

## A Summary of Discussion(10.23)

冯浩哲

2017 年 10 月 23 日

# 提纲

Danny Chen 教授针对我们在肺结节分割中所遇到的困难主要提出了两类建议，一是在研究过程中首先需要明确哪些问题，二是我们当前问题的发展方向与现有的可用方法。

## 需要明确的问题

Danny Chen 教授以发表文章的 Introduction 部分为例，指出我们在做研究之前应该明确以下几个问题：

- 我们着手解决的具体问题是什么？
- 解决这个问题有什么样的价值？

以肺结节为例，肺结节分割问题对于医生能起到什么样的帮助？它是否是解决一系列肺结节相关的问题中的关键步骤？

- 前人是如何尝试解决这个问题的？他们的解决方法存在什么样的缺陷，最重要的挑战是什么？
- 我们提出什么样的方法在哪一部分可以超越前人？
- 为了实现我们的方法，我们给出了什么样的假设与模型结构，并怎样去度量最后的结果？

## 需要明确的问题

针对这些问题，我们需要在接下来通过综述式的广泛阅读相关论文，从而弄清肺结节分割的具体作用与研究背景，找准某个可以突破的方向进行钻研。

同时，Danny Chen 指出我们可以以肺结节数据为一个例子，但是研究方向并不需要一定集中在肺结节上，教授同时为我们提出了新的研究方向。

## 当前问题的发展方向与现有的可用方法

Danny Chen 指出，对于样本量充足的肺结节分割问题我们只需要利用 3D FCN 以及 Graph Search 这两种成熟的技术基本就可以解决分割问题。这是一项适合入门的工程任务，可以帮助我们掌握神经网络的训练技巧以及熟练使用 Graph Search 解决问题。但是在科研方面，这不是重大创新。因此我们需要把注意力放到更重要的关联问题上去。教授提出了一个重点问题：如何解决训练集不足的问题？这个问题可以分为三个方面

- 如何解决训练集不足导致的模型训练问题
- 如何解决训练集不足使得整个数据空间缺失了某种特征的问题
- 如何增广训练集

## 当前问题的发展方向与现有的可用方法

这些问题是我们下一步研究的重点，而 Danny 教授也提出了以下可以探索的方向

- 直接采用模型本身进行样本空间增广

Danny Chen 教授提出的主要思路是可以先用已有样本进行模型训练，然后用已有的模型对没有标注的数据生成预测标注，并利用这些具有预测标注的数据进行模型训练。这个思想主要侧重于从新的数据中学习原数据没有的特征，在标注不是很正确的情况下进行样本空间的扩展，使得模型能接收更多特征的输入。但是它需要完善的地方是因为后面的数据所对应的标注都是初始数据所训练出模型的预测，因此这种做法很可能会让模型不断学习自己的误差。

- 采用 Graph Search 方法优化模型标注

Danny Chen 同时提出，我们也可以利用 Graph Search 的方法对原有的模型的标注进行优化，同时对上面方法中的预测标注进行优化。

- 采用多种模型来补全 MISS 的特征

对于同一个问题而言，不同的模型可能对于不同的特征有侧重，对于数据空间特征的 MISS 问题，有一些模型可以根据自身的泛化能力对缺失特征进行部分识别，而综合多个模型可以起到补全缺失特征的方法。因此可以选取多个模型进行缺失特征补全，同时利用多个模型生成无标注数据的预测标注，并用预测标注加入主要模型使得模型可以通过这种方法学习到缺失的特征

# 当前问题的发展方向与现有的可用方法

这些问题是我们下一步研究的重点，而 Danny 教授也提出了以下可以探索的方向

- 采用 Deep Adversarial Network 思想进行 Unannotated 数据学习

具体方法已在论文报告中进行叙述

- 采用 Suggest Annotation 方法寻求标注上的帮助

将 Suggest Annotation 方法扩展到 3D 也是我们正在做的方向

- 采用迁移学习的方法来获得 Miss 的特征

我认为我们同时也可以尝试迁移学习的方法，将其它类型数据集中出现的相似特征进行迁移学习的方法纳入模型中。

## 下一步的具体任务

在接下来的工作中，我们首先需要进行肺结节分割相关论文的综述性泛读以了解问题背景与主要困难。

同时我们要尝试训练一个 work 的 3D 语义分割网络，以及在我们的数据集上使用 Graph Search 方法，这两项都是亟需补足的工程能力。

在长远目标上，我们下一步会以 Suggest Annotation 为基础来进行 Danny 所提出的方向的探索与解决。