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
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]. 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. [2] 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

N. Navab et al. (Eds.): MICCAI 2015, Part III, LNCS 9351, pp. 234–241, 2015.
DOI: 10.1007/978-3-319-24574-4_28

U-Net：用于生物医学图像分割的卷积网络

Olaf Ronneberger、Philipp Fischer 和 Thomas Brox

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

摘要人们普遍认为，深度网络的训练需要成千上万标注的训练样本。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，以更有效地利用可用的标注样本。该架构包括一个用于捕获上下文的收缩路径和一个对称的扩展路径，该扩展路径能够实现精确的定位。我们证明，这样的网络可以从很少的图像中进行端到端训练，并在 ISBI 挑战赛（用于电子显微镜堆栈中神经元结构的分割）上优于之前最好的方法（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 2015 年 ISBI 细胞追踪挑战赛中的这些类别中取得了巨大的优势。此外，该网络速度很快。在最新的 GPU 上，对 512x512 图像的分割不到一秒。完整的实现（基于 Caffe）和训练好的网络可在 http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net。

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超越了当前技术水平，例如 [7]。虽然卷积网络已经存在很长时间 [8]，但由于可用的训练集大小和所考虑的网络大小，它们的成功有限。Krizhevsky 等人的突破 [7] 是由于在 ImageNet 数据集上对具有 8 层和数百万参数的大网络进行监督训练，该数据集有 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，其中图像的输出是一个单个类标签。然而，在许多视觉任务中，特别是在生物医学图像处理中，期望的输出应包括定位，即每个像素都应该分配一个类标签。此外，在生物医学任务中，通常数千张训练图像是无法企及的。因此，Ciresan 等人 [2] 在滑动窗口设置中训练了一个网络，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

N. Navab et al. (Eds.): MICCAI 2015, Part III, LNCS 9351, pp. 234–241, 2015.
DOI: 10.1007/978-3-319-24574-4_28

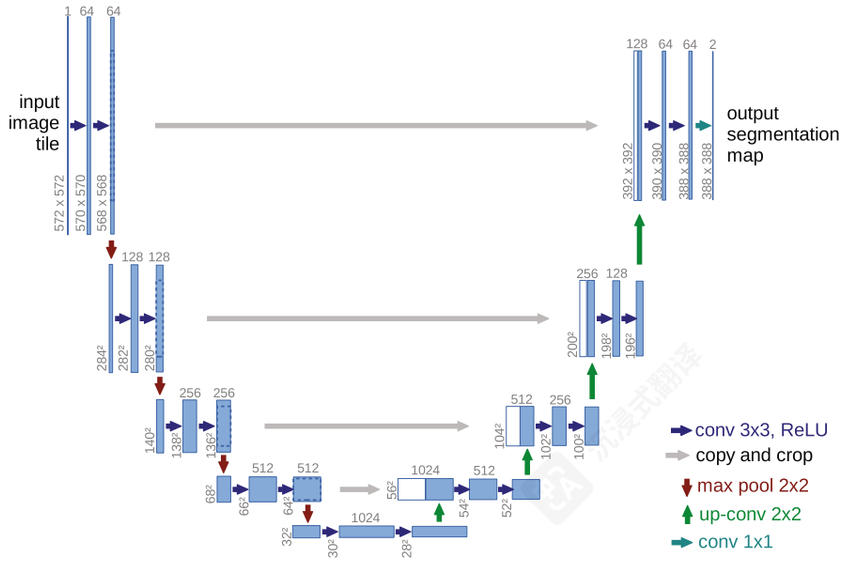


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows 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. [2] 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 output. A successive convolution layer can then learn to assemble a more precise output based on this information.

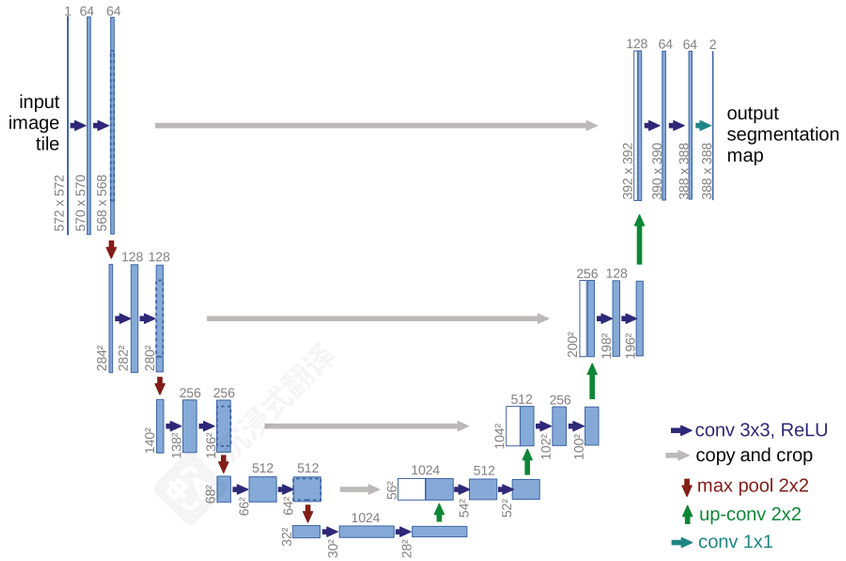


图1。 U-net架构（以32x32像素的最低分辨率为例）。每个蓝色框对应一个多通道特征图。通道数标在框的顶部。框的左下角提供了x-y尺寸。白色框表示复制的特征图。箭头表示不同的操作。

作为输入。首先，该网络可以进行定位。其次，以块为单位的训练数据远大于训练图像的数量。该网络在2012年ISBI的EM分割挑战中大幅获胜。

显然，Ciresan等人提出的策略有两个缺点。 [2] 首先，它相当慢，因为网络必须为每个块单独运行，并且由于块的重叠而存在大量冗余。其次，定位精度和上下文使用之间存在权衡。较大的块需要更多的最大池化层，这会降低定位精度，而小块允许网络只能看到很少的上下文。更近期的 [11,4] 方法提出了一种考虑多层特征的分类器输出。同时可以实现良好的定位和上下文使用。

在本文中，我们基于一种更优雅的架构，即所谓的"全卷积网络" [9]。我们修改并扩展这种架构，使其能够在很少的训练图像上工作并产生更精确的分割；参见图1。其核心思想在于 [9] 通过添加连续的层来补充一个常规的收缩网络，其中池化操作符被上采样操作符取代。因此，这些层增加了输出的分辨率。为了定位，收缩路径中的高分辨率特征与上采样输出相结合。然后，一个连续的卷积层可以根据这些信息学习如何组合出更精确的输出。

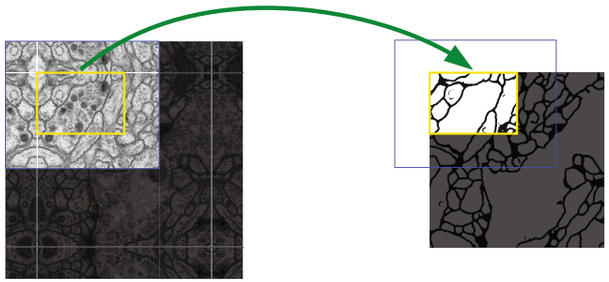


Fig. 2. Overlap-tile strategy for seamless segmentation of arbitrary large images (here segmentation of neuronal structures in EM stacks). Prediction of the segmentation in the yellow area, requires image data within the blue area as input. Missing input data is extrapolated by mirroring

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. [3] 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 outperformed the network of Ciresan et al. [2]. 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.

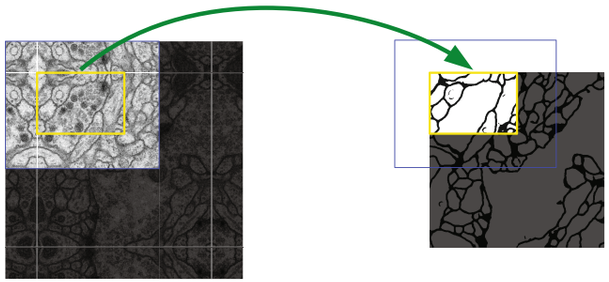


Fig.2. 重叠瓦片策略用于任意大图像的无缝分割（此处为EM堆栈中神经元结构的分割）。黄色区域的分割预测需要蓝色区域内的图像数据作为输入。缺失的输入数据通过镜像进行外推

我们架构的一个重要修改是在上采样部分我们也有大量的特征通道，这允许网络将上下文信息传播到更高分辨率的层。因此，扩张路径大致对称于收缩路径，并产生一个u形架构。网络没有全连接层，并且仅使用每个卷积的有效部分，即分割图仅包含在输入图像中具有完整上下文的像素。这种策略通过重叠瓦片策略（参见 图2）允许无缝分割任意大图像。为了预测图像边界的像素，通过镜像输入图像来外推缺失的上下文。这种瓦片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GPU内存的限制。

关于我们的任务，可用的训练数据很少，我们通过对可用的训练图像应用弹性变形来进行过度数据增强。这允许网络学习对这种变形的不变性，而无需在标注的图像语料库中看到这些变换。这在生物医学分割中尤其重要，因为变形曾经是组织和真实变形可以高效模拟的最常见变化。数据增强在学习不变性方面的价值已在 Dosovitskiy 等人 [3] 的无监督特征学习范围内得到证明。

在许多细胞分割任务中，另一个挑战是分离同一类接触的对象；参见 图 3。为此，我们提出使用加权损失，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。

生成的网络适用于各种生物医学分割问题。在本文中，我们在 EM 堆栈（一个于 2012 年 ISBI 开始进行的持续竞赛）中的神经元结构分割上展示了结果，在那里我们超过了 Ciresan 等人 [2] 的网络。此外，我们还展示了 ISBI 细胞跟踪挑战 2015 年从光显微镜图像中进行的细胞分割结果。在这里，我们在两个最具挑战性的 2D 透射光数据集上以较大优势获胜。

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

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

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

2 网络架构

网络架构如图1所示。它由一个收缩路径（左侧）和一个扩张路径（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（无填充卷积）的重复应用组成，每个卷积后接一个修正线性单元（ReLU）和一个步长为2的2x2最大池化操作用于下采样。在每次下采样步骤中，我们加倍特征通道的数量。扩张路径中的每一步都由特征图的上采样后接一个2x2卷积（“上卷积”）组成，该卷积将特征通道数量减半，然后与收缩路径中相应裁剪的特征图进行拼接，并接两个3x3卷积，每个卷积后接一个ReLU。裁剪是必要的，因为每次卷积都会丢失边界像素。在最后一层，使用一个1x1卷积将每个64组件特征向量映射到所需类别的数量。总共有23个卷积层。

为了使输出分割图（参见 图2）能够无缝平铺，选择输入瓦片大小以使所有2x2最大池化操作应用于具有偶数x和y大小的层非常重要。

3 训练

输入图像及其对应的分割图用于使用Caffe [6]的随机梯度下降实现来训练网络。由于没有填充的卷积，输出图像比输入图像小一个常数的边界宽度。为了最小化开销并最大限度地利用GPU内存，我们倾向于使用较大的输入瓦片而不是较大的批处理大小，因此将批处理减少为单个图像。相应地，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。

能量函数通过在最终特征图上进行逐像素softmax并结合交叉熵损失函数来计算。softmax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 $a_k(\mathbf{x})$ 表示特征通道 k 在像素位置 $\mathbf{x} \in \Omega$ 处的激活值，并且 $\Omega \subset \mathbb{Z}^2$。$K$ 是类别的数量，$p_k(\mathbf{x})$ 是近似最大函数。也就是说，$p_k(\mathbf{x}) \approx 1$ 对于具有最大激活 $a_k(\mathbf{x})$ 的 k，以及 $p_k(\mathbf{x}) \approx 0$ 对于所有其他 k。然后交叉熵在每个位置处惩罚 $p_{\ell(\mathbf{x})}(\mathbf{x})$ 与1的偏差，使用

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

where $\ell : \Omega \to \{1,\dots,K\}$ 是每个像素的真实标签，而 w: $\Omega \to$ R 是我们引入的权重图，用于在训练中给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，以补偿训练中来自某个类别的像素频率差异。

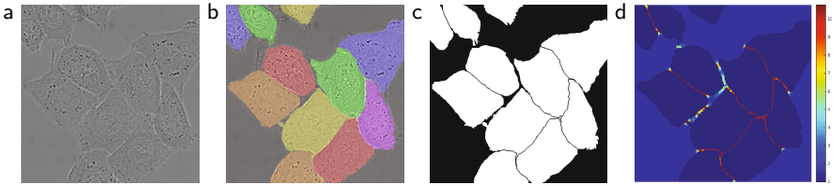


Fig. 3. HeLa cells on glass recorded with DIC (differential interference contrast) microscopy. (**a**) raw image. (**b**) overlay with ground truth segmentation. Different colors indicate different instances of the HeLa cells. (**c**) generated segmentation mask (white: foreground, black: background).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

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

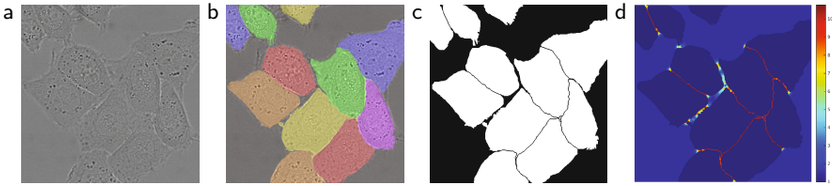
$$w(\mathbf{x}) = w_c(\mathbf{x}) + w_0 \cdot \exp\left(-\frac{(d_1(\mathbf{x}) + d_2(\mathbf{x}))^2}{2\sigma^2}\right) \tag{2}$$

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 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.



<样式 id='1'>图3。 </样式>在玻璃上记录的HeLa细胞，使用DIC（微分干涉对比）显微镜。(<样式 id='3'>a</样式>) 原始图像。(<样式 id='5'>b</样式>) 与真实分割叠加。不同颜色表示不同的HeLa细胞实例。(<样式 id='7'>c</样式>) 生成的分割掩码（白色：前景，黑色：背景）。(<样式 id='9'>d</样式>) 像素级损失权重图，以迫使网络学习边界像素。

数据集，并迫使网络学习我们引入的接触细胞之间的小间隔边界（参见<Figure 3>c和d）。

分离边界使用形态学操作计算。然后计算权重图。

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

其中 w_c：$\Omega \to \mathrm{R}$ 是用于平衡类别频率的权重图，d_1：$\Omega \to \mathrm{R}$ 表示到最近单元格边界的距离，以及 $d_2 : \Omega \to \mathbb{R}$ 到次近单元格边界的距离。在我们的实验中，我们设置了 $w_0 = 10$ 和 $\sigma \approx 5$ 像素。

在具有许多卷积层和不同网络路径的深度网络中，权重的良好初始化非常重要。否则，网络的一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，初始权重应该进行调整，以便网络中的每个特征图都具有近似于单位方差。对于具有我们架构（交替的卷积层和ReLU层）的网络，这可以通过从标准差为 $2/N$的高斯分布中抽取初始权重来实现，其中 N 表示一个神经元的输入节点数 [5]。例如，对于一个3x3卷积和前一层的64个特征通道 $N = 9 \cdot 64 = 576$。

3.1 数据增强

数据增强对于在只有少量训练样本可用时教会网络所需的不变性和鲁棒性属性至关重要。在显微图像的情况下，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分割网络的关键概念，只需很少的标注图像。我们使用随机位移向量在粗糙的3x3网格上生成平滑变形。

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 [2]	0.000420	0.0504	0.0613
4.	DIVE	0.000430	0.0545	**0.0582**
⋮				
10.	IDSIA-SCI	0.000653	**0.0189**	0.1027

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 [14,1] 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. [2], 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[1] applied to the probability map of Ciresan et al. [2].

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

表1。 在EM分割挑战中的排名 [14] (2015年3月6日)，按warping误差排序。

Rank	组名	Warping Error	Rand Error	Pixel Error
	** 人类价值观 **	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 [2]	0.000420	0.0504	0.0613
4.	DIVE	0.000430	0.0545	**0.0582**
⋮				
10.	IDSIA-SCI	0.000653	**0.0189**	0.1027

位移是从均值为10像素的标准差的高斯分布中采样的。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 个实验

我们展示了 u-net 在三个不同分割任务中的应用。第一个任务是分割电子显微镜记录中的神经元结构。数据集和我们的分割结果示例显示在 图 2。我们提供了完整的结果作为补充材料。该数据集由 EM 分割挑战 [14,1] 提供，该挑战始于 ISBI 2012 并仍然开放以接受新的贡献。训练数据是一组 30 张图像（512x512 像素），来自果蝇第一龄幼虫腹神经索（VNC）的连续切片透射电子显微镜图像。每张图像都附带一个相应的完整标注的地面真实验证分割图，用于细胞（白色）和膜（黑色）。测试集是公开的，但其分割图是保密的。可以通过向组织者发送预测的膜概率图来获得评估。评估是通过在 10 个不同级别上对图进行阈值处理并计算"变形误差"、"Rand 误差"和"像素误差"来完成的 [14]。

u-net（在输入数据的7个旋转版本上平均）无需任何进一步的前后处理即可实现0.0003529的变形误差（新的最佳分数，参见表1）和0.0382的rand误差。

这比Ciresan等人 [2], 的滑动窗口卷积网络结果要好得多，其最佳提交的变形误差为0.000420，rand误差为0.0504。在rand误差方面，这个数据集上表现更好的唯一算法是针对Ciresan等人 [2]的概率图应用了高度数据集特定的后处理方法[1]。

[1] 该算法的作者已提交了78个不同的解决方案以实现这一结果。

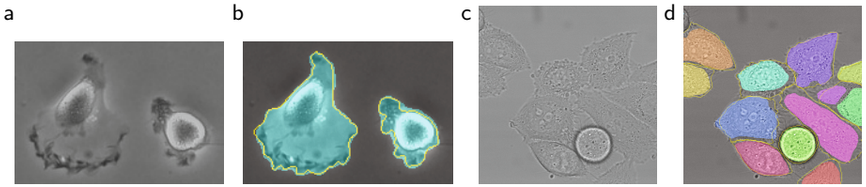


Fig. 4. Result on the ISBI cell tracking challenge. (**a**) part of an input image of the "PhC-U373" data set. (**b**) Segmentation result (cyan mask) with manual ground truth (yellow border) (**c**) input image of the "DIC-HeLa" data set. (**d**) Segmentation result (random colored masks) with manual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

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 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

[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.

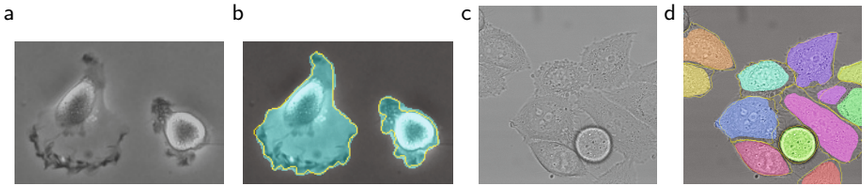


Fig.4. ISBI细胞追踪挑战赛的结果。(**a**)"PhC-U373"数据集的输入图像的一部分。(**b**)分割结果（青色掩码）与手动真实值（黄色边框）。(**c**)"DIC-HeLa"数据集的输入图像。(**d**)分割结果（随机颜色掩码）与手动真实值（黄色边框）。

表2. ISBI细胞追踪挑战赛2015年的分割结果（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	-
第二好 2015	0.83	0.46
u-net (2015)	**0.9203**	**0.7756**

我们也将在光显微图像中应用u-net进行细胞分割任务。这个分割任务是美国生物医学图像计算研究所（ISBI）细胞追踪挑战赛2014年的一部分和2015 [10,13]。第一个数据集"PhC-U373"[2]包含在聚丙烯酰胺基板上由相差显微镜记录的胶质母细胞瘤-星形细胞瘤U373细胞（见图 图4a,b和补充材料）。它包含35张部分标注的训练图像。在这里，我们实现了92%的平均IOU（"交并比"），这明显优于第二好的算法83%（见图表2）。第二个数据集"DIC-HeLa"[3]是平玻璃上由微分干涉对比（DIC）显微镜记录的HeLa细胞（见图 图3,图4c,d和补充材料）。它包含20张部分标注的训练图像。在这里，我们实现了77.5%的平均IOU，这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上实现了非常好的性能。由于使用弹性变形进行数据增强，它只需要非常少的标注图像，并且在 NVidia Titan GPU（6 GB）上只需要 10 小时的合理训练时间。我们提供

[2] 数据集由 Sanjay Kumar 博士提供。生物工程系，加州大学伯克利分校。伯克利，CA（美国）。[3] 数据集由 Gert vanCappellen 博士提供。埃拉斯姆斯医学中心。鹿特丹。荷兰。

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. Cardona, A., et al.: An integrated micro- and macroarchitectural analysis of the drosophila brain by computer-assisted serial section electron microscopy. PLoS Biol. 8(10), e1000502 (2010)
2. 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)
3. Dosovitskiy, A., Springenberg, J.T., Riedmiller, M., Brox, T.: Discriminative unsupervised feature learning with convolutional neural networks. In: NIPS (2014)
4. Hariharan, B., Arbeláez, 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., et al.: 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: 2013 IEEE International Conference on Computer Vision (ICCV), 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

完整 Caffe[6]-基于的实现和训练好的网络[4]。我们确信 u-net 架构可以轻松应用于许多更多任务。

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

参考文献

1. Cardona, A., et al.: An integrated micro- and macroarchitectural analysis of the drosophila brain by computer-assisted serial section electron microscopy. PLoS Biol. 8(10), e1000502 (2010)
2. 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)
3. Dosovitskiy, A., Springenberg, J.T., Riedmiller, M., Brox, T.: Discriminative unsupervised feature learning with convolutional neural networks. In: NIPS (2014)
4. Hariharan, B., Arbeláez, 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., et al.: 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: 2013 IEEE International Conference on Computer Vision (ICCV), 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 实现、训练好的网络和补充材料可在 http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net