

U-Net: Mạng kết hợp cho y sinh Phân đoạn hình ảnh

Olaf Ronneberger, Philipp Fischer và Thomas Brox

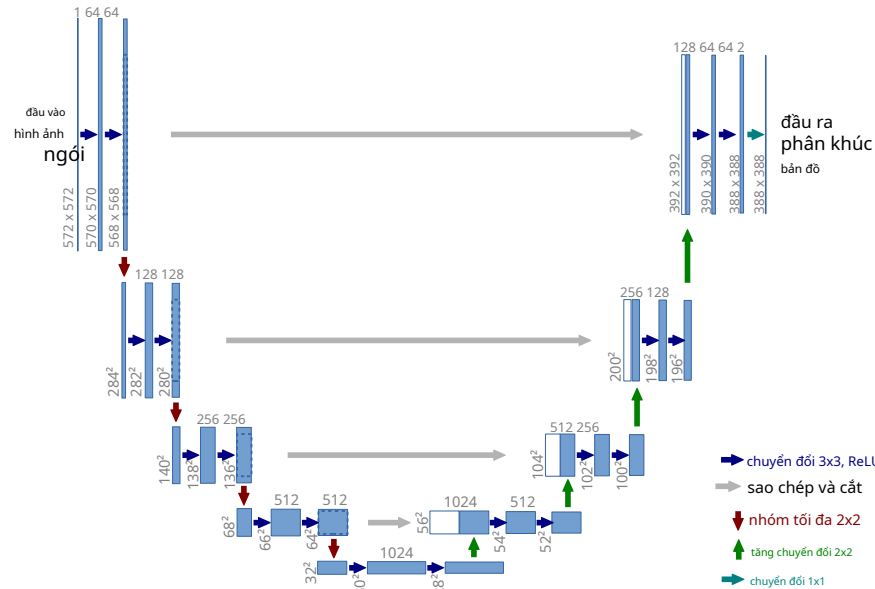
Khoa Khoa học Máy tính và Trung tâm Nghiên cứu Tín hiệu Sinh học BIOSS,
Đại học Freiburg, Đức
ronneber@informatik.uni-freiburg.de ,
Trang chủ WWW:<http://lmb.informatik.uni-freiburg.de/>

Trừu tượng. Có sự đồng ý lớn rằng việc đào tạo thành công các mạng sâu đòi hỏi hàng nghìn mẫu đào tạo có chú thích. Trong bài báo này, chúng tôi trình bày một chiến lược đào tạo và mạng dựa trên việc sử dụng mạnh mẽ công cụ tăng dữ liệu để sử dụng các mẫu chú thích có sẵn hiệu quả hơn. Kiến trúc bao gồm một đường dẫn hợp đồng để nắm bắt ngữ cảnh và một đường dẫn mở rộng đối xứng cho phép bản địa hóa 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 tích chập cửa sổ trượt) trong thử thách ISBI để phân đoạn cấu trúc nơ-ron trong ngăn xếp kính hiển vi điện tử. Sử dụng cùng một mạng được đào tạo về hình ảnh kính hiển vi ánh sáng truyền qua (độ tương phản pha và DIC), chúng tôi đã giành chiến thắng trong thử thách theo dõi tế bào ISBI năm 2015 ở các hạng mục này với tỷ số cách biệt lớn. Hơn thế nữa, mạng nhanh. Phân đoạn hình ảnh 512x512 mất chưa đến 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. Giới thiệu

Trong hai năm qua, các mạng tích chập sâu đã vượt trội so với công nghệ tiên tiến nhất trong nhiều nhiệm vụ nhận dạng hình ảnh, ví dụ: [7,3]. Mặc dù các mạng tích chập đã tồn tại từ lâu [số 8], thành công của họ bị hạn chế do kích thước của các tập huấn luyện có sẵn và kích thước của các mạng được xem xét. Bước đột phá của Krizhevsky et al. [7] là do đào tạo có giám sát của một mạng lớn với 8 lớp và hàng triệu tham số trên bộ dữ liệu ImageNet với 1 triệu hình ảnh đào tạo. Kể từ đó, các mạng thậm chí còn lớn hơn và sâu hơn đã được đào tạo [12].

Việc sử dụng điển hình của các mạng tích chập là trong các nhiệm vụ phân loại, trong đó đầu ra của một 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 phải bao gồm nội địa hóa, nghĩa là 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 từng pixel bằng cách cung cấp một vùng cục bộ (bản vá) xung quanh pixel đó

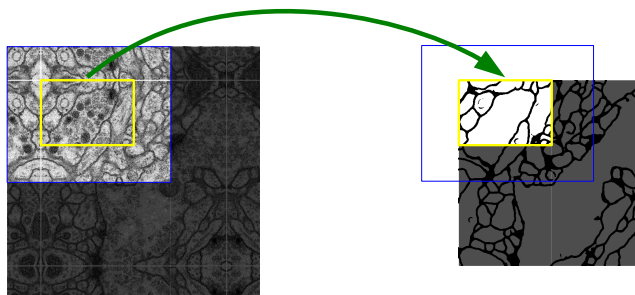


Hình 1. Kiến trúc U-net (ví dụ cho 32x32 pixel ở độ phân giải thấp nhất). Mỗi hộp màu xanh tương ứng với một bản đồ tính năng đa kênh. Số lượng kênh được biểu thị trên đầu hộp. Kích thước xy được cung cấp ở cạnh dưới bên trái của hộp. Các hộp màu trắng đại diện cho các bản đồ tính năng được sao chép.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

làm 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. Kết quả là mạng đã giành chiến thắng trong thử thách phân khúc EM tại ISBI 2012 với tỷ số cách biệt lớn.

Rõ ràng, chiến lược trong Ciresan et al. [1] có hai nhược điểm. Thứ nhất, nó khá chậm vì phải chạy mạng riêng cho từng bản vá và có rất nhiều dự phòng do các bản vá chồng chéo. Thứ hai, có **sự đánh đổi giữa độ chính xác của nội địa hóa và việc sử dụng ngữ cảnh**. Các bản vá lớn hơn yêu cầu nhiều **lớp tổng hợp tối đa** hơn làm giảm độ chính xác của bản địa hóa, trong khi các bản vá nhỏ cho phép mạng chỉ nhìn thấy một ngữ cảnh nhỏ. Các cách tiếp cận gần đây hơn [11,4] đã đề xuất một đầu ra của bộ phân loại có tính đến các tính năng từ nhiều lớp. Đồng thời có thể bản địa hóa tốt và sử dụng ngữ cảnh.

Trong bài báo này, chúng tôi xây dựng dựa trên một kiến trúc tao nhã hơn, cái gọi là “mạng tích chập hoàn toàn” [9]. Chúng tôi sửa đổi và mở rộng kiến trúc này sao cho nó hoạt động với rất ít hình ảnh đào tạo và tạo ra các phân đoạn chính xác hơn; nhìn thấy Hình 1. Ý tưởng chính trong [9] là để bổ sung cho mạng hợp đồng thông thường bằng các lớp liên tiếp, trong đó các toán tử tổng hợp 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 upsampling



Hình 2. Chiến lược xếp chồng lên nhau để phân đoạn liên mạch các hình ảnh lớn tùy ý (ở đây là phân đoạn cấu trúc nơ-ron trong ngăn xếp EM). Dự đoán phân đoạn trong vùng màu vàng, yêu cầu dữ liệu hình ảnh trong vùng màu xanh làm đầu vào. Thiếu dữ liệu đầu vào được ngoại suy bằng cách phản chiếu

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

Một sửa đổi quan trọng trong kiến trúc của chúng tôi là trong phần lấy mẫu, chúng tôi cũng có một số lượng lớn các kênh đặc trưng, cho phép mạng truyền thông tin ngữ cảnh đến các lớp có độ phân giải cao hơn. Kết quả là, đường dẫn mở rộng ít nhiều đối xứng với đường dẫn hợp đồng và tạo ra một kiến trúc hình chữ u. Mạng không có bất kỳ lớp được kết nối đầy đủ nào và chỉ sử dụng phần hợp lệ của mỗi tích chập, nghĩa là bản đồ phân đoạn chỉ chứa các pixel mà ngữ cảnh đầy đủ có sẵn trong hình ảnh đầu vào. Chiến lược này cho phép phân đoạn liên mạch các hình ảnh lớn tùy ý bằng chiến lược xếp chồng lên nhau (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 sắp xếp này rất quan trọng để áp dụng mạng cho các hình ảnh lớn, vì nếu không thì độ phân giải sẽ bị giới hạn bởi bộ nhớ GPU.

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

Một thách thức khác trong nhiều tác vụ phân đoạn ô là tách các đối tượng cảm ứng của cùng một lớp; nhìn thấy Hình 3. Để đạt được mục tiêu này, chúng tôi đề xuất sử dụng phương pháp giảm trọng số, trong đó các nhãn nền phân tách giữa các ô chạm nhau có trọng số lớn trong hàm mất mát.

Mạng kết quả có thể áp dụng cho các vấn đề phân đoạn y sinh khác nhau. Trong bài báo này, chúng tôi đưa ra kết quả về việc phân đoạn cấu trúc nơ-ron trong ngăn xếp EM (một cuộc thi đang diễn ra bắt đầu tại ISBI 2012), nơi chúng tôi vượt qua

thực hiện mạng của Ciresan et al. [1]. Hơn nữa, chúng tôi hiển thị kết quả cho việc phân đoạn tế bào trong hình ảnh kính hiển vi ánh sáng từ thử thách theo dõi tế bào ISBI năm 2015. Tại đây, chúng tôi đã giành chiến thắng với tỷ số cách biệt lớn trên hai bộ dữ liệu ánh sáng truyền qua 2D thử thách nhất.

2 Kiến trúc mạng

Kiến trúc mạng được minh họa trong Hình 1. Nó bao gồm một đường thu hẹp (bên trái) và một đường mở rộng (bên phải). Đường dẫn hợp đồng tuân theo kiến trúc điển hình của mạng tích chập. Nó bao gồm ứng dụng lặp đi lặp lại của hai tích chập 3x3 (tích chập không đệm), mỗi lần theo sau là một đơn vị tuyến tính được chỉnh lưu (ReLU) và thao tác tổng hợp tối đa 2x2 với bước 2 để lấy mẫu xuống. Ở mỗi bước lấy mẫu xuống, chúng tôi nhân đôi số lượng kênh đặc trưng. Mỗi bước trong đường dẫn mở rộng bao gồm lấy mẫu nâng cấp của bản đồ tính năng, sau đó là tích chập 2x2 ("tích chập lên") để giảm một nửa số lượng kênh tính năng, nối với bản đồ tính năng được cắt tương ứng từ đường dẫn hợp đồng và hai bước 3x3 tích chập, mỗi theo sau là một ReLU. Việc cắt xén là cần thiết do mất pixel đường viền trong mỗi lần tích chập. Ở lớp cuối cùng, tích chập 1x1 được sử dụng để ánh xạ từng vectơ đặc trưng 64 thành phần tới số lượng lớp mong muốn. Tổng cộng mạng có 23 lớp tích chập.

Để cho phép sắp xếp liền mạch bản đồ phân đoạn đầu ra (xem Hình 2), điều quan trọng là phải chọn kích thước ô đầu vào sao cho tất cả các hoạt động tổng hợp tối đa 2x2 được áp dụng cho một lớp có kích thước x và y chẵn.

3 Đào tạo

Các hình ảnh đầu vào và bản đồ phân đoạn tương ứng của chúng được sử dụng để huấn luyện mạng với việc triển khai giảm dần độ dốc ngẫu nhiên của Caffè [6]. Do các cấu chập không được đệm, hình ảnh đầu ra nhỏ hơn đầu vào với chiều rộng đường viền không đổi. Để giảm thiểu chi phí hoạt động và tận dụng tối đa bộ nhớ GPU, chúng tôi ưu tiên các ô đầu vào lớn hơn kích thước lô lớn và do đó giảm lô thành 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 đã thấy trước đó xác định bản cập nhật trong bước tối ưu hóa hiện tại.

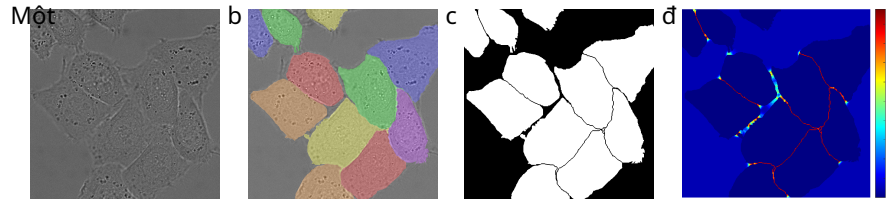
Hàm năng lượng được tính toán bằng soft-max theo từng pixel trên giá trị cuối cùng bản đồ đặc trưng kết hợp với (tổn thất entropy chéo) chức năng. Soft-max là

định nghĩa là $P_k(x) = \exp(\text{kinh nghiệm}(M_{\text{out}}(x))) / \sum_{k=1}^K \exp(\text{kinh nghiệm}(M_{\text{out}}(x)))$ nơi $M_{\text{out}}(x)$ biểu thị

kích hoạt trong kênh tính năng tại vị trí pixel $x \in \Omega$ với $\Omega \subset \mathbb{Z}^2$. K là số lớp và $P_k(x)$ là hàm tối đa gần đúng. $\mathbb{I} \in P_k(x) \approx 1$ cho k có kích hoạt tối đa $M_{\text{out}}(x)$ và $P_k(x) \approx 0$ cho tất cả những người khác k . Entropy chéo sau đó xử phạt ở mỗi vị trí độ lệch của

$P_k(x)$ từ 1 lần sử dụng

$$L = \sum_{x \in \Omega} w(x) \text{đăng nhập}(P_k(x)) \quad (1)$$



Hình 3. Các tế bào HeLa trên kính được ghi bằng kính hiển vi DIC (độ tương phản nhiễu vi sai). (a) hình ảnh thô. (b) lớp phủ với phân đoạn sự thật mặt đất. Các màu khác nhau biểu thị các phiên bản khác nhau của các ô HeLa. (c) mặt nạ phân đoạn được tạo (trắng: nền trước, đen: nền). (d) bản đồ với trọng lượng giảm theo pixel để buộc mạng tìm hiểu các pixel đường viền.

Ở đây $\Omega \rightarrow \{1, \dots, K\}$ là nhãn thật của từng pixel và $w: \Omega \rightarrow \mathbb{R}$ là một bản đồ trọng lượng mà chúng tôi đã giới thiệu để tăng tầm quan trọng của một số pixel trong quá trình đào tạo.

Chúng tôi tính toán trước bản đồ trọng số cho từng phân đoạn sự thật cơ bản để bù tần số pixel khác nhau từ một lớp nhất định trong tập dữ liệu huấn luyện và buộc mạng tìm hiểu các đường viền phân tách nhỏ mà chúng tôi giới thiệu giữa các ô chạm vào (Xem Hình 3c và d).

Đường viền phân tách được tính toán bằng các phép toán hình thái học. Bản đồ trọng lượng sau đó được tính là

$$w(x) = w_c(x) + w_o \cdot \frac{(\sigma(x) + \sigma(x-1))}{2\sigma} \quad (2)$$

Ở đây $w_c: \Omega \rightarrow \mathbb{R}$ là bản đồ trọng số để cân bằng tần số lớp, $\sigma: \Omega \rightarrow \mathbb{R}$ biểu thị khoảng cách đến đường viền của ô gần nhất và $\sigma-1: \Omega \rightarrow \mathbb{R}$ khoảng cách đến đường viền của ô gần thứ hai. Trong các thí nghiệm của chúng tôi, chúng tôi đặt $w_o=10$ và $\sigma \approx 5$ điểm ảnh.

Trong các mạng sâu với nhiều lớp tích chập và các đường dẫn khác nhau xuyên qua mạng, việc khởi tạo tốt các trọng số là vô cùng quan trọng. Mặt khác, 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 đặc trưng trong mạng có phương sai xấp xỉ đơn vị. Đối với một mạng có 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 tiêu chuẩn lệch $2/N$. Ở đây N biểu thị số lượng nút đến của một nơ-ron [5]. Ví dụ: tích chập 3×3 và 64 kênh đặc trưng ở lớp trước $N=9 \cdot 64 = 576$.

3.1 Tăng cường dữ liệu

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

hình ảnh hiển vi, chúng ta chủ yếu cần tính bất biến của chuyển động và xoay cũng như độ bền đối với các biến dạng và biến thể giá trị xám. Đặc biệt, các biến dạng đàn hồi 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 được chú thích. Chúng tôi tạo ra các biến dạng mịn bằng cách sử dụng các vectơ dịch chuyển ngẫu nhiên trên lưới thô 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 phép nội suy nhị phân. Các lớp bỏ qua ở cuối đường dẫn hợp đồng thực hiện thêm dữ liệu ẩn.

4 thí nghiệm

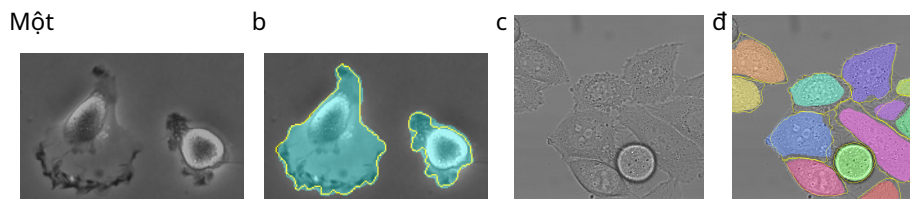
Chúng tôi chứng minh ứng dụng của u-net cho ba tác vụ phân đoạn khác nhau. Nhiệm vụ đầu tiên là phân đoạn các cấu trúc nơ-ron trong các bản ghi bằng kính hiển vi điện tử. Một ví dụ về tập dữ liệu và phân khúc thu được của chúng tôi được hiển thị trong [Hình 2](#). Chúng tôi cung cấp kết quả đầy đủ dưới dạng Tài liệu bổ sung. Tập dữ liệu được cung cấp bởi thử thách phân đoạn EM [\[14\]](#) đã được bắt đầu tại ISBI 2012 và vẫn đang 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 bụng ấu trùng *Drosophila* đầu tiên (VNC). Mỗi hình ảnh đi kèm với một bản đồ phân đoạn sự thật mặt đất được chú thích đầy đủ tương ứng cho các tế bào (màu trắng) và màng (màu đen). Bộ thử nghiệm được cung cấp công khai, nhưng bản đồ phân đoạn của nó được giữ bí mật. Có thể thu được đánh giá bằng cách gửi bản đồ xác suất màng dự đoán cho ban tổ chức. Việc đánh giá được thực hiện bằng cách lập ngưỡng bản đồ ở 10 cấp độ khác nhau và tính toán “lỗi cong vênh”, “lỗi Rand” và “lỗi pixel” [\[14\]](#).

Mạng u-net (trung bình trên 7 phiên bản xoay vòng của dữ liệu đầu vào) đạt được mà không cần xử lý thêm trước hoặc sau lỗi cong vênh là 0,0003529 (điểm tốt nhất mới, xem [Bảng 1](#)) và sai số rand là 0,0382.

Điều này tốt hơn đáng kể so với kết quả mạng tích chập cửa sổ trượt của Cirezan et al. [\[1\]](#), bài nộp tốt nhất có lỗi cong vênh là 0,000420 và lỗi rand là 0,0504. Về lỗi rand, chỉ hoạt động tốt hơn

Bảng 1. Xếp hạng trong thử thách phân khúc EM [\[14\]](#) (06/03/2015), sắp xếp theo lỗi cong vênh.

Thứ hạng	Tên nhóm	Lỗi cong vênh	Lỗi Rand	Lỗi điểm ảnh
	* giá trị con người**	0,000005	0,0021	0,0010
1.	mạng u	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.	Lặn	0,000430	0,0545	0,0582
...				
10.	IDSIA-SCI	0,000653	0,0189	0,1027



Hình 4. Kết quả về thử thách theo dõi tế bào ISBI. (MỘT) một phần của hình ảnh đầu vào của bộ dữ liệu "PhC-U373". (b) Kết quả phân đoạn (mặt nạ màu lục lam) với sự thật nền thủ công (viền màu vàng) (c) hình ảnh đầu vào của tập dữ liệu "DIC-HeLa". (d) Kết quả phân đoạn (mặt nạ có màu ngẫu nhiên) với sự thật mặt đất thủ công (viền màu vàng).

Ban 2. Kết quả phân đoạn (IOU) trong thử thách theo dõi tế bào ISBI 2015.

Tên	PhC-U373	DIC-HeLa
IMCB-SG (2014)	0,2669	0,2935
KTH-ĐN (2014)	0,7953	0,4607
HOUS-US (2014)	0,5323	-
nhất nhì 2015	0,83	0,46
u-net (2015)	0,9203	0,7756

các thuật toán trên tập dữ liệu này sử dụng các phương pháp xử lý hậu kỳ cụ thể của tập dữ liệu cao¹ áp dụng cho bản đồ xác suất của Ciresan et al. [1].

Chúng tôi cũng đã áp dụng u-net cho nhiệm vụ phân đoạn tế bào trong các hình ảnh hiển vi nhẹ. Nhiệm vụ phân đoạn này là một phần của thử thách theo dõi tế bào ISBI 2014 và 2015 [10,13]. Bộ dữ liệu đầu tiên "PhC-U373"² chứa các tế bào 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 IOU trung bình ("giao điểm trên liên kết") là 92%, tốt hơn đáng kể so với thuật toán tốt thứ hai với 83% (xem ban 2). Bộ dữ liệu thứ hai "DIC-HeLa"³ là các tế bào HeLa trên một tấm kính phẳng được ghi lại bằng kính hiển vi tương phản nhiễu vi sai (DIC) (xem Hình 3, hình 4c, d và Supp. Vật liệu). Nó chứa 20 hình ảnh đào tạo được chú thích một phần. Ở đây, chúng tôi đạt được IOU trung bình là 77,5%, tốt hơn đáng kể so với thuật toán tốt thứ hai với 46%.

5. Kết luận

Kiến trúc u-net đạt được hiệu suất rất tốt trên các ứng dụng phân đoạn y sinh rất khác nhau. Nhờ tăng cường dữ liệu với biến dạng đàn hồi-

¹Các tác giả của thuật toán này đã gửi 78 giải pháp khác nhau để đạt được điều này kết quả.

²Bộ dữ liệu được cung cấp bởi Tiến sĩ Sanjay Kumar. Khoa Công Nghệ Sinh Học Đại Học của California tại Berkeley. Berkeley CA (Mỹ)

³Bộ dữ liệu do Trung tâm Y tế Tiến sĩ Gert van Cappellen Erasmus cung cấp. Rotterdam. Hà Lan

mation, 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 đầy đủ Caffel[6] triển khai dựa trên J và các mạng được đào tạo⁴. Chúng tôi chắc chắn rằng kiến trúc u-net có thể được áp dụng dễ dàng cho nhiều tác vụ hơn nữa.

lời cảm ơn

Nghiên cứu này được hỗ trợ bởi Sáng kiến Xuất sắc của chính phủ Liên bang và Bang Đức (EXC 294) và BMBF (Fkz 0316185B).

Người giới thiệu

1. Ciresan, DC, Gambardella, LM, Giusti, A., Schmidhuber, J.: Mạng thần kinh sâu phân đoạn màng tế bào thần kinh trong hình ảnh kính hiển vi điện tử. Trong: NIPS. trang 2852–2860 (2012)
2. Dosovitskiy, A., Springenberg, JT, Riedmiller, M., Brox, T.: Học tập tính năng không giám sát phân biệt đối xử với các mạng thần kinh tích chập. Trong: NIPS (2014)
3. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Hệ thống phân cấp tính năng phong phú để phát hiện đối tượng chính xác và phân đoạn ngữ nghĩa. Trong: Kỷ yếu Hội nghị IEEE về Thị giác máy tính và Nhận dạng mẫu (CVPR) (2014)
4. Hariharan, B., Arbelaz, P., Girshick, R., Malik, J.: Hypercolumns for object segmentation and fine-grained localization (2014), arXiv:1411.5752 [cs.CV]
5. He, K., Zhang, X., Ren, S., Sun, J.: Đi sâu vào bộ chỉnh lưu: Vượt qua hiệu suất của con người về phân loại imagenet (2015), arXiv:1502.01852 [cs.CV]
6. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffel: Kiến trúc tích chập để nhúng tính năng nhanh (2014), arXiv:1408.5093 [cs.CV]
7. Krizhevsky, A., Sutskever, I., Hinton, GE: Phân loại Imagenet với các mạng thần kinh tích chập sâu. Trong: NIPS. trang 1106–1114 (2012)
8. LeCun, Y., Boser, B., Denker, JS, Henderson, D., Howard, RE, Hubbard, W., Jackel, LD: Lan truyền ngược áp dụng cho nhận dạng mã zip viết tay. Tính toán thần kinh 1(4), 541–551 (1989)
9. Long, J., Shelhamer, E., Darrell, T.: Mạng chập hoàn toàn cho phân đoạn ngữ nghĩa (2014), arXiv:1411.4038 [cs.CV]
10. Mask, M., (...), de Solorzano, CO: Điểm chuẩn để so sánh các thuật toán theo dõi tế bào. Tin sinh học 30, 1609–1617 (2014)
11. Seyedhosseini, M., Sajjadi, M., Tasdizen, T.: Phân đoạn hình ảnh với các mô hình phân cấp theo tầng và các mạng thông thường phân tách logistic. Trong: Thị giác máy tính (ICCV), Hội nghị quốc tế IEEE 2013 về. trang 2168–2175 (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 tế bào, http://www.codesolorzano.com/celltrackingchallenge/Cell_Tracking_Challenge/Welcome.html
14. WWW: Trang web của thử thách phân đoạn em, http://brainiac2.mit.edu/isbi_challenge/

⁴Triển khai U-net, mạng được đào tạo và tài liệu bổ sung có sẵn tại <http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net>