

U-Net - Convolutional Networks for Biomedical Image Segmentation 论文翻译 ——中英文对照

| 650

U-Net - Convolutional Networks for Biomedical Image Segmentation 论文翻译——中英文对照

文章作者：Tyan

博客：noahsnail.com | [CSDN](#) | [简书](#)

声明：作者翻译论文仅为学习，如有侵权请联系作者删除博文，谢谢！

翻译论文汇总：<https://github.com/SnailTyan/deep-learning-papers-translation>

U-Net: Convolutional Networks for Biomedical Image Segmentation

Abstract

There is large consent that successful training of deep networks requires many thousand annotated training samples. In this paper, we present a network and training strategy that relies on the strong use of data augmentation to use the available annotated samples more efficiently. The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. We show that such a network can be trained end-to-end from very few images and outperforms the prior best method (a sliding-window convolutional network) on the ISBI challenge for segmentation of neuronal structures in electron microscopic stacks. Using the same network trained on transmitted light microscopy images (phase contrast and DIC) we won the ISBI cell tracking challenge 2015 in these categories by a large margin. Moreover, the network is fast. Segmentation of a 512x512 image takes less than a second on a recent GPU. The full implementation (based on Caffe) and the trained networks are available at <http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net>.

摘要

许多人都赞同深度网络的成功训练需要大量标注的训练样本。在本文中，我们提出了一种网络及训练策略，它依赖于大量使用数据增强，以便更有效地使用获得的标注样本。这个架构包括捕获上下文的收缩路径和能够精

确定位的对称扩展路径。我们证明了这种网络可以从非常少的图像进行端到端训练，并且优于之前的ISBI赛挑战赛的最好方法（滑动窗口卷积网络），ISBI赛挑战赛主要是在电子显微镜堆叠中进行神经元结构分割。使用在透射光显微镜图像（相位衬度和DIC）上训练的相同网络，我们在这些类别中大幅度地赢得了2015年ISBI细胞追踪挑战赛。而且，网络速度很快。在最新的GPU上，分割一张512x512的图像不到一秒钟。网络的完整实现（基于Caffe）和预训练网络可在<http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net>上获得。

1 Introduction

In the last two years, deep convolutional networks have outperformed the state of the art in many visual recognition tasks, e.g. [7,3]. While convolutional networks have already existed for a long time [8], their success was limited due to the size of the available training sets and the size of the considered networks. The breakthrough by Krizhevsky et al. [7] was due to supervised training of a large network with 8 layers and millions of parameters on the ImageNet dataset with 1 million training images. Since then, even larger and deeper networks have been trained [12].

1 引言

在过去两年，深度卷积网络在许多视觉识别任务中的表现都优于当前的最新技术，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大小和所考虑网络的规模，它们的成功受到了限制。Krizhevsky等人[7]的突破是通过大型网络在ImageNet数据集上的监督训练实现的，其中大型网络有8个网络层和数百万参数，ImageNet数据集包含百万张训练图像。从那时起，即使更大更深的网络也已经得到了训练[12]。

The typical use of convolutional networks is on classification tasks, where the output to an image is a single class label. However, in many visual tasks, especially in biomedical image processing, the desired output should include localization, i.e., a class label is supposed to be assigned to each pixel. Moreover, thousands of training images are usually beyond reach in biomedical tasks. Hence, Ciresan et al. [1] trained a network in a sliding-window setup to predict the class label of each pixel by providing a local region (patch) around that pixel as input. First, this network can localize. Secondly, the training data in terms of patches is much larger than the number of training images. The resulting network won the EM segmentation challenge at ISBI 2012 by a large margin.

卷积网络的典型用途是分类任务，其中图像输出是单个的类别标签。然而，在许多视觉任务中，尤其是在生物医学图像处理中，期望的输出应该包括位置，即类别标签应该分配给每个像素。此外，生物医学任务中通常无法获得数千张训练图像。因此，Ciresan等人[1]在滑动窗口设置中训练网络，通过提供像素周围局部区域（patch）作为输入来预测每个像素的类别标签。首先，这个网络可以定位。其次，局部块方面的训练数据远大于训练图像的数量。由此产生的网络大幅度地赢得了ISBI 2012EM分割挑战赛。

Obviously, the strategy in Ciresan et al. [1] has two drawbacks. First, it is quite slow because the network must be run separately for each patch, and there is a lot of redundancy due to overlapping patches. Secondly, there is a trade-off between localization accuracy and the use of context. Larger patches require more max-pooling layers that reduce the localization accuracy, while small patches allow the network to see only little context. More recent approaches [11,4] proposed a classifier output that takes into account the features from multiple layers. Good localization and the use of context are possible at the same time.

显然，Ciresan等人[1]的策略有两个缺点。首先，它非常慢，因为必须为每个图像块单独运行网络，并且由于图像块重叠而存在大量冗余。其次，定位准确性与上下文的使用之间存在着权衡。较大的图像块需要更多的最大池化层，从而降低了定位精度，而较小的图像块则允许网络只能看到很少的上下文。许多最近的方法[11，4]提出了一种分类器输出，其考虑了来自多个层的特征。同时具有良好的定位和上下文的使用是可能的。

In this paper, we build upon a more elegant architecture, the so-called “fully convolutional network” [9]. We modify and extend this architecture such that it works with very few training images and yields more precise segmentations; see Figure 1. The main idea in [9] is to supplement a usual contracting network by successive layers, where pooling operators are replaced by upsampling operators. Hence, these layers increase the resolution of the output. In order to localize, high resolution features from the contracting path are combined with the upsampled output. A successive convolution layer can then learn to assemble a more precise output based on this information.

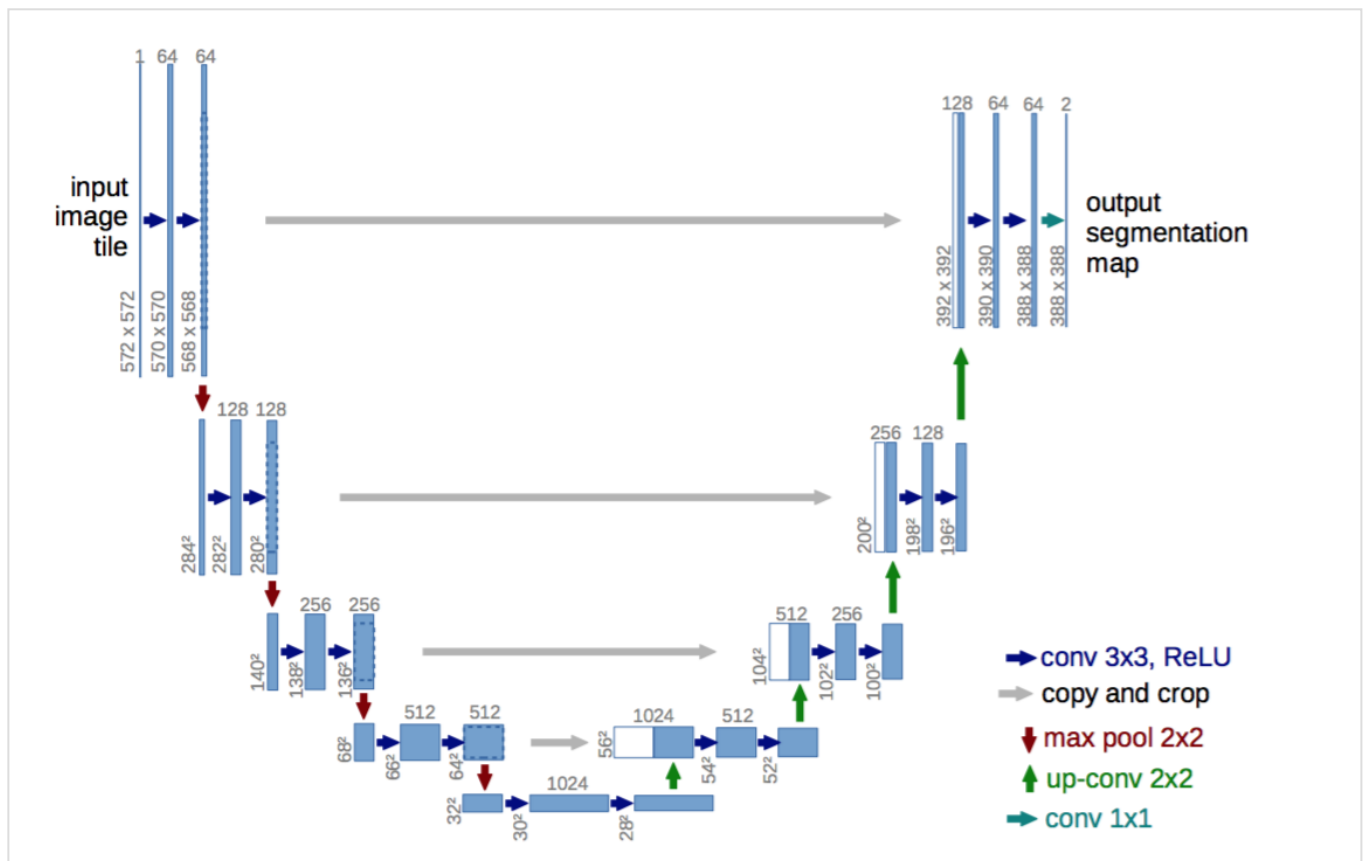


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows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

在本文中，我们构建了一个更优雅的架构，即所谓的“全卷积网络”[9]。我们对这种架构进行了修改和扩展，使得它只需很少的训练图像就可以取得更精确的分割；参见图1。[9]中的主要思想是通过连续层补充通常的收缩网络，其中的池化运算符由上采样运算符替换。因此，这些层增加了输出的分辨率。为了进行定位，来自收缩路径的高分辨率特征与上采样输出相结合。然后，后续卷积层可以基于该信息学习组装更精确的输出。

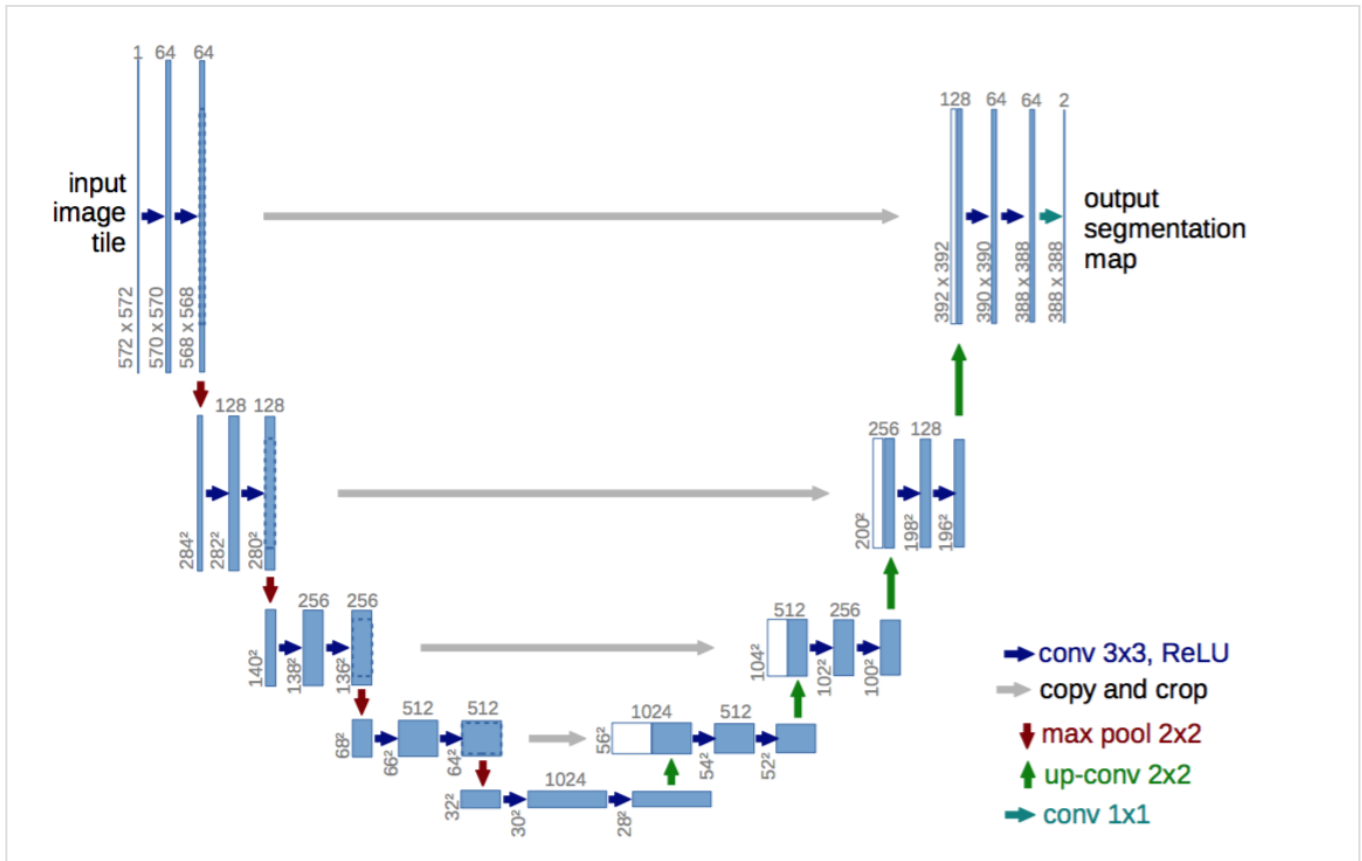


图1. U-net架构（最低分辨率为32x32像素的示例）。每个蓝色框对应于一张多通道特征映射。通道数在框的顶部。x-y 尺寸提供在框的左下边。白框表示复制的特征映射。箭头表示不同的操作。

One important modification in our architecture is that in the upsampling part we have also a large number of feature channels, which allow the network to propagate context information to higher resolution layers. As a consequence, the expansive path is more or less symmetric to the contracting path, and yields a u-shaped architecture. The network does not have any fully connected layers and only uses the valid part of each convolution, i.e., the segmentation map only contains the pixels, for which the full context is available in the input image. This strategy allows the seamless segmentation of arbitrarily large images by an overlap-tile strategy (see Figure 2). To predict the pixels in the border region of the image, the missing context is

extrapolated by mirroring the input image. This tiling strategy is important to apply the network to large images, since otherwise the resolution would be limited by the GPU memory.

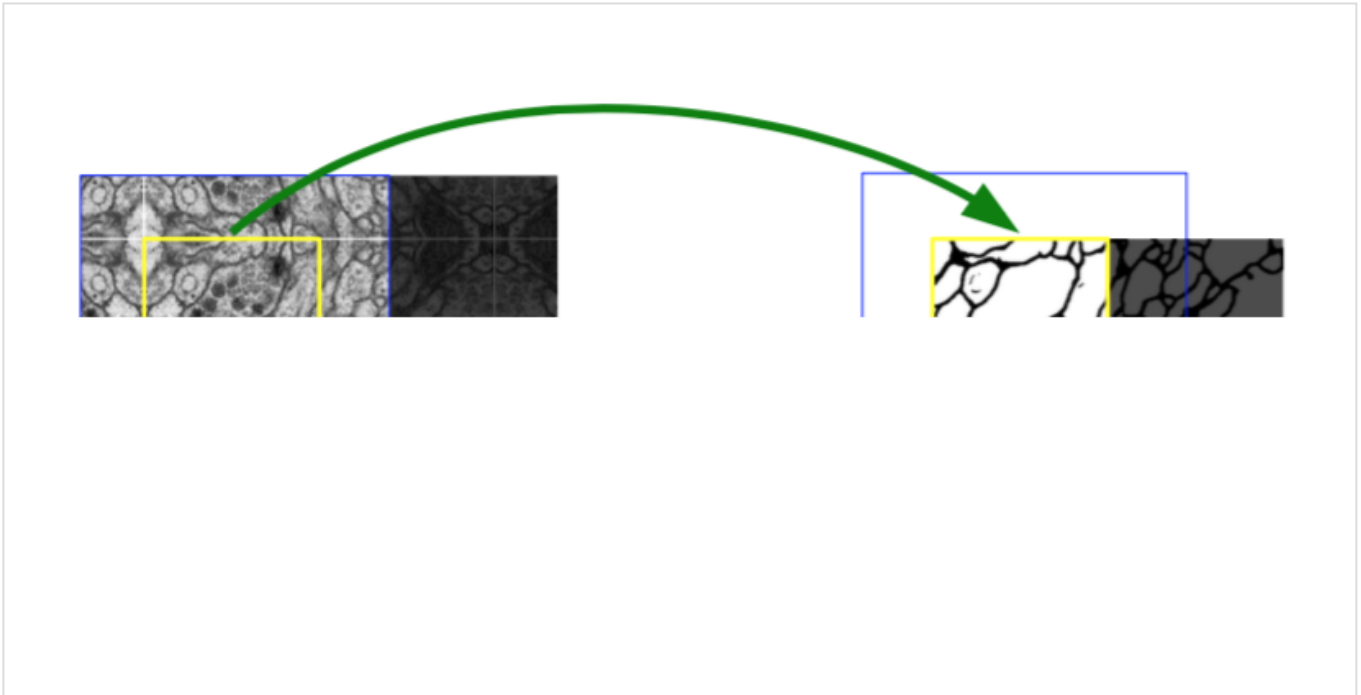


Fig. 2. Overlap-tile strategy for seamless segmentation of arbitrary large images (here segmentation of neuronal structures in EM stacks). Prediction of the segmentation in the yellow area, requires image data within the blue area as input. Missing input data is extrapolated by mirroring

我们架构中的一个重要修改是在上采样部分中我们还有大量的特征通道，这些通道允许网络将上下文信息传播到具有更高分辨率的层。因此，扩展路径或多或少地与收缩路径对称，并产生U形结构。网络没有任何全连接层，并且仅使用每个卷积的有效部分，即分割映射仅包含在输入图像中可获得完整上下文的像素。该策略允许通过重叠图像区策略无缝分割任意大小的图像（参见图2）。为了预测图像边界区域中的像素，通过镜像输入图像来外推缺失的上下文。这种图像块策略对于将网络应用于大的图像非常重要，否则分辨率将受到GPU内存的限制。

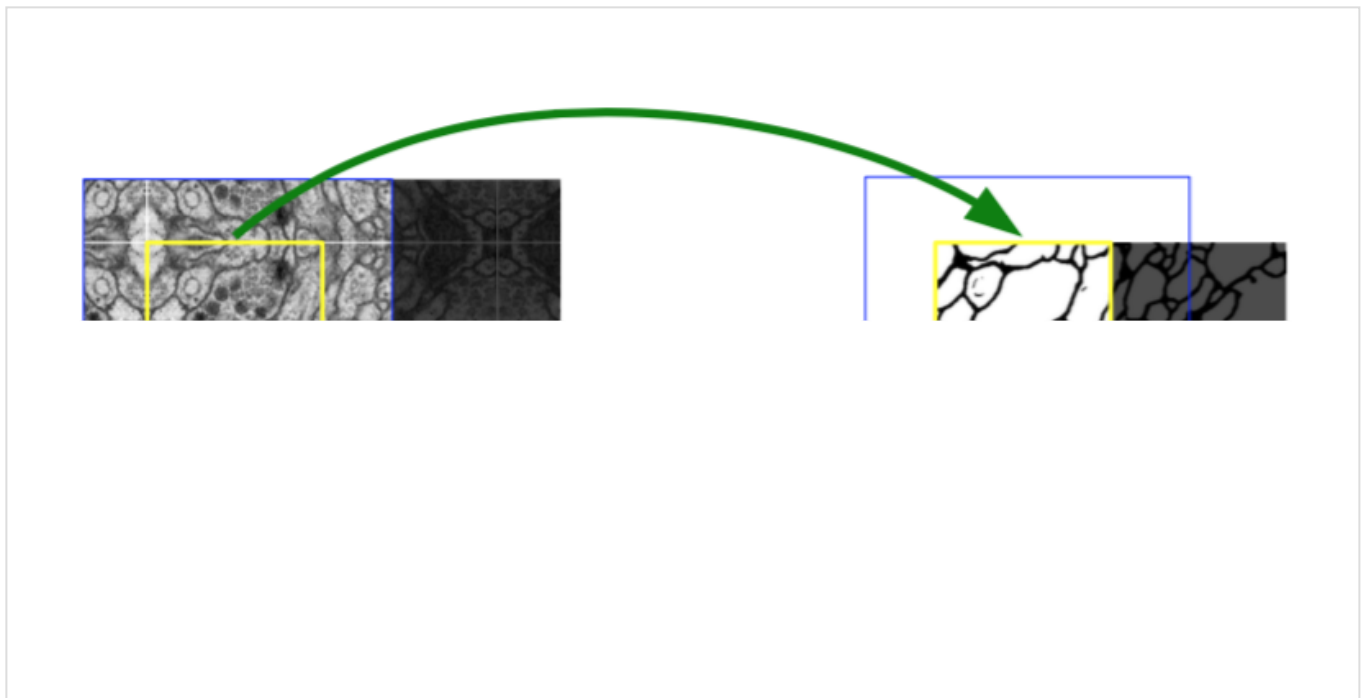


图2. 重叠图像块策略可以无缝分割任意大小的图像（EM堆叠中的神经元结构分割）。分割的预测在黄色区域，要求蓝色区域的图像数据作为输入。缺失的输入数据通过镜像外推。

As for our tasks there is very little training data available, we use excessive data augmentation by applying elastic deformations to the available training images. This allows the network to learn invariance to such deformations, without the need to see these transformations in the annotated image corpus. This is particularly important in biomedical segmentation, since deformation used to be the most common variation in tissue and realistic deformations can be simulated efficiently. The value of data augmentation for learning invariance has been shown in Dosovitskiy et al. [2] in the scope of unsupervised feature learning.

对于我们的任务，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更多的数据增强。这允许网络学习这种变形的不变性，而不需要在标注图像语料库中看到这些变形。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[2]在无监督特征学习的领域内已经证明了数据增强在学习不变性中的价值。

Another challenge in many cell segmentation tasks is the separation of touching objects of the same class; see Figure 3. To this end, we propose the use of a weighted loss, where the separating background labels between touching cells obtain a large weight in the loss function.

□

Fig. 3. HeLa cells on glass recorded with DIC (differential interference contrast) microscopy. (a) raw image. (b) overlay with ground truth segmentation. Different colors indicate different instances of the HeLa cells. (c) generated segmentation mask (white: foreground, black: background). (d) map with a pixel-wise loss weight to force the network to learn the border pixels.

许多细胞分割任务中的另一个挑战是分离同类的接触目标，见图3。为此，我们建议使用加权损失，其中接触单元之间的分离背景标签在损失函数中获得较大的权重。

□

图3. 用DIC（差异干涉对比）显微镜记录玻璃上的HeLa细胞。（a）原始图像。（b）覆盖的实际分割。不同的颜色表示不同的HeLa细胞实例。（c）生成分割掩码（白色：前景，黑色：背景）。（d）以像素损失权重的映射来迫使网络学习边界像素。

The resulting network is applicable to various biomedical segmentation problems. In this paper, we show results on the segmentation of neuronal structures in EM stacks (an ongoing competition started at ISBI 2012), where we outperformed the network of Ciresan et al. [1]. Furthermore, we show results for cell segmentation in light microscopy images from the ISBI cell tracking challenge 2015. Here we won with a large margin on the two most challenging 2D transmitted light datasets.

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堆叠中神经元结构的分割结果（从ISBI 2012开始的持续竞赛），其中我们的表现优于Ciresan等人[1]的网络。此外，我们展示了2015 ISBI细胞追踪挑战赛光学显微镜图像中的细胞分割结果。我们在两个最具挑战性的2D透射光数据集上以巨大的优势赢得了比赛。

2 Network Architecture

The network architecture is illustrated in Figure 1. It consists of a contracting path (left side) and an expansive path (right side). The contracting path follows the typical architecture of a convolutional network. It consists of the repeated application of two 3x3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling. At each downsampling step we double the number of feature channels. Every step in the expansive path consists of an upsampling of the feature map followed by a 2x2 convolution (“up-convolution”) that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU. The cropping is necessary due to the loss of border pixels in every convolution. At the final layer a 1x1 convolution is used to map each 64-component feature vector to the desired number of classes. In total the network has 23 convolutional layers.

2 网络架构

网络架构如图1所示。它由一个收缩路径（左侧）和一个扩展路径（右侧）组成。收缩路径遵循卷积网络的典型架构。它包括重复使用两个3x3卷积（无填充卷积），每个卷积后跟一个线性修正单元（ReLU）和一个2x2最大池化操作，步长为2的下采样。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特征映射的上采样，然后进行2x2卷积（“向上卷积”），将特征通道数量减半，与来自收缩路径的相应

裁剪特征映射串联，然后是两个3x3卷积，每个卷积后面接ReLU。由于每一次卷积都会丢失边界像素，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别上。网络总共有23个卷积层。

To allow a seamless tiling of the output segmentation map (see Figure 2), it is important to select the input tile size such that all 2x2 max-pooling operations are applied to a layer with an even x- and y-size.

为了允许输出分割映射的无缝平铺（参见图2），选择输入的图像块大小非常重要，这样所有的2x2最大池化操作都可以应用在具有偶数x和偶数y大小的层上。

3 Training

The input images and their corresponding segmentation maps are used to train the network with the stochastic gradient descent implementation of Caffe [6]. Due to the unpadded convolutions, the output image is smaller than the input by a constant border width. To minimize the overhead and make maximum use of the GPU memory, we favor large input tiles over a large batch size and hence reduce the batch to a single image. Accordingly we use a high momentum (0.99) such that a large number of the previously seen training samples determine the update in the current optimization step.

3 训练

使用输入图像及其相应的分割映射来训练带有随机梯度下降的网络，网络采用Caffe[6]实现。由于无填充卷积，输出图像比输入少恒定的边界宽度。为了最小化开销并最大限度地利用GPU内存，我们倾向于在大批量数据大小的情况下使用大的输入图像块，从而将批量数据大小减少到单张图像。因此，我们使用高动量值（0.99），大量先前看到的训练样本来决定当前优化步骤中的更新。

The energy function is computed by a pixel-wise soft-max over the final feature map combined with the cross entropy loss function. The soft-max is defined as $p_k(\mathbf{x}) = \exp(a_k(\mathbf{x})) / (\sum_{k'=1}^K \exp(a_{k'}(\mathbf{x})))$ where $a_k(\mathbf{x})$ denotes the activation in feature channel k at the pixel position $\mathbf{x} \in \Omega$ with $\Omega \subset \mathbb{Z}^2$. K is the number of classes and $p_k(\mathbf{x})$ is the approximated maximum-function. i.e. $p_k(\mathbf{x}) \approx 1$ for the k that has the maximum activation $a_k(\mathbf{x})$ and $p_k(\mathbf{x}) \approx 0$ for all other k . The cross entropy then penalizes at each position the deviation of $p_{l(x)}(\mathbf{x})$ from 1 using

$$E = \sum_{\mathbf{x} \in \Omega} \omega(\mathbf{x}) \log(p_{l(x)}(\mathbf{x})) \quad (1)$$

where $l : \Omega \rightarrow 1, \dots, K$ is the true label of each pixel and $\omega : \Omega \rightarrow \mathbb{R}$ is a weight map that we introduced to give some pixels more importance in the training.

能量函数由最终的特征映射上逐像素soft-max与交叉熵损失函数结合计算而成。soft-max定义为

$p_k(\mathbf{x}) = \exp(a_k(\mathbf{x})) / (\sum_{k'=1}^K \exp(a_{k'}(\mathbf{x})))$ ，其中 $a_k(\mathbf{x})$ 表示特征通道 k 中在像素位置 $\mathbf{x} \in \Omega$ 上的激活， $\Omega \subset \mathbb{Z}^2$ 。 K 是类别数量， $p_k(\mathbf{x})$ 是近似的最大化函数，即，对于有最大激活 $a_k(\mathbf{x})$ 的 k ， $p_k(\mathbf{x}) \approx 1$ ，对于其它的 k 有 $p_k(\mathbf{x}) \approx 0$ 。交叉熵在每个位置上使用

$$E = \sum_{\mathbf{x} \in \Omega} \omega(\mathbf{x}) \log(p_{l(\mathbf{x})}(\mathbf{x})) \quad (1)$$

来惩罚 $p_{l(\mathbf{x})}(\mathbf{x})$ 与1的偏差。其中， $l: \Omega \rightarrow 1, \dots, K$ 是每个像素的真实标签， $\omega: \Omega \rightarrow \mathbb{R}$ 是训练中我们引入的用来赋予某些像素更多权重的权重图。

We pre-compute the weight map for each ground truth segmentation to compensate the different frequency of pixels from a certain class in the training data set, and to force the network to learn the small separation borders that we introduce between touching cells (See Figure 3c and d).

我们为每一个真实分割预先计算了权重图，以补偿训练集里某个类别的像素的不同频率，并且迫使网络学习我们在相邻细胞间的引入小的分割边界（参见图3c和d）。

The separation border is computed using morphological operations. The weight map is then computed as

$$\omega(\mathbf{x}) = \omega_c(\mathbf{x}) + \omega_0 \bullet \exp\left(-\frac{(d_1(\mathbf{x}) + d_2(\mathbf{x}))^2}{2\sigma^2}\right) \quad (2)$$

where $\omega_c: \Omega \rightarrow \mathbb{R}$ is the weight map to balance the class frequencies, $d_1: \Omega \rightarrow \mathbb{R}$ denotes the distance to the border of the nearest cell and $d_2: \Omega \rightarrow \mathbb{R}$ the distance to the border of the second nearest cell. In our experiments we set $\omega_0 = 10$ and $\sigma \approx 5$ pixels.

分割边界使用形态学操作来计算。然后将权重图计算为

$$\omega(\mathbf{x}) = \omega_c(\mathbf{x}) + \omega_0 \bullet \exp\left(-\frac{(d_1(\mathbf{x}) + d_2(\mathbf{x}))^2}{2\sigma^2}\right) \quad (2)$$

其中， $\omega_c: \Omega \rightarrow \mathbb{R}$ 是用来平衡类频率的权重图， $d_1: \Omega \rightarrow \mathbb{R}$ 表示到最近细胞边界的距离， $d_2: \Omega \rightarrow \mathbb{R}$ 表示到次近细胞边界的距离。在我们的实验中，设置 $\omega_0 = 10$ ， $\sigma \approx 5$ 个像素。

In deep networks with many convolutional layers and different paths through the network, a good initialization of the weights is extremely important. Otherwise, parts of the network might give excessive activations, while other parts never contribute. Ideally the initial weights should be adapted such that each feature map in the network has approximately unit variance. For a network with our architecture (alternating convolution and ReLU layers) this can be achieved by drawing the initial weights from a Gaussian distribution with a standard deviation of $\sqrt{2/N}$, where N denotes the number of incoming nodes of one neuron [5]. E.g. for a 3x3 convolution and 64 feature channels in the previous layer $N = 9 \cdot 64 = 576$.

在具有许多卷积层和通过网络的不同路径的深度网络中，权重的良好初始化非常重要。否则，网络的某些部分可能会进行过多的激活，而其他部分永远不会起作用。理想情况下，初始化权重应该是自适应的，以使网络中的每个特征映射都具有近似的单位方差。对于具有我们架构（交替卷积和ReLU层）的网络，可以通过从标准偏差为 $\sqrt{2/N}$ 的高斯分布中绘制初始化权重来实现，其中 N 表示一个神经元[5]传入结点的数量。例如，对于前一层中3x3卷积和64个特征通道， $N = 9 \cdot 64 = 576$ 。

3.1 Data Augmentation

Data augmentation is essential to teach the network the desired invariance and robustness properties, when only few training samples are available. In case of microscopical images we primarily need shift and rotation invariance as well as robustness to deformations and gray value variations. Especially random elastic deformations of the training samples seem to be the key concept to train a segmentation network with very few annotated images. We generate smooth deformations using random displacement vectors on a coarse 3 by 3 grid. The displacements are sampled from a Gaussian distribution with 10 pixels standard deviation. Per-pixel displacements are then computed using bicubic interpolation. Drop-out layers at the end of the contracting path perform further implicit data augmentation.

3.1 数据增强

当只有少量训练样本可用时，对于教网络学习所需的不变性和鲁棒性而言，数据增强至关重要。对于显微镜图像，我们主要需要平移和旋转不变性，以及对形变和灰度值变化的鲁棒性。尤其是训练样本的随机弹性形变似乎是训练具有很少标注图像的分割网络的关键概念。我们使用在3x3粗糙网格上的随机位移矢量来生成平滑形变。从具有10个像素标准偏差的高斯分布中采样位移。然后使用双三次插值计算每个像素的位移。收缩路径末端的丢弃层执行进一步隐式数据增强。

4 Experiments

We demonstrate the application of the u-net to three different segmentation tasks. The first task is the segmentation of neuronal structures in electron microscopic recordings. An example of the data set and our obtained segmentation is displayed in Figure 2. We provide the full result as Supplementary Material. The data set is provided by the EM segmentation challenge [14] that was started at ISBI 2012 and is still open for new contributions. The training data is a set of 30 images (512x512 pixels) from serial section transmission electron microscopy of the *Drosophila* first instar larva ventral nerve cord (VNC). Each image comes with a corresponding fully annotated ground truth segmentation map for cells (white) and membranes (black). The test set is publicly available, but its segmentation maps are kept secret. An evaluation can be obtained by sending the predicted membrane probability map to the organizers. The evaluation is done by thresholding the map at 10 different levels and computation of the “warping error”, the “Rand error” and the “pixel error” [14].

4 实验

我们演示了U-Net在三个不同分割任务中的应用。第一个任务是在电子显微记录中分割神经元结构。图2显示了数据集样本以及获得的分割。我们在补充材料中提供了完整的结果。数据集由EM分割挑战赛[14]提供，该挑战始于ISBI 2012，目前仍在接受新的贡献。训练数据是一组来自果蝇一龄幼虫腹神经索（VNC）的连续切片透射电镜的30张图像（512x512像素）。每张图像都有一个对应的细胞（白色）和膜（黑色）完整标注的实际分割图。该测试集是公开可获得的，但其分割图像是保密的。评估可以通过将预测的膜概率图发送给组织者来获得。通过在10个不同级别对映射进行阈值化并计算“弯曲误差”，“兰德误差”和“像素误差”来进行评估[14]。

The u-net (averaged over 7 rotated versions of the input data) achieves without any further pre- or postprocessing a warping error of 0.0003529 (the new best score, see Table 1) and a of 0.0382.

Table 1. Ranking on the EM segmentation challenge [14] (march 6th, 2015), sorted by warping error.

□

U-Net（输入数据7个旋转版本的平均）在没有任何进一步的“前”或后处理的情况下实现了0.0003529的“弯曲误差”（新的最佳分数，参见表1）和0.0382的“兰德误差”。

表1：EM分割挑战14的排名，按warping error排序。

□

This is significantly better than the sliding-window convolutional network result by Ciresan et al., whose best submission had a warping error of 0.000420 and a rand error of 0.0504. In terms of rand error the only better performing algorithms on this data set use highly data set specific post-processing methods applied to the probability map of Ciresan et al. [1].

这比Ciresan等人[1]的滑动窗口卷积网络结果要好得多，其最佳提交的弯曲误差为0.000420，兰德误差为0.0504。就兰德误差而言，在该数据集上唯一表现更好的算法，其使用了应用到Ciresan等[1]概率图上的针对数据集的非常特定后处理方法。

We also applied the u-net to a cell segmentation task in light microscopic images. This segmenation task is part of the ISBI cell tracking challenge 2014 and 2015 [10,13]. The first data set “PhC-U373” contains Glioblastoma-astrocytoma U373 cells on a polyacrylimide substrate recorded by phase contrast microscopy (see Figure 4a,b and Supp. Material). It contains 35 partially annotated training images. Here we achieve an average IOU (“intersection over union”) of 92%, which is significantly better than the second best algorithm with 83% (see Table 2). The second data set “DIC-HeLa” are HeLa cells on a flat glass recorded by differential interference contrast (DIC) microscopy (see Figure 3, Figure 4c,d and Supp. Material). It contains 20 partially annotated

training images. Here we achieve an average IOU of 77.5% which is significantly better than the second best algorithm with 46%.

□

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).

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

□

我们还将U-Net应用于光学显微图像中的细胞分割任务。这个分割任务是ISBI细胞跟踪挑战赛2014和2015的一部分[10,13]。第一个数据集“PhC-U373”包含在聚丙烯酰亚胺基质上通过相衬显微技术记录的多形性胶质母细胞瘤U373细胞（参见图4a，b和补充材料）。它包含35个部分标注的训练图像。这里，我们取得了92%的平均IOU（“并集上的交集”），明显好于83%的次优算法（参见表2）。第二个数据集“DIC-HeLa”是通过微分干涉相差(DIC)显微镜记录的平板玻璃上的HeLa细胞（请参见图3，图4c，d和补充材料）。它包含20个部分标注的训练图像。这里，我们取得了77.5%的平均IOU，明显好于46%的次优算法。

□

图4：ISBI细胞跟踪挑战赛上的结果。(a)“PhC-U373”数据集的一张输入图像的一部分。(b)分割结果（蓝绿图像块）和实际结果（黄色边框）。(c)“DIC-HeLa”数据集的输入图像。(d)分割结果（随机颜色的图像块）和实际结果（黄色边框）。

表2：ISBI细胞跟踪挑战赛2015上的分割结果(IOU)。

□

5 Conclusion

The u-net architecture achieves very good performance on very different biomedical segmentation applications. Thanks to data augmentation with elastic deformations, it only needs very few annotated images and has a very reasonable training time of only 10 hours on a NVidia Titan GPU (6 GB). We provide the full Caffe[6]-based implementation and the trained networks. We are sure that the u-net architecture can be applied easily to many more tasks.

5 结论

U-Net架构在截然不同的生物医疗分割应用中取得了非常好的性能。由于具有弹性形变的数据增强，它仅需要非常少的标注图像，并且在NVidia Titan GPU (6 GB)上仅需要10个小时的合理训练时间。我们提供了完整的基

于Caffe[6]的实现以及训练之后的网络。我们确信U-Net架构可以很轻松地应用到更多的任务上。

Acknowledgements

This study was supported by the Excellence Initiative of the German Federal and State governments (EXC 294) and by the BMBF (Fkz 0316185B).

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。

References

1. Ciresan, D.C., Gambardella, L.M., Giusti, A., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images. In: NIPS. pp. 2852–2860 (2012)
2. Dosovitskiy, A., Springenberg, J.T., Riedmiller, M., Brox, T.: Discriminative unsupervised feature learning with convolutional neural networks. In: NIPS (2014)
3. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014)
4. Hariharan, B., Arbelaz, P., Girshick, R., Malik, J.: Hypercolumns for object segmentation and fine-grained localization (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: Convolutional architecture for fast feature embedding (2014), arXiv:1408.5093 [cs.CV]
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS. pp. 1106–1114 (2012)
8. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. Neural Computation 1(4), 541–551 (1989)

9. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation (2014), arXiv:1411.4038 [cs.CV]
10. Maska, M., (...), de Solorzano, C.O.: A benchmark for comparison of cell tracking algorithms. Bioinformatics 30, 1609–1617 (2014)
11. Seyedhosseini, M., Sajjadi, M., Tasdizen, T.: Image segmentation with cascaded hierarchical models and logistic disjunctive normal networks. In: Computer Vision (ICCV), 2013 IEEE International Conference on. pp. 2168–2175 (2013)
12. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014), arXiv:1409.1556 [cs.CV]
13. WWW: Web page of the cell tracking challenge, http://www.codesolorzano.com/celltrackingchallenge/Cell_Tracking_Challenge/Welcome.html
14. WWW: Web page of the em segmentation challenge, http://brainiac2.mit.edu/isbi_challenge/

如果有收获，可以请我喝杯咖啡！

赏

Deep Learning

◀ Linux登录信息查询

U-Net - Convolutional Networks for
Biomedical Image Segmentation论文翻译——中
文版 ▶

© 2016 - 2020 Tyan

👤 292393 | 👁 539883