

U-Net: 用于生物医学图像分割的卷积网络

Olaf Ronneberger、Philipp Fischer 和 Thomas Brox

计算机科学系和 BIOS 生物信号研究中心,
德国弗莱堡大学

ronneber@informatik.uni-freiburg.de、

万维网主页: <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力使用来更有效地使用可用的带注释样本。该架构由捕获上下文的收缩路径和实现精确定位的对称扩展路径组成。我们表明,这样的网络可以从很少的图像中进行端到端训练,并且在电子显微镜堆栈中神经元结构分割的 ISBI 挑战赛中优于先前的最佳方法(滑动窗口卷积网络)。使用在透射光显微镜图像(相差和 DIC)上训练的相同网络,我们在这些类别中以大幅优势赢得了 2015 年 ISBI 细胞追踪挑战赛。而且,网络速度很快。在最新的 GPU 上,分割 512x512 图像只需不到一秒。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>。

1 介绍

在过去的两年里,深度卷积网络在许多视觉识别任务中都超越了最先进的技术,例如视觉识别任务。[7, 3]。虽然卷积网络已经存在很长一段时间了[8],但由于可用训练集的大小和所考虑的网络的大小,它们的成功受到限制。Krizhevsky 等人的突破。[7]是由于在具有 100 万张训练图像的 ImageNet 数据集上对具有 8 层和数百万个参数的大型网络进行了监督训练。从那时起,更大、更深的网络开始被训练[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,特别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,在生物医学任务中,数以千计的训练图像通常是遥不可及的。因此,CIRESAN 等人。[1]在滑动窗口设置中训练网络,通过提供该像素周围的局部区域(补丁)来预测每个像素的类标签

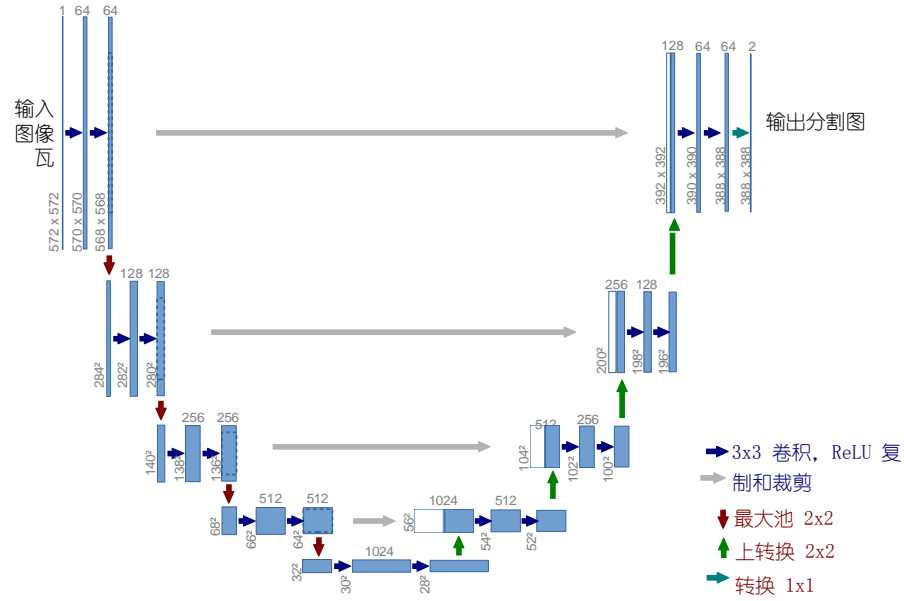


图 1. U-net 架构（最低分辨率下 32x32 像素的示例）。每个蓝色框对应一个多通道特征图。通道数显示在框的顶部。x-y-尺寸位于框的左下边缘。白框代表复制的特征图。箭头表示不同的操作。

作为输入。首先，这个网络可以定位。其次，训练数据的块数远大于训练图像的数量。由此产生的网络以大幅优势赢得了 ISBI 2012 的 EM 分割挑战赛。

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，并且由于补丁重叠而存在大量冗余。其次，定位精度和上下文的使用之间存在权衡。较大的补丁需要更多的最大池层，这会降低定位精度，而较小的补丁允许网络仅看到很少的上下文。最近的方法[11, 4]提出了一种考虑多层特征的分类器输出。良好的本地化和上下文的使用可以同时实现。

在本文中，我们构建了一种更优雅的架构，即所谓的“全卷积网络”[9]。我们修改和扩展了这种架构，使其能够使用很少的训练图像并产生更精确的分割；请参见图 1。[9]中的主要思想是通过连续层来补充通常的契约网络，其中池化算子被上采样算子取代。因此，这些层提高了输出的分辨率。为了定位，来自收缩路径的高分辨率特征与上采样的特征相结合

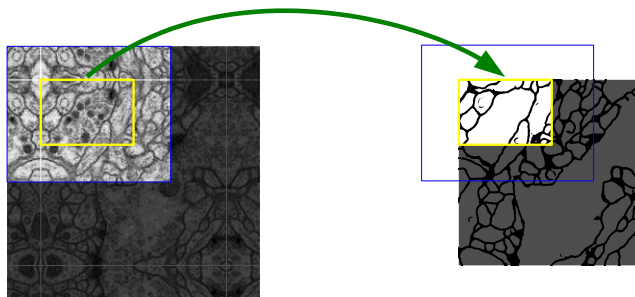


图 2. 用于任意大图像无缝分割的重叠平铺策略（此处是 EM 堆栈中神经元结构的分割）。黄色区域分割的预测需要蓝色区域内的图像数据作为输入。通过镜像推断丢失的输入数据

输出。然后，连续的卷积层可以学习根据这些信息组合更精确的输出。

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，扩张路径或多或少与收缩路径对称，并产生 U 形架构。该网络没有任何完全连接的层，仅使用每个卷积的有效部分，即分割图仅包含输入图像中可用的完整上下文的像素。该策略允许通过重叠平铺策略对任意大图像进行无缝分割（见图 2）。为了预测图像边界区域中的像素，通过镜像输入图像来推断丢失的上下文。这种平铺策略对于将网络应用于大图像非常重要，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。这使得网络能够学习此类变形的不变性，而无需在带注释的图像语料库中查看这些变换。这在生物医学分割中尤其重要，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同一类触摸对象的分离；参见图 3。为此，我们建议使用加权损失，其中接触细胞之间的分离背景标签在损失函数中获得较大的权重。

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（ISBI 2012 上开始的一项持续竞赛），其中我们胜出-

执行 Ciresan 等人的网络。[1]。此外，我们还展示了 2015 年 ISBI 细胞追踪挑战赛中光学显微镜图像中细胞分割的结果。在这里，我们在两个最具挑战性的 2D 透射光数据集上以大幅优势获胜。

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 3×3 卷积（未填充卷积）的重复应用组成，每个卷积后跟一个修正线性单元（ReLU）和一个步长为 2 的 2×2 最大池化操作，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都包含对特征图进行上采样，然后进行 2×2 卷积（“上卷积”），将特征通道数量减半，与收缩路径中相应裁剪的特征图进行串联，以及两个 3×3 卷积，每个卷积后跟一个 ReLU。由于每次卷积都会丢失边界像素，因此需要进行裁剪。在最后一层，使用 1×1 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平铺（参见图 2），选择输入平铺大小非常重要，以便将所有 2×2 最大池操作应用于具有均匀 x 和 y 大小的层。

3 训练

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，输出图像比输入小恒定的边框宽度。为了最大限度地减少开销并最大限度地利用 GPU 内存，我们更喜欢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决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逐像素 soft-max 与交叉熵损失函数相结合来计算的。软最大值是

定义为 $p(x) = \exp(a(x)) / \sum_{k=1}^K \exp(a^k(x))$ 其中 $a(x)$ 表示

在特征通道 k 的像素位置 $x \in \Omega$ 处激活，其中 $\Omega \subset \mathbb{Z}^2$ 。 K 是类别数， $p_k(x)$ 是近似最大函数。 \mathbb{E} 。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，交叉熵在每个位置处使用 $p_{f(x)}(x)$ 与 1 的偏差进行惩罚

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

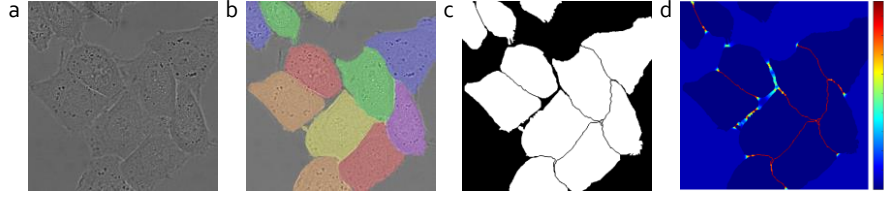


图 3. 用 DIC (微分干涉对比) 显微镜记录的玻璃上的 HeLa 细胞。(a) 原始图像。(b) 与地面实况分割重叠。不同的颜色表示 HeLa 细胞的不同实例。(c) 生成的分割掩模 (白色: 前景, 黑色: 背景)。(d) 使用逐像素损失权重进行映射, 以强制网络学习边界像素。

其中 $f: \Omega \rightarrow \{1, \dots, K\}$ 是每个像素的真实标签, $w: \Omega \rightarrow \mathbb{R}$ 是权重图, 我们引入它是为了在训练中赋予某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, 并迫使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)。

分离边界是使用形态学运算计算的。然后权重图计算为

$$w(x) = w_c(x) + \frac{12(d_1(x) + d_2(x))^2}{2\sigma^2} \quad (2)$$

w_0 ● 经验值

其中 $w_c: \Omega \rightarrow \mathbb{R}$ 是平衡类别频率的权重图, $d_1: \Omega \rightarrow \mathbb{R}$ 表示到最近单元格边界的距离, $d_2: \Omega \rightarrow \mathbb{R}$ 表示到第二个最近的单元格的边界。在我们的实验中, 我们设置 $w_0 = 10$ 和 $\sigma \approx 5$ 像素。

在具有许多卷积层和不同网络路径的深层网络中, 权重的良好初始化极其重要。否则, 网络的某些部分可能会产生过度的激活, 而其他部分则永远不会做出贡献。理想情况下, 应调整初始权重, 以使网络中的每个特征图具有大约单位方差。对于采用我们的架构 (交替卷积层和 ReLU 层) 的网络, 可以通过从标准差为 $2/N$ 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量 [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 数据增强

当只有很少的训练样本可用时, 数据增强对于教导网络所需的不变性和鲁棒性属性至关重要。的情况下

对于显微图像，我们主要需要平移和旋转不变性以及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具有很少注释图像的分割网络的关键概念。我们使用随机位移向量在粗略的 3×3 网格上生成平滑变形。位移是从标准差为 10 像素的高斯分布中采样的。然后使用双三次插值计算每像素位移。契约路径末端的退出层执行进一步的隐式数据增强。

4 实验

我们演示了 u-net 在三种不同分割任务中的应用。第一项任务是电子显微镜记录中神经元结构的分割。图 2 显示了数据集和我们获得的细分的示例。我们提供完整的结果作为补充材料。该数据集由 EM 分割挑战赛 [14] 提供，该挑战赛于 ISBI 2012 上启动，目前仍在接受新的贡献。训练数据是一组 30 张图像（512x512 像素），来自果蝇一龄幼虫腹神经索（VNC）的串行切片透射电子显微镜图像。每个图像都带有相应的完全注释的细胞（白色）和膜（黑色）的地面实况分割图。测试集是公开的，但其分割图是保密的。可以通过将预测的膜概率图发送给组织者来获得评估。评估是通过在 10 个不同级别对地图进行阈值处理并计算“扭曲误差”、“兰德误差”和“像素误差”[14] 来完成的。

u-net（对输入数据的 7 个旋转版本进行平均）无需任何进一步的预处理或后处理即可实现 0.0003529 的扭曲误差（新的最佳分数，请参阅表 1）和 0.0382 的兰德误差。

这明显优于 Cirezan 等人的滑动窗口卷积网络结果。[1]，其最佳提交的扭曲误差为 0.000420，兰德误差为 0.0504。就兰德误差而言，唯一表现更好的

表 1. EM 分割挑战排名 [14] (2015 年 3 月 6 日)，按扭曲误差排序。

秩	团队名字	扭曲误差	兰德误差	像素误差
	** 人类价值观 **	0.000005	0.0021	0.0010
1.	优网	0.000353	0.0382	0.0611
2.	潜水-SCI	0.000355	0.0305	0.0584
3.	伊迪西亚[1]	0.000420	0.0504	0.0613
4.	潜水	0.000430	0.0545	0.0582
.				
10.	IDSIA-SCI系列	0.000653	0.0189	0.1027

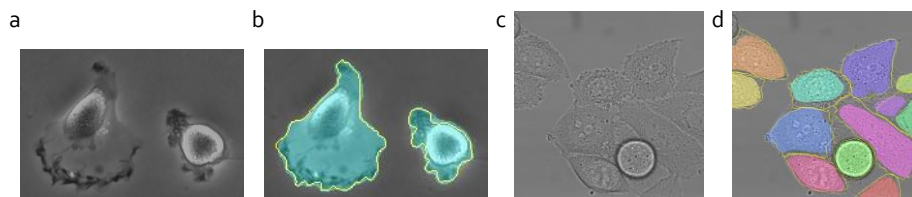


图 4. ISBI 细胞追踪挑战的结果。(a)“PhC-U373”数据集的输入图像的一部分。(b)手动地面实况(黄色边框)的分割结果(青色掩模)(c)“DIC-HeLa”数据集的输入图像。(d)具有手动地面实况(黄色边框)的分割结果(随机彩色掩模)。

表 2. 2015 年 ISBI 细胞追踪挑战赛的分割结果 (IOU)。

姓名	PHC-U373型	DIC-HeLa
IMCB-SG (2014 年)	0.2669	0.2935
KTH-SE (2014 年)	0.7953	0.4607
HOUS-美国 (2014)	0.5323	-
2015年第二好	0.83	0.46
优网 (2015)	0.9203	0.7756

该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 应用于光显微图像中的细胞分割任务。该分割任务是 2014 年和 2015 年 ISBI 细胞跟踪挑战赛的一部分 [10, 13]。第一个数据集“PhC-U373”²包含通过相差显微镜记录的聚丙烯酰亚胺基底上的胶质母细胞瘤-星形细胞瘤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 (“交集大于并集”)，这明显优于第二佳算法的 83% (见表 2)。第二个数据集“DIC-HeLa”³是通过微分干涉对比 (DIC) 显微镜记录的平板玻璃上的 HeLa 细胞(参见图 3、图 4c、d 和补充材料)。它包含 20 个部分注释的训练图像。这里我们实现了 77.5% 的平均 IOU，明显优于第二好的算法 46%。

5 结论

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。由于具有弹性变形的数据增强

¹该算法的作者提交了 78 种不同的解决方案来实现这一结果。

²数据集由 Sanjay Kumar 博士提供。加州大学伯克利分校生物工程系。加利福尼亚州伯克利(美国)

³数据集由 Gert van Cappellen Erasmus 医学中心博士提供。鹿特丹。荷兰人

它只需要很少的带注释的图像，并且在 NVidia Titan GPU (6 GB) 上的训练时间非常合理，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卓越计划 (EXC 294) 和 BMBF (Fkz 0316185B) 的支持。

参考文献

1. Ciresan, D.C.、Gambardella, L.M.、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)
2. Dosovitskiy, A.、Springenberg, J.T.、Riedmiller, M.、Brox, T.: 使用卷积神经网络进行判别性无监督特征学习。见: NIPS (2014)
3. Girshick, R.、Donahue, J.、Darrell, T.、Malik, J.: 用于精确对象检测和语义分割的丰富特征层次结构。见: IEEE 计算机视觉和模式识别会议 (CVPR) 会议记录 (2014)
4. Hariharan, B.、Arbelez, 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.: 深入研究整流器: 在 imagenet 分类上超越人类水平的表现 (2015), arXiv:1502.01852 [cs.CV]
6. Jia, Y.、Shelhamer, E.、Donahue, J.、Karayev, S.、Long, J.、Girshick, R.、Guadarrama, S.、Darrell, T.: Caffe: 用于快速特征嵌入的卷积架构 (2014 年), arXiv:1408.5093 [cs.CV]
7. Krizhevsky, A.、Sutskever, I.、Hinton, G.E.: 使用深度卷积神经网络进行 Imagenet 分类。在: NIPS。第 1106 - 1114 页 (2012)
8. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: 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.: 用于语义分割的全卷积网络 (2014), arXiv:1411.4038 [cs.CV]
10. Maska, M., (...), de Solorzano, C.O.: 细胞跟踪算法比较的基准。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析取正态网络进行图像分割。见: 计算机视觉 (ICCV), 2013 年 IEEE 国际会议。第 2168 - 2175 页 (2013)
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: 细胞跟踪挑战的网页, http://www.codesolorzano.com/celltrackingchallenge/Cell_Tracking_Challenge/Welcome.html
14. WWW: em 分段挑战的网页, http://brainiac2.mit.edu/isbi_challenge/

⁴U-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>