

# 文献阅读报告 I——U-Net

黄磊 计 702 2022E8013282156

## Short Summary

U-Net 是一种无监督深度神经网络，用于解决图像分割问题。它的主要思想是利用卷积神经网络和随机搜索技术，将输入图像分割为前景和背景，并在分割过程中保持完整性和连续性。其使用了自编码器和反向传播器来实现这一目标，在编码器和解码器中构建出一个完整的图像分割模型，从而有效提取图像的前景和背景。U-Net 在图像分割中表现出色，并能够保持图像的连续性和细节的完整性，被广泛应用于医学影像分析、自然语言处理、计算机视觉等领域。

## Introduction/background

在过去几年里，深度卷积网络不仅广泛地应用于各种视觉识别任务，同时也达到了非常好的效果。然而由于医学图像可用训练集和网络大小的限制，卷积网络难以取得更好的性能。

一些代表性方法（例如 Ciresan 提出的基于滑动窗口的网络）存在计算量大、运行效率低下的问题，并且在上下文使用和定位精度之间存在折衷。为了解决这一问题，人们开始考虑结合来自多层的特征，从而同时实现好的定位精度与上下文的使用，因此全卷积网络（FCN）被提出。其主要思想是利用连续层辅助收缩网络（contracting network，或者称为下采样网络），其中 pooling 运算被替换为上采样运算以增加输出分辨率。最后融合高分辨率特征与上采样输出，获得更精确的预测结果。

本文所提出的 U-Net 通过优化和扩展全卷积网络，使之从少量训练图像中能学习更精确的分割。

## Detailed summary

### 1. Network Architecture

U-Net 本质上是一个自编码器，由收缩路径（extracting path）和扩张路径（expansion path）组成，形似一个“U”，因此被称作 U-Net。

收缩路径对应左半区域，主要是进行下采样操作，其中重复使用了两个  $3 \times 3$  无 padding 卷积核，每个卷积后面跟随一个 ReLU 和步长为 2 的池化操作，以实现下采样。并且，在每次下采样的时候都会加倍特征通道数量。最低分辨率为  $32 \times 32$ 。

扩张路径对应右半区域，主要进行上采样操作。在上采样部分仍有大量的特征通道，这些特征通道允许网络将上下文信息传播到高分辨率层，因此扩张路径大部分相对于收缩路径对称，进而产生 U 形结构。该区域每层都包含一个  $2 \times 2$  反的卷积操作用于上采样，该操作会减半特征通道。同时包含了与同维度层剪切后的特征图的融合操作，其后添加两个带有 ReLU 的  $3 \times 3$  卷积。并且由于我们使用了无 padding 卷积，因此每次卷积中都会丢失边界像素，因此有必要进行裁剪。

最后一层，使用  $1 \times 1$  卷积核将含有 64 分量的特征向量映射到期望数量的类。

整体网络共有 23 个卷积层，不包含任何全连接层并且只使用卷积中的有效部分，这使得上下文信息均可以从图像中得到，在分割图中仅包含像素信息。因此，对于任意大的图像，可以使用 overlap-tile 的方法实现无缝分割；而对于边界的预测，丢失的像素信息可以使用输入图像的镜像补进行补充。

### 2. Training

生物医学图像分割中，可能需要对于相接触的同类的多个对象进行分割，因此训练时候使用了加权损失函数，使接触细胞之间的分离背景的标签在损失函数中获取较大权重，以实现较为精确的前背景分离。

输入图像使用 Caffe 中动量为 0.99 的随机梯度下降算法实现，每个 batch 输入一张图片，损失函数使用带权重的 softmax 和交叉熵损失：

$$p_k(x) = \frac{\exp(a_k(x))}{\sum_{k'=1}^K \exp(a_{k'}(x))}$$

其中， $K$  表示类别数量， $a_k(x)$  表示像素  $x$  在特征通道  $k$  上的激活值。该公式得到像素在某个特征通道的估计最大值，即给出像素  $x$  属于哪个特征通道。

权重通过以下公式给出：

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

可以看出，权重使用高斯分布进行初始化， $d_1(x)$  表示与像素  $x$  最近的对象（例如细胞）的距离， $d_2(x)$  表示与像素  $x$  第二近的对象（例如细胞）的距离， $w_c(x)$  用于平衡类别频率的权重。可以看出，当  $x$  离细胞越远，第二项越小，且逐渐趋近于 0，因此远离细胞的像素的权重相差不大，接近于  $w_c(x)$ ；相反，细胞边界周围的像素点权重则随着距离变化较为剧烈，从而进行区分。

惩罚函数用于防止过拟合，定义如下，其中  $l(x)$  是每个像素的真实类别，

$$E = \sum_{x \in \Omega} w(x) \log(p_{l(x)}(x))$$

### 3. Data Augmentation

由于医学图像数据量的限制，我们需要实现小数据样本训练获取较高性能，就必须使用数据增强以扩展数据量和多样性。但在做数据增强时候，需要满足网络关于数据具有平移和旋转不变性、以及对于变形和灰度值变化的鲁棒性。文中使用  $\sigma = 10$  的高斯分布采样随机位移矢量，之后在粗糙的  $3 \times 3$  网格上生成平滑变形，最后使用双三次插值计算每个像素的位移，以此实现随机弹性形变。

## Experiments

U-Net 在电子显微镜生成的神经元分割任务和光显微镜图像中的细胞分割任务（后者包含相差显微镜和差分干涉显微镜图像）均取得了非常高的性能指标，说明 U-Net 在医学图像的分割领域任务中表现十分优秀，在数据集有限的情况下依然能取得非常好的分割效果。

## References

- [1] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, Vol.9351: 234-241, 2015, available at arXiv:1505.04597 [cs.CV]