

U-net

Abstract

一般认为成功训练一个深度网络需要上千个标记的训练样本。本文给出了一个新的网络结构和策略，这种策略依靠数据增强能够更加高效的使用标记样本。这种网络结构包括一个用于获取上下文信息的contracting部分和能够进行经确定位的expanding部分，两者呈对称结构。另外这个网络是个端到端的，分割非常快，一种512*512的图像在GPU上分割少于1秒。

1.

深度卷积网络很早就出现，但是由于训练样本数量太少和对应的网络结构难以训练的限制，深度卷积网络一直没有成功。2012年的Alex Net 有8层，一百万个参数的训练打破了这个限制。

CNN一般用于分类，但是对于像生物学图像处理还需要定位，并且大量的生物学图像很难获得。

之前有人进行patch级别的像素分类和定位。但是这样有两个缺点：（1）非常慢，因为每个patch需要单独训练测试，并且由于patch之间的重叠使得数据冗余。（2）需要在定位精度和上下文信息之间做出权衡，因为大的patch需要很多的max-pooling，这就损失了定位精度，小的patch就损失了上下文信息。

本文中设计了一个更加简洁的FCN网络。对FCN进行了改进和扩展，具体可以查看

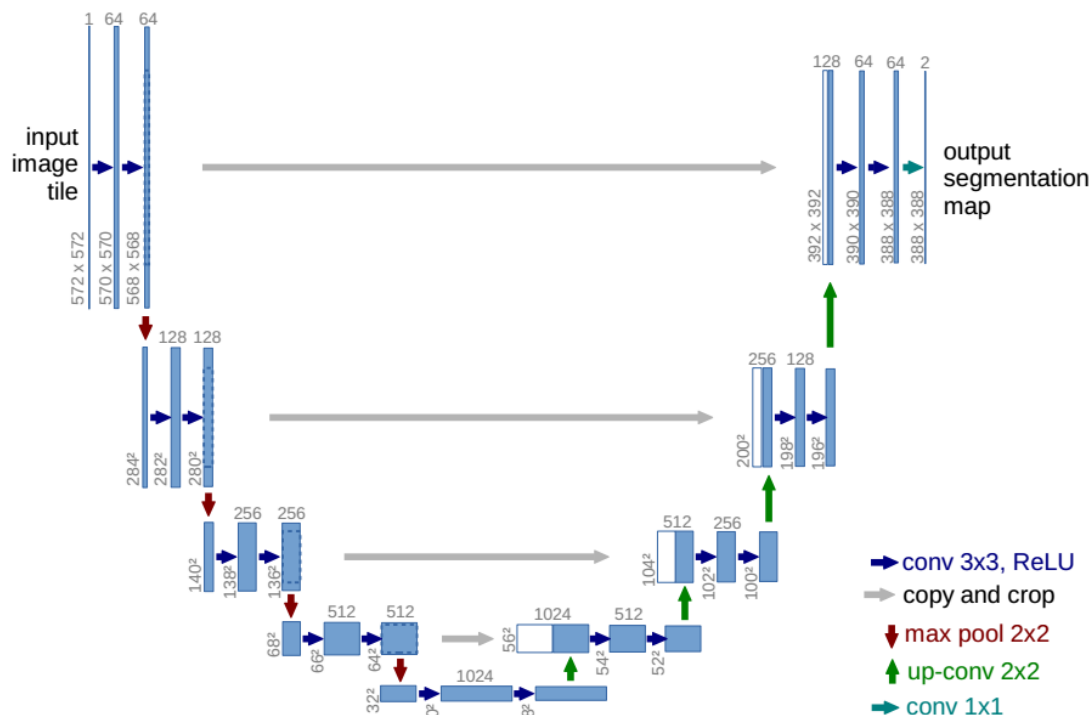


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

主要思想就是增加一个contracting 网络结构，并用上采样代替pooling操作。这个结构的主要作用就是增加输出的分辨率。为了使用局部信息，在网络收缩过程中产生的高分辨率特征，被连接到了修改后网络的上采样的结果上。基于这些信息，连续的卷积层能够学习到更加精确地输出。

与FCN相比，文中的上采样部分同样有大量的特征通道，这就能将上下文信息传播到高层。（这里的意思是不是expanding部分都采用了反卷积，相对于FCN只是将最后的全连接部分改成了反卷积）U-net只使用了每一个卷积的有效部分，并没有使用全连接层，也就是说分割图（特征

图??)只包含能够从输入图像中获得上下文信息的像素点(??)。这个策略通过一个重叠片方法能够对任意大小的图像进行无损分割,重叠片方法可以查看

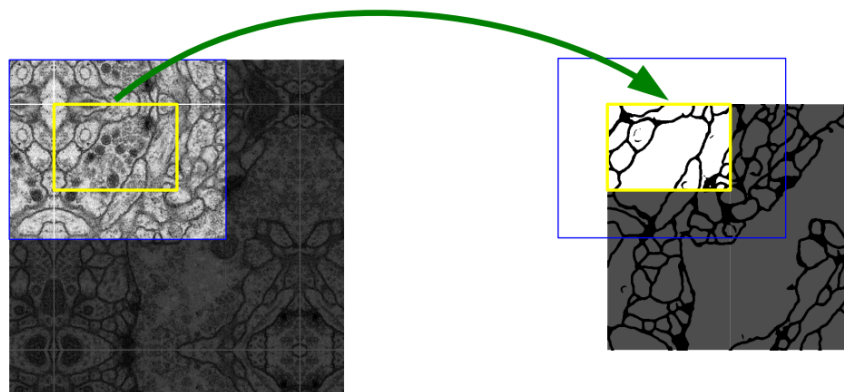


Fig. 2. Overlap-tile strategy for seamless segmentation of arbitrary large images (here segmentation of neuronal structures in EM stacks). Prediction of the segmentation in the yellow area, requires image data within the blue area as input. Missing input data is extrapolated by mirroring

对于图像边界区域的预测是通过对输入图像做镜像对称获得丢失的上下文信息的。这个方法对于网络大的输入图像非常重要,因为如果不这样的话分辨率将会受到GPU内存的限制(??)。

由于训练数据非常少,所以需要利用弹性形变方法对训练数据进行数据增强,这能够使网络学习到对各种形变的不变性。这对于生物医学图像非常有用,因为生物医学图像存在非常多的形变,通过数据增强能够有效地模拟实际中的这种形变。

细胞分割分割的另一个难点是同类目标的黏连,为此,本文建议使用加权的损失函数,也就是在黏连细胞之间分割的背景在损失函数中获得大的权重。

2.网络结构

网络结构由收缩路径(左边的部分)和扩张路径(右边的部分)组成。收缩路径遵循典型的卷积网络结构,包括多个重复出现的3X3的卷积(没有padding),每个卷积后面是一个Relu和一个用于下采样的2X2的最大池化层,池化步长为2。每经过一个下采样特征通道就加倍。扩张路径的每一步都包括上采样层,上采样后面是一个2X2的卷积层(上采样+卷积=反卷积,整体思想就是扩大特征图的分辨率,减少特征图的数量),将得到的结果和收缩路径中对应的特征图连接到一起(收缩路径中的特征图偏大,需要crop),之后就是两个3X3的卷积和Relu层。最后是一个1X1的卷积用于将每一个64通道的特征向量映射到所需的特征类数量。这个网络总共包括23个卷积层。

3.训练

使用随机梯度下降法训练,在Caffe框架下实现。动量设置为0.99,所以之前的训练样本对当前优化更新起着很重要的作用。

在最后的特征图上使用像素级的soft-max和交叉熵。交叉熵惩罚每一个位置相对于1的偏差。交叉熵中的权值 $\omega(x)$ 在训练过程中改变一些像素的重要性。

为了补偿训练集中某一类像素的频率差异(i.e.粘连的细胞之间的背景像素频率很少),我们为每一个金标准分割图制作了权值图,这就强迫网络去学习我们之前介绍的粘连细胞之间的小的分离边界。

分割边界是通过形态学操作获得的。权值图是通过以下公式获得：

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

where $w_c : \Omega \rightarrow \mathbb{R}$ is the weight map to balance the class frequencies, $d_1 : \Omega \rightarrow \mathbb{R}$ denotes the distance to the border of the nearest cell and $d_2 : \Omega \rightarrow \mathbb{R}$ the distance to the border of the second nearest cell. In our experiments we set $w_0 = 10$ and $\sigma \approx 5$ pixels.

w_c 是用于平衡类别频率的权值图， d_1 :代表到最近的细胞边界的距离， d_2 :代表到第二近细胞边界的距离。 w_0 和 σ 为常数，试验中设置 $w_0=10$ ， $\sigma=5$ 个像素。

深度学习中好的初始化权值非常重要，否则网络的一部分被过度激活，其他部分却没有起作用。理想状态的初始权值是网络中每一个特征图都有近似的单位方差。我们使用的初始权值为标准差为**根下2N**的高斯分布，N为每个神经元的输入节点数量。例如，对于一个上一层是64通道的3*3卷积核来说， $N=9*64=576$

3.1 数据增强

当使用很少的标记样本训练一个分割网络的时候，训练样本的随机弹性形变回一个很关键的概念。在一个3X3的粗糙网格上使用随机位移矢量生成光滑形变。位移从10个像素的标准差的高斯分布中取样。每一个位移使用**双三次插值**计算获得。网络结构的最后使用Drop-out进一步执行隐式的数据增强。

4.实验

训练集是30张512X512的神经元结构图。细胞是白色，细胞膜是黑色，测试集公开，但是测试集的分割mask是保密的。你需要将预测细胞膜的概率图发给组织方，组织方使用10种不同的阈值进行评价，并计算“warping error”等多误差。

U-net在没有进行预处理和后处理的情况下，实现warping误差0.0003529。

另外在“PHC”数据集上，有35张部分标记的训练图像。U-net的平均IOU是92%。在“DIC”上，有20张部分标记的训练图像，U-net的平均IOU是77.5%。

5.总结

U-net在不同的生物医学图像分割应用上都取得了很好的性能。多亏弹性形变的数据增强方法，U-net训练只需要很少的标记图像并且在NVidia Titan GPU（6G）上只花费了10个小时。