

3D U-Net Learning Dense Volumetric Segmentation from Dense Annotation

- **Abstract**

文章介绍了一种从稀疏标记的体积图像中进行体积分割的网络结构。分为两种方法：1，半自动化的形式，用户标记3D体积中的一些切片，网络进行学习并提供一种密集3D分割。2，全自动化形式，假设有一个具有代表性的稀疏标记的训练集，训练后网络能分割出一个图像。这是以前2D U-Net 的扩展，将2D操作替换成了3D操作。训练过程中在线进行弹性形变数据增强。该网络是端到端从头开始训练的，没有预训练模型。两种方法的测试性能都很好。

1. Introduction

生物医学数据分析中，3D体数据非常丰富。但是标记非常困难，枯燥，低效。为创建一个泛化性能非常好的巨大训练数据集，对3D数据进行完整的标记实在不是一个高效的方法。

2 Volumetric Segmentation with the 3D U-Net

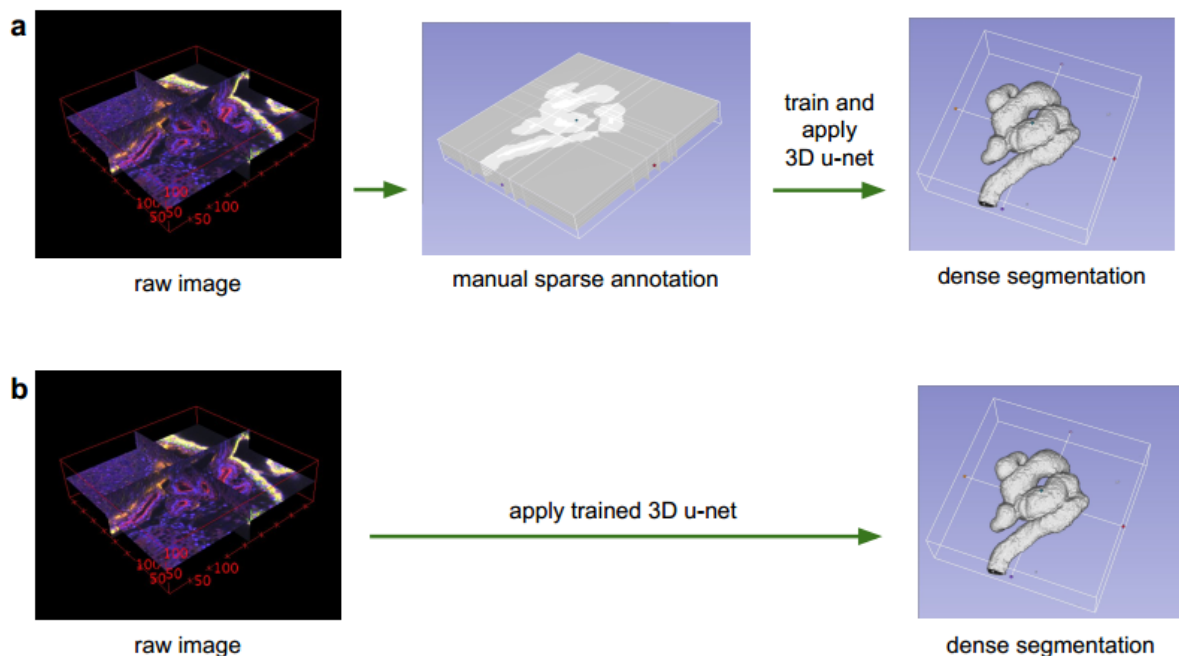


Fig.1: Application scenarios for volumetric segmentation with the 3D u-net. (a) Semi-automated segmentation: the user annotates some slices of each volume to be segmented. The network predicts the dense segmentation. (b) Fully-automated segmentation: the network is trained with annotated slices from a representative training set and can be run on non-annotated volumes.

本文给出了一个训练过程中只需要标记一些2D数据切片就能进行密集体积分割的深度网络结构。这种网络有两种使用方法，1：第一种方法旨在使稀疏标记的数据集稠密化（？？难道是将稀疏标记变成稠密标记（完整标记？？））；2：从多个稀疏标记的数据集中进行学习并产生一个新的数据。两种方法是强相关的。网络是基于之前的U-Net结构，输入是3D的体，相关操作也是3D的，例如，3D卷积，3D最大池化，3D反卷积。没有使用bottlenecks，但是使用了Batch Normalization加速收敛。生物医学图像应用中，训练一般需要很少数据就可以，这是因为每一个图像都包含了某个结构相关变化的多个结构（也就是包含了一个结构的多种形态）。这对体积图像更加显著，所以我们可以利用两个体积图像进行训练从而得到第三个体积。由于网络使用了加权的损失函数和特殊的数据增强方法，所以能够使用很少的标记切片进行训练。我们将这种方法成功应用在了很难对焦的显微镜数据上。首先我们展示了一些标记数据的稠密度的性能。这些结果支持定量评价（？？）。另外还做了一个关于标记切片数量对于网络性能的影响。

1.1 Related Work 现在CNN技术使得2D生物医学图像的分割精度接近了人类的水平。吧啦吧啦（各种之前的成果，略）。我们的网络结构是基于2D的U-Net结构，归功于U-net结构和各种数据增强方法，我们的网络能够从很少的标记样本中学习到很好的泛化性能。另外还基于这样的事实，对生物医学图像进行正确的刚性变换和轻微的弹性形变并不会改变其合理性（也就是说经过这些改变，生物医学图像并不会丢失原图中的信息）。本文的重点是能够在稀疏标记的体积上从头开始训练，并能对任意大小的体积进行训练，这主要是由于无缝拼接策略。

4 Volumetric Segmentation with the 3D U-Net

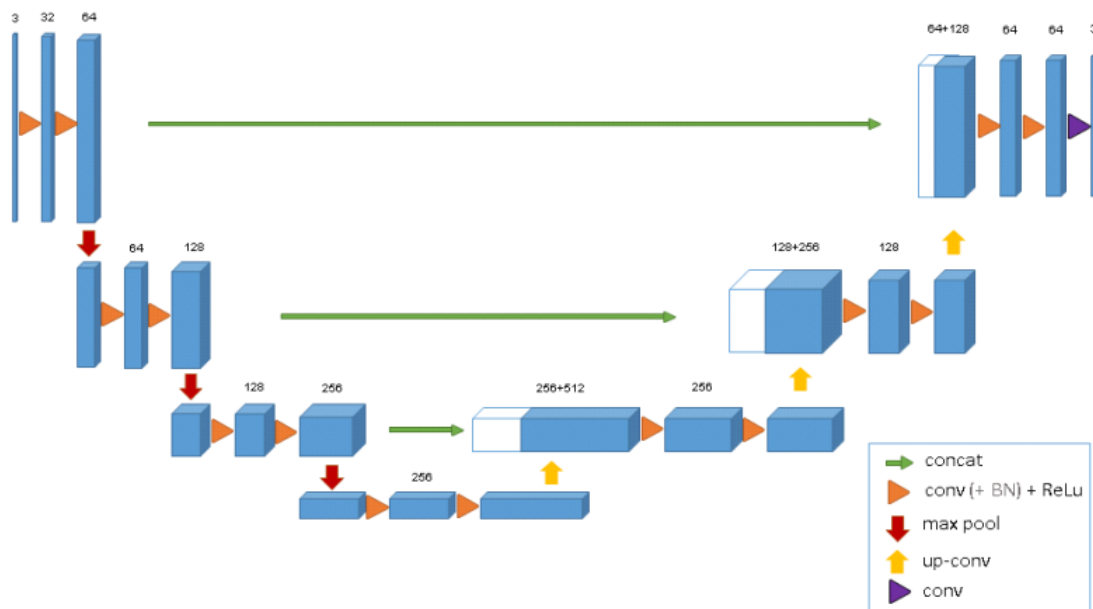


Fig. 2: The 3D u-net architecture. Blue boxes represent feature maps. The number of channels is denoted above each feature map.

2. New Architecture 结构如图2，包括分析和合成两个路径，每个路径有四个降采样步骤。分析路径中每一层包含两个3x3x3的卷积，之后是Relu，再往后是每个维度步长都是2的2x2x2的最大池化。在合成路径，每层包含一个反卷积，卷积核是2x2x2，每一个维度上的步长为2，之后是两个3x3x3的卷积，之后是Relu。具有相同分辨率的分析路径和合成路径之间有快捷连接，用于将分析路径中必要的高分辨率特征传递到合成路径中去。最后是一个1x1x1的卷积，用于将输出通道数降低到标签数，本文的标签数是3。整个结构有19069955个参数，分析路径中，在最大池化之前都将通道数加倍。网络输入是132x132x116，通道数为3的体素块。最后一层的输出是44x44x28的体素，x, y, z 顺序。体素大小是1.76x1.76x2.04 立方微米，在预测分割中的每一个体素就变成155x155x180立方微米（？？这里不懂）。所以每一个输出体素都有足够的上下文信息来进行学习。在【4】中每一个batch是通过这样的方法进行正则化的。但是我们只有一个batch的数据，很少的样本。所以在测试期间使用当前的统计信息也有很好的效果（？？？）。网络的最重要的部分是加权的softmax损失函数。将没有标记的像素

权值设置为0，这样就使网络只在标记的样本中进行学习，并泛化到整个像素体积（？）。**3 Implementatin Details** 3.1 Data 共有三个样本，3D数据记录在4个小片上，每一个小片都是三通道的。体素大小是 $0.88 \times 0.88 \times 1.02$ 立方微米。我们将这些小片拼接在一起，第一个通道是....第二个通道是....第三个通道是.....。我们手动标记了xy, xz, yz切片。标记位置尽可能均匀的分布在三个维度上。在肾小管内部的标记为“0”，肾小管标记为“1”，背景标记为“2”，没有标签的标记为“3”。那些所有像素都没有被标记的切片就标记为“3”。每个维度上都进行了2个像素的下采样。所以实验中使用的三个样本数据在x X y X z方向上分别是 $248 \times 248 \times 64$, $245 \times 244 \times 56$ 和 $246 \times 244 \times 59$ 。三个样本在 (yz, xz, xy) 上的标记切片数量分别是 (7, 5, 21,) , (6, 7, 12,) 和 (4, 5, 10) 。

3.2 Training

我们在数据和标签上除了使用旋转，缩放，灰度变换等增强方法外，还应用了一种平滑密集形变场。为此，在每一个维度上spacing都为32个体素的网格中进行标准差为4的正态分布采样（？？？），之后再行B样条插值。网络输出和标签之间使用加权交叉熵损失的softmax函数，这个加权交叉熵损失函数减少背景的权重，增加肾小管的权重，这样就使肾小管和背景体素对loss的影响达到均衡。标签为3的权值为0，对loss计算不起作用。训练使用随机梯度下降，使用了cuDNN，在线进行数据增强。在TitanX GPU上共训练了70000个迭代，花费了三天。

4 Experiments

4.1 Semi-Automated Segmentation

对于半自动分割，我们假设用户需要对一些3维图像进行完整分割，并且用户没有进行先验分割。用户在每一个3维图像中标记了一些切片，之后使用网络进行密集体积分割。

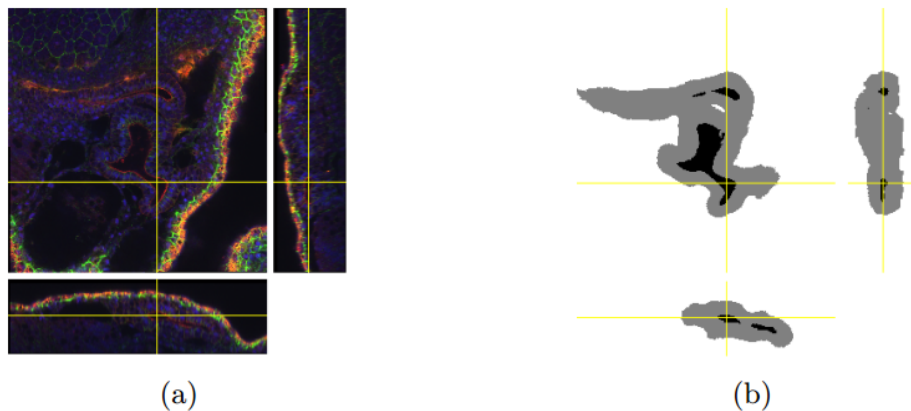


Fig. 3: **(a)** The confocal recording of our 3rd *Xenopus* kidney. **(b)** Resulting dense segmentation from the proposed 3D u-net with batch normalization.

为了进行定性评估，我们在所有的三个稀疏标记样本中进行训练。图3展示了第三个样本的分割结果。为了进行量化评估半自动分割性能，我们将三个样本中的77个切片均分成了三个子集进行交叉验证，分为使用Batch Normalization和不使用Batch Normalization 两种情况。为此，筛出测试切片并没有标记标签。这模拟了用户对样本提供更加稀疏的标记。我们将使用完整3D上下文信息得到的结果和只使用2D图像的结果进行比较，比较结果在表1。比较精度使用IOU，定义： $\text{true positives} / (\text{true positives} + \text{false negatives} + \text{false positives})$ ，结果表明我们的方法能够使用很少的标记切片生成一个精确的3D分割。我们还分析了标记切片数量对于网络性能的影响。为此，模拟了一个样本的半自动分割。开始在每一个正交方向上只使用一个标记切片，之后再慢慢增加切片数量。表2给出了每一个样本 (s1, s2, s3) 在不同“GT”切片数量时网络的性能增益。该网络利用Batch Normalization训练了10个小时。实验的每一步都没有使用测试切片。

Table 1: Cross validation results for semi-automated segmentation (IoU)

test	3D	3D	2D
slices	w/o BN	with BN	with BN
subset 1	0.822	0.855	0.785
subset 2	0.857	0.871	0.820
subset 3	0.846	0.863	0.782
average	0.842	0.863	0.796

Table 2: Effect of # of slices for semi-automated segmentation (IoU)

GT	GT	IoU	IoU	IoU
slices	voxels	S1	S2	S3
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

Table 3: Cross validation results for fully-automated segmentation (IoU)

test	3D	3D	2D
volume	w/o BN	with BN	with BN
1	0.655	0.761	0.619
2	0.734	0.798	0.698
3	0.779	0.554	0.325
average	0.723	0.704	0.547

4.2 Fully-automated Segmentation

全自动分割假设用户想要分割大量的图像。我们假设一个具有代表性的训练集能够被组装起来。

为了评估性能，我们使用两个进行训练，对第三个进行分割。我们对三种组合都进行测试，表3展示了IOU的结果。本次实验中，BN只有在第三次实验中没有起作用，其他两次都对提高结果起了作用。这主要是因为数据集之间的不同造成的。全自动分割在大数据集上作用能加明显，这要是因为相同的数据标签将会分布在更多的数据上，从而能够获得更具代表性的训练集。

5 Conclusion

略