

U-Net:用于生物医学的卷积网络 图像分割

Olaf Ronneberger, Philipp Fischer 和 Thomas Brox

德国弗莱堡大学计算机科学系和 BIOSS 生物信号研究中心 ronneber@informatik.uni-freiburg.de, WWW
主页: <http://lmb.informatik.uni-freiburg.de/>

抽象的。人们普遍认为,深度网络的成功训练需要数千个带注释的训练样本。在本文中,我们提出了一种网络和训练策略,该策略依赖于数据增强的强大使用来更有效地使用可用的注释样本。该体系结构包括用于捕获上下文的收缩路径和实现精确定位的对称扩展路径。我们表明,这样的网络可以从很少的图像进行端到端训练,并且在 ISBI 挑战赛中优于先前的最佳方法(滑动窗口卷积网络),用于分割电子显微镜堆栈中的神经元结构。使用在透射光显微镜图像(相衬和 DIC)上训练的相同网络,我们在这些类别中赢得了 2015 年 ISBI 细胞跟踪挑战赛。此外,网络速度很快。在最近的 GPU 上分割 512x512 图像只需不到一秒的时间。完整的实现(基于 Caffe)和经过训练的网络可在<http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net>获得。

1 简介

在过去的两年里,深度卷积网络在许多视觉识别任务中的表现都优于现有技术,例如 [7, 3]。虽然卷积网络已经存在了很长时间 [8],但由于可用训练集的限制,考虑网络的大小等之间的突破。[7] 是由于在 ImageNet 数据集上对具有 8 层和数百万个参数的大型网络进行监督训练,具有 100 万张训练图像。从那时起,已经训练了更大更深的网络[12]。

卷积网络的典型用途是分类任务,其中图像的输出是单个类标签。然而,在许多视觉任务中,尤其是在生物医学图像处理中,所需的输出应该包括定位,即应该为每个像素分配一个类标签。此外,数以千计的训练图像在生物医学任务中通常是遥不可及的。

因此,Ciresan 等人。[1] 在滑动窗口设置中训练一个网络,通过在该像素周围提供一个局部区域(补丁)来预测每个像素的类别标签

2

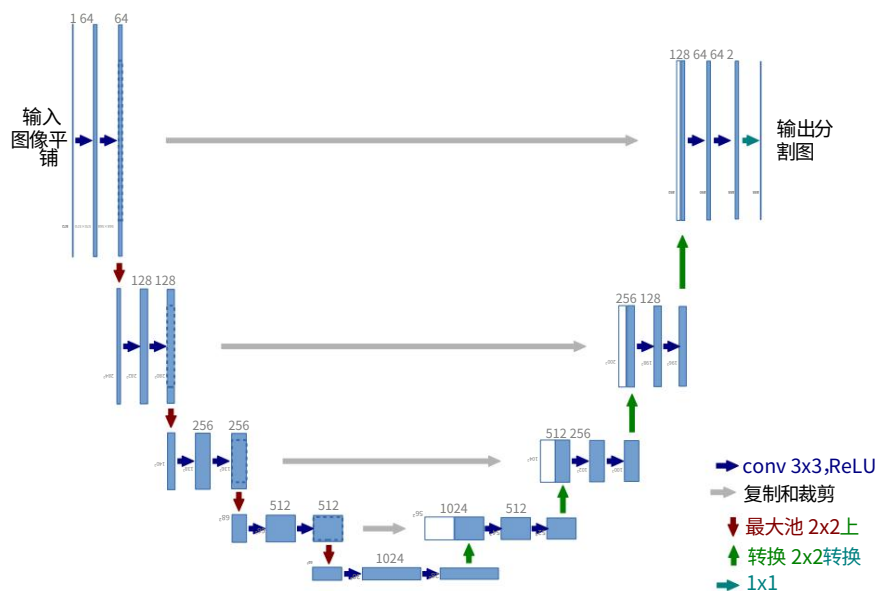


图 1. U-net 架构（最低分辨率下的 32x32 像素示例）。每个蓝色框对应一个多通道特征图。通道数显示在框的顶部。xy 尺寸位于框的左下边缘。白框代表复制的特征图。箭头表示不同的操作。

作为输入。首先,这个网络可以本地化。其次,就补丁而言,训练数据远大于训练图像的数量。由此产生的网络在 ISBI 2012 的 EM 分割挑战中大获全胜。

显然,Ciresan 等人的策略。[1]有两个缺点。首先,它非常慢,因为网络必须为每个补丁单独运行,并且由于重叠补丁而存在大量冗余。其次,在定位准确性和上下文使用之间存在权衡。较大的补丁需要更多的最大池化层,这会降低定位精度,而小补丁允许网络只能看到很少的上下文。最近的方法[11,4]提出了一个分类器输出,它考虑了来自多个层的特征。良好的本地化和上下文的使用是可能的。

在本文中,我们建立在一个更优雅的架构上,即所谓的“全卷积网络”[9]。我们修改和扩展了这种架构,使其适用于很少的训练图像并产生更精确的分割;参见图1。[9]中的主要思想是通过连续层来补充通常的承包网络,其中池化算子被上采样算子取代。

因此,这些层增加了输出的分辨率。为了进行定位,将收缩路径中的高分辨率特征与上采样的特征相结合

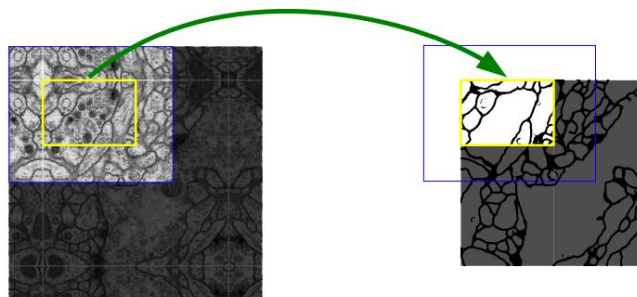


图 2. 任意大图像无缝分割的重叠平铺策略（这里是 EM 堆栈中神经元结构的分割）。黄色区域的分割预测,需要蓝色区域内的图像数据作为输入。通过镜像推断丢失的输入数据

输出。然后一个连续的卷积层可以学习根据这些信息组装一个更精确的输出。

我们架构中的一个重要修改是,在上采样部分,我们还有大量特征通道,这允许网络将上下文信息传播到更高分辨率的层。因此,扩展路径或多或少与收缩路径对称,并产生 u 形架构。网络没有任何全连接层,只使用每个卷积的有效部分,即分割图只包含输入图像中完整上下文可用的像素。

该策略允许通过重叠平铺策略无缝分割任意大的图像（参见图2）。为了预测图像边界区域的像素,通过镜像输入图像来推断缺失的上下文。

这种平铺策略对于将网络应用于大图像很重要,因为否则分辨率将受到 GPU 内存的限制。

至于我们的任务,可用的训练数据非常少,我们通过对可用的训练图像应用弹性变形来使用过多的数据增强。这允许网络学习对此类变形的不变性,而无需在带注释的图像语料库中查看这些转换。这在生物医学分割中尤为重要,因为变形曾经是组织中最常见的变化,并且可以有效地模拟真实的变形。Dosovitskiy 等人已经证明了数据增强对于学习不变性的价值。[2]在无监督特征学习的范围内。

许多细胞分割任务中的另一个挑战是分离同一类的触摸对象。参见图3。为此,我们建议使用加权损失,其中触摸单元之间的分离背景标签在损失函数中获得很大的权重。

由此产生的网络适用于各种生物医学分割问题。在本文中,我们展示了 EM 堆栈中神经元结构分割的结果（从 ISBI 2012 开始的正在进行的竞赛）,其中我们超越了

4

执行 Ciresan 等人的网络。[1]。此外,我们展示了来自 ISBI 细胞跟踪挑战 2015 的光学显微镜图像中的细胞分割结果。在这里,我们在两个最具挑战性的 2D 透射光数据集上大获全胜。

2 网络架构

网络架构如图1所示。它由收缩路径(左侧)和扩展路径(右侧)组成。收缩路径遵循卷积网络的典型架构。它由两个 3x3 卷积(未填充卷积)的重复应用组成,每个卷积后跟一个整流线性单元(ReLU)和一个 2x2 最大池化操作,步幅为 2,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都包括对特征图进行上采样,然后将特征通道数量减半的 2x2 卷积(“上卷积”),与收缩路径中相应裁剪的特征图的连接,以及两个 3x3 卷积,每个卷积后面跟着一个 ReLU。由于在每个卷积中都会丢失边界像素,因此裁剪是必要的。在最后一层,使用 1x1 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。

为了允许输出分割图的无缝平铺(参见图2),重要的是选择输入平铺大小,以便将所有 2x2 最大池化操作应用于具有偶数 x 和 y 大小的层。

3 培训

输入图像及其相应的分割图用于通过 Caffe [6] 的随机梯度下降实现来训练网络。

由于未填充的卷积,输出图像比输入图像小一个恒定的边框宽度。为了最大限度地减少开销并最大限度地利用 GPU 内存,我们倾向于使用大输入图块而不是大批量大小,从而将批量减少为单个图像。因此,我们使用高动量(0.99),以便大量先前看到的训练样本确定当前优化步骤中的更新。

能量函数是通过最终特征图上的像素级 soft-max 结合交叉熵损失函数来计算的。soft-max 定义为 $p_k(x) = \exp(a_k(x)) / \sum_{k=0}^K \exp(a_k(x))$ 其中 $a_k(x)$ 表示

特征通道 k 在像素位置 $x \in \Omega$ 处的激活,其中 $\Omega \subset Z$ 是类数, $p_k(x)$ 是近似的最大函数。到数。即对于具有最大激活 $a_k(x)$ 的 k, $p_k(x) \approx 1$, 对于所有其他 k, $p_k(x) \approx 0$ 。然后交叉熵在每个位置惩罚 $p_k(x)$ 与 1 的偏差,使用

$$E = \sum_{x \in \Omega} w(x) \log(p_k(x)) \quad (1)$$

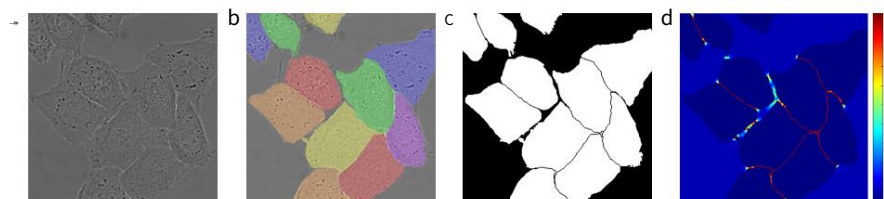


图 3. 用 DIC (微分干涉对比) 显微镜记录的玻璃上的 HeLa 细胞。(a) 原始图像。(b) 与地面实况分割叠加。不同的颜色表示 HeLa 细胞的不同实例。(c) 生成的分割掩码 (白色:前景,黑色:背景)。(d) 使用逐像素损失权重进行映射,以强制网络学习边界像素。

其中 $\hat{y} : \Omega \rightarrow \{1, \dots, K\}$ 是每个像素的真实标签, $w : \Omega \rightarrow \mathbb{R}$ 是我们引入的权重图,以便在训练中赋予一些像素更多的重要性。

我们预先计算每个地面实况分割的权重图,以补偿训练数据集中某个类别的像素的不同频率,并强制网络学习我们在接触单元之间引入的小分离边界 (见图 3c 和 d)。

使用形态学运算计算分离边界。这然后将权重图计算为

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

其中 $w_c : \Omega \rightarrow \mathbb{R}$ 是平衡类频率的权重图, $d_1 : \Omega \rightarrow \mathbb{R}$ 表示到最近单元格边界的距离, $d_2 : \Omega \rightarrow \mathbb{R}$ 表示到第二最近单元格边界的距离。在我们的实验中,我们设置 $w_0 = 10$ 和 $\sigma \approx 5$ 个像素。

在具有许多卷积层和通过网络的不同路径的深度网络中,良好的权重初始化非常重要。否则,网络的某些部分可能会提供过多的激活,而其他部分则永远不会做出贡献。理想情况下,应该调整初始权重,使得网络中的每个特征图都具有近似单位方差。对于具有我们架构的网络 (交替卷积和 ReLU 层),这可以通过从标准差为 $\sqrt{2/N}$ 的高斯分布中提取初始权重来实现,其中 N 表示一个神经元的传入节点数 [5]。例如,对于前一层中的 3×3 卷积和 64 个特征通道, $N = 9 \cdot 64 = 576$ 。

3.1 数据增加

当只有很少的训练样本可用时,数据增强对于教会网络所需的不变性和鲁棒性是必不可少的。的情况下

显微图像我们主要需要平移和旋转不变性以及
对变形和灰度值变化的鲁棒性。特别是训练样本的随机弹性变形似乎是训练的关键概念

带有很少注释图像的分割网络。我们生成平滑
在粗略的 3 x 3 网格上使用随机位移矢量进行变形。这
位移是从具有 10 像素标准的高斯分布中采样的
偏差。然后使用双三次插值计算每像素位移。收缩路径末端的 drop-out 层进一步隐含
日期增加。

4个实验

我们演示了 u-net 在三种不同分割中的应用
任务。第一项任务是电子显微镜记录中神经元结构的分割。数据集示例和我们获得的分割

显示在图2 中。我们提供完整结果作为补充材料。
数据集由启动的 EM 分割挑战[14]提供
在 ISBI 2012 上,仍然对新的贡献开放。训练数据是一组
来自串行截面透射电子显微镜的 30 张图像 (512x512 像素)
果蝇第一龄幼虫腹神经索 (VNC)。每一张图片都来了
具有相应的完整注释的单元格地面实况分割图
(白色)和膜 (黑色)。测试集是公开的,但它的分割图是保密的。可以通过发送来获得评估

预测到组织者的膜概率图。评估完成
通过在 10 个不同级别对地图进行阈值处理并计算 “翘曲”
误差”、“随机误差”和 “像素误差” [14]。
u-net (输入数据的 7 个旋转版本的平均值)无需任何进一步的预处理或后处理即可实现
0.0003529 的翘曲误差 (新
最好的分数,见表1)和 0.0382 的随机误差。
这明显优于滑动窗口卷积网络
Ciresan 等人的结果。 [1],其最佳提交的翘曲误差为 0.000420
和 0.0504 的兰德误差。就 rand 误差而言,唯一表现更好的

表 1. EM 分割挑战排名[14] (2015 年 3 月 6 日) ,排序
通过翘曲错误。

等级 组名	翘曲误差	随机误差	像素误差
** 人类价值观 **	0.000005	0.0021	0.0010
1. u-net 2.	0.000353	0.0382	0.0611
DIVE-SCI 3. IDSIA	0.000355	0.0305	0.0584
[1]	0.000420	0.0504	0.0613
4. 潜水	0.000430	0.0545	0.0582
⋮			
10.IDSIA-SCI	0.000653	0.0189	0.1027

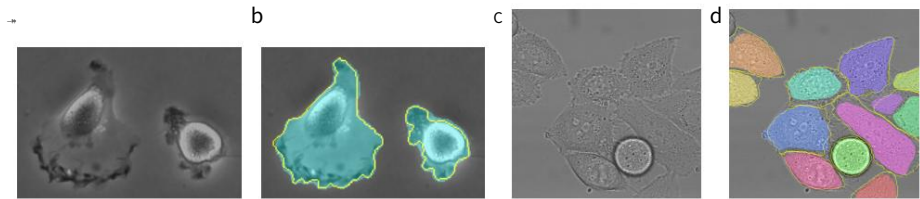


图 4. ISBI 小区跟踪挑战的结果。(a) 输入图像的一部分
“PhC-U373”数据集。(b) 具有手动地面实况的分割结果（青色蒙版）
（黄色边框）(c) “DIC-HeLa”数据集的输入图像。(d) 分割结果
（随机彩色面具）与手动地面实况（黄色边框）。

表 2. 2015 年 ISBI 小区跟踪挑战赛的分割结果 (IOU)。

姓名	PhC-U373	DIC-HeLa
IMCB-SG (2014) 0.2669 KTH-SE		0.2935
(2014) 0.7953 HOUS-US (2014)		0.4607
0.5323 第二好 2015 0.83		-
		0.46
你网 (2015)	0.9203	0.7756

此数据集上的算法使用高度数据集特定的后处理方法¹
应用于 Ciresan 等人的概率图。[1]。
我们还将 u-net 应用于光学显微图像中的细胞分割任务。此分割任务是 2014 年 ISBI
单元跟踪挑战的一部分，
2015 [10,13]。第一个数据集 “PhC-U373”²包含胶质母细胞瘤-星形细胞瘤
相差显微镜记录的聚丙烯酰亚胺基底上的 U373 细胞
（参见图4a,b 和补充材料）。它包含 35 个部分注释的训练图像。在这里,我们实现
了 92% 的平均 IOU（“intersection over union”），
这明显优于 83% 的第二好的算法（见表2）。第二个数据集 “DIC-HeLa”³是记录在平板
玻璃上的 HeLa 细胞
通过微分干涉对比 (DIC) 显微镜（见图3、图4c,d
和补充材料）。它包含 20 个部分注释的训练图像。在这里,我们
平均IOU达到77.5%,明显优于第二
46% 的最佳算法。

5 结论

u-net 架构在非常不同的生物医学分割应用程序上实现了非常好的性能。多亏了弹性 deform
的数据增强

¹ 该算法的作者提交了 78 种不同的解决方案来实现这一目标
结果。
² 数据集由 Sanjay Kumar 博士提供。生物工程大学系
加州伯克利分校。加州伯克利（美国）
³ 数据集由 Dr. Gert van Cappellen Erasmus 医疗中心。鹿特丹。
荷兰人

在 NVidia Titan GPU (6 GB) 上,它只需要很少的注释图像,并且非常合理的训练时间仅为 10 小时。我们提供完整的基于 Caffe[6] 的实现和经过训练的网络⁴。我们确信 u-net 架构可以轻松应用于更多任务。

致谢

这项研究得到了德国联邦和州政府的卓越计划 (EXC 294) 和 BMBF (Fkz 0316185B) 的支持。

参考

1. Ciresan, DC, Gambardella, LM, Giusti, A., Schmidhuber, J.: 深度神经网络在电子显微镜图像中分割神经元膜。在: NIPS。第 2852–2860 页 (2012 年)
2. Dosovitskiy, A., Springenberg, JT, Riedmiller, M., Brox, T.: 卷积神经网络的判别无监督特征学习。在: NIPS (2014)
3. Girshick, R., Donahue, J., Darrell, T., Malik, J.: 用于准确对象检测和语义分割的丰富特征层次结构。在: 计算机视觉和模式识别 (CVPR) IEEE 会议论文集 (2014)
4. Hariharan, B., Arbelaz, P., Girshick, R., Malik, J.: 用于对象分割和细粒度定位的超列 (2014), arXiv:1411.5752 [cs.CV]
5. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human level performance on imagenet classification (2015), arXiv:1502.01852 [cs.CV]
6. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: 快速特征的卷积架构嵌入 (2014), arXiv:1408.5093 [cs.CV]
7. Krizhevsky, A., Sutskever, I., Hinton, GE: 使用深度卷积神经网络的 Imagenet 分类。在: NIPS。第 1106–1114 页 (2012 年)
8. LeCun, Y., Boser, B., Denker, JS, Henderson, D., Howard, RE, Hubbard, W., Jackel, LD: 反向传播应用于手写邮政编码识别。神经计算 1(4), 541–551 (1989)
9. Long, J., Shelhamer, E., Darrell, T.: 语义的全卷积网络分割 (2014), arXiv:1411.4038 [cs.CV]
10. Maska, M., (...), de Solorzano, CO: 细胞跟踪比较的基准算法。生物信息学 30, 1609–1617 (2014)
11. Seyedhosseini, M., Sajjadi, M., Tasdizen, T.: 使用级联分层模型和逻辑分离法线网络进行图像分割。在: 计算机视觉 (ICCV), 2013 年 IEEE 国际会议上。第 2168–2175 页 (2013 年)
12. Simonyan, K., Zisserman, A.: 用于大规模图像识别的非常深的卷积网络 (2014), arXiv:1409.1556 [cs.CV]
13. WWW: 细胞跟踪挑战网页, http://www.codesolorzano.com/celltrackingchallenge/Cell_Tracking_Challenge/Welcome.html
14. WWW: em 分割挑战的网页, http://brainiac2.mit.edu/isbi_challenge/

⁴ U-net 实施、训练有素的网络和补充材料可在 <http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net> 获得