

U-Net: Convolutional Networks for Biomedical Image Segmentation

≡ 标签	CV
☑ 状态	已读
📅 时间	2024/04/03
≡ 来源	2015 MICCAI
☑ 复现	已复现
≡ 题目信息	U-Net: 用于生物医学图像分割的卷积网络
≡ 作者信息	Olaf Ronneberger 等，来自德国弗赖堡大学
+ 新增属性	

研究主题：利用深度卷积神经网络进行生物医学图像的分割任务

主要贡献：提出 U-Net

实验总结：

论文中的实验部分展示了 U-Net 在多个生物医学图像分割任务上的应用和性能。具体来说：

1. 在 ISBI 挑战赛中，U-Net 在电子显微镜图像的神经结构分割任务上取得了最佳性能，超越了之前最好的方法。
2. 在 ISBI 细胞跟踪挑战赛 2015 中，U-Net 在两个最具挑战性的 2D 透射光数据集上取得了显著的胜利，大幅领先于其他参赛算法。

内容

研究动机：在生物医学图像处理领域，精确的图像分割对于理解图像内容至关重要。然而，由于可用的标注样本数量有限，传统的深度学习方法在训练时面临挑战。此外，生物医学图像通常具有高度的变异性和复杂性，这增加了分割任务的难度。因此，研究者们需要开发一种能够在有限数据下仍能保持高准确度的图像分割方法。

研究现状：

具体贡献：

1. U-Net 架构：论文提出了一种新颖的网络架构，它包含一个收缩路径（用于捕捉上下文信息）和一个对称的扩展路径（用于精确定位）。这种设计使得网络能够在很少的训练图像上进行端到端的训练，并在多个生物医学图像分割任务中取得了优异的性能。
2. 数据增强策略：为了克服训练样本不足的问题，论文采用了一种基于弹性变形的数据增强方法。这种方法通过模拟组织中的自然变形，使网络能够学习到对这些变形的不变性，从而提高了模型的泛化能力。
3. 有效的训练策略：论文介绍了一种有效的训练策略，包括使用高动量值的随机梯度下降、像素级交叉熵损失函数以及权重图的计算，这些策略共同提高了网络的训练效率和分割精度。
4. 快速分割能力：U-Net 的设计允许在 GPU 上快速进行图像分割，处理 512x512 像素的图像仅需不到一秒的时间。

模型设计

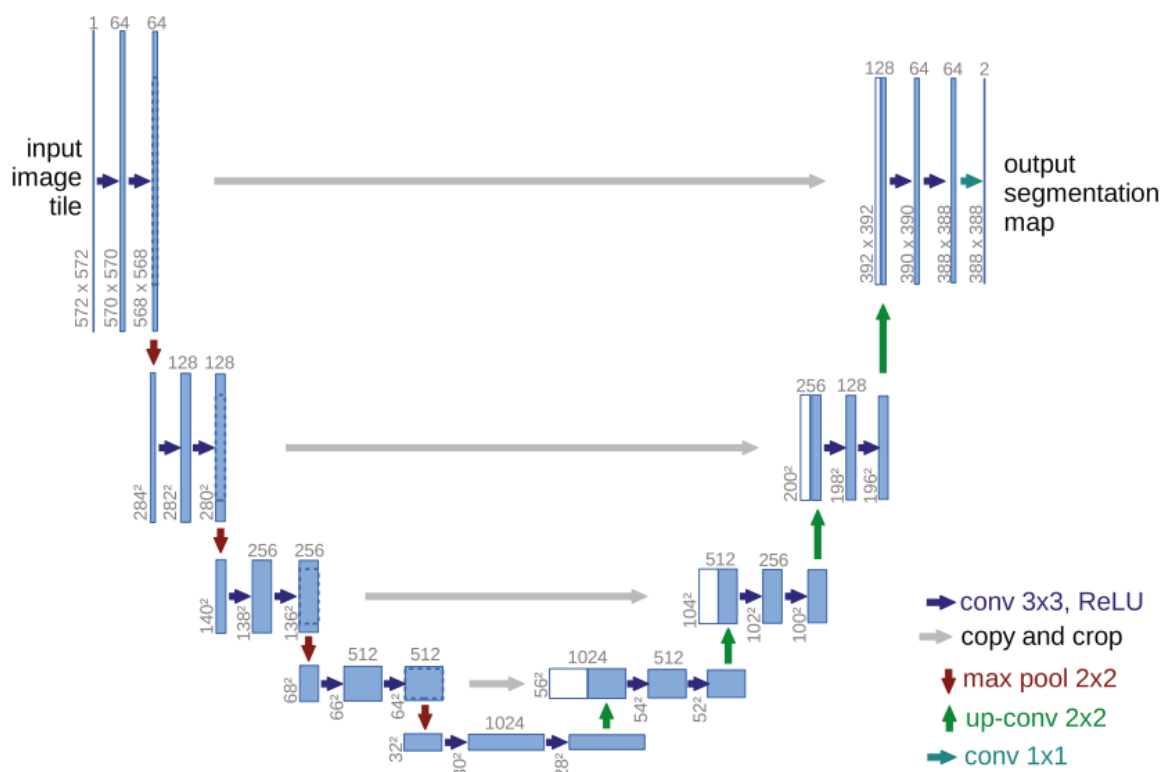


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows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

Unet 网络结构是对称的，形似英文字母 U 所以被称为 Unet。整张图都是由蓝/白色框与各种颜色的箭头组成，其中，蓝/白色框表示 feature map；蓝色箭头表示 3x3 卷积，用于特征提取；灰色箭头表示 skip-connection，用于特征融合；红色箭头表示池化 pooling，用于降低维度；绿色箭头表示上采样 upsample，用于恢复维度；青色箭头表示 1x1 卷积，用于输出结果。其中灰色箭头 **copy and crop** 中的 **copy** 就是 **concatenate** 而 **crop** 是为了让两者的长宽一致

UNet 是 Encoder-Decoder 结构，结构很简单但很有效。其中 Encoder 负责特征提取，Decoder 负责将 Encoder 提取的抽象特征重新映射到原始图像的空间分辨率，并生成一个详细的、像素级的分割图。

- Encoder

Encoder 由卷积操作和下采样操作组成，文中所用的卷积结构统一为 **3x3 的卷积核**，padding 为 0，striding 为 1。没有 padding 所以每次卷积之后 feature map 的 H 和 W 变小了，在 skip-connection 时要注意 feature map 的维度(其实也可以将 padding 设置为 1 避免维度不对应问题)。

两次卷积之后是一个 stride 为 2 的 max pooling，输出大小变为 $1/2 * (H, W)$

上面的步骤重复 5 次，最后一次没有 max-pooling，直接将得到的 feature map 送入 Decoder。

- Decoder

feature map 经过 Decoder 恢复原始分辨率，该过程除了卷积比较关键的步骤就是 upsampling 与 skip-connection。

Upsampling 逐步放大特征图的空间尺寸，从而恢复到接近原始输入图像的尺寸，上采样常用的方式有两种：1. FCN 中介绍的反卷积；2. 插值。文中使用双线性插值的方式。

skip-connection 这一步融合了底层信息的位置信息与深层特征的语义信息，这里需要注意的是，FCN 中深层信息与浅层信息融合是通过对对应像素相加的方式，而 Unet 是通过拼接的方式。

那么这两者有什么区别呢，其实在 ResNet 与 DenseNet 中也有一样的区别，Resnet 使用了对应值相加，DenseNet 使用了拼接。个人理解在相加的方式下，feature map 的维度没有变化，但每个维度都包含了更多特征，对于普通的分类任务这种不需要从 feature map 复原到原始分辨率的任务来说，这是一个高效的选择；而拼接则保留了更多的维度/位置信息，这使得后面的 layer 可以在浅层特征与深层特征自由选择，这对语义分割任务来说更有优势。

- 损失函数

1. 交叉熵损失 (Cross-Entropy Loss) :

- 交叉熵损失是图像分割中最常用的损失函数之一，特别是在分类问题中。对于分割任务，它通常以像素级别计算，每个像素的预测类别概率与真实类别之间的交叉熵差异被累加起来。
- 对于二分类问题，交叉熵损失可以表示为： $L = - \sum_i [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)]$ 其中 y_i 是第 i 个像素的真实标签 (0 或 1)， p_i 是模型预测该像素为正类的概率。

2. Dice 系数损失 (Dice Coefficient Loss) :

- Dice 系数是衡量两个样本相似度的指标，也常用于图像分割任务中。Dice 系数损失函数鼓励模型预测与真实标注更接近的分割结果。
- Dice 系数可以定义为： $Dice = \frac{2 \times |X \cap Y|}{|X| + |Y|}$ 其中 X 是预测的分割区

域， Y 是真实的分割区域， $|X \cap Y|$ 表示 X 和 Y 的交集的大小， $|X|$ 和 $|Y|$ 分别表示 X 和 Y 的大小。

- Dice 损失通常以 1 减去 Dice 系数的形式表示： $Dice = 1 - Dice$

3. 像素级损失 (Pixel-wise Loss) :

- 像素级损失是指对每个像素的预测结果进行评估的损失函数。除了交叉熵损失，还可以使用其他像素级损失函数，如均方误差 (MSE) 或绝对误差 (MAE) 。

复现

<https://nn.labml.ai/unet/index.html>

实验结果

1. 电子显微镜图像中的神经结构分割：

- U-Net 被应用于电子显微镜图像中神经结构的分割任务，这是 ISBI 挑战赛中的一个任务。
- 论文中提到，U-Net 在没有任何预处理或后处理的情况下，通过平均 7 个旋转版本的输入数据，达到了 0.0003529 的变形误差（warping error），这是当时的最佳成绩。
- 与之前的最佳方法（Ciresan 等人的滑动窗口卷积网络）相比，U-Net 在变形误差和随机误差（rand error）上都有显著的改进。具体来说，U-Net 的随机误差为 0.0382，而之前最佳方法的随机误差为 0.0504。

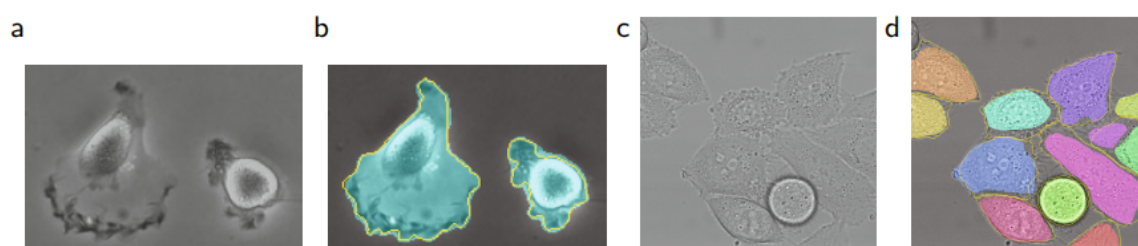


Fig. 4. Result on the ISBI cell tracking challenge. (a) part of an input image of the “PhC-U373” data set. (b) Segmentation result (cyan mask) with manual ground truth (yellow border) (c) input image of the “DIC-HeLa” data set. (d) Segmentation result (random colored masks) with manual ground truth (yellow border).

2. ISBI 细胞跟踪挑战赛 2015：

- U-Net 参加了 ISBI 细胞跟踪挑战赛 2015，并在两个最具挑战性的二维透射光数据集上取得了显著的成果。
- 在“PhC-U373”数据集上，U-Net 实现了 92% 的平均交并比（Intersection over Union, IOU），这显著优于第二名的 83%。
- 在“DIC-HeLa”数据集上，U-Net 实现了 77.5% 的平均 IOU，同样显著优于第二名的 46%。

Table 2. Segmentation results (IOU) on the ISBI cell tracking challenge 2015.

Name	PhC-U373	DIC-HeLa
IMCB-SG (2014)	0.2669	0.2935
KTH-SE (2014)	0.7953	0.4607
HOUS-US (2014)	0.5323	-
second-best 2015	0.83	0.46
u-net (2015)	0.9203	0.7756

总结

底层(深层)特征:

经过多次下采样后的低分辨率信息。能够提供分割目标在整个图像中上下文语义信息，可理解为反应目标和它的环境之间关系的特征。这个特征有助于物体的类别判断（所以分类问题通常只需要低分辨率/深层信息，不涉及多尺度融合）

高层(浅层)特征:

经过 concatenate 操作从 encoder 直接传递到同高度 decoder 上的高分辨率信息。能够为分割提供更加精细的特征，如梯度等。

意义：UNet 最早的提出是用于在医疗影像领域，由于网络的架构，UNet 的 Encoder 替换为各类特征提取的网络，可以广泛应用于各类语义分割任务。