U-Net: Convolutional Networks for Biomedical Image Segmentation

Olaf Ronneberger, Philipp Fischer, and Thomas Brox

Computer Science Department and BIOSS Centre for Biological Signalling Studies,
University of Freiburg, Germany
ronneber@informatik.uni-freiburg.de,
WWW home page: http://lmb.informatik.uni-freiburg.de/

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.

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].

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

U-Net: Convolutional Networks for Biomedical Image Segmentation

奥拉夫·罗内伯格、菲利普·菲舍尔和托马斯·布罗克斯

弗莱堡大学计算机科学系与生物信号研究中心（BIOSS）
ronneber@informatik.uni-freiburg.de，WWW主页：
http://lmb.informatik.uni-freiburg.de/

Abstract. 普遍认为，深度网络的成功训练需要数千个标注样本。本文提出了一种网络和训练策略，通过大量使用数据增强技术，更有效地利用现有标注样本。该架构包含捕捉上下文的收缩路径和实现精确定位的对称扩展路径。我们证明，这种网络可以从极少量图像进行端到端训练，并在ISBI电子显微镜堆栈神经元结构分割挑战中超越了先前最佳方法（滑动窗口卷积网络）。使用相同网络在透射光显微镜图像（相差显微镜和微分干涉对比显微镜）上训练后，我们在2015年ISBI细胞追踪挑战赛中以显著优势赢得了这些类别的冠军。此外，该网络速度极快，在最新GPU上分割512x512图像仅需不到一秒。完整实现（基于Caffe）及训练好的网络可在http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net获取。

1 Introduction

在过去的两年中，深度卷积网络在许多视觉识别任务中超越了现有技术水平，例如[7,3]。尽管卷积网络早已存在[8]，但由于可用训练集的规模和所考虑网络的规模限制，其成功曾受限。Krizhevsky等人[7]的突破在于，在包含100万张训练图像的ImageNet数据集上，通过监督训练一个拥有8层和数百万参数的大型网络得以实现。此后，更大更深的网络也相继被训练出来[12]。

卷积网络的典型用途是分类任务，即对图像输出单个类别标签。然而，在许多视觉任务中，尤其是在生物医学图像处理领域，期望的输出应包含定位信息，也就是说，需要为每个像素分配一个类别标签。此外，在生物医学任务中，通常难以获取数千张训练图像。因此，Ciresan等人[1]采用滑动窗口设置训练网络，通过提供每个像素周围的局部区域（图像块）来预测该像素的类别标签。

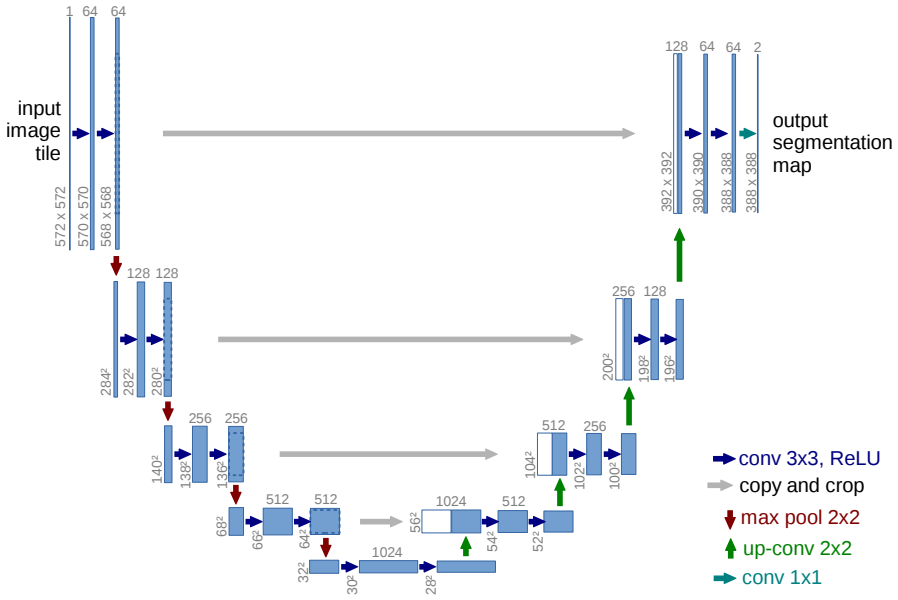

Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows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

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.

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.

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

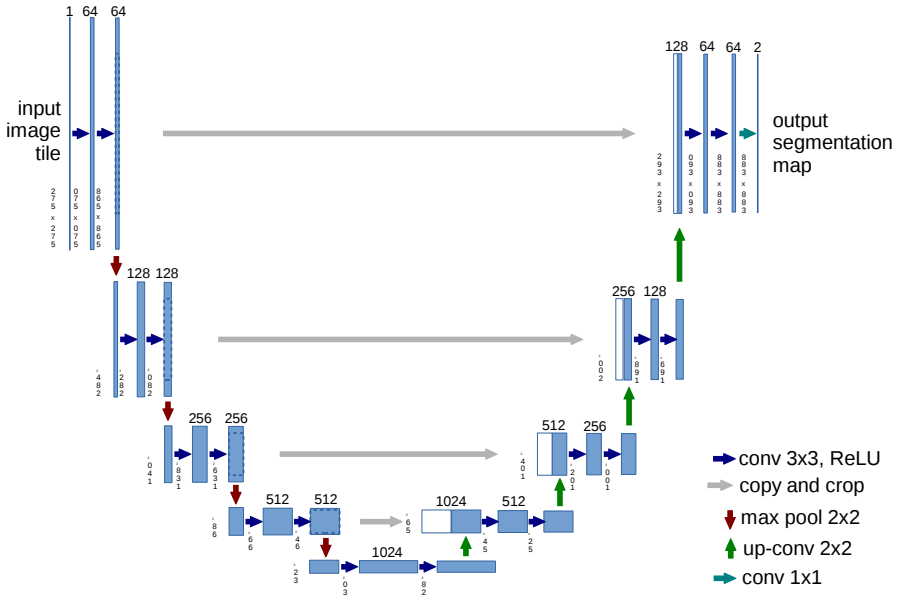

Fig. 1. U-net架构（以最低分辨率32x32像素为例）。每个蓝色框对应一个多通道特征图，通道数标注于框体上方，x-y尺寸标注于框体左下角。白色框表示复制的特征图，箭头指示不同的运算操作。

作为输入。首先，该网络能够进行定位。其次，以图像块形式存在的训练数据量远大于训练图像的数量。该网络以显著优势赢得了2012年ISBI的EM分割挑战赛。

显然，Ciresan等人[1]的策略存在两个缺点。首先，它相当缓慢，因为网络必须为每个图像块单独运行，且由于图像块重叠导致大量冗余。其次，定位精度与上下文利用之间存在权衡。较大的图像块需要更多的最大池化层，这会降低定位精度，而较小的图像块则让网络只能看到很少的上下文。更近期的研究[11, 4]提出了一种分类器输出方法，该方法综合考虑了来自多个层的特征。这样，良好的定位能力和上下文利用可以同时实现。

本文基于一种更为优雅的架构——即所谓的“全卷积网络”[9]。我们对此架构进行修改与扩展，使其能够在极少量训练图像下运行，并生成更精确的分割结果（见图1）。文献[9]的核心思想是在常规收缩网络基础上叠加连续层级，其中池化算子被上采样算子替代。这些层级因此提升了输出分辨率。为实现精确定位，收缩路径中的高分辨率特征将与上采样后的{v*}特征相结合。

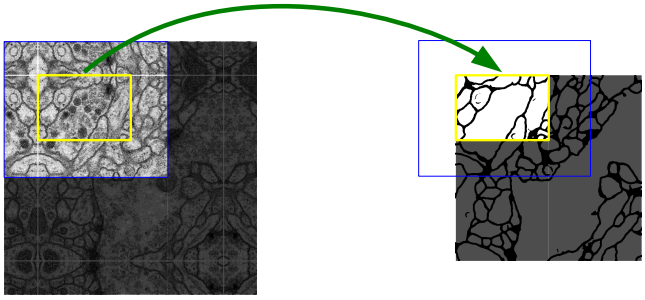

Fig. 2. Overlap-tile strategy for seamless segmentation of arbitrary large images (here segmentation of neuronal structures in EM stacks). Prediction of the segmentation in the yellow area, requires image data within the blue area as input. Missing input data is extrapolated by mirroring

output. A successive convolution layer can then learn to assemble a more precise output based on this information.

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.

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.

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.

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 out-

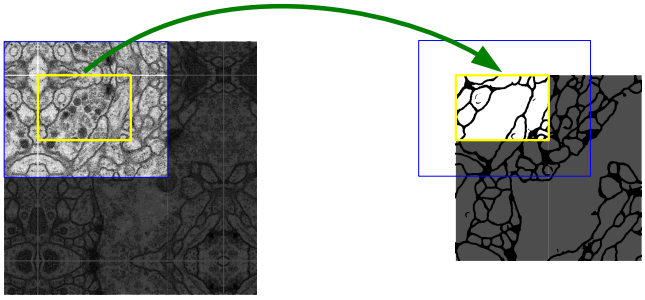

Fig. 2. 用于任意大图像无缝分割的重叠平铺策略（此处为电子显微镜堆栈中神经元结构的分割）。黄色区域分割的预测需要蓝色区域内的图像数据作为输入。缺失的输入数据通过镜像进行外推。

输出。随后的卷积层可以基于这些信息学习组装出更精确的输出。

我们架构中的一个重要修改是在上采样部分也拥有大量的特征通道，这使得网络能够将上下文信息传播到更高分辨率的层中。因此，扩展路径与收缩路径大致对称，从而形成了一个U形架构。该网络没有任何全连接层，并且只使用每个卷积的有效部分，即分割图仅包含那些在输入图像中具有完整上下文的像素。这种策略通过重叠平铺策略（见图2）实现了对任意大图像的无缝分割。为了预测图像边界区域的像素，缺失的上下文通过镜像输入图像进行外推。这种平铺策略对于将网络应用于大图像至关重要，否则分辨率将受限于GPU内存。

由于我们的任务可用的训练数据非常少，我们通过对现有训练图像施加弹性形变来进行过度的数据增强。这使得网络能够学习对此类形变的不变性，而无需在标注图像库中看到这些变换。这在生物医学分割中尤为重要，因为形变曾是组织中最常见的变化，且真实的形变可以被高效模拟。数据增强对于学习不变性的价值已在Dosovitskiy等人[2]的无监督特征学习研究中得到证实。

在许多细胞分割任务中，另一个挑战是分离同一类别中相互接触的物体；参见图3。为此，我们提出使用加权损失函数，其中接触细胞之间的分隔背景标签在损失函数中获得较大的权重。

所得网络适用于多种生物医学分割问题。本文展示了在电子显微镜堆栈中神经元结构分割的结果（这是自2012年ISBI会议启动的一项持续竞赛），我们在该任务中取得了超越——

performed 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.

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.

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.

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.

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 $p_k(\mathbf{x}) = \exp(a_k(\mathbf{x}))/\left(\sum_{k'=1}^{K}\exp(a_{k'}(\mathbf{x}))\right)$ 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 $p_{\ell(\mathbf{x})}(\mathbf{x})$ from 1 using

$$E = \sum_{\mathbf{x}\in\Omega} w(\mathbf{x})\log(p_{\ell(\mathbf{x})}(\mathbf{x})) \tag{1}$$

我们实施了Ciresan等人[1]的网络。此外，我们展示了在ISBI 2015细胞追踪挑战赛中，对光学显微镜图像进行细胞分割的结果。在此，我们在两个最具挑战性的二维透射光数据集上以显著优势获胜。

2 Network Architecture

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，包含重复应用两次的3x3卷积（无填充卷积），每次卷积后接一个线性整流单元（ReLU）以及步长为2的2x2最大池化操作进行下采样。在每个下采样步骤中，我们将特征通道数量加倍。扩展路径中的每个步骤包含特征图的上采样，随后进行2x2卷积（“上卷积”）将特征通道数减半，再与收缩路径中对应裁剪的特征图进行拼接，最后经过两个3x3卷积（每个卷积后接一个ReLU）。由于每次卷积会丢失边界像素，裁剪操作是必要的。在最后一层使用1x1卷积将64维特征向量映射到目标类别数。该网络总共包含23个卷积层。

为了实现输出分割图的无缝平铺（见图2），选择输入图块大小时需确保所有2x2最大池化操作都在x和y方向尺寸为偶数的层上执行。

3 Training

输入图像及其对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。由于未使用填充卷积，输出图像比输入图像小一个恒定的边框宽度。为了最小化开销并最大限度地利用GPU内存，我们倾向于使用较大的输入图块而非较大的批次大小，因此将批次减少为单张图像。相应地，我们采用较高的动量（0.99），使得大量先前见过的训练样本决定当前优化步骤中的更新。

能量函数通过对最终特征图进行逐像素的soft-max并结合交叉熵损失函数计算得到。soft-max定义为 $p_k(\mathbf{x}) = \exp(a_k(\mathbf{x}))/\left(\sum_{k'=1}^{K}\exp(a_{k'}(\mathbf{x}))\right)$，其中 $a_k(\mathbf{x})$ 表示在像素位置 $\mathbf{x} \in \Omega$ 处特征通道 k 的激活值（满足 $\Omega \subset \mathbb{Z}^2$）。$K$ 是类别总数，$p_k(\mathbf{x})$ 是近似的最大值函数。即当 k 对应的激活值 $a_k(\mathbf{x})$ 最大时 $p_k(\mathbf{x}) \approx$ 为1，其余所有 k 对应的 $p_k(\mathbf{x}) \approx$ 为0。随后交叉熵函数在每个位置通过下式对 $p_{\ell(\mathbf{x})}(\mathbf{x})$ 偏离1的情况进行惩罚：

$$E = \sum_{\mathbf{x}\in\Omega} w(\mathbf{x})\log(p_{\ell(\mathbf{x})}(\mathbf{x})) \tag{1}$$

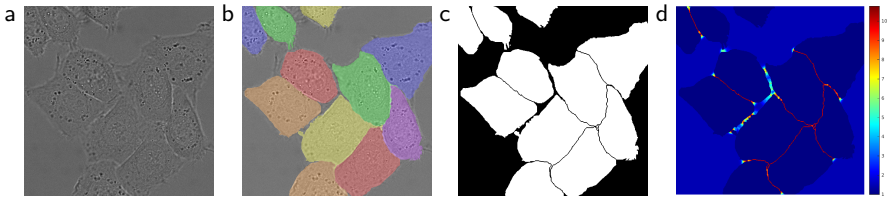

Fig. 3. HeLa cells on glass recorded with DIC (differential interference contrast) microscopy. (**a**) raw image. (**b**) overlay with ground truth segmentation. Different colors indicate different instances of the HeLa cells. (**c**) generated segmentation mask (white: foreground, black: background).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

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

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).

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

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

where $w_c : \Omega \to \mathbb{R}$ is the weight map to balance the class frequencies, $d_1 : \Omega \to \mathbb{R}$ denotes the distance to the border of the nearest cell and $d_2 : \Omega \to \mathbb{R}$ the distance to the border of the second nearest cell. In our experiments we set $w_0 = 10$ and $\sigma \approx 5$ pixels.

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 $N = 9 \cdot 64 = 576$.

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

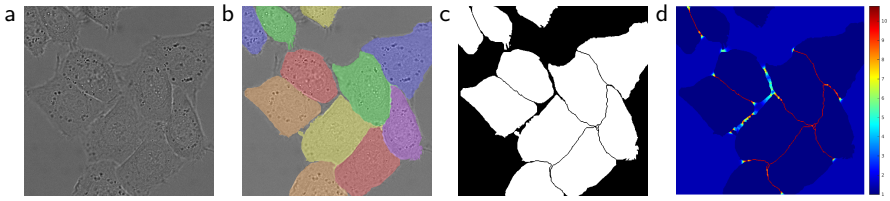

Fig. 3. 使用DIC（微分干涉对比）显微镜记录的玻璃上的HeLa细胞。**(a)** 原始图像。**(b)** 与真实分割叠加的效果图。不同颜色表示HeLa细胞的不同实例。**(c)** 生成的分割掩码（白色：前景，黑色：背景）。**(d)** 带有逐像素损失权重的映射图，用于强制网络学习边界像素。

其中 $\ell : \Omega \rightarrow \{1, \ldots, K\}$ 是每个像素的真实标签，而 $w : \Omega \rightarrow \mathrm{R}$ 是我们引入的权重图，用于在训练中赋予某些像素更高的重要性。

我们预先计算每个真实分割的权重图，以补偿训练数据集中特定类别像素的不同频率，并迫使网络学习我们在接触细胞之间引入的细小分隔边界（参见图3 c和d）。

分离边界是通过形态学操作计算得出的。随后，权重图的计算公式为

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

其中 $w_c : \Omega \rightarrow \mathrm{R}$ 是用于平衡类别频率的权重图，$d_1 : \Omega \rightarrow \mathrm{R}$ 表示到最近细胞边界的距离，而 $d_2 : \Omega \rightarrow \mathrm{R}$ 表示到第二近细胞边界的距离。在我们的实验中，我们设定 $w_0 =$ 为10像素，$\sigma \approx$ 为5像素。

在具有许多卷积层和不同网络路径的深度网络中，权重的良好初始化至关重要。否则，网络某些部分可能产生过度激活，而其他部分则始终无法发挥作用。理想情况下，初始权重的设定应使网络中每个特征图具有近似单位方差。对于采用我们这种架构（卷积层与ReLU层交替）的网络，可通过从标准差为$\sqrt{2/N}$的高斯分布中抽取初始权重来实现，其中N表示单个神经元的输入节点数[5]。例如对于3x3卷积且前一层具有64个特征通道的情况，N =9·64=576。

3.1 Data Augmentation

数据增强对于教会网络所需的**不变性**和**鲁棒性**至关重要，尤其是在训练样本数量有限的情况下。当

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.

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].

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 rand-error of 0.0382.

This is significantly better than the sliding-window convolutional network result by Ciresan et al. [1], 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

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

Rank	Group name	Warping Error	Rand Error	Pixel Error
	** human values **	0.000005	0.0021	0.0010
1.	u-net	**0.000353**	0.0382	0.0611
2.	DIVE-SCI	0.000355	0.0305	0.0584
3.	IDSIA [1]	0.000420	0.0504	0.0613
4.	DIVE	0.000430	0.0545	**0.0582**
⋮				
10.	IDSIA-SCI	0.000653	**0.0189**	0.1027

在显微图像处理中，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。特别是对训练样本施加随机弹性形变，似乎是利用极少标注图像训练分割网络的关键方法。我们通过在粗糙的3×3网格上生成随机位移向量来创建平滑形变，位移量从标准差为10像素的高斯分布中采样，随后通过双三次插值计算每个像素的位移。收缩路径末端的Drop-out层则进一步执行隐式的数据增强。

4 Experiments

我们展示了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镜记录中分割神经元结构。数据集的一个示例及我们获得的分割结果如图2所示。完整结果作为补充材料提供。该数据集由EM分割挑战赛[14]提供，该挑战始于ISBI 2012，目前仍接受新的贡献。训练数据包含30张图像（512x512像素），这些图像来自果蝇一龄幼虫腹侧神经索（VNC）的连续切片透射电子显微镜。每张图像都配有对应的完全标注的真实分割图，标注了细胞（白色）和膜（黑色）。测试集公开可用，但其分割图保密。通过将预测的膜概率图发送给组织者，可以获得评估结果。评估通过对概率图在10个不同阈值水平进行二值化，并计算“形变误差”、“兰德误差”和“像素误差”[14]来完成。

U-net（对输入数据的7个旋转版本取平均）在无需任何额外预处理或后处理的情况下，实现了0.0003529的形变误差（当前最佳成绩，见表1）以及0.0382的兰德误差。

这显著优于Ciresan等人[1]的滑动窗口卷积网络结果，其最佳提交的形变误差为0.000420，随机误差为0.0504。就随机误差而言，唯一表现更优的

Table 1. 在EM分割挑战赛[14]（2015年3月6日）上的排名，按扭曲误差排序。

Rank	Group name	Warping Error	Rand Error	Pixel Error
	** human values **	0.000005	0.0021	0.0010
1.	u-net	**0.000353**	0.0382	0.0611
2.	DIVE-SCI	0.000355	0.0305	0.0584
3.	IDSIA [1]	0.000420	0.0504	0.0613
4.	DIVE	0.000430	0.0545	**0.0582**
⋮				
10.	IDSIA-SCI	0.000653	**0.0189**	0.1027

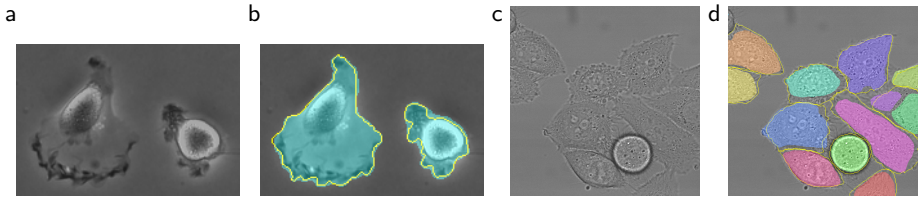

Fig. 4. Result on the ISBI cell tracking challenge. (**a**) part of an input image of the "PhC-U373" data set. (**b**) Segmentation result (cyan mask) with manual ground truth (yellow border) (**c**) input image of the "DIC-HeLa" data set. (**d**) Segmentation result (random colored masks) with manual ground truth (yellow border).

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

Name	PhC-U373	DIC-HeLa
IMCB-SG (2014)	0.2669	0.2935
KTH-SE (2014)	0.7953	0.4607
HOUS-US (2014)	0.5323	-
second-best 2015	0.83	0.46
u-net (2015)	**0.9203**	**0.7756**

algorithms on this data set use highly data set specific post-processing methods[1] applied to the probability map of Ciresan et al. [1].

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"[2] 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"[3] 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%.

5 Conclusion

The u-net architecture achieves very good performance on very different biomedical segmentation applications. Thanks to data augmentation with elastic defor-

[1] The authors of this algorithm have submitted 78 different solutions to achieve this result.

[2] Data set provided by Dr. Sanjay Kumar. Department of Bioengineering University of California at Berkeley. Berkeley CA (USA)

[3] Data set provided by Dr. Gert van Cappellen Erasmus Medical Center. Rotterdam. The Netherlands

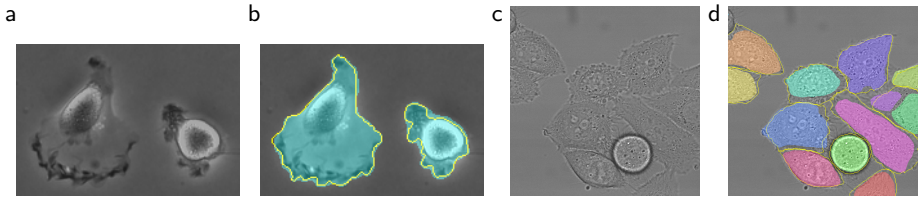

Fig. 4. ISBI细胞追踪挑战赛的结果。**(a)** “PhC-U373”数据集的部分输入图像。**(b)** 分割结果（青色掩膜）与人工标注真值（黄色边框）的对比。**(c)** “DIC-HeLa”数据集的输入图像。**(d)** 分割结果（随机彩色掩膜）与人工标注真值（黄色边框）的对比。

Table 2. ISBI 2015细胞追踪挑战赛上的分割结果（IOU）。

Name	PhC-U373	DIC-HeLa
IMCB-SG (2014)	0.2669	0.2935
KTH-SE (2014)	0.7953	0.4607
HOUS-US (2014)	0.5323	-
second-best 2015	0.83	0.46
u-net (2015)	**0.9203**	**0.7756**

在该数据集上使用的算法采用了高度针对数据集的特定后处理方法[1]，这些方法应用于Ciresan等人[1]的概率图。

我们还将U-Net应用于光学显微镜图像中的细胞分割任务。该分割任务是ISBI 2014和2015细胞追踪挑战赛的一部分[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胺基底上的胶质母细胞瘤-星形细胞瘤U373细胞（见图4a、b及补充材料）。该数据集包含35张部分标注的训练图像。在此我们实现了92%的平均IOU（交并比），显著优于第二佳算法的83%（见表2）。第二个数据集“DIC-HeLa”[3]是通过微分干涉对比显微镜记录的平板玻璃上的HeLa细胞（见图3、图4c、d及补充材料）。该数据集包含20张部分标注的训练图像。在此我们实现了77.5%的平均IOU，显著优于第二佳算法的46%。

5 Conclusion

U-net架构在多种生物医学分割应用中均表现出色。这得益于弹性形变数据增强技术的应用。

1 该算法的作者提交了78种不同的解决方案以实现这一结果。[2] 数据集由加州大学伯克利分校生物工程系的Sanjay Kumar博士提供。伯克利，加利福尼亚州（美国）[3] 数据集由鹿特丹伊拉斯姆斯医学中心的Gert van Cappellen博士提供。荷兰

mations, 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[4]. We are sure that the u-net architecture can be applied easily to many more tasks.

Acknowlegements

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

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., Arbelez, 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/

[4] U-net implementation, trained networks and supplementary material available at http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net

在标注图像极少的情况下，它仅需在NVidia Titan GPU（6 GB显存）上进行10小时的合理训练时长。我们提供了完整的基于Caffe[6]的实现及训练好的网络[4]。我们确信u-net架构能够轻松应用于更多任务中。

Acknowlegements

本研究得到了德国联邦和州政府卓越计划（EXC 294）以及联邦教育与研究部（项目编号 0316185B）的支持。

References

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。见：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式无监督特征学习。见：NIPS (2014)3. Girshick, R., Donahue, J., Darrell, T., Malik, J.: 用于精确目标检测和语义分割的丰富特征层次结构。见：IEEE计算机视觉与模式识别会议 (CVPR) 论文集 (2014)4. Hariharan, B., Arbelez, P., Girshick, R., Malik, J.: 用于目标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性能 (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), 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。见：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.: 细胞追踪算法比较的基准测试。Bioinformatics 30, 1609–1617 (2014)
11. Seyedhosseini, M., Sajjadi, M., Tasdizen, T.: 基于级联分层模型和逻辑析取正态网络的图像分割。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极深卷积网络 (2014), arXiv:1409.1556 [cs.CV]
13. WWW：细胞追踪挑战的网页，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 implementation, trained networks and supplementary material available at http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net