

# U-Net: Convolutional Networks for Biomedical Image Segmentation

Ronneberger, Olaf and Fischer, Philipp and Brox, Thomas

## 摘要

结肠镜检查是一种有效的结直肠息肉检测技术，而结直肠息肉与结直肠癌密切相关。在临床实践中，从结肠镜图像中分割息肉具有重要意义，因为它为诊断和手术提供了非常有价值的信息。基于深度学习的肠息肉分割方法十分火热，相较传统图像分割方法表现优良。而 U-net 作为十分经典的轻量级网络模型，以较小的样本需求以及较好的分割性能在医学图像分割领域著名。本文将利用 U-net 网络实现开源数据集 Kvasir-SEG 肠息肉数据集分割预测。

**关键词：**U-net；肠息肉图像分割；Kvasir-SEG

## 1 引言

结直肠癌（CRC）是全球第三大常见癌症类型。因此，通过筛查和切除癌前病变（结直肠腺瘤）预防结直肠癌是非常关键的，并已成为全球公共卫生的优先事项。肠息肉是一种常见的消化系统疾病，它可以在肠道内形成一个或多个肿块，这些肿块可能会导致症状，如腹痛、腹胀、恶心、呕吐等。肠息肉的诊断主要依赖于医学影像技术，如 CT、MRI 等。结肠镜检查是筛查和预防结直肠癌的有效技术，因为它可以提供结直肠息肉的位置和外观信息，使医生能够在它们发展成结直肠癌之前将其切除。许多研究表明早期结肠镜检查有助于降低 30% 的 CRC 发病率<sup>[1]</sup>。因此，在临床设置中，准确的息肉分割是非常重要的。然而，由于两个主要原因，这是一项具有挑战性的任务。首先，息肉通常在外观上变化，例如，大小、颜色和纹理，即使它们是相同的类型。第二，在结肠镜检查图像中，息肉及其周围粘膜之间的边界通常是模糊的，并且缺乏分割方法所需的强烈对比度。这些问题导致了息肉分割的不准确，有时甚至会造成息肉的漏检。因此，急需一种高效准确的图像分割方法，以极大的减轻 CT 影像医师对结肠癌诊断的压力、减少医生的阅片数量、提高医生对结肠癌诊断的及时性与准确性以及降低医生的学习成本。

随着由于深度学习算法近年来在分类问题上表现突出，效果可观。产生了一系列的深度学习网络，在这其中由 long 等人提出的全卷积神经网络 (FCN)<sup>[2]</sup>是最早应用于图像分割的深度网络之一。此网络经过端到端和像素到像素的训练，以进行语义分割。CNN 架构的所有 FC 层都被卷积和反卷积取代，以保持原始的空间分辨率。因此，在网络执行分割任务时，将恢复特征图的原始空间维度，FCN 经常用于医学和生物医学图像的分割。Ronneberger 等人<sup>[3]</sup>将此体系结构扩展到 U-Net，利用大量训练数据的需求实现了良好的分割结果。他们的网络由编码和解码路径组成。在编码路径中，从输入数据中提取了大量降维的特征图。解码路径用于通过执行上卷积来生成分段图 (与输入大小相同)。U-net 架构在非常不同的生物医学分割应用中实现了非常好的性能。由于使用弹性变形进行数据扩增，它只需要很少的注释图像，并且其训练时间也很可观。因此，本工作将基于 U-net 网络对肠息肉医疗图像进行分割。

## 2 相关工作

### 2.1 传统基于阈值的分割方法

阈值法的基本思想是基于图像的灰度特征来计算一个或多个灰度阈值，并将图像中每个像素的灰度值与阈值作比较，最后将像素根据比较结果分到合适的类别中。因此，该方法最为关键的一步就是按照某个准则函数来求解最佳灰度阈值。图像若只有目标和背景两大类，那么只需要选取一个阈值进行分割，此方法成为单阈值分割；但是如果图像中有多个目标需要提取，单一阈值的分割就会出现作物，在这种情况下就需要选取多个阈值将每个目标分隔开，这种分割方法相应的成为多阈值分割。

### 2.2 全卷积神经网络 FCN

FCN<sup>[2]</sup>是第一次将端到端的卷积网络推广到了语义分割任务当中，它可以接受任意尺寸的输入图像，采用反卷积层对最后一个卷积层的 feature map 进行上采样, 使它恢复到输入图像相同的尺寸，从而可以对每个像素都产生了一个预测, 同时保留了原始输入图像中的空间信息, 最后在上采样的特征图上进行逐像素分类 (最后逐个像素计算 softmax 分类的损失, 相当于每一个像素对应一个训练样本)。另一方面，采用了跳跃连接的方式，跳跃连接结构：这是特征融合的一种方式，未加入跳跃结构的模型是特别粗糙的，因为特征图直接扩大了 32 倍，所以论文中有一个改进，就是利用跳跃结构将网络中间的输入联合起来，即逐点相加，再进行反卷积，这样能够依赖更多的信息，将深层的信息与浅层的信息相结合，这样可以提高模型分割的准确性。

## 3 本文方法

### 3.1 本文方法概述

所采用的 U-net<sup>[3]</sup>网络整体结构如图 1所示：

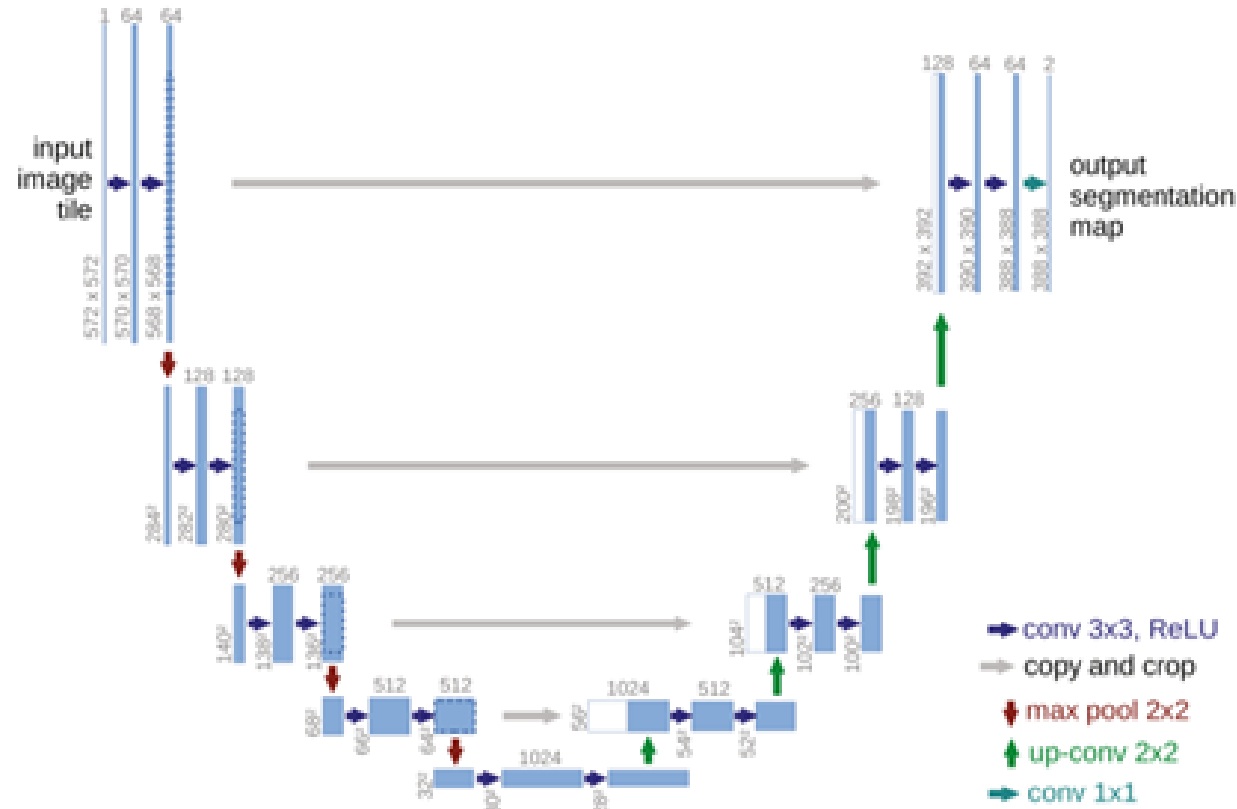


图 1: Unet 网络整体结构

它由收缩路径（左侧）和扩张路径（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  $3\times 3$  卷积（未填充卷积），每个卷积后面都有一个整流线性单元（ReLU）和一个跨距为 2 的  $2\times 2$  最大池操作，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包括特征图的上采样，之后是将特征通道的数量减半的  $2\times 2$  卷积（“上卷积”）、与来自收缩路径的相应裁剪的特征图的级联、以及两个  $3\times 3$  卷积，每个卷积之后是 ReLU。由于在每个卷积中丢失边界像素，因此需要进行裁剪。在最后一层，使用  $1\times 1$  卷积将每个 64 分量特征向量映射到所需数量的类。该网络总共有 23 个卷积层。为了实现输出分割图的无缝分块，选择输入分块大小非常重要，以便所有  $2\times 2$  最大池操作应用于  $x$  和  $y$  大小相等的层。Unet 网络结构的主要优点在于：1）采取将低级特征图与后面的高级特征图进行融合操作。2）完全对称的 U 型结构使得前后特征融合更为彻底，使得高分辨率信息与低分辨率信息在目标图片中增加，特征信息更加丰富。3）结合了下采样时的低分辨率信息（提供物体类别识别依据）和上采样时的高分辨率信息（提供精准分割定位依据），此外还通过融合操作（跳跃结构）填补底层信息以提高分割精度（分辨率就是图片的尺寸）。

### 3.2 损失函数选用

所采用的损失函数是 Pytorch 中的交叉熵损失函数 `nn.CrossEntropyLoss()`，该函数主要结合了 `nn.LogSoftmax()` 和 `nn.NLLLoss()` 两个函数。实质是将 softmax-log-NLLoss 合并到一块得到的结果。Softmax 后的数值都在 0 1 之间，所以  $\ln$  之后值域是负无穷到 0。然后将 Softmax 之后的结果取  $\log$ ，将乘法改成加法减少计算量，同时保障函数的单调性。NLLoss 的结果就是把上面的输出与 Label 对应的那个值拿出来，去掉负号，再求均值。

## 4 复现细节

### 4.1 与已有开源代码对比

复现过程使用了 milesial 的 Pytorch-UNet 的项目代码（原作者在该项目中用 Unet 实现对 Kaggle's Carvana Image Masking Challenge 数据集中汽车的语义分割），包括其 `train.py` (用于网络的训练)、`predict.py` (用于生成预测图片)、`utils` 文件夹下的 `dataloading` 等文件 (数据的准备)。为了深入学习 Unet 网络结构，复现中自己依照原论文的 Unet 网络结构完成了 `Unet.py`，同时因为所采用的开源肠息肉数据集 Kvasir-SEG 图像源格式为 TIF 格式，需转换为 png 格式，所以补充了 `TIF2PNG.py` 的编写。

### 4.2 实验环境搭建

Python 版本: 3.8.5 Pytorch 版本: 1.12.0+cu102 CUDA 版本: 10.2

### 4.3 创新点

将原作者使用的 Kaggle's Carvana Image 汽车数据集更换为自己研究方向的肠息肉医疗图像数据集。

## 5 实验结果分析

本部分对实验所得结果进行分析，详细对实验内容进行说明，实验结果进行描述并分析。

### 5.1 实验内容

1) 数据集的准备，分别从 Kvasir-SEG 和 CVC-ClinicDB 的官网上下载数据集，其中 Kvasir-SEG 是一个开放的胃肠道息肉图像和相应的分割掩码数据集，是由一名医生手工注释，然后由一名经验丰

富的胃肠病学家进行验证所得到的，它由约 1000 张息肉图像以及对应的标注所组成；CVC-ClinicDB 包含有 612 张，是分别从 29 个不同的内窥镜视频片段中提取出来的。分别从中按照 7：3（训练集：测试集）的比例提取部分数据，构成实验所需的数据集。2）将训练集放在 U-Net 模型上进行训练。3）将保存好的已训练模型用测试集进行验证。4）最后，得到分割预测图与分割评估指标参数。

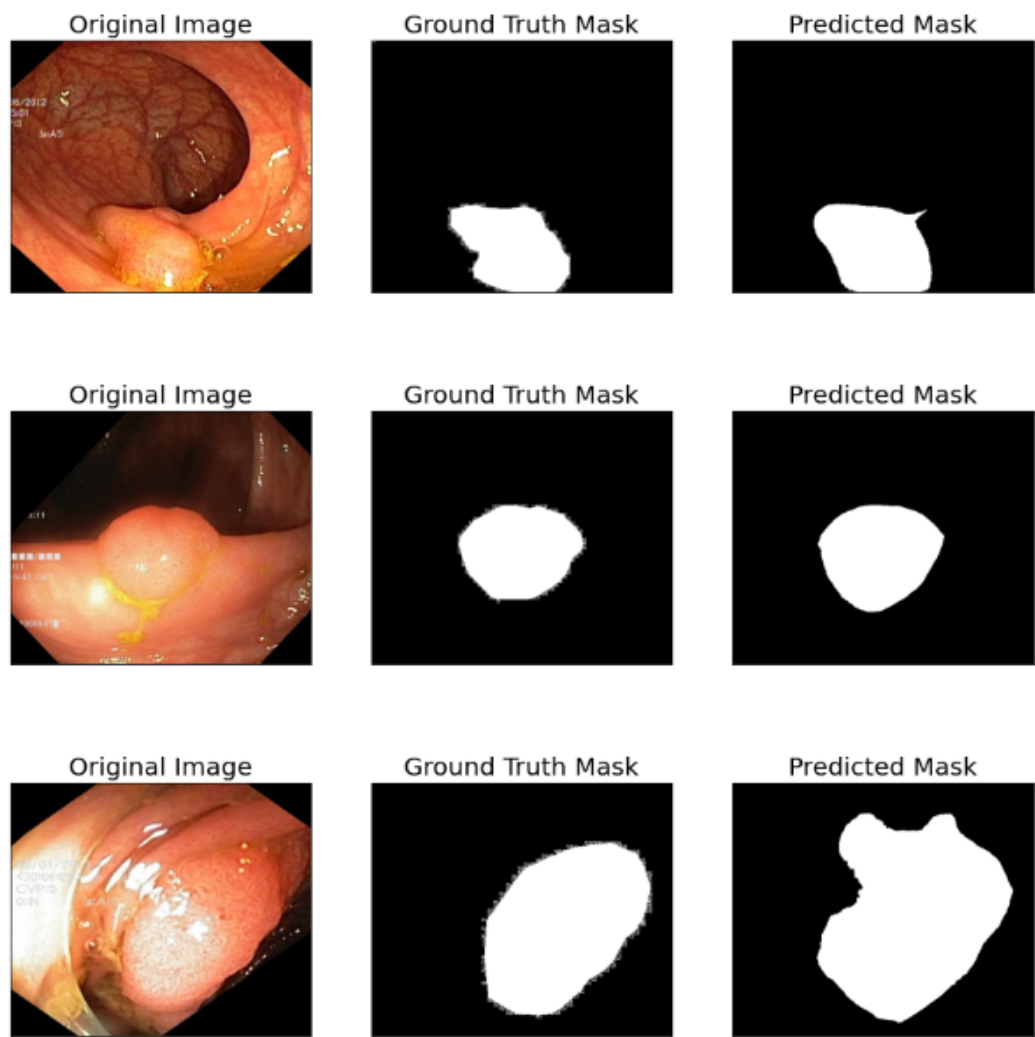


图 2: 实验结果示意

Method	Dice	Iou
U-Net	0.728	0.741

评估参数说明：Dice 系数是一种集合相似度量函数，通常用于计算两个样本的相似度，取值范围在 [0,1]，系数越大说明与测试集相似度越高，评估结果越好。公式如下：

$$dice(A,B)=\frac{2|A\cap B|}{|A|+|B|}$$

其中  $|A\cap B|$  是 A 和 B 之间的交集，对于语义分割问题而言，A 为 GT 分割图 (Ground Truth)，B 是预测的分割图像。其中，分子的系数为 2，是因为分母存在重复计算 A 和 B 之间的共同元素的原因。

IoU 和 Dice 参数类似，也是衡量测试结果相似度的，与 Dice 参数的值呈正相关，其计算公式如下：

$$IOU(A,B)=\frac{|A\cap B|}{|A\cup B|}$$

由实验结果可知，U-Net 作为经典轻量级网络，其网络结构简单，适用于许多医疗图像分割任务，在数据量较少的情况下仍然能够取得比较好的效果。

## 6 总结与展望

由于是跨考生，专业知识和代码水平较为薄弱，十分遗憾没能独自复现，参考了项目源代码。同时由于论文 U-Net: Convolutional Networks for Biomedical Image Segmentation 是于 2015 年发表于 MICCAI 的，后续已经有了许多改善和衍生版本，十分遗憾未能将这些改进加入到此代码中去实际学习体会。未来进一步的研究方向将是对目前开源肠息肉数据集的 *sota* 方法的学习以及尝试复现。

## 参考文献

- [1] HAGGAR F A, BOUSHEY R P. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors[J]. Clinics in colon and rectal surgery, 2009, 22(04): 191-197.
- [2] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.
- [3] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C] // Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. 2015: 234-241.