U-Net: Convolutional Networks for Biomedical Image Segmentation

Abstract

Introduction

Network Architecture

Training

Data Augmentation

Experiment

Conclusion

之前普遍认为训练深度网络需要数千个标记的训练样本。本文提出了一种网络和训练策略，依赖于对数据扩充的应用，对有限的样本更有效地使用。结构包含获得context的收缩路径(contracting path)和确保精确定位的扩展路径(expanding path)。这种网络可以端到端地进行训练，使用从非常少的图像。网络表现效果很好，速度也很快。

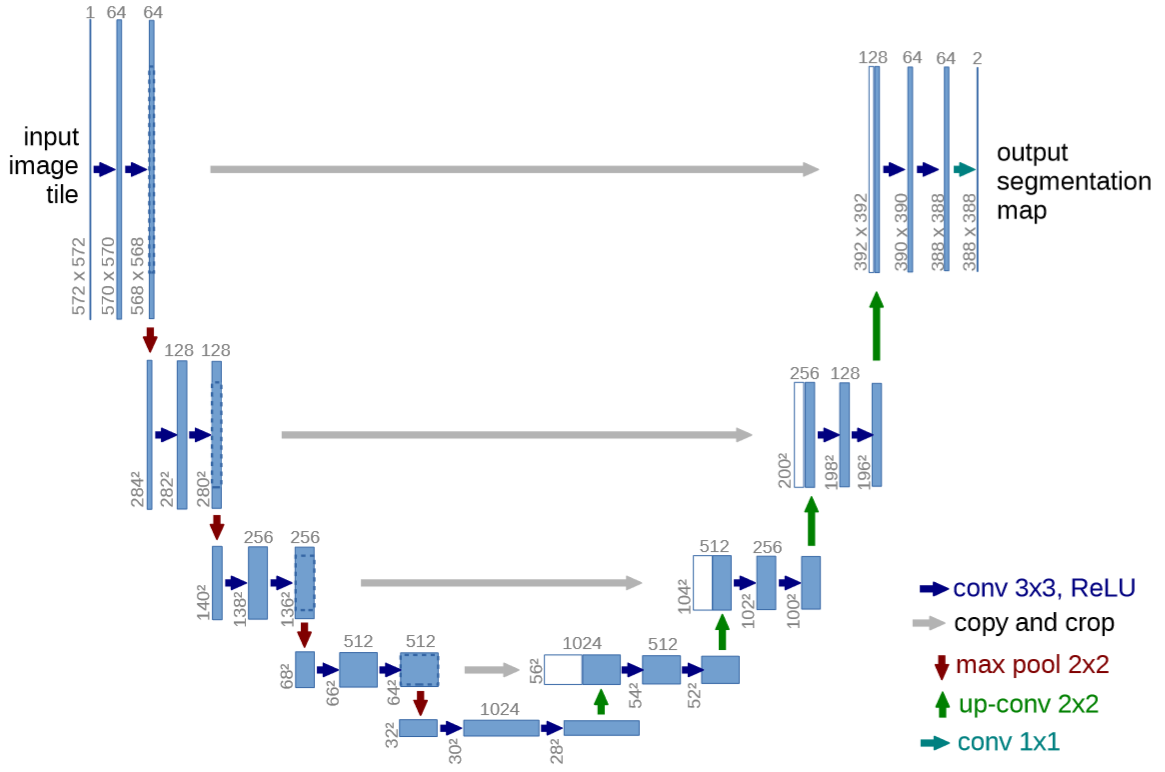
Introduction

在生物医学图像处理中，期望的输出应该包括定位，即，应该将类标签分配给每个像素。

生物医学任务中无法获得很多的训练数据。

滑动窗口卷积网络通过在该像素周围提供局部区域（patch）作为输入来预测每个像素的类别标签。网络可以局部化；补丁方面的训练数据远远大于训练图像的数量。

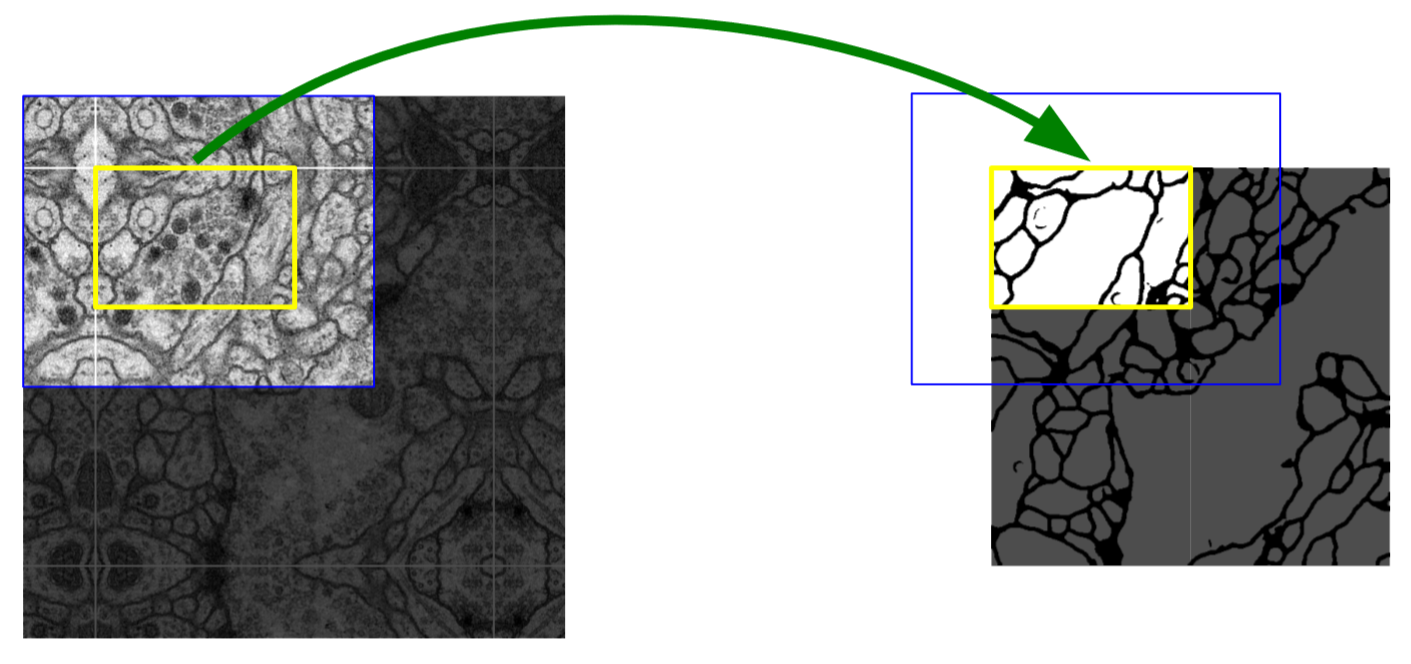
滑动窗口网络的缺点。首先，它非常慢，因为必须为每个补丁单独运行网络，并且由于补丁之间重叠而存在大量冗余。其次，定位精度与内容的使用之间存在着一种权衡。较大的patch需要更多的最大池化层，从而降低了定位精度，而使用小补丁的话，网络就只能看到很少的内容。

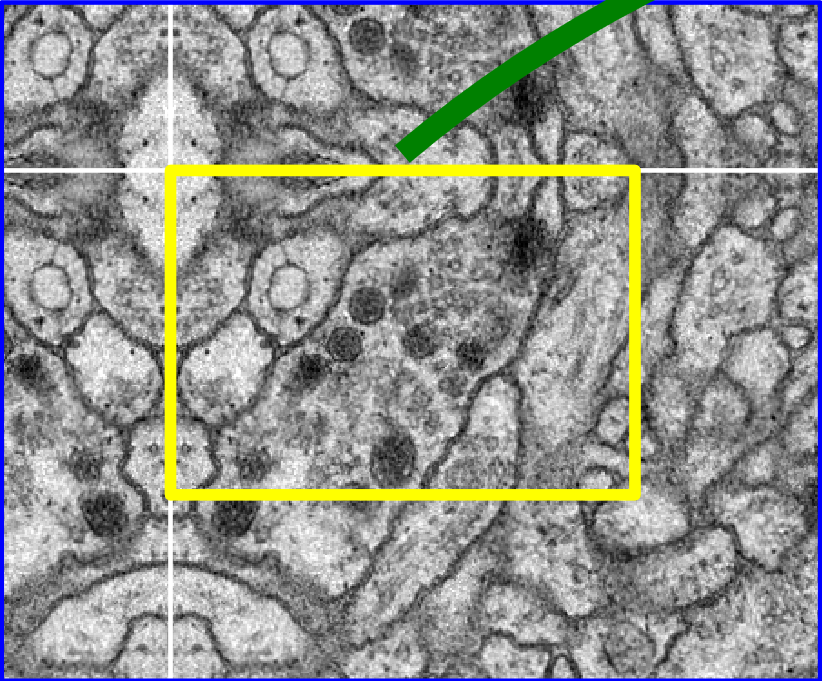


在本文中，我们建立了一个更优雅(elegant)的架构，即所谓的“完全卷积网络”。我们对这种架构进行了修改和扩展，使其能够处理很少的训练图像，并产生更精确的分割。主要思想是通过连续的层来增加常用的的收缩网络，其中池化操作由上采样操作替换。因此，这些层增加了输出的分辨率。为了进行本地化，来自收缩路径的高分辨率特征与上采样输出相结合。然后，连续卷积层可以学习，并基于该信息集合更精确的输出。

我们架构中的一个重要修改是，在上采样部分，我们也有大量的特征通道，允许网络将内容信息传播到更高分辨率的层。因此，扩展路径或多或少地与收缩路径对称，并产生U形结构。网络没有任何完全连接的层，并且仅使用每个卷积的有效部分，即，分割图仅包含在输入图像中可获得完整内容的像素。该策略允许通过重叠区块策略(overlap-tile)无缝分割任意大的图像为了预测图像的边界区域中的像素，通过镜像输入图像来推测丢失的内容。这种平铺策略对于将网络应用于大图像很重要，否则分辨率将受到GPU内存的限制。

至于我们的任务，可用的训练数据非常少，我们通过对可用的训练图像应用弹性形变来进行数据扩充。这允许网络学习这种形变的不变性，而不需要在标注的图像语料库中看到这些变换。这在生物医学分割中尤其重要，因为形变是组织中最常见的变化，并且可以有效地模拟真实的变形。





Network Architecture

网络架构如图1所示。它由一个收缩路径（左侧）和一个扩展路径（右侧）组成。收缩路径遵循卷积网络的典型架构。它包括重复应用两个3x3卷积（没有padding），每个卷积后跟一个线性修正单元（ReLU）和一个2x2最大池化操作，步长2用于下采样。在每个下采样步骤中，我们将特征通道的数量加倍。扩展路径中的每一步都包括对特征图进行上采样，然后进行2x2卷积（“up-convolution”），将特征通道数量减半，与来自收缩路径的相应裁剪特征图连结，以及两个3x3卷积，还有ReLU。由于每个卷积中边界像素的丢失，裁剪是必要的。在最终层，使用1x1卷积将每个64分量特征向量映射到所需数量的类。总的来说，网络有23个卷积层。

为了实现输出分割图的无缝平铺（参见图2），选择输入图块大小非常重要，这样所有2x2最大池操作都应用于具有偶数x和y尺寸的图层。

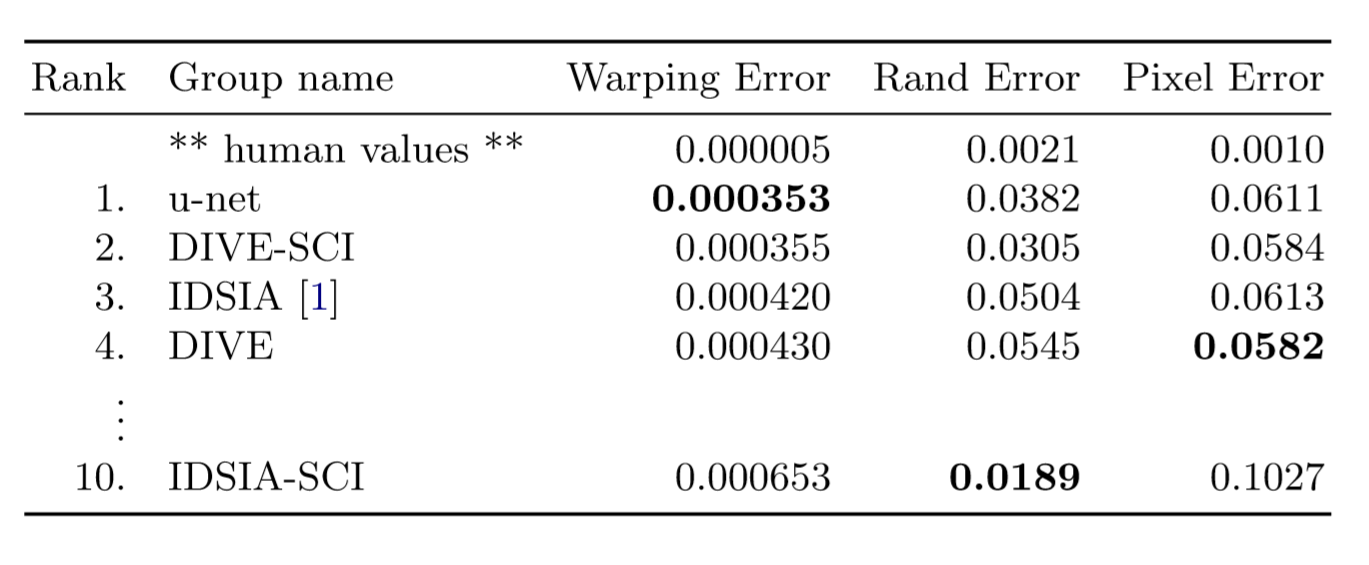
Training

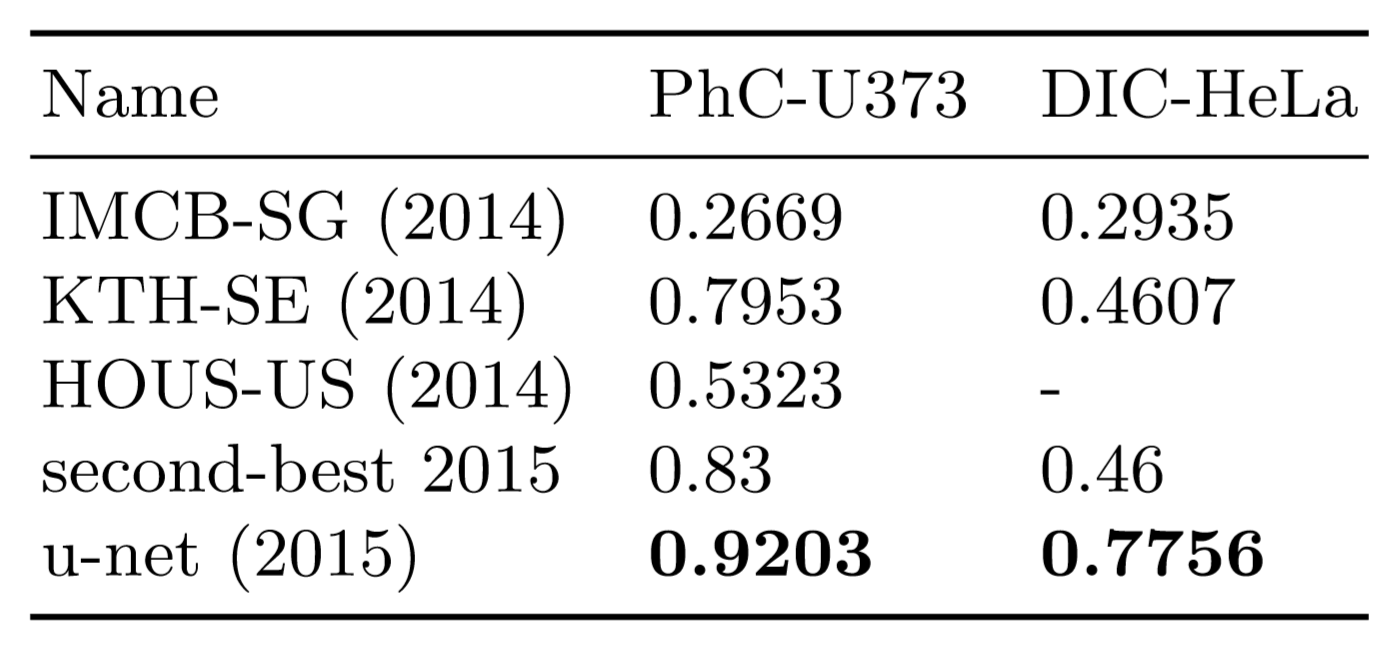
Data Augmentation

在训练样本很少的时候，数据扩充对于网络学习不变性和鲁棒性十分重要

在显微镜图片中我们主要需要移位和旋转不变性以及对形变和灰度值变化的鲁棒性。 特别是训练样本的随机弹性形变似乎是训练具有极少注释图像的分割网络的关键概念。 我们使用粗糙的3×3网格上的随机位移矢量生成平滑变形。 位移是从具有10个像素标准差的高斯分布中采样的。 然后使用双三次插值计算每像素的位移。

Experiment





The ﬁrst data set “PhC-U373”2 contains Glioblastoma-astrocytoma U373 cells on a polyacrylimide substrate recorded by phase contrast microscopy (see Figure 4a,b and Supp. Material). It contains 35 partially annotated training images.