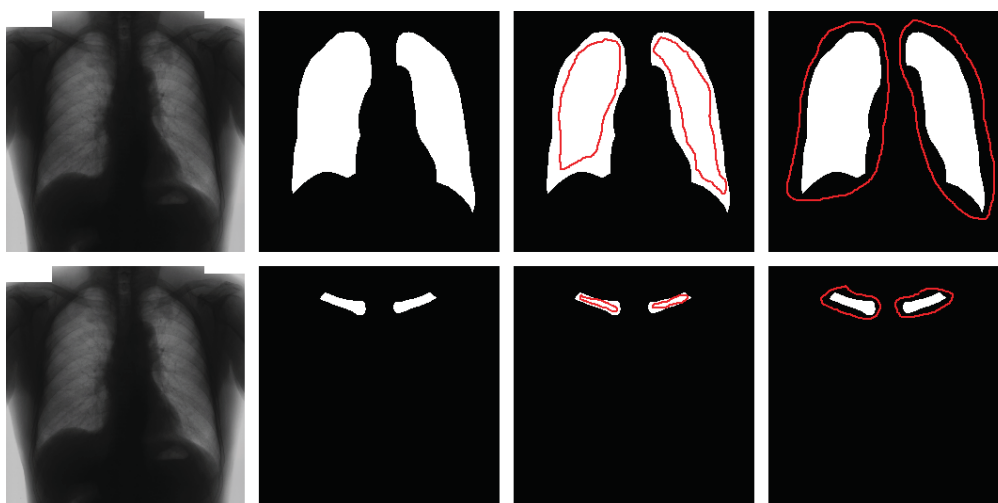


图 3.1 含噪声标注样本实例。第 2 列为准确标注，第 3、4 列红圈代表收缩和扩张标注

## 3.2 方法介绍

我们在这一小节将对于我们处理在含噪声标注的训练集下训练的方法进行介绍。对于分割问题而言，标注噪声可能存在有两者：扩张和收缩。我们希望借助于网络在无需额外的对于数据质量的标注下，对于正常标注和噪声样本标注之间进行有效的区分，使得网络能够在训练过程中挑选出精确标注的那些样本进行训练，从而避免噪声标注的样本对于模型训练造成的影响。如果网络训练中加入了含噪声标注的样本，网络的 loss 值将会下降速度更慢，并且最终稳定的 loss 值会更高。这是因为噪声程度不同的样本对应的标注标准是不同的，网络需要更多的迭代次数来对于不同的标准进行理解或者找到一个平衡点，这在扩张和收缩两种完全矛盾的标注存在下是更为明显的；而如果网络在标注标准完全统一的精确标注的训练集中进行训练的话，loss 下降速度相比而言更快，并且最终稳定的 loss 值也会更低。

借助于以上的噪声标注性质，我们提出了一个如图 3.2 所示的网络结构。我们设定送入的图像是  $x$ ，而对应的标注为  $y$ 。我们的网络功能是计算在同一个批里的各个样本之间的相对得分来挑选出质量较高的那些样本，从而进行训练。网络一共分为三个模块：分割模块、质量检测模块（QAM）和防过拟合模块（OCM）。

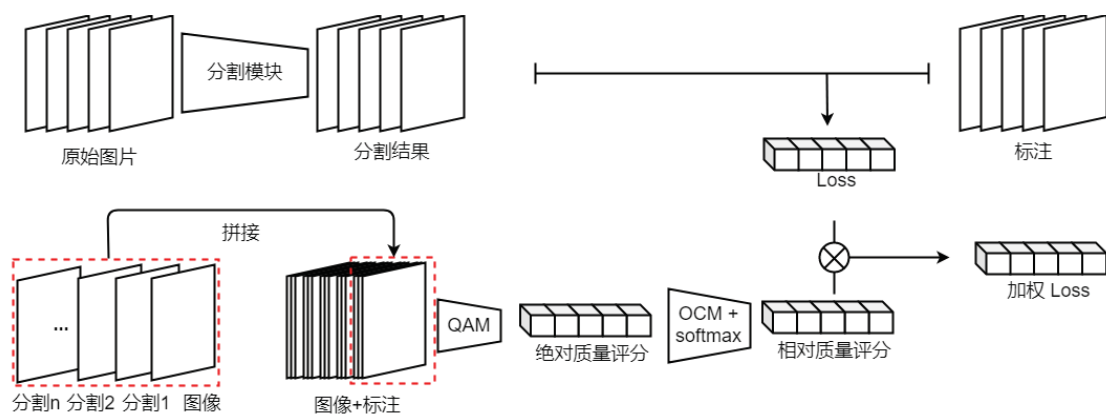


图 3.2 含噪声标注数据集训练的网络结构

### 3.2.1 分割模块

在我们的网络中，分割模块部分和现有的监督学习方法部分相同。在这一模块里网络通过映射  $x \rightarrow y_{\text{pred}}$  的方式实现对于原有图像的每一个像素点的最终类别的预测，从而实现分割预测结果。在训练中，通过计算  $y_{\text{pred}}$  和  $y$  之间的 loss 数

值后进行梯度回传实现网络的优化。在我们的网络中，我们使用了 UNet 的网络结构作为我们的分割模块的深度学习模型，并选择了二值交叉熵作为模型的损失函数。

### 3.2.2 质量评估模块

质量评估模块（Quality Awareness Module，简称 QAM）是对于处于不同的样本之间进行质量的相对值的评估的模块。模块通过对于在同一个批里面的样本进行打分的方式，挑选出在同一个批中相对而言质量更高的那些样本给予更高的权值进行训练，使得网络能够保持较好的性能与准确性。作为质量评估标准网络，我们采用的是把图片的本身和训练所用的标注拼接起来的形式，送入一个类似于 VGG 形状的质量检测网络  $\Phi(\cdot)$  进行高维信息提取。我们将 VGG 的输入通道数修改为  $n+1$ ，其中  $n$  是标注的总类别数，并且将最后一层用一个平均池化层进行替代，从而得到对于每个样本产生的分数的绝对评估值：

$$AbS = \Phi(x; y) \quad (3-1)$$

之后通过一个 Softmax 层将所有数值归一化到  $[0,1]$  区间范围内，从而产生最终对于同一个批里的数据质量的相对评估值的打分  $\Theta(x, y)$ 。在最终的 Loss 中，我们通过对于监督网络的  $L_i$  进行加权

$$\begin{aligned} \text{Loss} &= \sum_{i=1, \dots, N} \Theta(x_i, y_i) \cdot L_i \\ \text{s. t. } \sum_{i=1, \dots, N} \Theta(x_i, y_i) &= 1, 0 \leq \Theta(x_i, y_i) \leq 1 \end{aligned} \quad (3-2)$$

其中  $i$  代表的是同一个批中的样本编号。这一 loss 会作为最终的 loss 值进行梯度回传，同时对于分割网络和质量检测网络进行优化。

### 3.2.3 防过拟合模块

考虑到 Softmax 函数的性质，单纯的质量评估模块确实能够挑选出质量的相对的评估值较高的标注样本进行训练，但是在训练过程中质量评估模块会倾向于对个别样本赋予极高的权值，从而使得整个网络在少量的数据上严重过拟合。同时，如果某个样本在训练初始状态时被赋予了过高或者过低的权值的绝对评估值，Softmax 层在这一位置处的梯度已经近似于 0，会使得训练后期网络无法对于这些样本的权值进行有效的修改，造成了最终选择固化，加重最终的过拟合程度。为



了解决这一问题，我们在 Softmax 层和质量评估模块之间增加了一层防过拟合模块（Overfitting Control Module，简称 OCM）。使用方程

$$\Gamma(t) = \lambda \tanh(t) \quad (3-3)$$

将样本的分数的绝对评估值 AbScore 从原先 $(-\infty, \infty)$ 范围缩小到 $(-\lambda, \lambda)$ 。其中 $\lambda$ 是超参数， $t$ 是质量评估模块产生的质量的绝对评估值。这样操作后，在通过 Softmax 层之后，处于同一批中的质量相对值比值从最开始的 $\infty$ 降低到 $e^{2\lambda}$ 。这一数值一方面能够保证质量不同的样本之间的比值依旧较明显，同时可以解决部分的样本被赋予过高的权值以及早期的错误挑选结果使得网络在后面的训练中权值过高或者过低的难题。

### 3.3 实验结果

为了验证我们的方法并和现有方法进行对比，我们在这一部分借助公开数据集进行测试。我们选择了 JSRT 数据集<sup>[30]</sup>进行测试。JSRT 数据集一共包含有 247 张图片，每张图片都有 3 种分割标注：肋骨、肺部和心脏。每张图片都是  $1024 \times 1024$  大小的灰度值图像，我们把每张图片调整到了  $256 \times 256$  像素大小，并把数据集随机划分出 165 张进行训练，余下 82 张进行测试。

试验中我们对于人为设计的噪声设计了两组 $n_i$ 范围值，0-8 和 5-13，来仿真低噪声和高噪声这两种情况。同时我们在试验中使用了 0%、25%、50%和 75%四种不同的噪声比例进行不同噪声比例下的实验比对。试验中我们对于质量评估模块的学习率设定在 0.0001， $\lambda$ 设定为 2，批大小设置为 32。我们得到的实验数据如表 3.1 所示，其中第一组是最新方法在无噪声污染的训练集上得到的结果。

表 3.1 OCM 和 QAM 作用下的噪声标注分割结果

噪声占比	噪声程度	试验方法	肺部分割	心脏分割	锁骨分割	平均分割
无	-	InvertedNet <sup>[31]</sup>	0.972	0.935	0.902	0.936
无	-	基础方法	0.943	0.941	0.862	0.915
无	-	QAM	0.939	0.923	0.831	0.898
无	-	QAM+OCM	0.941	0.940	0.852	0.911
25%	1-8 像素	基础方法	0.868	0.888	0.538	0.765

25%	1-8 像素	QAM	0.925	0.926	0.748	0.866
25%	1-8 像素	QAM+OCM	<b>0.936</b>	<b>0.925</b>	<b>0.823</b>	<b>0.895</b>
50%	1-8 像素	基础方法	0.873	0.884	0.539	0.765
50%	1-8 像素	QAM	0.922	0.925	0.726	0.857
50%	1-8 像素	QAM+OCM	<b>0.936</b>	<b>0.929</b>	<b>0.828</b>	<b>0.895</b>
75%	1-8 像素	基础方法	0.820	0.828	0.512	0.720
75%	1-8 像素	QAM	0.898	0.825	0.536	0.753
75%	1-8 像素	QAM+OCM	<b>0.937</b>	<b>0.939</b>	<b>0.809</b>	<b>0.895</b>
25%	5-13 像素	基础方法	0.865	0.857	0.422	0.715
25%	5-13 像素	QAM	0.893	0.835	0.615	0.781
25%	5-13 像素	QAM+OCM	<b>0.935</b>	<b>0.935</b>	<b>0.801</b>	<b>0.890</b>
50%	5-13 像素	基础方法	0.755	0.807	0.393	0.652
50%	5-13 像素	QAM	0.828	0.853	0.491	0.714
50%	5-13 像素	QAM+OCM	<b>0.942</b>	<b>0.942</b>	<b>0.853</b>	<b>0.912</b>
75%	5-13 像素	基础方法	0.745	0.738	0.381	0.621
75%	5-13 像素	QAM	0.770	0.772	0.366	0.636
75%	5-13 像素	QAM+OCM	<b>0.938</b>	<b>0.937</b>	<b>0.801</b>	<b>0.892</b>

### 3.3.1 和基础方法的对比

我们在 JSRT 数据集上进行了我们的实验。我们在存在有不同程度的噪声比例和噪声程度的标注的训练集上对于模型进行了训练，并且在没有噪声标注的测试集上进行了测试。在无噪声的数据集上，无论是基础方法，QAM 还是 QAM+OCM 都有很高的分割准确率。但是随着噪声的体积和噪声程度的逐步加大，基础方法性能的准确性急剧下降。对于更小部位的分割，例如锁骨，受到的影响更为明显。在我们引入了 QAM 和 OCM 两个模块之后，相比较于基础方法在噪声情况下的表现，网络的准确性能够得到足够的提升，甚至能够保持它在不含噪声情况下训练得到的模型的准确率。图 3.3 展示了对于不同的噪声情况下模型的类均值分割准确率以及训练过程中得到的 loss 变化曲线。loss 曲线设定为在含噪声 50% 的 5-13 像素值噪声污染下训练的变化曲线。相比较于基础方法的高 loss 值，QAM 能够在训练样本中找到相对更好的样本，并且在训练的过程中体现出了相比较而言更低的 loss 值。

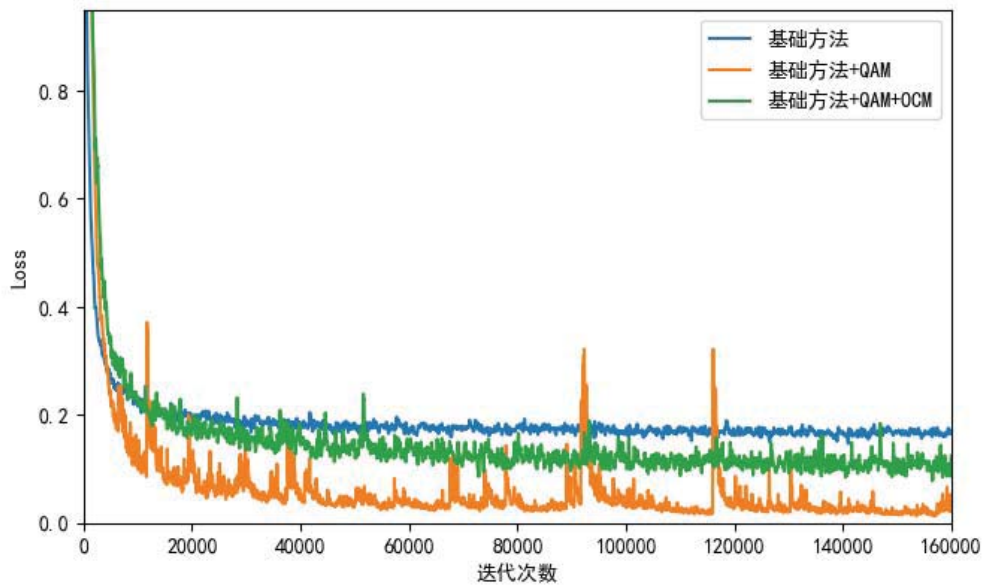
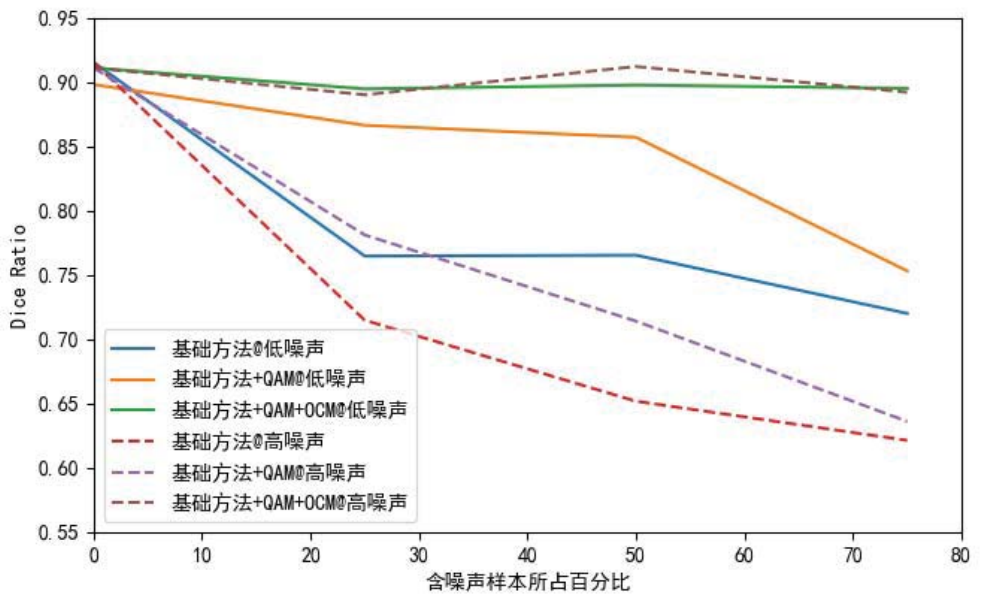


图 3.3 含噪声标注训练集下 Dice Ratio 值与训练中 loss 值变化

### 3.3.2 防过拟合模块评估

我们同样测试、衡量了试验中的防过拟合模块的影响。在没有 OCM 的情况

下, QAM 能够在训练过程中体现出更快的 loss 下降速度和相比较而言更低的 loss 值, 但是在测试集上, 没有 OCM 的 QAM 模块虽然能够比基础方法的准确率高, 却依旧会有随着噪声的程度、百分比的上升、分割的准确率严重下滑的情况, 并且数值比含有 OCM 的网络要低。当噪声标注的程度逐渐增多和严重时, loss 值的剧烈变化也证实网络出现了训练不稳定的情况。例如对于锁骨而言, 当训练集中有 75% 的样本受到了 5 到 13 像素值的噪声的影响的时候, 没有 OCM 的 QAM 模块的分割准确率从 0.862 降到了 0.366。这一数值甚至比同等噪声情况下的基础方法的分割准确率要更低, 但是一旦加入了 OCM, 同等噪声情况下锁骨的分割准确率可以保持在 0.801 这一较高水平。在噪声严重程度逐步增大的时候, 含有 OCM 模块的网络能够持续地保持在较高的分割准确率, 而一旦删去 OCM, 实验结果证明在噪声标注下训练的网络会有很严重的过拟合现象, 使得它失去自己的泛化能力。

### 3.3.3 质量评估模块评估

因为 QAM 的本质是从同一个批中挑选出来相对而言质量更高的标注的样本进行训练, 本文同样对于训练过程中的 QAM 的挑选机制的效果进行了测试和评估。在不同的训练迭代次数下, 我们在一个批中送入了 50% 的 5 到 13 像素值的噪声标注, 余下的都是无噪声的标注。我们希望观察 QAM 给两类不同的标注情况的数据的加权情况。图 3.4 展示了两类不同的标注情况下的样本获得权值均值和同类中样本的权值方差变化情况。在训练初期, 网络不能有效区分两种不同的数据, 但是给没有噪声影响的干净数据的权值随着训练轮数的增加而逐渐变高, 并且无噪声污染的标注样本和有噪声污染的标注样本之间的权值比例逐渐变大, 体现出 QAM 能够在训练开始阶段就挑选出部分的无噪声标注的样本赋予较大的权值。当训练进一步深入后, QAM 给无噪声标注的训练集中计算得到的权值的方差值开始降低并且最终在一个接近于 0 的数值处保持稳定, 证明网络在训练轮数提升后能够找到一个统一的标准来进行标注质量高低的评估工作, 并且逐步提升其召回率。在整个训练中, QAM 能够在少量迭代后迅速鉴别出标注质量低的样本, 并且在少量的有较高的标注质量的样本上进行了过拟合, 但是随着训练轮数的加深, QAM 能够把越来越多的标注质量高的样本考虑进来, 模型的泛化能力有进一步的提升, 并且网络的最终表现也越来越稳定。

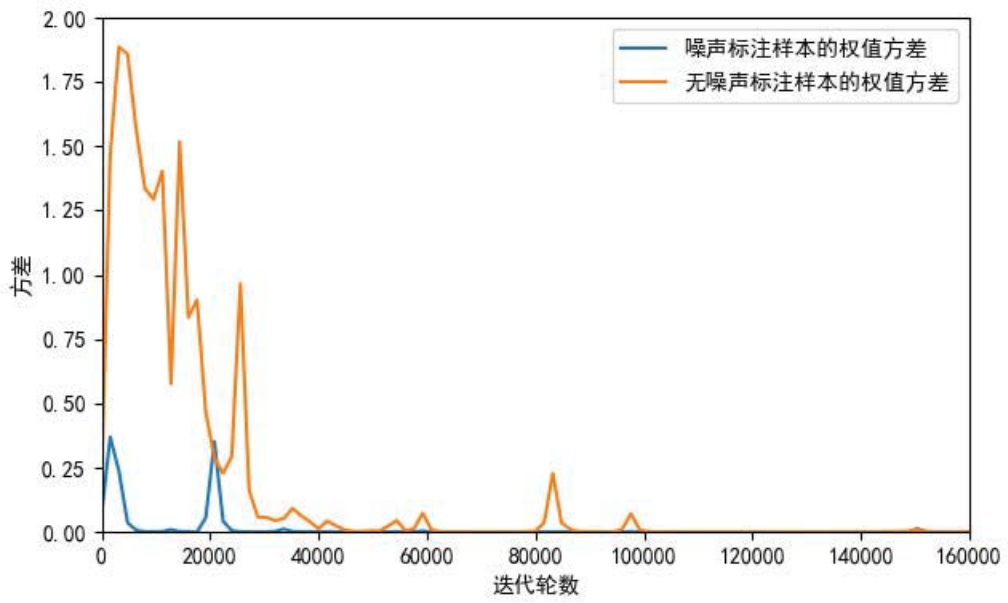
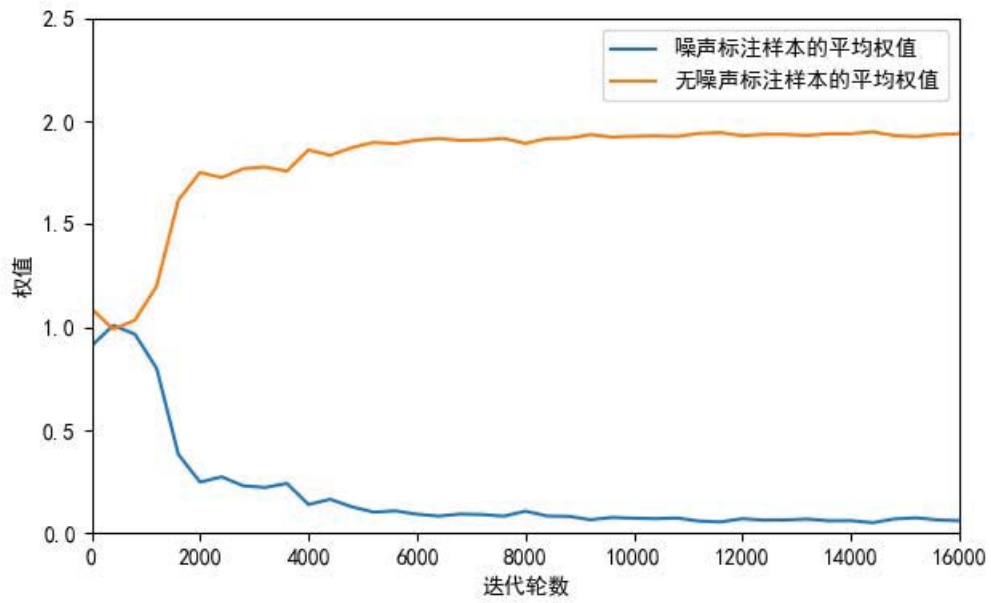


图 3.4 噪声标注样本和无噪声标注样本的权值和组内方差变化曲线

## 第4章 系统设计与实现

### 4.1 需求分析

基于上述的自动靶区勾画算法，我们设计系统使得这一算法能够更好推广，从而提供给医院方面在线操作。系统需要能够符合医院的需求，并且在原有的自动勾画功能模块上有良好的交互能力，因此在需求方面，本系统介绍如下：

前端可交互性页面。系统上线给医院使用的过程中，需要设计完成前端网页或者软件的交互。用户对于已经成型的系统上传需要进行分割的图片，系统收到后自动进行分割并且提供最终的靶区勾画结果给用户分析处理，并且提供下载功能。

后端自动分割模型。在收到用户传来的待分割的图片后，系统自动对于图片进行重新设置成规定的尺寸、归一化等一系列预处理操作，并且对于原数据进行自动的靶区勾画区域的勾定和处理。分割结束后，系统自动保存分割结果并且将分割结果提供给前端进行显示。

模型的高性能及稳定性要求。我们的模型需要最终构成系统的核心模块，因此对于系统的稳定性和性能具有一定的要求。按照医院方面的指示，我们的系统需要给放疗医生一定的辅助参考，数值化结果必须满足医院方面要求才能使用。

模型对于噪声的抗干扰性能要求。靶区勾画作为一个需要大量经验的医学任务，每个医生都需要按照个人的理解对于勾画目标进行标注。不同的医生和专家对于任务都有一定的个人理解，因此在数据集中不免有大量的噪声数据。如何对于训练集中含噪声的标注图片进行处理和训练对于得到一个稳定的模型具有很大的提升作用。这一需求作为额外任务，同样可以在其他含噪声标注样本训练过程中进行应用。

### 4.2 系统用例图

根据系统需求，医生在使用系统的时候具有上传原始数据、逐张勾画结果单一显示、整体勾画结果显示和下载勾画结果四个功能。医生在上传原始数据后，系统通过对于原始数据进行保存管理，并且进行预处理后生成勾画结果。之后系统对于勾画结果进行可视化操作以及勾画结果保存。医生通过对上传的数据和可

可视化结果进行对比进行逐张的结果观察，并且也可以直接基于可视化结果来观察整体分割结果。同时医生也可以通过勾画结果管理方式实现对于最终的勾画结果的下载。系统的用例图如图 4.1 所示。

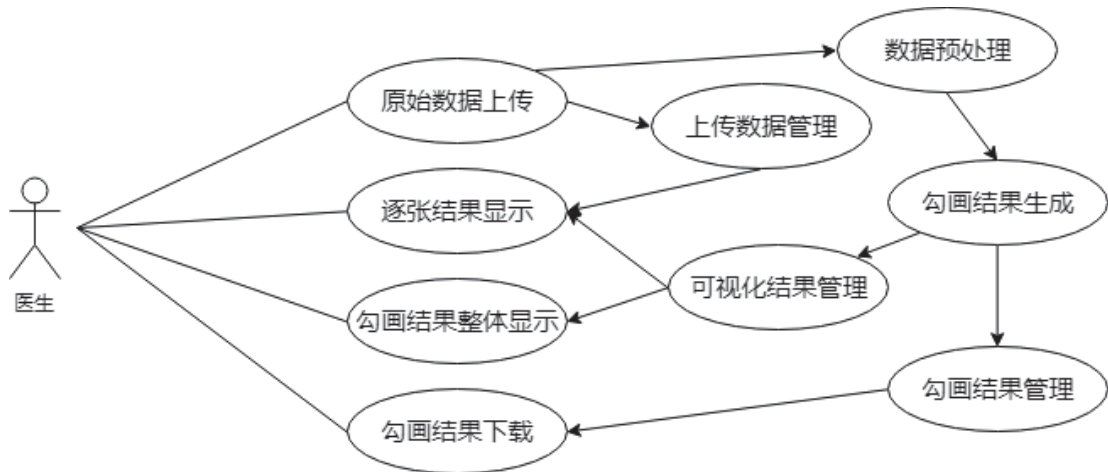


图 4.1 系统用例图

### 4.3 系统动态建模

对于整个系统部分，本文建立了如图 4.2 所示的动态建模图。因为系统进行数据转化的时间开销过大，我们把数据转化及预处理部分放在本地。医生通过在本地将 DICOM 数据转化成为 hdf5 数据后可以上传 hdf5 来实现自动分割，从而提升使用者的体验满意度。在使用系统过程中，医生首先上传数据系统会初步判断数据是否符合标准，例如所上传的数据是否包含有效的 CT 图片以及数据上传是否格式正确等。数据监测结果如果错误，系统会自动报错，并且重新回到首页。如果数据监测正确，前端将整个数据传输给后端处理系统。

后端系统在收到前端系统的信息后，进行双进程处理：一方面对于原始的数据进行存储，另一方面则是进行自动勾画操作。勾画结果生成后，系统将存储勾画结果，同时勾画端同时开启两个新的进程：一方面与原始数据存储部分合并，用户可以进行数据的可视化和逐层比对；另一方面用户可以下载和存储原始勾画结果。这一部分也是结果显示及下载部分。用户使用结束后，如果对于页面进行刷新，则用户可以回到系统的初始主页上，进行第二个需要勾画结果的上传和处理；如果用户不选择刷新，则整个流程结束。

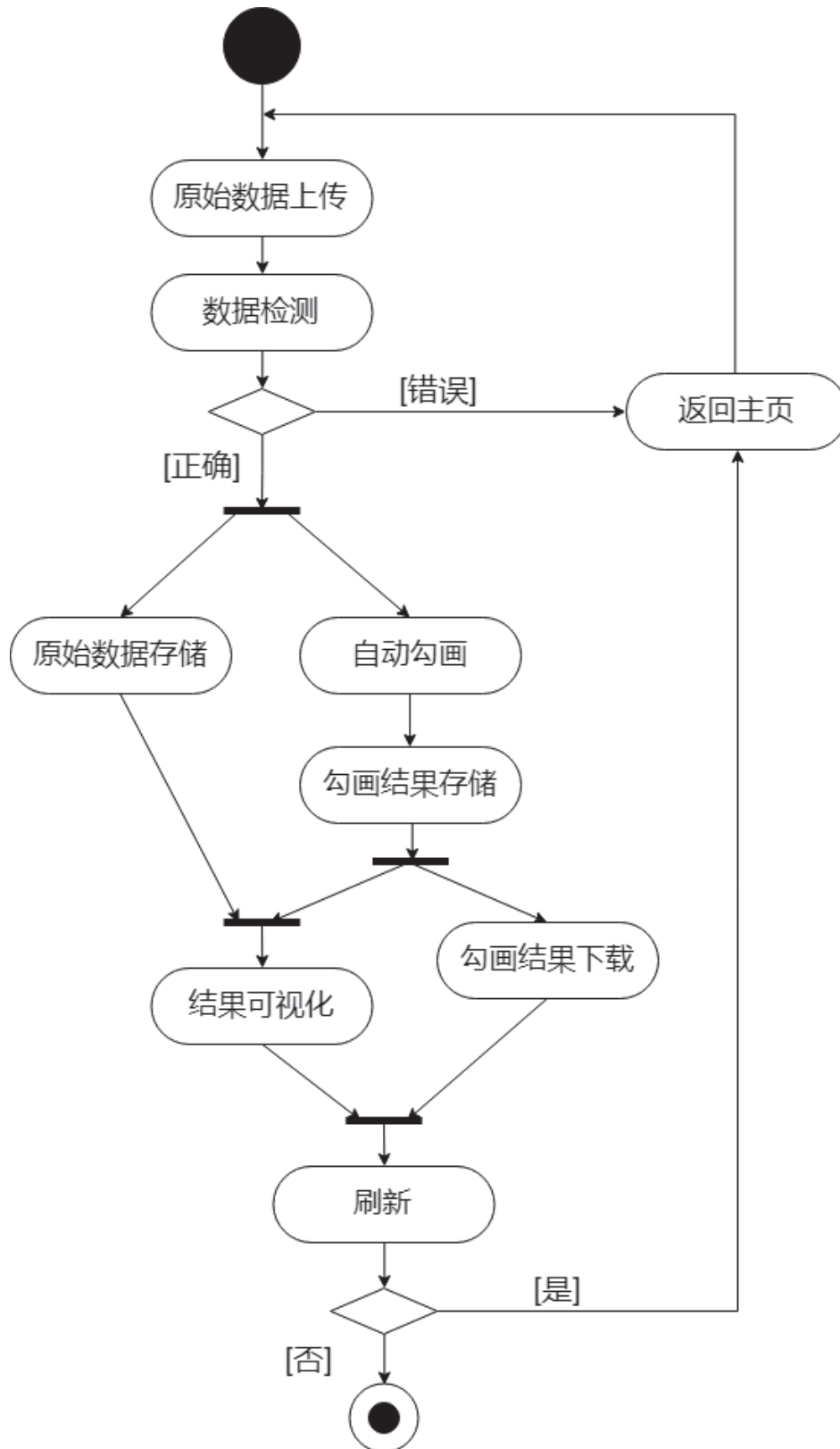


图 4.2 系统动态建模



## 4.4 前端模块需求

结合以上需求及所示的用例图、动态建模过程，我们发现系统需要一个和用户进行实施交互的前端系统，从而实现用户和后端系统之间的信息交互。作为交互的媒介，细化前端系统后，关键的功能模块有三个：上传功能模块、结果显示与对比模块、和勾画结果下载模块。前端三个核心模块与用户及后端之间的关系如图 4.3 所示。

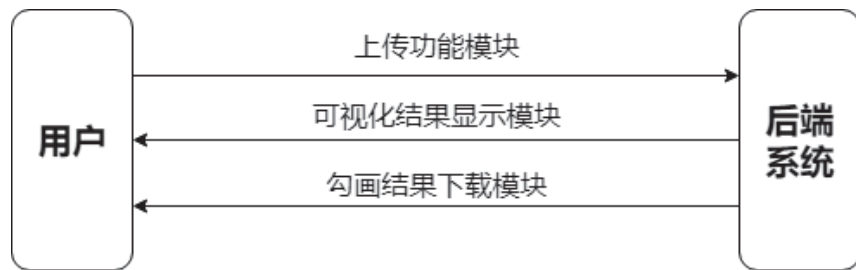


图 4.3 前端模块功能示意图

### 4.4.1 上传功能模块

上传功能模块在最终的模型中有非常重要的作用，它作为桥梁实现用户和主要功能的后端分割模块之间的交互。用户上传信息，通过简单地方式实现将本地的数据传输给在线系统的功能。上传功能模块通过提供给用户一个交互界面，使得用户通过简单的操作将本地的待勾画的数据样本上传给服务器，从而提供给后端勾画系统进行之后的操作。

### 4.4.2 可视化结果显示模块

勾画结果显示和对比模块的功能是将后端的靶区勾画结果呈现给前端进行显示。因为所上传的数据是三维的，普通的图片功能没有办法对于原始的结果和勾画结果进行显示，因此我们需要特殊的模块显示勾画结果，同时比对产生的勾画结果和原始上传数据，提供给用户最有效的可视化效果显示，从而提升用户满意度和易操作性。

### 4.4.3 勾画结果下载模块

勾画结果下载模块是提供给用户的对于后端产生的结果进行下载功能的模块。下载功能能够使得用户在本机保存勾画结果，同时让用户在本机进一步的分析和

后处理所生成的勾画结果。用户通过简单的交互界面即可从后端系统中获取需要的数据。这一模块和上传模块一样，是连接后端和用户的重要桥梁。

## 4.5 后端模块需求

后端模块包含有本系统的核心功能。通过前端系统来和用户之间进行连接，后端系统能够将从用户中获取到的原始图片进行靶区勾画结果生成，同时也要兼顾对于不同的上传和下载的结果进行版本管理工作。因此后端部分也是整个自动化系统的核心所在。基于对于后端系统的需求与性能分析，后端系统的核心部分主要由四个部分组成：预处理与数据转换模块、自动读取和勾画结果生成模块、可视化结果生成模块和数据上传与下载管理模块。

### 4.5.1 预处理与数据转换模块

预处理与数据转换模块是将原始数据进行数据格式转换和预处理后传送给勾画结果生成模块进行勾画结果生成的重要模块。在系统中，数据转换模块将数据从原始 CT 得到的 DICOM 格式转化为数据处理中更为密集标注的 hdf5 格式后，通过预处理模块实现清洗和预处理功能。这一模块的功能在第 3 章中有更为详细的对应叙述。因为数据转化和预处理模块本身需要大量的时间开销，如果放在网页端很有可能造成网络连接中断的错觉，导致用户自行多次刷新中断结果生成，因此这个模块被独立出来放在本地进行脚本运行，以改善用户的体验度。用户通过本地直接运行数据转换和预处理的脚本即可生成满足深度学习模块所需要的格式和分布的待勾画数据。

### 4.5.2 勾画结果生成模块

自动读取和生成勾画结果模块是整个系统的核心部分，其功能是针对用户上传的数据进行自动的靶区勾画结果的生成。通过一个深度学习网络，系统自动读取最新的上传的图片，并通过深度网络对于图片进行勾画结果的生成。该模块在第 2 章中有详细的介绍。

### 4.5.3 可视化结果生成模块

可视化结果生成模块是和前端系统配合的对于生成的勾画结果以及原始图片进行可视化操作的模块。系统中收到的和进行处理的数据均为三维数据，无法通过现有的图片显示方法将其在电脑上进行结果的显示。可视化结果生成模块通过

后端对于结果进行处理后，将可以通过二维方式显示的序列图片生成后保存，提供给前端对于图片序列进行显示。

#### 4.5.4 数据上传和下载管理模块

数据的上传和下载模块是通过版本管理对于用户上传的原始数据和后端生成的数据进行一一版本对应的模块。这一模块将生成结果和原始图片对应起来，以防止可能出现的版本错误。如果用户得到的最终结果和原始上传图像不对应，则会产生错误的诊断结果，造成严重医疗事故。通过上传和下载的数据进行一对一的对应与管理，系统能够保证显示的以及提供给用户下载的结果最终能够和用户上传的数据相对应。

### 4.6 前端实现

在需求分析和前段模块功能分析后，本文设计实现了前端部分相应模块的功能。前端模块通过连接用户和后端的功能模块。为了和前端进行良好的衔接，我们在试验中对于前段采用的是 `django` 的整体框架配合 `html` 语言进行用户交互行为的通信功能。

#### 4.6.1 上传功能模块实现

我们在前端的上传模块使用的是 `POST` 的方式进行文件的传输。当用户选择了一个文件之后发出 `POST` 的请求方式，前端可以自动收到这一请求方式并且分块将用户提供的问题件进行上传和写入，覆盖原有的图片数据。上传结束后，系统自动调用后端模块生成最终的勾画结果并且自动切片，以给之后的可视化结果提供帮助，并且自动跳转到展示结果的页面。相反，如果用户没有上传文件，前端则不会收到 `POST` 请求，则此时前端会有刷新的效果回到页面的初始状态。

上传模块同时拥有对于不合格或者错误的数据进行识别的功能。对于上传模块，系统同样对于错误的文件格式或者文件数据进行报错处理。因为系统只有在完整接收所有的数据后才会进行数据是否符合标准的判断，在上传模块进行完整的数据上传后，系统会尝试按照特定的数据格式对于原始数据进行读取。一旦读取报错，系统会在网页端进行明显的提示，并且停止对于数据的进一步处理，以此来给用户提示。用户进行刷新后，上传功能才会恢复，让用户提供一个新的数据进行进一步处理和分析。

#### 4.6.2 结果显示与对比模块实现

因为结果显示和对比需要有很强的实时性要求，临时进行数据的读取和切片工作会在网络连接延时之外产生大量的时间开销，我们选择在生成结果结束后立刻对于原数据和勾画结果进行切片并保存。这样在显示部分，前端只需要对于处于相同层数的原始图片、勾画结果和数据标注的三张图片进行读取和传输即可，而不需要对于整个三维数据集进行传输。这样一来单层的传输时间能够控制在 100ms 以内，用户不会有过于明显的延时体验感，可以体现出良好的实时性效果。

显示部分我们分为了两种显示方式：手动切页和自动切页。自动切页上，考虑到网络延时等原因，我们设置的自动切页的相邻两张时间间隔为 250ms，用户可以选择自动切页的方式实现连续的分割结果逐层对比，可以看出良好的分割效果的连续性比较。而在手动切页上，用户可以通过自行选择页数的方式进行专一性某层的分割结果对比，同样也能够通过键盘上快捷键进行相邻层的结果的比对和反复观察，给用户对于少量层数进行仔细观察的方式。两种方式相结合可以给用户提供多样化的可视化方式。前端结果实现后，我们可以得到如图 5.1 所示的结果比对显示和网页功能。



图 4.4 前端系统架构结果

### 4.6.3 勾画结果下载模块实现

在后端生成了勾画结果后，原先的分割结果会被自动覆盖，以保证当前下载请求获得的分割结果一定是最新的上传的数据对应的分割结果。下载功能仅仅会在用户进入可视化结果后才会出现，而可视化结果只有在分割结果保存并且切片同样得到保存后才会出现，这样可以保证用户请求的下载的结果一定是最新的结果。当用户发送了下载请求后，我们通过以 512 字节为一个块的方式进行分块快速传输，这样可以很大程度上加快结果的下载速度，使得下载所需要的时间明显减少。当用户下载结束后，页面依旧会保持在结果对比和显示页面，从而满足用户可能的“先下载后观察”的工作模式。

## 4.7 后端实现

在前端的用户交互的功能模块的基础上，我们对于后端的的功能进行了设计和搭建。为了和前端进行良好的衔接，我们后端采用的是 PyTorch 的深度学习框架，并且主要通过 Python 语言进行设计。

### 4.7.1 勾画结果生成模块实现

因为之前在算法设计部分，我们已经讨论了数据预处理方法以及深度学习的算法实现。我们直接将这一部分整合进入我们系统中。我们最终选择了 UNet+残差模块+SE 模块+循环损失函数+Dice 损失函数的网络结构，并在系统中使用了 Dice Ratio 为 0.763 的模型的参数。在用户上传了 CT 横扫数据并且保存在系统上后，后端会首先进行预处理和归一化工作，之后会自动将 3 维 CT 图片送入如图 2.4 所示的网络结构中，生成最终的自动化勾画结果并保存，以提供给可视化以及下载操作使用。

### 4.7.2 可视化结果生成模块实现

为了能够可视化最终结果，我们选择通过对于三维图像切片的方式实现最终的结果可视化。我们对于生成的结果进行切片后保存了两种格式，一种是原始生成结果的三维的勾画结果，另一种则是每一层的切片后的结果。通过对于这两种数据进行保存，我们能够同时实现对于生成结果下载和可视化的数据管理。两者在存储时都是独立存储，同时将切片画的每一张结果都存储成为二维的形式，实现最终将三维数据转化成为序列的二维图片的降维处理。

#### 4.7.3 数据上传和下载管理模块实现

作为上传和下载数据管理，后端需要给前端提供对应的图片和分割结果进行上传和下载。因为我们系统仅仅是作为小范围内使用，不需要进行大量的并行结果考虑，因此我们每次生成结果以及接收到新的上传文件的时候，我们都对于前一次的结果进行覆盖，每一次试图下载结果前一定会重新对于上传的图像进行勾画结果的生成，从而保证每次下载的结果必然是最新的，并且保证了两者必然是对应的。

## 第 5 章 总结与展望

### 5.1 工作总结

本次的毕业设计工作中，我们设计得到了一个可推广的靶区勾画系统。这一系统能够在 CT 横扫得到三维图像之后自动对于宫颈癌部分的靶区进行勾画工作。在我们的测试集上，这一系统在 Dice Ratio 这一指标值达到了 0.763，满足了医生所需的 0.64 的推广要求。在临床医学诊断上，这一模型可以很大程度上降低放疗医生的工作量，并且提取出来的高层次信息能够为治疗方案的设计、射线照射的剂量选取等工作提供帮助。

本次毕业设计工作同样对于靶区勾画系统等生物学图像分割任务中常遇到的标注质量较低的问题提出了一个解决方式。据我们了解，这是第一个对于分割问题中标注质量较低的问题的解决方法。借助于我们提出的质量评估模块和防过拟合模块，我们能够在含有噪声标注的训练集中挑选出在同一个批中相对质量较高的那些样本进行训练。这一方法能够在存在有噪声标注的数据集中训练时，使得模型的性能受到噪声标注的影响减少，甚至性能能够和在完全无噪声的数据集下训练得到的模型相匹配。

### 5.2 系统架构及方法描述

在最终试验之后，我们确定的靶区勾画系统可以上线。系统的用例图如图 4.1 所示。用户通过上传需要进行特定勾画的 CT 横扫图片，系统就能够自动生成对应的勾画部位，并且提供可视化以及下载功能。而对应于系统中最重要的一部分，我们在第 2 章进行了多组试验后使用了如图 2.4 对应的网络结构，并且使用了满足需求的 Dice Ratio 为 0.763 的模型参数进行自动化生成。

作为靶区勾画系统的拓展部分，我们设计的在含噪声标注的训练集下的训练方法通过 QAM 和 OCM 联合作用，对于处于同一个批中的样本进行打分来挑选出同一样本中标注质量的相对的评估值较高的样例进行更有力的训练。这样能够在没有额外的数据标注的质量的标注情况下，通过网络挑选得到更有利于网络训练的样本，使得网络的分割性能能够达到与没有噪声标注的数据集中训练得到的模型相当的水平，并不损失模型的泛化能力。QAM 中的参数在训练中与分割模

块相互作用,生成同一个 loss,从而一次回传的同时对于两个网络同时进行优化。因为 QAM 本身和分割模块并行的关系,在测试中只需使用分割模块即可,而不会受到 QAM 和 OCM 的影响。网络结构图如 3.2 所示。

### 5.3 问题与改进

对于靶区勾画系统而言,系统在并行运行上存在有一些小的问题。因为本次系统只提供给医生进行小范围研究使用,因此本系统只能够提供给单一用户进行运行,无法进行并行结果生成操作;同时因为上传下载保存机制的问题,用户没有办法对于之前的某个样本进行再次勾画,每次只能够对于新上传的样本进行勾画、显示结果和下载,管理方面同样有待进一步的改进。

对于靶区勾画类似的标注中存在噪声的训练方法而言,因为这一模型需要对于同一个批中的样本标注的质量相对的评估值进行评估,其最大的问题是这一模型当前只能够对于批大小大于 1 的网络结构进行使用。而当网络结构较大或者分割的原始图片尺寸较大时,例如本次模型中我们设计的三维靶区勾画系统,因为显存限制等原因,这一方法就没有办法进行使用。这也造成了我们的这一改进方法最终没有整合进入到系统中,试图生成一些更好的模型进行评测,因此未来还有进一步的提升空间。



## 插图索引

图 1.1 靶区示意图.....	2
图 1.2 ResNet 中的残差模块 .....	4
图 1.3 普通卷积和 n=2 的空洞卷积操作对比.....	5
图 1.4 Rossenberger 等人提出的 UNet 网络结构 .....	6
图 2.1 数据集中的样例图片可视化.....	8
图 2.1 UNet3D 基础网络结构图 .....	10
图 2.2 残差模块和 Squeeze-and-Excitation 模块.....	11
图 2.3 三组标注和预测结果。 .....	13
图 2.4 自动化靶区勾画最终使用模型的网络结构图.....	13
图 3.1 含噪声标注样本实例.....	14
图 3.2 含噪声标注数据集训练的网络结构.....	15
图 3.3 含噪声标注训练集下 Dice Ratio 值与训练中 loss 值变化.....	19
图 3.4 噪声标注样本和无噪声标注样本的权值和组内方差变化曲线.....	21
图 4.1 系统用例图.....	23
图 4.2 系统动态建模.....	24
图 4.3 前端模块功能示意图.....	25
图 4.4 前端系统架构结果.....	28

## 表格索引

表 2.1 勾画生成模块实验结果.....	12
表 3.1 OCM 和 QAM 作用下的噪声标注分割结果 .....	17

## 公式索引

公式 2-1 .....	9
公式 2-2 .....	11
公式 2-3 .....	12
公式 3-1 .....	16
公式 3-2 .....	16
公式 3-3 .....	17

## 参考文献

- [1] 王远瑞. 基于形变配准的交互式靶区勾画系统的设计与实现[D]. 中山大学软件工程, 2013.
- [2] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [3] Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge[J]. International journal of computer vision, 2015, 115(3): 211-252.
- [4] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.
- [5] Sánchez J, Perronnin F. High-dimensional signature compression for large-scale image classification[C]//CVPR 2011. IEEE, 2011: 1665-1672.
- [6] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [7] 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.
- [8] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.
- [9] Shi J, Malik J. Normalized cuts and image segmentation[J]. Departmental Papers (CIS), 2000: 107.
- [10] Tobias O J, Seara R. Image segmentation by histogram thresholding using fuzzy sets[J]. IEEE transactions on Image Processing, 2002, 11(12): 1457-1465.
- [11] 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.
- [12] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs[J]. arXiv preprint arXiv:1412.7062, 2014.
- [13] Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1520-1528.

- [14] 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.
- [15] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//*International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015: 234-241.
- [16] Çiçek Ö, Abdulkadir A, Lienkamp S S, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]//*International conference on medical image computing and computer-assisted intervention*. Springer, Cham, 2016: 424-432.
- [17] Goldberger J, Ben-Reuven E. Training deep neural-networks using a noise adaptation layer[J]. 2016.
- [18] Patrini G, Rozza A, Krishna Menon A, et al. Making deep neural networks robust to label noise: A loss correction approach[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017: 1944-1952.
- [19] Jiang L, Zhou Z, Leung T, et al. Mentornet: Regularizing very deep neural networks on corrupted labels[J]. *arXiv preprint arXiv:1712.05055*, 2017, 4.
- [20] Tanaka D, Ikami D, Yamasaki T, et al. Joint optimization framework for learning with noisy labels[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018: 5552-5560.
- [21] Veit A, Alldrin N, Chechik G, et al. Learning from noisy large-scale datasets with minimal supervision[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017: 839-847.
- [22] Dgani Y, Greenspan H, Goldberger J. Training a neural network based on unreliable human annotation of medical images[C]//*2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018: 39-42.
- [23] Xue C, Dou Q, Shi X, et al. Robust Learning at Noisy Labeled Medical Images: Applied to Skin Lesion Classification[J]. *arXiv preprint arXiv:1901.07759*, 2019.
- [24] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]//*European conference on computer vision*. Springer, Cham, 2014: 740-755.
- [25] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018: 7132-7141.
- [26] Mosinska A, Marquez-Neila P, Koziński M, et al. Beyond the pixel-wise loss for topology-aware delineation[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018: 3136-3145.

- [27] 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.
- [28] Paszke A, Gross S, Chintala S, et al. Automatic differentiation in pytorch[J]. 2017.
- [29] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
- [30] Shiraishi J, Katsuragawa S, Ikezoe J, et al. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists'detection of pulmonary nodules.[J]. AJR. American journal of roentgenology, 2000,174(1):71-74.
- [31] Novikov A A, Lenis D, Major D, et al. Fully convolutional architectures for multiclass segmentation in chest radiographs[J]. IEEE transactions on medical imaging, 2018, 37(8): 1865-1876.

## 致谢

感谢导师吴及教授和石佳琳学姐的帮助和悉心指导，为我的毕设思路和方法构思上指引了方向，并提出了大量指导性的建议。

感谢安徽医科大学附属第一医院给我的毕设工作提供原始数据及标注。

## 声明

本人郑重声明：所呈交的学位论文，是在导师指下独立进行研究工作所取得的成果。尽我知，除文中已经注明引用内容外本学位论文研究成果不包含任何他人享有著作权的内容。对本论文所涉及研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签名： 朱海东 日期： 2019.6.13



## 附录 A 书面翻译

### A.1 U-Net: 为生物医学图像分割设计的卷积网络

Olaf Ronneberger, Philipp Fischer和Thomas Brox

德国弗莱堡大学计算机系及BIOSS生物信号研究中心

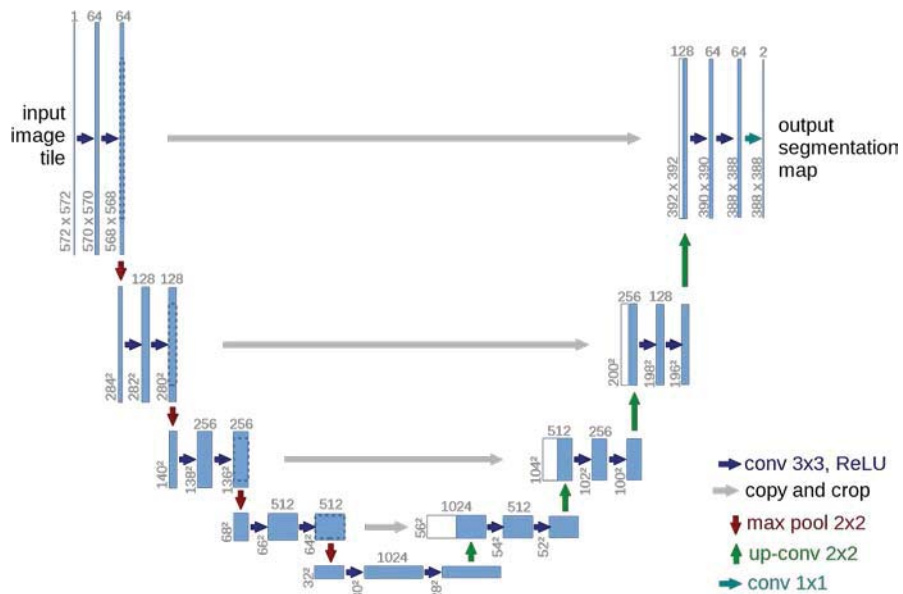
**摘要:** 如何成功训练出一个性能优异深度学习的网络呢? 在这一个问题上, 大量的训练样本的必要性是学术界的共识。在这篇文章里, 我们同时提出一个新的网络结构和我们用来训练的方式, 使得我们的训练能够通过大量的数据增强使得能够更为有效地使用现有数量的标记的样本。我们的网络结构包括了一个特征提取路径来获取高维度信息, 并且有一个相似的扩充网络来保证每个分割的准确的位置。我们证明了这个网络可以在很少的图片的情况下进行端到端的训练, 并且在ISBI电镜成像细胞分割挑战中超过了先前的最好的方法(带有滑窗的卷积网络)。使用在光学显微镜下成像的图像训练的网络, 我们在2015年的ISBI细胞追踪竞赛中获得了胜利。同时, 我们的网络运行更快。对于一张512x512大小的图片进行分割, 仅需要在新GPU上耗时不到1秒。网络结构(基于Caffe编写的)以及与训练模型已经开源, 可以在网站上获取。

#### A.1.1 介绍

追溯学术界前两年的时间里, 卷积神经网络在很多的问题取得了杰出的成果, 尤其是许多的计算机视觉任务, 例如<sup>[7,3]</sup>。虽然用于深度学习的网络基础诞生时间较早, 历史悠久<sup>[8]</sup>, 因为外界因素的影响, 例如存在良好标注的训练用样本数目以及显存大小的限制等, 它们的成功性依旧受到额外的限制。Krizhevsky et al.在<sup>[7]</sup>中的突破是因为在一个有8层的上百万的参数的网络上通过近100万张图片的Imagnet数据集进行训练。从那以后, 更大更深的网络就开始被训练<sup>[12]</sup>。

卷积网络的经典应用是分类任务, 这一任务的一张输入图片的结果是它对应的类别标签。但是, 在当前许许多多的图像分析任务中, 我们希望得到的输出不仅仅具备有类别的信息, 同样希望它是包含位置信息的, 这一需求在医学领域是格外迫切的。对于医生而言, 大数据量的医学图像一方面对于用户隐私进行侵犯,

另一方面对于标注医生而言是一种困扰，实现的可能性极低。因此，Ciresan et al.<sup>[1]</sup>通过一个滑窗设定训练了一个网络来，从而预测某个数据集中体素或者像素对应的分类信息，从而对于这个体素或者像素的分类任务进行指导作用。这样而来，第一这个网络可以在局部进行信息处理；第二，通过图像块进行训练得到的图像块的数量远多于可使用图像进行训练的数量。这个结构的网络在ISBI 2012年的内镜成像图像比赛中以较大优势获得胜出。

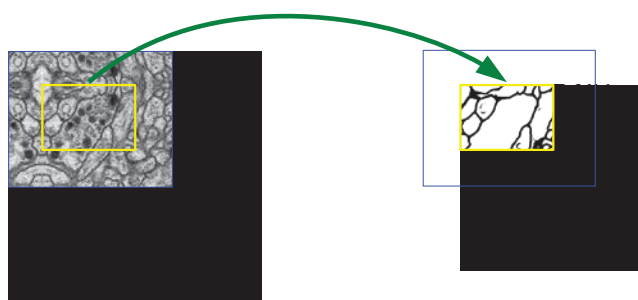


图A.1.1 UNet网络结构。每个蓝色方盒代表一个多通道的特征图，白色方块代表复制的特征图，不同的箭头代表不同的操作

非常明显，Ciresan et al.<sup>[1]</sup>使用的策略有两个缺点。第一，网络比较慢，因为网络必须分别对于每一个图像块进行处理，并且有大量的重复浪费，因为在图像的重叠部分处进行运行；另一个则是在局部准确性上以及输入使用的信息上有一个平衡问题。更大的图像块需要更多的最大池化层而降低局部的准确性，然而较小的图像块会让网络仅仅看见局部的图像信息。更近一些的方法<sup>[11,3]</sup>设计了一个分类器输出使得特征从不同的层的信息都会被纳入考虑。好的定位和使用图像的信息量上在相同的实验中成为了可能。

在这篇论文里，我们搭建了一个更为优雅的网络结构，叫做“全卷积网络”<sup>[9]</sup>。我们对于网络的结构进行了一定的修改并且扩充使得它能够在非常少的图片上依旧训练得到很高的分割精度，参见图表1。文献<sup>[9]</sup>中的主要想法是通过使用一

个常见的拼接网络作为后续的几层，这些网络里面均没有池化操作层，取而代之的是在相同位置加入的特定的上采样层，从而能够使得在最终的网路里，输出的预测的分割结果的精度得到更好的提高。为了能够进行定位操作，从拼接处产生的高精度的特征被用来和上采样的结果拼在一起。一个后面的卷积可以基于这些信息学习组合出更近准的输出。



图A.1.2 试验中对于交叠部分的分割操作。网络的道德最终的分割结果区域是黄框所示区域，网络的输入对应的区域是左图蓝框所示区域

我们的网络与众不同，其中最主要的创新点，是我们的上采样部分将卷积核通道的数目进行了提升，从而满足更高精度分割结果的需求。因此，上采样的路径在一定程度上关于连接呈现为镜像结构，并且形成了一个类似于U型的上开下和的结构。我们删除了网络中全部的全连接层，卷积核采用了有效区域（即valid）部分的结果，这样网络可以有较大的感受野，对于输入的图像的每一个像素，网络都能够感知与处理。这一策略使得网络在交叠部位进行了冗余分割，产生了多组结果，从而能够产生把图像中的断裂现象解决（如图A.1.2）。按照原有算法，边界部位的分割结果会被损失，我们最终采用了这一方法能够良好解决这一问题。在图像尺寸比较大的情况下，这一种堆叠方式是极其重要的，因为如果不这样的话，GPU的显存大小会限制网络大小，最终使得我们得到的结果更为模糊。

对于我们的任务，训练的数据集数量非常有限，我们使有对于输入图像的弹性变形产生额外的数据增强。这样可以让网络学习到更多的变量而不需要让这些变化需要在标注的图像中出现。刚性变换是在医学问题上对于训练数据有限情况下获取更多的数据的最有效的手段，同时也是很常见的细胞学现象。提供用来学习参数的数值在Dosovitskiy et al.<sup>[2]</sup>的非监督特征学习的图中已经被展示出来。

另一个在细胞分割任务中的挑战是把属于同一类的接触到一起的东西分割开来，参考图A.1.3.在这一种情况下，我们提出使用一种带参数的损失函数，在接触

到的细胞的边界处的背景标签可以在损失函数里得到大的权重。

我们的得到的网络绝不仅限于上文描述，该网络有很广阔的使用空间。作为结果，我们实现了在光学显微镜下产生的神经元的分割结构分割（一种自从ISBI2012后一直持续的竞赛）。我们在这一任务上超过了Ciresan等人提出的结构。更重要的是，我们展示了我们的在光学显微镜下成像的ISBI2015细胞分割的结果。这里我们以较大的优势赢得了当前最具挑战性的两个2D光学数据集。

### A. 1. 2 网络结构

我们的网络结构如图1所示，它包含了一条上卷积路径（左半条）和下卷积途径（右半条）。拼接路径我们仿照了传统的卷积神经网络的结构，包含有相同配置的两个 $3 \times 3$ 的卷积核（没有扩充的版本），每个卷积后面都连接了一个非线性激活单元（ReLU函数）和一个 $2 \times 2$ 最大采样的操作并且步长为2进行下采样。每次下采样的步长里，下一层的通道数都是上一层的两倍。在含有上采样的道路中，每一步包含了一个对于特征图的 $2 \times 2$ 的上采样方式并且把通道数减一半进行处理，并通过拼接路径进行一个拼接操作带有对应的裁剪，并且通过两次 $3 \times 3$ 的卷积和ReLU激活函数操作。裁剪过程是很有必要的，因为在卷积过程中会因为每次卷积过程中，卷积边界的影响造成边界像素的损失。在最后一层上，我们使用了一个 $1 \times 1$ 的卷积来把64个特征组成向量从而实现最终的类别数的分类。综上，网络的卷积层数目为23，并且为全卷积结构。

我们选择了一个更为合适的输入图像尺寸，即使得输入的每一个下采样层的特征大小都是偶数。这个尺寸的结果对于我们最后生成冗余的分割结果和最后生成原精度大小的分割结果是极其重要的。

### A. 1. 3 训练

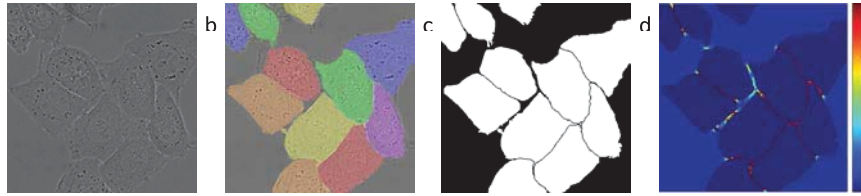
在训练过程中，我们使用了Caffe进行网络的架构。因为卷积操作使用的是我们的有效卷积核部分，如蓝框和黄框所示，我们的网络最终输出的图像比我们的输入图像要小一个固定值。为了最小化这种边界空隙的影响，相比较更大的批数目，我们更倾向于使用更大的图片块来尽可能完全地使用GPU资源，因此我们把图片的每次训练批数目设置为1。同样的，我们使用了高的动量值0.99使得之前见过的多个图片来共同影响当前步骤的更新状态。

在参数优化情况中，我们选择了随机梯度下降（SGD）进行参数优化。关于能量方程，我们使用了在最后的特征图片上的像素的Softmax结合交叉熵损失函数。

Softmax被定义为 $p_k(x) = \exp(a_k(x)) / (\sum \exp(a_{k'}(x)))$ ，其中 $a_k(x)$ 是对于每个特征通道 $k$ 在像素 $x$ 位置处的激活值， $x$ 是二维整数。 $K$ 是类别数目， $p_k(x)$ 是最大似然函数。我们让 $p_k(x) \approx 1$ 对于那些有最大的激活值的 $a_k(x)$ ，并且对于其他的 $k$ 都有 $p_k(x) \approx 0$ 。交叉熵函数对于每一个位置的倒数进行计算，使用

$$E = \sum w(x) \cdot \log(p_{l(x)}(x)) \quad (A-1)$$

这里 $l$ 是每个像素的真实标签，并且 $w$ 是我们引入的对于不同像素进行计算得权重图。



图A.1.3 通过DIC获取的HeLa细胞图像。a为原始图像，b为细胞和标注的交叠，c为我们生成的细胞分割结果图像，d为像素级别的权重加成

我们在训练集上对于真值标注的分割图像上进行每个类别出现的像素数目的频率进行预计算，并且对于涂总标注比较细的部分（如图中的细胞边界部分），我们会赋予更大的权值，从而让它能够更加重视这一些边界的结果。

细胞分割的边界是通过形态学操作来进行计算得，权重图像被

$$w(x) = w_c(x) + w_0 \cdot \exp\left(-\frac{(d_1(x) + d_2(x))^2}{2\sigma^2}\right) \quad (A-2)$$

进行计算。公式中出现的 $w_c$ 是为了权衡出现的频率，特意设计的与其相关的权值。参数设定我们按照原文进行，使用了距离边界最近的两个细胞的距离 $d_1$ 和 $d_2$ 进行 $w_0$ 的权衡和调控，并且按照原文结果进行超参设置 $w_0$ 为10， $\sigma$ 为5。

在有很多的卷积层和不同的分支连通的深度网络里，一个好的权重初始化是非常重要的，否则网络的部分结构会被过度激活，而另一部分可能会永远没有任何贡献。如果我们的设置合理的话，我们最终得到的网络的每一层的特征上，其分布应当满足方差为1的高斯分布情况。为了实现我们的实验设定，我们设计从这

方差为 $\sqrt{2/N}$ 的高斯分布中获得我们网络所需的分布，其中N代表的是下一层的特征图中，每一个像素或者体素所能够看到的上一层的感受野的大小。例如对于一个 $3 \times 3$ 卷积和64通道的卷积层而言， $N=9 \times 64=576$ 。

### A. 1. 3. 1 数据增强

对于训练一个具有稳定和鲁棒性较好的网络来说，如果获取的样本数目非常少的时候，数据增强是非常必要的。这样的操作一方面能够很大程度上增加我们的数据的数量，同时也可以增加数据的可变成都，尤其是对应的翻转、平移、变换等操作。我们对于随机位移图像增加了一个 $3 \times 3$ 感受也得平滑操作。这一替代事从高斯分布上获取方差为10个像素值的数据。单一像素数值通过变换前图像进行二项插值来进行。随机失活在最后的拼接层上能够提供更多的数据增强。

### A. 1. 4 实验

我们在三个不同的分割任务中阐述了UNet的效果。第一个任务是电镜成像下的神经元结构的分割。图像2展示了一个从数据集中获取的数据的例子与我们获得的分割结果。我们把全部的结果作为附加材料进行了提交。这一数据集是从2012年开始的ISBI显微电镜分割挑战中获取的，这一数据集依旧可以进行新的结果的测试。训练数据是一组30张照片（ $512 \times 512$ 像素）从不同位置的果蝇第一龄幼虫腹侧神经索的光学显微镜图片中获得的。每个图片来自一个对应的有完全标注的分割图像，标注区域中明亮的部分为细胞体，而较为灰暗的部分为细胞间质或者细胞膜对应部位。我们在提交一个细胞膜分布概率图给竞赛的组织方后可以得到一个评价数值。这一评价标准是在10个不同的阈值的时候对于整个图进行二值化处理，并且计算了拼接误差、随机误差和像素误差三个不同的指标。

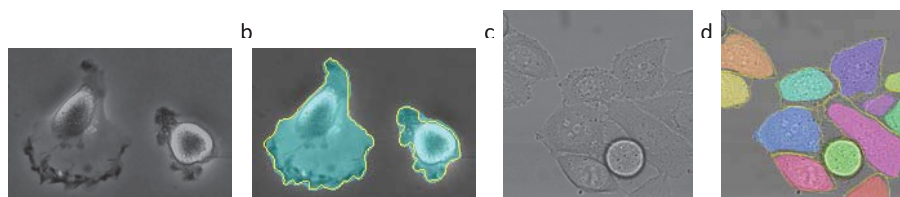
表 A.1.1. EM 数据集上分割结果

排名	组名	Warping 误差	Rand 误差	像素误差
	<b>**人类结果 **</b>	5e-6	21e-4	10e-4
1.	u-net	353e-6	382e-4	611e-4
2.	D-S 算法	355e-6	305e-4	584e-4
3.	IDSIA 算法 <sup>[1]</sup>	420e-6	504e-4	613e-4
4.	DIVE 算法	430e-6	545e-4	582e-4
10	I-SCI 算法	653e-6	189e-4	1027e-4

我们的UNet结构（在对于输入数据进行7次不同的旋转后取各个版本的平均

值)能够在没有任何的预处理或者后处理的情况下达到了0.0003529的拼接误差(最新的最好的结果,参见表1)和0.0382的随机误差。

这一结果明显比Ciresan等人提出的滑窗卷积的结果要更好,他们的最好的版本拼接误差为0.000420,随机误差为0.0504。在随机误差这一项上,唯一比我们在这个数据集上表现更好的算法使用了大量专门设计的对于Ciresan等人产生的概率图谱得到的后处理方法。



图A.1.4 ISBI细胞追踪挑战结果。a为送入的原始图像, b为分割结果, c是送入的DIC-HeLa图像, d为DIC-HeLa图像的分割结果

我们同样把unet是用在了光学显微镜成像的细胞分割任务上。这个分割任务是ISBI2014和ISBI2015提出的细胞跟踪挑战赛。第一个数据集,“PhC-U373”,包括了一个光学对比显微镜记录下的聚丙烯酰胺基质。它包括了35个部分标注的训练图片。这里我们达到了92%的IoU数值,比当前第二好的算法的83%的结果明显更优。第二个数据集是“DIC-HeLa”,这些是HeLa细胞在一个玻璃玻片上通过差分干涉对比度显微镜获取的图像。它包括了20张部分标注的训练图片,在这里我们达到了77.5%的IoU数值,同样比第二的46%的算法明显更高。

表 A.1.2. ISBI 2015 细胞追踪比赛中的分割结果

方法名称	PhC 数据集	DH 数据集
I-SG (2014 年提出)	0.27	0.29
K-SE (2014 年提出)	0.80	0.46
H-US (2014 年提出)	0.53	没有测试
2015 第二好的方法	0.83	0.46
我们的方法 (2015 年提出)	0.92	0.78

### A. 1.5 结论

UNet结构能够在不同的生物医学图像分割任务上都有很好的结果。感谢数据

增强以及图形学变换, UNet只需要很少的标注图像就能够在单英伟达GPU (6GB) 上用10小时极其合理的训练时间得到很好的结果。我们提供了一个完整地caffe版本的实施以及预训练的网络, 我们同样很确信UNet结构也能够轻松在其他更多的任务上进行实施。

## 参考文献

- [1] Ciresan D, Giusti A, Gambardella L M, et al. Deep neural networks segment neuronal membranes in electron microscopy images[C]//Advances in neural information processing systems. 2012: 2843-2851.
- [2] Dosovitskiy A, Springenberg J T, Riedmiller M, et al. Discriminative unsupervised feature learning with convolutional neural networks[C]//Advances in neural information processing systems. 2014: 766-774.
- [3] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.
- [4] Hariharan B, Arbeláez P, Girshick R, et al. Hypercolumns for object segmentation and fine-grained localization[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 447-456.
- [5] 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.
- [6] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding[C]//Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014: 675-678.
- [7] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.
- [8] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural computation, 1989, 1(4): 541-551.
- [9] 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.
- [10] Maška M, Ulman V, Svoboda D, et al. A benchmark for comparison of cell tracking algorithms[J]. Bioinformatics, 2014, 30(11): 1609-1617.
- [11] Seyedhosseini M, Sajjadi M, Tasdizen T. Image segmentation with cascaded hierarchical models and logistic disjunctive normal networks[C]//Proceedings of the IEEE International



Conference on Computer Vision. 2013: 2168-2175.

[12] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[13] WWW: Web page of the cell tracking challenge, [http://www.codesolorzano.com/celltrackingchallenge/Cell\\_Tracking\\_Challenge/Welcome.html](http://www.codesolorzano.com/celltrackingchallenge/Cell_Tracking_Challenge/Welcome.html)

[14] WWW: Web page of the em segmentation challenge, [http://brainiac2.mit.edu/isbi\\_challenge/](http://brainiac2.mit.edu/isbi_challenge/)

## A. 2 3D UNet: 从稀疏标注中学习密集的体分割

Özgun Çiçek, Ahmed Abdulkadir, Soeren Lienkamp, Thomas Brox, Olaf Ronneberger

**概要:** 这篇论文介绍了一个从稀疏的体标注的图片中得到密集的体分割的网络模型。我们标注出这种方式存在的两项吸引力: (1) 在一个半自动的设置中, 用户可能会标注部分要分割的体积。网络能够从这些稀疏标注中学习到信息并且生成一个密集分割的结果。(2) 在全自动分割的设置中, 我们假设一部分稀疏标注的训练集作为代表是存在的。在这些数据集上训练的网络模型能够生成新的密集体分割图像。我们提出的网络拓展了UNet的网络结构, 并且把所有的2D操作都用3D的操作进行了替代。该实现执行动态弹性变形, 以在训练期间有效地增加数据。这个端到端的网络训练模型无需预训练网络。我们在一个复杂的高度形变的3D结构上测试了我们提出的方法, 并且在多种情况下都获得了不错的结果。

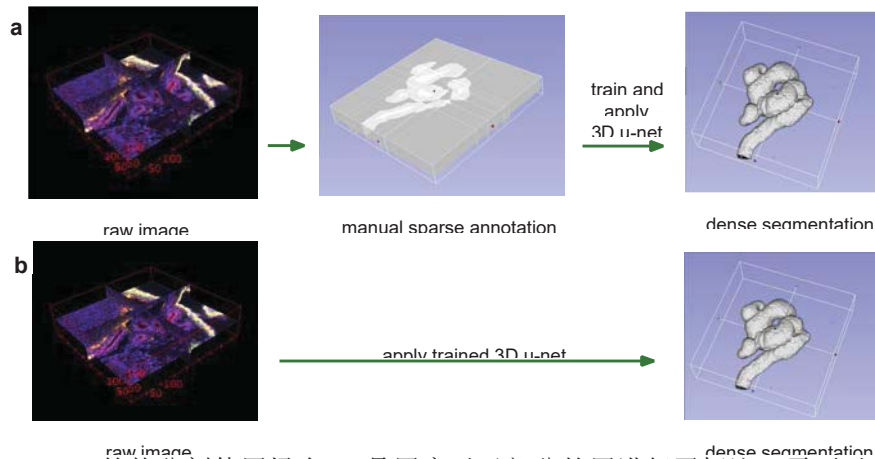
**关键词:** 卷积神经网络, 3D, 生物学体积图像分割, 半自动, 全自动, 稀疏标注

### A. 2.1 引言

体积数据在生物数据分析中是很丰富的。进行这种的分割任务的标注会有很大的困难, 因为只有2D的层会被显示在电脑的屏幕上。因此, 进行逐层的人工标注这种体数据是非常劳累的工作, 同样是效率很低的, 因为邻近的层几乎能够共享所有的信息。这个对于需要训练的方法来说极为重要, 这些大量的完全标注的数据并不是很容易获得的, 但是只有大量的数据才能够让这类模型效果良好。

在这篇论文里, 我们提出了一种深度网络, 它可以学着生成体密集分割, 但

是仅仅需要少量的2D的层标注来提供训练。这个网络能够和图A.2.1一样使用在两种方式：第一种应用仅仅是专注于对于稀疏标注的数据进行密集化分割，第二种是从多重的稀疏标注的数据中学习后重新生成一个新的数据。这两种应用都高度相关。



图A.2.1 3D UNet的体分割使用场合。a是用户对于部分的层进行了标注，需要对于整个体数据进行分割；b是用户对于一个训练集上的数据进行训练后再一个完全无标注的数据上进行分割

网络结构是基于UNet的网络结构，它包含有一个提取的编码部分来分析整个图片，以及一个接下来的延展的解码结构来生成原分辨率的网络分割<sup>[11]</sup>。虽然UNet是一个完完全全的2D网络结构，为这篇论文提出的网络能够以3D的体数据作为输入，并且用对应的3D操作来进行处理，细致来说，包括3D卷积、3D最大池化和3D上卷积层。另外，我们避免了网络结构中的瓶颈结构<sup>[13]</sup>并且使用了批正则化<sup>[4]</sup>以更快收敛。

在很多的生物应用中，仅有少量的图片被使用来训练一个网络并且要求生成一个合理的结果。这是因为每张图片已经组成了代表性的结构，并且有对应的变化。在体数据集中，这个影响被着重提出，因此我们能够在两个提及的图片中训练得到一个网络来生成第三个体分割。一个加权的损失函数和特殊的数据增强让我们能够仅仅在少量的几层数据标注情况下训练得到网络，例如从我们稀疏标注的训练数据。

我们在Xenopus肾脏数据集这一从显微镜获取数据集上成功地实现了我们的方法。在它的发展中，Xenopus肾脏形成了复杂的结构<sup>[7]</sup>，这个会限制预定义的参数的模型方法的使用性能。我们首先提供了一个量化的结果来为这个非常稀疏标

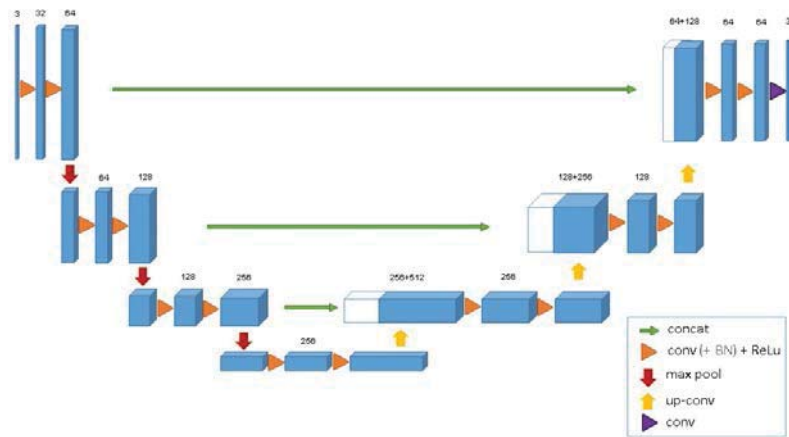
注的数据生成密集的结果。这个结果被量化评估标准所支持。我们同样提供了实验来展示标注的层的数目在我们的网络的效果上会有一些影响。我们的网络使用Caffe框架<sup>[5]</sup>并且进行了开源。

#### A. 2. 1. 1 相关工作

具有挑战性的2D生物医学图像能够通过一个网络实现近似于人类的精确的分割结果<sup>[11,12,3]</sup>。因为这一种成功，之前已经有许多的尝试在试图用2D的卷积方法应用在生物医学体数据上。Milletari等人<sup>[9]</sup>提出了一种结合了Hough投票方法的卷积方法来进行3D分割。然而他们的方法并不是端到端的，并且仅仅在一些气泡形状的分割中是有效果的。Kleesiek等人<sup>[6]</sup>的方法展现了一个少有的端到端的3D卷积神经网络的方法。但是他们的网络不够深，并且仅有一个最大池化层。因此用这个网络很难来进行多参差的结构分析衡量。我们的工作是基于2D UNet<sup>[11]</sup>的结构，这个结构在2015年已经赢得了许多的国际分割和最终比赛。UNet的网络结构和数据增强方法使得我们能够在仅有的很少的标注的样本中训练得到一个具有很好的泛化能力的模型。它利用了刚性变换和弹性变形后，图像依旧是符合生物医学标准的图像这一特性。上卷积结构类似于语义分割<sup>[8]</sup>的全卷积结构，并且UNet依旧没有很广的推广，我们只知道一个类似于用这种3D<sup>[14]</sup>的方式进行的尝试。在这个工作里，Tran等人的结构被用来分析影响并且完全的标注可以被用来进行训练。我们这份论文能够被训练在稀疏标注的体积数据航，并且能够在任意大的数据集上进行使用，因为它的生成方式，是我们的论文的亮点。

#### A. 2. 2 网络结构

图2展示了我们的网络结构。类似于标准的UNet的网络结构，他有一个分析和一个合成部分，每个都有四种分辨率的步骤。在分析路径中，每一层包含了两个 $3 \times 3 \times 3$ 的卷积层，并连接了一个ReLU层。接下来一个 $2 \times 2 \times 2$ 的最大池化层使用了步长为2的情况来对于每一维度进行处理。在合成层，每一层包含了一个 $2 \times 2 \times 2$ 的上卷积结构并且承接了两个 $3 \times 3 \times 3$ 的卷积层和一个ReLU层。旁路分支从每一层的等分辨率的分析层传递给了合成层来提供更加高层的重要信息给合成从。在最后 $1 \times 1 \times 1$ 的卷积被使用来降低输出的通道数，并且限制输出通道数目为3，以符合我们的例子。整个结构一共有19069955个参数。和<sup>[13]</sup>建议的一样，我们避免了瓶颈结构，并且通过将通道数目在最大池化前进行乘两倍来解决。我们同样在合成层中使用了这种方法。



图A.2.2 3D UNet的网络结构。蓝色的盒子代表特征图，通道数目在图中标出

网络的输入是一个 $132 \times 132 \times 116$ 的三维图像，并且有3个通道。我们的输出在x, y, z轴上是 $44 \times 44 \times 28$ 体素。这个体素大小为 $1.76 \times 1.76 \times 2.04$ 立方微米。在这种大小的体素上，我们最终能够得到的分割结果的每个体素的感受野是 $155 \times 155 \times 180$ 立方微米。因此每个输入的体数据能够提供足够的信息来进行细致的学习。

我们同样引入了批正则化在ReLU层前进行操作。在论文4中，每个批中数据都被正则化，使得在训练过程中它的均值和方差以及全局数据对于这些数值进行了实时更新。之后跟着一层来进行度量学习和偏置层。在测试阶段，正则化通过对于全度化数据，偏置，尺度进行计算。但是我们有批尺寸为1和极少的数据。在这个应用中，在测试和训练中使用当前的数值是最有效的。

网络的重要部分，就是网络的允许我们进行稀疏训练的部分是加权的Softmax损失函数。我们设置未被标注的体素的权值为0来使得它仅仅从标注了的数据中进行学习，以此来生成整个立体的体积。

## A. 2. 3 实施细节

### A. 2. 3. 1 数据

我们在Xenopus肾脏胚胎的三个样本中<sup>[10]</sup>进行了实验。其中一个被展示在图A.2.3中（左边），这一组数据时在四组的三通道数据中进行特殊操作生成的。我们通过使用XuvTools<sup>[1]</sup>把不同的快拼接成为更大的体积。第一个通道展现了Tomato Lectin结合到Fluorescein在488纳米的刺激波纹结果。第二个通道展现了DAPI的细胞在405纳米的回应，第三个通道展现的是在564纳米刺激情况下的细胞的反

应,体现出了细胞膜。我们人为在这些体数据中使用Slicer3D<sup>[2]</sup>标注这些垂直的xy, xz和yz层的分割结果。标注的位置是特定的选择的有代表性的部位获取的数据,例如标注层是从3通道数据中均衡得到的那些数据。对于标签0被赋予了另一层结构的含义:小管内; 1: 小管; 2: 背景以及3: 未必标注。所有的没有得到标注的样本都被标注成为“未标注”。我们将我们的实验在最终的原是本变绿上得到了降采样系数为2的分辨率。因此我们实验中使用的数据为 $248 \times 244 \times 64$ ,  $245 \times 244 \times 56$ 和 $246 \times 244 \times 59$ 在xyz三个维度上给1, 2, 3不同的数据。我们对于标注的层在yz, xz和xy上被人为设定为(7, 5, 21), (6, 7, 12)和(4, 5, 10)给三个不同的数据样本。

#### A. 2. 3. 2 训练

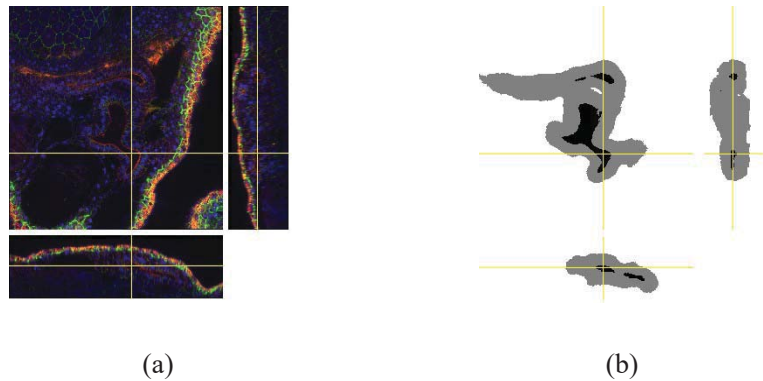
除去旋转,尺寸重设和灰度值增强后,我们在数据和标注中使用了一个圆润化分割的方法进行处理。对于这个,我们从一个方差为4的正则分布在一个32体素的范围内所有的方向中随机采样了一些向量来使用在B-spline插值。网络的输出和标注的标签通过Softmax函数结合一个带系数的交叉熵损失函数进行比较,我们在这里对于常见的背景进行了权值下降,并且对于小管内部的数据进行权值提升,以此来达到训练样本均衡化的效果。我们在训练过程中被标注为3的样本不会对于loss计算有帮助,因为它们权值为0。我们使用了随机梯度下降的方法和Caffe<sup>[5]</sup>的框架来训练整个网络。为了能够在大的3D网络中进行训练我们使用了cuDNN的卷积层的设置。数据增强是当场做的,这个是的训练图片在不同的训练回合是不同的。我们一共在英伟达的TitanX GPU上跑了70000个训练循环,总共耗时3天左右。

#### A. 2. 4 实验

##### A. 2. 4. 1 半自动分割

与自动分割不同,半自动分割只具有少量的标注,网络需要在这些标注上生成更为广阔通用的分割结果。我们的网络并且没有先验分割。我们提出的方法能够让用户进行及其稀疏的标注就能够让网络生成最终的密集分割结果。

作为质量指标衡量,我们在三个稀疏标注的样本中都分别训练了我们的网络,从而进行对比。其中,为了进行更好的结果可视化,对于序号为3的样本里,网络的预测如图3所示。这个玩了国能够发现3D的分割结果在少量标注的层上学习后能够将专家从密集的总体积标注中解放出来。



图A.2.3 (a)是我们的3D的肾脏数据集的图像，b是通过3D UNet生成分割结果

为了评估量化指标，我们归一化了77个人为标注的层并且进行了三折验证并且没有进行批正则化。在这一部分我们把测试的层数都移除了，并且让他们都保持未标注的状态。这个能够刺激出一个新的应用，从而使得使用更为系数的标注就能够生成密集分割结果。为了衡量从全3D内容上标注得到的提升，我们比对了在纯粹的2D图片上的分割得到的结果，那些把所有的标注都当成是单一独立的图片处理的方法。我们将我们的结果呈现在表格A.2.1中。IoU方法被用来对于我们的3D体分割结果进行衡量。从IoU的结果我们可以分析得到，虽然我们能够使用的数据集极少极小，但是在一个额外稀疏的结果上，我们的网络生成最终的分割结果有很高的水平，能够生成得到尺寸大并且位置精确的分割结果。

我们同样在层数的影响上对于网络的效果进行了一定的分析。在这一块中，我们模仿了一个采样的半自动分割方法。我们从每个方向上使用1张被标注的图片开始，并且逐步增加了各个方向使用的图片的张数。我们对于每一个样本都汇报了很高的数据结果提升，虽然仅仅额很少量的额外标注的数据。这些数据从网络中获得，并且带上正则化训练了10个小时。作为测试，我们使用了在本实验中从来没有使用过的那些层的样本。

#### A. 2. 4. 2 全自动分割

全自动分割的设立是假设操作者需要对于一个大的数据的图片进行分割。我们认定了这些可以具有代表性的数据是能够拼合在一起的。

为了预估我们的实验的表现结果，我们在两个部分标注的肾脏体积上训练了我们的模型，并且使用第三个体积进行测试。我们在所有的可能组合的训练测试

上都汇报了我们的结果。表格A.2.3总结了前一部分的IoU结果，并且和完全标注的2D图片进行了对比。在这个实验中批正则化同样对于饰演的结果有一定的提高，除了第三组实验，在哪组数据中效果是相反的。我们认为数据的分布不同是造成结果恶化的主要原因。作为全监督学习，一个最主要的使用时在一个更大的数据样本上，仅有少量的稀疏化标注的样本中依旧能够获得更多的数据来得到更多的又代表的训练数据集。

表A.2.1: 半监督分割的交叉验证结果IoU数值

测试用数据集	3D无批正则	3D有批正则	2D有批正则
子集1	0.822	0.855	0.785
子集2	0.857	0.871	0.820
子集3	0.846	0.863	0.782
平均数值	0.842	0.863	0.796

表A.2.2: 半监督分割标注张数对于结果的影响

标注张数	标注体素	第一个数据IoU值	第二个数据IoU值	第三个数据IoU值
1,1,1	2.5%	0.331	0.483	0.475
2,2,1	3.3%	0.676	0.579	0.738
3,3,2	5.7%	0.761	0.808	0.835
5,5,3	8.9%	0.856	0.849	0.872

表A.2.3: 全监督分割的交叉验证结果IoU数值

测试用数据集	3D无批正则	3D有批正则	2D有批正则
子集1	0.655	0.761	0.619
子集2	0.734	0.798	0.698
子集3	0.779	0.554	0.325
平均数值	0.723	0.704	0.547

## A. 2. 5 结论

我们介绍了一种端到端的方法来进行半自动和全自动的稀疏化标注的3D体分割方法。它能够在一个具有高度不确定性的Xenopus kidney数据集上提供一个准确的分割结果。我们在3折的半自动交叉验证实验上获取了一个高达0.863的IoU结果。在全监督的实验设定上，3D结构的表现比2D结果要更好。这个网络通过从单层进行训练，并且没有为了这个应用而进行特定的优化。我们期望这个能够在

许多其他的生物医学三维分割任务中能够有效。它的设定现在已经被完全开源。

## 参考文献

- [1] Emmenlauer M, Ronneberger O, Ponti A, et al. XuvTools: free, fast and reliable stitching of large 3D datasets[J]. *Journal of microscopy*, 2009, 233(1): 42-60.
- [2] Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network[J]. *Magnetic resonance imaging*, 2012, 30(9): 1323-1341.
- [3] Hariharan B, Arbeláez P, Girshick R, et al. Hypercolumns for object segmentation and fine-grained localization[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015: 447-456.
- [4] Normalization B. Accelerating deep network training by reducing internal covariate shift[J]. *CoRR*.–2015.–Vol. abs/1502.03167.–URL: <http://arxiv.org/abs/1502.03167>, 2015.
- [5] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding[C]//*Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014: 675-678.
- [6] Kleesiek J, Urban G, Hubert A, et al. Deep MRI brain extraction: a 3D convolutional neural network for skull stripping[J]. *NeuroImage*, 2016, 129: 460-469.
- [7] Lienkamp S, Ganner A, Boehlke C, et al. Inversin relays Frizzled-8 signals to promote proximal pronephros development[J]. *Proceedings of the National Academy of Sciences*, 2010, 107(47): 20388-20393.
- [8] 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.
- [9] Milletari F, Ahmadi S A, Kroll C, et al. Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound[J]. *Computer Vision and Image Understanding*, 2017, 164: 92-102.
- [10] Tarin, D.: Normal Table of *Xenopus laevis* (Daudin)[J]. *Journal of Anatomy*, 1968, 103: 578
- [11] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//*International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015: 234-241.
- [12] Seyedhosseini M, Sajjadi M, Tasdizen T. Image segmentation with cascaded hierarchical models and logistic disjunctive normal networks[C]//*Proceedings of the IEEE International Conference on Computer Vision*. 2013: 2168-2175.



- [13] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.
- [14] Tran D, Bourdev L, Fergus R, et al. Deep end2end voxel2voxel prediction[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2016: 17-24.

## 在学期间参加课题研究成果

- [1] Brian Matejek, Daniel Haehn, Haidong Zhu, Donglai Wei, Toufiq Parag, and Hanspeter Pfister, Biologically-Constrained Graphs for Global Connectomics Reconstruction, accepted by CVPR 2019.