PHÂN VÙNG TỔN THƯƠNG DA TỪ ẢNH SOI DA BẰNG MÔ HÌNH SEGUNET

SKIN LESION SEGMENTATION FROM DERMOSCOPIC IMAGES BY SEGUNET NEURAL NETWORK

Phạm Văn Trường[1]*, Trần Thị Thảo[1]

[1]Trường Đại học Bách khoa Hà Nội

*Tác giả liên hệ: truong.phamvan@hust.edu.vn

(Nhận bài: 04/01/2021; Chấp nhận đăng: 11/4/2021)

Tóm tắt - Phân tích ảnh soi da là một trong các kỹ thuật được quan tâm trong nghiên cứu ung thư da. Trong phân tích ảnh soi da, việc phân vùng chính xác vùng da bị tổn thương đóng một vai trò quan trọng. Nghiên cứu này đề xuất một mô hình phân vùng tổn thương da từ ảnh soi da bằng mô hình học sâu-SegUnet. Mô hình đề xuất kế thừa những ưu điểm của hai mô hình UNet và SegNet, như khả năng trích chọn các thông tin thô và tinh từ ảnh đầu vào của U-Net; Tính hiệu quả trong tính toán của SegNet. Chúng tôi cũng đề xuất sử dụng phép chuẩn hóa trung bình-phương sai để thay cho phép toán chuẩn hóa theo mẻ như trong các mô hình gốc để giảm số tham số của mô hình. Mô hình được áp dụng trên bộ dữ liệu ISIC 2017 gồm 2000 ảnh huấn luyện và được đánh giá trên một bộ dữ liệu thử nghiệm gồm 600 ảnh. Kết quả cho thấy, mô hình SegUNet cho độ chính xác cao nhất 93,1%, hệ số Dice 0,851, đã chứng minh tính hiệu quả của phương pháp đề xuất.

Từ khóa - Mạng nơ-ron SegNet; Mạng nơ-ron UNet; Phân vùng tổn thương da; Phân vùng ảnh; Mạng Nơ-ron

Abstract - Skin cancer is one of the most widespread cancer types all over the world, but it can be treated if early detected. Nowadays, analyzing the dermoscopic images is a crucial approach in skin cancer study. In particular, accurate segmentation of skin lesion from dermoscopic images play an important role in skin cancer analysis. In this paper, we present an approach for skin lesion segmentation by a deep neural network model namely SegUNet. The proposed model takes the advantages of both SegNet and UNet models, i.e., the ability of capturing fine image information of UNet, and computational efficiency of SegNet. In particular, instead of using batch normalization as in the original model, we propose using mean-variance normalization operation in order to reduce parameters of the network. The model is applied on the ISIC 2017 dataset including 2000 training images and validated on 600 test images. Experimental results show that the proposed model can reach 93.1% of accuracy and obtained the Dice coefficient of 0.851, which demonstrate the effective performances of the proposed approach.

Key words - SegNet; UNet; Skin lesion segmentation; Image segmentation; Deep neural networks

1. Đặt vấn đề

Ung thư da là một vấn đề sức khỏe cộng đồng lớn. Tại Hoa Kỳ, chỉ tính riêng trong năm 2020, có hơn 108 420 ca mới được chẩn đoán và 11 480 ca tử vong [1]. Trong đó, nguy hiểm bậc nhất phải kể đến là ung thư hắc tố da, tên tiếng Anh là Melanoma [1]. Ung thư hắc tố da là do sự xuất hiện của các khối u ác tính ở các tế bào hắc tố, các tế bào được tìm thấy trong lớp biểu bì da với nhiệm vụ sản xuất melanin-sắc tố giúp mang lại màu sắc cho da và mắt. Ung thư hắc tố da chiếm tỷ lệ chủ yếu trong các ca tử vong do ung thư da gây ra. Mặc dù, tỷ lệ tử vong cao, nhưng khi được phát hiện sớm, tỷ lệ sống của người mắc ung thư hắc tố da có thể lên đến 92% [2].

Khi các tổn thương sắc tố xảy ra trên bề mặt da, các khối u ác tính có thể được phát hiện sớm bằng cách kiểm tra bằng trực quan của các chuyên gia, và cũng có thể được phát hiện bằng các phương pháp phân tích hình ảnh. Hiện nay, một trong các kỹ thuật được quan tâm là phân tích ảnh soi da. Kỹ thuật này hiện đang được sử dụng phổ biến trong các bệnh viện và trung tâm chẩn đoán lâm sàng [3]. Trong phân tích ảnh soi da thì việc phát hiện hay phân vùng chính xác vùng bị tổn thương da từ các ảnh soi da đóng một vai trò quan trọng, là tiền đề cho các bước tiếp theo [4].

Mặt khác, trong những năm gần đây, trí tuệ nhân tạo đặc biệt là mạng nơron tích chập (CNN) đã được ứng dụng trong phân tích cũng như phân vùng các hình ảnh y tế. Ví dụ như: Glavan và Holbal [5] sử dụng mạng nơron tích chập để lấy vùng phân vùng sườn từ ảnh chụp X-ray của ngực. Melinscak và các cộng sự [6] dùng mạng nơ-ron để phân vùng những mạch máu. Tiếp đến có thể kể đến các phương pháp dựa trên nền tảng mạng CNN để phát triển các mô hình phù hợp trong các bài toán phân vùng ảnh. Ví dụ, Long và các cộng sự [7] đã đề xuất phương pháp tích chập hoàn toàn Fully Convolutional Network (FCN), bằng cách thay thế lớp kết nối đầy đủ trong bài toán phân loại ảnh bằng lớp tích chập. Với sự ra đời của FCN, các phương pháp dùng kỹ thuật học sâu cho bài toán phân vùng ảnh đã thu được nhiều thành tự nổi bật [8]. Đã có rất nhiều mô hình được phát triển từ mô hình FCN này, như mô hình FCN [9] cho phân vùng ảnh tim của Phi-Vu Tran, mô hình DeconvNet [10], và đặc biệt là U-Net [11] và SegNet [12].

Trong lĩnh vực phân vùng ảnh ở vùng tổn thương da, cũng có nhiều nghiên cứu được đề xuất [13, 14]. Ví dụ, Yu và các cộng sự [13], Bi và các cộng sự [3]. Gần đây Yuan và cộng sự [14] đề xuất sử dụng mô hình FCN với hàm tối ưu là Jaccard. Trong nghiên cứu của Ibtehaz và cộng sự [15], Tang và cộng sự [16], các tác giả đề xuất các mô hình phân vùng tổn thương da sử dụng mô hình Unet với một số cải tiến trong mô hình và kết hợp thêm một số bước xử lý nhằm tăng độ chính xác cho mô hình. Tuy có nhiều nghiên cứu trong hướng sử dụng học sâu, các kết quả cho thấy, vẫn còn chưa được cải tiến nhiều do những thách thức đặc thù của

[1] Hanoi University of Science and Technology (Van-Truong Pham, Tran Thi Thao)

bài toán. Vì vậy bài toán phân vùng ảnh tổn thương da vẫn là một bài toán mở, vẫn được quan tâm nghiên cứu hiện nay.

Trong phạm vi bài báo này, các tác giả đề xuất một mô hình phân vùng tổn thương da từ ảnh soi da bằng mô hình học sâu-mạng SegUNet. Mô hình đề xuất kế thừa những ưu điểm của hai mô hình UNet và SegNet như khả năng trích chọn các thông tin thô và tinh từ ảnh đầu vào của U-Net, tính hiệu quả trong tính toán của SegNet. Mô hình được huấn luyện thông qua một bộ dữ liệu ISIC 2017 bao gồm 2000 ảnh soi da của bảy chủng bệnh da liễu khác nhau. Các tham số hiệu quả được đánh giá dựa trên một bộ dữ liệu thử nghiệm bao gồm 600 hình ảnh. Kết quả thu được cho thấy, mô hình SegUNet cho độ chính xác cao nhất ở 93,1%, hệ số Dice 0,851, đã chứng minh tính hiệu quả của phương pháp đề xuất.

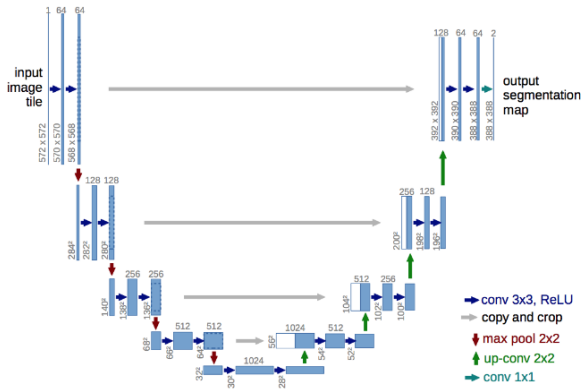
2. Phân vùng ảnh sử dụng kỹ thuật học sâu

Trong những năm gần đây, các mạng nơ ron, nhất là mạng nơ ron tích chập, đã được sử dụng rộng rãi trong những ứng dụng liên quan tới phân vùng ảnh. Trong mạng nơ ron tích chập, các lớp tích chập (convolutional layer) kết hợp với các lớp giảm chiều (pooling) có xu hướng làm cho kích thước của dữ liệu đầu vào bị thu nhỏ lại. Trong bài toán phân vùng ảnh, đầu ra của mô hình là một ảnh khác, hay gọi là mặt nạ (mask) – phân chia chính xác từng vùng riêng biệt, tương ứng với đó là các vật thể trong ảnh, có chiều dài và rộng có thể bằng với ảnh đầu vào. Và để làm được điều này, một trong những cách đó chính là người ta xóa bỏ từ lớp mạng liên kết đầy đủ trở về sau, chỉ giữ lại phần tích chập và giảm chiều, đồng thời thêm vào cuối đó các lớp tích chập khác, và các lớp mở rộng, hay còn gọi là upsampling. Cụ thể, từ các lớp tích chập và giảm chiều đầu vào, ta thu được một loạt các biểu đồ đặc tính (feature map) và từ những biểu đồ đặc tính ấy, thông qua việc mở rộng, tính toán với các lớp tích chập khác, ta sẽ có đầu ra là mặt nạ tương ứng. Như vậy, mô hình có thể được tóm gọn lại với hai phần chính, phần mã hóa (encoder), với đầu ra là các biểu đồ đặc tính, và phần giải mã (decoder) chính là phần mở rộng để thu được mặt nạ phân vùng. Các mô hình như này có thể dễ dàng được cải tiến từ các mô hình mạng CNN vốn đã rất hiệu quả như VGG [17], ResNet [18], DenseNet [19], GoogLeNet [20]…

2.1. Mô hình mạng UNet

Mô hình mạng UNet [11] được giới thiệu bởi Ronnenberger và cộng sự dành cho phân vùng ảnh y sinh. Unet được cải tiến và phát triển dựa trên mô hình mạng nơ ron tích chập toàn phần. Mô hình có tên là UNet do cấu trúc đối xứng tạo thành hình chữ U, như mô tả ở Hình 1.

Cấu trúc của mạng UNet phần mã hóa encoder cũng tương tự như các mạng nơ ron tích chập truyền thống, gồm các lớp tích chập và các lớp giảm chiều nhằm thu được những đặc trưng của ảnh. Trong Hình 2 có thể thấy, các các lớp tích chập sử dụng cửa sổ trượt có kích thước 3x3, lớp giảm chiều với cửa sổ trượt là 2x2, giảm kích thước đầu vào đi 2 lần cả về chiều dài và rộng. Qua các lần tích chập, giảm chiều, kích thước dài rộng của ảnh sẽ giảm dần, do đó là cần phải tăng lên về chiều sâu, số lượng kênh (chanel) thông qua việc tăng lên số cửa sổ trượt ở các lớp tích chập về cuối của phần mã hóa. Việc này giúp tăng thêm hiệu quả học, lượng thông tin, đặc tính trích xuất được. Điều nổi bật của mạng Unet chính là lớp giải mã decoder. Trong kiến trúc mạng Unet, phần decoder gần như đối xứng với encoder. Trong phần decoder, số lần mở rộng chiều sẽ tương ứng với số lần giảm chiều ở các lớp trước đó. Sau một lần mở rộng chiều, ta sẽ kết hợp đầu ra này với đầu ra của lớp tích chập ngay trước lớp giảm chiều tương ứng, rồi tính toán qua một số lớp tích chập khác trước khi đến bước mở rộng tiếp theo. Trong phần decoder ngoài việc mở rộng ta còn thực hiện kết nối đối xứng với các layer phần encoder cho đến tận layer cuối cùng. Việc kết nối đối xứng với phần encoder sẽ giúp ta phục hồi lại thông tin đã mất tại các lớp giảm chiều pooling. So với mô hình FCN, mô hình UNet cho kết quả phân vùng ảnh tốt hơn. Tại thời điểm công bố, năm 2015, mô hình đã đạt được độ chính xác đứng đầu bảng trong thử thách về phân vùng ảnh chụp từ kính hiển vi điện tử, phân vùng tế bào trong ảnh được tổ chức bởi Hội nghị quốc tế về hình ảnh y sinh năm 2012.



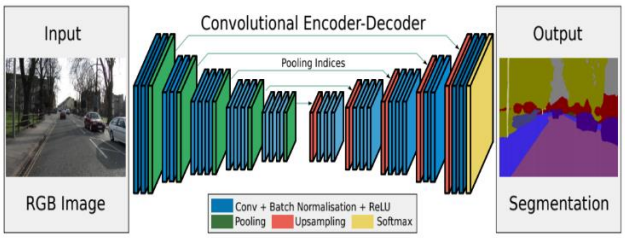
Hình 1. *Mô hình mạng UNet*

2.2. Mô hình mạng SegNet

Mô hình mạng SegNet, được nghiên cứu và phát triển bởi nhóm nghiên cứu về robot và thị giác máy tính của đại học Cambrigde và được công bố lần đầu năm 2016 [12] dành cho bài toán phân vùng ảnh đa lớp. Mô hình SegNet cũng tương tự như mô hình UNet, có phần mã hóa được xây dựng như mạng nơ ron tích chập thông thường, với các lớp tích chập, giảm chiều, và có thể được xây dựng dựa trên những mô hình mạng nổi tiếng và hiệu quả như VGG16, VGG19, ResNet… với phần cuối là lớp mạng liên kết đầy đủ được loại bỏ. Như minh họa trong Hình 2 người ta sử dụng phần mã hóa của mạng VGG16, với 13 lớp tích chập và 5 lớp giảm chiều. Và tại phần giải mã của SegNet cũng gồm có những lớp mở rộng và các lớp tích chập, nhưng điểm đặc biệt hơn và cũng là làm nên hiệu quả của mạng chính là tại những lớp mở rộng, đầu ra được mở rộng dựa vào vị trí ban đầu của các pixel tương ứng ở các lớp giảm chiều trước đó. Để thực hiện điều này, ngay tại phần mã hóa, ở các lớp giảm chiều, vị trí của các điểm ảnh, pixel lớn nhất ở đầu vào sẽ được lưu giữ lại. Những pixel đầu ra của lớp giảm chiều, qua quá trình tính toán, đi qua các lớp tích chập, khi đến lớp mở rộng sẽ được đưa về đúng vị trí của nó khi chưa thực hiện giảm chiều.

Khi so sánh với UNet, ở đây không có sự kết nối với những biểu đồ đặc trưng của lớp trước, nhằm đưa những thông tin quan trọng tới được các lớp sau, nhưng trong mô hình SegNet có những thông tin được lưu trữ và chuyển tới các lớp sau chính là vị trí của các pixel. Điều này thể hiện một ưu điểm khá nổi bật của SegNet, đó là trong những ứng dụng, thiết bị

nhúng, sẽ tốn ít bộ nhớ để lưu trữ hơn. Nhờ có lớp giải mã lớn, phức tạp với nhiều tầng tích chập, mô hình mạng SegNet trong nhiều thử nghiệm thực tế đã cho ra kết quả khả quan, nhất là trong bài toán phân vùng tổn thương da.

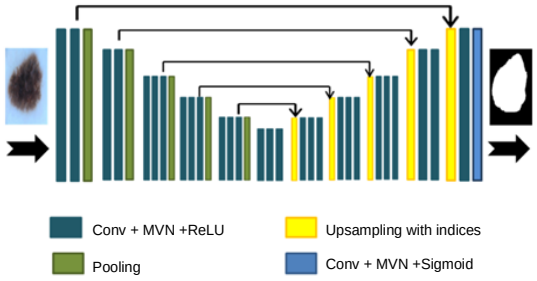


Hình 2. *Mô hình mạng SegNet*

3. Mô hình đề xuất và phương pháp đánh giá

3.1. Mô hình SegUNet

Trong nghiên cứu này, ta sẽ tích hợp mô hình SegNet vào mô hình UNet để có mô hình SegUNet dùng cho bài toán phân vùng da tổn thương. Do đó, kiến trúc của mô hình SegUNet là sự kết hợp giữ hai kiến trúc mạng SegNet và UNet được thể hiện ở Hình 3.



Hình 3. *Mô hình SegUNet được đề xuất*

Mô hình SegUNet được cấu tạo bởi 38 lớp, được chia làm hai phần là phần mã hóa (Encoder) và phần giải mã (Decoder). Giống với mô hình SegNet đã trình bày ở phần trước, phần mã hóa được thừa kế là từ mô hình VGG16 với 13 lớp tích chập và 5 lớp giảm chiều. Các lớp tích chập được sử dụng trong bài toán là tích chập 2D. Các lớp tích chập sử dụng cửa sổ trượt có chiều dài và rộng là 3x3, bước trượt bằng 1 và được thêm lề 0 tại mỗi lớp sao cho kích thước sau mỗi lần tính tích chập là không đổi. Đầu ra của lớp tích chập là biểu đồ đặc tính, có chiều sâu bằng số lượng cửa sổ trượt và lớp tích chập đó sử dụng, và đúng với nguyên tắc, chiều dài và rộng giảm đi. Theo sau của mỗi lớp tích chập là một lớp Batch Normalization, nhằm chuẩn hoá phân phối giá trị đầu ra của lớp tích chập, chuyển về dạng dữ liệu phân phối chuẩn quanh điểm 0 và một lớp nữa là lớp hàm kích hoạt. Ở đây hàm kích hoạt được sử dụng là ReLU, biến các giá trị nhỏ hơn 0 về bằng 0. Bên cạnh lớp tích chập, phần mã hóa của mô hình bao gồm 5 lớp giảm chiều. Các lớp giảm chiều sử dụng cửa sổ trượt có kích thước 2x2, bước trượt bằng 2 và sử dụng phương thức giảm chiều theo giá trị cực đại. Đầu ra của lớp giảm chiều sẽ bị giảm đi còn một nửa về chiều dài và chiều rộng so với đầu vào. Và đặc trưng của mô hình SegNet thể hiện ở đây chính là việc vị trí của các pixel có giá trị lớn nhất sẽ được lưu lại.

Phần giải mã bao gồm 5 lớp mở rộng, tương ứng với 5 lớp giảm chiều và cũng gồm 13 lớp tích chập, 12 lớp đầu cũng sử dụng cửa sổ trượt 3x3, bước trượt 1 và được thêm lề giá trị 0 để kích thước về dài rộng là không đổi. Đi liền với mỗi lớp tích chập là lớp Batch normalization cùng với lớp hàm kích hoạt. Và trong các lớp mở rộng, đầu vào sẽ được mở rộng gấp 2 lần kích thước dài rộng với vị trí của các pixel max được trích xuất từ lớp giảm chiều tương ứng. Điểm đặc biệt so với SegNet thông thường và cũng là đặc trưng của mô hình UNet đó chính là đầu ra của lớp mở rộng sẽ được kết hợp với đầu ra của lớp tích chập đằng trước lớp giảm chiều tương ứng rồi sau đó mới đưa vào lớp tích chập tiếp theo để tính toán. Trong các lớp tích chập trước, hàm kích hoạt là ReLU thì tại lớp tích chập này, hàm kích hoạt được sử dụng là hàm Sigmoid, sẽ thực hiện nhiệm vụ dự đoán nhãn cho mỗi pixel đầu ra của lớp tích chập này. Đây là đối với những bài toán phân vùng đầu ra hai nhãn, có thể biểu diễn được bởi đầu ra một kênh duy nhất. Còn đối với bài toán mà đầu ra phân vùng từ ba nhãn trở nên, số lượng cửa sổ trượt của lớp cuối này sẽ phải tương ứng với số nhãn đầu ra, và khi đó hàm kích hoạt được sử dụng là hàm Softmax.

3.2. Đánh giá kết quả và huấn luyện mạng

Để phân loại cho một ảnh hay một pixel với nhãn tương ứng, người ta sử dụng bốn khái niệm đánh giá sự phân loại đúng hay sai một mẫu, ảnh nào đó. Đó là dương tính thật (True Positive - TP), dương tính giả (False Positive - FP), âm tính thật (True Negative – TN), và âm tính giả (False Negative – FN).

- Dương tính thật: Chỉ những ảnh, mẫu là đúng vật thể cần dự đoán và được mô hình dự đoán đúng.

- Dương tính giả: Chỉ những mẫu không phải là vật thể cần dự đoán nhưng được mô hình dự đoán, phân loại là đúng.

- Âm tính thật: Chỉ những mẫu không phải là vật thể cần dự đoán và được mô hình dự đoán là không phải.

- Âm tính giả: Chỉ những mẫu là vật thể cần dự đoán nhưng bị mô hình phân loại là không phải.

Trong bài toán phân vùng ảnh, khi xét tới phân vùng 2 nhãn, phần vật thể mà mô hình bao đúng là dương tính thật, phần mô hình khoanh thừa, không phải vật là dương tính giả, phần nền mà mô hình không khoanh là âm tính thật, phần của vật thể mà mô hình không khoanh là âm tính giả.

Như vậy, để đánh giá một mô hình phân loại có chính xác hay không, ta đến với tiêu chí đánh giá đầu tiên, cũng là phổ biến nhất khi huấn luyện mạng nơ ron, đó chính là độ chính xác (Accuracy). Độ chính xác được tính bằng tỉ số giữa số mẫu phân loại đúng trên tổng tất cả số mẫu đưa vào phân loại, có nghĩa là bằng tổng số mẫu âm tính thật và dương tính thật chia cho tổng của âm tính thật, dương tính thật, âm tính giả và dương tính giả.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Trong phân vùng ảnh, dữ liệu được dự đoán là từng pixel của ảnh đầu ra. Như vậy, độ chính xác sẽ bằng tỉ số giữa số lượng pixel mà mô hình phân loại đúng trên tổng pixel của ảnh.

Bên cạnh đó, để đánh giá một mô hình phân vùng ảnh, hai tiêu chí khác cũng rất hay được sử dụng, chính là hệ số Dice (Dice Coefficient) và chỉ số Jaccard (Jaccard Index). Dice Coefficient là chỉ số đánh giá độ tương tự của hai đối tượng. Trong bài toán phân vùng ảnh, thì chính là độ tương tự vùng vật thể được xác định bởi mô hình và vùng thực tế. Chỉ số này đã trở thành tiêu chuẩn phổ biến nhất trong việc

đánh giá các mô hình học sâu trong nhiệm vụ phân vùng hình ảnh. Trong lý thuyết về tập hợp, Dice Coefficient được xác định bởi tỉ số giữa hai lần phần trùng nhau của hai tập hợp so với tổng của cả hai tập, ở đây ý nói về số lượng phần tử. Biểu diễn bởi công thức, ta có:

$$\text{Dice Coefficient} = \frac{2*|X \cap Y|}{|X|+|Y|} \quad (2)$$

Trong đó, |X| và |Y| là biểu thị số lượng phần tử của tập X và Y. Từ đó suy ra với ví dụ về phân vùng ảnh, ta có Dice Coeffiecient sẽ được xác định bởi:

$$\text{Dice Coefficient} = \frac{2*TP}{2*TP+FP+FN} \quad (3)$$

Jaccard Index, cũng là một trong những chỉ số để đánh giá độ tương tự của các đối tượng, được xác định như sau:

$$\text{Jaccard Index} = \frac{TP}{TP+FP+FN} \quad (4)$$

Ta có thể nhận thấy, không có quá nhiều khác biệt giữa Jaccard Index là Dice Coefficient, đều là sự đánh giá mức độ tương tự, hay phần trùng nhau của phần bao vật thể giữa dự đoán và thực tế. Nhưng Dice coefficient tập trung nhiều vào độ tương tự của hai mẫu, thì Jaccard Index lại cho cái nhìn rõ ràng hơn về sai khác giữa phần được bao của mô hình so với thực tế. Cùng với Accuracy, đây là sẽ những tiêu chí được sử dụng để đánh giá mô hình được triển khai trong bài báo này.

Trong nghiên cứu này, mô hình được thực thi bằng ngôn ngữ Python trong gói Keras với Tensorflow. Máy tính sử dụng bộ vi xử lý E5−1620 Intel ® Xeon ®, 3.5 GHZ, 64GB RAM. Card đồ họa sử dụng loại Geforce RTX 2080 với 8 GB RAM. Để huấn luyện mạng, thuật toán ta sử dụng chính là Gradient giảm dần với với hệ số học được chọn là 0.01, theo các khuyến nghị của các nghiên cứu trong lĩnh vực phân vùng ảnh. Số lượng Epoch sử dụng là 100 và batch size bằng 8 để đảm bảo việc hội tụ của thuật toán và hiệu quả tính toán. Sau mỗi epoch ta sẽ thực hiện lưu bộ weights để chắc chắn rằng tìm được bộ tham số tốt nhất của mô hình. Hàm mất mát được sử dụng với mô hình này là hàm Dice loss được định nghĩa như sau:

$$\text{Dice loss} = 1 - \text{Dice Coefficient} = 1 - \frac{2*TP}{2*TP+FP+FN} \quad (5)$$

4. Kết quả thí nghiệm

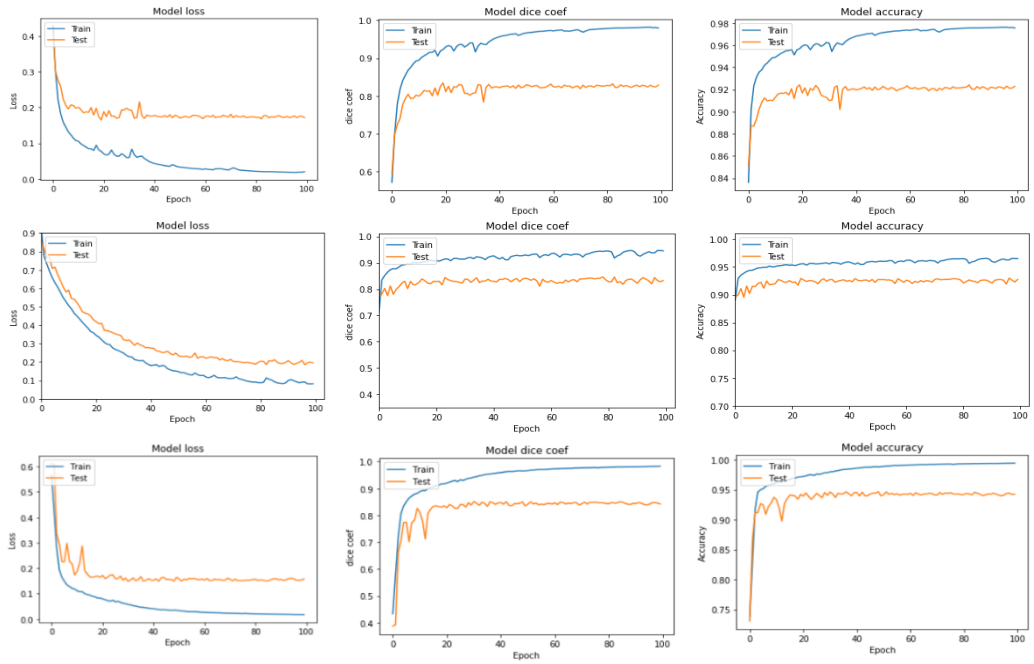
4.1. Cơ sở dữ liệu phân vùng ảnh tổn thương da ISIC2017

Trong nghiên cứu này, để phân vùng ở ảnh ung thư da, dữ liệu được sử dụng sẽ được lấy từ cuộc thi của ISIC 2017, thử thách 1. Dữ liệu gồm ảnh vùng da bị tổn thương và mặt nạ phân vùng da tổn thương được khoanh bởi bác sỹ. Các ảnh được thu thập từ nhiều nguồn với kích thước khác nhau, từ 540x722 tới 4499x6748 pixels. Bộ dữ liệu này được chia thành ba tập tương ứng với tập huấn luyện (Training set), tập đánh giá (Validation set) và tập kiểm tra (Test set). Tập huấn luyện bao gồm 2000 ảnh kèm theo mặt nạ phân vùng, tập đánh giá là 150 ảnh và tập kiểm tra là 600 ảnh. Ảnh đầu vào cùng mặt nạ trước khi đưa vào mô hình để phân vùng sẽ được chuẩn hóa kích cỡ về 256x256. Phần gia tăng dữ liệu được thực hiện với tập huấn luyện bằng một số phép đơn giản như xoay, lật, dịch. Cụ thể là, trong nghiên cứu này nhóm tác giả đã sử dụng phép gia tăng dữ liệu online với phép xoay ảnh 180 độ, phép lật theo phương thẳng đứng; phép thay đổi kích thước (tăng thêm 10% so với ảnh gốc) theo phương ngang. Còn tập đánh giá và kiểm tra thì dữ liệu được giữ nguyên nhằm tạo được hiệu quả tốt nhất cho việc kiểm định mô hình.

4.2. Kết quả phân vùng ảnh tổn thương da

4.2.1. Kết quả của mô hình SegUNet

Để minh họa hiệu quả của mô hình SegUNet cho phân vùng tổn thương da, ta có thể quan sát trên các đường cong học (Learning Curve) trên Hình 4. Trên hình này biểu diễn kết quả huấn luyện trên tập Training set gồm 2000 ảnh và thử nghiệm trên 600 ảnh của tập ISIC 2017.



Hình 4. *Learning curve của mô hình UNet (hàng 1), SegNet (hàng 2), và SegUNet (hàng 3) trên tập ISIC 2017. Cột 1: Hàm mất mát; Cột 2: hệ số Dice; Cột 3: Giá trị accuracy*

Sau quá trình huấn luyện, ta sẽ thấy được sự biến đổi hàm mất mát, chính là hàm Loss giảm dần. Độ chính xác của mô hình là Accuracy, cùng với đó là Dice Coefficient và Jaccrad Index sẽ tăng lên qua các epoch trên cả hai tập huấn luyện và đánh giá. Điều này thể hiện, mô hình đang học và học có hiệu quả. Cụ thể, trên đường cong học này chúng ta có thể thấy mô hình cho kết quả hội tụ chỉ sau khoảng 70 vòng lặp. Các kết quả của tập thử nghiệm cho thấy, độ chính xác Accuracy và hệ số Dice cao và ổn định sau 30 vòng lặp. Cũng cần nói thêm rằng trong các đường cong học, nhóm tác giả không đưa chỉ số Jaccard do sự thay đổi của chỉ số này theo các vòng lặp là tương đồng với chỉ số Dice trên Hình 4.

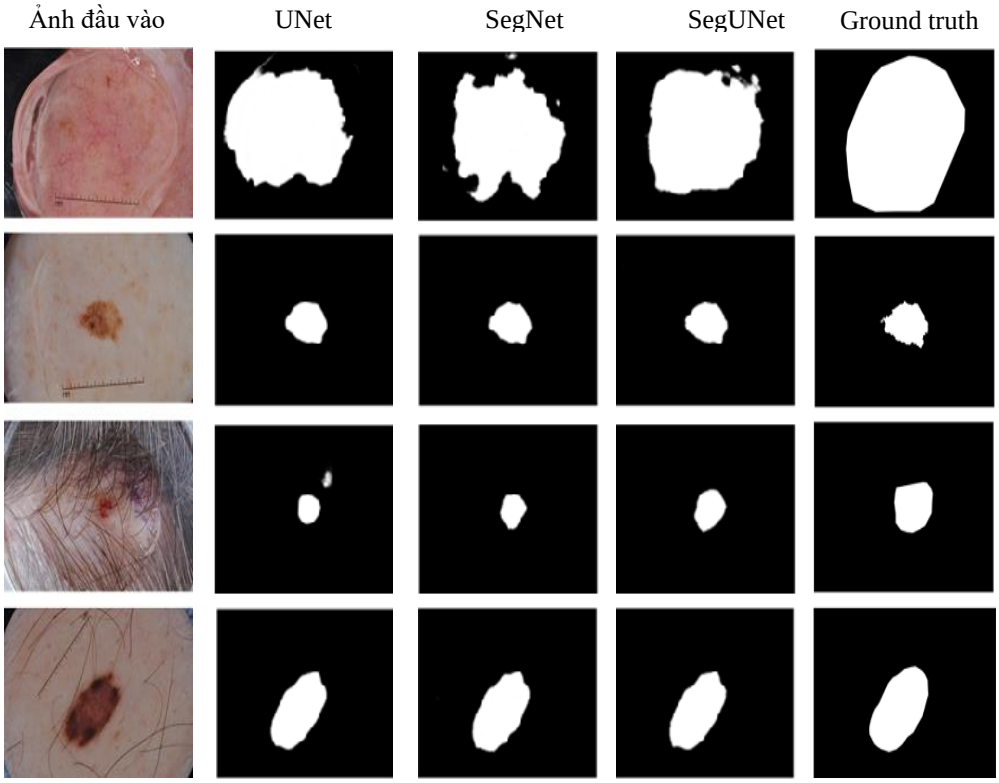
Như đồ thị ở Hình 4 có thể thấy, sự biến đổi của Loss, Accuracy, và Dice Coefficient trên tập đánh giá khá trơn, không có nhiều nhấp nhô, mô hình đạt đến hội tụ khá sớm, cũng như sự biến đổi của các thành phần này trên cả hai tập khá tương tự nhau, cho thấy mô hình được học khá hiệu quả, và sự nhấp nhô ít khi mô hình đã đạt hội tụ. Thông qua kết quả của bộ trọng số sau mỗi epoch, ta sẽ lựa chọn ra được những bộ trọng số có kết quả cao nhất để đánh giá với tập kiểm tra. Đối với mô hình SegUNet, kết quả trên tập kiểm tra lần lượt là: Loss – 0,1491, Accuracy – 0,9305, Dice Coefficient – 0,8508, Jaccard Index – 0,7649.

4.2.2. So sánh với SegNet và UNet

Mô hình SegUNet là sự kết hợp giữa SegNet và UNet, nhằm tận dụng được lợi thế của cả hai mô hình, tạo nên một mô hình mới tốt hơn. Và thực tế khi huấn luyện mô hình UNet và SegNet, với cùng kích thước ảnh vào, gia tăng dữ liệu giống nhau và thuật toán học học như nhau, ta thu được kết quả tốt hơn SegNet và UNet.

Để minh họa kết quả phân vùng, nhóm tác giả đưa ra một số kết quả phân vùng đại diện như trên Hình 5. Trên hình 5, ảnh đầu vào được đặt ở cột đầu tiên, kế đến là kết quả dự đoán lần lượt bởi các mô hình: UNet, SegNet, SegUNet.

Cột cuối cùng là kết quả phân vùng chuẩn được thực hiện bởi các bác sỹ và chuyên gia (thường được gọi là ground truth) đã cung cấp cùng với tập dữ liệu ISIC 2017. Từ kết quả này trên Hình 5 ta thấy, trong các kết quả dự đoán, kết quả do mô hình SegUNet thu đươc gần với ground truth nhất.



Hình 5. *Một số hình ảnh phân vùng ảnh bởi UNet, SegNet, SegUNet, so với phân vùng chuẩn (ground truth)*

Để so sánh một cách định lượng hiệu quả của các mô hình cho cả tập dữ liệu thử nghiệm của tập ISIC 2017, chúng tôi cung cấp các chỉ số kết quả trung bình trong Bảng 1. Trong đó có các thông tin kết quả gồm Accuracy, Dice hệ số Dice, và chỉ số Jaccard. Từ bảng này ta thấy mô hình SegUNet cho kết quả cao hơn so với UNet và SegNet. Về mặt thời gian huấn luyện, với ảnh kích thước 256x256, UNet mất 172,3 phút, trong khi đó mô hình SegNet và SegUNet huấn luyện trong thời gian lâu hơn, tương ứng với 210,3 và 216,4 phút, như trong Bảng 1.

Bảng 1. *Kết quả của các mô hình UNet, SegNet và SegUnet trên tập kiểm tra của ISIC 2017*

Phương pháp	Time (minutes)	Dice Coefficient	Jaccard Coefficient	Accuracy
UNet	172,3	0,812	0,716	0,917
SegNet	210,3	0,837	0,748	0,927
SegUNet (proposed)	216,4	**0,851**	**0,765**	**0,931**

4.2.3. So sánh với một số nghiên cứu trước

Để biểu thị hiệu quả của phương pháp đề xuất, trong phần này nhóm tác giả đưa ra bảng so sánh với kết quả của một số phương pháp trước đó. Các nghiên cứu này cùng được huấn luyện trên tập Training và cùng thử nghiệm trên tập test của tập cơ sở dữ liệu ISIC 2017. Các kết quả của các nghiên cứu này đưa ra bao gồm hệ số Dice và chỉ số Jaccard. Các phương pháp so sánh bên cạnh UNet và SegNet còn bao gồm nghiên cứu của Bi và cộng sự [4], Xue và cộng sự [21], và Yuan và cộng sự [22]. Thêm vào đó, mô hình đề xuất cũng được so sánh với các mô hình Unet+ [23], Unet++ [24], U-SegNet [25], Res-SegNet [25]. Kết quả so sánh được biểu diễn trên Bảng 2.

Từ Bảng 2 ta thấy, mô hình đề xuất SegUNet cho kết quả phân vùng tốt hơn so với một số nghiên cứu trước đó.

Bảng 2. *Kết quả so sánh với một số nghiên cứu trên tập kiểm tra của ISIC 2017*

Phương pháp	Dice Coefficient	Jaccard Coefficient
Bi và cộng sự	0,834	0,731
Xue và cộng sự	0,839	0,749
Yuan và cộng sự	0,849	**0,765**
UNet	0,812	0,716
SegNet	0,837	0,748
UNet+	0,833	0,743
UNet++	0,829	0,738
U-SegNet	0,844	0,759
Res-SegNet	0.841	0,755
SegUNet (proposed)	**0,851**	**0,765**

5. Kết luận

Nghiên cứu này đề xuất phương án phân vùng ảnh ở vùng tổn thương da sử dụng kỹ thuật học sâu. Dựa trên nghiên cứu và kiểm chứng các phương pháp khác nhau, bài báo đã đề xuất mô hình SegUNet là sự kết hợp hai mô hình UNet và SegNet để cho kết quả tốt hơn, đồng thời đề xuất thay thế phép toán chuẩn hóa trung bình - phương sai thay cho chuẩn hóa theo mẻ để giảm số tham số của mô hình. Trên cơ sở đó, nghiên cứu đã xây dựng được mô hình mạng học sâu cho bài toán phân vùng tổn thương da. Mô hình đã được thử nghiệm trên tập cơ sở dữ liệu ISIC 2017 và đạt độ chính xác cao hơn so với một số mô hình trước đó. Mô hình đề xuất có thể được ứng dụng trong các ứng dụng phân vùng ảnh khác, cùng trên ý tưởng trích xuất, phát hiện vị trí, hình dạng vật thể.

TÀI LIỆU THAM KHẢO

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer Statistics, 2020", *CA CANCER J CLIN* vol. 70, pp. 7–30, 2020.

[2] R. Siegel, K. Miller, and A. Jemal, "Cancer statistics, 2018", *CA Cancer J. Clin.*, vol. 68, pp. 7–30, 2018.

[3] L. Bi, J. Kim, E. Ahn, A. Kumar, M. Fulham, and D. Feng, "Dermoscopic image segmentation via multi-stage fully convolutional networks", *IEEE Trans. Biomed. Eng.*, vol. 64, pp. 2065-2074, 2017.

[4] L. Bi, J. Kim, E. Ahn, and D. Feng, "Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks", *in arXiv:1703.04197*, Available: https://arxiv.org/abs/1703.04197, 2017.

[5] C. Cernazanu-Glavan and S. Holban, "Segmentation of bone structure in X-ray images using convolutional neural network", *Adv. Electr. Comput. Eng.*, vol. 13, pp. 87-94, 2013.

[6] M. Melinšˇcak, O. Prentaši´c, and S. Lonˇ cari´c, "Retinal vessel segmentation using deep neural networks", in *Proc. 10th International Conference on Computer Vision Theory and Applications*, 2015, pp. 577–582.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation", *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, 2015.

[8] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation", *arXiv: 1704.06857*, 2017.

[9] P. V. Tran, "A fully convolutional neural network for cardiac segmentation in short-axis MRI", *Available: https://arxiv.org/abs/1604.00494*, 2016.

[10] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation", in *Proc. of the IEEE International Conference on Computer Vision*, 2015, pp. 1520-1528.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234-241.

[12] V. Badrinarayanan, A. Kendall, and R. Cipolla, "E: A deep convolutional encoder-decoder architecture for image segmentation", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, pp. 2481–2495, 2017.

[13] L. Yu, H. Chen, Q. Dou, J. Qin, and P. A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks", *IEEE Trans. Med. Imaging*, vol. 36, pp. 994-1004, 2017.

[14] Y. Yuan, M. Chao, and Lo, Y. C., "Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance", *IEEE Trans. Med. Imaging* vol. 36, pp. 1876-1886, 2017.

[15] N. Ibtehaz and M. S. Rahman, "Multiresunet: Rethinking the U-Net architecture for multimodalbiomedical image segmentation", *Available: https://arxiv.org/abs/1902.04049*, 2019.

[16] Y. Tang, F. Yang, S. Yuan, and C. A. Zhan, "A multi-stage framework with context information fusion structure for skin lesion segmentation", in *Proc. IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 1407–1410.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", pp. arXiv preprint arXiv:1409.1556, 2014.

[18] K. He, X. Zhang, S. Ren, and S. Sun, "Deep residual learning for image recognition", *[Online]. Available: https://arXiv:1512.03385*, 2015.

[19] G. Huang, Z. Liu, and K. Q. Weinberger, "Densely connected convolutional networks", *arXiv:1608.06993*, 2016.

[20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, *et al.*, "Going deeper with convolutions", in *Proc. IEEE Conf. Comput. Vis. Pattern Recognition*, 2015, pp. 1-9.

[21] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, "SegAN: Adversarial network with multi-scale L 1 loss for medical image segmentation", *Neuroinformatic*, vol. 16, pp. 383-392, 2018.

[22] Y. Yuan and Y.-C. Lo, "Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks", *IEEE J. Biomed. Health Inform.*, vol. 23, pp. 519-526, 2019.

[23] S. M. K. Hasan and C. A. Linte, "U-NetPlus: A modified encoder-decoder U-Net architecture for semantic and instance segmentation of surgical instruments from laparoscopic images", *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, pp. 7205-7211, Jul. 2019.

[24] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation", *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856-1867, Jun. 2020.

[25] D. Daimary, M.B. Bora, K. Amitab, D. Kandar, "Brain tumor segmentation from MRI images using hybrid convolutional neural networks", *Procedia Comput. Sci.*, vol. 167, pp. 2419-2428, 2020.