

¹Institutetext: Computer Science Department and BIOS Centre for Biological Signalling Studies, University of Freiburg, Germany

計算機科學系與生物信號研究中心 · 弗賴堡大學 · 德國

¹email: ronneber@informatik.uni-freiburg.de

WWW home page: <http://lmb.informatik.uni-freiburg.de/>

WWW 主頁: <http://lmb.informatik.uni-freiburg.de/>

U-Net: Convolutional Networks for Biomedical Image Segmentation

U-Net: 生物醫學影像分割的卷積網絡

Olaf Ronneberger

Philipp Fischer

Thomas Brox

Abstract 摘要

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>.

在成功訓練深度網絡需要大量標註訓練樣本這一點上存在廣泛共識。本文提出了一種依賴於強大數據增強技術來更有效地使用可用標註樣本的網絡和訓練策略。該架構包括一個收縮路徑來捕捉上下文和一個對稱擴展路徑以實現精確定位。我們展示了這樣的網絡可以從極少量的圖像中進行端到端訓練，並且在 ISBI 挑戰賽中超越了之前最好的方法（一個滑動窗口卷積網絡）。在電子顯微鏡堆疊中的神經結構分割方面表現更佳。使用在透射光顯微鏡圖像（相位對比和差分干涉顯微鏡）上訓練的相同網絡，我們在 ISBI 2015 細胞追蹤挑戰賽中在這些類別中取得了巨大優勢。此外，該網絡運行速度很快。在近期的 GPU 上，512x512 圖像的分割時間不到一秒。完整的實現（基於 Caffe）和訓練好的網絡可以在 <http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net> 上獲得。

1 Introduction 1 引言

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].

在過去兩年中，深度卷積網絡在許多視覺識別任務中超越了最先進的技術，例如 [7, 3]。雖然卷積網絡已經存在了很長時間 [8]，但由於可用訓練集的大小和考慮中的網絡大小，其成功仍然有限。Krizhevsky 等人 [7] 的突破來自於在擁有 100 萬張訓練圖像的 ImageNet 數據集上，對一個具有 8 層和數百萬個參數的大型網絡進行有監督的訓練。自那時以來，更大更深的網絡已經被訓練 [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.

卷積網絡的典型使用是進行分類任務，其中對於圖像的輸出是一個單一的類別標籤。然而，在許多視覺任務中，特別是在生物醫學圖像處理中，期望的輸出應包括定位，即每個像素應分配一個類別標籤。此外，在生物醫學任務中，數千張訓練圖像通常無法獲得。因此，Ciresan 等人[1] 在滑動窗口設置中訓練了一個網絡，以通過提供圍繞該像素的局部區域（區塊）作為輸入來預測每個像素的類別標籤。首先，這個網絡可以進行定位。其次，基於區塊的訓練數據比訓練圖像的數量大多得多。這個網絡在 ISBI 2012 的 EM 分割挑戰中以巨大優勢獲勝。

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.

顯然，Ciresan 等人 [1] 的策略有兩個缺點。首先，由於網絡必須為每個區塊單獨運行，這樣會非常緩慢，並且由於重疊的區塊，會有大量冗餘。其次，定位準確度與上下文使用之間存在權衡。較大的區塊需要更多的最大池化層，這會降低定位準確度，而較小的區塊則使網絡只能看到很少的上下文。最近的方法 [11, 4] 提出了考慮來自多層的特徵的分類器輸出。這樣可以在同時實現良好的定位和上下文使用。

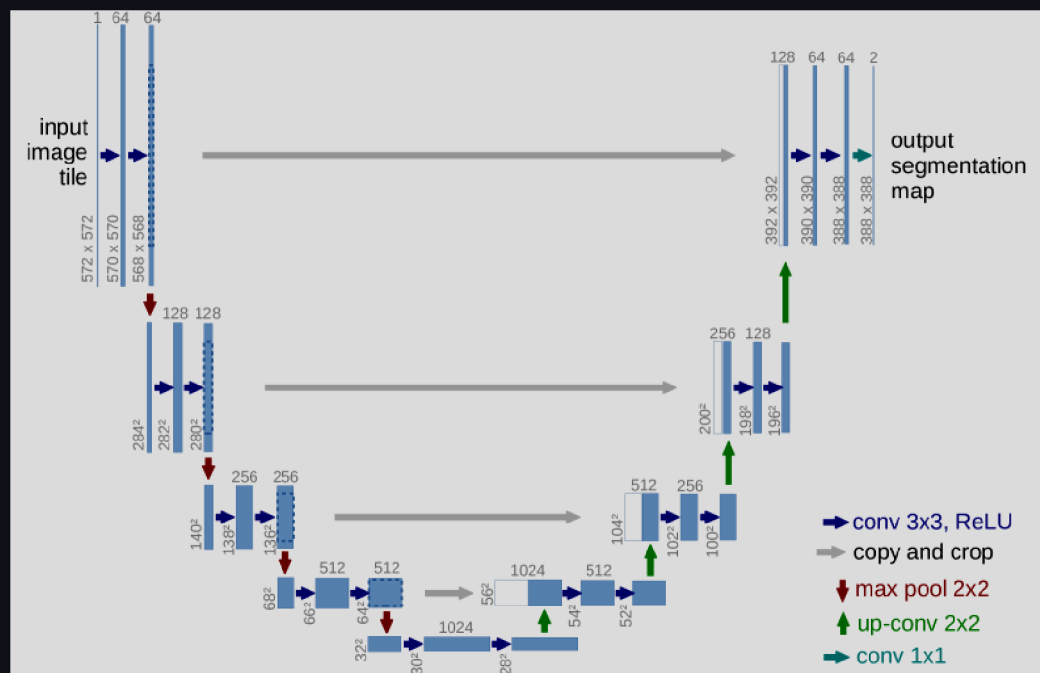


Figure 1: U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows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

圖 1：U-net 架構（最低解析度下的 32x32 像素範例）。每個藍色方框對應到一個多通道特徵圖。通道數量標示在方框上方。x-y 尺寸提供在方框的左下邊緣。白色方框表示複製的特徵圖。箭頭表示不同的操作。

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.

在本文中，我們建立在一個更優雅的架構上，即所謂的「全卷積網絡」[9]。我們修改並擴展了這個架構，使其能夠在很少的訓練圖像下工作並產生更精確的分割；見圖 1。[9]中的主要思想是通過連續的層來補充通常的收縮網絡，其中池化操作符被上採樣操作符取代。因此，這些層提高了輸出的解析度。為了進行定位，來自收縮路徑的高解析度特徵與上採樣的輸出相結合。然後，連續的卷積層可以基於這些信息學習組裝更精確的輸出。

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.

我們架構中的一個重要修改是，在上採樣部分我們也擁有大量的特徵通道，這允許網絡將上下文信息傳播到更高解析度的層。因此，擴展路徑或多或少是對稱於收縮路徑，形成一個 U 形架構。網絡沒有任何完全連接的層，僅使用每個卷積的有效部分，即分割圖僅包含輸入圖像中具有完整上下文的像素。這種策略允許通過重疊拼接策略無縫地分割任意大的圖像（見圖 2）。為了預測圖像邊界區域的像素，通過鏡像輸入圖像來推斷缺失的上下文。這種拼接策略對於將網絡應用於大圖像非常重要，因為否則解析度將受到 GPU 記憶體的限制。

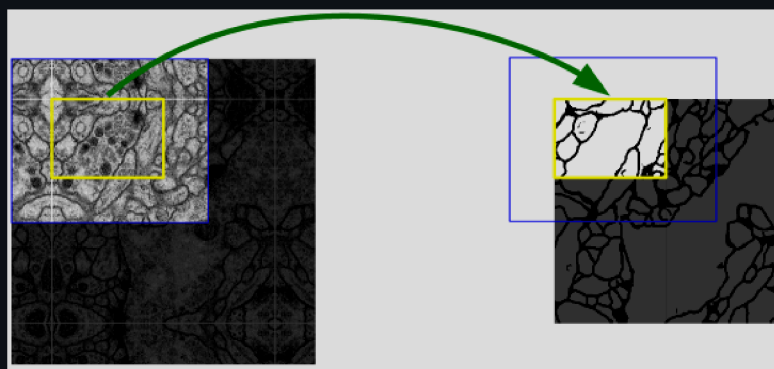


Figure 2: Overlap-tile strategy for seamless segmentation of arbitrary large images (here segmentation of neuronal structures in EM stacks). Prediction of the segmentation in the yellow area, requires image data within the blue area as input. Missing input data is extrapolated by mirroring

圖 2: 用於無縫分割任意大型圖像的重疊瓦片策略（此處為 EM 堆疊中的神經結構分割）。在黃色區域進行分割預測，需要藍色區域內的圖像數據作為輸入。缺失的輸入數據通過鏡像進行外推。

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.

由於我們的任務中可用的訓練數據非常有限，我們通過對可用訓練圖像應用彈性變形來進行過度數據增強。這使得網絡可以學習對這些變形的變不變性，而無需註釋圖像集上看到這些變換。這在生物醫學分割中特別重要，因為變形曾經是組織中最常見的變化，並且可以有效地模擬現實變形。Dosovitskiy 等人 [2] 在無監督特徵學習的範疇中已經顯示了數據增強對學習不變性的價值。

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.

許多細胞分割任務中的另一個挑戰是分離同一類別的接觸物體；見圖 3。為此，我們建議使用加權損失，其中接觸細胞之間的分隔背景標籤在損失函數中獲得較大權重。

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.

所得到的網絡適用於各種生物醫學分割問題。在本文中，我們展示了在 EM 堆疊中神經結構分割的結果（這是一個自 ISBI 2012 開始的持續競賽），我們超越了 Ciresan 等人 [1] 的網絡。此外，我們展示了來自 ISBI 細胞追蹤挑戰 2015 的光學顯微鏡圖像中的細胞分割結果。在這裡，我們在兩個最具挑戰性的 2D 傳遞光數據集上贏得了巨大的優勢。

2 Network Architecture 2 網絡架構

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 recti-

fied 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.

網絡架構如圖 1 所示。它由一個收縮路徑（左側）和一個擴展路徑（右側）組成。收縮路徑遵循卷積網絡的典型架構。它包括兩個 3x3 卷積（無填充卷積）的重複應用，每個卷積後跟隨一個修正線性單元（ReLU）和一個 2x2 的最大池化操作，步幅為 2，用於下採樣。在每次下採樣步驟中，我們將特徵通道數量加倍。擴展路徑中的每一步包括對特徵圖的上採樣，接著是 2x2 卷積（“上卷積”），該卷積將特徵通道數量減半，然後與來自收縮路徑的相應裁剪特徵圖進行連接，並進行兩次 3x3 卷積，每次卷積後跟隨一個 ReLU。裁剪是由於每次卷積中邊界像素的丟失。在最後一層，使用 1x1 卷積將每個 64 維特徵向量映射到所需的類別數量。總體而言，該網絡具有 23 層卷積層。

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.

為了實現輸出分割地圖的無縫鋪排（見圖 2），選擇輸入圖塊的大小是非常重要的，以確保所有 2x2 的最大池化操作都應用於具有偶數 x 和 y 尺寸的層。

3 Training 3 訓練

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.

輸入圖像及其對應的分割地圖用於訓練網絡，使用 Caffe [6] 的隨機梯度下降實現。由於未填充的卷積，輸出圖像比輸入圖像小一個固定的邊界寬度。為了最小化開銷並最大限度地利用 GPU 內存，我們偏好使用大型輸入圖塊而非大批量大小，因此將批量減少為單張圖像。因此，我們使用高動量（0.99），以使大量先前見過的訓練樣本決定當前優化步驟中的更新。

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 $p_k(\mathbf{x}) = \exp(a_k(\mathbf{x})) / \left(\sum_{k=1}^K \exp(a_k(\mathbf{x})) \right)$ 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

能量函數是通過對最終特徵圖進行逐像素的 soft-max 計算，並結合交叉熵損失函數來計算的。soft-max 定義為 $p_k(\mathbf{x}) = \exp(a_k(\mathbf{x})) / \left(\sum_{k=1}^K \exp(a_k(\mathbf{x})) \right)$ ，其中 $a_k(\mathbf{x})$ 表示像素位置 $\mathbf{x} \in \Omega$ 中特徵通道 k 的激活值，具有 $\Omega \subset \mathbb{Z}^2$ 。 K 是類別數量， $p_k(\mathbf{x})$ 是近似的最大函數。即 $p_k(\mathbf{x}) \approx 1$ 代表激活值最大 k 的 $a_k(\mathbf{x})$ ，而 $p_k(\mathbf{x}) \approx 0$ 代表所有其他 k 。交叉熵則在每個位置對 $p_{\ell(\mathbf{x})}(\mathbf{x})$ 偏離 1 進行懲罰，使用

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

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

其中 $\ell : \Omega \rightarrow \{1, \dots, K\}$ 是每個像素的真實標籤， $w : \Omega \rightarrow \mathbb{R}$ 是我們引入的權重圖，用於在訓練中給予某些像素更高的重要性。

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).

我們預先計算每個真實標註分割的權重圖，以補償訓練數據集中來自某個類別的像素頻率不同，並迫使網絡學習我們在接觸細胞之間引入的小分隔邊界（見圖 3c 和 d）。

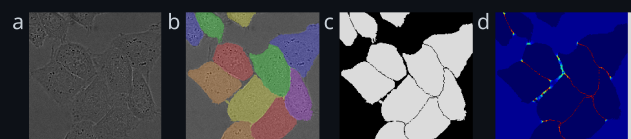


Figure 3: HeLa cells on glass recorded with DIC (differential interference contrast) microscopy. (a) raw image. (b) overlay with ground truth segmentation. Different colors indicate different instances of the HeLa cells. (c) generated segmentation mask (white: foreground, black: background). (d) map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

圖 3: 在玻璃上錄製的 HeLa 細胞，使用 DIC (差分干涉對比) 顯微鏡。(a) 原始圖像。(b) 與真實分割結果疊加。不同顏色表示 HeLa 細胞的不同實例。(c) 生成的分割掩碼 (白色: 前景, 黑色: 背景)。(d) 帶有逐像素損失權重的圖，以強制網絡學習邊界像素。

The separation border is computed using morphological operations. The weight map is then computed as
 分隔邊界是通過形態學操作來計算的。然後計算權重圖為

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

where $w_c: \Omega \rightarrow \mathbb{R}$ 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 $w_0 = 10$ 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 $2/N$, 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 $N = 9 \cdot 64 = 576$.

在深度網絡中，擁有許多卷積層和不同的網絡路徑，權重的良好初始化是極其重要的。否則，網絡的某些部分可能會產生過多的激活，而其他部分則永遠不會貢獻。理想情況下，初始權重應該進行調整，使得網絡中的每個特徵圖具有大約單位方差。對於我們的架構 (交替的卷積層和 ReLU 層) 的網絡，這可以通過從均值為 $2/N$ 的高斯分佈中抽取初始權重來實現，其中 N 表示一個神經元的輸入節點數量 [5]。例如，對於 3x3 卷積和前一層中的 64 個特徵通道 $N = 9 \cdot 64 = 576$ 。

3.1 Data Augmentation 3.1 數據擴增

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.

資料增強對於在訓練樣本有限的情況下教導網絡所需的不變性和魯棒性特徵至關重要。對於顯微鏡圖像，我們主要需要位移和旋轉不變性，以及對變形和灰度值變化的魯棒性。特別是，隨機彈性變形的訓練樣本似乎是用非常少量標註圖像訓練分割網絡的關鍵概念。我們通過在粗略的 3 by 3 網格上使用隨機位移向量來生成平滑的變形。這些位移是從標準差為 10 像素的高斯分佈中抽樣的。然後，使用雙三次插值計算每個像素的位移。在收縮路徑的末端，Drop-out 層進一步執行隱式數據增強。

4 Experiments 4 實驗

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].

我們展示了 u-net 在三個不同分割任務中的應用。第一個任務是對電子顯微鏡記錄中的神經結構進行分割。數據集和我們獲得的分割結果的示例顯示在圖 2 中。我們提供了完整的結果作為補充材料。數據集由 EM 分割挑戰 [14] 提供，該挑戰於 ISBI 2012 啟動，並仍然開放接受新的貢獻。訓練數據是一組來自果蠅第一齡幼蟲腹側神經索（VNC）串聯切片透射電子顯微鏡的 30 張圖像（512x512 像素）。每張圖像都有一個相應的完全標註的地面真實分割圖，標註了細胞（白色）和膜（黑色）。測試集公開可用，但其分割圖保密。可以通過將預測的膜概率圖發送給組織者來獲得評估。評估是通過在 10 個不同水平上對圖進行閾值處理，計算“變形誤差”、“Rand 誤差”和“像素誤差” [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.

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

u-net（對輸入數據的 7 個旋轉版本取平均）在未經任何進一步預處理或後處理的情況下，達到 0.0003529 的變形誤差（新最佳分數，見表 1）和 0.0382 的 Rand 誤差。

| Rank Table 1:Ranking on the EM segmentation challenge [14] (march 6th, 2015), sorted by warping error. Group name 表 1 : EM segmentation challenge [14] (march 6th, 2015), sorted by warping error. Group name | |
|--|--------------------|
| | ** human values ** |
| | ** 人類價值 ** |
| | 1. u-net |
| | 2. DIVE-SCI |
| | 3. IDSIA [1] |
| | 4. DIVE |
| | ⋮ |
| | 10. IDSIA-SCI |

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 algorithms on this data set use highly data set specific post-processing methods¹

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

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

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 algorithms on this data set use highly data set specific post-processing methods ¹ applied to the probability map of Ciresan et al. [1]. text (Traditional Chinese): 這比 Ciresan 等人 [1] 的滑動窗口卷積網絡結果要好得多，他們的最佳提交的變形錯誤為 0.000420，rand 錯誤為 0.0504。在 rand 錯誤方面，該數據集上唯一表現更好的算法使用了高度特定於數據集的後處理方法 ¹ 應用於 Ciresan 等人 [1] 的概率圖。

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

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

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.

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. text (Traditional Chinese): 我們還將 u-net 應用於光學顯微鏡圖像中的細胞分割任務。這項分割任務是 ISBI 細胞追蹤挑戰 2014 和 2015 [10, 13] 的一部分。第一個數據集 “PhC-U373” ² 包含在聚丙烯酰胺基板上的膠質母細胞瘤-星形膠質細胞 U373，通過相位差顯微鏡錄製（見圖 4a,b 和補充材料）。它包含 35 張部分標註的訓練圖像。

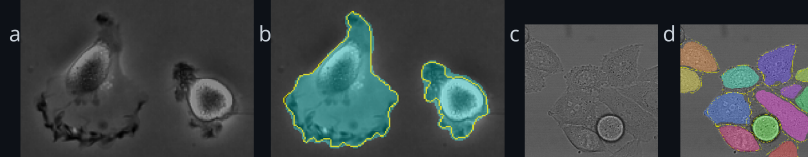


Figure 4: Result on the ISBI cell tracking challenge. (a) part of an input image of the “PhC-U373” data set. (b) Segmentation result (cyan mask) with manual ground truth (yellow border) (c) input image of the “DIC-HeLa” data set. (d) Segmentation result (random colored masks) with manual ground truth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

圖 4：ISBI 細胞追蹤挑戰結果。(a) “PhC-U373”數據集的輸入圖像部分。(b) 分割結果(青色遮罩)與手動真實標註(黃色邊框)(c) “DIC-HeLa”數據集的輸入圖像。(d) 分割結果(隨機顏色的遮罩)與手動真實標註(黃色邊框)。

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).

我們達到了 92% 的平均 IOU (“交集聯合”)，這比第二好的算法 83% 好得多 (見表 2)。

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

表 2：ISBI 細胞追蹤挑戰 2015 的分割結果 (IOU)。

| Name 名稱 | PhC-U373 | DIC-HeLa |
|---------------------------------|---------------|---------------|
| IMCB-SG (2014) IMCB-SG (2014) | 0.2669 | 0.2935 |
| KTH-SE (2014) KTH-SE (2014) | 0.7953 | 0.4607 |
| HOUS-US (2014) HOUS-US (2014) | 0.5323 | - |
| second-best 2015 第二名 2015 | 0.83 | 0.46 |
| u-net (2015) u-net (2015) | 0.9203 | 0.7756 |

The second data set “DIC-HeLa”³

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

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%.

第二組數據集「DIC-HeLa」³ 是通過差分干涉對比 (DIC) 顯微鏡記錄的平面玻璃上的 HeLa 細胞 (見圖 3、圖 4c、d 及補充資料)。它包含 20 張部分註釋的訓練圖像。在這裡，我們達到了 77.5% 的平均 IOU，這顯著優於第二名算法的 46%。

5 Conclusion 5 結論

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 networks⁴

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

. We are sure that the u-net architecture can be applied easily to many more tasks.

u-net 架構在非常不同的生物醫學分割應用中表現非常好。由於使用了彈性變形的數據增強，它只需要非常少量的標註圖像，並且在 NVidia Titan GPU (6 GB) 上只有 10 小時的訓練時間。我們提供了完整的基於 Caffe[6] 的實現和訓練好的網絡⁴。我們相信 u-net 架構可以輕鬆應用於更多的任務。

Acknowledgements 致謝

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., 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., (...), 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/

