U-Net: Mạng tích chập cho Y sinh học Phân đoạn hình ảnh

Olaf Ronneberger, Philipp Fischer và Thomas Brox

Khoa học máy tính và Trung tâm nghiên cứu tín hiệu sinh học BIOSS, Đại học Freiburg, Đức
ronneber@informatik.uni-freiburg.de
http://lmb.informatik.uni-freiburg.de/

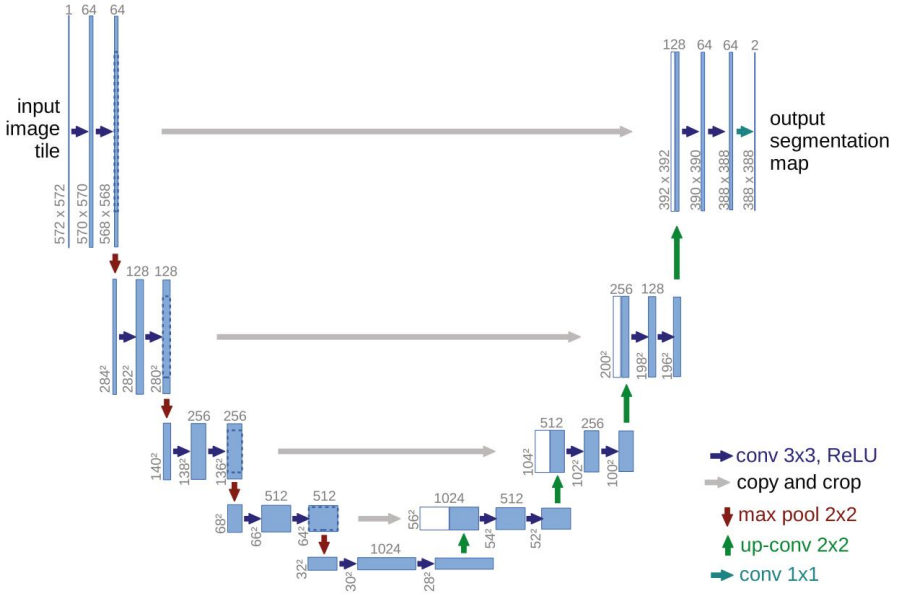
Tóm tắt. Có sự đồng thuận lớn rằng việc đào tạo thành công các mạng lưới sâu đòi hỏi hàng nghìn mẫu đào tạo có chú thích. Trong bài báo này, chúng tôi trình bày một mạng lưới và chiến lược đào tạo dựa trên việc sử dụng mạnh mẽ việc tăng cường dữ liệu để sử dụng các mẫu có chú thích hiệu quả hơn. Kiến trúc bao gồm một đường dẫn co lại để nắm bắt ngữ cảnh và một đường dẫn mở rộng đối xứng cho phép định vị chính xác. Chúng tôi chỉ ra rằng một mạng lưới như vậy có thể được đào tạo từ đầu đến cuối từ rất ít hình ảnh và vượt trội hơn phương pháp tốt nhất trước đây (mạng tích chập cửa sổ trượt) trong thử thách ISBI về phân đoạn các cấu trúc nơ-ron trong các ngăn xếp kính hiển vi điện tử. Sử dụng cùng một mạng lưới được đào tạo trên các 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 biên độ lớn. Hơn nữa, mạng lưới này rất nhanh. Phân đoạn hình ảnh 512x512 mất chưa đến một giây trên GPU mới nhất. Bản triển khai đầy đủ (dựa trên Caffe) và các mạng lưới 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 hơn so với công nghệ tiên tiến nhất trong nhiều tác vụ nhận dạng hình ảnh, ví dụ [7]. Mặc dù các mạng tích chập đã tồn tại trong một thời gian dài [8], nhưng thành công của chúng bị hạn chế do quy mô của các tập huấn luyện có sẵn và quy mô của các mạng được xem xét. Bước đột phá của Krizhevsky và cộng sự [7] là do đào tạo có giám sát một mạng lớn với 8 lớp và hàng triệu tham số trên tập 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 mạng tích chập là trên các tác vụ phân loại, trong đó đầu ra cho một hình ảnh là một nhãn lớp duy nhất. 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 định vị, tức là, một nhãn lớp được cho là được gán cho mỗi pixel. Hơn nữa, hàng nghìn hình ảnh đào tạo thường nằm ngoài tầm với trong các tác vụ y sinh.
Do đó, Ciresan et al. [2] 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) xung quanh pixel đó

cộng sự (Biên tập viên): MICCAI 2015, Phần III, LNCS 9351, trang 234–241, 2015.
DOI: 10.1007/978-3-319-24574-4_28

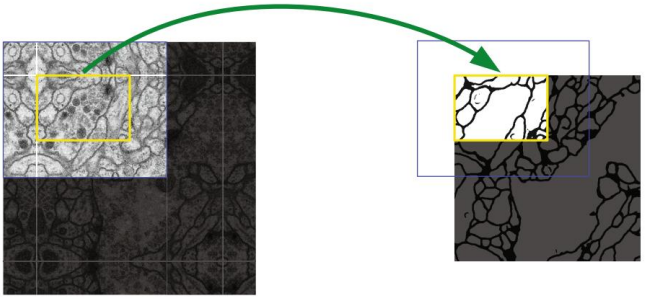


Hình 1. Kiến trúc U-net (ví dụ cho 32x32 pixel ở độ phân giải thấp nhất). Mỗi hộp màu xanh tư ơ ng ứng với một bản đồ tính năng đa kênh. Số lư ợng kênh đư ợc biểu thị ở trên cùng của hộp. Kích thư ớc xy đư ợc cung cấp ở cạnh dư ới bên trái của hộp. Các hộp màu trắng biểu thị các bản đồ tính năng đã sao chép. Các mũi tên biểu thị các thao tác khác nhau.

làm đầu vào. Đầu tiên, mạng này có thể định vị. Thứ hai, dữ liệu đào tạo về các bản vá lớn hơ n nhiều so với số lư ợng hình ảnh đào tạo. Mạng kết quả đã giành chiến thắng trong thử thách phân đoạn EM tại ISBI 2012 với biên độ lớn.

Rõ ràng, chiến lư ợc trong Ciresan et al. [2] có hai như ợc điểm. Đầu tiên, nó khá chậm vì mạng phải chạy riêng cho từng bản vá và có rất nhiều sự trùng lặp do các bản vá chồng chéo. Thứ hai, có sự đánh đổi giữa độ chính xác định vị và việc sử dụng ngữ cảnh. Các bản vá lớn hơ n yêu cầu nhiều lớp max-pooling hơ n làm giảm độ chính xác định vị, trong khi các bản vá nhỏ chỉ cho phép mạng thấy ít ngữ cảnh. Các phư ơ ng pháp tiếp cận gần đây hơ n [11,4] đã đề xuất một đầu ra phân loại có tính đến các tính năng từ nhiều lớp. Có thể định vị tốt và sử dụng ngữ cảnh cùng một lúc.

Trong bài báo này, chúng tôi xây dựng dựa trên một kiến trúc thanh lịch hơ n, 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 hơ n; xem Hình 1. Ý tư ởng chính trong [9] là bổ sung cho mạng hợp đồng thông thư ờng bằng các lớp liên tiếp, trong đó các toán tử gộp đư ợc thay thế bằng các toán tử lấy mẫu tăng. Do đó, các lớp này làm tăng độ phân giải của đầu ra. Để định vị, các đặc điểm có độ phân giải cao từ đư ờng dẫn co lại đư ợc kết hợp với đầu ra đư ợc lấy mẫu tăng. Sau đó, một lớp tích chập kế tiếp có thể học cách lắp ráp đầu ra chính xác hơ n dựa trên thông tin này.



Hình 2. Chiến lư ợc chồng chéo để phân đoạn liền mạch các hình ảnh lớn tùy ý (ở đây là phân đoạn các cấu trúc nơ -ron trong các ngăn xếp EM). Dự đoán phân đoạn trong vùng màu vàng, yêu cầu dữ liệu hình ảnh trong vùng màu xanh lam làm đầu vào. Dữ liệu đầu vào bị thiếu đư ợc ngoại suy bằng cách phản chiếu

Một thay đổi quan trọng trong kiến trúc của chúng tôi là trong phần upsampling, chúng tôi cũng có một số lư ợng lớn các kênh đặc điểm, cho phép mạng truyền thông tin ngữ cảnh đến các lớp có độ phân giải cao hơ n. Do đó, đư ờng dẫn mở rộng ít nhiều đối xứng với đư ờng dẫn thu hẹp và tạo ra kiến trúc hình chữ U. Mạng không có bất kỳ lớp nào đư ợc kết nối đầy đủ và chỉ sử dụng phần hợp lệ của mỗi phép tích chập, tức là bản đồ phân đoạn chỉ chứa các pixel mà ngữ cảnh đầy đủ có sẵn trong hình ảnh đầu vào.

Chiến lư ợc này cho phép phân đoạn liền mạch các hình ảnh lớn tùy ý bằng chiến lư ợc chồng chéo (xem Hình 2). Để dự đoán các điểm ảnh trong vùng viền của hình ảnh, ngữ cảnh bị thiếu đư ợc ngoại suy bằng cách phản chiếu hình ảnh đầu vào.
Chiến lư ợc lát gạch này rất quan trọng để áp dụng mạng cho những hình ảnh lớn, 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 cư ờng 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 đư ợc tính bất biến đối với các biến dạng như vậy, mà không cần phải xem các chuyển đổi này trong ngữ liệu hình ảnh đư ợc chú thích. Điều này đặc biệt quan trọng trong phân đoạn y sinh học, vì biến dạng từng là biến thể phổ biến nhất trong mô và các biến dạng thực tế có thể đư ợc mô phỏng hiệu quả. Giá trị của việc tăng cư ờng dữ liệu để học tính bất biến đã đư ợc thể hiện trong Dosovitskiy et al. [3] trong phạm vi học tính năng không giám sát.

Một thách thức khác trong nhiều tác vụ phân đoạn tế bào là việc tách các đối tư ợng chạm vào nhau cùng một lớp; xem Hình 3. Để đạt đư ợc mục đích này, chúng tôi đề xuất sử dụng một tổn thất có trọng số, trong đó các nhãn nền tách biệt giữa các tế bào chạm vào nhau có trọng số lớn trong hàm tổn thất.

Mạng lư ới kết quả có thể áp dụng cho nhiều vấn đề phân đoạn y sinh học khác nhau. Trong bài báo này, chúng tôi trình bày kết quả về phân đoạn các cấu trúc nơ -ron trong các ngăn xếp EM (một cuộc thi đang diễn ra bắt đầu tại ISBI 2012), trong đó chúng tôi đã vư ợt trội hơ n mạng lư ới của Ciresan et al. [2]. Hơ n nữa, chúng tôi trình bày kết quả về phân đoạn tế bào trong hình ảnh kính hiển vi quang học từ thử thách theo dõi tế bào ISBI 2015. Ở đây, chúng tôi đã giành chiến thắng với biên độ lớn trong hai tập dữ liệu ánh sáng truyền 2D đầy thách thức nhất.

2 Kiến trúc mạng

Kiến trúc mạng đư ợc minh họa trong Hình 1. Nó bao gồm một đư ờng dẫn co lại (bên trái) và một đư ờng dẫn mở rộng (bên phải). Đư ờng dẫn co lại tuân theo kiến trúc điển hình của mạng tích chập. Nó bao gồm ứng dụng lặp lại của hai phép tích chập 3x3 (các phép tích chập không đệm), mỗi phép theo sau là một đơ n vị tuyến tính chỉnh lư u (ReLU) và một hoạt động gộp tối đa 2x2 với bư ớc 2 để giảm mẫu. Tại mỗi bư ớc giảm mẫu, chúng tôi nhân đôi số kênh đặc trư ng. Mỗi bư ớc trong đư ờng dẫn mở rộng bao gồm một phép lấy mẫu tăng của bản đồ đặc trư ng theo sau là một phép tích chập 2x2 (“phép tích chập tăng”) làm giảm một nửa số kênh đặc trư ng, một phép nối với bản đồ đặc trư ng đư ợc cắt tư ơ ng ứng từ đư ờng dẫn co lại và hai phép tích chập 3x3, mỗi phép theo sau là một ReLU. Việc cắt xén là cần thiết do mất các pixel đư ờng viền trong mọi phép tích chập. Ở lớp cuối cùng, tích chập 1x1 đư ợc sử dụng để ánh xạ từng vectơ đặc trư ng 64 thành phần thành số lớp mong muốn. Tổng cộng, mạng có 23 lớp tích chập.

Để cho phép ghép liền mạch bản đồ phân đoạn đầu ra (xem Hình 2), điều quan trọng là phải chọn kích thư ớc ô đầu vào sao cho tất cả các hoạt động gộp tối đa 2x2 đều đư ợc áp dụng cho một lớp có kích thư ớc x và y đều.

3 Đào tạo

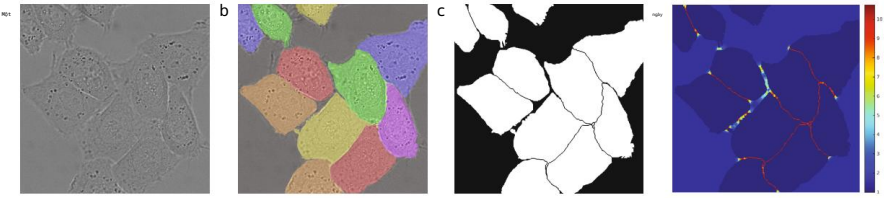
Các hình ảnh đầu vào và bản đồ phân đoạn tư ơ ng ứng của chúng đư ợc sử dụng để đào tạo mạng với việc triển khai phư ơ ng pháp giảm dần độ dốc ngẫu nhiên của Caffe [6]. Do các phép tích chập không đệm, hình ảnh đầu ra nhỏ hơ n hình ảnh đầu vào theo chiều rộng đư ờng viền không đổi. Để giảm thiểu chi phí chung và tận dụng tối đa bộ nhớ GPU, chúng tôi ư u tiên các ô đầu vào lớn hơ n kích thư ớc lô lớn và do đó giảm lô xuống còn một hình ảnh duy nhất. Theo đó, chúng tôi sử dụng động lư ợng cao (0,99) sao cho một số lư ợng lớn các mẫu đào tạo đã thấy trư ớc đó xác định bản cập nhật trong bư ớc tối ư u hóa hiện tại.

Hàm năng lư ợng đư ợc tính bằng soft-max từng pixel trên bản đồ đặc trư ng cuối cùng kết hợp với hàm mất entropy chéo. Soft-max đư ợc định nghĩa là $p_k(x) = \exp(a_k(x)) / \sum_{k=1}^{K} \exp(a_k(x))$ trong đó $a_k(x)$ biểu thị sự kích hoạt trong kênh đặc trư ng k tại vị trí pixel $x \in \Omega$ với $\Omega \subset \mathbb{Z}^2$. K là số lớp và $p_k(x)$ là hàm cực đại xấp xỉ. Tức là $p_k(x) \approx 1$ đối với k có sự kích hoạt cực đại $a_k(x)$ và $p_k(x) \approx 0$ đối với tất cả k khác. Sau đó, entropy chéo sẽ phạt tại mỗi vị trí độ lệch của $p_{\ell(x)}(x)$ so với 1 bằng cách sử dụng

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

trong đó: $\Omega \subset \mathbb{Z}^2$ là nhãn thực của mỗi pixel và $w: \Omega \to \mathbb{R}$ là bản đồ trọng số mà chúng tôi giới thiệu để tăng thêm tầm quan trọng cho một số pixel trong quá trình đào tạo.

Chúng tôi tính toán trư ớc bản đồ trọng số cho mỗi phân đoạn thực tế để bù đắp tần suất khác nhau của các pixel từ một lớp nhất định trong quá trình đào tạo



Hình 3. Tế bào HeLa trên kính đư ợc ghi lại bằng kính hiển vi DIC (độ tư ơ ng phản giao thoa khác biệt). (a) hình ảnh thô. (b) phủ lớp phân đoạn thực tế. Các màu khác nhau chỉ ra các trư ờng hợp khác nhau của tế bào HeLa. (c) mặt nạ phân đoạn đư ợc tạo ra (trắng: tiền cảnh, đen: hậu cảnh). (d) bản đồ với trọng số mất mát từng pixel để buộc mạng phải học các pixel viền.

tập dữ liệu và buộc mạng phải học các ranh giới phân tách nhỏ mà chúng ta đư a vào giữa các ô tiếp xúc (Xem Hình 3c và d).

Đư ờng biên phân cách đư ợc tính toán bằng các phép toán hình thái. Trọng lư ợng bản đồ sau đó đư ợc tính toán như

$$w(x) = wc(x) + w0 \cdot \exp\left(\frac{(d1(x) + d2(x))2}{2 \text{ giây } 2}\right) \quad (2)$$

trong đó wc : Ω R là bản đồ trọng số để cân bằng tần số lớp, d1 : Ω R biểu thị khoảng cách đến đư ờng viền của ô gần nhất và d2 : Ω R là khoảng cách đến đư ờng viền của ô gần thứ hai. Trong các thí nghiệm của chúng tôi, chúng tôi đặt w0 = 10 và σ ≈ 5 pixel.

Trong các mạng sâu có nhiều lớp tích chập và các đư ờng dẫn khác nhau qua mạng, việc khởi tạo trọng số tốt là cực kỳ quan trọng. Nếu không, một số phần của mạng có thể cung cấp quá nhiều kích hoạt, trong khi các phần khác không bao giờ đóng góp. Lý tư ởng nhất là các trọng số ban đầu nên đư ợc điều chỉnh sao cho mỗi bản đồ đặc điểm trong mạng có phư ơ ng sai xấp xỉ đơ n vị. Đối với mạng có kiến trúc của chúng tôi (xen kẽ các lớp tích chập và ReLU), điều này có thể đạt đư ợc bằng cách rút các trọng số ban đầu từ phân phối Gaussian với độ lệch chuẩn là 2/N, trong đó N biểu thị số nút đến của một nơ -ron [5]. Ví dụ đối với tích chập 3x3 và 64 kênh đặc điểm trong lớp trư ớc đó N = 9 · 64 = 576.

3.1 Tăng cư ờng dữ liệu

Việc tăng cư ờng dữ liệu là điều cần thiết để dạy cho mạng các thuộc tính bất biến và độ mạnh mong muốn, khi chỉ có một vài mẫu đào tạo. Trong trư ờng hợp hình ảnh hiển vi, chúng ta chủ yếu cần bất biến dịch chuyển và quay cũng như độ mạnh đối với biến dạng và các biến thể giá trị xám. Đặc biệt, biến dạng đàn hồi ngẫu nhiên của các mẫu đào tạo có vẻ là khái niệm chính để đào tạo mạng phân đoạn với rất ít hình ảnh đư ợc chú thích. Chúng tôi tạo ra các biến dạng trơ n tru bằng cách sử dụng các vectơ dịch chuyển ngẫu nhiên trên lư ới thô 3 x 3.

Bảng 1. Xếp hạng về thách thức phân đoạn EM [14] (ngày 6 tháng 3 năm 2015), đư ợc sắp xếp do lỗi cong vênh.

Xếp hạng Tên nhóm	Lỗi cong vênh	Lỗi Rand	Lỗi điểm ảnh
** giá trị con ngư ời **	0,000005	0,0021	0,0010
1. u-net	0,000353	0,0382	0,0611
2. DIVE-SCI 3.	0,000355	0,0305	0,0584
IDSIA [2]	0,000420	0,0504	0,0613
4. LẶN	0,000430	0,0545	0,0582
⋮			
10. IDSIA-SCI	0,000653	0,0189	0,1027

Các dịch chuyển đư ợc lấy mẫu từ phân phối Gaussian với độ lệch chuẩn 10 pixel. Các dịch chuyển trên mỗi pixel sau đó đư ợc tính toán bằng cách sử dụng phép nội suy bicubic. Các lớp drop-out ở cuối đư ờng dẫn co lại thực hiện thêm tăng cư ờng dữ liệu ngầm.

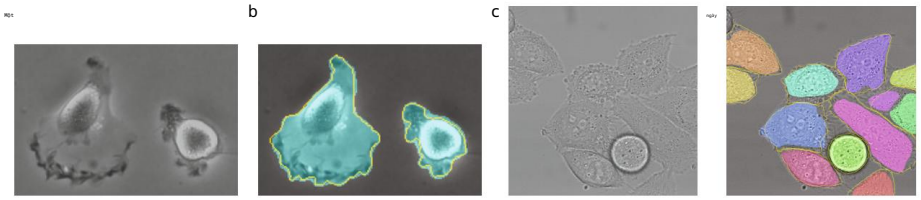
4 Thí nghiệm

Chúng tôi trình bày ứng dụng của u-net cho ba phân đoạn khác nhau Nhiệm vụ đầu tiên là phân đoạn các cấu trúc nơ -ron trong electron bản ghi vi mô. Một ví dụ về tập dữ liệu và phân đoạn thu đư ợc của chúng tôi đư ợc hiển thị trong Hình 2. Chúng tôi cung cấp kết quả đầy đủ dư ới dạng Tài liệu bổ sung. Tập dữ liệu đư ợc cung cấp bởi thử thách phân đoạn EM [14,1] đã bắt đầu tại ISBI 2012 và vẫn mở cửa cho những đóng góp mới. Dữ liệu đào tạo là một tập hợp 30 hình ảnh (512x512 pixel) từ phần truyền điện tử nối tiếp kính hiển vi của dây thần kinh bụng ấu trùng tuổi đầu tiên của ruồi giấm (VNC). Mỗi hình ảnh đi kèm với phân đoạn thực tế có chú thích đầy đủ tư ơ ng ứng bản đồ cho các tế bào (màu trắng) và màng (màu đen). Bộ thử nghiệm đư ợc công khai, như ng bản đồ phân đoạn của nó đư ợc giữ bí mật. Đánh giá có thể đư ợc thực hiện bằng gửi bản đồ xác suất màng dự đoán cho những ngư ời tổ chức. Đánh giá đư ợc thực hiện bằng cách ngư ỡng bản đồ ở 10 mức độ khác nhau và tính toán "lỗi cong vênh", "lỗi Rand" và "lỗi pixel" [14].

u-net (trung bình trên 7 phiên bản xoay của dữ liệu đầu vào) đạt đư ợc lỗi cong vênh 0,0003529 (mới) mà không cần bất kỳ quá trình tiền xử lý hoặc hậu xử lý nào. điểm số cao nhất, xem Bảng 1) và sai số rand là 0,0382.

Điều này tốt hơ n đáng kể so với mạng tích chập cửa sổ trư ợt kết quả của Ciresan et al. [2], bài nộp tốt nhất của họ có lỗi cong vênh là 0,000420 và lỗi rand là 0,0504. Về lỗi rand thì chỉ có hiệu suất tốt hơ n các thuật toán trên tập dữ liệu này sử dụng các phư ơ ng pháp xử lý hậu kỳ rất cụ thể cho tập dữ liệu[1] đư ợc áp dụng cho bản đồ xác suất của Ciresan et al. [2].

[1] Các tác giả của thuật toán này đã đư a ra 78 giải pháp khác nhau để đạt đư ợc điều này kết quả.



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

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

Tên	PhC-U373	DIC-HeLa
IMCB-SG (2014)	0,2669	KTH-SE (2014) 0,7953 HOUS-US (2014) 0,5323 thứ hai tốt nhất 2015 0,83	0,2935	0,4607 - 0,46
mạng u (2015)	0,9203	0,7756

Chúng tôi cũng áp dụng u-net vào nhiệm vụ phân đoạn tế bào trong hình ảnh hiển vi ánh sáng. Nhiệm vụ phân đoạn này là một phần của thử thách theo dõi tế bào ISBI năm 2014 và 2015 [10,13]. Bộ dữ liệu đầu tiên "PhC-U373"2 chứa Glioblastoma-astrocytoma Tế bào U373 trên chất nền polyacrylimide đư ợc ghi lại bằng kính hiển vi tư ơ ng phản pha (xem Hình 4a, b và Tài liệu bổ sung). Nó chứa 35 hình ảnh đào tạo đư ợc chú thích một phần. Ở đây chúng tôi đạt đư ợc IOU trung bình ("giao điểm trên hợp") là 92%, tốt hơ n đáng kể so với thuật toán tốt thứ hai với 83% (xem Bảng 2). Bộ dữ liệu thứ hai "DIC-HeLa"3 là các tế bào HeLa trên một tấm kính phẳng đư ợc ghi lại bằng kính hiển vi tư ơ ng phản giao thoa vi sai (DIC) (xem Hình 3, Hình 4c,d và Tài liệu bổ sung). Nó chứa 20 hình ảnh đào tạo đư ợc chú thích một phần. Ở đây chúng tôi đạt đư ợc IOU trung bình là 77,5%, tốt hơ n đáng kể so với thuật toán tốt thứ hai với 46%.

5 Kết luận

Kiến trúc u-net đạt đư ợc hiệu suất rất tốt trên các ứng dụng phân đoạn y sinh học rất khác nhau. Nhờ tăng cư ờng dữ liệu bằng biến dạng đàn hồi, nó chỉ cần rất ít hình ảnh đư ợc chú thích và có một thời gian đào tạo chỉ 10 giờ trên GPU NVidia Titan (6 GB). Chúng tôi cung cấp

[2] Bộ dữ liệu đư ợc cung cấp bởi Tiến sĩ Sanjay Kumar. Khoa Kỹ thuật sinh học, Đại học của California tại Berkeley. Berkeley CA (Hoa Kỳ).

[3] Bộ dữ liệu đư ợc cung cấp bởi Tiến sĩ Trung tâm y tế Gert van Cappellen Erasmus. Thành phố Rotterdam. Hà Lan.

triển khai đầy đủ dựa trên Caffe[6] 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.

Lời cảm ơn. Nghiên cứu này được hỗ trợ bởi Sáng kiến Xuất sắc của Chính quyền Liên bang và Tiểu bang Đức (EXC 294) và BMBF (Fkz 0316185B).

Tài liệu tham khảo

1. Cardona, A., et al.: Phân tích kiến trúc vi mô và vĩ mô tích hợp của não drosophila bằng kính hiển vi điện tử cắt lớp hỗ trợ máy tính. PLoS Sinh học. 8(10), e1000502 (2010)
2. Ciresan, DC, Gambardella, LM, Giusti, A., Schmidhuber, J.: Mạng lưới nơ-ron sâu phân đoạn màng nơ-ron trong hình ảnh kính hiển vi điện tử. Trong: NIPS, trang 2852–2860 (2012)
3. Dosovitskiy, A., Springenberg, JT, Riedmiller, M., Brox, T.: Học tính năng không giám sát phân biệt với mạng nơ-ron tích chập. Trong: NIPS (2014)
4. Hariharan, B., Arbel´aez, P., Girshick, R., Malik, J.: Siêu cột để phân đoạn đối tượng và định vị chi tiết (2014), arXiv:1411.5752 [cs.CV]
5. He, K., Zhang, X., Ren, S., Sun, J.: Đi sâu vào bộ chỉnh lưu: Vượt trội hiệu suất ở cấp độ 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.: Caffe: 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 mạng nơ-ron tích chập sâu. Trong: NIPS, trang 1106–1114 (2012)
8. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, LD: Backpropagation áp dụng cho nhận dạng mã bưu chính viết tay. Thần kinh Tính toán 1(4), 541–551 (1989)
9. Long, J., Shelhamer, E., Darrell, T.: Mạng tích chập hoàn toàn cho ngữ nghĩa phân đoạn (2014), arXiv:1411.4038 [cs.CV]
10. Maska, M., et al.: Một chuẩn mực để so sánh các thuật toán theo dõi tế bào. Bioin- định dạng 30, 1609–1617 (2014)
11. Seyedhosseini, M., Sajjadi, M., Tasdizen, T.: Phân đoạn hình ảnh với thác đổ mô hình phân cấp và mạng chuẩn logistic disjunctive. Trong: Hội nghị quốc tế về tầm nhìn máy tính IEEE năm 2013 (ICCV), trang 2168–2175 (2013)
12. Simonyan, K., Zisserman, A.: Mạng tích chập rất sâu cho quy mô lớn nhận dạng hình ảnh (2014), arXiv:1409.1556 [cs.CV]
13. WWW: Trang web của thử thách theo dõi tế bào, http://www.codesolorzano.com/celltrackingchallenge/Thử thách_theo dõi tế bào/Welcome.html
14. WWW: Trang web của thử thách phân đoạn em, http://brainiac2.mit.edu/isbi_challenge/

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