**密级： 保密期限：**



**硕士学位论文**



**题目：基于深度学习的冠脉造影图像分割**

**学 号：**

**姓 名：**

**专 业： 信息与通信工程**

**导 师：**

**学 院：信息与通信工程学院**

**2018年11月29日**

独创性（或创新性）声明

本人声明所呈交的论文是本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京邮电大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

本人签名： 日期：

关于论文使用授权的说明

学位论文作者完全了解北京邮电大学有关保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属北京邮电大学。学校有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许学位论文被查阅和借阅；学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。（保密的学位论文在解密后遵守此规定）

保密论文注释：本学位论文属于保密在 年解密后适用本授权书。

非保密论文注释：本学位论文不属于保密范围，适用本授权书。

本人签名： 日期：

导师签名： 日期：

基于深度学习的冠脉造影图像分割

摘 要

近年来，心脏血管疾病已经变成人类致死率最高的疾病之一，它具有着高患病率、高致残率和高死亡率等特点。而在我国心血管病的发病率和死亡率也呈逐步上升趋势，因此我们提出通过计算机视觉与医疗相结合的方式，用更加高效的图像处理方法从血管影像中提取精确的血管结构，减少诊断中的人工交互、降低对医疗人员的依赖，提升疾病诊断的效率，并且为之后基于心脏冠状动脉造影图的医疗诊断系统提供辅助策略，如：钙化检测、狭窄检测等。目前成功运用于冠状动脉造影图像的分割算法并不是很多，表现也不能达到要求，而结合深度学习进行冠脉造影分割任务的相关研究也几乎为零，在此背景下，本论文提出使用深度神经网络的方法对冠脉造影图像进行分割。

论文的主要成果如下：第一，我们将传统分割算法用于冠脉造影图像进行分割实验，成功实现了部分主流的经典分割算法，包括：Otsu算法、GrowCut算法，并根据实验结果分析传统算法在冠脉造影分割任务中的表现以及不足之处；第二，在研究了卷积神经网络和深度学习的相关知识和核心思路之后，我们提出了一个基于深度学习的冠脉造影图像完整分割算法流程，包括图像的预处理、网络的选用、具体结构以及训练和测试方法等，成功实现了经典网络FCN、PSPNet在分割任务中的实验，并根据结果重点分析了卷积神经网络优于传统算法的核心技术点；第三，我们从零开始建立了心脏冠脉造影图像数据库，包括左右冠在内细分七个体位的图像，并在运用PSPNet对各个体位进行实验之后，提出了精确率、召回率和F1 score三个评估指标对实验结果进行分析与讨论。最终的实验结果验证了基于深度学习的PSPNet网络在心脏冠脉造影图像分割任务上的优势和有效性。

关键词：图像分割 心脏冠脉造影 深度学习 卷积神经网络

CORONARY ANGIOGRAPHY SEGMENTATION

BASED ON DEEP LEARNING

ABSTRACT

In recent years, cardiovascular disease has become one of the most lethal diseases in human beings. It has the characteristics of high morbidity, high disability and high mortality. In our country, the morbidity and mortality of cardiovascular diseases are also on the rise. Therefore, we propose that by combining computer technology with medical treatment, we can extract precise vascular structures from vascular images with more efficient image processing methods, and reduce the manual interaction in diagnosis, reduce the dependence on doctor, and improve the efficiency of disease diagnosis. It also provides assistant strategies for subsequent medical diagnosis based on coronary angiography, such as calcification detection, stenosis detection and so on. At present, the performance of traditional image segmentation methods in medical images are not good enough, and also the research of coronary angiography based on deep learning is just a beginning. In this context, we propose a deep neural network method for coronal angiography image segmentation.

The main work of this paper are as the following parts: Firstly we use traditional algorithms to segment coronary angiographic images, and successfully implement some of the mainstream classical segmentation algorithms, including: Otsu algorithm, GrowCut algorithm. Based on the results, we analyze the performance and shortcomings of these algorithms in this task; Secondly after studying the relevant knowledge and core ideas of convolutional neural network and deep learning, we design an algorithm used for coronal angiography image segmentation based on deep learning, including image preprocessing, network selection and specific structure, training and testing methods, we have successfully implemented the experiments of classical network FCN and PSPNet in the task of segmentation. According to the results, we focus on the core technology points of convolutional neural network which is superior to the traditional algorithm. Thirdly we established the database of cardiac coronary angiography from scratch, including the image of left and right coronary arteries subdivided into seven views, In the experiment, we used PSPNet and achieved excellent results in every view of angiography. And proposed three evaluation indicators: accuracy, recall and F1 score to analysis and discuss the results. The results also proved the advantages and effectiveness of PSPNet network based on deep learning in image segmentation tasks of coronary angiography.

KEY WORDS: Image Segmentation; Coronary angiography; Deep Learning; Convolutional Neural Network;

目录

第一章 绪论 1

1.1 研究背景和意义 1

1.2 国内外研究现状 1

1.3 论文主要研究内容 3

1.4 论文组织结构 3

第二章 基于传统算法的目标分割 5

2.1 图像分割的概念 5

2.1.1 传统分割方法的分类和介绍 6

2.2 基于阈值的Otsu分割算法 8

2.2.1 Otsu算法介绍 8

2.2.2 使用Otsu算法进行冠脉造影图像的分割 9

2.2.3 实验结果及分析 10

2.3 基于区域分割算法-种子区域生长、GrowCut算法 10

2.3.1 基于区域生长图像分割算法的原理 10

2.3.2 GrowCut算法 11

2.3.3 使用GrowCut算法进行冠脉造影图像分割 12

2.3.4 实验结果及分析 13

2.4 本章小结 13

第三章 基于深度学习的语义分割 14

3.1 人工神经网络 14

3.2 卷积神经网络 16

3.2.1 卷积神经网络综述 16

3.2.2 卷积神经网络的特点 17

3.2.3 卷积神经网络的结构 18

3.3 基于卷积神经网络的语义分割 21

3.3.1 语义分割简述 21

3.3.2 FCN简述 22

3.4 本章小结 23

第四章 基于PSPNet的冠脉造影图像分割 25

4.1 数据集的建立 25

4.1.1 数据集的分类 25

4.1.2 数据的标注 27

4.1.3 数据的预处理 29

4.2 基于PSPNet的冠脉造影分割 30

4.2.1 FCN的不足 30

4.2.2 金字塔池化 32

4.2.3 网络结构 33

4.3 其他设计 34

4.4 本章小结 35

第五章 实验结果及分析 37

5.1 实验流程 37

5.2 实验环境及训练方法 37

5.2.1 实验环境 37

5.2.2 训练和测试方法 37

5.2.3 目标分割效果的评估方法 38

5.3 实验结果记录和分析 40

5.4 本章小结 45

第六章 总结与展望 46

6.1 论文工作总结 46

6.2 讨论与展望 47

参考文献 49

致谢 53

作者攻读学位期间发表的学术论文目录 54

# 第一章 绪论

## 研究背景和意义

近年来，心脏血管疾病已经变成人类致死率最高的疾病之一，它具有着高患病率、高致残率和高死亡率等特点。在《世界卫生报告》中指出全球每年死于心血管疾病的人数高达1750万，其中冠状动脉引发的疾病及其并发症的死亡率已超过了所有癌症之和，并成为排名第一的死亡病因。在我国，随着人民生活水平的提高，饮食结构的改变和人口老龄化的不断加剧，心血管病的发病率和死亡率也呈现了逐步上升趋势，在每5例死亡中就有2例死于心血管病，其中冠心病约占10-20%。

目前，血管造影技术被认为是血管成像及诊断相关病症的“金标准”，它的原理是：血管造影利用电离辐射，碘化血管内造影剂和X线来成像，造影剂可以根据时间不同而显影，能够观察到侧支循环和供应血管的来源而不包括骨骼和其他组织。血管造影技术也是诊断冠状动脉粥样硬化性心脏病（冠心病）的一种常用而有效的方法，已广泛应用于临床。统计数据显示，约有90%的冠心病诊断数据来自于冠状动脉影像，且近年来我国医学影像的数据仍在以高达30%的增长率逐年增长。然而，我国相应专业医疗人员的数量基数小且增长速度缓慢，无法有效应对庞大的冠心病病人群体及其医学影像数据。因此，亟需开展关于冠状动脉医学影像的图像分析及检测技术研究，提高我国冠心病的诊断及治疗水平。我们希望通过运用计算机的图像处理方法从血管造影图像中提取精确的血管结构，作为疾病辅助诊断的依据，在初期的技术实现中达到辅助医生诊断的目标，并逐步迭代完善，提高冠状动脉影像的计算机分析效率及计算机辅助医疗水平。

## 国内外研究现状

图像分割是从图像处理到图像分析的关键步骤之一，也是进一步进行图像理解的基础，分割结果的好坏直接影响到后续图像分析结果的优劣。通过图像分割可以提取出图像中用户关心的目标并为以后的图像分析提供必要的数据。目前传统的图像分割方法主要分基于阈值的分割方法[1]、基于区域的分割方法、基于边缘的分割方法以及基于特定理论的分割方法等。

阈值分割方法在约40年前就已经被证明提出，现已转化出大量算法。主要适用于目标和背景的灰度差距比较大的图像，其拥有着计算简单、运算效率高、速度快等特点，在重视运算效率的应用场景中得到了广泛采纳。而目前的研究大多集中在优化阈值的确定上。

基于区域的分割方法实际就是把性质相似的像素连接起来，变成最终的分格区域。1990年Meyer提出的分水岭法（Watershed Algorithm）[2]是一种基于拓扑理论学的数学形态的分割算法，对于微弱的边缘响应比较敏锐，但较高的噪声会导致过分分割现象；2002年Gonzalez提出了区域分裂合并法，适用于灰度图像和纹理图像的分割应用。

基于边缘的分割方法利用在边缘图像局部不连续的特点，构造了对图像灰度阶跃变化敏感的差分算子来对图像进行分割，如Robert算子、Sobel算子等。其难点在于边缘检测时抗噪性和检测精度的不确定性。

随着人们对图像处理技术研究的深入和医学技术的不断发展，图像分割技术开始大量应用于医学领域，成为决定医学图像在临床治疗中提供可靠依据的关键。文献[3]中成功将区域生长、阈值分割技术结合起来，应用于分割卵巢细胞质图像；Angelina针对恶性黑色素瘤的诊断，在2012年将遗传算法引入图像分割，解决了利用非侵入性诊断工具对皮肤癌进行早期诊断耗时较长的问题。

自20世纪70年代末起，CNN神经网络开始被用于医学图像的分析。最早是在1995年Lo等人[4]运用四层CNN检测x射线图像中的结节。而后在2012年AlexNet[5]获得ImageNet挑战赛的冠军以来，CNN逐渐成为了医疗图像分析界的主流。最初在医学图像领域中，CNN主要关注在无监督的预训练和网络架构的问题上，如文献[6][7]基于MRI图像对阿尔茨海默氏病患者进行了不同分类，文献[8]对深度学习在AD/MCI的分类中进行了实验。在医学图像的检测方面，Yang等人[9]对三组独立的二维 MRI切片使用常规CNN方法来确定股骨远端表面的标志。在此研究基础上，De Vos等[10]又进一步的在切片的区域中进行了感兴趣区域相关（ROI）的研究与分析。而文献[11][12][13]则使用了将预训练的CNN结构和RBM一同结合的方法，这克服了数据的缺乏从而导致无法得到更好特征表达的问题。当然CNN也可用于视频或者关键帧视频中的物体检测，Baumgartner[14]对视频帧的部分数据进行了CNN训练，通过对多达12个标准化平面扫描来检测孕妇的不同妊娠时期。此外，CNN也可以使用显著图像获得大脑或者脊柱图像中感兴趣物体的大体定位。Kong[15]利用医学视频中包含的时间序列作为另一种类型的高维数据，检测了心脏MRI图像中的舒张帧和收缩帧。而最近很多团队开始使用像素级的分类来进行物体的检测与定位，Hwang和Kim[16]提出了一种弱监督的深度学习方法进行像素级的标注，用以在乳腺X射线图中检测结节。 但是，目前关于深度学习在医学影像方面的应用主要集中在CNN网络及其相关变种[17]，很难同时满足病变特征实时检测并达到一定准确率的要求，所以仍然是较难的课题。而且，目前基于深度学习的医学影像处理技术，尚处于起步阶段，除了对于CNN的研究及应用之外，缺乏对能更加高效应用在医学影像的深度学习网络的研究和拓展。更主要的是，目前对于冠状动脉造影的图像处理技术研究及数据严重匮乏，不同部位医学影像的特征也迥然不同，X射线图、妊娠影像、心脏MRI等方面的医学影像处理技术几乎不可能直接迁移到冠状动脉造影图像中。因此，目前亟需开展基于深度学习的冠状动脉造影的图像处理技术，包括数据的获取、分析、处理及图像内的目标分割等。

## 论文主要研究内容

应基于深度学习的医疗项目发展趋势及企事业单位的应用需求，本论文针对基于真实病例的冠状动脉造影的图像，分别采用了传统目标分割方法和深度学习的方法进行实验研究，目的为利用较为前沿的计算机技术，对所获取的数据进行处理，为医生的疾病辅助诊断提供依据，并逐步迭代完善，提高冠状动脉影像的计算机分析效率及计算机辅助医疗水平。主要成果如下：

1.学习研究传统主流的目标分割算法。首先了解传统目标分割算法的种类和主要思想，然后选取其中经典传统算法进行具体分析，使用不同的工具进行实现、并使用多个体位下真实的冠脉造影图像用于分割实验。基于得出的实验数据，验证传统的分割方法主体思路、计算量、时间复杂度等，分析每种算法的优点、缺点。

2.阐述深度神经网络的算法思想并解释其中最有代表性的卷积神经网络的相关知识。设计一个适用于分割心脏血管的完整系统，包括图像预处理、数据增强、用于冠脉造影的卷积神经网络以及图像的后续输出处理。其次运用深度学习网络PSPNet实现冠状动脉影像的血管分支及病变区域等的识别和分割，通过设定的多个指标衡量方法的准确率、召回率、F1 score等，对不同的体位结果进行分析，并提出改进方向。

3.从零开始建立心脏冠脉造影图的数据库：联系阜外心血管医院搜集整理心脏冠脉造影图像数据，与医生商讨出一套合适的标注标准，设计实现标注工具，并进行双重标注，由医生进行简单的粗标注，研究人员进行二次的精标注，从多个体位对感兴趣的血管分支进行描述，进而做预处理，为本论文实验做准备。

## 论文组织结构

本文的组织结构如下：

第一章：本章为绪论，首先介绍了研究课题的宏观背景和国内外现阶段的研究情况，然后阐述了目标分割的相关概念、人工神经网络的主要思想以及部分算法在医疗领域的应用趋势。

第二章：本章介绍了几种传统的目标分割算法，阐述了算法的具体思想，并尝试使用不同的工具进行实验，根据结果分析算法的优劣。

第三章：本章介绍了深度神经网络的算法思想并解释其中最有代表性的卷积神经网络的相关知识。简要介绍了语义分割的相关概念。

第四章：本章介绍了本论文的冠状动脉造影数据库的建立及实验中所使用的网络模型PSPNet，详细介绍了网络结构中较为核心的部分，并设计、介绍了实验的整体流程。

第五章：本章是对基于PSPNet网络分割实验中得出的测试结果进行分析，对各个体位的结果进行对比，并做出总结。

第六章：论文最后一章针对论文涉及到的理论内容进行了整体的总结梳理，并针对未来的工作与研究方向给出了展望。

# 第二章 基于传统算法的目标分割

## 图像分割的概念

一般情况下，人们使用计算机进行数字图像处理的目的有两种，一种是产生更适合人类视觉观察和识别的图像，二是期望计算机能够自动进行识别和理解图像及其中包含的信息。无论是出于何种目的，图像处理的关键一步是对包含有大量各式各样景物信息的图像进行分解。最终分解的结果就是图像被划分成多个具有各类特征的最小组成，这些组成就称为图像的基元。产生这些基元的整个过程就是图像分割。图像分割作为图像处理领域中极为重要的内容之一，是实现计算机对图像分析与理解的基础。从概念上来说，所谓图像分割就是按照一定的原则将一幅图像或景物分为若干个部分或子集的过程。我们目前在图像处理系统中只能得到二维图像信息，还不能得到三维信息，因此只能进行图像分割而不是景物分割，因为景物是三维信息；图像分割也可以理解成把图像中有意义的特征区域或者需要应用的特征区域提取出来，这些特征区域可以是像素的灰度值、各类型像素的区域、物体轮廓曲线、纹理特性等，也可以是图像的空间频谱或直方图特征等。在图像中用来表示特定一类物体的区域，其特征都是大体相近或相同的，但是不同物体的区域之间，特征就不会大体相同，会出现急剧变化。

现阶段已经提出了很多种的图像分割方法，从分割依据的角度来看，图像的分割方法可以分为相似性分割和非连续性分割。相似性分割的思路是将具有同一灰度级或相同组织结构的像素聚集在一起，从而形成图像的不同区域；非连续性分割就是首先检测局部不连续性，然后将它们连接在一起形成边界，这些边界能够把图像分成不同的区域。基于不同种类的图像，拥有不同的应用场景，需要提取的图像特征并不相同，对应的图像特征提取方法也就不同，因此并不存在一种普遍适应的最优方法。从结构上说，图像分割方法又可分为结构分割方法和非结构分割方法两大类。结构分割方法的思路是：根据图像局部区域的像素特征来进行图像分割，如阈值的分割、区域生长、边缘检测、纹理分析等方法，这些方法假设了我们提前知道这些区域的不同特性，或者在算法过程中能够得到这些特性，进而一定能够找到各种形态像素群并进行研究。而非结构分割法分为：统计模式识别、神经网络方法或其它利用景物的先验知识实现的方法等等。目标分割的任务属于图像分割的一种，研究人员通常只对图像中某些区域感兴趣。我们在实验中也只关注其中的一部分区域或目标点，称之为：感兴趣区域，而其他区域称为背景区域[18]。针对一般的光学图像来说，对于像素的分割是一种比较多见的任务目标，即：选择一部分像素用以表述已知目标。这种分割可以当作是一个分类问题，即针对每一个像素赋予一个标签label，并在其中提取出感兴趣的那一类label的像素；也可以当作是聚类问题，即不知道标签label，但是需要满足一些最优解条件，如类间相似度最小等。

对于心脏冠脉造影图像的分割来说，既是一个广义的图像分割问题，同时更应该算是一个二分类的目标分割问题：对于区域中的每一个像素点，我们通过算法实验判断其是属于血管（目标、正样例）还是属于非血管（背景、负样例），二分类实验如下图2.1

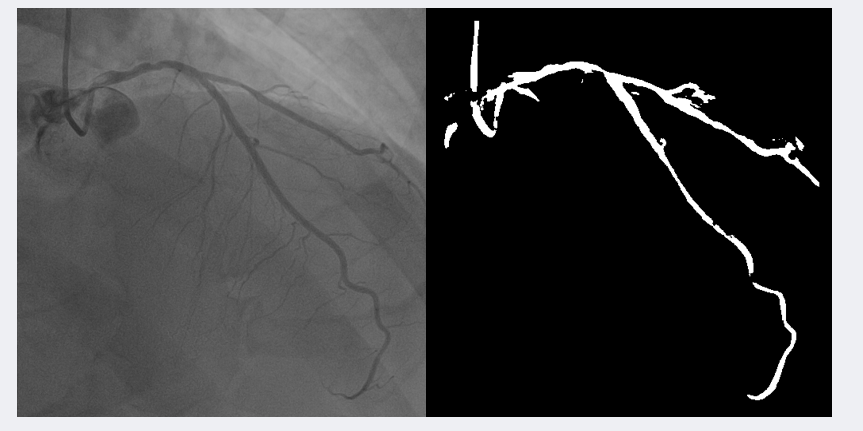


图2.1 二分类血管分割示意图

### 传统分割方法的分类和介绍

现在传统图像分割的方法主要有基于阈值的分割方法、基于区域的分割方法、基于边缘的分割方法以及基于特定理论的分割方法等多种。而随着研究的进步，分割的算法又被分成基于像素的分类、阈值的分割、边缘的检测、彩色图像分割、深度图像分割和基于模糊集的方法等。在这里我们先对前一种阈值分割法进行分析、讨论。

 阈值分割法是一种基于区域的图像分割算法，其基本思想是用图像的特征来计算一个或多个特定阈值，并将图像中每个像素的特征值与阈值相比较，然后将每个像素根据比较结果分到其相对应的合适类别中：

 式（2.1）

鉴于实际得到的图像目标和背景有可能不一定分布在两个特征值范围内，公式可以变形为：

 式（2.2）

常用的特征值可以是原始图像的灰度或彩色特征值，也可以由原始灰度或彩色值变换得到的特征[19]。阈值分割法实际就是假设一个准则函数，然后利用它求出最优阈值解的过程，经常使用的阈值选取办法有最大类间的方差法、最大熵自动阈值法、灰度直方图法等。

而基于区域的分割方法可以分为区域生长和分裂合并两大类。区域生长算法的基础思路是人为的把图像分成n个相似的区域，然后将某些已知的区域根据一种判断准则逐步地对相邻区域进行连接合并。区域分裂算法刚好与之相反，没有人为划分区域的过程，首先将输入的图像当做一个完整的区域，然后开始分裂，最后对相似区域进行合并。这两种算法都是串行的算法，即：分割的过程是按照一定顺序的多个步骤，后面的操作是基于前一步操作的结果然后继续进行的。

鉴于目标和背景是按照边缘划分分界，因此在基于边缘的分割算法中最重要的操作是边缘检测算法。边缘检测是对比目标和背景在灰度、颜色、纹理等特征上存在的不同来实现检测。在边缘区域两侧的像素点一般有着较大的微分值，而且因为微分计算是带有方向性运算：当微分的方向和边缘垂直时微分的值最大。所以在实际应用中我们一般利用差分运算代替微分：

 式（2.3）

 式（2.4）

   式（2.5）

其中是像素点的特征值，和分别代表特征值在两个方向上的变化速率，表示特征值在方向上的变化速率。目前常见的差分算子有梯度算子、拉普拉斯算子等。

阈值分割算法的优点是计算简单、运算效率高、速度快。在重视运算效率的应用场合(如用于硬件实现)，它已经得到了广泛应用。目前，图像的阈值分割也已被应用于很多的领域，例如，在红外视觉技术应用场景中，红外无损检测[的红外](https://baike.baidu.com/item/%E4%B8%AD%E7%BA%A2%E5%A4%96/5133220)热图像分割，[红外成像](https://baike.baidu.com/item/%E7%BA%A2%E5%A4%96%E6%88%90%E5%83%8F/5264257)跟踪系统的目标分割；在[遥感控制应用](https://baike.baidu.com/item/%E9%81%A5%E6%84%9F%E5%BA%94%E7%94%A8/5133027)中，用来[合成孔径雷达](https://baike.baidu.com/item/%E5%90%88%E6%88%90%E5%AD%94%E5%BE%84%E9%9B%B7%E8%BE%BE/1449279)图像的目标分割等；在农业项目的实际应用中，水果品质无损检测过程中的水果图像与背景的分割。在工业生产应用中，机器视觉运用于产品质量检测等；并且在医学应用中，血液的细胞图像分割，磁共振中图像的分割也都已经使用了阈值分割算法。

## 基于阈值的Otsu分割算法

### Otsu算法介绍

Otsu算法[20]又被称为最大类方差算法、大津算法，是一种图像回溯自适应的阈值分割算法，是日本学者大津(Nobuyuki Otsu)在1979年提出并命名的。大津法是按照图像上的灰度值分布，将图像分成背景和前景（目标）两部分。背景和目标之间的类间方差越大，说明构成图像的两部分像素的差别越大，当部分目标错分为背景或背景错分为目标都会导致两部分的类间方差变小。因此，使类间方差最大的分割意味着错分概率最小。因为计算简单，不受图像亮度和对比度的影响，Otsu算法被认为是图像分割中阈值选取的最佳算法。

Otsu算法首先从图像的灰度直方图中得到各分割特征的发生概率，并以阈值变量t将分割特征值分为两类，然后求出每一类的类内和类间方差，选取最佳阈值t使得类间方差最大、类内方差最小。具体算法流程如下：

首先设图像的原始灰度级为M，灰度级为i的像素点个数为，灰度值i的像素数为，对灰度直方图进行归一化，得到

 式（2.6）

对总的两类像素和，每一类出现的概率分别为：

 式（2.7）

 式（2.8）

每一类的平均灰度值为

 式（2.9）

 式（2.10）

其中：

 式（2.11）

 式（2.12）

式（2.11）是整体图像的灰度平均值；式（2.12）是阈值为时的平均灰度值，所以全部采样的灰度平均值为

 式（2.13）

那么两组间的方差为

 式（2.14）

Otsu算法，就是遍历灰度级从0到m-1，直到找到某个灰度级使得为最大，那么该灰度级就是最佳分割阈值。

Otsu算法有着较强的抗噪性能、且处理速度较快的特点。

### 2.2.2 使用Otsu算法进行冠脉造影图像的分割

图2.2原始数据 图2.3实验分割结果

Otsu算法为全自动算法，实验中我们输入的是原始图像2.2，图2.3是计算机输出的实验分割结果。

### 2.2.3 实验结果及分析

从实验结果中我们可以得出Otsu算法对于灰度范围不算很广的心脏冠脉造影图像来说，虽然它对类间方差为单峰的图像有着较好的分割效果，且计算效率比较高，不需要考虑时间复杂度等问题，但是Otsu算法对噪声和目标的大小较为敏感，心脏造影图像的噪声干扰非常高，仅利用灰度直方图得到的阈值并不能使实验分割得到满意的结果。另外，对于冠脉造影图像来说，图像中的背景和目标的灰度值经常会较为接近，类间方差准则函数图像大概率是双峰或多峰，便可能会产生比较明显的误判，因此造成实验效果并不是很好。

## 基于区域分割算法-种子区域生长、GrowCut算法

### 2.3.1 基于区域生长图像分割算法的原理

图像分割算法一般是基于灰度值的两个基本特性之一：不连续性和相似性。前一种性质的应用途径是基于图像灰度的不连续变化分割图像，比如图像的边缘。第二种性质的主要应用途径是依据实验指定的准则将图像分割为相似的区域。区域生长算法就是基于图像的第二种性质，即图像灰度值的相似性。种子区域生长算法的基本思想是将具有相似性的像素集合起来构成区域。首先对每个需要分割的区域找出一个种子像素作为生长的起点，然后将种子像素周围邻域中与种子有相同或相似性质的像素（根据事先确定的生长或相似准则来确定）合并到种子像素所在的区域中[21]。而新的像素继续作为种子向四周生长，直到再没有满足条件的像素可以包括进来，这样一个区域就生长而成了。

种子区域生长算法的设计主要有以下三个关键点：生长种子点的确定，区域生长的条件，区域生长停止的条件：

1.种子点的个数根据不同的任务目标可以选择一个或者多个，并且根据具体的问题不同可以采用自动确定或者人为设定。

2.区域生长的条件实际上就是根据像素灰度间的连续性，进行定义的一些相似性原则。

3.区域生长停止的条件则是定义了一个停止规则，我们设定，在没有像素满足加入某个区域的条件的时候，区域生长就会停止。

在具体实现过程中，基本流程如下：

1.首先选取种子点p（x0，y0），用堆栈来表示种子区域，将种子点push到种子堆栈中。

2.将种子堆栈中第一个种子点pop出堆栈，并以该点为中心，遍历该中心距离为8的邻域像素。

3. 判断遍历像素点是否已经在种子区域中，如果否，判断遍历像素点是否满足相邻种子点相似性，如果像素点（x，y）满足相似性，将（x，y）push到堆栈中。

4.重复步骤 2-3，直到种子堆栈为空为止。

### 2.3.2 GrowCut算法

GrowCut算法[22]继承了种子区域生长算法思想，以细胞自动机（Cellular Automaton，CA）作为分割框架，将类别信息转化为细胞的状态，依据位置、颜色等特征，把位置相邻与颜色相近的像素点改变演化至同一种类别。

我们把一幅图像作为一张m\*n的网格，图像中的每个像素点与细胞自动机空间中的细胞一一对应，这些细胞被分为前景点、背景点、不确定点或其他类别，随着算法演进，细胞们相互竞争来占领图片，而占据图片的能力和该像素强度有关。

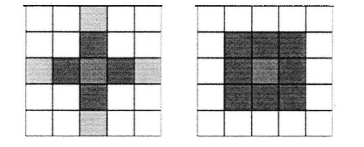
GrowCut算法的主要思想是计算当前的种子细胞状态及其邻域之间细胞的特征值，以判断它们之间的关联强度从而来决定下一时刻的状态，通过生长和反复竞争来改变细胞状态，直至完成分割。细胞的状态修改通过领域模型迭代完成，常见的领域系统有冯诺依曼领域和摩尔领域（Moore）两种。

图2.4 冯诺依曼领域和摩尔领域

在GrowCut算法中我们采用Moore型邻域实现迭代：

 式（2.15）

公式定义为：对于中心细胞来说，其相邻半径为R的区域内的细胞，为该细胞的邻居，、表示邻居细胞的行列坐标值，表示中心细胞的行列坐标值，因此当相邻半径R=1,2,3时，中心细胞分别含有8、24、48个邻居，分别对应Moore、Moore、Moore邻域。

在算法进行时，我们需要人工标记目标和背景的初始种子以及它们的标签。一张图像中细胞的初始状态为：

，， 式（2.16）

其中为像素的标记，为像素属于当前判定的标签的强度，是像素的特征向量，是像素在图像中的三通道色彩向量值。然后种子细胞开始攻击其临近细胞。如果防守细胞的防守强度高于攻击细胞的攻击强度则攻击失败。反之则攻击成功，当攻击成功时防守细胞才会改变标签，被攻击细胞占领。如此反复迭代直到所有细胞的状态都不再发生变化时算法达到收敛，即算法结束。

### 2.3.3 使用GrowCut算法进行冠脉造影图像分割

图2.4原始数据 图2.5 实验分割结果

GrowCut算法需要人工交互。我们首先手动在造影图像血管和背景的关键部位分别用白色和黑色像素点做一些种子（图2.4中红色选框），没有标注的像素点默认为不确定的点。算法从红色选框中的种子开始迭代，最后输出为图2.5。

此外，对于同一张图片，选取不同的位置种子点，实验得到的结果也并不不同。在原始图2.6中，随机选取两组，每组各20个种子点，即得到不同的分割结果如图2.7。

图2.6原始数据 图2.7 不同实验的分割结果

### 2.3.4 实验结果及分析

Growcut算法有着以下几种优点：1.具有良好的交互性；2.Growcut算法能够降低部分噪声和模糊对图像的影响，鲁棒性较好；3.算法的扩展性较好；4.算法本身并不复杂，易于实现。但是同理算法有着一定的缺陷：首先算法需要人工设定种子点，不同的种子点的数目和位置设定对最终的结果影响巨大，使得算法并不客观；其次对于区域面积较大的图片进行分割时算法计算缓慢，时间复杂度较高，不同的种子和区域需要多次的计算和改变状态，从而造成实验中单张造影图片的分割有着较大的延迟；最后，如果待分割的目标结构比较复杂，例如造影图像中细小的血管、或者图像本身噪声较高，便会极大降低分割结果的准确度。

## 本章小结

在本章节中我们先梳理了图像分割以及目标分割的部分知识，然后挑选了其中两种比较有代表性的传统目标分割算法进行研究分析，并使用源于真实病例的心脏冠脉造影图像进行分割实验。实验证明，虽然传统的分割方法思路简洁，计算量小、计算速度快，且每种算法都有其独特的优点，但算法也有着较为明显的缺点。这些传统的目标分割算法存在的缺陷主要体现在以下几个方面：

1.对于某种特定图像提出的传统分割算法并不适用于其他类型的图像。

2.造影图像中血管的分割一般是依据像素点的灰度特征进行提取，在一定范围区域内目标和背景的灰度值差别并不是很大；而另外一些区域血流速度低且流向复杂，导致梯度值不够高。

3.心脏造影图像中的噪声普遍较高，对于传统算法实验的影响较大。

4.例如Growcut算法等是属于需要人工交互式算法，人工选取的起始点对于实验结果影响较大，导致结果并不是很客观。

5.上述部分分割算法在对于心脏造影图的分割任务中计算量较大，算法复杂度较高，导致实验有一定的延迟，很难达到实时。

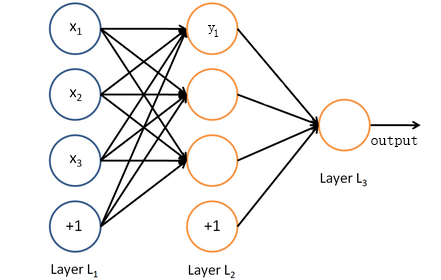
# 第三章 基于深度学习的语义分割

图像分割一直是计算机视觉的关键问题。而本文的目标是对冠状动脉造影图进行分割，这是图像处理在应用下的语义分割任务，方法采用的是人工神经网络的深度学习模型——卷积神经网络。得益于深度学习和图像处理技术的发展与进步，出现了很多高精准度的语义分割的论文与工程可以供于本论文学习与参考。因此本章节先介绍人工神经网络、卷积神经网络等，最后对语义分割进行简单的介绍。

## 人工神经网络

人工神经网络（Artificial Neural Network，即ANN ），是20世纪80 年代以来兴起的人工智能领域中的研究热点。它模拟人脑神经元的网络并对其进行抽象化，进而建立一种简单模型，按不同的连接方式组成不同的网络。在工程与学术界也常直接简称为神经网络或类神经网络。神经网络是一种运算模型，是由大量的节点（或称神经元）之间相互连接构成。每个节点代表一种特定的输出函数，定义为激励函数（activation function）。在每两个节点之间的连接都表示对于通过该连接信号的加权值，定义为权重，这类似于人工神经网络的记忆。网络的输出则依照网络的连接方式、权重值和激励函数的多样而不同。对于网络本身，一般都是对自然界某种算法或者函数的逼近，也可以是对一种逻辑策略的表达。20世纪50年代末期，Rosenblatt首先提出了单层感知器模型[23]，即：拥有输入层、输出层和一个隐含层。输入的特征向量通过隐含层变换到达输出层，在输出层得到分类结果。但是单层感知器模型有个缺陷，它对较为复杂的函数无法处理，甚至都不能进行抑或运算的拟合。随着数学的进步，多层感知机的发明才将这个问题克服。1986年，Rumelhart等人[24]提出了经典的BP算法（Back Propagation），BP算法的优点在于只利用一次前向传播就可以得到所有的偏导数，被人们看做是一种经典高效的算法。2006年，Hinton等人[25]利用预训练的办法克服了局部最优解的问题，将网络的隐含层提升到了7层，自此神经网络才拥有了它的深度。

人工神经网络有着可以自适应的学习的优点，经研究人员发现，在人脑中，神经元传递信息时主要依靠的部位是突触，因此在建立人工神经网络的模型时，研究人员参照人脑中的突触设计了很多方法来模拟，比较具有代表性的是调整权重。人工神经网络依照学习方式的不同，可以分为监督学习和非监督学习。本次研究使用的是监督学习的方法，这种学习方法的方式是，将输入数据通过神经网络一层一层的学习权重参数、最后在输出层根据与正确结果的比较将误差反向传播回网络，网络通过误差值调整权重，从而自适应学习。这种学习方式与非监督学习是不同的，非监督学习方法是根据类别位置的训练样本来解决模式识别中的各种问题。最具代表性的是聚类方法，该方法的规则是在一堆数据中寻找一种“自然分组”，我们希望同组（类别）的样本较为相似，而不同组的样本之间有着明显的不同。

监督学习，它是对人脑神经元进行抽象化建立的模型，每个神经元对应网络里的一个节点，所有节点通过不同的连接方式相互连接，节点连接都有一个加权值，即为权重（weight），权重是网络需要学习和记忆的值，然后通过激励函数（activation function）计算得到每个节点的一个输出值，如图3.1。Layer1层为输入层，有4个神经元也即节点，,,为输入数据，+1为偏置节点b；Layer2层为隐藏层，可以有一层或者多层，层和层的每个连接线都有一个权重Wij（i为输入层节点的序号，j为隐藏层节点的序号），层每个节点的激励值（输出值）为，所有与之相连接的层节点值对权重求和然后再通过激励函数的值，即yj为：

式（3.1）

图3.1神经网络模型示意图

从上可以得出，人工神经网络由节点数、层数、连接方式等元素组成了其结构。这种网络结构可以提取图像、文字、语音等数据中包含的信息或特征，然后通过计算机自行学习数据之间的差异，进而在实验中的实现：检测、分割等不同的任务目标。

考虑到人体的生理组织和我们能获得的真实医疗病例图片的稀缺性和复杂性，以及疾病自身多变性和个体差异性，我们通常需要处理海量的数据，因此也较为适合人工神经网络的应用。所以，近几年人工神经网络与医疗紧密结合，在心血管病研究方面也取得了很多成果，美国现近已经有用ANN实现的心电图检测设备，命名为AliveCor，该设备可以通过检测患者心电图并结合患者自身年龄等其他医学指标来初步预测测试者得心血管病的可能性大小。

## 卷积神经网络

卷积神经网络属于人工神经网络的一种，它是一种典型的前馈神经网络，可以用图像直接作为网络的输入，避免了传统机器学习算法中的特征提取和数据重建步骤，对于大型图像或者多维图像处理有出色的表现[26]。同时，卷积网络的基础是一个多层感知器，最初是为识别二维形状而特殊设计的，这种网络结构对平移、比例缩放、倾斜或者其他的变形具有高度不变性。

### 3.2.1 卷积神经网络综述

卷积神经网络（Convolutional Neural Network, CNN）是一种带有卷积结构的前馈型网络，1959年，Hubel等人发表了对猫视觉神经的研究文章[27]，人们便开始对研究通过神经网络处理图像、视频等数据，之后出现的神经认知机[28]，实现了对特征的初步处理和识别。后来的研究人员提出了一种新的人工神经网络模型，核心思想是：多层感知器（Multilayer Perceptron），这种网络包括至少一个隐藏层，解决了单层感知器只能学习线性函数而不能学习非线性函数的问题，而且运用了随机梯度下降法训练网络，实现了机器代替人工提取特征的设计。

20世纪80年代，LeCun等人尝试在手写数字识别任务中，提出在神经网络中加入BP算法和下采样（Udersampling）与权值共享（Weigth shared）的办法提供了新的思路，于是在实验之后，他们提出了改进优化过后的LeNet-5[29]，LeNet-5经过在MINST数据集上的训练和测试，实验正确率达到了惊人的97%，并且该网络的结构保持了一定的平移、尺度不变形。Lecun等人实验的成功，让研究人员对卷积神经网络有了新的认识，但是当时计算机硬件计算能力的限制，CNN并没有掀起研究的热潮，沉寂多年之后，随着计算机性能的发展，GPU的出现大大提升了计算机对图像处理的效率，并且在2012年的ImageNet比赛上，Alex Krizhevsky[5]等人提出的AlexNet网络，在对数据集中多达1000类的图片进行分类的任务中，实现了突破，Top5的错误率比上一年的冠军降低了10个百分点。AlexNet采用了GPU来进行训练处理百万数量级的ImageNet数据集。它包括5个卷积层和3个全连接层，并使用了全新的ReLu（Rectified linear unit）非线性激活函数，其效果优于传统的tanh函数，后来的很多新的卷积网络模型都采用了Relu函数，AlexNet还引入了防止过拟合的Dropout层：减小神经元之间的共适应关系以及数据增强等方法。2013年，Matthew设计出了ZFNet[30]，ZFNet相较于AlexNet中11\*11的卷积核，其采用7\*7的卷积核用来保留更多原始信息。2014年的VGG网络[31]拥有20层的网络结构，推动了卷积神经网络往更深层发展，它的结构连接非常简单，只是基于AlexNet做了简单改变，但是这一次的卷积核再次缩小到3\*3，分类的错误率降到7.3%左右。之后的2015年，出现了同为24层网络的MSRANet和谷歌的GoogleNet[32]，而后者取得了当年的ImageNet冠军，其为卷积模块引入了Inception的串联结构，即一种网络中的小网络，这种结构并行计算卷积，可以融合更多的特征信息。2016年，何凯明等人提出的恒等映射深度残差网络是网络层数加深、卷积模块增强的结合，网络通过去掉ReLU层，残差项跨层可加，损失层的梯度可以直接传到任意一层，使得1001层网络的训练成为可能。此外，Faster-RCNN[33]、FCN[34]、LSTM[35]等网络的提出为卷积神经网络增加了目标检测、多标签输入输出、时序输入输出等新功能。

上述的网络是几种用于分类的卷积神经网络，它们为后来的研究人员提供了在层数和创新结构上的不同思路。之后很多图像处理的任务，如物体检测、目标识别以及语义分割，大多都是基于这些新型网络进行参数微调、网络结构优化等，然后与实际问题相结合进行的。

### 3.2.2 卷积神经网络的特点

在图像处理中，图像一般被认为是像素的向量：一张大小的图片用一个大小的向量表示。我们假设一个普通神经网络的隐含层数目和输入层相同，那么这个神经网络的参数就有多达个，几乎无法训练。因此卷积神经网络与传统的网络相比，最大的特点就是可以减少参数数量。

首先，卷积神经网络中加入了局部感知野。在人类视觉皮层的神经元就是局部进行信息的接收，神经元只响应某些特定区域的刺激。受此神经学中的视觉结构的启发，卷积神经网络中的神经元也只对局部进行感知和接受，而并非对全局图像的感知，然后通过多层感知，最终将获取的特征信息结合起来就是全图的特征。这一步对局部图像的感知就是卷积操作。

其次，为了继续减少参数，卷积神经网络对于卷积的操作引入了权值共享的办法。其根据是图像局部区域的统计特性在各部分是大体相同的，即图片的底层特征与特征在图片中的位置是并不相关的。比如边缘特征，无论是在图片上方或是下方的边缘特征，还是在图片中间区域的边缘特征，都可以用微分特征提取器提取。即通俗来讲，当在一个尺寸较大的图像中随机选取一块，网络从这个小块样本中学习到了某一些特征，我们就可以把这些特征应用到这个图像的任意区域中去。特别的，我们可以用这些特征对原本的图像进行卷积计算，进行遍历，从而对这个图像上的任一位置都获得一个不同特征的激活值。网络还可以使用多个卷积核作为滤波器对图像进行卷积，来提取图像的更多特征，每种滤波器都有其不同的参数，代表它可以提出图像的不同特征，这样每种滤波器对卷积图像就得到对图像的不同特征的映射，从而生成多种特征图。

基于上文的叙述，卷积神经网络具有许多良好性能，但也包括如下的一些约束：

1.特征提取。每一层神经元都是从上一层得到输入，然后提取到局部特征。在获得被提取的局部特征之后，特征所代表的原始位置信息将成为冗余被损失。

2.特征映射。卷积神经网络的每一层都是多个特征映射组合而成，神经元在约束下共享着相同的卷积核的权值参数。

3.子抽样。为了降低特征映射的分辨率，进一步降低网络对于各种变形的敏感度，每一个卷积层之后一般都接一个实现局部平均或者子抽样的计算层。

基于以上这些特性，神经网络对平移变换、比例缩放、倾斜形式的变形具有高度不变性质，因此能够在图像处理领域解决和处理很多复杂的问题。

### 3.2.3 卷积神经网络的结构

卷积神经网络是一个由多层组成的神经网络，每层网络包含多个二维平面，每个平面由多个独立神经元组成。总体来说，卷积神经网络的基本结构组成都是大体类似的，包含：卷积层（Convolutional layer）、池化层（Pooling layer）、全连接层（Fully-connected layer）、上采样（Up sampling）、激活层（激活函数activation function）、BN层（Batch Normalization）等，在一个网络结构中，前两种层普遍存在，其余结构视情况增减。

1.卷积层

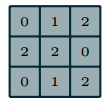
卷积层的主要作用是对输入的数据进行卷积操作来得到数据的特征。卷积，是对输入做一种线性变换，这种变换是稀疏的，输出并非由所有输入数据得到，并且它可以多次利用参数的性质，即：同一个卷积核权重可以计算所有位置的数据，等同于同一张图不同坐标位置的相同类别的物体所提取的特征是几乎一样的，这就是参数共享（parameters shared）的概念。对于卷积核来说，一个卷积核只能提取一种特征，因此在卷积神经网络中都采用多个卷积核卷积的操作来提取图片不同的特征，卷积核大小一般相同。对输入做卷积的卷积核（kernal size）以及其他参数如步长（strides）、零边框（zero padding）等相互作用决定卷积层输出的大小。

图3.2 一个3\*3卷积核

卷积是一种线性变换，以二维图片为例子，假设图片的大小为h\*w，高为h、宽为w，图片通道数为RGB三个。卷积核大小设为k\*k（常见的卷积核都为方阵），卷积核与重叠部分的图片做线性变换输出一个值并记录，然后卷积核按照步长大小平移，计算新的输出，直到遍历此图为止。例如h=5,w=5,k=3，卷积核数值如式3.2，卷积过程如图3.3所示：假设卷积过程是从图片左上角开始，令卷积核每个值为 *(0≤i≤9)，*图3-3中与卷积核重叠的值为 *(0≤i≤9)*，则一个输出为

式（3.2）

可以看出，每次卷积产生一个值（图3.3中第二、四、六列中每个颜色加深的值），然后卷积核在图片上向右平移1格，计算下一个值，直到遍历此图。所以，当图片尺寸很大、数量很多时，卷积的计算量便十分庞大，硬件的计算能力并没有跟上，这也是90年代限制卷积神经网络发展的原因之一。

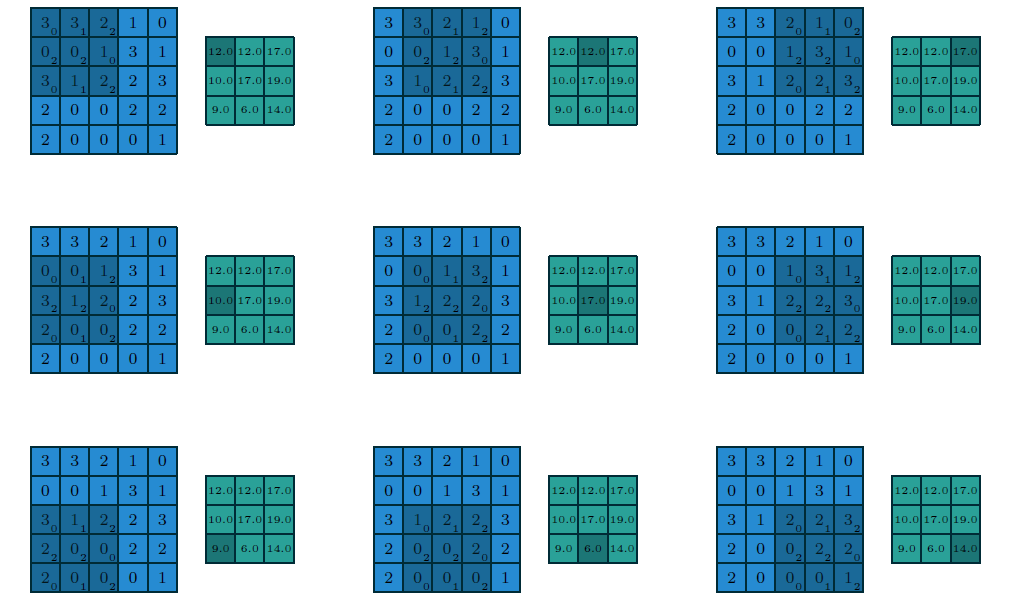
上一个例子中，卷积核每次能恰好平移覆盖整张图，但是当padding=2时，图片大小和卷积核大小并不能满足完整覆盖，所以要在图片四周加零，即令padding=1，卷积过程如图3.4。

图3.3 卷积过程示意图(stride=1，padding=0)

记strides为s, padding为p，那么一次卷积后输出图片高为（宽同理）：

h’=（h-k+2p）/s+1 式（3.3）

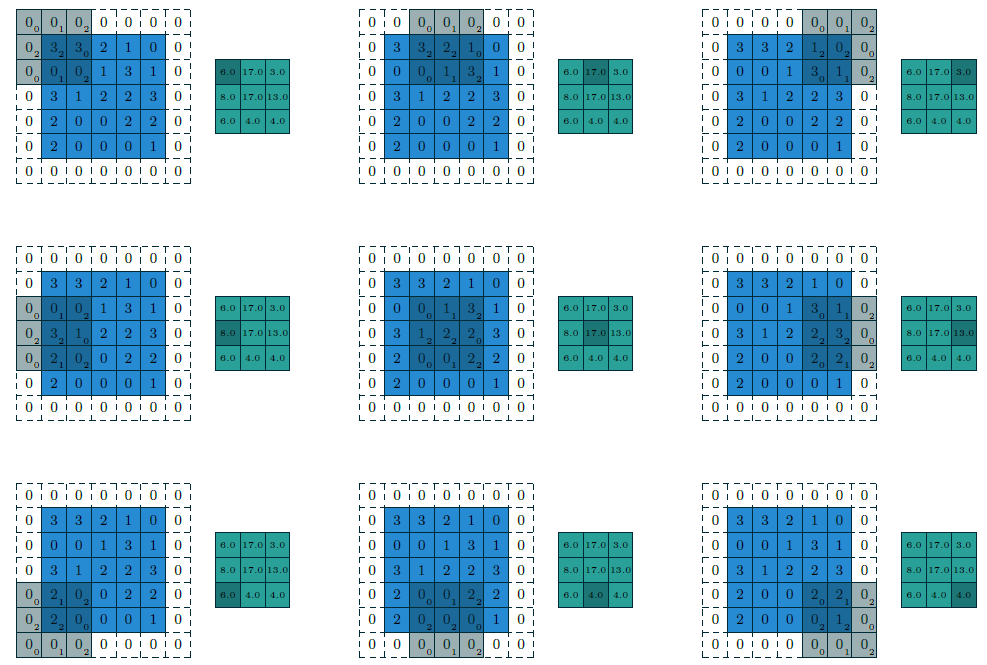
一张图经过卷积层处理时可以用N个（N≥1）卷积核随机地提取特征，所以每个有N个卷积核的卷积层将输出通道数为N的特征面（feature map）。

图3.4 卷积过程示意图(stride=2，padding=1)

2.池化层

池化层，又称下采样层，是根据图片的“静态性”设计的，它主要的功能是对输入的特征图像进行压缩，一方面使特征图变小，简化网络的计算复杂度；一方面进行特征压缩，从而提取主要的特征。池化层的输入是上一个卷积层的feature map，在不改变通道数的前提下，池化操作通过滑动窗的操作用一些函数对feature map的子区域求和，其计算方式与卷积类同，只是将卷积时的线性求和换成其他函数，常用的有最大池化（max pooling）、平均池化（average pooling）和金字塔池化（SPP, Spatial Pyramid Pooling）等。本次研究使用的网络结构中池化层就是SPP。但是池化层的操作中使得特征图缩小，就有可能影响网络的准确度，因此之后的网络中有使用增加特征图的深度来弥补。

3.全连接层

全连接层（FC, Fully-connected layer）是连接所有的特征[36]，将输出值送给分类器（如softmax分类器等）。在AlexNet中最后三层是全连接层，整合信息后传递给逻辑回归（softmax）层，实验输出的是属于1000类的各概率值。不同于卷积层参数量较小、计算率较大的特点，全连接层的参数过多，且大多数没有得到充分利用从而有着不小的计算冗余。之前还有很多网络直接使用全连接层充当网络最后的“分类器”，近几年的网络中全连接层几乎都被别的结构替代。

在AlexNet中还提出Dropout操作来防止参数过多和过拟合，即以50%的概率随机将隐层神经元的输出值暂时置0，使这些节点不再参加后续计算，但是对于每一个minibatch，被丢弃的神经元都是随机，所以网络实际上是学习不同的参数。

4.激活函数

激活函数将非线性特征引入到神经网络中，使得网络不仅可以学习线性的特征，也可以学习不同隐藏层之间的非线性的特征。在AlexNet论文发表之前，研究者们采用过Sigmoid、Logistic函数以及Tanh（双曲正切函数）来激活神经元，但Krizhevsky等人发现并引入了ReLU（Rectified linear units线性修正单元），它的收敛性相较于Tanh函数提升了6倍，只要R(x)=max(0,x)，即若x<0，则R(x)=0，若x>0，则R(x)=x，它很好的避免和纠正了梯度消失的问题。后来常用激活函数还有: Leakly ReLU、Parametric ReLU、Randomized ReLU、ELU等。

5.BN层

BN层batch normalization，顾名思义，即批规范化，通过规范化操作将输出信号规范以保证网络的稳定性。在神经网络训练时遇到收敛速度很慢，或梯度爆炸等无法训练的状况时可以尝试BN来解决。另外，在一般使用情况下也可以加入BN来加快训练速度，提高模型精度等。

## 基于卷积神经网络的语义分割

### 3.3.1 语义分割简述

在计算机视觉领域，最基本的工作就是图像分割，也就是将把图片中相似的部分分成连续的区域块，如图像块的粗糙度、对比度、方向度、紧凑度等，以此辅助图像分割并提升效果。简单来说，语义分割就是给定一张图片，对图片上像素进行分类。早期研究者们使用传统的机器学习方法，如决策树以及深度学习方法，也就是卷积神经网络（全连接层型）。直到2015年CVPR最佳论文[34] “Fully convolutional networks for semantic segmentation”（FCN网络），将网络最后的几层全连接层变成了全卷积层，实现了图像级别分类到像素级别分类的扩展。近几年针对语义分割的网络也基本基于FCN。其他语义分割的网络结构还有Dilated Convolutions 、PSPNet[37]和Deeplab[39]。

### 3.3.2 FCN简述

传统的基于CNN的分割方法：为了对一个像素分类，使用了该像素周围的一个图像块作为CNN网络的输入来训练和预测。这种方法有几个缺点：一是存储开销很大。例如对每个像素使用的图像块的大小为15\*15，然后不断滑动窗口，每次滑动的窗口给网络进行判别分类，则所需的存储空间根据滑动窗口的次数和大小急剧上升。二是计算效率低下。相邻的像素块基本上是重复的，那么针对每个像素块逐个计算卷积，这种计算就有很大程度上的浪费。三是像素块大小的限制了感知区域的大小。通常像素块的大小比整幅图像的大小小很多，只能提取一些局部的特征，从而导致分类的性能受到限制。

而FCN则是从抽象的特征中恢复出每个像素所属的类别。即从图像级别的分类进一步延伸到像素级别的分类。它的核心思路主要使用了以下三种技术

1.全卷积

FCN最重要的观点就是提出了将AlexNet、VGG和GoogleNet的全连接层替换成全卷积层。全连接层和全卷积层之间唯一的不同就是卷积层中的神经元只与输入数据中的一个局部区域连接，并且在卷积列中的神经元共享参数。在两类层中，神经元都是计算点积，它们的函数形式是一样的。所以，将此两者相互转化是可能的：对于任一个卷积层，都存在一个能实现和它一样的前向传播函数的全连接层。同理，任何全连接层都可以被转化为卷积层。其中权重矩阵是一个巨大的矩阵，除了某些特定块，其余部分都是0。在大部分块中，元素都是相等的。比如，一个 K=4096 的全连接层，输入数据的尺寸是 7∗7∗512，这个全连接层可以被等效地看做一个 F=7,P=0,S=1,K=4096 的卷积层。换句话说，就是将滤波器的尺寸设置为和输入数据体的尺寸一致。只有一个单独的深度列覆盖并滑过输入数据体，所以输出将变成 1∗1∗4096，可以使得结果与初始的全连接层一致。

全连接层与卷积层的相互转换中，将全连接层转化为卷积层在实际运用中更加有用。假设一个卷积神经网络的输入是 224\*224\*3 的图像，一系列的卷积层和下采样层将图像数据变为尺寸为 7\*7\*512 的激活数据体。AlexNet使用了两个尺寸为4096的全连接层，最后由一个含有1000个神经元的全连接层用于计算分类评分。我们可以将这3个全连接层中的任意一个转化为卷积层：针对第一个连接区域是[7\*7\*512]的全连接层，令其滤波器尺寸为F=7，这样输出数据体就为[1\*1\*4096]了。针对第二个全连接层，令其滤波器尺寸为F=1，这样输出数据体为[1\*1\*4096]。对最后一个全连接层也做类似的操作，令其F=1，最终输出为[1\*1\*1000]。FCN中将全连接层的权重重塑成卷积层的滤波器，这样可以让网络接受任意大小的图片而且节省了不必要的内存开销，加速了训练过程，具体方式是在最后几层使用1\*1的卷积核对图片卷积，不改变图片大小，设定卷积核个数以后依然可以做分类网络。最后将输出通过上采样得到精细的分割。

2.上采样

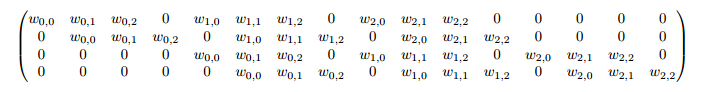
众所周知，图片经过卷积层后输出的尺寸会发生变化，如果不刻意保持图片尺寸，经过连续的5次卷积/池化后，h\*w的原图会变成h/32 \* w/32，特征的位置信息也会变得模糊，所以为了将抽象的输出变成稠密的像素预测，作者采用了上采样方式和一个跳跃连接来恢复信息。FCN中的上采样可以看作一种转置卷积（transposed convolution），设定一次卷积，输入图片为2n\*2n，卷积后得到n\*n的输出。把输入输出矩阵展开为一维向量，记做I和O，那么卷积可以认为是O=WI，其中W为图3.5中的矩阵

图3.5 系数矩阵W

对于上采样操作，文章中称为反卷积（deconvolution），相当于输入n\*n的图片得到2n\*2n的图片。同理，将输入输出展开为一维向量，记做I’和O’，那么上采样后，输出为O’=I’。 也就是说，如果在卷积操作中，输入图像的某一位置 p 及输出图像的某一位置 q 通过权重Wpq或者0连接，那么对应的在反卷积操作中，输入图像的位置 q 与输出图像的位置 p 也通过权重W’pq或0连接。

3. 忽略连接结构（又称跳跃结构）

FCN中还提出了一种新的结构，称为跳跃结构。这个结构的作用就在于优化结果，因为如果将全卷积之后的结果直接上采样得到的结果是很粗糙的，所以作者将不同池化层的结果进行上采样之后来优化输出。

## 本章小结

本节首先介绍了人工神经网络、卷积神经网络的发展，较为细致的讲述了卷积神经网络的特点、结构以及部分重要的卷积计算方法。最后介绍了基于神经网络的语义分割研究的发展和语义分割任务中较为重要的论文Fully convolutional networks for semantic segmentation的基本概要，重点的阐述了该论文提出的FCN网络结构中的主要技术和创新点。

# 第四章 基于PSPNet的冠脉造影图像分割

## 4.1 数据集的建立

冠状动脉是医学中的专业术语，是指分布在心外膜下和心肌壁内，维持心肌供血的血管，因此建立本研究课题的数据库需要专业医生的辅助，我们和阜外心血管病医院的医生进行了学习和请教，并由医生教授相关的医学知识并协助完成数据的粗标注流程。

### 4.1.1 数据集的分类

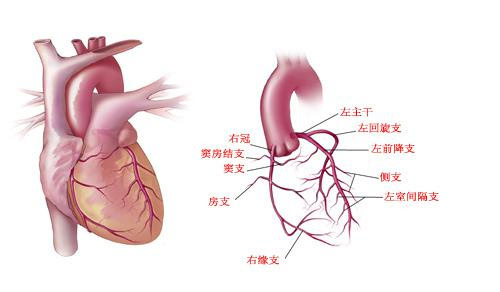
1. 心脏的形状是不规则的，但冠脉就如同手指和手掌从心脏下方握住心脏，手指就像冠脉血管，只不过产生了很多分支，但是远看动脉的形状就像王冠一般，这就是其名称由来，见下图4.1。冠状动脉和静脉形成冠脉循环，供给心脏营养。医学上一般划分为左右冠 冠状动脉。

图4.1冠状动脉示意图

左冠状动脉一般是一短干，自左主动脉窦起，经肺动脉起始部位和左心耳间，沿[冠状沟](https://baike.baidu.com/item/%E5%86%A0%E7%8A%B6%E6%B2%9F)向左前行3～5mm后，走势分叉为左室间支和左旋支。前室间支沿前室间沟下行，绕过心尖切迹至心的膈面与右冠状动脉的后室间支相吻合。我们主要研究的分支有：

a) 左主干：起源于左冠窦，一般高于右冠开口，多数长度在0-11mm之间，少数人左主干缺失，左前降支及左回旋支分别起源于左冠窦。

b) 左前降支：从左主干发出后沿前室间沟下行，约76%可到达心尖部。沿途分支：对角支（一般为1- 3支）、间隔支。供血范围占左心室的40- 50%。

c) 左回旋支：从左主干发出后沿左侧房室沟下行至后十字交叉。沿途分支：钝缘支、窦房结支。左优势型时可达45- 55%。供血范围占左心室的10-20%。

右冠状动脉分支简介

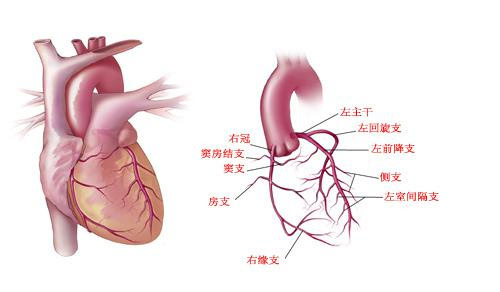
右冠状动脉起源于右冠窦，沿房室沟下行至后十字交叉处，然后分为右室后支及后降支，供血范围占右心室的20-30%。沿途分支主要有：圆锥支、窦房结支、右室支（锐缘支）、房室结支。

图4.2冠状动脉血管分支的分类

冠状动脉造影图在拍摄时有以下几种投照体位，如图4.3：1）AP，正位2）RAO，右前斜3）LAO，左前斜4）CRA，头位5）CAU，足位。临床上通常需要将RAO/LAO及CRA/CAU结合，如 RAO+CRA、LAO+CAU等。最终我们按照在不同体位下冠状动脉拍摄的方式的不同，将整个冠状动脉造影图数据集分为以下几类：其中，左冠状部分包括CAU、CRA、LAO\_CAU、LAO\_CRA、RAO\_CAU、RAO\_CRA六类数据集；右冠状部分LAO、LAO\_CAU(右)、LAO\_CRA(右)、CRA(右)，在数据集中，我们注意到几乎所有体位下右冠的造影图形状都大体类似，形如C型，所以我们把右冠部分的数据统一集合成了一个总的数据集。

图4.3冠状动脉造影图体位的分类

### 4.1.2 数据的标注

1.标注工具的设计与实现

标注工具的基本功能包括血管分支标注和病变标注。

由于需要使用标注工具标注数万例影像，满足需求的标注工具能够明显的提高数据标注效率，因此标注工具的便利性显得格外重要。并且标注工具要能够从DICOM格式的影像文件中解析出造影的体位信息，让医生和标注人员根据造影的体位信息和血管的走形来判断各个血管分支。使用标注工具标注完成的影像及标注信息要能够方便的输出给深度学习网络进行训练使用，所以标注工具要具有导出标注信息的功能，即带有部分病变信息的标注功能。最后，为了方便医生和标注人员使用，标注工具需要移植到surface或ipad等平板电脑上使用平板电脑的触控笔进行标注。于是我们设计实现了如下图4.4的标注工具。

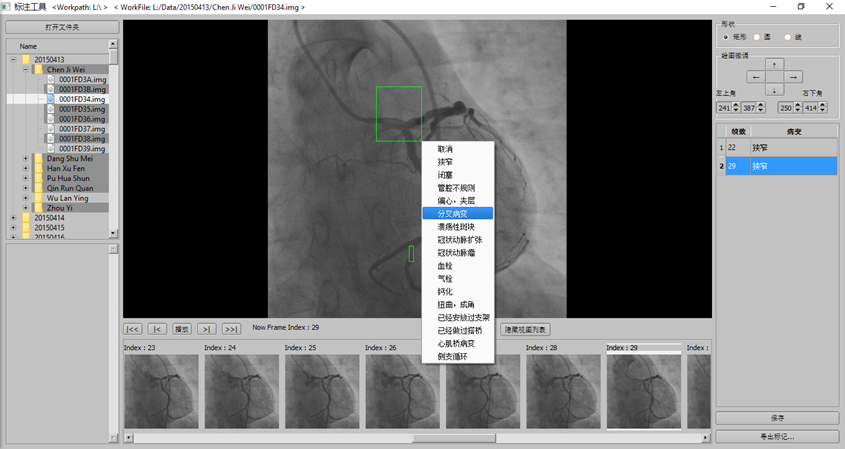


图4.4标注工具图

2.医生的粗略标注

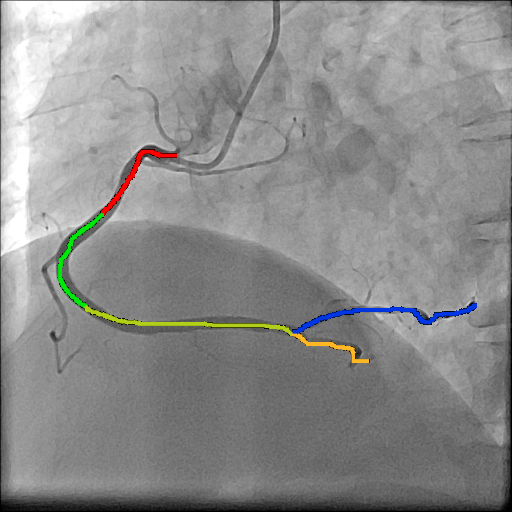
在完成真实病例中的冠状动脉造影图像从Dicom文件中到标注工具中具体图片的转换后。医生从已有的图像数据库中，挑选合适的目标图像，进行简单的人工粗略特征标注。即：由医生在静态帧图片上进行简单的描画，标出当前体位下能观察到的主要血管及其走向，忽略部分不重要血管及过细的血管，其中不同颜色代表不同的主要血管，使之变成本课题研究所需要的基本粗标注数据，见下图4.5。

图4.5医生粗略标注示意图

3.电脑端利用影像标注工具的精细标注

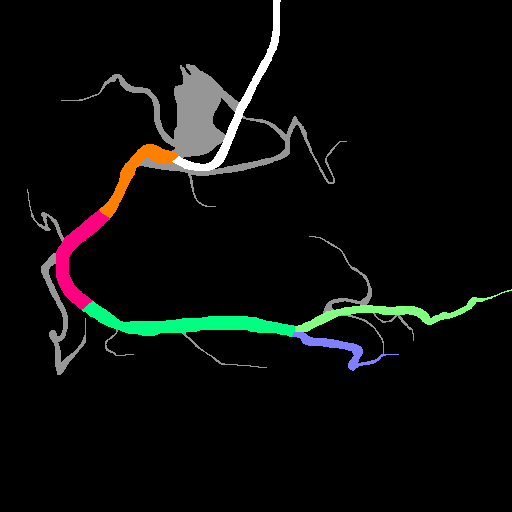
在专业医生完成简单的粗略标注后，我们再对获取的数据在电脑端使用Photoshop进行精细标注。根据医生所绘制的粗略走势图，对其中的主要血管逐个像素进行标注，对被忽略的其他血管及干扰项使用灰色RGB值（150，150，150）标注，背景设为黑色RGB值（255，255，255）。完成后如下图4.6

图4.6像素级精标注示意图

4. “多人标注，可信度最优”的策略

鉴于大量精细的人工标注信息是为了用于基于深度学习的冠状动脉影像识别与检测，从而获得更加可靠和科学检测结果。因此在课题中，我们采用了“多人标注，可信度最优”的策略对图像数据进行标注。即，多人对同一份数据分别进行标注，然后采用相应的策略得到一份可信度最高的标注结果。其中，可信度的判断原则如下：

（1）少数服从多数原则。当标注数据出现分歧时，以多数人同意的标注结果为准。

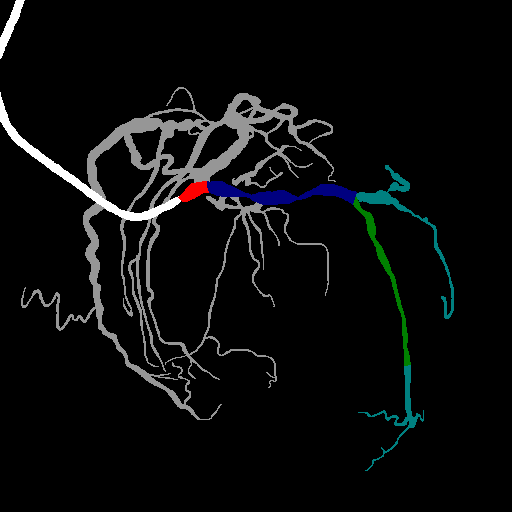
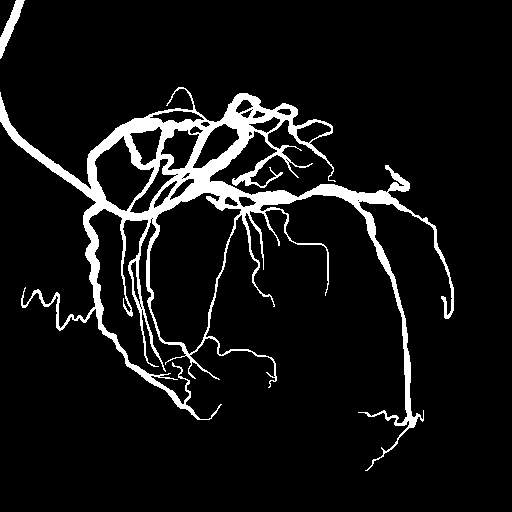
（2）可信度之和最优原则。对每个标注者维护一个可信度列表，其标注的数据正确的越多，可信度就越高。在标注数据出现分歧，且排名前二的两份标注的同意者的人数很相近时，则计算每份标注同意者的可信度之和，以可信度之和高的标注为准。

### 4.1.3 数据的预处理

考虑到后续训练时，网络只能接受灰度图作为输入，所以我们按照血管重要程度的评分将血管RGB值再转为0到N-1的灰度值，N为各自体位下的血管种类。考虑到有些血管的作用相近，所以转为相同的灰度值，转换后的N值见表格4-1。

表4-1 灰度转换N值对照表

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 体位 | CAU | CRA | RAO\_CAU | RAO\_CRA | LAO\_CRA | LAO\_CAU | RIGHT |
| N | 10 | 11 | 10 | 11 | 11 | 15 | 8 |

转换后的灰度图下图，转换后的灰度图与原图成对输入网络。

a：精细标注数据 b：转换后的灰度图

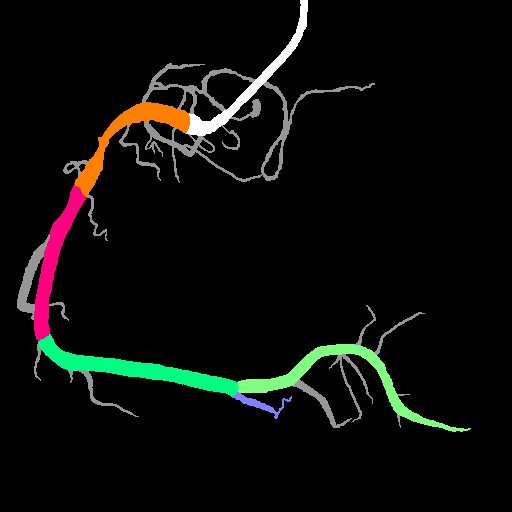
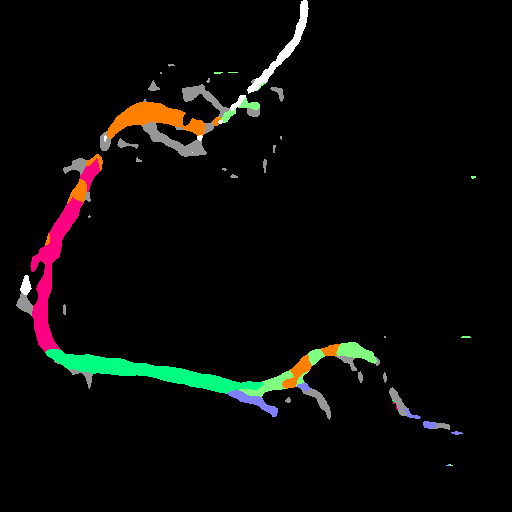
图4.7 LAO\_CAU精标注图转灰度图

## 4.2 基于PSPNet的冠脉造影分割

### 4.2.1 FCN的不足

依上文中的叙述，语义分割是计算机视觉的一个很重要话题，它的目的是将图片中的每个像素点都做分类，然后计算机能通过文字标注或者不同颜色对图中的区域做分割。在FCN网络中，虽然做到了输入任意尺寸的图片就可以训练，但是FCN的分割的效果并不够精细，原文中使用对比了8倍上采样、16倍上采样以及32倍上采样都进行了实验，从实验结果来看，8倍的效果要远远优于32倍，且上采样的结果还是比较模糊和平滑，对图像中的细节不敏感。此外，FCN虽然实现了对每个像素进行分类，但是没有充分考虑像素与像素之间的关系。忽略了在通常的基于像素分类的分割方法中使用的空间规则化（spatial regularization）步骤，使得网络缺乏空间一致性。

在2016年ImageNet 场景分割挑战赛中，获得冠军的是由Hengshuang Zhao等人提出的PSPNet（Pyramid Scene Parsing Network）[37]，论文中提出，对于准确的场景来说，知识图依赖于场景上下文的先验信息。而FCN模型的主要问题就是缺乏恰当的策略来利用全局场景中的类别线索，在本次图像分割任务中，我们也尝试使用了FCN8s网络来进行实验分割血管，结果如下图4.8。



（a）精细标注原图 （b）实验分割结果

图4.8基于右冠数据的FCN8s实验分割

总的来说，FCN有以下几点问题：

1.关系不匹配

对于语义分割，背景场景中的信息对于图像的定义是十分重要的，尤其处理特殊场景的图片，如冠状动脉造影图，背景是人体心脏及其轮廓，伴随有少许会有脊柱骨骼、导管等干扰项，但是绝对不会出现其他人体生理结构或者不属于人体的生理结构的干扰因素。在PSPNet论文中做了一组用FCN的对比实验，实验结果中FCN网络出现了明显的误判，即：把湖边停泊的船分类为汽车，而汽车出现在湖面上这一情况在现实生活里是几乎不会发生的，鉴于此证明，FCN网络缺少理解图片背景潜在信息的能力。

2.类别混乱

最初的设想中，我们希望计算机能够识别分类出所有现实生活中的物体，正如我们人眼一样能够很好的区分我们所处的世界千奇百怪的事物。例如：我们人类能够很好的区分事物细分的类别，狗这种动物可以细分为雪纳瑞、泰迪、德国牧羊犬等等不同品种。这个问题也是语义分割任务中需要面对的巨大挑战，即：类别之间十分相近，并没有清晰的界限。PSPNet的论文中研究发现，当事物具有相似的外观，如田野和土地、高山与山丘、普通楼房与摩天大楼，FCN网络的表现并不是很好，使得分类较难，实验中FCN将一座摩天大楼的下半部分划分为另一幢普通楼房，也证明了其缺少利用类别关系的能力，对于图像中类别的细节并不敏感。

3.忽略不明显类别

图像分割一般是对区域的划分，因此图片中存在的尺寸较大的物体会因为占据的区域及像素点数量巨大，造成可能超过了FCN中的感受野范围，当同一张图中又存在有一些尺寸较小的目标物体，比如花纹、枕头时，就会使得网络造成了不连续的分割。PSPNet论文也针对这一情况做了实验，结果表明，FCN会因为枕头的花纹与床单一致而没有将枕头与床单区分，其次，因为柜子尺寸较大且的花纹是动物形状，从而将柜子整体进行了忽略。

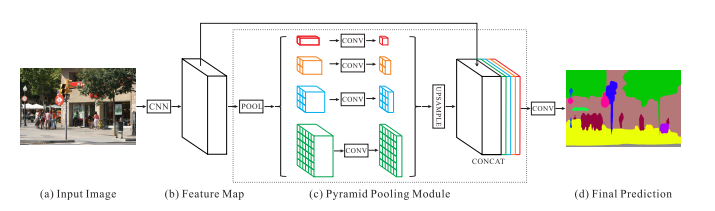
在PSPNet网络中旨在解决FCN网络对于背景信息的忽略和对不同感受野信息的融合，即：拥有合适的全局场景中先验信息的深度学习网络有可能提升语义分割的效果。

### 4.2.2 金字塔池化

传统的全局平均池化是将所有像素融合成一个向量，这样操作虽然较好的利用了全局背景中的先验信息，但是会失去部分空间信息从而造成歧义，所以对于有相似类别以及特征的物体，结果会出现混淆。在一个卷积神经网络中，感受野的大小可以大致说明网络结背景信息的程度，感受野（receptive field）是输入空间中特定CNN特征的一个范围区域的定义，它可以决定所在层输出结果（feature map）中一个像素所对应的输入图像的区域大小或映射。普通的卷积输出的特征图尺寸会有所改变，因此无法确定该特征在原图中的位置和区域大小，即：来源感受野的中心位置及其尺寸大小，造成失去了背景信息和空间信息。

金字塔池化模型是一个能够掌握全局背景中的先验信息的高效模型结构。2015年何凯明[38]等人提出的空间金字塔池化操作（Spatial Pyramid Pooling，SPP），考虑到之前网络输入要求是固定尺寸的图片来满足全连接层的要求，是使用了各类裁剪技术对图片进行处理，但是这种操作会出现部分重复，导致某些区域的权重变化，所以论文中空间金字塔池化的操作在于统一了图片大小，从而解决输入图片大小不一的情况。具体的说，空间金字塔池化是将前一个卷积和池化层输出的每个feature maps图片做3种卷积操作，包括：原本的特征图，分成大小是4的特征图和分成尺寸是16的特征图，将这3种特征图相加得到了21种的特征图。

一般来说空间金字塔池化有以下三个优点：1.对特征面的多角度特征进行提取接聚合，从而让网络拥有较好的鲁棒性；2.解决了输入图片大小不一的问题；3.提高了目标检测的精度。其中SPP结构对输入的每一张特征图都进行了多尺度的特征提取，基于此优点，后来大量的检测类网络都运用了SPP结构。

后来的网络大致分为两个研究方向，第一个方向是结构预测（structure prediction），例如deeplab中引入的条件随机场（conditional random field, CRF ）[39]；另一个方向是特征整合（Feature Ensembling），它又分为：多级（multi-level）特征整合和多尺度（multi-scale）特征整合，金字塔池化是多尺度特征融合的一种。为了降低背景信息在不同子区域的的loss值，PSPNet作者参考空间金字塔池化模型提出了一种多层次型全局先验信息的结构金字塔池化，包括了不同子区域和不同尺度的信息。

（a）：输入图片 (b)：特征图 (c)：金字塔池化 (d)：最终预测

图4.9 PSPNet的网络结构

金字塔池化模型结合了四种不同尺度的特征。图4.9-(c)中红色的立方体是一个全局池化，产生了单个输出，橙黄色立方体将特征图分成4个子区域（尺寸为2\*2），蓝色的尺寸为3\*3，绿色为6\*6，它们均以不同位置的集合池的形式表示。金字塔池化输出的4中特征图是不同尺寸的。为了保持全局特征的权重，论文在每个金字塔的层级输出后面使用了大小为1\*1的卷积层。这样可以将原来在金字塔中维度为N的图片减少到1/N的维度大小。之后，作者直接对低维度的特征图做双线性插值上采样至相同的尺寸。最后将不同级别的特征进行拼接输出。这种金字塔池化模型还有一个优点就是可以任意修改每个级别尺寸的大小，它们和输入模型的特征图大小相关。金字塔池化模型将特征多尺寸化，丰富了分割的细节和区域。

### 4.2.3 网络结构

PSPNet中预训练使用的是ResNet和扩张网络（dilated network）相结合的方式，首先对输入的数据进行特征提取，如图4.9-(b)，这样提取出的特征图的尺寸为原来输入图片的八分之一。再将特征图输入金字塔池化模块，最后级联得到特征面，然后通过计算loss和反向传播算法学习并调整参数。这样的结构相比global pooling，该模型获取多尺寸全局信息效果更好。计算上相比dilated FCN网络也不会增加很多，global pyramid pooling模型与FCN特征提取模型可以同时进行训练优化。

1.data层及预处理

针对本次实验要用到的冠状动脉造影图数据集，我们重新编写了data层，首先对数据做了预处理。首先进行图片的裁剪，将大部分背景去除，实验中选择了基于经验的尺寸大小473。然后减去图片均值，把图片映射到一定范围内。最后设置几个不同的scale\_chang\_factor对输入数据的大小进行转变。并且在训练过程中每一次迭代都打乱图片顺序保证数据的随机性。

2.网络初始化

我们知道，一般在训练网络时会赋值给权重一个初始值，合适的初始值可以帮助网络快速收敛，缩短找到最优值的时间，从而加快网络训练的速度。实验中我们选择了2012年VOC数据集在PSPNet上训练后的权重值作初始化操作。

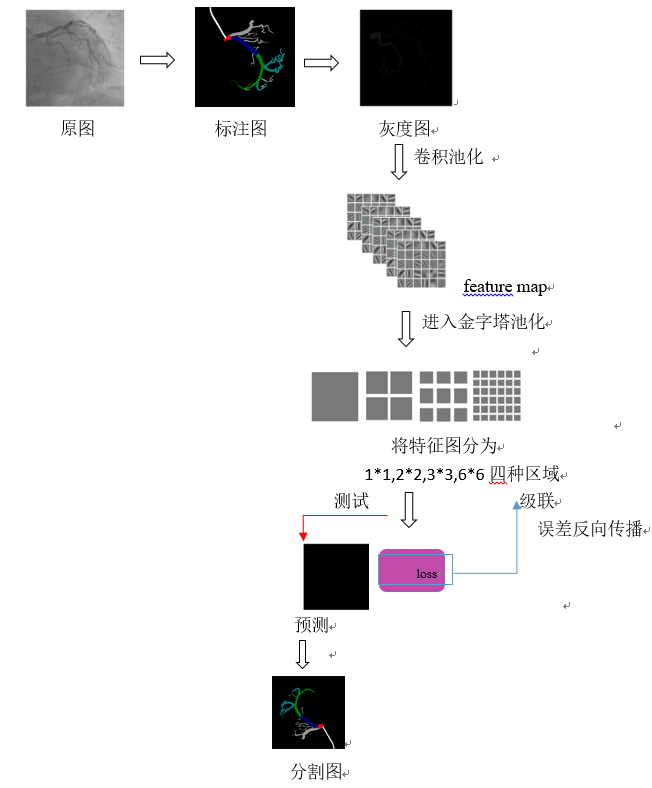
3.其他层参数的调整

鉴于不同体位的血管类别并不相同，实验中将“conv5”、“conv5\_128”、“conv6”、“conv6\_256”、“conv7”输出的参数做了不同的修改。

## 4.3 其他设计

基于PSPNet的冠脉造影图分割的网络结构如图4.10。

本次实验的流程如下，首先对冠状动脉造影图进行数据的标注：分为粗标注、精标注两步，然后按照医学上对体位的划分分别进行分类和训练测试。然后，将标注好的图处理成灰度图并输入到caffe框架，图片经过网络中卷积池化等组合操作后提取出feature map，然后将其放入金字塔池化层。金字塔池化层会把特征图分为1\*1,2\*2,3\*3,6\*6几种进行不同尺寸卷积，然后级联整合输出，再经过其他卷积和Dropout等操作，最后使用softmax分类器对像素进行分类，并实现最终的预测，完成分割实验。

图4.10 基于PSPNet的冠脉造影图像分割

## 4.4 本章小结

本章中，首先介绍本次实验使用的冠状动脉造影图数据集的建立，包括原始数据的选取、粗标注、精标注，以及将其按照医学上的常识根据不同体位进行分类，然后介绍了用于网络训练的造影图像精标注的部分流程以及为了训练所做的预处理转换为灰度图。其次介绍了本次实验所选用的网络模型PSPNet，细致的介绍了其中重要的技术创新点，金字塔池化模块的构成和优点。然后介绍了本次实验中所修改的PSPNet网络结构，并设计了基于PSPNet冠状动脉造影图的分割实验的流程，准备开始实验。

# 第五章 实验结果及分析

## 5.1 实验流程

图5.1 实验流程图

## 5.2 实验环境及训练方法

### 5.2.1 实验环境

基于PSPNet的分割实验使用的是Ubuntu14.04环境、CAFFE框架；硬件设备使用的是GTX1080显卡。

### 5.2.2 训练和测试方法

训练和测试的原始数据集来自医院病例中的冠状动脉造影Dicom文件，进过处理帧选取和标注后，统一图片尺寸为512\*512，通道数为2，即原始图为灰度图。经过4.1章节对数据的分类和标注操作，实验将图片输入网络进行训练和测试。

1.训练网络具体如图5.2：Conv1-5五个正方形浅灰色框代表卷积模块，参考依据是ResNet结构，conv5-7长方形深灰色的框代表卷积层，之间通过interp和fuse连接，PSPNet的框架是基于ResNet101的框架上修改而来的，因此在Conv4输出到con\_aux一个辅助的卷积层来计算辅助loss值loss\_aux和在Conv5模块后，加入PSPNet的多尺寸的金字塔池化模块，进行多尺度特征提取的操作。在最后，网络经过Loss层进行反向传播学习。训练和测试使用的都是ReLU函数作为激励函数，采用跟Alexnet相同的Dropout=0.5来减少参数防止结果过拟合。

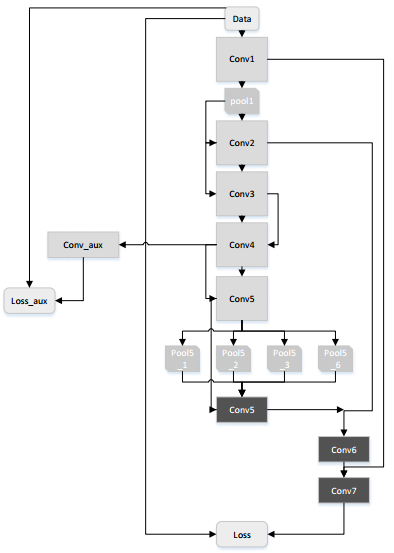
2.最终测试与训练时不同，我们没有使用辅助的loss\_aux，而是经过最后一个Conv7层后直接输出结果。权重参数是第1步中操作训练和验证得到的，预测结束，统一对图片做尺寸的调整然后输出结果。

图5.2 训练时网络结构

### 5.2.3 目标分割效果的评估方法

在实验进行后，我们需要用一个合适的方法对分割效果进行评估。虽然人工肉眼可以大致评判出分割效果的好坏，但这种主观的评价并不准确，也不能直观地量化算法的优劣，因此我们需要找到一个较为规范的量化标准对目标分割的结果进行评估。

1. 准确率（Accuracy）、精确率（Precision）和召回率（Recall）

假设原始样本中有两类，其中共有P个类别为1的正样本和N个类别为0的负样本；经过分类后有TP个标签为1的样本被正确判定、FN个标签为1的样本被错误判定，有TN个标签为0的样本被正确判定、FP个标签为0的样本被错误判定，如表5-1所示：

|  |  |  |
| --- | --- | --- |
|  | 正样本（P） | 负样本（T） |
| 预测为正 | TP | FP |
| 预测为负 | FN | TN |

表5-1

定义三个基础指标Accuracy（准确率）、Precision（精确率）和Recall（回归率）：

 式（5.1）

准确率表示所有分对的样例占整个分类数据的比重。

 式（5.2）

精确率表示被分割出的正样例中真正的正样本的比重。

 式（5.3）

回归率表示了被正确分割的正样例占全部正样例的比重。

即：在血管分割试验中，准确率表示了所有被分对的代表血管像素点所占的比例；精确率表示了被判定为血管的像素中真正是血管的像素比例，即血管分割的精确程度；召回率表示了被判定为血管的像素占所有标定为血管的像素的比例，即有多少血管没有被漏检。总之，准确率关注正负样本分别有没有分对，精确率关注正样本有没有分对，回归率关注正样本漏检了多少。

2. F1分数评价

F1就是综合考虑血管像素分割的召回率和精确率，把二者相结合，这样做能平衡二者之间的差异，得到一个较为平均的估值，计算方式如下：

 式（5.4）

在本次课题中取a=1，则

 式（5.5）

## 5.3 实验结果记录和分析

表5-2 配置文件参数

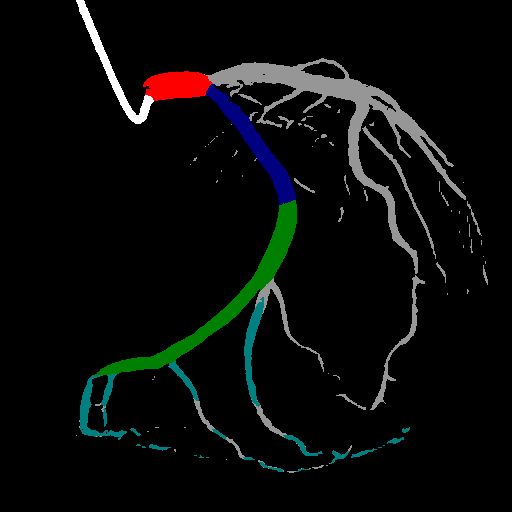
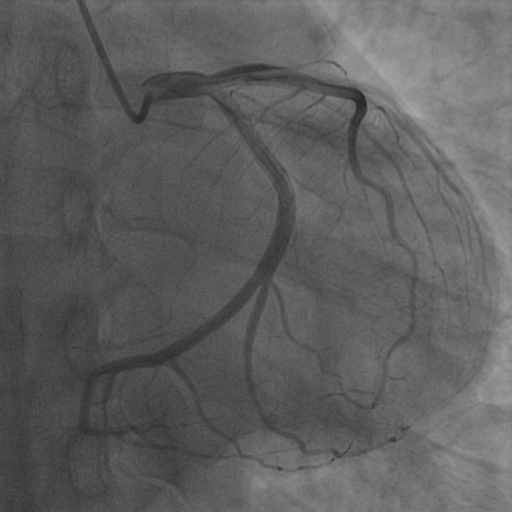
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 参数 | test\_interval | base\_lr | gamma | | momentum | | weight\_decay | type |
| 数值 | 100 | 0.0001 | | 0.1 | | 0.9 | 0.0001 | SGD |

训练时学习率的调整规则遵循如下规则，即

 式（5.6）

1.左冠

a.CAU体位

训练集848张，包括290张粗标注数据（其余为精标注数据），测试集50张，训练迭代40000次时效果达到最好，精确率76.1%，召回率67.1%，F1为71.3%，配置文件参数如上表5-2。在实验结果中，表现最佳的一张图片如图5.4中的分割结果，该图出现的问题主要是血管末端的枝节分割并不连续，以及出现了部分的误判。

（a）原始数据 （b）实验分割结果

图5.4 CAU体位分割结果

b.CRA体位

训练集528张，包括130张粗标注数据，测试集50张，训练迭代105000次时效果达到最好，精确率77.1%，召回率59.9%，F1为67.4%。在实验结果中，表现最佳的一张图片如图5.5中的分割结果，该图出现的问题主要是其他血管（灰色）没有很好的预测以及血管末端预测不充分，出现该问题的原因是其他血管在末端十分纤细，细节容易被忽略，在实际运用当中，其他血管所占的比重也不是很大，因此能够接受。

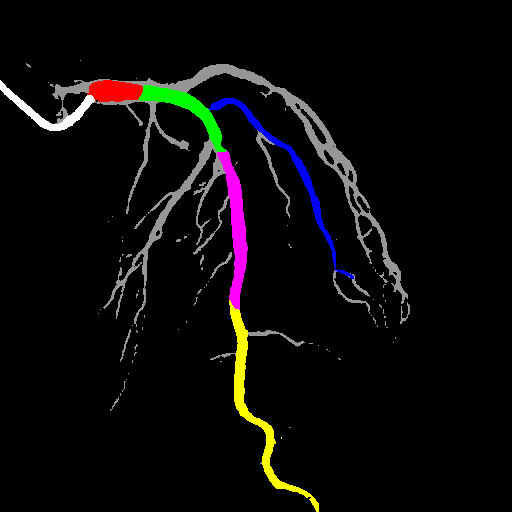
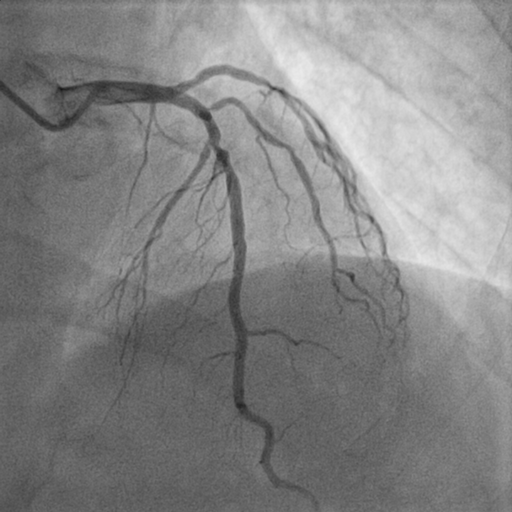
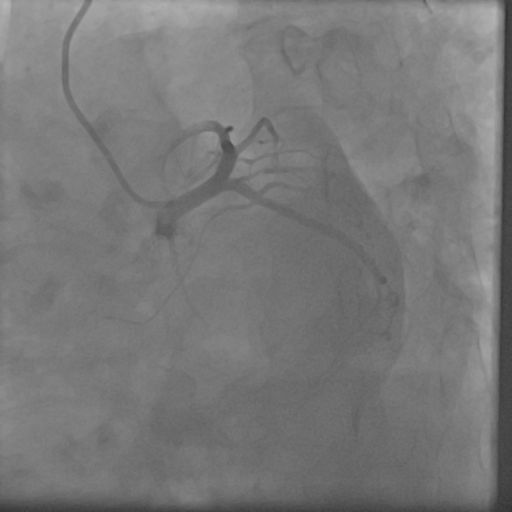
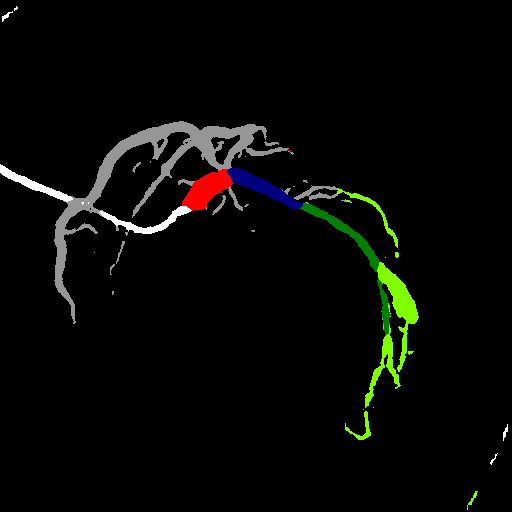
（a）原始数据 （b）实验分割结果

图5.5 CRA体位分割结果

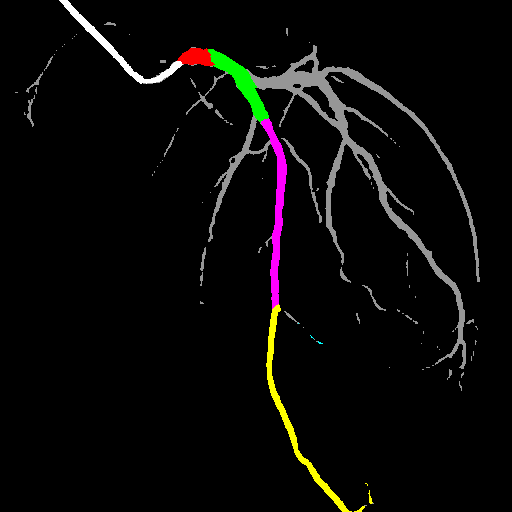
c.LAO\_CAU体位

训练集569张，包括120张粗标注数据，测试集50张，训练迭代200000次，精确率74.7%，召回率64.1%，F1为68.9%。在实验结果中，出现问题最有代表性的结果如图5.6中，该图出现的问题主要是导管（白色）部分检测不完全，被其他血管覆盖；后降支（若草色）出现了明显的空洞，且与左室后侧支（墨绿色）相互覆盖，出现这个问题的主要原因是后降支与左室后侧支在距离上十分接近，形状类似，本身区分定位就比较困难，所以网络在预测时不好分辨位置信息与特征信息，导致了混淆，造成误判。

（a）原始数据 （b）实验分割结果

图5.6 LAO\_CAU体位分割结果

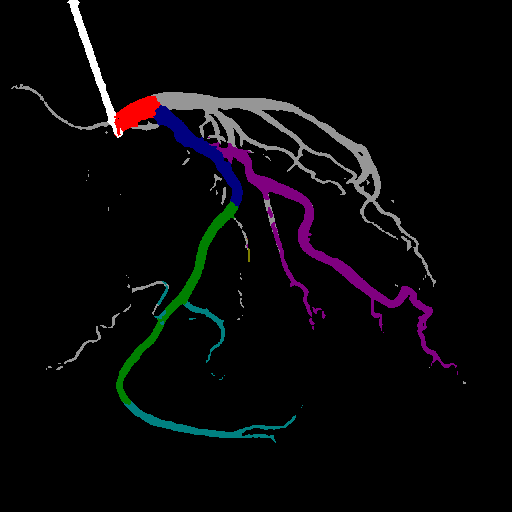
d.LAO\_CRA体位

训练集905张，包括637张粗标注数据，测试集50张，训练迭代50000次时表现最好，精确率74.9%，召回率61.7%，F1为67.0%。如下图5.7所示，该图出现的问题主要是左主干（红色）与前降支（浅绿色）边缘分割并不平滑；前降支远端（黄色）与其他血管（灰色）末端分割不连续，造成这个结果的原因是毛细血管的末端一般比较纤细，对于精细分割的任务是较难的问题。

（a）原始数据 （b）实验分割结果

图5.7 LAO\_CRA体位分割结果

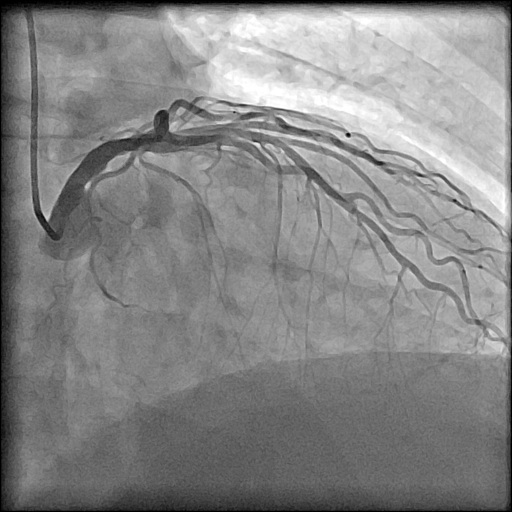
e.RAO\_CAU体位

训练集772张，包括583张粗标注数据，测试集50张，训练迭代10000次是表现最好，精确率81.6%，召回率66.1%，F1为73.0%。如下图5.8所示，该图出现的问题主要是第一钝圆支（暗紫色）在分割时出现了空洞，被背景像素点覆盖。这是因为在PSPNet网络进行上采样时使用的双线性插值方法导致零散的背景预测点被插值，导致背景（黑色）点范围扩大，从而出现了空洞。

（a）原始数据 （b）实验分割结果

图5.8 RAO\_CAU体位分割结果

f.RAO\_CRA体位

训练集1379张，包括1034张粗标注数据，测试集50张，训练迭代35000次是表现最好，精确率72.7%，召回率60.8%，F1为66.2%。如下图5.9所示，该图出现的问题主要是第一对角支（蓝色）血管在分割过程中的缺失，出现这个问题的原因主要是在RAO\_CRA数据集中，含有第一对角支的样本图片数量不足，出现了不同血管样本之间数量不均衡的现象，因此对这种类别的血管的学习是不充足的，从而导致了缺失检测的问题。

（a）原始数据 （b）实验分割结果

图5.9 RAO\_CRA体位分割结果

2.右冠

训练集1500张，包括504张粗标注数据，测试集50张。训练迭代175000次时效果达到最好，配置文件参数见表5-2。分割结果见图5.10，右冠类血管由于类别比较少而且右主冠脉总体呈“C”字形状，而且分支的血管位置形状相较于左冠血管差别并不是那么大，所以对关注血管的分割效果还是比较好，175000次迭代的参数进行测试后，精确率为78.5%，召回率为68.5%，F1为73.2%。但是从图中也可看出，对其他血管（灰色）、漏液（灰色）的分割还并不是很充分。鉴于实际项目中，一般医疗诊断时通常只利用关注血管，对其他血管不是主要供血的血管的并不关注，所以其误差影响不太大。

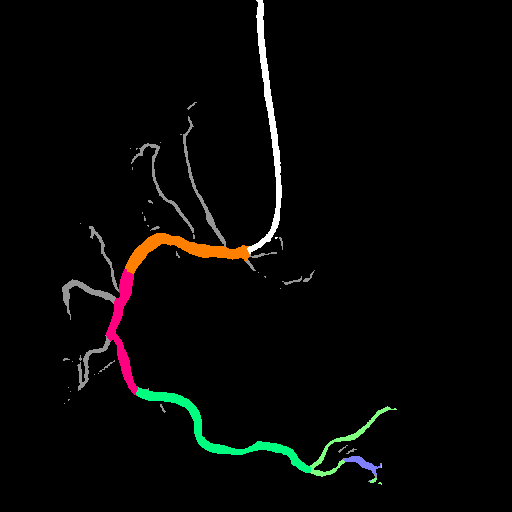
（a）原始数据 （b）实验分割结果

图5.10 CAU血管分割图

3.各个体位分割结果汇总

试验完成后对各个体位的分割结果做了汇总如下表5-3

表5-3 各个体位分割结果汇总

|  |  |  |  |
| --- | --- | --- | --- |
| **体位** | **最大F1值** | **精确率** | **召回率** |
| **CAU** | 71.3% | 76.1% | 67.2% |
| **CRA** | 67.4% | 77.1% | 59.9% |
| **LAO\_CAU** | 69.0% | 74.7% | 64.1% |
| **LAO\_CRA** | 67.0% | 74.9% | 60.6% |
| **RAO\_CAU** | 73.0% | 81.6% | 66.1% |
| **RAO\_CRA** | 66.2% | 72.6% | 60.8% |
| **Right** | 73.2% | 78.5% | 68.5% |

总的来说存在以下几个问题：

1. 训练时某些血管数据相对较少，是造成对该几类分类效果不佳的原因。想要增强对该几类的分类效果，可以采用强化训练。在数据的源头，增强该几类的数据，或者对含有该类血管的数据做旋转、镜像翻转、放大与缩小尺寸等增加数据量。

2. 位置相对固定且被覆盖较少的部分血管分类效果较好，细小而繁杂的或形态较为特殊的血管分类效果较差，如血管末端、其他血管、毛细血管等。这给后面做狭窄以及病变检测时的syntax评分造成了麻烦，目前可提供的解决方案有使用滤波器对图片进行降噪，还可以使用插值法让血管变得平滑。

## 5.4 本章小结

本章介绍了实验搭建的系统环境和所运用的深度学习运行框架以及硬件显卡设备，然后详细介绍了实验训练的网络结构及流程。随后阐述了本次实验的评估方法和指标，最后展示了冠脉造影图像多个体位，包含CAU、CRA、LAO\_CAU、LAO\_CRA、RAO\_CRA、RAO\_CAU、右冠等分割成果，并对实验结果中出现的问题进行了分析。总的来说，基于PSPNet的深度学习网络能够大体准确的把握不同体位下、不同分段血管的相对位置关系，基本满足了实验设计的期望，成功分离出导管(白色部分)、关注血管(彩色部分)、以及不关注的部分(灰色部分)和背景(黑色部分)，为后期的病变检测及syntax系统的评分做了铺垫。

# 第六章 总结与展望

## 6.1 论文工作总结

目标分割是图像处理的基础，以及图像分析的关键；心脏冠脉造影图像的分割应用是计算机视觉在医学领域的重要实践和创新。冠状动脉造影技术因为安全和精确的特性，也已广泛应用于临床，被认为是诊断冠状动脉心脏病的“金标准”。基于冠脉造影的图像分割在未来不但能够辅助医生做临床疾病诊断和治疗，也可以成为血管疾病量化描述的前提，还能为发展心脏血管三维重建建立基础。因此，实验对冠脉造影图像的精确分割具有非常重大的意义。论文深入探讨了基于冠脉造影图像的几种目标分割方法，主要的研究工作和创新性成果如下：

1.首先对现今世界上和中国心脏病概况做了调查，了解了国内相关社会环境和心脏病患病的原因，了解了心脏的组织结构和血管对于心脏是如何供血的等医学常识。了解了冠脉造影图的工作原理以及各个体位下的血管分支相关的医学知识。并调查研究了我国使用的血管分割的方法的种类及一些方法的优劣。

2.从原理上分析了几种传统的目标分割方法的主要思想，结合医学图像数据，研究并指出算法的不足之处，用实验加以证明。目前传统的图像分割方法主要有基于阈值的分割方法、基于区域的分割方法、基于边缘的分割方法以及基于特定理论的分割方法等。实验证明，虽然传统的分割方法计算速度快、处理图像便捷，但算法的缺点也非常突出，主要是难以抵抗高强度的噪声和分辨较为复杂精细的血管结构。

3.阐述了卷积神经网络的发展历程和几大优点，提出了使用深度学习的方法，更快、更简便、更精确地对心脏造影图像进行目标分割。深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征，以发现数据的分布式特征表示，因此可以学到传统方法难以得到的非线性结构，又因为其引入了大数据的收集，使得算法具有更好的鲁棒性。

4.结合之前卷积神经网络的知识，提出了完整的用于冠脉造影图像分割的算法流程，包括图像的预处理、网络的选用及具体结构、训练和测试方法等。并从零开始建立心脏冠脉造影图的数据库：联系医生搜集整理心脏冠脉造影图像数据，并商讨出一套合适的标注标准，设计实现标注工具，进行双重标注，由医生进行简单的粗标注，研究人员进行二次的精标注，从多个体位对感兴趣的血管分支进行描述，进而做预处理，为实验的开展做准备。

5. 最后，成功地用PSPNet对冠状动脉造影图进行了分割任务，在各个体位取得较为优异的效果，并撰写了测试的脚本，构造目标分割准确性的测度，用三个评估指标（精确率、召回率和F1）来评估分割的结果，使得分割算法的优劣能够被精确量化，并利用这些指标作为实验结果的输出，证明了基于深度学习的PSPNet网络在心脏造影图像分割任务上的优势。最后对结果进行了展示，用指标对预测进行了比对分析，对各个体位下出现的问题、不足以及有可能造成的原因进行了分析。

## 6.2 讨论与展望

本文使用PSPNet成功分割了冠状动脉造影图，但是用深度学习进行造影图像的分割任务还处于刚起步阶段，面对复杂情况下的各体位数据，目前还没有一个完美的网络能适应所有的分割任务。本研究项目在以下方面还有存在有一些问题：

1. 论文中的网络仅仅针对单帧的静态图像，并没有考虑动态序列各帧之间的关系。实际的心脏造影图像数据来源是连续的视频dicom文件，实验是将其转换为单帧图片进行分割，对连续运动序列信息的忽略降低了实验分割的质量。

2.虽然在右冠数据集上PSPNet网络得到了很好的分割效果，但是对于LAO\_CRA、LAO\_CAU等类别较多、结构较为复杂的造影图像，以及血管类型不均匀的的图片，出现了血管边缘分割并不平滑、部分血管空洞、部分血管类别混淆的现象，因而网络的分割效果还并不是很好，这一部分拥有很大的优化和改进空间。

3.目前数据集的量级还远远不够达到深度学习网络动辄几万数十万的数据规模，众所周知，数据量的增加会大大提升网络的实验结果，相信伴随着标注工作的进行，数据集量不断增大，分割结果会有很大的提升。

4.目前实验中研究的深度学习网络只是众多表现优异网络中的一个，相信伴随着深度学习技术的发展，会出现更多深层次的分割网络以及对小规模数据及医疗图像表现更好的网络，在之后的学习工作中可以不断进行尝试。

综上，下一步的工作主要包括：研究如何有效地分割结构复杂的体位以及更细小精密的血管、利用连续帧信息提高血管的分割质量、不断扩充增加数据集、尝试新的深度学习网络、细化目前使用网络的参数及结构精修。

相信在不远的未来，在精确分割出血管结构的基础上，就能实现人工智能+医疗项目给予医生辅助诊断的策略、基于冠脉造影图像syntax评分系统中深度学习良好的表现，以及基于深度学习算法的疾病检测系统的建立。

# 参考文献

1. 夏良正，李晓兵，李久贤. 基于模糊图像的分割方法[J].福州大学学报,1993,10(2):35-4.
2. Bergstrom S. Computer Models of Watershed Hydrology[J]. Computer Models of Watershed Hydrology, 1995:443-476.
3. Wu HS, Gil J, Deligdisch L. Region Growing Segmentation of Chromatin Clumps of Ovarian Cells Using Adaptive Gradients[J]. Journal of Imaging Science & Technology, 2004, 48(1):22-27.
4. Lo, S.-C., Lou, S.-L., Lin, J.-S., Freedman, M. T., Chien, M. V., Mun, S. K., 1995. Artificial convolution neural network techniques and applications for lung nodule detection. IEEE Trans Med Imaging 14, 711–718.
5. Krizhevsky, A., Sutskever, I., Hinton, G., 2012. Imagenet classifi- cation with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. pp. 1097–1105.
6. Brosch, T., Tam, R., 2013. Manifold learning of brain MRIs by deep learning. In: Med Image Comput Comput Assist Interv. Vol. 8150 of Lect Notes Comput Sci. pp. 633–640.
7. Plis, S. M., Hjelm, D. R., Salakhutdinov, R., Allen, E. A., Bockholt, H. J., Long, J. D., Johnson, H. J., Paulsen, J. S., Turner, J. A., Calhoun, V. D., 2014. Deep learning for neuroimaging: a validation study. Frontiers in Neuroscience.
8. Suk, H.-I., Shen, D., 2013. Deep learning-based feature representation for AD/MCI classification. In: Med Image Comput Comput Assist Interv. Vol. 8150 of Lect Notes Comput Sci. pp. 583–590.
9. Yang D, Zhang S, Yan Z, et al. Automated anatomical landmark detection ondistal femur surface using convolutional neural network[C]//Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on. IEEE, 2015: 17-21.
10. De Vos, Bob D., et al. "2D image classification for 3D anatomy localization: employing deep convolutional neural networks." Medical Imaging: Image Processing. 2016.
11. Cai Y, Landis M, Laidley D T, et al. Multi-modal vertebrae recognition using transformed deep convolution network[J]. Computerized Medical Imaging and Graphics, 2016, 51: 11-19.
12. Chen H, Ni D, Qin J, et al. Standard plane localization in fetal ultrasound via domain transferred deep neural networks[J]. IEEE journal of biomedical and health informatics, 2015, 19(5): 1627-1636.
13. Kumar A, Sridar P, Quinton A, et al. Plane identification in fetal ultrasound images using saliency maps and convolutional neural networks[C]//Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on. IEEE, 2016: 791-794.
14. Baumgartner C H, Kamnitsas K, Matthew J, et al. Real-time standard scan plane detection and localisation in fetal ultrasound using fully convolutional neural networks[J]. 2016.
15. Kong B, Zhan Y, Shin M, et al. Recognizing end-diastole and end-systole frames via deep temporal regression network[C]//International conference on medical image computing and computer-assisted intervention. Springer International Publishing, 2016: 264-272.
16. Hwang S, Kim H E. Self-transfer learning for fully weakly supervised object localization[J]. ARXiV preprint arXiv:1602.01625, 2016.
17. 余凯, 贾磊, 陈雨强, 等. 深度学习的昨天, 今天和明天[J]. 计算机研究与发展, 2015, 50(9): 1799-1804.
18. 许志涛. 基于CCSDS的遥感图像感兴趣区域压缩研究[D]. 中国科学院研究生院(长春光学精密机械与物理研究所), 2014.
19. Li Peng, Yuan X. Device and method for determining boundary of target region of medical image: wo, wo/2013/131420[P]. 2013.
20. Otsu N. A Threshold Selection Method from Gray-Level Histograms[J]. IEEE Transactions on Systems Man & Cybernetics, 2007, 9(1):62-66.
21. 郑健, 董月芳, 戴亚康, et al. Retina image processing method and device:, CN 103198321 A[P]. 2013.
22. Vezhnevets V, Konouchine V. "GrowCut" - Interactive Multi-Label ND Image Segmentation By Cellular Automata[J]. Graphicon-2005, Novosibirsk Akademgorodok, 2005.
23. Rosenblatt F. The perception: a probabilistic model for information storage and organization in the brain[M]// Neurocomputing: foundations of research. MIT Press, 1988:386-408.
24. Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[M]// Neurocomputing: foundations of research. MIT Press, 1988:533-536.
25. Hinton G, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. Science, 2006, 313(5786):504-507.
26. 王添翼. 基于卷积网络的三维模型特征提取[D]. 吉林大学, 2006.
27. Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex.[J]. Journal of Physiology, 1962, 160(1):106-154.
28. Fukushima K, Miyake S. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position[J]. Pattern Recognition, 1982, 15(6):455-469.
29. Lecun Y, Boser B, Denker J S, et al. Backpropagation Applied to Handwritten Zip Code Recognition[J]. Neural Computation, 2014, 1(4):541-551.
30. Zeiler M D, Fergus R. Visualizing and Understanding Convolutional Networks[J]. 2013, 8689:818-833.
31. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.
32. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[J]. 2014:1-9.
33. Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, PP(99):1-1.
34. Wu X. Fully Convolutional Networks for Semantic Segmentation[J]. Computer Science, 2015.
35. D'Informatique D E, Ese N, Esent P, et al. Long Short-Term Memory in Recurrent Neural Networks[J]. Epfl, 2001, 9(8):1735 - 1780.
36. Sainath T N, Mohamed A, Kingsbury B, et al. Deep convolutional neural networks for IVCSR//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver, Canada, 2013: 8614-8618.
37. Zhao H, Shi J, Qi X, et al. Pyramid Scene Parsing Network[J]. 2016.
38. He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(9):1904-1916.
39. L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L.Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arXiv:1606.00915, 2016. 5, 7, 8, 9.

# 致谢

值此论文完成之际，我要向指导老师致以最诚挚的谢意，他在我完成论文的过程中给出过许多非常宝贵的建议，衷心感谢老师在我毕业设计期间给予的悉心指导；感谢阜外心血管医院的专业医生，在我课题研究中给予的相关医学教授和指导；最后还要感谢在建立数据库过程中参与标注的人员，数据的标注工作是整个实验的基石与铺垫，谢谢他们的付出。

最后，还要特别感谢在百忙之中抽出宝贵时间，不辞辛劳参加答辩的各位老师。

# 作者攻读学位期间发表的学术论文目录