

UNet++: A Nested U-Net Architecture for Medical Image Segmentation

Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, et al.

摘要

在本文中，我们提出了 UNet++，一个用于医学图像分割的全新更强大的架构。我们的架构本质上是一个深度监督的编码器-解码器网络，其中编码器和解码器子网络通过一系列密集而嵌套的跳跃连接。重新设计的跳过路径旨在减少编码器和解码器子网络的特征图之间的语义差距。我们认为，当解码器和编码器网络的特征图在语义上相似时，优化器会更容易处理学习任务。我们在多个医学图像分割任务中对 UNet++ 与 U-Net 和 wide U-Net 架构进行了评估：胸部低剂量 CT 扫描中的结节分割、显微镜图像中的细胞核分割、腹部 CT 扫描中的肝脏分割以及结肠镜检查视频中的息肉分割。我们的实验表明，带有深度监督的 UNet++ 比 U-Net 和 wide U-Net 分别实现了 3.9 和 3.4 分的平均 IoU 增益。

关键词：图像分割；医学图像；U-Net

1 引言

最先进的图像分割模型是编码器-解码器结构的变种，如 U-Net 和全卷积网络（FCN）。这些用于分割的编码器-解码器网络有一个关键的相似之处：跳跃连接，它将解码器子网络的深层、语义、粗粒度的特征图与编码器子网络的浅层、低层、细粒度的特征图相结合。事实证明，跳跃连接在恢复目标物体的细粒度细节方面是有效的；即使在复杂的背景上也能生成具有细粒度的分割掩码。跳跃连接也是实例级分割模型（如 Mask-RCNN）成功的基础，它可以对被遮挡的物体进行分割。可以说，自然图像中的图像分割已经达到了令人满意的水平，但这些模型还不能满足医学图像的严格分割要求。

医学图像中的病变或异常情况的分割要求比自然图像中的精确度更高。虽然精确的分割掩码在自然图像中可能并不关键，但在医学图像中，即使是边缘性的分割错误也会导致临床环境中糟糕的用户体验。例如，结节周围细微的刺状图案可能表明结节的恶性。因此，从临床角度看，将它们从分割掩码中排除会降低模型的可信度。此外，不准确的分割也可能导致随后的计算机生成的诊断出现重大变化。例如，在纵向研究中，对结节生长的错误测量会导致对筛查病人分配一个错误的肺结节类别。因此，人们希望设计出更有效的图像分割架构，以有效恢复医学图像中目标对象的精细细节。

为了解决医学图像中更精确的分割需求，我们提出了 UNet++，一个基于嵌套和密集跳跃连接的新分割架构。我们架构背后的基本假设是，当来自编码器网络的高分辨率特征图在与来自解码器网络的相应富含语义的特征图融合之前，该模型可以更有效地捕获前景对象的细粒度细节。我们认为，当解码器和编码器网络的特征图在语义上相似时，网络会更容易处理学习任务。这与 U-Net 中常用的普通跳跃连接不同，后者直接将高分辨率的特征图从编码器网络快速转到解码器网络，导致语义不同的特征图被融合在一起。根据我们的实验，我们的架构是有效的，比 U-Net 和 wide U-Net 有明显的性能提升。

2 相关工作

Long 等人^[1]首先介绍了全卷积网络 (FCN)，而 UNet 是由 Ronneberger 等人^[2]介绍的。它们都有一个关键的想法：跳跃连接。在 FCN 中，上采样的特征图与编码器跳跃得到的特征图相加，而 U-Net 将它们串联起来，并在每个上采样步骤之间增加卷积和非线性。事实证明，跳跃的连接有助于在网络输出端恢复完整的空间分辨率，使全卷积方法适用于语义分割。受 DenseNet 架构^[3]的启发，Li 等人提出了用于肝脏和肝脏肿瘤分割的 H-denseunet。本着同样的精神，Drozdzalet 等人系统地研究了跳过连接的重要性，并在编码器内引入了短跳过连接。尽管上述架构之间存在微小的差异，但它们都倾向于融合来自编码器和解码器子网络的语义不同的特征图，根据我们的实验，这可能会降低分割性能。

另外两个最近的相关工作是 GridNet 和 Mask-RCNN。GridNet 是一个编码器-解码器架构，其中的特征图是以网格的方式连接的，融合了几个经典的分割架构。然而，GridNet 在跳过连接之间缺乏上采样层；因此，它并不代表 UNet++。Mask-RCNN 可能是物体检测、分类和分割的最重要的元框架。我们想指出的是，UNet++ 可以很容易地部署为 Mask-RCNN 的骨干架构，只需用我们的嵌套密集跳跃路径取代普通跳跃连接即可。

3 本文方法

3.1 本文方法概述

本文提出的 UNet++ 作为基于 UNet 架构的模型，研究出了一种能提供更加丰富语义信息，捕获更多编码器-解码器的细粒度细节的骨干网络。本质上是一个深度监督的编码器-解码器网络，其中编码器和解码器子网络通过一系列密集而嵌套的跳跃连接。比起直接将高分辨率的特征图从编码器网络快速转到解码器网络，我们可以更好将语义不同的特征图被融合在一起。在医学图像领域，UNet++ 对于病变或异常情况的分割能得到更高的精确度，常用于结肠镜检查视频中的息肉分割、腹部 CT 扫描中的肝脏分割以及显微镜图像中的细胞核分割。由图 3 显示的 U-Net、wide U-Net 和 UNet++ 之间的定性比较，可以看出 UNet++ 在息肉、肝脏和细胞核数据集的分割结果取得更高的准确率。

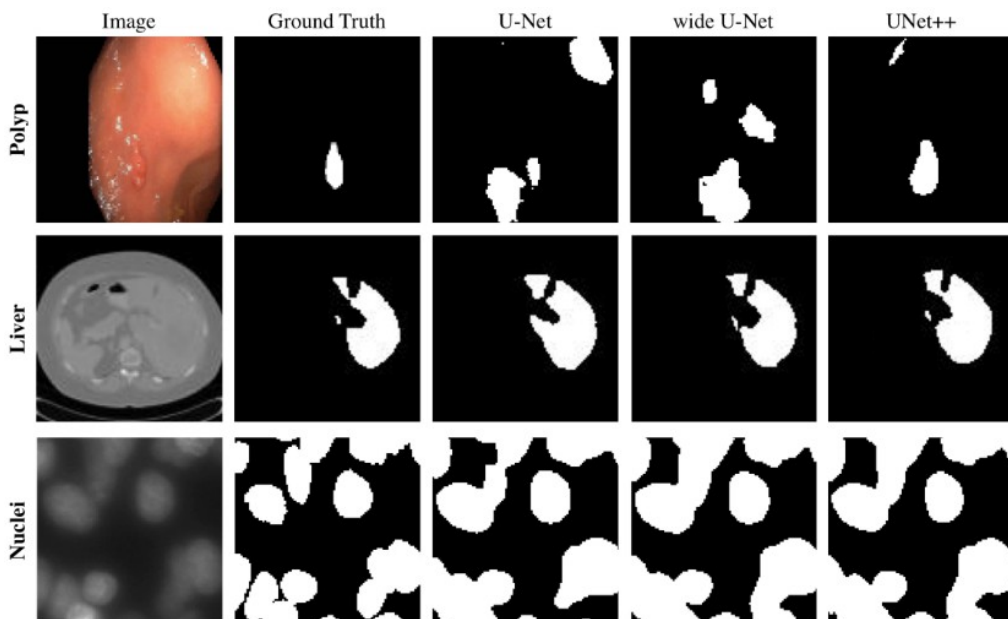


图 1: 息肉、肝脏和细胞核数据集的分割结果

3.2 特征提取模块

重新设计的跳跃路径改变了编码器和解码器子网络的连接。在 U-Net 中，编码器的特征图被直接接收到解码器中；然而，在 UNet++ 中，它们要经过一个密集的卷积块，其卷积层的数量取决于金字塔级别。例如，节点 $X^{0,0}$ 和 $X^{1,3}$ 之间的跳跃路径由一个具有三个卷积层的密集卷积块组成，每个卷积层之前都有一个连接层，将同一密集块的前一个卷积层的输出与较低密集块的相应上采样输出相融合。从本质上讲，密集卷积块使编码器特征图的语义水平更接近于解码器中等待的特征图。我们的假设是，当收到的编码器特征图和相应的解码器特征图在语义上相似时，优化器就只需要处理一个相对容易的优化问题。形式上，我们将跳跃路径表述如下：让 $X^{i,j}$ 表示节点 $X^{i,j}$ 的输出，其中 i 索引是编码器的下采样层， j 索引是跳跃路径的密集块的卷积层。 $X^{i,j}$ 所代表的所有特征图的计算方法：

$$x^{i,j} = \begin{cases} \mathcal{H}(x^{i-1,j}), & j = 0 \\ \mathcal{H}\left(\left[\left[x^{i,k}\right]_{k=0}^{j-1}, \mathcal{U}(x^{i+1,j-1})\right]\right), & j > 0 \end{cases}$$

其中，函数 $\mathcal{H}(-)$ 是一个卷积运算后的激活函数， $\mathcal{U}(-)$ 表示上采样层， $[\]$ 表示连接层。基本上， $j=0$ 层的节点只接收来自编码器前一层的一个输入； $j=1$ 层的节点接收两个输入，都来自编码器子网络，且都是连续的两层； $j>1$ 层的节点接收 $j+1$ 个输入，其中 j 个输入是同一跳过通路中前 j 个节点的输出，最后一个输入是下层跳过通路的上采样输出。我们希望沿着每个跳跃路径利用密集的卷积块，所以需要所有先前的特征图累积起来并到达当前节点。

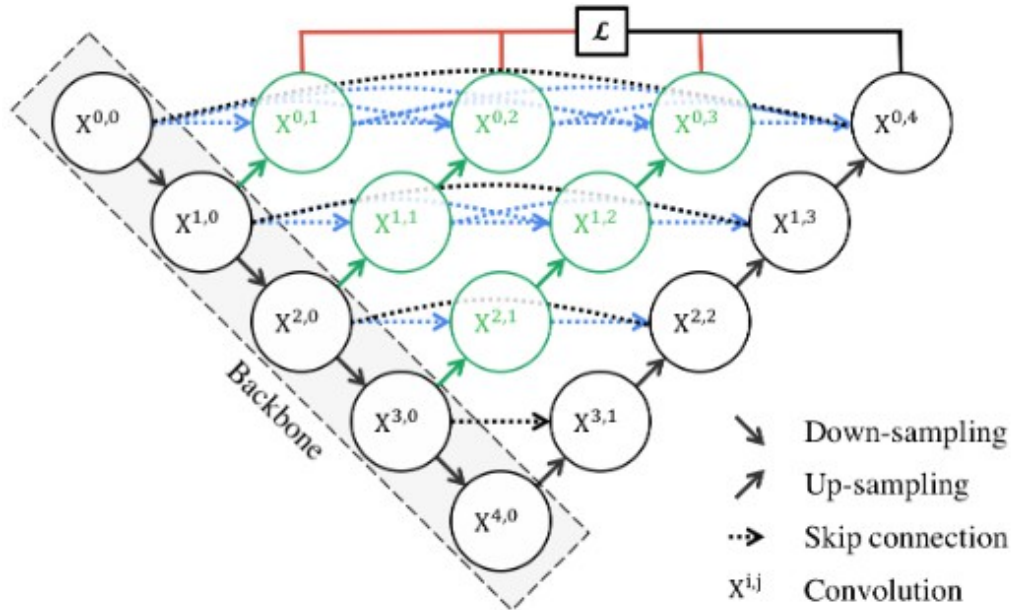


图 2: 由编码器和解码器组成，它们通过一系列嵌套的密集卷积块连接

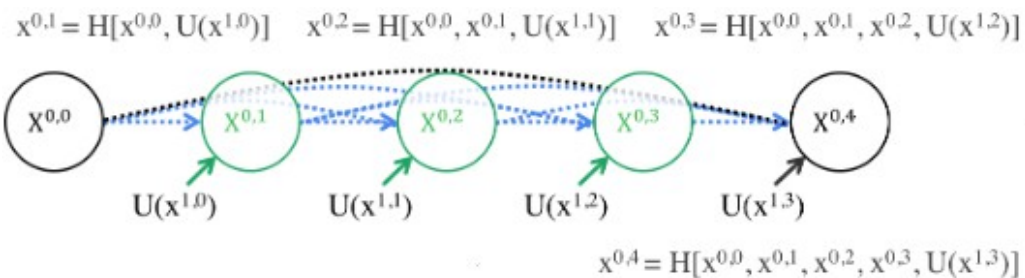


图 3: UNet++ 第一个跳跃路径详解

3.3 损失函数定义

本文使用了 Binary Cross-Entropy 和 Dice 系数作为模型的损失函数。Binary Cross-Entropy 如公式：

$$\mathcal{L}(Y, \hat{Y}) = -\frac{1}{N} \sum_{b=1}^N \left(\frac{1}{2} \cdot Y_b \cdot \log \hat{Y}_b + \frac{2 \cdot Y_b \cdot \hat{Y}_b}{Y_b + \hat{Y}_b} \right)$$

其中 \hat{Y}_b 和 Y_b 分别表示 b^{th} 图像展平后的预测概率和展平后的基本事实，N 表示批大小。

4 复现细节

本次复现是基于 pytorch 框架实现的，代码实现过程也是在熟练掌握论文所提及的重要模块以及相应公式下完成的。在研究初期，为了对复杂模型框架的进一步理解，有研读过源代码模型框架的部分编写思路，而在逐渐深入明白框架原理后，最终代码是使用了我个人所理解方式去编写，各模块的内容也是对照论文所提到的架构顺序编写，而论文没有详细说明了深度监督部分，在本次复现代码也没有实现。

5 实验结果分析

复现最后得到的结果如图 5 下所示，结果所呈现的标准是计算预测值与 ground truth 的重叠面积，并除以预测值与 ground truth 的并集面积，而 IOU 指标则是度量来计算预测的 mask 与 ground truth mask 的准确性。对比论文所给出的实验结果与我本身复现的结果不难看出，UNet++ 模型在医学图像切割领域准确性是极高的，但在数值上与原论文仍有差距，我猜测原因大概是缺少了深度监督部分的提升。

Architecture	Params	Dataset			
		cell nuclei	colon polyp	liver	lung nodule
U-Net [9]	7.76M	90.77	30.08	76.62	71.47
Wide U-Net	9.13M	90.92	30.14	76.58	73.38
UNet++ w/o DS	9.04M	92.63	33.45	79.70	76.44
UNet++ w/ DS	9.04M	92.52	32.12	82.90	77.21

图 4: 原论文实验准确率

```
100%|████████████████████| 9/9 [00:19<00:00, 2.21s/it, loss=0.171, iou=0.833]
loss 0.1980 - iou 0.7997 - val_loss 0.1707 - val_iou 0.8332

Process finished with exit code 0
```

图 5: 复现实验准确率

6 总结与展望

本次复现论文为期时间较长，给了我更充足研究论文细节的机会，加深了我图像切割技术在医学领域的运用，掌握了 UNet 模型以及其在空间上延展的变种 UNet++，让我对模型框架的创新上多了一种新的思路。同时，希望以这次复现论文为基础，阅读更多在医学图像领域的创新论文，以后可以有属于自己的想法去创造新模型。

参考文献

- [1] J. LONG E S, DARRELL. T. Fully convolutional networks for semantic segmentation[J]. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2015: 3431-3440.
- [2] O. RONNEBERGER P F, BROX. T. U-net: Convolutional networks for biomedical image segmentation [J]. In International Conference on Medical image computing and computer-assisted intervention, 2015: 234-241.
- [3] M. DROZDZAL G C, E. Vorontsov. Densely connected convolutional networks[J]. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 3.