

# FCN、Unet、Unet++ + 医学图像分割那点事儿

1. 相关知识点解释
2. FCN 网络算法的理解
3. Unet 网络算法的理解
4. Unet++ 网络算法的理解
5. Unet+++ 网络算法的理解
6. DeepLab v3+ 算法简阅
7. Unet在医学图像上的适用与CNN分割算法的简要总结

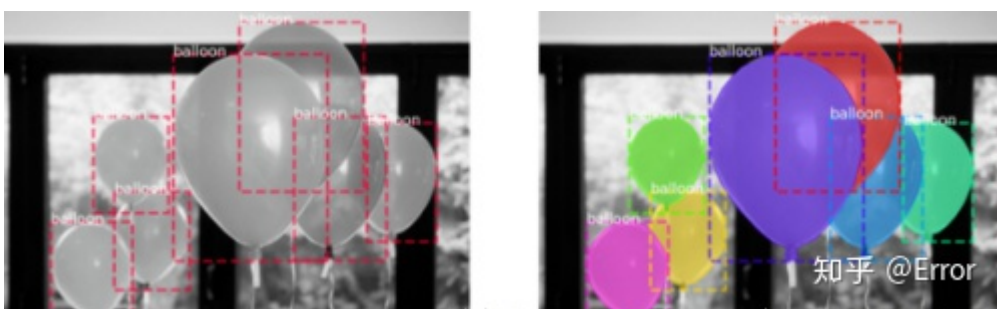
## 一、相关知识点解释

### 1. 图像分割中几种定义的区别

- **语义分割**(Semantic Segmentation): 就是对一张图像上的**所有像素点**进行分类。(eg: FCN/Unet/Unet++/...)
- **实例分割**(Instance Segmentation): 可以理解为**目标检测**和**语义分割**的结合。(eg: Mask R-CNN/...)
  - 相对目标检测的边界框，实例分割可精确到物体的边缘；
  - 相对语义分割，实例分割需要标注出图上同一物体的不同个体。
- **全景分割**(Panoptic Segmentation): 可以理解为**语义分割**和**实例分割**的结合。
  - 实例分割只对图像中的object进行检测，并对检测到的object进行分割，
  - 全景分割是对图中的所有物体包括背景都要进行检测和分割。



图像分类（左），语义分割（右）



目标检测（左），实例分割（右）

图像分类：图像中的气球是一个类别。<sup>[1]</sup>

语义分割：分割出气球和背景。

目标检测：图像中有7个目标气球，并且检测出每个气球的坐标位置。

实例分割：图像中有7个不同的气球，在像素层面给出属于每个气球的像素。



## 2. CNN特征学习的优势

**高分辨率特征(较浅的卷积层)**感知域较小，有利于feature map和原图进行对齐的，也就是我说的可以提供更多的位置信息。

**低分辨率信息(深层的卷积层)**由于感知域较大，能够学习到更加抽象一些的特征，可以提供更多的上下文信息，即强语义信息，这有利于像素的精确分类。

## 3. 上采样（意义在于将小尺寸的高维度feature map恢复回去）

上采样（upsampling）一般包括2种方式：

- **Resize**，如双线性插值直接对图像进行缩放（这种方法在原文中提到）
- **Deconvolution（反卷积）**<sup>[2]</sup>，也叫Transposed Convolution(转置卷积)，可以理解为卷积的逆向操作。

## 4. 医学影像语义分割的几个评估指标<sup>[3]</sup>

### (1) Jaccard(IoU)

用于比较有限样本集之间的相似性与差异性。**Jaccard值越大，样本相似度越高。**

$$IOU = \frac{TP}{FP + TP + FN}$$

关于对TP、FP、TN、FN的理解，可参考我的另一篇目标检测中mAP计算的博文：

Error：轻松计算目标检测评价指标  
mAP

[zhuanlan.zhihu.com](https://zhuanlan.zhihu.com)



### (2) Dice相似系数

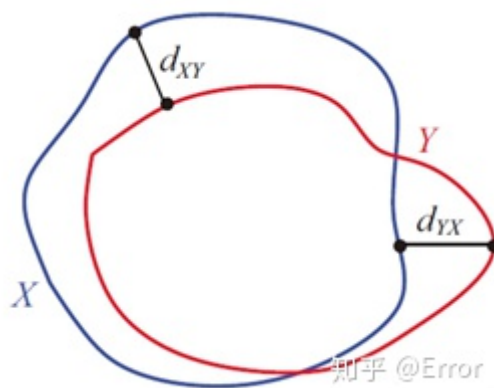
一种集合相似度度量指标,通常用于计算两个样本的相似度,值的范围0~1,分割结果最好时值为1,最差时值为0。**Dice相似系数对mask的内部填充比较敏感。**

$$DSC(A, B) = \frac{2 \times |\text{Intersection}|}{|A| + |B|}$$

$$\text{Dice}(P, T) = \frac{|P_1 \wedge T_1|}{(|P_1| + |T_2|)/2} \Leftrightarrow \text{Dice} = \frac{2TP}{FP + 2TP + FN}$$

### (3) Hausdorff 距离 (豪斯多夫距离)

描述两组点集之间相似程度的一种量度，对分割出的边界比较敏感。



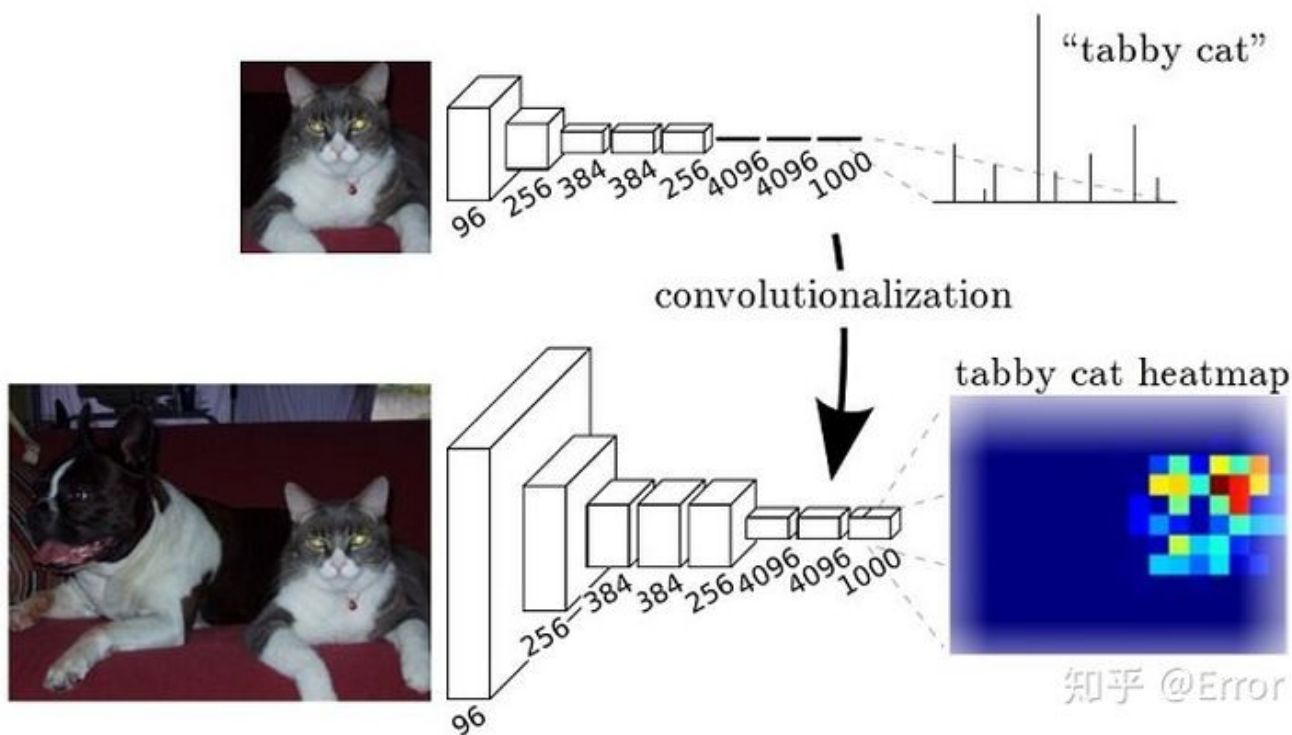
$$d_H(X, Y) = \max \{d_{XY}, d_{YX}\} = \max \left\{ \max_{x \in X} \min_{y \in Y} d(x, y), \max_{y \in Y} \min_{x \in X} d(x, y) \right\}$$

### (4) F1-score

用来衡量二分类模型精确度的一种指标,同时考虑到分类模型的准确率和召回率，可看做是准确率和召回率的一种加权平均。

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

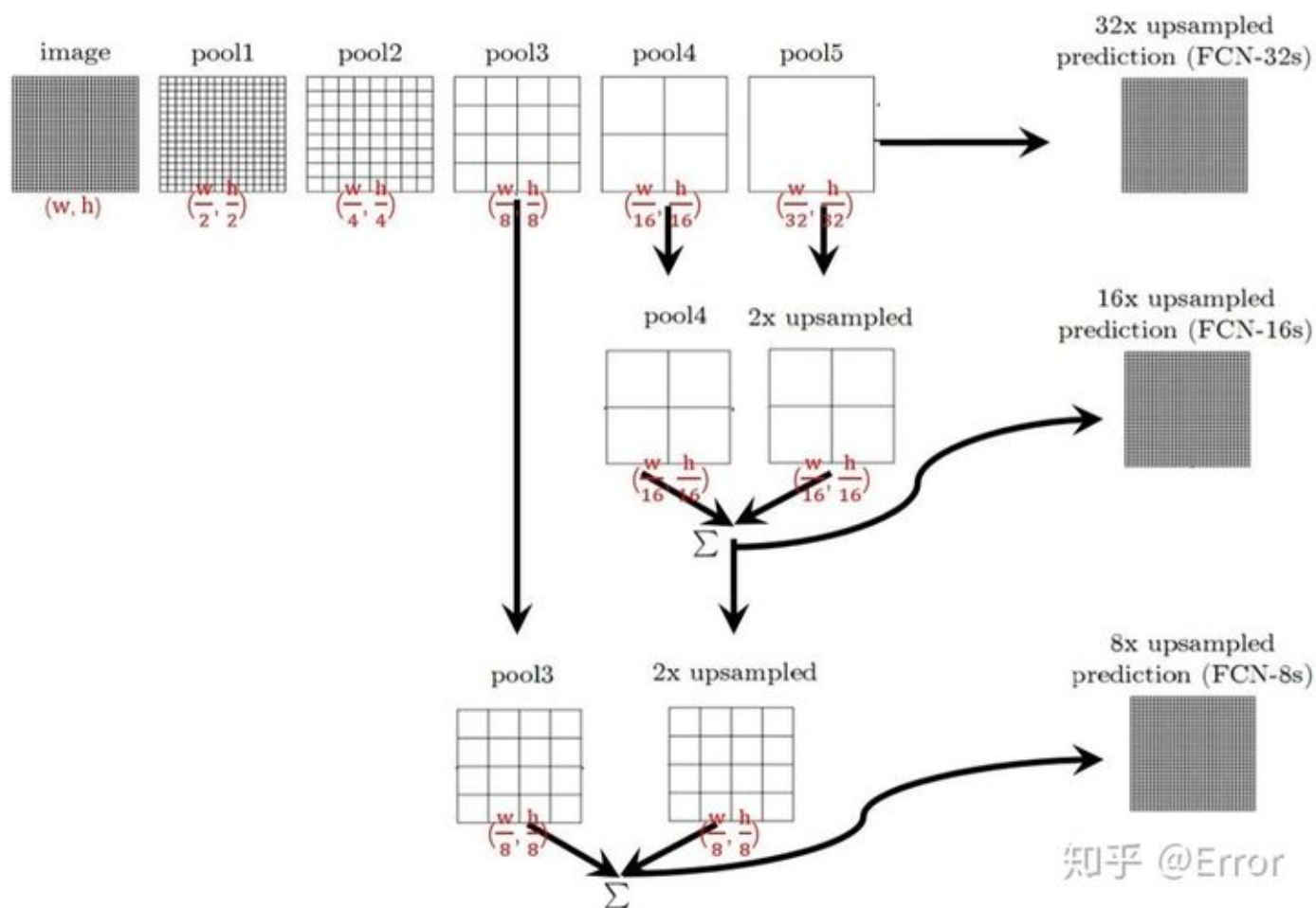
## 二、FCN网络的理解



FCN将一般的经典的分类网络模型（VGG16...）的最后一层的FC层（全连接）换成卷积，这样可以通过二维的特征图，后接softmax获得每个像素点的分类信息，从而解决了分割问题。

核心思想：

- 不含全连接层(fc)的全卷积(fully conv)网络。可适应任意尺寸输入。
- 增大数据尺寸的反卷积(deconv)层。能够输出精细的结果。
- 结合不同深度层结果的跳级(skip)结构。同时确保鲁棒性和精确性。



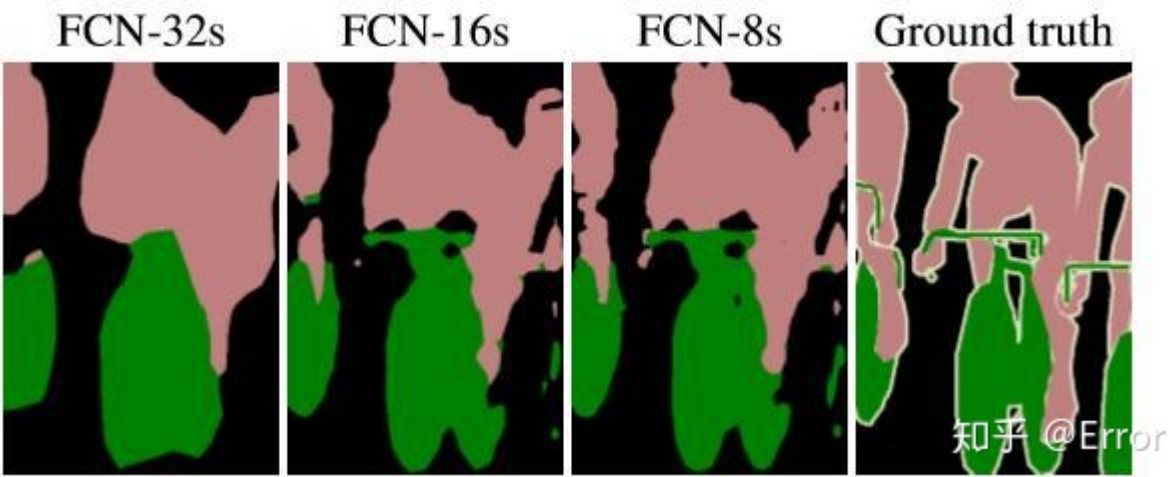
FCN结构示意图



对于FCN-32s，直接对pool5 feature进行32倍上采样获得32x upsampled feature，再对32x upsampled feature每个点做softmax prediction获得32x upsampled feature prediction（即分割图）。

对于FCN-16s，首先对pool5 feature进行2倍上采样获得2x upsampled feature，再把pool4 feature和2x upsampled feature逐点相加，然后对相加的feature进行16倍上采样，并softmax prediction，获得16x upsampled feature prediction。

对于FCN-8s，首先进行pool4+2x upsampled feature逐点相加，然后又进行pool3+2x upsampled逐点相加，即进行更多次特征融合。



FCN-32s & FCN-16s & FCN-8s，也就是：使用多层 feature融合有利于提高分割准确性。

**FCN缺点：**

- 1. 结果不够精细。进行8倍上采样虽然比32倍的效果好了很多，但是上采样的结果还是比较模糊和平滑，对图像中的细节不敏感。
- 2. 对各个像素进行分类，没有充分考虑像素与像素之间的关系。忽略了在通常的基于像素分类的分割方法中使用的空间规整（spatial regularization）步骤，缺乏空间一致性。

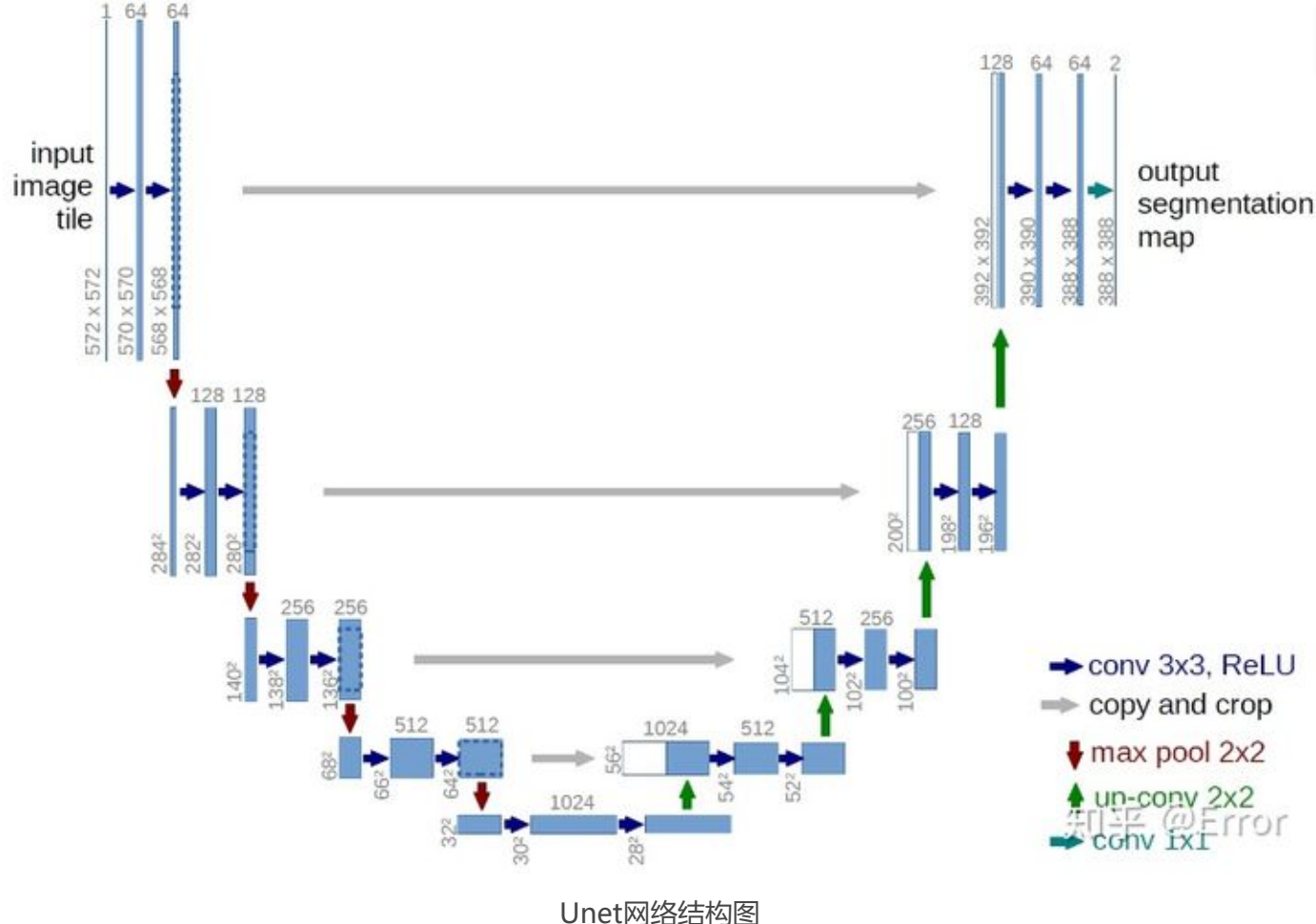
附：

FCN论文地址  
[arxiv.org](https://arxiv.org)



**三、U-net网络的理解**





U-Net网络结构图

整个U-Net网络结构类似于一个大型的字母U，与FCN都是很小的分割网络，既没有使用空洞卷积，也没有后接CRF，结构简单。

1. 首先进行Conv+Pooling下采样；
2. 然后反卷积进行上采样，crop之前的低层feature map，进行融合；
3. 再次上采样。
4. 重复这个过程，直到获得输出388x388x2的feature map，
5. 最后经过softmax获得output segment map。总体来说与FCN思路非常类似。

UNet的encoder下采样4次，一共下采样16倍，对称地，其decoder也相应上采样4次，将encoder得到的高级语义特征图恢复到原图片的分辨率。

它采用了与FCN不同的特征融合方式：

1. FCN采用的是逐点相加，对应tensorflow的tf.add()函数
2. U-Net采用的是channel维度拼接融合，对应tensorflow的tf.concat()函数

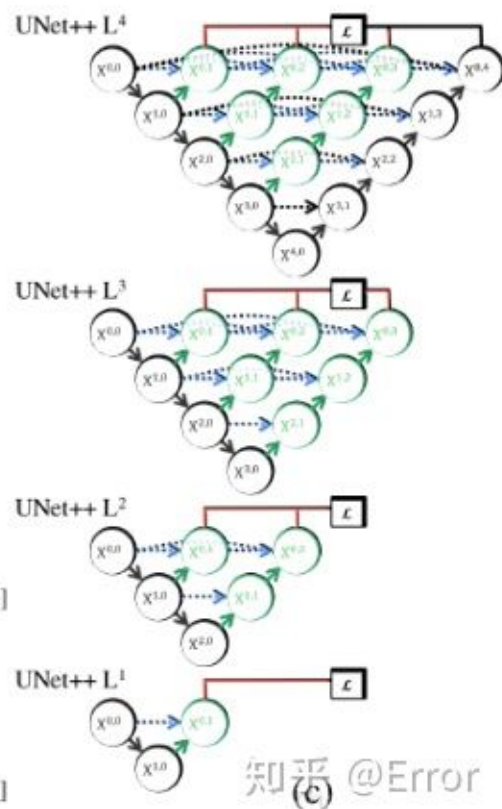
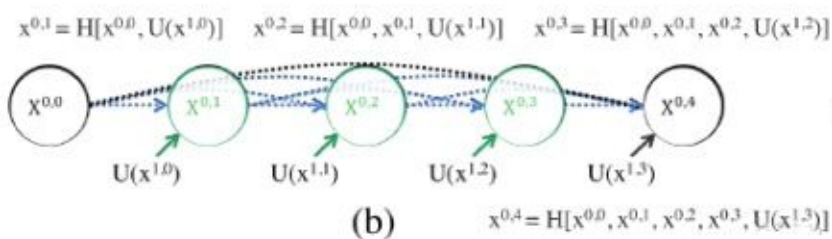
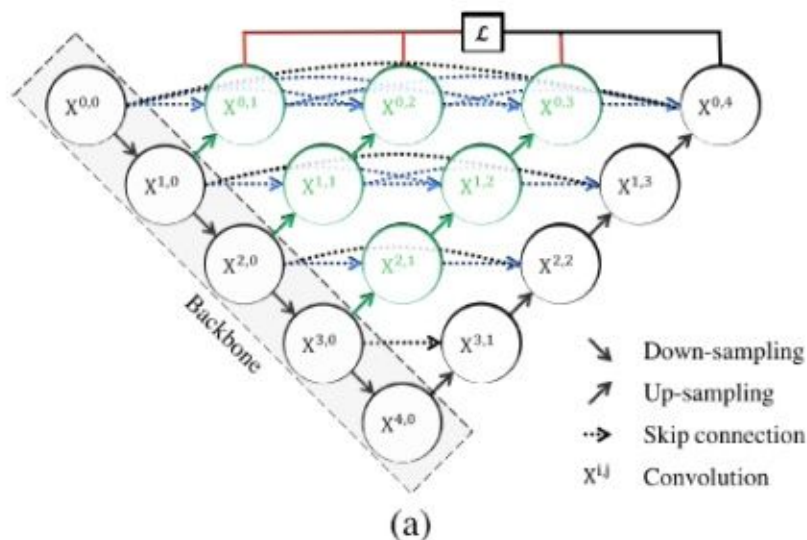
附：



## 四、Unet++ + 网络的理解<sup>[4]</sup>

文章对Unet改进的点主要是skip connection，作者认为skip connection 直接将unet中encoder的浅层特征与decoder的深层特征结合是不妥当的，会产生semantic gap。

文中假设：当所结合的浅层特征与深层特征是semantically similar时，网络的优化问题就会更简单，因此文章对skip connection的改进就是想bridge/reduce 这个semantic gap。



附：

Unet++论文地址

[arxiv.org](https://arxiv.org)

MrGiovanni/UNetPlusPlus

[github.com](https://github.com)



## 五、Unet++ + 算法的理解<sup>[5]</sup>

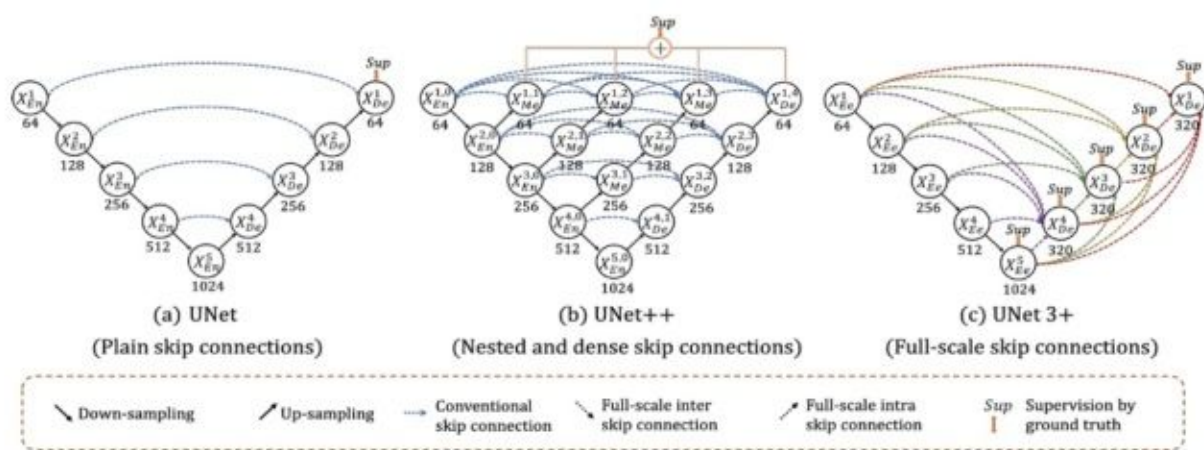


Fig .1: Comparison of UNet (a), UNet++(b) and proposed UNet 3+ (c). The depth of each node is presented below the circle.

为了弥补UNet和UNet++的缺陷，UNet 3+中的每一个解码器层都融合了来自编码器中的小尺度和同尺度的特征图，以及来自解码器的大尺度的特征图，这些特征图捕获了全尺度下的细粒度语义和粗粒度语义。

附：

U-net+++论文地址  
[arxiv.org](https://arxiv.org)



## 六、DeepLab v3+算法简阅<sup>[6]</sup>

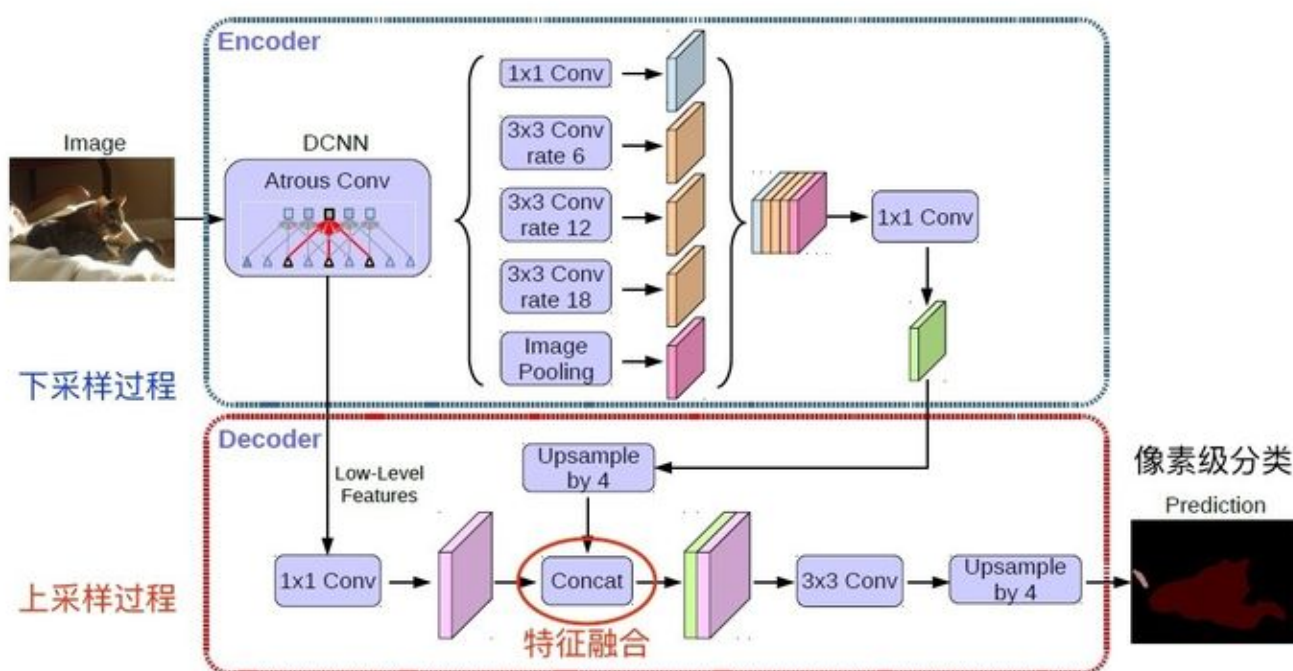


Figure 2. Our proposed DeepLabv3+ extends DeepLabv3 by employing an encoder-decoder structure. The encoder module encodes multi-scale contextual information by applying atrous convolution at multiple scales, while the simple yet effective decoder module refines the segmentation results along object boundaries.

DeepLab v3+结构图

Encoder部分



Encoder就是原来的DeepLabv3,

需要注意点:

1. 输入尺寸与输出尺寸比 (output stride = 16) , 最后一个stage的膨胀率rate为2
2. Atrous Spatial Pyramid Pooling module (ASPP) 有四个不同的rate, 额外一个全局平均池化

## Decoder部分

先把encoder的结果上采样4倍, 然后与resnet中下采样前的Conv2特征concat一起, 再进行3x3的卷积, 最后上采样4倍得到最终结果。

需要注意点: 融合低层次信息前, 先进行1x1的卷积, 目的是降通道 (例如有512个通道, 而encoder结果只有256个通道)

附:

DeepLab v3+论文地址

[arxiv.org](https://arxiv.org/abs/1706.05424)



## 七、Unet在医学图像上的适用与CNN分割算法的简要总结

### 1. Unet结构特点

UNet相比于FCN和Deeplab等, 共进行了4次上采样, 并在同一个stage使用了skip connection, 而不是直接在高级语义特征上进行监督和loss反传, 这样就保证了最后恢复出来的特征图融合了更多的low-level的feature, 也使得不同scale的feature得到了融合, 从而可以进行多尺度预测和DeepSupervision。4次上采样也使得分割图恢复边缘等信息更加精细。

### 2. 为什么适用于医学图像? [7]

1. 因为医学图像边界模糊、梯度复杂, 需要较多的高分辨率信息。高分辨率用于精准分割。
2. 人体内部结构相对固定, 分割目标在人体图像中的分布很具有规律, 语义简单明确, 低分辨率信息能够提供这一信息, 用于目标物体的识别。

UNet结合了**低分辨率信息** (提供物体类别识别依据) 和**高分辨率信息** (提供精准分割定位依据), 完美适用于医学图像分割。

### 3. 分割算法改进总结:

1. **下采样+上采样**: Convolution + Deconvolution / Resize
2. **多尺度特征融合**: 特征逐点相加 / 特征channel维度拼接
3. **获得像素级别的segment map**: 对每一个像素点进行判断类别

1. ^ 可以参考一下卷积运算的几种操作 [https://blog.csdn.net/attitude\\_yu/article/details/84697181](https://blog.csdn.net/attitude_yu/article/details/84697181)

2. ^ 周纵苇的研习Unet <https://zhuanlan.zhihu.com/p/44958351>

3. ^ UNet3+(UNet+++ )论文解读 <https://zhuanlan.zhihu.com/p/136164721>

4. ^ DeepLab系列理解 <https://www.jianshu.com/p/755b001bfe38>

5. ^ Unet神经网络为什么会在医学图像分割表现好? <https://www.zhihu.com/question/269914775>

