

# PHÂN ĐOẠN ẢNH POLYP DỰA TRÊN CẢI TIẾN PHƯƠNG PHÁP RECURRENT RESIDUAL U-NET

Võ Thị Hồng Tuyết, Nguyễn Thanh Bình

Khoa Công Nghệ Thông Tin, Trường Đại học Ngoại ngữ-Tin học TP.HCM

tuyetvth@huflit.edu.vn, binh@huflit.edu.vn

**TÓM TẮT**— Polyp là dạng khối u ở đại tràng mà người bệnh có thể gặp phải ở mọi giới tính. Việc xác định được hình dạng và kích thước của polyp trong hình ảnh y khoa sẽ hỗ trợ rất nhiều cho quá trình chẩn đoán và điều trị lâm sàng. Từ đó hỗ trợ các bác sĩ phát hiện kịp thời và có phác đồ điều trị phù hợp nhằm tăng khả năng cứu chữa thành công cho người bệnh. Bài toán phân đoạn được thực hiện nhằm hỗ trợ quá trình xác định này. Tuy nhiên, việc lựa chọn bản đồ đặc trưng cho phân đoạn là vấn đề luôn cần cải tiến và tối ưu trong lĩnh vực thị giác máy tính. Bài báo này trình bày phương pháp Recurrent Residual U-Net (R2U-Net) được hiệu chỉnh hệ thống mã hóa-giải mã để phân đoạn ảnh polyp. Quá trình cải tiến được thực hiện bằng cách thay thế các tầng Convolution bằng các tầng Recurrent Convolution và áp dụng phần dư khối trong mỗi khối của nó. Phương pháp đề xuất được thử nghiệm trên tập dữ liệu Kvasir-SEG và EndoTect 2020. Kết quả đánh giá được so sánh với các phương pháp gần đây bằng chỉ số Jaccard Index.

**Từ khóa**— Ảnh polyp, phân đoạn, Recurrent Residual U-Net, U-Net.

## I. GIỚI THIỆU

Ung thư đại trực tràng là một loại ung thư ảnh hưởng đến đại tràng (ruột già) hoặc trực tràng. Đây là một trong những loại ung thư phổ biến nhất trên thế giới. Nó có thể gây tổn hại nghiêm trọng và tử vong. Nguy cơ ung thư đại trực tràng tăng theo tuổi tác. Hầu hết các trường hợp ảnh hưởng đến những người trên 50 tuổi. Vào năm 2020, ước tính có hơn 1,9 triệu ca ung thư đại trực tràng mới và hơn 930.000 ca tử vong do ung thư đại trực tràng trên toàn cầu. Đến năm 2040, gánh nặng ung thư đại trực tràng được dự đoán sẽ tăng lên 3,2 triệu ca mắc mới mỗi năm (tăng 63%) và 1,6 triệu ca tử vong mỗi năm (tăng 73%) [1]. Chẩn đoán kịp thời, điều trị thích hợp và chăm sóc theo dõi thường xuyên là yếu tố quan trọng để cải thiện tỷ lệ sống sót và chất lượng cuộc sống. Phân đoạn ảnh polyp đại tràng giúp ích cho quá trình chẩn đoán và điều trị ung thư đại trực tràng. Đây là công việc đối mặt nhiều khó khăn và thách thức:

Đầu tiên, hình dạng polyp không đều. Polyp có thể được phân loại theo các loại hình dạng khác nhau: không cuống, có cuống hoặc phẳng và sự biến đổi về hình dạng lớn.

Thứ hai, kích thước polyp rất khác nhau. Kích thước polyp dao động từ 5 mm đến hơn 50 mm.

Thứ ba, các khu vực xung quanh rất phức tạp. Đại tràng có thể bị căng quá mức. Có thể có phân và chất lỏng còn sót lại ở khu vực xung quanh. Các nếp gấp Haustral và các cấu trúc khác có thể rất khó phân đoạn.

Ứng dụng khoa học công nghệ vào lĩnh vực khoa học sức khỏe được đẩy mạnh trong thời gian qua, nhất là ứng dụng công nghệ xử lý thông minh và trí tuệ nhân tạo trong chẩn đoán và điều trị bệnh. Cụ thể, các khâu ứng dụng bao gồm: dự đoán, sàng lọc, phân tích, giải mã dữ liệu hình ảnh y khoa có chứa bất thường. Bài toán phân vùng ảnh (image segmentation) là bài toán phổ biến trong thị giác máy tính. Việc lựa chọn khu vực đối tượng trước đây phụ thuộc vào hình thái học (morphology) sẽ dễ dàng bị ảnh hưởng bởi các yếu tố cường độ xám từ hình ảnh y tế. Theo các nghiên cứu gần đây, công việc phân đoạn dựa trên các chiết xuất đặc trưng sẽ khắc phục các giới hạn của hình thái học. Các mạng học sâu được áp dụng để thực hiện vai trò này, nhưng khi áp dụng lại phải vận dụng các ngữ cảnh và tham số phù hợp để thu kết quả khả quan.

Bài báo này đề xuất một phương pháp phân đoạn ảnh polyp dựa trên Recurrent Residual U-Net đã được hiệu chỉnh. Phương pháp đề xuất tập trung vào việc nâng cao kết quả chiết xuất đặc trưng của mô hình học sâu nhằm phục vụ cho khả năng phân vùng ảnh polyp tốt hơn để phục vụ cho các hệ thống nhận dạng và hỗ trợ chẩn đoán tự động. Bố cục của bài báo bao gồm: giới thiệu tổng quan bài báo, được trình bày ở phần 1; phần 2 trình bày các nghiên cứu liên quan cũng như ưu và nhược điểm của chúng; phần 3 trình bày phương pháp đề xuất để phân đoạn polyp; phần 4 trình bày thí nghiệm và đánh giá kết quả; phần cuối cùng là kết luận cũng như hướng mở rộng tương lai.

## II. CÁC NGHIÊN CỨU LIÊN QUAN

Các thuật toán phân đoạn polyp nội soi cũng đã được nghiên cứu rộng rãi trong những năm gần đây, với nhiều mô hình khác nhau được đề xuất để phân đoạn polyp chính xác và hiệu quả. Nhóm tác giả Hong [2] đã sử dụng dữ liệu hình ảnh nội soi là dữ liệu cho mô hình học U-Net thực hiện quá trình mã hóa (encoder) và giải mã (decoder) nhằm quyết định vùng lựa chọn. Sự cải tiến framework dựa vào ResNet do nhóm tác giả Afify [3] đề xuất nhằm giảm thiểu số block thực thi mã hóa của mô hình học sâu. Trong [4], ResUNet++ được kết hợp với điều kiện cho tham số ngẫu nhiên đã giúp kết quả phân đoạn giảm thiểu mất mát thông tin về hình dạng của đối tượng. Ngoài sự kết hợp này, atrous convolution cũng được dùng để hỗ trợ lựa chọn đặc trưng tối ưu cho U-Net mang tên A-DenseUNet [5]. Mạng ngữ nghĩa (semantic network) cũng tham gia vào quá trình chiết xuất đặc trưng của ảnh khối u polyp [7] cũng như

tiến trình phát hiện và nhận dạng [6]. Vào năm 2022, mô hình phân đoạn hỗ trợ phát hiện ung thư đại trực tràng đã được Akilandeswari [8] đề xuất dựa trên mạng nơron tích chập và cải thiện số lượng lớp tuyến tính cũng như lưới biến đổi. Điều rút ra được từ các công trình trên chính là các nhóm nghiên cứu tìm cách kết hợp hoặc vận dụng điều kiện theo ngữ cảnh để tăng cường đặc trưng tốt cho quá trình học dữ liệu.

ResUNet++ [9] bổ sung một số khối như khối bóp và kích thích, nhóm kim tự tháp không gian atrous (atrous spatial pyramid pooling- ASPP), khối attention block và khối residual block vào khung U-Net hiện có. Tuy nhiên, việc thu thập thông tin chi tiết khó khăn khi độ sâu của mạng lưới thần kinh tăng lên do kích thước bản đồ đặc trưng giảm. Để giải quyết vấn đề này, ASPP được sử dụng trong ResUNet++ để duy trì thông tin chi tiết và cho phép dự đoán chính xác theo đơn vị pixel. Để cải thiện hơn nữa hiệu suất của mô hình, Jha và cộng sự [9] đã thêm trường ngẫu nhiên có điều kiện (conditional random field -CRF) và tăng thời gian thử nghiệm (test-time augmentation) vào ResUNet++.

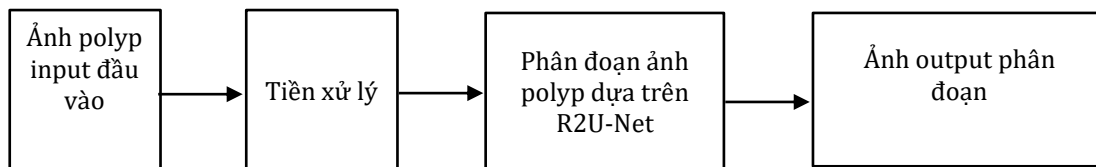
Để giải quyết vấn đề các cạnh phân đoạn bị mờ, Zhao và cộng sự [10] đã đề xuất mạng trừ đa tỷ lệ (Multi-Scale Subtraction Network), kết hợp trừ đơn vị để trích xuất các đặc trưng khác biệt giữa các mức liên kề trong bộ mã hóa. Mạng này chỉ định các trường tiếp nhận khác nhau cho các cấp độ khác nhau của các đơn vị này theo mô hình kim tự tháp, cho phép trích xuất đặc trưng khác biệt đa dạng. Wei và cộng sự [11] đã đề xuất mạng Shallow Attention, tận dụng các đặc trưng cấp thấp để giảm thiểu trùng lặp do lấy mẫu nhiều lần. Họ cũng đề xuất một phương pháp đổi màu (color-swapping) để giảm sự phụ thuộc vào màu sắc bằng cách trao đổi dữ liệu thống kê màu sắc. Md Mostafijur và cộng sự [12] đã đề xuất bộ giải mã chú ý theo tầng (Cascade Attention Decoder) để giải quyết hiệu quả vấn đề kích thước đặc trưng không nhất quán bằng chuyển đổi cấu trúc phân cấp (hierarchical structure of transformers) và mô-đun tích chập để tổng hợp các đặc trưng đa cấp toàn cục và cục bộ.

Để xử lý các biến thể về kích thước, Tao và cộng sự [13] đã đề xuất mạng tổng hợp tính năng chéo cấp (Cross-Level Feature Aggregation Network), sử dụng mạng phân đoạn dựa trên cấu trúc luồng kép và chiến lược hợp nhất từng lớp để xử lý hiệu quả các biến thể quy mô và tích hợp thông tin ngữ nghĩa cấp cao với các tính năng cấp thấp.

Các nghiên cứu trên gặt hái nhiều kết quả tốt. Tuy nhiên, hiện tại kết quả nghiên cứu còn nhược điểm là độ phức tạp của phương pháp đề xuất cao và độ chính xác chưa cao do sự tác động của các yếu tố màu sắc, ngữ cảnh và ánh sáng.

### III. PHƯƠNG PHÁP ĐỀ XUẤT PHÂN ĐOẠN ẢNH POLYP

Như đã trình bày ở phần trên, hầu hết khác biệt giữa các đặc trưng bất thường của ảnh polyp là nhỏ. Vì vậy, chúng tôi chọn phương pháp phân đoạn phù hợp với sự khác biệt này. Phương pháp phân đoạn trong trường hợp này phải đảm bảo trích xuất đặc trưng chính xác, tránh mất thông tin trong ảnh. Để phân đoạn ảnh polyp, phương pháp đề xuất được trình bày như hình 1 dưới đây bao gồm hai giai đoạn như: tiền xử lý, phân đoạn dùng Recurrent Residual U-Net (R2U-Net).



Hình 1. Phương pháp đề xuất phân đoạn ảnh polyp

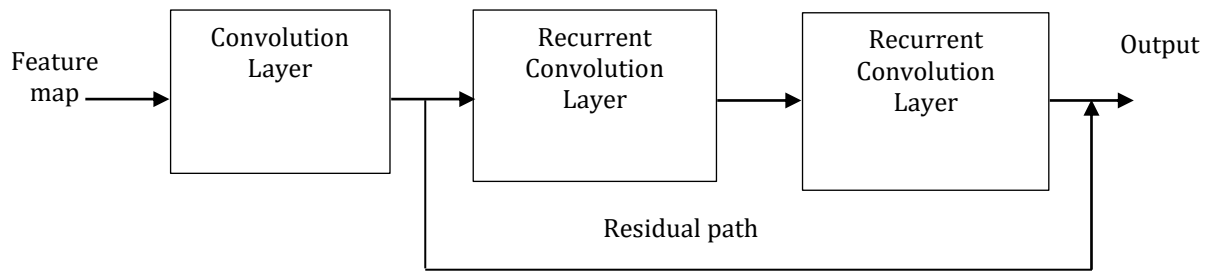
#### A. TIỀN XỬ LÝ

Ảnh polyp có thể có kích thước khác nhau và có thể có một số thành phần bị nhiễu bởi các yếu tố tương đồng với khu vực xung quanh, ảnh hưởng đến kết quả của bất kỳ mô hình phân đoạn nào. Trong quá trình xử lý trước ảnh, bước đầu tiên là thay đổi kích thước ảnh. Vì mỗi hình ảnh trong tập dữ liệu có kích thước khác nhau nên hình ảnh được thay đổi kích thước thành  $416 \times 416$  pixel. Kích thước này được chọn sau khi thực hiện phân tích thử nghiệm; hình ảnh có kích thước cao hơn sẽ tăng thời gian thực hiện cũng như bộ nhớ cần thiết.

#### B. PHÂN ĐOẠN ẢNH POLYP DỰA TRÊN R2U-NET

Mô hình R2U-Net được cải tiến từ mô hình U-Net. Quá trình cải tiến được thực hiện bằng cách thay thế các tầng Convolution bằng các tầng Recurrent Convolution và áp dụng phần dư khối trong mỗi khối của nó. Mô hình R2U-Net có hai phần chính: bộ mã hóa (encoder) và bộ giải mã (decoder). Nhiệm vụ của bộ mã hóa trong mô hình R2U-Net là trích xuất các đặc trưng của ảnh và bộ giải mã là khôi phục ảnh về kích thước ban đầu. Khôi phục ảnh được thực hiện bằng cách nối các bản đồ đặc trưng (feature maps) từ bộ mã hóa đến bộ giải mã. Cả bộ mã hóa và bộ giải mã đều được xây dựng từ khối Recurrent Residual Convolution (RRC) [14].

Khối RRC được tạo thành từ hai tầng Identical Recurrent Convolution giống hệt nhau và một đường dẫn để kết hợp các kết quả thu được khi thực hiện các phép tính trên hai tầng Recurrent Convolution và đầu vào của chính khối. Hình 2 bên dưới minh họa cấu trúc của khối RRC.



Hình 2. Cấu trúc của khối RRC

Trong tầng Recurrent Convolutions có chứa tầng Convolution. Vì vậy, kích thước và số lượng bộ lọc của tầng Recurrent Convolutional là kích thước và số lượng bộ lọc của tầng Convolution bên trong nó. Kiến trúc của mô hình R2U-Net được xây dựng như sau:

- Giả sử ảnh polyp có kích thước  $512 \times 512 \times 1$  là ảnh đầu vào của mô hình R2U-Net. Ảnh đi qua khối RRC có kích thước  $3 \times 3$ , 32 bộ lọc và tạo ra một bản đồ đối tượng có kích thước  $512 \times 512 \times 32$ . Các đặc trưng này được chuyển qua tầng Pooling có kích thước  $2 \times 2$  để tạo ra bản đồ đặc trưng mới có kích thước  $256 \times 256 \times 32$ . Chúng tôi tiếp tục sử dụng thêm 3 bộ nữa với mỗi bộ gồm 1 khối RRC kích thước  $3 \times 3$  và 1 Pooling layer kích thước  $2 \times 2$  với số bộ lọc của khối RRC lần lượt là 64, 128, 256. Đầu ra của một bản đồ đặc trưng là  $32 \times 32 \times 256$ . Bản đồ đặc trưng này đi qua khối RRC có kích thước  $3 \times 3$  và số bộ lọc là 512 để có được bản đồ đặc trưng mới có kích thước  $32 \times 32 \times 512$ . Sau đó, bản đồ đặc trưng này được thay đổi kích thước ảnh.
- Khi được thực thi trên bộ giải mã đầu tiên, bản đồ đặc trưng trên được chuyển qua một tầng tích chập chuyển vị (transpose convolution) có kích thước  $2 \times 2$  và 256 bộ lọc. Bằng cách nối bản đồ đặc trưng này với bản đồ đặc trưng từ bộ mã hóa đối xứng (symmetric encoder), chúng tôi nhận được bản đồ đặc trưng mới có kích thước  $64 \times 64 \times 512$ . Tiếp tục sử dụng khối RRC  $3 \times 3$  với 256 bộ lọc, bản đồ đặc trưng  $64 \times 64 \times 256$  được tạo có cùng kích thước với bản đồ đặc trưng trong bộ mã hóa đối diện (opposite encoder). Chúng tôi tiếp tục thực hiện công việc này với ba bộ giải mã: thông qua tầng Transpose Convolution, kết hợp các đặc trưng trong bộ mã hóa đối xứng (symmetric encoder) và một khối Recurrent Residual Convolution  $3 \times 3$  [16].
- Tầng Convolution Transpose và RRC có cùng số lượng bộ lọc, lần lượt là 128, 64 và 32. Vào cuối quá trình này, chúng tôi nhận được bản đồ đặc trưng có kích thước  $512 \times 512 \times 32$ . Ảnh được phân đoạn là kết quả của bản đồ đặc trưng vừa thu được khi đi qua tầng Convolution  $1 \times 1$  với số bộ lọc là 1. Và ảnh này có kích thước  $512 \times 512 \times 1$ . Tất cả các tầng Convolution được theo sau bởi tầng Batch normalization và hàm kích hoạt ReLU (ReLU activation function), chỉ tầng Convolution cuối cùng sử dụng hàm Sigmoid khi hàm này được kích hoạt.

#### IV. THÍ NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

Các chương trình thử nghiệm của chúng tôi được phát triển bằng ngôn ngữ Python trên Google Colab. Để đánh giá phương pháp đề xuất, bộ dữ liệu Kvasir-SEG [17] và EndoTect 2020 [18] đã được sử dụng. Bộ dữ liệu Kvasir-SEG chứa 1000 hình ảnh polyp và ground-truth tương ứng. Kích thước của Kvasir-SEG thay đổi từ  $332 \times 487$  đến  $1920 \times 1072$  pixel với JPEG và JSON cho định dạng mẫu. Bộ dữ liệu EndoTect 2020 chứa 10662 hình ảnh polyp với định dạng JPEG và kích thước từ  $720 \times 576$  đến  $768 \times 576$ . Chúng tôi đã sử dụng 1000 hình ảnh polyp của Kvasir-SEG và 1000 hình ảnh polyp của EndoTect 2020 với 70% cho huấn luyện và 30% cho thử nghiệm. Một số ảnh trong các tập dữ liệu trên được minh họa như hình 3 bên dưới.

Để đánh giá kết quả phân đoạn một cách định lượng, chúng tôi đã tính toán chỉ số Jaccard Index (JI) giữa vùng được trích xuất và vùng thực, được xác định thủ công bởi chuyên gia y tế [19].

$$JI(A, B) = \frac{|A \cap B|}{|A \cup B|} * 100 \quad (1)$$

trong đó, A là vùng được phân đoạn, B là vùng phân đoạn thực (ground-truth). Nếu A và B đều trống, chúng ta xác định  $JI(A, B) = 100$ . Chỉ số JI nằm trong khoảng từ 0 đến 100%. Chỉ số JI càng cao, kết quả phân vùng càng chính xác hơn. Trong trường hợp A là vùng được phân đoạn toàn ảnh, B là vùng phân đoạn thực, chỉ số Jaccard thu được là chỉ số Jaccard của toàn bộ mọi vùng được phân đoạn của quá trình phân đoạn.



a. Bộ dữ liệu Kvasir-SEG [17]



b. Bộ dữ liệu EndoTect 2020 [18]

Hình 3. Một số ảnh trong tập dữ liệu Kvasir-SEG [17] và EndoTect 2020 [18]



a. Ảnh gốc



JI = 96,75%

JI = 97,04%

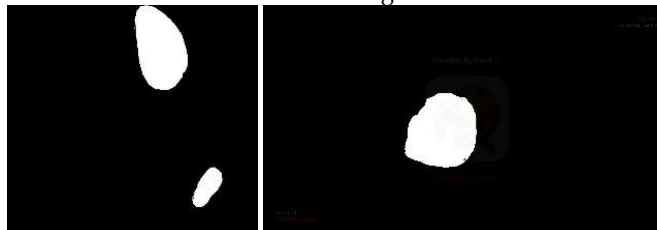
JI = 96,88%

b. Kết quả phân đoạn

Hình 4. Một số ảnh kết quả sau phân đoạn trong tập dữ liệu Kvasir-SEG



a. Ảnh gốc



JI = 97,79%

JI = 98,11 %

b. Kết quả phân đoạn

Hình 5. Một số ảnh kết quả sau phân đoạn trong tập dữ liệu EndoTect 2020

Chúng tôi thực nghiệm phân đoạn ảnh trong các tập dữ liệu Kvasir-SEG và EndoTect 2020 nêu trên bằng phương pháp đề xuất và sử dụng thang đo JI để đánh giá kết quả sau khi phân đoạn. Một số ví dụ kết quả điển hình được trình bày ở hình 4 và hình 5. Phương pháp đề xuất được thực nghiệm trên 2 tập dữ liệu trên. Kết quả thử nghiệm giá trị JI trung bình của phương pháp đề xuất trên tập dữ liệu ảnh tương ứng như bảng 1.

Bảng 1. Giá trị JI trung bình của phương pháp đề xuất trên tập dữ liệu ảnh tương ứng

Tập dữ liệu	Chỉ số JI (%) trung bình của phương pháp đề xuất
Tập dữ liệu Kvasir-SEG	97,52
Tập dữ liệu EndoTect 2020	98,05

Bảng 1 trình bày giá trị JI trung bình của phương pháp đề xuất trên tập dữ liệu Kvasir-SEG là 97,52 %, trên tập dữ liệu EndoTect 2020 là 98,05. Cả hai quá trình thực nghiệm đều đảm bảo tỉ lệ huấn luyện và thử nghiệm như đầu phần III đã nêu rõ.

Để đánh giá khách quan kết quả của phương pháp cải tiến, chúng tôi so sánh kết quả phương pháp cải tiến với các mô hình máy học khác nhau như: U-Net [20], SegU-Net [21]. Bảng 2 trình bày giá trị JI trên tập dữ liệu Kvasir-SEG của phương pháp đề xuất, phương pháp U-Net và phương pháp SegU-Net lần lượt là 97,52; 94,31 và 96,02, trong khi độ chính xác trên tập dữ liệu EndoTect 2020 lần lượt là 98,05; 95,53 và 93,25.

Bảng 2. Giá trị JI giữa các mô hình máy học thí nghiệm

Phương pháp/ tập dữ liệu	Chỉ số JI (%)		
	Phương pháp U-NET [20]	Phương pháp SegU-Net [21]	Phương pháp đề xuất
Tập dữ liệu Kvasir-SEG	94,31	96,02	97,52
Tập dữ liệu EndoTect 2020	93,25	95,53	98,05

Từ bảng 1 và bảng 2, chúng tôi thấy rằng, phương pháp đề xuất cho kết quả tốt và tốt hơn phương pháp phương pháp U-Net và phương pháp SegU-Net trên tập dữ liệu thử nghiệm Kvasir-SEG và EndoTect 2020.

Như đã trình bày ở trên, kiến trúc R2U-Net được cải thiện bằng cách thay thế các tầng Convolution bằng các tầng Recurrent Convolution và áp dụng phần dư khối trong mỗi khối của nó. Vì vậy, chúng tôi thu được nhiều đặc trưng hơn để cải thiện việc phân đoạn.

## V. KẾT LUẬN VÀ HƯỚNG NGHIÊN CỨU MỞ RỘNG

Trong y khoa, các khối u có thể âm thầm phát triển và gây ảnh hưởng rất lớn đến sức khỏe và tính mạng của bệnh nhân. Các khối u trên cơ thể có thể lành tính nhưng qua thời gian vì nhiều nguyên nhân hoàn toàn có thể chuyển hóa sang ác tính và trở thành ung thư. Polyp là dạng khối u ở đại tràng diễn ra với mọi giới tính. Việc xác định được hình dạng và kích thước của polyp trong hình ảnh y khoa sẽ hỗ trợ rất nhiều cho chẩn đoán và điều trị lâm sàng. Từ đó hỗ trợ các y bác sĩ phát hiện kịp thời và có phác đồ điều trị phù hợp để cứu chữa cho người bệnh.

Bài báo này đề xuất một phương pháp phân đoạn ảnh polyp dựa trên Recurrent residual U-Net đã được điều chỉnh. Kết quả của phương pháp đề xuất tốt hơn so sánh kết quả của các phương pháp gần đây. Để tăng độ chính xác của phương pháp được đề xuất, trong tương lai sẽ cải thiện kiến trúc học sâu và thử nghiệm trên các bộ dữ liệu khác. Kết quả phân đoạn ảnh polyp giúp xác định hình dạng, vị trí khối u này trên các bộ phận cơ thể. Điều này hỗ trợ cho các chuyên gia y tế chẩn đoán và điều trị lâm sàng; mặt khác còn là cơ sở để hỗ trợ phân loại bệnh của các hệ thống chẩn đoán tự động.

## VI. LỜI CẢM ƠN

Nghiên cứu được tài trợ bởi Trường Đại học Ngoại ngữ-Tin học Thành phố Hồ Chí Minh trong khuôn khổ Đề tài mã số H2023-03.

## VII. TÀI LIỆU THAM KHẢO

- [1] World Health Organization (2023), *World Health Organization*, Colorectal cancer, [https://www.who.int/news-room/fact-sheets/detail/colorectal-cancer?gad\\_source=1&gclid=EAIaIQobChMIyt-mlKibhQMVrtYWBR2jawi9EAAYASAAEgJkPfd\\_BwE](https://www.who.int/news-room/fact-sheets/detail/colorectal-cancer?gad_source=1&gclid=EAIaIQobChMIyt-mlKibhQMVrtYWBR2jawi9EAAYASAAEgJkPfd_BwE) (lần truy cập cuối 26/03/2024)
- [2] Le Thi Thu Hong, Nguyen Chi Thanh and Tran Quoc Long (2020), Polyp segmentation in colonoscopy images using ensembles of u-nets with efficientnet and asymmetric similarity loss function, *in 2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*, IEEE, pp. 1–6.
- [3] H. M. Afify, K. K. Mohammed and A. E. Hassanien (2021), An improved framework for polyp image segmentation based on SegNet architecture, *Int. Journal Imaging System Technology*, Vol. 31, No. 3, pp. 1741-1751.

- [4] D. Jha, P. H. Smedsrud, D. Johansen, T. de Lange, H. D. Johansen, P. Halvorsen and M. A. Riegler (2021), A comprehensive study on colorectal polyp segmentation with ResUNet++, conditional random field and test-time augmentation, *IEEE J. Biomed. Health Information*, Vol. 25, No. 6, pp. 2029-2040.
- [5] S. Safarov and T. K. Whangbo (2021), A-DenseUNet: Adaptive densely connected UNet for polyp segmentation in colonoscopy images with atrous convolution, *Sensors*, Vol. 21, No. 4: pp. 1-16.
- [6] Viraj Jiwane, Anubhav DattaGupta, Arunkumar Chauhan and Vidya Patil (2021), Detecting diabetic retinopathy using deep learning technique with Resnet-50, *ICDSMLA 2020, Lecture notes in Electrical Engineering, Springer*, Vol. 783, pp. 45 – 55.
- [7] Wenchao Zhang, Chong Fu, Yu Zheng, Fangyuan Zhang, Yanli Zhao, Chiu-Wing Sham (2022), HSNet: A hybrid semantic network for polyp segmentation, *Computers in Biology and Medicine*, vol. 150.
- [8] A. Akilandeswari, D. Sungeetha, Christeena Joseph, K. Thaiyalnayaki, K. Baskaran, R. Jothi Ramalingam, Hamad Al-Lohedan, Dhaifallah M. Al-dhayan, Muthusamy Karnan and Kibrom Meansbo Hadish (2022), Automatic Detection and Segmentation of Colorectal Cancer with Deep Residual Convolutional Neural Network, *Evidence-Based Complementary and Alternative Medicine, Hindawi*, Vol. 2022.
- [9] Jha, D., Smedsrud, P.H., Riegler, D., Lange, T.D., Halvorsen, P., Johansen, H.D (2019), "ResUNet++: An Advanced Architecture for Medical Image Segmentation", *In Proceedings of the 2019 IEEE International Symposium on Multimedia (ISM)*, pp. 9-11.
- [10] X. Zhao, L. Zhang, and H. Lu (2021), Automatic polyp segmentation via multi-scale subtraction network, *in MICCAI. Springer*, pp. 120-130.
- [11] J. Wei, Y. Hu, R. Zhang, Z. Li, S. K. Zhou, and S. Cui (2021), Shallow attention network for polyp segmentation, *in MICCAI. Springer*, pp. 699-708.
- [12] Md Mostafijur Rahman, Radu Marculescu (2023), Medical image segmentation via cascaded attention decoding", *2023 IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 6222- 6231.
- [13] Tao Zhou, Yi Zhou, Kelei He, Chen Gong, Jian Yang, Huazhu Fu and Dinggang Shen (2023), Cross-level Feature Aggregation Network for Polyp Segmentation, *Pattern Recognition*, Vol. 140, No. 109555, <https://doi.org/10.1016/j.patcog.2023.109555>.
- [14] Jiaxin Mei, Tao Zhou, Kaiwen Huang, Yizhe Zhang, Yi Zhou, Ye Wu and Huazhu Fu (2024), A Survey on Deep Learning for Polyp Segmentation: Techniques, Challenges and Future Trends, <https://arxiv.org/abs/2311.18373>.
- [15] Jha, D., Smedsrud, P.H., Johansen, D., de Lange, T., Johansen, H.D., Halvorsen, P. and Riegler (2021), M.A. A comprehensive study on colorectal polyp segmentation with resunet++, conditional random field and test-time augmentation, *IEEE J. Biomed. Health Inform*, Vol. 25, pp. 2029-2040.
- [16] Mubashar, M., Ali, H., Grönlund and Shoaib Azmat (2022), R2U++: a multiscale recurrent residual U-Net with dense skip connections for medical image segmentation, *Neural Comput & Applic* 34, pp. 17723-17739, DOI:10.1007/s00521-022-07419-7.
- [17] Jha, Debesh and Smedsrud, Pia H and Riegler, Michael A and Halvorsen, P.I and de Lange, Thomas and Johansen, Dag and Johansen, Havad D (2020), *Kvasir SEG*, Kvasir SEG segmented dataset for computer aided gastrointestinal disease detection, <https://datasets.simula.no/kvasir-seg/> (lần truy cập cuối 12/04/2024)
- [18] EndoTect Challenge (2020), *Endotect polyp segmentation challenge dataset*, The EndoTect 2020 Challenge: Evaluation and Comparison of Classification, Segmentation and Inference Time for Endoscopy, <https://endotect.com/> (lần truy cập cuối 10/04/2024)
- [19] Tran Duc Tam and Nguyen Thanh Binh, Efficient Pancreas Segmentation in Computed Tomography Based Region-Growing (2014), *Proceedings of the International Conference on Nature of Computation and Communication, Springer*, Vol. 144:332-340, DOI: 10.1007/978-3-319-15392-6\_31.
- [20] George Batchkala and Sharib Ali (2020), *CEUR Workshop proceedings*, Real-Time Polyp Segmentation Using U-Net with IoU Loss Multimedia Benchmark Workshop 2020, <https://ceur-ws.org/Vol-2882/paper30.pdf> (lần truy cập cuối 16/04/2024)
- [21] Dinthisrang Daimary, Mayur Bhargab Bora, Khwairakpam Amitab and Debdatta Kandar (2020), Brain tumor segmentation from MRI Images using Hybrid convolutional neural networks", *Procedia computer science, Elsevier*, Vol. 167, pp. 2419 – 2428.

## POLYP IMAGE SEGMENTATION BASED ON THE RECURRENT RESIDUAL U-NET IMPROVEMENT METHOD

Vo Thi Hong Tuyet, Nguyen Thanh Binh

**ABSTRACT**— Polyps are tumors in the colon that can occur in people of all genders. Determining the shape and size of polyps in medical images will greatly support the clinical diagnosis and treatment process. It helps doctors detect promptly and have appropriate treatment regimens to increase the likelihood of successfully treating patients. The segmentation problem is

implemented to support this determination process. However, choosing feature maps for segmentation is an issue that always needs improvement and optimization in the field of computer vision. This paper presents the Recurrent Residual U-Net (R2U-Net) method with a modified encoder-decoder system to segment polyp images. The improvement is done by replacing Convolution layers with Recurrent Convolution layers and applying block residuals in each of its blocks. The proposed method is tested on the Kvasir-SEG and EndoTect 2020 datasets. The evaluation results are compared with recent methods using the Jaccard Index.

**Keywords**— Polyp image, Segmentation, Recurrent Residual U-Net, U-Net.



**ThS. Võ Thị Hồng Tuyết** (<https://orcid.org/0000-0002-9153-2883>) tốt nghiệp Cử nhân ngành Sư phạm Tin học tại Trường Đại học Sư phạm TP.HCM vào năm 2011 và Thạc sĩ Kỹ thuật chuyên ngành Khoa học máy tính tại Trường Đại học Bách Khoa - ĐH Quốc Gia TP.HCM vào năm 2015. Từ 2019 đến nay, cô là nghiên cứu sinh ngành Khoa học máy tính tại

Khoa Khoa học và Kỹ thuật máy tính của Trường Đại học Bách Khoa - Đại học Quốc Gia TPHCM. Hiện cô là giảng viên bộ môn Khoa học máy tính, Khoa Công nghệ thông tin, trường Đại học Ngoại ngữ-Tin học TPHCM. Lĩnh vực nghiên cứu chính của cô liên quan đến các vấn đề về thị giác máy tính, xử lý ảnh và nhận dạng.



**PGS.TS Nguyễn Thanh Bình** (<https://orcid.org/0000-0003-0136-874X>) nhận bằng kỹ sư tại Trường Đại học Bách khoa-ĐHQG TP.HCM năm 2000, và nhận bằng Thạc sĩ và Tiến sĩ ngành khoa học máy tính từ Đại học Allahabad - Ấn Độ vào năm 2005 và 2011. Hiện nay, ông là Phó Giáo sư ngành Công nghệ

thông tin tại Trường Đại học Ngoại ngữ-Tin học TP.HCM. Lĩnh vực nghiên cứu: nhận dạng, xử lý ảnh, hệ thống thông tin đa phương tiện, hệ thống hỗ trợ quyết định và dữ liệu chuỗi thời gian.

