

U-Net: Convolutional Networks for Biomedical Image Segmentation

Olaf Ronneberger, Philipp Fischer và Thomas Brox

Khoa Khoa học Máy tính và Trung tâm nghiên cứu tín hiệu sinh học, Đại học Freiburg, Đức
c ronneber@informatik.uni-freiburg.de, Trang chủ www:
http://lmb.informatik.uni-freiburg.de/

Abstract. Có sự đồng ý lớn rằng đào tạo thành công các công trình net sâu yêu cầu nhiều ngàn mẫu đào tạo chú thích. Trong phần này, chúng tôi trình bày một mạng lưới và chiến lược đào tạo phụ thuộc vào việc sử dụng mạnh mẽ việc tăng dữ liệu để sử dụng các mẫu được chú thích có sẵn một cách hiệu quả hơn. Kiến trúc bao gồm một con đường hợp đồng để nắm bắt bối cảnh và một con đường mở rộng đối xứng cho phép địa phương chính xác. Chúng tôi cho thấy rằng một mạng như vậy có thể được đào tạo từ đầu đến cuối từ rất ít hình ảnh và vượt trội so với phương pháp tốt nhất trước đó (mạng chập cửa sổ trượt) trên Thách thức ISBI để phân đoạn các cấu trúc neu trong các n găn xếp kính hiển vi điện tử. Sử dụng cùng một công việc mạng được đào t ạo trên các hình ảnh kính hiển vi ánh sáng truyền qua (độ tương phản pha và DIC), chúng tôi đã giành được Thử thách theo dõi tế bào ISBI 2015 trong các cấu trúc này bằng một biên độ lớn. Hơn nữa, mạng là nhanh. Phân đoạn của hình ảnh 512x512 mất ít hơn một giây trên GPU gần đây. Việc triển khai đầy đủ (dựa trên Caffe) và các mạng được đào tạo có sẵn tại <http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net>.

1 Introduction

Trong hai năm qua, các mạng tích chập sâu đã vượt trội so với trạng thái của nghệ thuật trong nhiều nhiệm vụ nhận dạng trực quan, ví dụ: [7,3]. Mặc dù các mạng tích chập đã tồn tại trong một thời gian dài [8], thành công của chúng bị hạn chế do quy mô của các bộ đào tạo có sẵn và quy mô của các mạng được xem xét. Bước đột phá của Krizhevsky et al. [7] là do đào tạo có giám sát một mạng lưới lớn với 8 lớp và hàng triệu thông số trên bộ dữ liệu ImageNet với 1 triệu hình ảnh đào tạo. Kể từ đó, thậm chí các mạng lớn hơn và sâu hơn đã được đào tạo [12].

Việc sử dụng điển hình của các mạng tích chập nằm trong các tác vụ phân loại, trong đó đầu ra cho hình ảnh là một nhãn lớp đơn. Tuy nhiên, trong nhiều tác vụ trực quan, đặc biệt là trong xử lý hình ảnh y sinh, đầu ra mong muốn nên bao gồm nội địa hóa, tức là, một nhãn lớp được cho là được gán cho mỗi pixel. Hơn nữa, hàng ngàn hình ảnh đào tạo thường nằm ngoài tầm với trong các nhiệm vụ y sinh. Do đó, Ciresan et al. [1] đã đào tạo một mạng trong thiết lập cửa sổ trượt để dự đoán nhãn lớp của mỗi pixel bằng cách cung cấp một vùng cục bộ (bản vá)

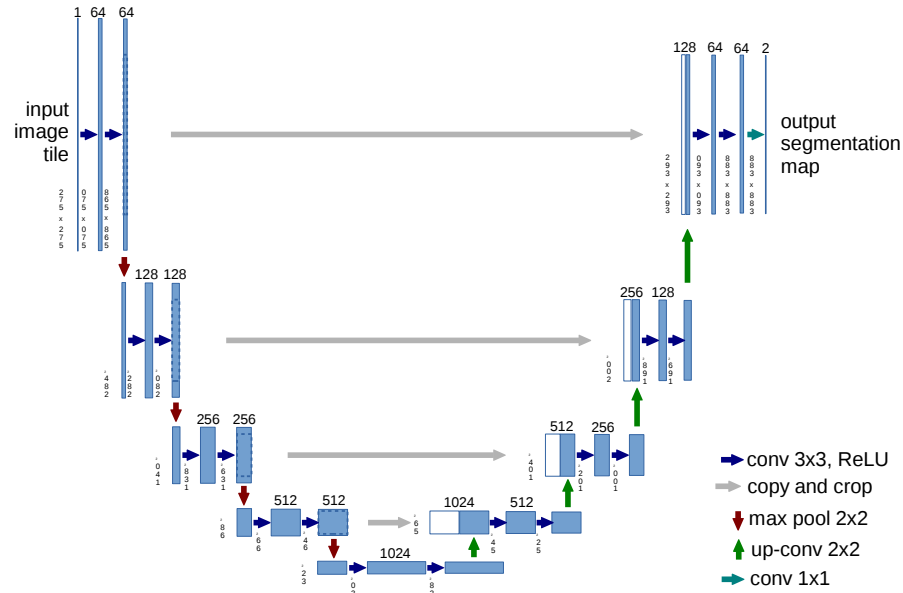


Fig. 1. Kiến trúc U-NET (ví dụ cho 32x32 pixel ở độ phân giải thấp nhất). Mỗi hộp màu xanh tương ứng với bản đồ tính năng đa kênh. Số lượng kênh được biểu thị trên đầu hộp. Kích thước X-Y được cung cấp ở cạnh dưới bên trái của hộp. Hộp trắng đại diện cho bản đồ tính năng sao chép. Các mũi tên biểu thị các hoạt động khác nhau.

như đầu vào. Đầu tiên, mạng này có thể bản địa hóa. Thứ hai, dữ liệu đào tạo về các bản vá lớn hơn nhiều so với số lượng hình ảnh đào tạo. Mạng kết quả đã giành được Thử thách phân khúc EM tại ISBI 2012 bằng một biên độ lớn.

Rõ ràng, chiến lược trong Ciresan et al. [1] có hai nhược điểm. Đầu tiên, nó khá chậm vì mạng phải được chạy riêng cho mỗi bản vá và có rất nhiều dự phòng do các bản vá chồng chéo. Thứ hai, có một sự thương mại giữa độ chính xác nội địa hóa và việc sử dụng bối cảnh. Các bản vá lớn hơn yêu cầu các lớp nhóm tối đa hơn làm giảm độ chính xác nội địa hóa, trong khi các bản vá nhỏ cho phép mạng chỉ thấy ít bối cảnh. Các cách tiếp cận gần đây hơn [11,4] đã đề xuất một đầu ra phân loại có tính đến các tính năng từ nhiều lớp. Nội địa hóa tốt và việc sử dụng bối cảnh là có thể cùng một lúc.

Trong bài báo này, chúng tôi xây dựng dựa trên một kiến trúc thanh lịch hơn, cái gọi là mạng tích chập hoàn toàn của Google [9]. Chúng tôi sửa đổi và mở rộng kiến trúc này sao cho nó hoạt động với rất ít hình ảnh đào tạo và mang lại các phân đoạn chính xác hơn; Xem Hình 1. Ý tưởng chính trong [9] là bổ sung một mạng hợp đồng thông thường bằng các lớp liên tiếp, trong đó các toán tử gộp được thay thế bằng các toán tử lấy mẫu. Do đó, các lớp này làm tăng độ phân giải của đầu ra. Để bản địa hóa, các tính năng có độ phân giải cao từ đường dẫn hợp đồng được kết hợp với UPSampled

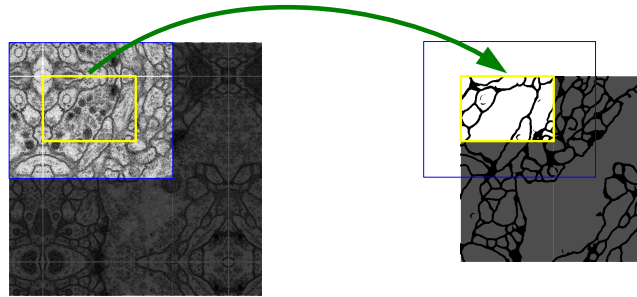


Fig. 2. Chiến lược gạch chồng chéo để phân đoạn liên mạch các hình ảnh lớn tùy ý (ở đây phân đoạn cấu trúc tế bào thần kinh trong các ngăn xếp EM). Dự đoán phân đoạn trong khu vực màu vàng, yêu cầu dữ liệu hình ảnh trong khu vực màu xanh làm đầu vào. Thiếu dữ liệu đầu vào được ngoại suy bằng cách phản chiếu

đầu ra. Một lớp tích chập liên tiếp sau đó có thể học cách lắp ráp một đầu ra chính xác hơn dựa trên thông tin này.

Một điều chỉnh quan trọng trong kiến trúc của chúng tôi là trong phần UPSampling, chúng tôi cũng có một số lượng lớn các kênh tính năng, cho phép mạng truyền bá thông tin ngữ cảnh đến các lớp độ phân giải cao hơn. Kết quả là, con đường mở rộng ít nhiều đối xứng với đường dẫn hợp đồng và mang lại một kiến trúc hình chữ U. Mạng không có bất kỳ lớp kết nối đầy đủ nào và chỉ sử dụng phần hợp lệ của mỗi lần chập, tức là bản đồ phân đoạn chỉ chứa các pixel, trong đó bối cảnh đầy đủ có sẵn trong hình ảnh đầu vào. Chiến lược này cho phép phân đoạn liên mạch của các hình ảnh lớn tùy ý bằng một chiến lược gạch chồng chéo (xem Hình 2). Để dự đoán các pixel trong vùng viền của hình ảnh, bối cảnh bị thiếu được ngoại suy bằng cách phản chiếu hình ảnh đầu vào. Chiến lược ốp lát này rất quan trọng để áp dụng mạng vào các hình ảnh lớn, vì nếu không độ phân giải sẽ bị giới hạn bởi bộ nhớt GPU.

Đối với các nhiệm vụ của chúng tôi, có rất ít dữ liệu đào tạo có sẵn, chúng tôi sử dụng tăng dữ liệu quá mức bằng cách áp dụng các biến dạng đàn hồi cho các thời gian đào tạo có sẵn. Điều này cho phép mạng học được bất biến đối với các biến dạng như vậy, mà không cần phải xem các biến đổi này trong kho chứa hình ảnh được chú thích. Điều này đặc biệt quan trọng trong phân đoạn y sinh, vì biến dạng được sử dụng là biến thể phổ biến nhất trong mô và biến dạng thực tế có thể được mô phỏng một cách hiệu quả. Giá trị của việc tăng dữ liệu cho sự bất biến học tập đã được thể hiện trong Dosovitskiy et al. [2] trong phạm vi học tập tính năng không giám sát.

Một thách thức khác trong nhiều nhiệm vụ phân đoạn tế bào là tách các đối tượng ngấm ứng của cùng một lớp; Xem Hình 3. Đến đầu này, chúng tôi đề xuất việc sử dụng một sự mất trọng có trọng số, trong đó các nhân nền tách biệt giữa các ô chàm có trọng lượng lớn trong hàm mất.

Mạng kết quả được áp dụng cho các phân đoạn y sinh khác nhau. Trong bài báo này, chúng tôi cho thấy kết quả về phân đoạn cấu trúc tế bào thần kinh trong các ngăn xếp EM (một cuộc thi đang diễn ra bắt đầu tại ISBI 2012), nơi chúng tôi ra ngoài

Thực hiện mạng lưới của Cirean et al. [1]. Hơn nữa, chúng tôi hiển thị kết quả phân đoạn tế bào trong hình ảnh kính hiển vi ánh sáng từ tế bào ISBI theo dõi Challenge 2015. Ở đây chúng tôi đã giành chiến thắng với một biên độ lớn trên hai bộ dữ liệu ánh sáng 2D thử thách nhất.

2 Network Architecture

Kiến trúc mạng được minh họa trong Hình 1. Nó bao gồm một đường dẫn hợp đồng (bên trái) và một đường dẫn mở rộng (bên phải). Đường dẫn hợp đồng tuân theo kiến trúc điển hình của một mạng lưới tích chập. Nó bao gồm ứng dụng lặp đi lặp lại của hai lần chập 3×3 (tích chập không được sử dụng), mỗi lần theo sau là một đơn vị tuyến tính trực tràng (RELU) và hoạt động gộp tối đa 2×2 với Stride 2 để lấy mẫu xuống. Ở mỗi bước Downsampling, chúng tôi tăng gấp đôi số lượng kênh tính năng. Mỗi bước trong đường dẫn mở rộng bao gồm việc lấy mẫu của bản đồ tính năng theo sau là tích chập 2×2 (cách mạng UP), giảm một nửa số lượng kênh tính năng, kết nối với bản đồ tính năng được cắt tương ứng từ đường dẫn hợp đồng và hai lần chập số 3×3 , mỗi lần giảm dần. Việc cắt xén là cần thiết do mất các pixel biên giới trong mỗi lần kết hợp. Ở lớp cuối cùng, tích chập 1×1 được sử dụng để ánh xạ từng vectơ tính năng thành phần 64 cho số lượng lớp mong muốn. Tổng cộng mạng có 23 lớp chập.

Để cho phép một ộp lát liền mạch của bản đồ phân đoạn đầu ra (xem Hình 2), điều quan trọng là chọn kích thước gạch đầu vào sao cho tất cả các hoạt động của nhóm tối đa 2×2 được áp dụng cho một lớp có kích thước X và Y chẵn.

3 Training

Các hình ảnh đầu vào và các bản đồ phân đoạn tương ứng của chúng được sử dụng để đào tạo mạng với việc thực hiện giảm độ dốc ngẫu nhiên của Caffee [6]. Do các chập chập không được bảo vệ, hình ảnh đầu ra nhỏ hơn đầu vào của chiều rộng đường viền không đổi. Để giảm thiểu chi phí và sử dụng tối đa bộ nhớ GPU, chúng tôi ủng hộ các ô đầu vào lớn trên một kích thước lô lớn và do đó giảm lô xuống một hình ảnh duy nhất. Theo đó, chúng tôi sử dụng động lượng cao (0.99) sao cho một số lượng lớn các mẫu đào tạo được thấy trước đây xác định bản cập nhật trong bước tối ưu hóa hiện tại.

Hàm năng lượng được tính toán bởi một max mềm thông minh trên bản đồ tính năng cuối cùng kết hợp với chức năng mất entropy chéo. Phần mềm mềm được định nghĩa là $p_k(\mathbf{x}) = \exp(a_k(\mathbf{x})) / \left(\sum_{k'=1}^K \exp(a_{k'}(\mathbf{x})) \right)$ trong đó $a_k(\mathbf{x})$ biểu thị kích hoạt trong kênh tính năng k ở vị trí pixel $\mathbf{x} \in \Omega$ với $\Omega \subset \mathbb{Z}^2$. K là số lớp và $p_k(\mathbf{x})$ là chức năng tối đa gần đúng. Tức là $p_k(\mathbf{x}) \approx 1$ cho k có kích hoạt tối đa $a_k(\mathbf{x})$ và $p_k(\mathbf{x}) \approx 0$ cho tất cả các k khác. Entropy chéo sau đó xử phạt ở mỗi vị trí độ lệch của $p_{\ell(\mathbf{x})}(\mathbf{x})$ từ 1 bằng cách sử dụng

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

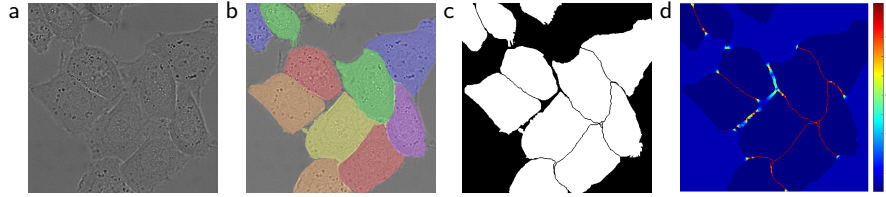


Fig. 3. Các tế bào HeLa trên thủy tinh được ghi lại bằng DIC (độ tương phản nhiễu di động) Mi- croscopy. (a) Hình ảnh thô. (b) Lớp phủ với phân đoạn sự thật mặt đất. Màu sắc khác nhau chỉ ra các trường hợp khác nhau của các tế bào HeLa. . .

Trong đó $\ell: \Omega \rightarrow \{1, \dots, K\}$ là nhãn thực của mỗi pixel và $w: \Omega \rightarrow \mathbb{R}$ là bản đồ trọng lượng mà chúng tôi giới thiệu để cung cấp cho một số pixel quan trọng hơn trong đào tạo.

Chúng tôi tổng hợp bản đồ trọng lượng cho từng phân đoạn sự thật mặt đất để kết hợp tần số khác nhau của các pixel từ một lớp nhất định trong tập dữ liệu đào tạo và buộc mạng phải tìm hiểu các đường viền phân tách nhỏ mà chúng tôi giới thiệu giữa các ô chập (xem Hình 3C và D).

Đường viền phân tách được tính toán bằng cách sử dụng các hoạt động hình ảnh. Bản đồ trọng lượng sau đó được tính toán là

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

trong đó $w_c: \Omega \rightarrow \mathbb{R}$ là bản đồ trọng lượng để cân bằng các tần số lớp, $d_1: \Omega \rightarrow \mathbb{R}$ biểu thị khoảng cách đến đường viền của ô gần nhất và $d_2: \Omega \rightarrow \mathbb{R}$ khoảng cách đến đường viền của ô gần nhất. Trong các thử nghiệm của chúng tôi, chúng tôi đặt $w_0 = 10$ và $\sigma \approx 5$ pixel.

Trong các mạng sâu với nhiều lớp chập và các đường dẫn khác nhau qua mạng, việc khởi tạo tốt các trọng số là vô cùng quan trọng. Otherwise, các phần của mạng có thể kích hoạt quá mức, trong khi các phần khác không bao giờ đóng góp. Lý tưởng nhất là các trọng số ban đầu nên được điều chỉnh sao cho mỗi bản đồ tính năng trong mạng có phương sai gần đúng. Đối với một mạng với kiến trúc của chúng tôi (các lớp tích chập và relu xen kẽ), điều này có thể đạt được bằng cách vẽ các trọng số ban đầu từ phân phối Gaussian với độ lệch chuẩn của $\sqrt{2/N}$, trong đó N biểu thị số lượng nút đến của một neuron [5]. Ví dụ. Đối với các kênh tích chập 3x3 và 64 kênh trong lớp trước $N = 9 \cdot 64 = 576$.

3.1 Data Augmentation

Tăng cường dữ liệu là điều cần thiết để dạy cho mạng các thuộc tính bất biến và mạnh mẽ mong muốn, khi chỉ có một vài mẫu đào tạo có sẵn. Trong trường hợp đầu tiên

Hình ảnh kính hiển vi mà chúng ta chủ yếu cần sự thay đổi và xoay bất biến cũng như sự độ mạnh đối với các biến dạng và biến thể giá trị xám. Đặc biệt là các biến dạng ngẫu nhiên của các mẫu đào tạo đường như là khái niệm chính để đào tạo một mạng phân đoạn với rất ít hình ảnh chú thích. Chúng tôi tạo ra các biến dạng trơn tru bằng cách sử dụng các vectơ chuyển vị ngẫu nhiên trên lưới 3×3 . Các chuyển vị được lấy mẫu từ phân phối Gaussian với độ lệch chuẩn 10 pixel. Các chuyển vị trên mỗi pixel sau đó được tính toán bằng cách sử dụng nội suy bicubic. Các lớp bỏ học ở cuối đường dẫn hợp đồng thực hiện tăng cường dữ liệu ngầm.

4 Experiments

Chúng tôi chứng minh việc áp dụng U-NET cho ba nhiệm vụ phân đoạn khác nhau. Nhiệm vụ đầu tiên là phân đoạn cấu trúc tế bào thần kinh trong các bản ghi âm điện não tử. Một ví dụ về tập dữ liệu và phân đoạn thu được của chúng tôi được hiển thị trong Hình 2. Chúng tôi cung cấp kết quả đầy đủ dưới dạng tài liệu bổ sung. Bộ dữ liệu được cung cấp bởi Thử thách phân đoạn EM [14] đã được bắt đầu tại ISBI 2012 và vẫn mở cho những đóng góp mới. Dữ liệu đào tạo là một bộ gồm 30 hình ảnh (512×512 pixel) từ kính hiển vi điện tử truyền qua phần nối tiếp của dây thần kinh tâm thất Instar ấu trùng của *Drosophila* (VNC). Mỗi hình ảnh đi kèm với một bản đồ phân đoạn sự thật được chú thích đầy đủ tương ứng cho các tế bào (màu trắng) và màng (màu đen). Bộ thử nghiệm có sẵn công khai, nhưng bản đồ phân tích của nó được giữ bí mật. Một đánh giá có thể thu được bằng cách gửi bản đồ xác suất màng dự đoán cho ban tổ chức. Việc đánh giá được thực hiện bằng cách ngưỡng bản đồ ở mức 10 mức và tính toán của lỗi cong vênh trên mạng, lỗi Rand Rand và lỗi pixel pixel [14].

U-NET (tính trung bình hơn 7 phiên bản xoay của dữ liệu đầu vào) đạt được bất kỳ lỗi trước hoặc sau xử lý trước là 0,0003529 (điểm mới nhất, xem Bảng 1) và lỗi RAND là 0,0382.

Điều này tốt hơn đáng kể so với kết quả mạng chập cửa sổ trượt của Cireşan et al. Về mặt lỗi rand, hoạt động tốt hơn duy nhất

Table 1. Xếp hạng về Thử thách phân khúc EM [14] (ngày 6 tháng 3 năm 2015), được sắp xếp theo lỗi cong vênh.

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

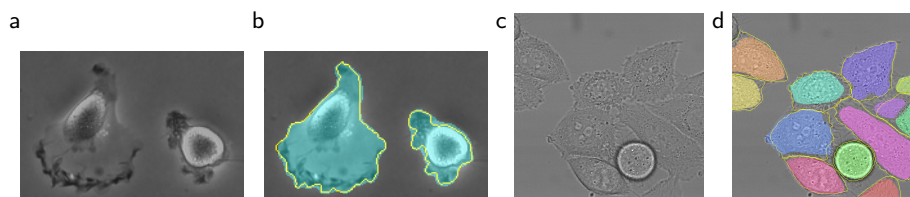


Fig. 4. Kết quả trên Thử thách theo dõi ô ISBI. . . .

Table 2. Kết quả phân đoạn (IOU) trên Thử thách theo dõi ô ISBI 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

Các thuật toán trên tập dữ liệu này Sử dụng các phương thức xử lý hậu xử rất cao về các phương pháp xử lý sau ¹ được áp dụng cho bản đồ xác suất của Cirešan et al. [1]. Chúng tôi cũng đã áp dụng NET U vào một nhiệm vụ phân đoạn tế bào trong các thời gian hiển vi bằng kính hiển vi ánh sáng. Nhiệm vụ Segmentation này là một phần của Thử thách theo dõi ô ISBI 2014 và 2015 [10,13]. Bộ dữ liệu đầu tiên của Phc Phc-U373, ² chứa các tế bào U373 của Glioblastoma-astrocytoma U373 trên chất nền polyacrylimide được ghi lại bằng kính hiển vi tương phản pha (xem Hình 4A, B và Supp. Vật liệu). Nó chứa 35 hình ảnh đào tạo được chú thích một phần. Ở đây, chúng tôi đạt được mức IOU trung bình (giao lộ trên Union Union) là 92%, tốt hơn đáng kể so với thuật toán tốt thứ hai với 83% (xem Table 2). Bộ dữ liệu thứ hai tập hợp DIC DIC-HELA ³ là các tế bào HeLa trên một kính được ghi lại bằng kính hiển vi tương phản nhiễu (DIC) (xem Hình 3, Hình 4C, D và Supp. Vật liệu). Nó chứa 20 hình ảnh đào tạo chú thích một phần. Ở đây chúng tôi đạt được mức trung bình IOU là 77,5%, tốt hơn đáng kể so với thuật toán tốt thứ hai với 46%.

5 Conclusion

Kiến trúc U-NET đạt được hiệu suất rất tốt trên các ứng dụng phân đoạn y học rất khác nhau. Nhờ gia tăng dữ liệu với độ đàn hồi-

¹ Các tác giả của thuật toán này đã gửi 78 giải pháp khác nhau để đạt được kết quả này. ² Bộ dữ liệu được cung cấp bởi Tiến sĩ Sanjay Kumar. Khoa Đại học Sinh học California tại Berkeley. Berkeley CA (Hoa Kỳ) ³ Bộ dữ liệu được cung cấp bởi Trung tâm y tế bác sĩ Bert van Cappellen Erasmus. Rotterdam. Hà Lan

Mations, nó chỉ cần rất ít hình ảnh chú thích và có thời gian đào tạo rất hợp lý chỉ 10 giờ trên GPU NVIDIA Titan (6 GB). Chúng tôi cung cấp thực hiện dựa trên Caff e đầy đủ [6] và các mạng được đào tạo ⁴. Chúng tôi chắc chắn rằng kiến trúc U-NET có thể được áp dụng dễ dàng cho nhiều nhiệm vụ hơn.

Acknowledgements

Nghiên cứu này được hỗ trợ bởi Sáng kiến Xuất sắc của Chính phủ Liên bang và Tiểu bang Đức (Exc 294) và BMBF (FKZ 0316185B).

References

1. Trong: NIP. Trang 2852 Từ 2860 (2012) 2. Dosovitskiy, A., Springenberg, J.T., Riedmiller, M., Brox, T. Trong: NIPS (2014) 3. Girshick, R., Donahue, J., Darrell, T., Malik, J. : Phân cấp tính năng phong phú để phát hiện đối tượng và phân đoạn ngữ nghĩa. Trong: Thủ tục tổ tụng của Hội nghị IEEE về Tầm nhìn máy tính và nhận dạng mẫu (CVPR) (2014) 4. Hariharan, B., Arbelaz, P., Girshick, R., Malik, J. J. : Xuất sắc vào trực tràng: Vượt qua hiệu suất cấp độ con người trên ImageNet Classification (2015), Arxiv: 1502.01852 [CS.CV] 6. Jia, Y., S helhamer, E., Donahue, J. Những tính năng nhanh (2014), ARXIV: 1408.5093 [CS.CV] 7. Kri zhevsky, A., Sutskever, L., Hinton, G.E. : Phân loại ImageNet với mạng lưới thần kinh liên tục sâu. Trong: NIP. Trang 1106 Từ 1114 (2012) 8. Tính toán thần kinh I (4), 541 Từ 551 (1989) 9
10. Maska, M., (...), De Solorzano, C.O. : Một chuẩn mực để so sánh các thuật toán theo dõi tế bào. Tin sinh học 30, 1609 Từ 1617 (2014)
11. Seyedhosseini, M., Sajjadi, M., Tasdizen, T. : Phân đoạn hình ảnh với các mô hình phân cấp xếp tầng và các mạng bình thường khác nhau. Trong: Tầm nhìn máy tính (ICCV), Hội nghị quốc tế IEEE 2013 về. trang 2168 Từ 2175 (2013)
12. Simonyan, K., Zisserman, A. : Mạng tích chập rất sâu để nhận dạng hình ảnh quy mô lớn (2014), ARXIV: 1409.1556 [CS.CV]
13. WWW: Trang web của Thử thách theo dõi ô, http://www.codesolorzano.com/celltrackingchallenge/Cell_Tracking_Challenge/Welcome.html
- 14.

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