

# PHÂN VÙNG HÌNH ẢNH NỘI SOI POLYP TRỰC TRẮNG - THỰC TRẠNG VÀ MỘT SỐ HƯỚNG PHÁT TRIỂN

Nguyễn Hoàng Long<sup>1</sup>, Nguyễn Đình Công<sup>2</sup>, Lê Thị Hồng Hà<sup>2</sup>

## TÓM TẮT

*Phân vùng hình ảnh y tế liên quan đến việc trích xuất các vùng được quan tâm trong hình ảnh. Với sự phát triển của công nghệ trí tuệ nhân tạo, các phương pháp phân vùng hình ảnh y tế đã đạt được những kết quả đáng mong đợi. Trong bài báo này, chúng tôi cung cấp một cái nhìn tổng thể về các phương pháp phân vùng hình ảnh nội soi polyp. Tiếp theo đó, các phương pháp này sẽ được đánh giá trên các tập dữ liệu tiêu chuẩn. Cuối cùng, chúng tôi thảo luận về những thách thức và các hướng phát triển trong tương lai của chủ đề nghiên cứu.*

**Từ khóa:** CNN, transformer, polyp, phân vùng hình ảnh.

**DOI:** <https://doi.org/10.70117/hdujs.2.2024.743>

## 1. ĐẶT VẤN ĐỀ

Phân vùng hình ảnh là một trong những hướng nghiên cứu chính của lĩnh vực thị giác máy tính. Mục tiêu của phân vùng là chia hình ảnh thành các nhóm con có nhãn thuộc về một lớp tương ứng. Phân vùng hình ảnh được ứng dụng trong nhiều lĩnh vực như phân tích hình ảnh y học, xe tự hành,... Trong lĩnh vực y tế, phân vùng hình ảnh y tế đóng vai trò quan trọng trong việc chẩn đoán lâm sàng và lập kế hoạch điều trị. Phân vùng hình ảnh y tế nhằm xác định cơ quan hoặc các mô tổn thương từ các hình ảnh y tế như nội soi, CT (Computed Tomography) hoặc MRI (Magnetic Resonance Imaging). Polyp là một dạng tổn thương bên trong đại trực tràng, có hình dạng khối u, có cuống hoặc không. Polyp ban đầu là các khối u lành tính, tuy nhiên nếu không phát hiện và loại bỏ kịp thời có thể dẫn đến ung thư đại trực tràng, gây nguy hiểm tới tính mạng. Phân vùng polyp là một vấn đề trong phân vùng hình ảnh và đã thu hút được sự chú ý từ cộng đồng nghiên cứu trong thời gian gần đây. Việc phát hiện và đánh giá sớm các polyp đóng vai trò quan trọng trong quá trình chẩn đoán và điều trị sớm ung thư đại trực tràng. Các phương pháp phân vùng hình ảnh nội soi polyp cung cấp một giải pháp hiệu quả nhằm hỗ trợ bác sĩ lâm sàng trong việc xác định và khoanh vùng chính xác các vùng polyp. Tuy nhiên, để phân vùng hình ảnh nội soi polyp phát huy vai trò trong thực tế là một nhiệm vụ đầy thách thức với hai lý do chính. Thứ nhất, các polyp có hình dạng khác nhau về kích thước, màu sắc và kết cấu. Thứ hai, trong hình ảnh nội soi, đường biên giữa polyp và các vùng xung quanh thường có độ tương

<sup>1</sup> Trung tâm CNTT&TT, Trường Đại học Hồng Đức; Email:nguyenhoanglong@hdu.edu.vn

<sup>2</sup> Khoa Công nghệ thông tin và Truyền thông, Trường Đại học Hồng Đức

phản thấp hoặc bị mờ, vì vậy gây khó khăn cho các phương pháp phân vùng, dẫn tới hiện tượng phân vùng sai hoặc phân vùng thiếu các vùng polyp.

Trong bài viết này, chúng tôi trình bày một nghiên cứu khảo sát về các phương pháp phân vùng polyp dựa trên các mạng học sâu. Ngoài ra, chúng tôi tiến hành đánh giá hiệu năng của các mô hình vùng đoạn polyp đại diện khác nhau và phân tích lợi thế tương ứng của chúng. Hơn nữa, chúng tôi sẽ thảo luận về tương lai những thách thức và hướng nghiên cứu tiềm năng.

Phần còn lại của bài báo được trình bày như sau: Phần 2 trình bày một số nghiên cứu liên quan trong bài toán phân vùng polyp. Tiếp theo, phần 3 mô tả các tập dữ liệu và các phương pháp đánh giá. Sau đó, phần 4 thảo luận về kết quả thực nghiệm. Cuối cùng, phần 5 kết luận chủ đề và đưa ra các định hướng tương lai.

## 2. TỔNG QUAN VỀ TÌNH HÌNH NGHIÊN CỨU

Những phương pháp truyền thống về phân vùng hình ảnh polyp chủ yếu dựa trên các đặc trưng cấp thấp như hình dạng, màu sắc và kết cấu [1] [2]. Tuy nhiên, các đặc trưng cấp thấp được trích chọn một cách thủ công với đặc điểm độ đa dạng về dữ liệu, độ tương đồng giữa vùng có polyp và vùng niêm mạc xung quanh gây nhiều khó khăn cho việc phân vùng. Bên cạnh đó, các phương pháp truyền thống có hiệu quả thấp, tính tổng quát hoá không cao. Do đó, dẫn đến việc phát hiện sai hoặc phát hiện thiếu những vùng polyp.

Với sự phát triển của công nghệ trí tuệ nhân tạo, các mô hình học sâu đã được áp dụng vào các bài toán phân vùng hình ảnh y tế, đặc biệt là các mô hình CNN. Unet [3] là một kiến trúc CNN tiên phong trong bài toán xử lý hình ảnh y tế. Unet bao gồm hai thành phần chính: bộ mã hoá và bộ giải mã. Bộ mã hoá có nhiệm vụ trích chọn đặc trưng đa cấp và đa tỉ lệ. Sau khi đi hết bộ mã hoá, những đặc trưng này sẽ được chuyển trực tiếp sang bộ giải mã để tiếp tục trích chọn ở mức cao hơn, đồng thời độ phân giải qua mỗi tầng trong bộ giải mã sẽ tăng dần đến khi kích thước thu được bằng với kích thước ảnh gốc ban đầu của nhãn. Ngoài ra, các đặc trưng đã được trích chọn tại bộ mã hoá sẽ được đưa trực tiếp sang bộ giải mã tương ứng thông qua nhánh tắt. Hai phần mã hoá và giải mã của kiến trúc U-net tương ứng nhau, có dạng chữ U. Unet đã mang tới một bước đột phá mới trong xử lý hình ảnh y học, tuy vậy, Unet có hạn chế khi gặp các đối tượng có kích thước nhỏ. Để giải quyết vấn đề này, các biến thể của Unet như Unet++ [4], DoubleUnet [5], ResUnet++ [6] đã được đề xuất và cho kết quả tốt hơn so với ý tưởng ban đầu.

Mô hình PraNet [7] được đề xuất gồm ba phần: mã hoá, giải mã và tinh chỉnh. Phần mã hoá sử dụng Res2Net dùng để trích chọn đặc trưng đa cấp và đa tỉ lệ. Các đặc trưng thu được sau đó được chuyển qua bộ giải mã song song nhằm kết hợp các thông tin ngữ cảnh và thu được bản đồ đặc trưng toàn cục. Bên cạnh đó, đặc trưng từ phần mã hoá sẽ được đưa vào khối tinh chỉnh (khối chú ý) nhằm thiết lập mối quan hệ giữa các khu vực polyp và đường biên của chúng. Mặc dù đạt được những kết quả vượt trội trong bài toán phân vùng polyp, PraNet vẫn có những hạn chế với các vùng polyp có kích thước nhỏ. CaraNet [8] đã khắc phục những hạn chế của PraNet bằng cách đề xuất một khối tinh chỉnh mới có

tên gọi Axial Reverse Attention - ARA. Bên cạnh đó, mô hình CaraNet còn sử dụng mô-đun trích xuất đặc trưng kim tự tháp theo kênh (CFP) nhằm giúp mô hình học được nhiều chi tiết hơn, lấy được nhiều đặc trưng từ các tỉ lệ khác nhau. Kết quả của phân giải mã thu được là một bản đồ toàn cục có thể xác định vị trí gần đúng của polyp. Các tác giả đã sử dụng bản đồ toàn cục kết hợp với bản đồ đặc trưng thu được tại cấp thứ  $i$  sau khi đi qua mô-đun trích xuất đặc trưng kim tự tháp theo kênh để làm đầu vào của khối ARA. Mô-đun ARA được dựa trên cơ chế tự chú ý, ánh xạ một truy vấn và các cặp khoá giá trị tới một kết quả đầu ra. Bên cạnh đó, các tác giả đã đề xuất chuẩn hoá cơ chế chú ý 2D sang hai cơ chế chú ý theo chiều cao và chiều rộng của dữ liệu đầu vào và thay đổi hàm kích hoạt từ Softmax sang Sigmoid nhằm giảm bớt tài nguyên tính toán.

Với sự thành công của Transformer trong lĩnh vực xử lý ngôn ngữ tự nhiên, các mô hình xử lý hình ảnh dựa trên Transformer đã được áp dụng vào bài toán phân vùng hình ảnh y tế một cách mạnh mẽ nhằm khắc phục những hạn chế về nắm bắt thông tin toàn cục của CNN. Polyp-PVT [9] đã được Dong và các cộng sự đề xuất dựa trên ý tưởng sử dụng một mạng Transformer kim tự tháp kết hợp với các mô-đun hợp nhất xếp hàng (CFM), mô-đun nhận dạng nguy trang (CIM) và mô-đun tổng hợp tương đương (SAM). Các đặc trưng được trích xuất từ mạng cơ sở có thể khai thác các thông tin toàn cục kết hợp với mô-đun CFM nhằm thu thập các thông tin ngữ nghĩa và sử dụng các đặc trưng cấp cao để xác định các vị trí của polyp. Mô-đun CIM được áp dụng để loại bỏ nhiễu và tăng cường biểu diễn các đặc trưng cấp thấp. Mô-đun SAM giúp tổng hợp các đặc trưng thu được từ hai mô-đun CFM và CIM. Ngoài những phương pháp sử dụng thuần CNN hoặc các mô hình dựa trên Transformer, các tác giả [10] đã đề xuất mô hình Transfuse sử dụng song song kết hợp CNN và Transformer nhằm thu thập các thông tin cục bộ và các thông tin toàn cục để nâng cao đáng kể hiệu suất phân vùng.

### 3. THỰC NGHIỆM VÀ ĐÁNH GIÁ

#### 3.1. Tập dữ liệu

Trong phần này, chúng tôi nghiên cứu các bộ dữ liệu tiêu chuẩn trong bài toán phân vùng polyp trong hình ảnh nội soi gồm: Kvasir [11], CVC-ClinicDB [12], CVC-ColonDB [13], ETIS-Larib [14]. Chi tiết về các bộ dữ liệu sẽ được mô tả chi tiết tại Bảng 1 dưới đây:

**Bảng 1. Chi tiết về các bộ dữ liệu tiêu chuẩn trong bài toán phân vùng polyp.**

Bộ dữ liệu	Số lượng ảnh	Kích thước
CVC-ClinicDB [12]	612	576x768
CVC-ColonDB [13]	380	500x572
ETIS-Larib [14]	196	1226x996
Kvasir [11]	1000	720x576 đến 1920x1072

#### 3.2. Phương pháp đánh giá

Trong bài báo này, chúng tôi sẽ sử dụng hai độ đo phổ biến trong bài toán phân vùng hình ảnh là độ đo Jaccard (IoU) và hệ số Dice. Một số thông số phổ biến như dương tính đúng (True Positive - TP), dương tính sai (False Positive - FP), âm tính đúng (True Negative - TN) và âm tính sai (False Negative - FN) cũng sẽ được sử dụng.

*Độ đo Jaccard:* Độ đo Jaccard, hay còn gọi là độ đo IoU, được sử dụng để đo sự trùng lặp của kết quả dự đoán và nhãn thực tế. Nó thường được sử dụng trong việc xử lý dữ liệu mất cân bằng.

$$IoU = \frac{TP}{TP + FP + FN}$$

*Hệ số Dice:* Hệ số Dice là một phương pháp đánh giá thường được sử dụng để đo lường sự giống nhau giữa kết quả dự đoán và nhãn thực tế. Hệ số Dice có giá trị trong khoảng từ 0 đến 1, trong đó 1 biểu thị kết quả dự đoán và nhãn thực tế hoàn toàn trùng nhau, và 0 biểu thị không có kết quả nào phù hợp.

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN}$$

### 3.3. Kết quả và thảo luận

#### 3.3.1. Một số kết quả

Trong bài báo này, chúng tôi thực hiện đánh giá hiệu năng của các mô hình Unet [3], Unet++ [4], PraNet [7], CaraNet [8], Polyp-PVT [9] và TransFuse [10]. Các mô hình được huấn luyện từ đầu và đánh giá sử dụng framework Pytorch, quá trình huấn luyện các mô hình sử dụng Google Colab với các thông số phần cứng như sau: GPU-NVIDIA Tesla T4; Bộ nhớ trong 16GB; Số phép tính dấu phẩy động/giây 8.1 TFLOPS. Các mô hình được huấn luyện với 20 epochs và kích thước mỗi lô (batch) là 8. Các hình ảnh đầu vào đã được chuẩn hoá về kích thước 352x352.

Để đánh giá hiệu năng của các mô hình, chúng tôi sử dụng dữ liệu huấn luyện là sự kết hợp giữa hai tập dữ liệu Kvasir và CVC-ClinicDB được chia theo tỉ lệ 80% số lượng ảnh mỗi tập dữ liệu dành cho tập huấn luyện, 10% cho tập dữ liệu đánh giá và 10% kết hợp với ba bộ dữ liệu là CVC-ColonDB, ETIS-Larib, CVC-T dành cho quá trình kiểm thử. Kết quả thực nghiệm so sánh các mô hình đã được nghiên cứu cho bài toán phân vùng polyp được mô tả chi tiết trong Bảng 2.

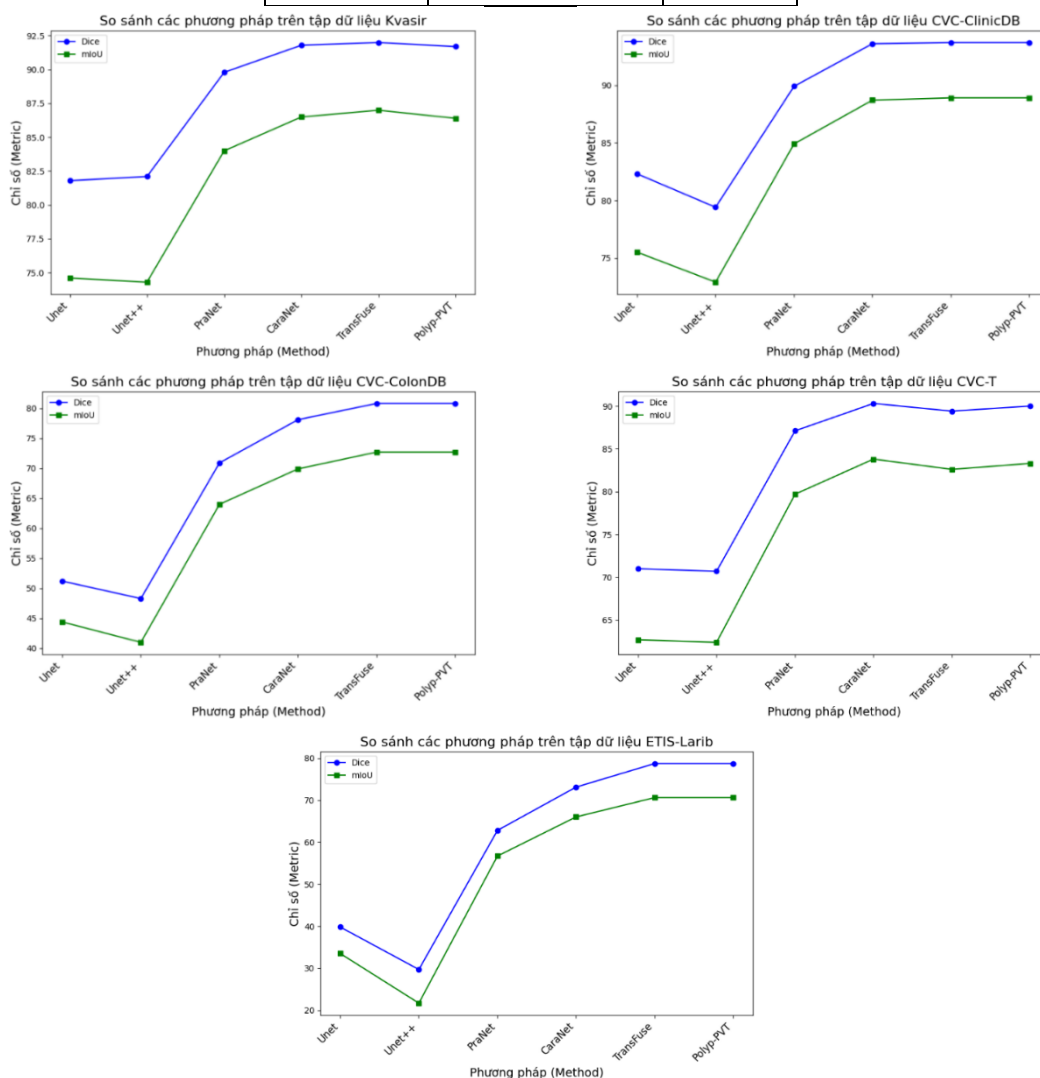
**Bảng 1. Bảng so sánh kết quả của các mô hình trên 5 bộ dữ liệu.**

Phương pháp	Kvasir		ClinicDB		ColonDB		CVC-T		ETIS	
	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU
Unet	81.8	74.6	82.3	75.5	51.2	44.4	71.0	62.7	39.8	33.5
Unet++	82.1	74.3	79.4	72.9	48.3	41.0	70.7	62.4	29.7	21.7
PraNet	89.8	84.0	89.9	84.9	70.9	64.0	87.1	79.7	62.8	56.7
CaraNet	91.8	86.5	<u>93.6</u>	<u>88.7</u>	78.1	69.9	90.3	83.8	73.1	66.0

TransFuse	92.0	87.0	93.7	88.9	80.8	72.7	<u>89.4</u>	<u>82.6</u>	78.7	70.6
Polyp-PVT	91.7	86.4	93.7	88.9	80.8	72.7	<u>90.0</u>	<u>83.3</u>	78.7	70.6

**Bảng 2. Bảng so sánh số lượng tham số và GFLOPs của các mô hình.**

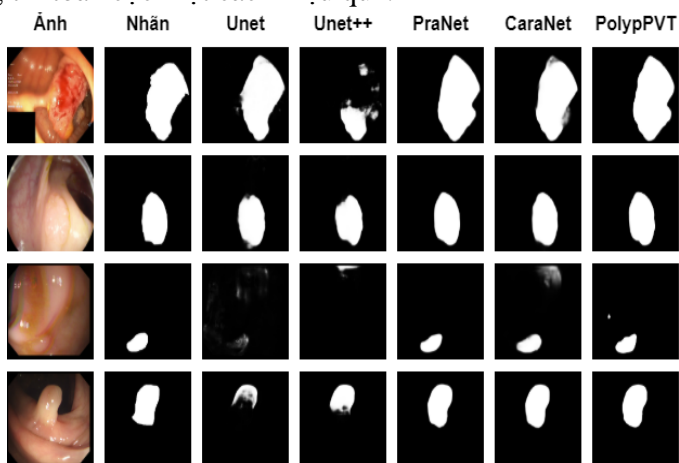
Mô hình	Số lượng tham số	GFLOPs
Unet	34.52	123.87
Unet++	36.63	262.16
PraNet	32.55	13.11
CaraNet	46.64	21.69
Polyp-PVT	25.11	10.02



**Hình 1. Trục quan kết quả so sánh các mô hình trên các tập dữ liệu kiểm tra**

3.3.2. Thảo luận

Bảng 2 trình bày kết quả của các mô hình đã nghiên cứu trên 5 tập dữ liệu kiểm tra. Các mô hình dựa được chia thành hai nhóm: các mô hình dựa trên CNN (Unet, Unet++, PraNet, CaraNet) và các mô hình dựa trên kiến trúc Transformer (TransFuse, Polyp-PVT). CaraNet đạt kết quả cao nhất trong số các mô hình dựa trên CNN với 2%-33.3% ở độ đo mDice và 1.5%-32.5% ở độ đo mIoU so với các mô hình dựa trên CNN khác. Ta có thể thấy độ chính xác của các mô hình dựa trên kiến trúc Transformer cao hơn đa số các tập dữ liệu. Với mô hình Polyp-PVT và TransFuse cao hơn mô hình đứng thứ hai là CaraNet với 2.7% ở độ đo mDice và 2.8% ở độ đo mIoU trên tập dữ liệu CVC-ColonDB. Mô hình CaraNet đạt độ chính xác trên tập CVC-T cao hơn mô hình đứng thứ hai 0.3% và 0.5% lần lượt trên hai độ đo mDice và mIoU. Ngoài ra, các mô hình dựa trên kiến trúc Transformer cho hiệu năng tốt hơn do cấu trúc mạng trích chọn đặc trưng dựa trên cơ chế tự chú ý có thể nắm bắt thông tin toàn cục một cách hiệu quả.



**Hình 2. Một số kết quả phân vùng.**

Hình 2 trình bày kết quả phân vùng của các mô hình đã nghiên cứu trên các tập dữ liệu kiểm thử. Mặc dù các mô hình đã đạt được những kết quả đáng mong đợi, vẫn còn rất nhiều thách thức cho các nhà nghiên cứu để nâng cao hiệu năng của các mô hình. Một trong số những nguyên nhân dẫn đến điều này đến từ việc hình dạng và kích thước của các polyp phong phú và các đường biên giữa polyp và niêm mạc xung quanh không rõ ràng. Ở hàng 1, đa số các mô hình đã phân vùng được vùng lớn hơn nhân thực do độ tương phản thấp giữa vùng chứa polyp và vùng niêm mạc xung quanh, điều này gây nhầm lẫn cho các mô hình. Ở hàng 3 và hàng 4, các mô hình Unet và Unet++ không thể phân vùng được do vùng polyp bị đổ bóng và là nơi giao thừa giữa các vùng sáng tối. Ngoài ra, ở hàng 3, mô hình CaraNet cũng tồn tại nhiều do đổ bóng, làm cho mô hình nhận nhầm đây là vùng polyp. Mô hình Polyp-PVT cho kết quả ít nhiễu hơn so với các phương pháp còn lại, tuy nhiên, hình dạng của polyp cũng chưa được chính xác so với nhân thực.

Bên cạnh đó, một số khó khăn khác trong bài toán phân vùng polyp có thể kể đến như hạn chế về dữ liệu, không có bộ dữ liệu chuyên biệt với lượng dữ liệu lớn và dữ liệu được thu thập từ nhiều loại thiết bị nội soi khác nhau. Vì vậy, một hướng nghiên cứu mới

đã được đề xuất bởi [15] sử dụng một mạng thích ứng cho nhiều bộ dữ liệu polyp khác nhau nhằm làm tăng hiệu quả của mô hình trên các bộ dữ liệu mục tiêu khác nhau.

Ngoài ra, Bảng 3 đã trình bày về số lượng tham số và GFLOPs của các mô hình học sâu dựa trên mạng nơ-ron tích chập và kiến trúc Transformer đã tìm hiểu. Có thể thấy GLOPs của mô hình Unet và Unet++ cao hơn nhiều so với các mô hình khác với lần lượt là 123.87 GLOPs và 262.16 GFOPs, mô hình Polyp-PVT dựa trên kiến trúc Transformer có số lượng tham số và GLOPs với 25.11 triệu tham số và 10.02 GFLOPs. Các thông số này cho thấy các mạng học sâu đòi hỏi một độ phức tạp tính toán lớn, khiến chúng khó có thể áp dụng trong các ứng dụng thực tiễn đòi hỏi thời gian thực. Do đó, hướng nghiên cứu phát triển các mô hình học sâu có kiến trúc và độ phức tạp nhỏ mà không cần đánh đổi hiệu suất của mô hình là một vấn đề mở cho các nhà nghiên cứu, giúp làm tăng tính ứng dụng thực tế của các mô hình học sâu trong bài toán phân vùng polyp trong ảnh nội soi.

#### 4. KẾT LUẬN

Các mô hình đã nghiên cứu đã đạt được những kết quả đáng mong đợi với mô hình TransFuse và Polyp-PVT đã đạt 93.7% ở độ đo mDice và 88.9% ở độ đo mIoU. Tuy nhiên, vẫn còn những hạn chế và những tiềm năng có thể tối ưu trong những mô hình đã tìm hiểu. Trong các nghiên cứu kể trên, phép toán nội suy song tuyến tính được sử dụng để tăng kích thước các bản đồ đặc trưng, điều này có thể gây mất mát một số thông tin hữu ích và dẫn đến việc đường biên có thể không chính xác. Điều này có thể được tối ưu bằng cách sử dụng các khối nội suy hai khối hoặc sử dụng các khối phân tách tích chập (deconvolution) để tăng kích thước của các bản đồ đặc trưng. Để cải tiến khả năng phân vùng, chúng tôi dự kiến sẽ nghiên cứu một số kiến trúc mạng cơ sở nhằm trích chọn các đặc trưng, một số phương pháp làm mịn nhằm tăng cường khả năng nhận biết các vùng biên. Bên cạnh đó, thiết kế mô hình phân vùng chính xác với một số lượng tham số và GFLOPs của mô hình nhỏ cũng là một hướng nghiên cứu mà nhóm tác giả quan tâm.

#### TÀI LIỆU THAM KHẢO

- [1] Jianhua Yao and Miller, M. and Franaszek, M. and Summers, R.M. (2004), *Colonic polyp segmentation in CT colonography-based on fuzzy clustering and deformable models*, IEEE Transactions on Medical Imaging, vol.23, 1344-1352.
- [2] M. Ganz, X. Yang and G. Slabaugh (2012), *Automatic Segmentation of Polyps in Colonoscopic Narrow-Band Imaging Data*, IEEE Transactions on Biomedical Engineering, vol.59, 2144-2151.
- [3] O Ronneberger, P Fischer, T Brox (2015), *U-Net: Convolutional Networks for Biomedical Image Segmentation*, in Medical Image Computing and Computer-Assisted Intervention.
- [4] Zongwei Zhou and Md Mahfuzur Rahman Siddiquee and Nima Tajbakhsh and Jianming Liang (2018), *UNet++: A Nested U-Net Architecture for Medical Image Segmentation*, Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : 4th International Workshop, DLMIA 2018, and 8th

- International Workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, S..., vol.11045, 3-11.
- [5] Debesh Jha and Michael A. Riegler and Dag Johansen and Pål Halvorsen and Håvard D. Johansen (2020), *DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation*, 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS).
- [6] D Jha, PH Smedsrud, MA Riegler, D Johansen, T De Lange, P Halvorsen, HD Johansen (2019), *Resunet++: An advanced architecture for medical image segmentation*, IEEE international symposium on multimedia (ISM).
- [7] Fan, Deng-Ping; Ji, Ge-Peng; Zhou, Tao; Chen, Geng; Fu, Huazhu; Shen, Jianbing Shao, Ling (2020), *PraNet: Parallel Reverse Attention Network for Polyp Segmentation*, in MICCAI.
- [8] Lou, Ange; Guan, Shuyue; Loew, Murray (2023), *Caranet: Context axial reverse attention network for segmentation of small medical objects*, Journal of Medical Imaging.
- [9] B Dong, W Wang, DP Fan, J Li, H Fu, L Shao (2021), *Polyp-pvt: Polyp segmentation with pyramid vision transformers*, arXiv.
- [10] Y Zhang, H Liu, Q Hu (2021), *Transfuse: Fusing transformers and cnns for medical image segmentation*, in Medical Image Computing and Computer Assisted Intervention, 2021.
- [11] Debesh Jha, Pia H. Smedsrud, Michael A. Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, Håvard D. Johansen (2020), *Kvasir-SEG: A Segmented Polyp Dataset*, in International Conference on Multimedia Modeling.
- [12] Bernal J, Sánchez FJ, Fernández-Esparrach G, Gil D, Rodríguez C, Vilariño F (2015), *WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians*, Comput Med Imaging Graph.
- [13] Tajbakhsh N, Gurudu SR, Liang J. (2016), *Automated Polyp Detection in Colonoscopy Videos Using Shape and Context Information*, IEEE Trans Med Imaging.
- [14] Juan Silva, Aymeric Histace & Olivier Romain (2014), *Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer*, International Journal of Computer Assisted Radiology and Surgery , p.283-293.
- [15] C. Yang, X. Guo, M. Zhu, B. Ibragimov, Y. Yuan (2021), *Mutual-Prototype Adaptation for Cross-Domain Polyp Segmentation*, IEEE Journal of Biomedical and Health Informatics, vol.25, 3886-3897.



- [16] A Vaswani, N Shazeer, N Parmar, J Uszkoreit, L Jones, AN Gomez, Ł Kaiser, I Polosukhin (2017), *Attention is all you need*, in Advances in neural information processing systems.

## **POLYP SEGMENTATION - FACTS AND FUTURE DEVELOPMENT**

**Nguyen Hoang Long, Nguyen Dinh Cong, Le Thi Hong Ha**

### **ABSTRACT**

*Medical image segmentation involves extracting regions of interest in the image. With the development of artificial intelligence technology, medical image segmentation methods have achieved significant results. In this article, we provide an overview of polyp segmentation methods. Next, these methods will be evaluated on standard datasets. Finally, we discuss the challenges and future development directions of the research topic.*

**Keywords:** *CNN, transformer, polyp, medical image.*

\* Ngày nộp bài: 12/3/2024; Ngày gửi phản biện: 17/3/2024; Ngày duyệt đăng: 15/11/2024