**PHÂN ĐOẠN ẢNH MRI NÃO DỰA TRÊN MÔ HÌNH SEMANTIC SEGMENTATION**

TRƯỜNG ĐẠI HỌC NGOẠI NGỮ - TIN HỌC TP. HỒ CHÍ MINH

KHOA CÔNG NGHỆ THÔNG TIN

**TP. HỒ CHÍ MINH – THÁNG 08 NĂM 2024**

GIẢNG VIÊN HƯỚNG DẪN: PGS. TS. Nguyễn Thanh Bình

SINH VIÊN THỰC HIỆN: Lê Đức Phú - 21DH114491

**KHÓA LUẬN TỐT NGHIỆP**

**PHÂN ĐOẠN ẢNH MRI NÃO DỰA TRÊN MÔ HÌNH SEMANTIC SEGMENTATION**

TRƯỜNG ĐẠI HỌC NGOẠI NGỮ - TIN HỌC TP. HỒ CHÍ MINH

KHOA CÔNG NGHỆ THÔNG TIN

**TP. HỒ CHÍ MINH – THÁNG 08 NĂM 2024**

GIẢNG VIÊN HƯỚNG DẪN: PGS. TS. Nguyễn Thanh Bình

SINH VIÊN THỰC HIỆN: Lê Đức Phú - 21DH114491

**KHÓA LUẬN TỐT NGHIỆP**

**LỜI CẢM ƠN**

Chúng tôi xin gửi lời cảm ơn sâu sắc đến PGS. TS. Nguyễn Thanh Bình, người đã tận tình hướng dẫn và hỗ trợ chúng tôi trong suốt quá trình thực hiện khóa luận này.

Chúng tôi cũng xin cảm ơn Khoa Công Nghệ Thông Tin, Trường Đại học Ngoại Ngữ - Tin Học TP. Hồ Chí Minh đã tạo điều kiện thuận lợi và cung cấp các tài liệu cần thiết cho quá trình nghiên cứu.

Cuối cùng, chúng tôi xin cảm ơn các đồng nghiệp và những người bạn đã cùng làm việc và chia sẻ kinh nghiệm trong suốt thời gian thực hiện khóa luận.

Xin chân thành cảm ơn.

**LỜI CAM ĐOAN**

Chúng tôi xin cam đoan rằng toàn bộ nội dung của luận văn này là kết quả của quá trình nghiên cứu và làm việc nghiêm túc của chúng tôi. Các tài liệu, dữ liệu và kết quả nghiên cứu được sử dụng trong luận văn đều được trích dẫn đầy đủ và rõ ràng theo đúng quy định.

Chúng tôi cam kết rằng luận văn này không sao chép bất kỳ phần nào từ các công trình nghiên cứu khác mà không có sự cho phép hoặc không được trích dẫn nguồn gốc rõ ràng. Nếu phát hiện bất kỳ sai sót hoặc vi phạm nào trong quá trình nghiên cứu và trình bày, chúng tôi xin hoàn toàn chịu trách nhiệm.

Chúng tôi cũng cam đoan rằng toàn bộ những gì được trình bày trong luận văn này đều là những kết quả và đóng góp của chúng tôi, không vi phạm các quy định về bản quyền và đạo đức nghiên cứu khoa học.

Xin trân trọng cảm ơn.

MỤC LỤC

[MỤC LỤC i](#_Toc175183672)

[DANH MỤC CÁC KÝ HIỆU, CHỮ VIẾT TẮT VÀ THUẬT NGỮ iii](#_Toc175183673)

[DANH MỤC CÁC BẢNG iv](#_Toc175183674)

[DANH MỤC CÁC HÌNH VẼ, ĐỒ THỊ v](#_Toc175183675)

[CHƯƠNG 1: GIỚI THIỆU 1](#_Toc175183676)

[1.1. Giới thiệu đề tài 1](#_Toc175183677)

[1.2. Mục tiêu và nội dung của đề tài 1](#_Toc175183678)

[1.3. Giới hạn đề tài 2](#_Toc175183679)

[1.4. Phương pháp nghiên cứu 2](#_Toc175183680)

[1.5. Đóng góp của luận văn 3](#_Toc175183681)

[1.6. Cấu trúc luận văn 3](#_Toc175183682)

[CHƯƠNG 2: CƠ SỞ LÝ THUYẾT 5](#_Toc175183683)

[2.1. Cơ sở lý thuyết 5](#_Toc175183684)

[2.2. Các nghiên cứu liên quan. 9](#_Toc175183685)

[CHƯƠNG 3: PHƯƠNG PHÁP ĐỀ XUẤT PHÂN ĐOẠN ẢNH MRI U NÃO 13](#_Toc175183686)

[3.1. Mô tả bài toán. 13](#_Toc175183687)

[3.2. Phương pháp đề xuất. 14](#_Toc175183688)

[3.2.1. Unet 14](#_Toc175183689)

[3.2.2. Unet cải tiến 17](#_Toc175183690)

[3.2.3. ResUnet 19](#_Toc175183691)

[3.2.4. ResUnet++ 21](#_Toc175183692)

[3.2.5. MultiResUnet 26](#_Toc175183693)

[3.3. Phương pháp đánh giá 28](#_Toc175183694)

[3.3.1. Dice Score 29](#_Toc175183695)

[3.3.2. Jaccard Index 29](#_Toc175183696)

[CHƯƠNG 4: THÍ NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ. 31](#_Toc175183697)

[4.1. Tập dữ liệu và cấu hình phần cứng. 31](#_Toc175183698)

[4.2. Thí nghiệm và đánh giá kết quả. 32](#_Toc175183699)

[CHƯƠNG 5: KẾT LUẬN 38](#_Toc175183700)

[5.1. Kết quả đạt được. 38](#_Toc175183701)

[5.2. Ưu và nhược điểm của phương pháp đề xuất. 39](#_Toc175183702)

[5.3. Hướng mở rộng tương lai 40](#_Toc175183703)

[TÀI LIỆU THAM KHẢO 41](#_Toc175183704)

[PHỤ LỤC 43](#_Toc175183705)

DANH MỤC CÁC KÝ HIỆU, CHỮ VIẾT TẮT VÀ THUẬT NGỮ

|  |  |  |
| --- | --- | --- |
| STT | Chữ viết tắt và thuật ngữ | Ý nghĩa |
| 1 | MRI | Magnetic Resonance Imaging (Hình ảnh Cộng hưởng Từ) |
| 2 | CNN | Convolutional Neural Network (Mạng Nơ-ron Tích chập) |
| 4 | Ground Truth | Dữ liệu chính xác để so sánh với dự đoán của mô hình |
| 5 | IoU | Intersection over Union (Giao nhau trên Liên kết) |
| 6 | TP | True Positive (Dương tính thật) |
| 7 | FP | False Positive (Dương tính giả) |
| 8 | TN | True Negative (Âm tính thật) |
| 9 | FN | False Negative (Âm tính giả) |
| 10 | Dice Coefficient | Hệ số Dice, một công cụ thống kê đo lường sự tương đồng giữa hai tập dữ liệu |

DANH MỤC CÁC BẢNG

[Bảng 4.1: Trực quan hóa bộ dữ liệu 31](#_Toc175183658)

[Bảng 4.2: Các tham số để huấn luyện mô hình. 33](#_Toc175183659)

[Bảng 4.3: Kết quả dự đoán của các mô hình. 35](#_Toc175183660)

DANH MỤC CÁC HÌNH VẼ, ĐỒ THỊ

[Hình 2.1: Vị trí của một lát cắt MRI trong bộ ảnh MRI [1] 5](#_Toc175183641)

[Hình 2.2: Kiến trúc mô hình Unet [7] 6](#_Toc175183642)

[Hình 2.3: Kiến trúc mô hình ResUnet [8] 7](#_Toc175183643)

[Hình 2.4: Khối tương tự Inception [6] 8](#_Toc175183644)

[Hình 2.5: Đường dẫn Residual [6] 9](#_Toc175183645)

[Hình 2.6: Full-scale skip connections [6] 9](#_Toc175183646)

[Hình 3.1: Hình ảnh đầu vào (Input) 13](#_Toc175183647)

[Hình 3.2: Hình ảnh đầu ra (Output) 13](#_Toc175183648)

[Hình 3.3: Hình ảnh mô hình Unet [7] 14](#_Toc175183649)

[Hình 3.4: Hình ảnh mô hình Unet cải tiến 17](#_Toc175183650)

[Hình 3.5: Hình ảnh mô hình ResUnet [12] 20](#_Toc175183651)

[Hình 3.6: Squeeze và excitation blocks [14] 22](#_Toc175183652)

[Hình 3.7: Tích chập Atrous và tích chập CNN 23](#_Toc175183653)

[Hình 3.8: Hình ảnh mô hình ResUnet++ [12] 24](#_Toc175183654)

[Hình 3.9: MultiResUnet [6] 26](#_Toc175183655)

[Hình 3.10: Hình ảnh mô hình MultiResUnet [6] 27](#_Toc175183656)

[Hình 4.1: Dice Score và Jaccard của tất cả các mô hình. 33](#_Toc175183657)

# GIỚI THIỆU

## Giới thiệu đề tài

Trong lĩnh vực y học, việc chẩn đoán và điều trị các bệnh lý não là một thách thức lớn đối với các bác sĩ và chuyên gia. Ảnh não được thu thập từ các phương pháp chẩn đoán hình ảnh như MRI (Magnetic Resonance Imaging) và CT (Computed Tomography) đóng vai trò quan trọng trong việc phát hiện và theo dõi các tổn thương, khối u và các dị tật khác trong não. Tuy nhiên, việc phân tích và xử lý các ảnh này đòi hỏi kỹ năng chuyên môn cao và tiêu tốn nhiều thời gian.

Phân đoạn ảnh y tế, đặc biệt là phân đoạn ảnh não là quá trình tách rời các vùng quan trọng trong ảnh để giúp nhận dạng và đánh giá các cấu trúc và tổn thương. Việc này giúp các chuyên gia y tế có được cái nhìn chi tiết và chính xác hơn về tình trạng của bệnh nhân, từ đó hỗ trợ quá trình chẩn đoán và điều trị.

Trong những năm gần đây, với sự phát triển mạnh mẽ của trí tuệ nhân tạo (AI) và các phương pháp học sâu (Deep Learning) đã mở ra nhiều cơ hội mới trong việc tự động hóa quá trình phân đoạn ảnh y tế này. Đặc biệt là Semantic Segmentation, một kỹ thuật trong học sâu đã chứng minh được khả năng vượt trội trong việc phân đoạn các đối tượng trong ảnh với độ chính xác cao. Semantic Segmentation không chỉ giúp phân đoạn các vùng cụ thể trong ảnh mà còn gán nhãn cho từng pixel, giúp nhận diện chính xác hơn các cấu trúc và tổn thương trong não.

Với những tiềm năng và lợi ích to lớn mà Semantic Segmentation mang lại, đề tài "Phân đoạn ảnh não dựa trên mô hình Semantic Segmentation" được nhóm chọn nhằm nghiên cứu và phát triển một mô hình phân đoạn ảnh não hiệu quả. Đề tài này không chỉ tập trung vào việc phát triển mô hình mà còn ứng dụng vào các bộ dữ liệu ảnh não thực tế, đánh giá hiệu quả và đề xuất các phương pháp cải tiến để nâng cao độ chính xác cũng như tính ứng dụng trong thực tế.

## Mục tiêu và nội dung của đề tài

Mục tiêu của đề tài là xây dựng và phát triển một mô hình phân đoạn ảnh não dựa trên kỹ thuật học sâu Semantic Segmentation. Cụ thể, ảnh MRI sọ não được sử dụng trong bài toán này.

Để đạt được mục tiêu trên, đề tài thực hiện các công việc sau đây:

1. Tìm hiểu đặc trưng ảnh MRI u não.
2. Tìm hiểu các nghiên cứu liên quan và phân tích ưu nhược điểm của chúng.
3. Tìm hiểu tập dữ liệu ảnh MRI u não.
4. Đề xuất phương pháp phân đoạn ảnh MRI u não.
5. Thực nghiệm, đánh giá kết quả và so sánh với các phương pháp khác.

## Giới hạn đề tài

Trong đề tài này, nhóm xây dựng và phát triển mô hình phân đoạn ảnh não MRI dựa trên kỹ thuật học sâu Semantic Segmentation. Vì đây là một bài toán có quy mô lớn và rộng nên trong khuôn khổ luận văn này, nhóm đã giới hạn lại các công việc cần thiết để sử dụng cho bài toán trên:

1. Hạn chế về dữ liệu: Do nhóm sử dụng tập dữ liệu ảnh MRI não có sẵn nên bị hạn chế về số lượng cũng như tính đa dạng và chất lượng hình ảnh.
2. Hạn chế về tài nguyên phần cứng và thời gian: Trong quá trình huấn luyện mô hình Semantic Segmentation đòi hỏi phải có tài nguyên tính toán lớn kèm theo đó là thời gian huấn luyện mô hình dài.
3. Khả năng đánh giá của mô hình: mô hình được nhóm xây dựng chỉ dự đoán và đánh giá độ chính xác của bài toán. (nhóm chưa thực hiện đánh giá hiệu suất xử lý của bài toán).

## Phương pháp nghiên cứu

Phương pháp nghiên cứu lý thuyết: Tìm hiểu các khái niệm, lý thuyết và các công trình nghiên cứu liên quan đến phân đoạn ảnh não dựa trên mô hình Semantic Segmentation. Đây là giai đoạn nghiên cứu và phân tích các khái niệm cơ bản, các mô hình và phương pháp đã được đề xuất trong tài liệu chuyên ngành.

Phương pháp thực nghiệm khoa học: Xây dựng và đánh giá hiệu quả của mô hình phân đoạn ảnh não. Quá trình này bao gồm thu thập dữ liệu, chuẩn bị dữ liệu, thiết kế và huấn luyện mô hình, đánh giá kết quả và tối ưu hóa mô hình dựa trên các thử nghiệm thực tế và phân tích kết quả thu được.

## Đóng góp của luận văn

Đóng góp về mặt khoa học: Luận văn góp phần vào việc nghiên cứu và đánh giá các phương pháp phân đoạn ảnh não dựa trên mô hình Semantic Segmentation. Hơn nữa, luận văn cung cấp một cái nhìn sâu sắc vào các kỹ thuật và mô hình hiện đại trong lĩnh vực này. Nghiên cứu có thể đề xuất các cải tiến và xây dựng mô hình có khả năng tổng quát hóa tốt hơn trên các tập dữ liệu đa dạng.

Đóng góp về mặt thực tiễn: Luận văn góp phần vào việc nghiên cứu áp dụng giải thuật phân đoạn ảnh MRI não có thể sử dụng ngoài thực tế trong tương lai, giúp cải thiện quá trình chẩn đoán và điều trị bệnh nhân bằng cách cung cấp các công cụ hỗ trợ phân đoạn và phân tích ảnh MRI u não. Các kết quả nghiên cứu có thể đóng vai trò quan trọng trong việc phát triển công nghệ hỗ trợ y tế và cải thiện chất lượng chăm sóc sức khỏe.

## Cấu trúc luận văn

Để đạt được mục tiêu trên, đề tài được tổ chức thành 5 chương có cấu trúc như sau:

***Chương 1: Giới thiệu.*** Trong chương này, nhóm trình bày tổng quan về vấn đề nghiên cứu, mục tiêu và phương pháp nghiên cứu cùng với phạm vi và ý nghĩa của đề tài.

***Chương 2: Cơ sở lý thuyết và các nghiên cứu liên quan.*** Trong chương này, nhóm trình bày các kiến thức lý thuyết cơ bản liên quan đến phân đoạn ảnh não và các mô hình Semantic Segmentation. Đồng thời, phân tích các nghiên cứu liên quan để đặt ra nền tảng cho phương pháp nghiên cứu của luận văn.

***Chương 3: Phương pháp đề xuất.*** Trong chương này, nhóm mô tả chi tiết phương pháp nghiên cứu được áp dụng trong luận văn, bao gồm quy trình tiền xử lý dữ liệu, thiết kế và triển khai mô hình Semantic Segmentation, cùng với các phương pháp đánh giá hiệu quả.

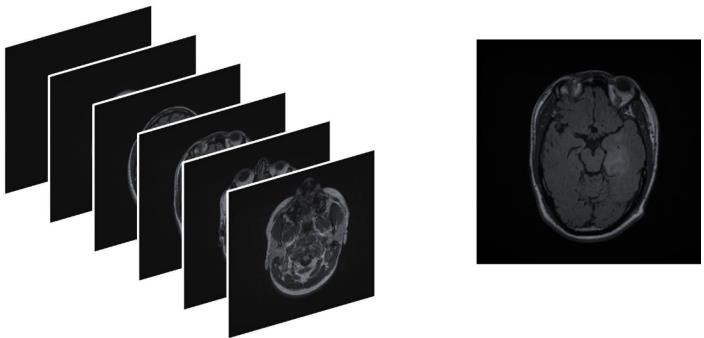
***Chương 4: Hiện thực và đánh giá kết quả.*** Trong chương này, nhóm trình bày quá trình thực hiện thí nghiệm, kết quả thu được và phân tích so sánh giữa kết quả dự đoán của mô hình và kết quả thực tế.

***Chương 5: Kết luận.*** Trong chương này, nhóm tổng kết các công việc đã thực hiện, ưu nhược điểm của mô hình đề xuất cũng như đồng thời đề xuất hướng mở rộng trong tương lai để nâng cao hiệu quả và ứng dụng của mô hình Semantic Segmentation trong lĩnh vực y tế.

# CƠ SỞ LÝ THUYẾT

## Cơ sở lý thuyết

Ảnh MRI não là tập hợp nhiều ảnh xám, mỗi ảnh gọi là lát cắt thể hiện ảnh chụp tại một vị trí cắt ngang của não được thể hiện như Hình 2.1. Tùy vào cấu hình máy chụp, mỗi bộ ảnh MRI có thể có 16, 32 hoặc 64 lát cắt. Vùng bất thường trên mỗi ảnh có màu sắc, cấu trúc, hình dạng khác với vùng ảnh thông thường. Để khảo sát các dạng vùng bất thường khác nhau (các bệnh khác nhau của não), người ta chụp với các chuỗi xung khác nhau tạo ra các bộ ảnh MRI khác nhau như FLAIR, DWI và STIR [1].

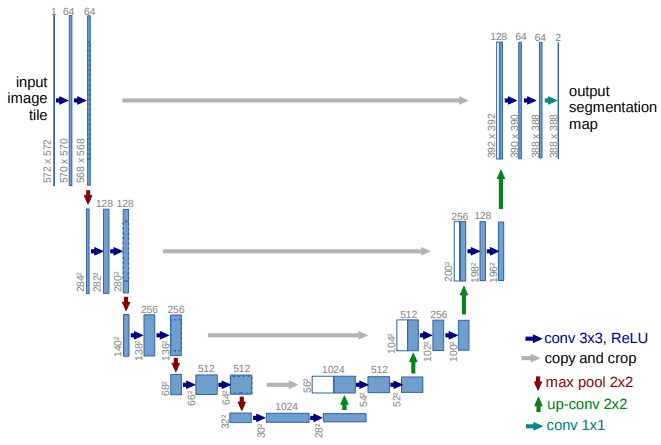


Vị trí của một lát cắt MRI trong bộ ảnh MRI [1]

Ảnh MRI có độ tương phản cao giữa các mô khác nhau trong não như chất trắng, chất xám và dịch não tủy, giúp phân biệt rõ ràng các cấu trúc khác nhau. Hơn nữa, độ phân giải cao của ảnh MRI cho phép nhận diện chi tiết các cấu trúc nhỏ, điều này rất quan trọng trong việc phát hiện và phân đoạn các vùng bất thường. Thông tin đa chuỗi từ các chuỗi xung khác nhau cung cấp thêm nhiều khía cạnh khác nhau về cấu trúc và tình trạng của não, làm tăng độ chính xác trong việc chẩn đoán và điều trị [2].

Ban đầu các phương pháp phát hiện vùng bất thường chủ yếu dựa vào các thuật toán máy học truyền thống [3]. Với sự phát triển của CNN, Unet được đề xuất trong [4], để phân đoạn hình ảnh. Unet rất phù hợp cho nhiệm vụ phân đoạn ảnh MRI não vì khả năng xử lý tốt các đặc trưng như độ tương phản cao giữa các mô, độ phân giải cao, và thông tin đa chuỗi từ các chuỗi xung khác nhau. Cấu trúc đối xứng của Unet giúp tăng cường khả năng học và phục hồi các đặc trưng quan trọng từ ảnh, làm cho nó trở thành một công cụ mạnh mẽ trong việc phát hiện và phân đoạn các vùng bất thường trong ảnh MRI não. Với sự đơn giản và hiệu suất vượt trội của cấu trúc hình chữ U, nhiều biến thể khác của Unet như ResUnet [[5],](#_bookmark4) MultiResUnet [6] cũng đã được phát triển[.](#_bookmark5)

Unet là một kiến trúc mạng nơ-ron sâu được giới thiệu bởi Ronneberger vào năm 2015 [7] để giải quyết vấn đề phân đoạn ảnh y tế. Kiến trúc của Unet có dạng hình chữ U được thể hiện như Hình 2.2.



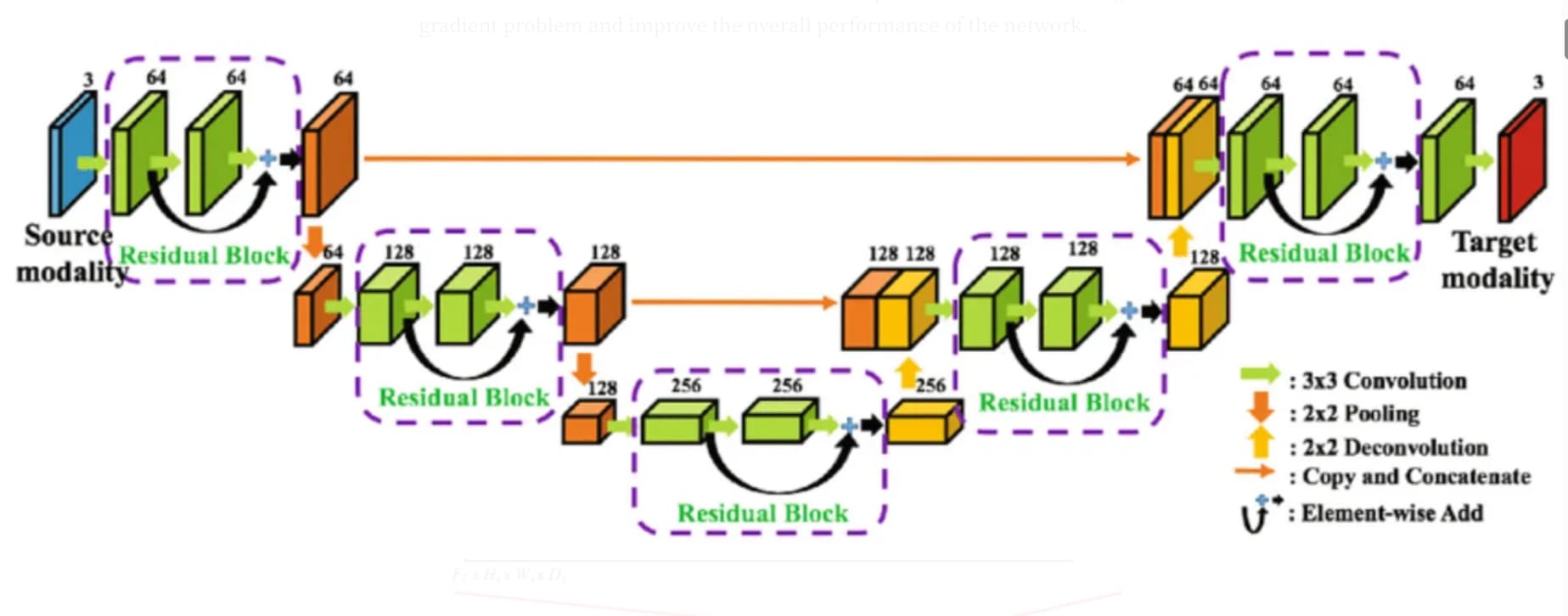
Kiến trúc mô hình Unet [7]

Kiến trúc Unet trên được mô tả bao gồm Encoder (con đường nén phía bên trái) và Decoder (con đường mở rộng phía bên phải). Trong đó, Encoder sử dụng các lớp tích chập và các lớp max-pooling để giảm kích thước không gian của ảnh đầu vào, trong khi Decoder sử dụng các lớp tích chập và up-sampling để khôi phục kích thước ban đầu của ảnh đầu vào. Điểm đặc biệt của Unet là nó sử dụng các Skip connections (Kết nối bỏ qua) giữa các lớp tương ứng của Encoder và Decoder, giúp bảo tồn thông tin không gian chi tiết từ ảnh đầu vào [7].

Các thành phần chính của Unet bao gồm:

* Encoder (bộ mã hóa): Chuỗi các khối tích chập và max-pooling để trích xuất các đặc trưng từ ảnh đầu vào. Mỗi khối tích chập bao gồm hai lớp tích chập 3×3, theo sau là một lớp max-pooling 2×2 để giảm kích thước không gian.
* Bottleneck (cổ chai): Khối tích chập nằm ở giữa mạng, có nhiệm vụ trích xuất các đặc trưng cao cấp từ ảnh đầu vào.
* Decoder (bộ giải mã): Chuỗi các khối up-sampling và tích chập để khôi phục kích thước không gian của ảnh đầu ra. Mỗi khối up-sampling bao gồm một lớp up-sampling theo sau bởi hai lớp tích chập 3×3.
* Skip connections (kết nối bỏ qua): Các kết nối trực tiếp giữa các lớp tương ứng của encoder và decoder, giúp bảo tồn thông tin chi tiết từ ảnh đầu vào.

ResUnet là một biến thể của Unet kết hợp các khối residual của ResNet vào kiến trúc Unet để cải thiện khả năng huấn luyện và hiệu suất của mạng. Các khối residual cho phép mạng học được các đặc trưng sâu mà không gặp phải vấn đề về gradient biến mất, nhờ các kết nối tắt (shortcut connections). Điều này giúp mạng dễ dàng huấn luyện hơn và đạt được hiệu suất cao hơn trong các tác vụ phân đoạn ảnh [5].



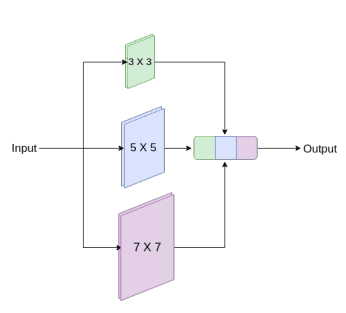
Kiến trúc mô hình ResUnet [8]

Kiến trúc ResUnet ở Hình 2.3 trên bao gồm:

* Residual blocks (khối residual): Các khối residual được thêm vào encoder và decoder của Unet. Mỗi khối residual bao gồm hai lớp tích chập 3×3 và một kết nối tắt (shortcut connection) giữa đầu vào và đầu ra của khối.
* Bottleneck with residual blocks (cổ chai với các khối residual): Khối cổ chai của ResUnet cũng bao gồm các khối residual để trích xuất các đặc trưng cao cấp từ ảnh đầu vào

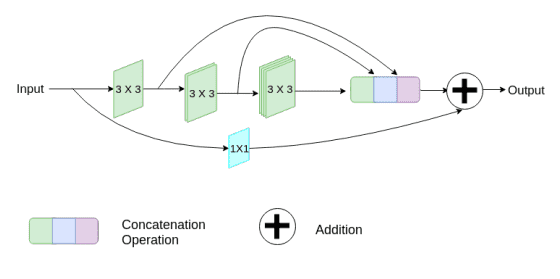
MultiResUnet là một cải tiến khác của Unet, nhằm mục đích kết hợp thông tin đa tỷ lệ để cải thiện khả năng phân đoạn của mạng. MultiResUnet sử dụng các khối MultiRes, kết hợp các lớp tích chập với kích thước kernel khác nhau để nắm bắt các đặc trưng ở các tỷ lệ khác nhau được thể hiện thông qua hình. Ngoài ra, MultiResUnet còn sử dụng các kết nối Residual để cải thiện khả năng huấn luyện của mạng [6]. Một kiến trúc MultiResUnet bao gồm 3 phần:

* + MultiRes Block: MultiRes Block của MultiResUNet sử dụng một khối tương tự Inception nhằm mục đích nắm bắt các đặc trưng ở nhiều tỷ lệ khác nhau được mô tả qua Hình 2.4. Cụ thể, khối Inception này bao gồm nhiều lớp tích chập với các kích thước kernel khác nhau như 3×3, 5×5 và 7×7. Việc kết hợp các lớp tích chập này trong một khối giúp mạng có khả năng xử lý và nhận diện các đối tượng với kích thước và hình dạng đa dạng, từ đó cải thiện độ chính xác và hiệu suất của mô hình phân đoạn ảnh y tế.



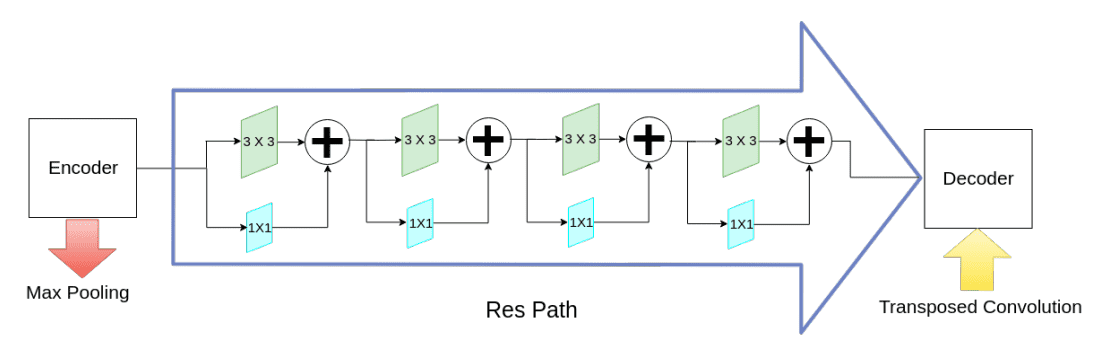
Khối tương tự Inception [6]

* + Residual Path: Đường dẫn Residual được mô tả ở Hình 2.5 là một cải tiến quan trọng trong khối MultiRes. Tăng dần số bộ lọc trong các lớp tích chập kế tiếp để giảm bớt lượng bộ nhớ được sử dụng ở các lớp ban đầu. Cuối cùng, bổ sung một kết nối bỏ qua ngắn dưới dạng residual và đưa vào một lớp tích chập pointwise (kích thước 1 x 1), để giúp mô hình nắm bắt thêm thông tin không gian.



Đường dẫn Residual [6]

* + Full-scale skip connections: MultiResUNet sử dụng các kết nối bỏ qua (skip connections) ở nhiều cấp độ khác nhau trong mạng như Hình 2.6, từ đầu vào đến đầu ra. Các kết nối này giúp bảo toàn thông tin chi tiết trong quá trình lan truyền ngược và cải thiện độ chính xác của phân đoạn.



Full-scale skip connections [6]

## Các nghiên cứu liên quan.

Nghiên cứu của Tsai và các đồng nghiệp [3] tập trung vào phương pháp phân đoạn hình ảnh y học bằng cách sử dụng các đường đẳng mức dựa trên hình dạng. Nghiên cứu này đề xuất một phương pháp sử dụng các đường đẳng mức để tự động hoặc bán tự động phân đoạn các cấu trúc trong hình ảnh y học. Mục tiêu nhằm cải thiện quá trình phân đoạn hình ảnh y học bằng cách áp dụng các đường đẳng mức, giúp giảm thiểu sự phụ thuộc vào người thực hiện và tăng tính chính xác của quá trình phân đoạn. Nghiên cứu đã áp dụng phương pháp này vào việc phân đoạn các cấu trúc y tế như gan, não và các cơ quan khác trên hình ảnh y học, đánh giá hiệu quả thông qua so sánh với các phương pháp truyền thống khác. Ưu điểm chính của phương pháp này là khả năng giảm thiểu sự phụ thuộc vào người thực hiện, tăng độ chính xác của quá trình phân đoạn, và hỗ trợ cho nhiều loại cấu trúc y tế khác nhau. Tuy nhiên, nhược điểm của phương pháp là trong một số trường hợp phức tạp, nó vẫn yêu cầu sự can thiệp của con người để đạt được kết quả chính xác, đặc biệt là khi đối mặt với các cấu trúc phức tạp hoặc có biên giới không rõ ràng. Kết quả cho thấy phương pháp đề xuất có khả năng phân đoạn chính xác hơn và ít phụ thuộc vào sự can thiệp của con người.

Nghiên cứu của Held và các đồng nghiệp [9] trình bày phương pháp phân đoạn hình ảnh MRI não bằng mô hình Markov random field (MRF). Nghiên cứu này đề xuất sử dụng MRF để cải thiện quá trình phân đoạn các cấu trúc não trên hình ảnh y học.Mục tiêu nhằm áp dụng mô hình MRF để tự động hoá quá trình phân đoạn hình ảnh MRI não, giúp tăng tính chính xác và độ tin cậy của kết quả phân đoạn so với các phương pháp truyền thống. Các tác giả đã phát triển một mô hình dựa trên lý thuyết MRF để mô hình hóa sự tương tác giữa các điểm ảnh trong hình ảnh MRI, từ đó áp dụng các thuật toán phân đoạn dựa trên độ tương tự và xác suất. Phương pháp này có ưu điểm lớn là khả năng mô hình hóa sự tương tác giữa các điểm ảnh dựa trên lý thuyết xác suất và độ tương tự, từ đó mang lại kết quả phân đoạn chính xác và ổn định hơn so với các phương pháp truyền thống. Tuy nhiên, một nhược điểm lớn của MRF là đòi hỏi rất nhiều tài nguyên tính toán và có cấu trúc thiết lập phức tạp, khiến cho việc triển khai thực tế trở nên khó khăn. Kết quả của nghiên cứu cho thấy phương pháp này mang lại kết quả phân đoạn chính xác và ổn định.

Nghiên cứu của Le và các đồng nghiệp [4] sử dụng mô hình U-Net để phát hiện các vùng bất thường trên hình ảnh MRI não. Nghiên cứu này được trình bày tại Hội nghị Quốc gia về Nghiên cứu Cơ bản và Ứng dụng Công nghệ Thông tin (FAIR) tại Nha Trang, Việt Nam. Nghiên cứu nhằm áp dụng mô hình U-Net, một mô hình mạng học sâu được sử dụng phổ biến trong xử lý hình ảnh y tế, để tự động phát hiện và phân đoạn các vùng bất thường trên hình ảnh MRI não. Các tác giả đã áp dụng mô hình U-Net vào bộ dữ liệu hình ảnh MRI để phát hiện các khu vực có dấu hiệu bất thường như khối u, tổn thương hoặc các biến đổi cấu trúc. Mô hình được huấn luyện để nhận biết các đặc trưng của các bệnh lý thường gặp trên hình ảnh MRI. Ưu điểm của U-Net là dễ triển khai, hiệu quả cao trong việc nhận diện các đặc trưng bất thường, và được sử dụng rộng rãi trong xử lý hình ảnh y tế. Tuy nhiên, nhược điểm của phương pháp này là nó phụ thuộc rất nhiều vào chất lượng của dữ liệu huấn luyện, và thường gặp khó khăn trong việc phân đoạn các vùng biên phức tạp hoặc có biến dạng lớn.

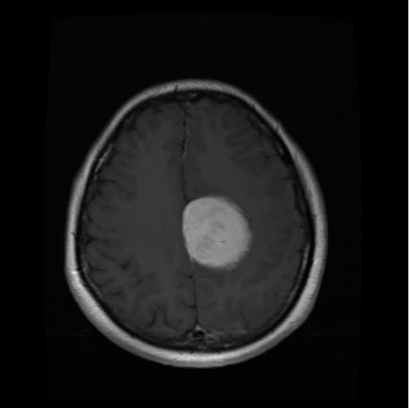
Nghiên cứu của Hatamizadeh và các đồng nghiệp [10] tập trung vào ứng dụng của mô hình Swin UNetR trong phân đoạn u não trên hình ảnh MRI. Mô hình này sử dụng Swin Transformers, một kiến trúc mạng nổi bật trong lĩnh vực xử lý ngôn ngữ tự nhiên và thị giác máy tính, để cải thiện khả năng phân đoạn hình ảnh y tế. Kiến trúc Swin UNetR kết hợp giữa Swin Transformer và kiến trúc UNet, một mô hình mạng học sâu được sử dụng phổ biến trong phân đoạn hình ảnh y tế. Swin Transformer được chọn vì khả năng xử lý thông tin không gian rộng và khả năng học tập sâu về mối quan hệ bất kể vị trí, giúp cải thiện độ chính xác của phân đoạn so với các phương pháp truyền thống. Ứng dụng trong phân đoạn u não trên MRI là mục tiêu chính của nghiên cứu, nhằm cung cấp các công cụ hỗ trợ chẩn đoán y tế chính xác và nhanh chóng. Ưu điểm nổi bật của Swin Transformer là khả năng xử lý thông tin không gian rộng và học tập sâu về mối quan hệ bất kể vị trí, giúp cải thiện độ chính xác trong phân đoạn hơn so với các phương pháp truyền thống. Tuy nhiên, Swin UNetR có nhược điểm là đòi hỏi phần cứng mạnh mẽ và nhiều tài nguyên tính toán, điều này có thể gây khó khăn trong việc triển khai thực tế tại các cơ sở y tế có hạ tầng hạn chế. Việc sử dụng Swin UNetR được đánh giá là có thể giúp cải thiện quá trình phân tích hình ảnh, từ đó hỗ trợ các bác sĩ và nhân viên y tế trong việc phát hiện và đánh giá kích thước, vị trí của u não trên MRI một cách chính xác và toàn diện hơn.

Nghiên cứu của Xue Li và đồng nghiệp [5] tập trung vào việc phát triển một phương pháp phân đoạn ảnh não để nâng cao hiệu quả trong việc phát hiện và phân loại các khối u não từ MRI. Để đạt được điều này, tác giả đề xuất một phiên bản cải tiến của mạng Res-UNet, một kiến trúc mạng nơ-ron tích chập (CNN) nổi bật trong lĩnh vực phân đoạn ảnh y tế. Phương pháp cải tiến của họ kết hợp các khối residual với kiến trúc UNet, nhằm cải thiện khả năng học của mạng và giảm thiểu các vấn đề liên quan đến phân đoạn ảnh phức tạp. Mô hình mới này được tinh chỉnh để xử lý hiệu quả hơn các đặc điểm của hình ảnh MRI, từ đó nâng cao độ chính xác và tính nhạy cảm trong việc phát hiện u não. Ưu điểm của mô hình là khả năng cải thiện đáng kể độ chính xác phân đoạn nhờ vào việc kết hợp khối residual với UNet, giúp nhận diện và phân loại các đặc điểm phức tạp trong ảnh MRI. Tuy nhiên, nhược điểm của phương pháp là yêu cầu tài nguyên tính toán cao, điều này có thể làm tăng thời gian và chi phí đào tạo mô hình. Ngoài ra, mô hình có thể không tổng quát tốt cho các tập dữ liệu khác và cần điều chỉnh siêu tham số phức tạp. Kết quả thí nghiệm cho thấy mô hình cải tiến này vượt trội hơn các phương pháp hiện có về độ chính xác và hiệu suất phân đoạn, điều này được minh chứng qua các chỉ số như độ chính xác, độ nhạy, và độ đặc hiệu. Bài báo cũng thảo luận về ứng dụng của phương pháp này trong việc hỗ trợ chẩn đoán y tế và lập kế hoạch điều trị, đồng thời đề xuất các hướng nghiên cứu tiếp theo để mở rộng và tối ưu hóa hơn nữa phương pháp phân đoạn này.

# PHƯƠNG PHÁP ĐỀ XUẤT PHÂN ĐOẠN ẢNH MRI U NÃO

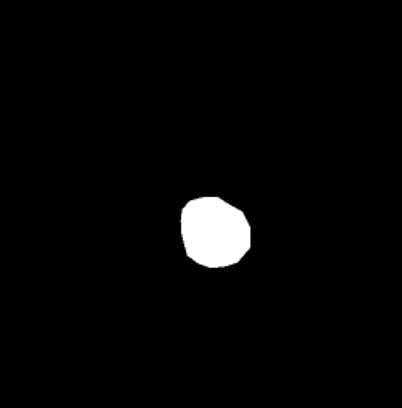
## Mô tả bài toán.

Nhóm thực hiện giải quyết bài toán phân đoạn ảnh não dựa trên các mô hình Semantic Segmentation. Mục đích của bài toán này là xác định vị trí và phân đoạn các vùng bị ảnh hưởng bởi khối u trong các hình ảnh trong các ảnh cộng hưởng từ MRI của não. Hình ảnh đầu vào (Input) của bài toán sẽ là ảnh cộng hưởng từ của não chưa được phân đoạn, thể hiện một lát cắt ngang qua não tương tự như Hình 3.1 dưới đây:



Hình ảnh đầu vào (Input)

Hình ảnh đầu ra (output) của bài toán sẽ là hình ảnh đã phân đoạn cho ảnh MRI đầu vào tương tự như Hình 3.2, cho biết vị trí và các vùng bị ảnh hưởng bởi khối u trong não.

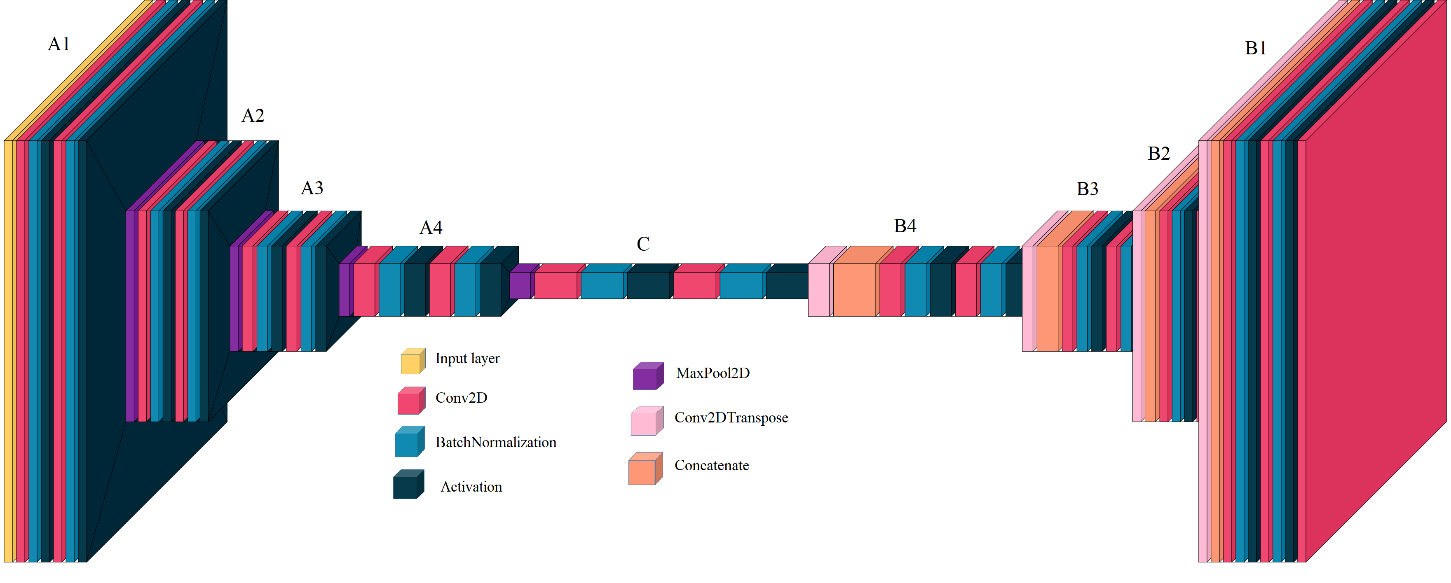


Hình ảnh đầu ra (Output)

## Phương pháp đề xuất.

Dựa trên đặc trưng của ảnh MRI và để xử lý được các hình ảnh y học với mục tiêu không chỉ là phân loại ảnh mà còn xác định được vị trí. Với mô hình CNN thì không thể hoàn thành tốt yêu cầu của bài toán, nên các nhà khoa học đã nghiên cứu ra các mô hình semantic segmentation với nhiệm vụ là phân đoạn và xác định vị trí. Đầu tiên là mô hình Unet được phát minh và chứng minh chúng rất hiệu quả trong phân đoạn ảnh y tế, và sau này đã có nhiều nghiên cứu phát triển dựa trên mô hình Unet gốc đưa ra kết quả tốt hơn. Từ những thông tin đó nên nhóm quyết định chọn các mô hình Semantic segmentation sau để giải quyết vấn đề của mình. Cụ thể, mô hình đề xuất của nhóm là mô hình Unet cải tiến. Mặt khác, nhóm cũng tiến hành đề xuất thêm mô hình Unet, ResUnet, ResUnet++ và MultiResUnet. Mỗi mô hình đều có ưu và nhược điểm riêng của chúng. Các mô hình này được làm rõ ở phần bên dưới.

### Unet



Hình ảnh mô hình Unet [7]

Kiến trúc Unet được dựa theo [7] như mô tả hình 3.3 gồm có 3 phần: encoder, bottleneck và decoder.

1. **Encoder**: Có nhiệm vụ là trích xuất các đặc trưng của ảnh đầu vào. Trong mô hình Unet bao gồm có 4 lớp encoder, mỗi lớp encoder là một chuỗi các lớp Convolution kích thước 3×3, ReLU, lớp max-pooling 2×2 để downsample và tăng gấp đôi số bộ lọc trong các lớp tích chập sau khi max-pooling. Chúng có vai trò chủ chốt là xác định vật thể trong hình ảnh đầu vào là cái gì. Bốn lớp encoder được đánh dấu chỉ số tương ứng từ A1 đến A4 trong hình.
2. **Bottleneck**: gồm có lớp Convolution và hàm activation ReLU. Chúng giúp duy trì thông tin bằng cách kết nối các bản đồ đặc trưng (feature map) từ encoder qua decoder tương ứng thông qua skip connections. Khối bottleneck được kí hiệu là C tương ứng như trong hình.
3. **Decoder**: có nhiệm vụ là khôi phục lại kích thước và chi tiết của ảnh từ các đặc trưng trích xuất được từ phần encoder. Trong mô hình Unet bao gồm có 4 lớp decoder, mỗi lớp decoder là một chuỗi các lớp Convolution kích thước 3×3 với hàm activation ReLu, thực hiện tăng kích thước ảnh kết hợp với lớp tích chập 2×2 và giảm một nửa số bộ lọc trong các lớp tích chập sau khi tăng kích thước ảnh. Chúng có vai trò chủ chốt là xác định vị trí của vật thể. Bốn lớp decoder được đánh dấu chỉ số tương ứng từ B4 đến B1 trong hình.

Bắt đầu của mô hình là một lớp ảnh đầu vào có kích thước (256, 256, 3), lớp ảnh đầu vào sẽ đi vào khối encoder A1. Trong khối A1 bao gồm 2 lớp tích chập (Convolution layers) với 64 bộ lọc và kích thước kernel 3×3, theo sau đó là có hàm activation ReLU và hàm chuẩn hóa BatchNormalization sau mỗi lớp tích chập, cuối cùng là một lớp pooling có kích thước là 2×2, tới đây lớp ảnh đầu vào sẽ có kích thước là (128, 128, 64), rồi tiếp tục qua khối A2.

Trong khối encoder A2 bao gồm 2 lớp tích chập (Convolution layers) với 128 bộ lọc với kích thước kernel 3×3, theo sau đó tiếp tục là hàm activation ReLU và hàm chuẩn hóa BatchNormalization sau mỗi lớp tích chập, cuối cùng là một lớp pooling có kích thước là 2×2, sau khi đi qua lớp pooling thì ảnh đầu vào có kích thước là (64, 64, 128) rồi chuyển qua khối A3. Trong khối decoder A3 cũng tương tự như khối A1 và A2, nhưng sử dụng 2 lớp tích chập (Convolution layers) với 256 bộ lọc với kích thước kernel 3×3 rồi đi đến khối A4. Khối encoder A4 tương tự như khối A3 nhưng khối này sử dụng 2 lớp tích chập (Convolution layers) với 512 bộ lọc với kích thước kernel 3×3, kích thước của ảnh đầu vào sau khi đi qua khối A3 và A4 lần lượt là (32, 32, 256) và (16, 16, 512). Như vậy thì sau khi đi qua bốn khối encoder từ A1 đến A4 thì kích thước của lớp ảnh đầu vào sẽ là (16, 16, 512).

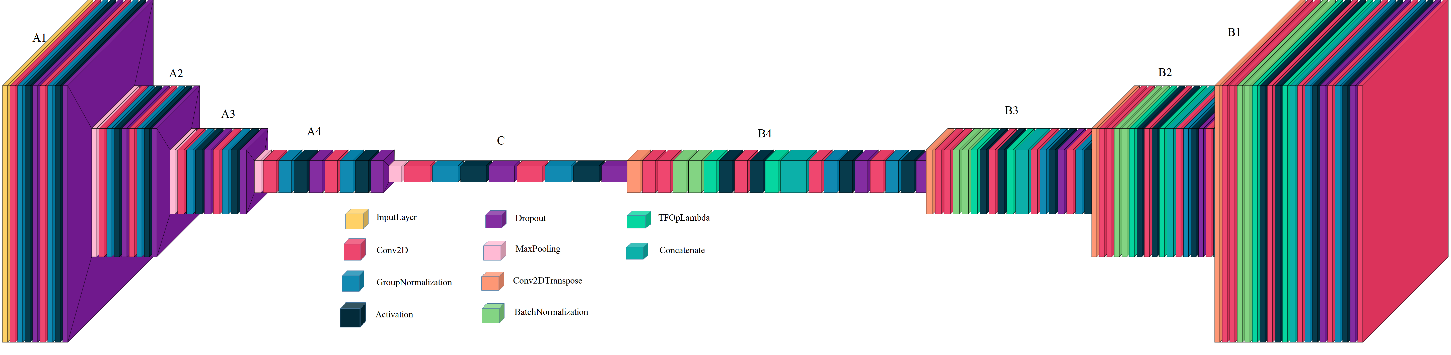
Kết thúc khối encoder sẽ tới khối bottleneck C, trong khối bottleneck bao gồm 2 lớp tích chập (Convolution layers) với 1024 bộ lọc và kích thước kernel 3×3, theo sau đó là có hàm activation ReLU và hàm chuẩn hóa BatchNormalization sau mỗi lớp tích chập, sau khi đi qua khối bottleneck thì kích thước của lớp ảnh đầu vào sẽ là (16, 16, 1024).

Sau khi đi qua khối bottleneck thì ảnh đầu vào sẽ tới khối decoder và sẽ đi từ khối decoder B4 tới B1. Trong mỗi khối decoder sẽ sử dụng thêm 2 lớp đó là Conv2DTranspose và concatenate, lớp Conv2DTranspose được gọi là lớp tích chập ngược, chúng có nhiệm vụ là khôi phục lại kích thước của ảnh trước khi chúng bị giảm kích thước trong khối encoder. Lớp concatenate dùng để kết nối các feature map từ khối encoder đến khối decoder tương ứng, giúp cải thiện khả năng khôi phục các đặc trưng để giữ được thông tin quan trọng và cải thiện độ chính xác trong phân đoạn ảnh.

Trong khối decoder B4 bao gồm một lớp Conv2DTranspose với 512 bộ lọc và kích thước kernel 2×2, một lớp concatenate nối với các đặc trưng từ khối encoder A4, hai lớp tích chập (Convolution layers) với 512 bộ lọc và kích thước kernel 3×3, theo sau đó là có hàm activation ReLU và hàm chuẩn hóa BatchNormalization sau mỗi lớp tích chập và sau khi đi qua khối decoder B4 thì lớp ảnh đầu vào sẽ có kích thước là (32, 32, 512). Tiếp tục qua khối decoder B3, trong khối này tương đồng với khối decoder B4 nhưng có sự thay đổi là sử dụng lớp Conv2DTranspose với 256 bộ lọc, lớp concatenate nối với các đặc trưng từ khối encoder A3, hai lớp tích chập (Convolution layers) với 512 bộ lọc và kích thước của ảnh đầu vào sau khi đi qua khối B3 là (64, 64, 256) rồi di chuyển lần lượt qua khối decoder B2 và B1. Hai khối encoder B2 và B1 cũng sử dụng các lớp tương tự như khối decoder B4 và B3 nhưng có thay đổi đó là sử dụng lớp Conv2DTranspose lần lượt với 128 bộ lọc và 64 bộ lọc, lớp concatenate nối với các đặc trưng lần lượt với khối encoder A2 và A1, mỗi khối sử dụng 2 lớp tích chập lần lượt là 128 bộ lọc và 64 bộ lọc, kích thước ảnh đầu vào thay đổi sau khi đi qua khối decoder B2 là (128, 128, 128) và sau khi đi qua khối decoder B1 là (256, 256, 64), rồi cuối cùng là áp dụng một lớp đầu ra (output layers) gồm một lớp tích chập với 1 bộ lọc với kích thước kernel 1×1 và hàm kích hoạt sigmoid. Vậy sau khi đi qua mô hình Unet thì ảnh sẽ có kích thước là (256, 256, 1), với 1 là đại diện cho số lớp.

### Unet cải tiến

Mặc dù Unet đã chứng minh được hiệu quả, nhưng vẫn còn tồn tại một số hạn chế như mất mát thông tin không gian chi tiết trong quá trình down-sampling và hiện tượng overfitting. Do đó, nhóm đã cải tiến Unet để khắc phục những hạn chế và nâng cao độ chính xác. Nhóm đã thêm cơ chế Attention Mechanism [11], như tác giả đã trình bày trong bài báo của mình thì cơ chế này hiểu đơn giản là chúng sẽ loại bỏ các vùng không quan trọng và làm nổi bật các vùng quan trọng trong ảnh, giúp mô hình tập trung vào các vùng quan trọng với hình dạng và kích thước khác nhau. Trong mô hình Unet cải tiến, nhóm sẽ sử dụng cơ chế Attention Mechanism vào các khối decoder để làm nổi bật các vùng quan trọng trong skip connection và loại bỏ các vùng không quan trọng, từ đó sẽ giúp mô hình cải thiện được độ chính xác phân đoạn. Nhóm sử dụng thêm Dropout để ngăn chặn overfitting, dùng GroupNormalization thay BatchNormalization, giúp mô hình ổn định hơn trong khi dùng batch kích thước nhỏ. Dưới đây sẽ là hình ảnh trực quan của mô hình Unet sau khi được nhóm tinh chỉnh lại.



Hình ảnh mô hình Unet cải tiến

Trên hình 3.4 là hình ảnh trực quan của mô hình Unet cải tiến, chúng cũng có 3 phần chính tương tự như mô hình Unet là bao gồm: encoder, bottleneck, decoder. Ba phần chính này cũng có cấu trúc, nhiệm vụ tương tự như Unet, các khối encoder được đánh dấu từ A1 đến A4, bottleneck là C và decoder từ B4 đến B1.

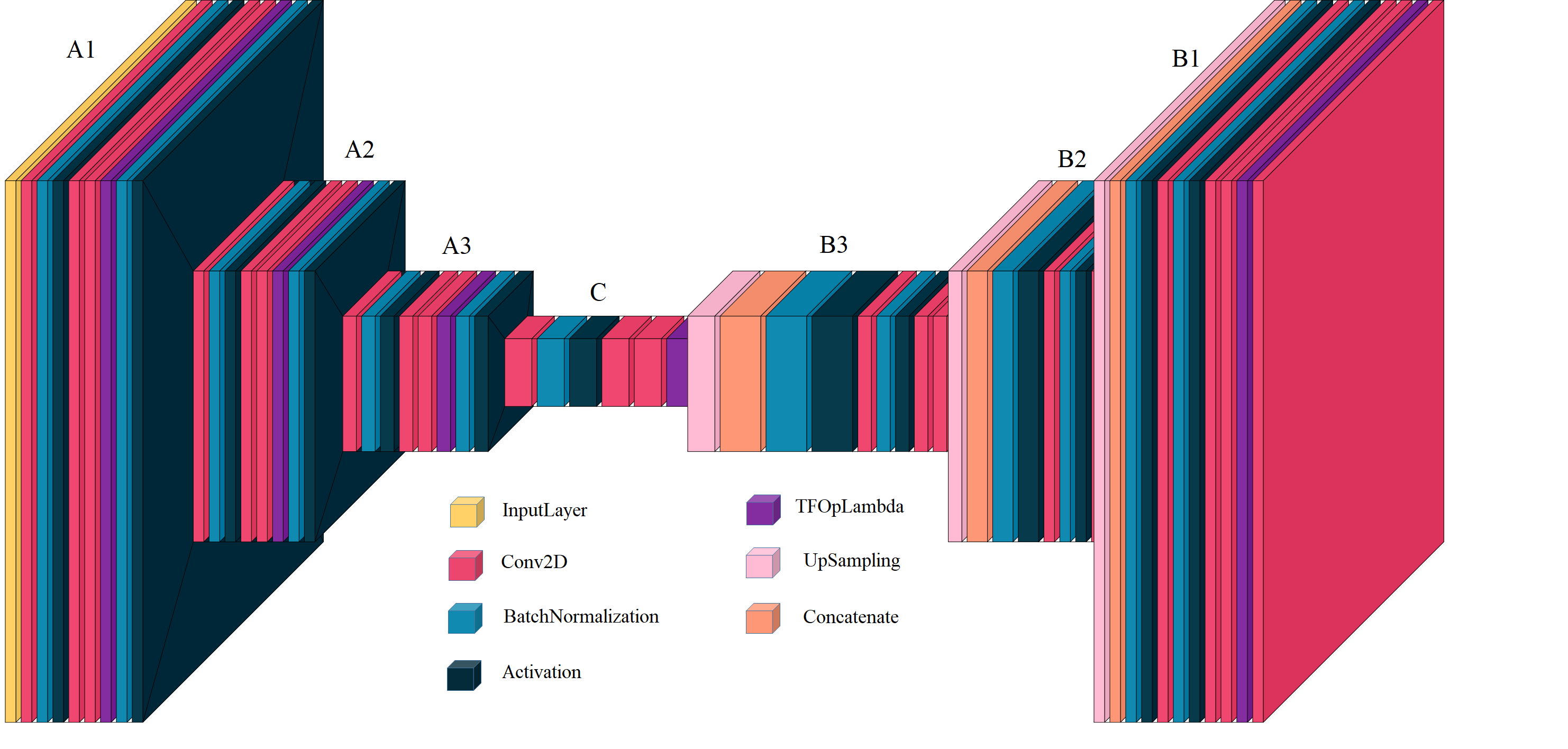
Bắt đầu là một lớp đầu vào là hình ảnh có kích thước (256, 256, 3), chúng sẽ đi vào các khối encoder từ A1 đến A4. Trong khối A1 bao gồm hai lớp tích chập (Convolution layers) với 64 bộ lọc và kích thước kernel 3×3, 2 lớp GroupNormalization có tham số là groups = 8 nghĩa là các đặc trưng của lớp sẽ được chia thành 8 nhóm và mỗi nhóm sẽ được chuẩn hóa riêng biệt, tiếp theo là 2 hàm activation ReLu và sử dụng hàm dropout sau mỗi hàm activation, cuối cùng là một lớp pooling có kích thước là 2×2. Sau khi đi qua khối A1 thì lớp đầu vào sẽ có kích thước là (128, 128, 64), rồi tiếp tục qua khối A2. Trong khối encoder A2 bao gồm 2 lớp tích chập với 128 bộ lọc với kích thước kernel 3×3, 2 lớp GroupNormalization, theo sau cũng tiếp tục là hàm activation ReLU và sử dụng hàm dropout sau mỗi lớp tích chập, cuối cùng là một lớp pooling có kích thước là 2×2, sau khi đi qua lớp pooling thì ảnh đầu vào có kích thước là (64, 64, 128) rồi chuyển qua khối A3. Trong khối encoder A3 cũng tương tự như khối A1 và A2 nhưng thay vào đó là sử dụng 2 lớp tích chập (Convolution layers) 256 bộ lọc với kích thước kernel 3×3 và sau khi đi qua khối A3 thì lớp đầu vào sẽ có kích thước là (32, 32, 256), rồi di chuyển qua khối encoder cuối cùng là A4. Trong khối A4 sử dụng các lớp tương tự như A3 nhưng thay đổi là sử dụng 2 lớp tích chập (Convolution layers) với 512 bộ lọc với kích thước kernel 3×3, sau khi đi qua khối A4 thì lớp đầu vào có kích thước là (16,16, 512). Như vậy thì sau khi đi qua 4 khối encoder thì kích thước ảnh còn lại sẽ là (16, 16, 512).

Sau khối encoder thì sẽ di chuyển đến khối bottleneck có kí hiệu là C. Trong khối C này bao gồm 2 lớp tích chập sử dụng 1024 bộ lọc với kích thích thước kernel 3×3, 2 lớp GroupNormalization với tham số groups là 8, theo sau đó là có hàm activation ReLU và dropout sau mỗi lớp tích chập, sau khi đi qua khối bottleneck thì kích thước của lớp ảnh đầu vào sẽ là (16, 16, 1024).

Cuối cùng thì lớp đầu vào sẽ di chuyển vào các khối decoder từ B4 đến B1. Trong khối decoder, lớp Conv2DTranspose được sử dụng để tăng kích thước không gian của các feature map. Sau đó kết quả được kết nối (concatenated) với các feature maps từ khối encoder thông qua hàm attention gate. Trong hàm này bao gồm 3 lớp tích chập (Convolution layers) với kernel kích thước 1×1, 2 lớp BatchNormalization, 2 lớp activation (ReLU và sigmoid. Bắt đầu từ khối decoder B4, trong khối này bao gồm 1 lớp Conv2DTranspose với 512 bộ lọc, 1 hàm attention gate, 1 lớp concatenate nối với các đặc trưng từ khối encoder A4, hai lớp tích chập (Convolution layers) với 512 bộ lọc và kích thước kernel 3×3, theo sau đó là có hàm chuẩn hóa GroupNormalization, hàm activation ReLU và dropout sau mỗi lớp tích chập. Sau khi đi qua khối decoder B4 thì ảnh sẽ có kích thước đầu ra là (32, 32, 512) rồi tiếp tục chuyển qua khối B3. Trong khối B3 cũng bao gồm các lớp tương tự như khối B4 nhưng sử dụng 256 bộ lọc trong các lớp tích chập và lớp Conv2DTranspose, lớp concatenate nối với các đặc trưng từ khối encoder A3, kích thước của ảnh đầu vào sau khi đi qua khối B3 là (64, 64, 256) rồi di chuyển lần lượt qua khối decoder B2 và B1. Hai khối encoder B2 và B1 cũng sử dụng các lớp tương tự như khối decoder B4 và B3 nhưng có thay đổi đó là sử dụng lớp Conv2DTranspose lần lượt với 128 bộ lọc và 64 bộ lọc, lớp concatenate nối với các đặc trưng lần lượt với khối encoder A2 và A1, mỗi khối sử dụng 2 lớp tích chập lần lượt là 128 bộ lọc và 64 bộ lọc, kích thước ảnh đầu vào thay đổi sau khi đi qua khối decoder B2 là (128, 128, 128) và sau khi đi qua khối decoder B1 là (256, 256, 64), rồi cuối cùng là áp dụng một lớp đầu ra (output layers) gồm một lớp tích chập với 1 bộ lọc với kích thước kernel 1×1 và hàm kích hoạt sigmoid, kích thước ảnh đầu ra là (256, 256, 1). Vậy sau khi sử dụng mô hình Unet cải tiến thì ảnh đầu ra sẽ có kích thước là (256, 256, 1).

### ResUnet

Sau khi nhóm đã triển khai xong mô hình Unet và Unet cải tiến, nhóm tiếp tục triển khai mô hình là ResUnet. ResUnet là mô hình kết hợp giữa Unet và Resnet, mô hình này sử dụng thêm các khối Residual để cải thiện hiệu suất và tăng sự ổn định cho mô hình như đã trình bày ở chương 2. Kiến trúc của mô hình được dựa [12] theo bao gồm ba phần chính: encoder, bottleneck, decoder. Các khối encoder được kí hiệu từ A1 đến A3, bottleneck là C và các khối decoder có kí hiệu từ B3 đến B1 tương ứng với Hình 3.5 sau:

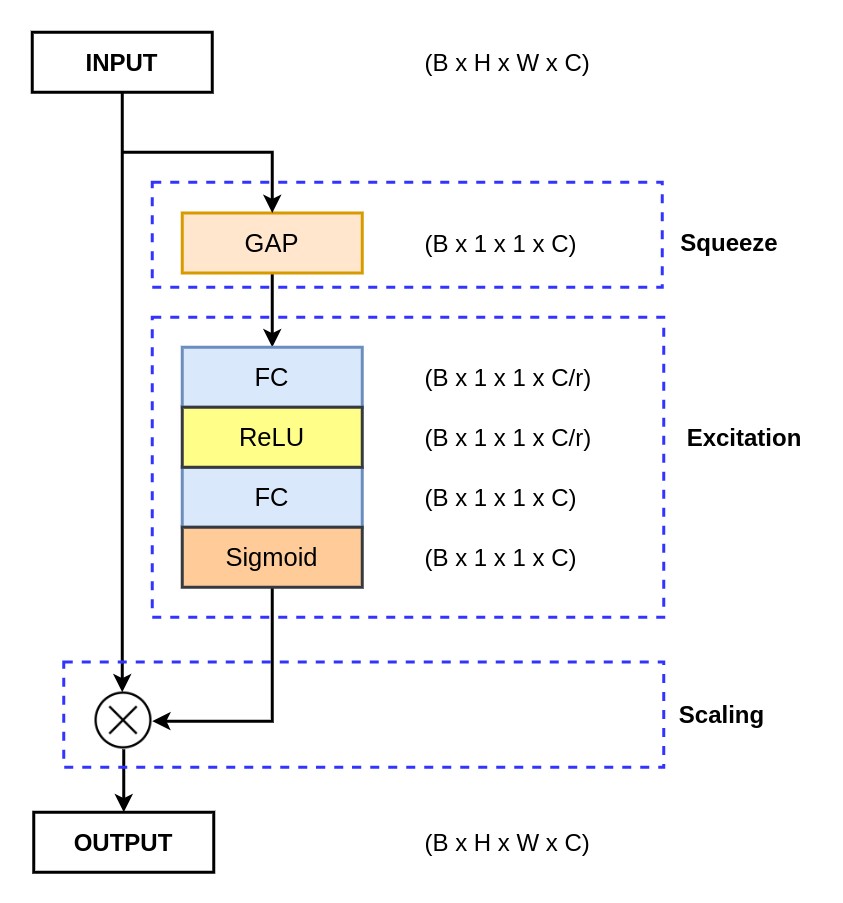


Hình ảnh mô hình ResUnet [12]

Bắt đầu là một lớp input layers có kích thước là (256, 256, 3), lớp đầu vào sẽ đi qua các khối encoder từ A1 đến A3. Bắt đầu với khối A1 là sử dụng 1 lớp tích chập 64 bộ lọc với kích thước kernel 3×3, theo sau là hàm activation ReLU với hàm chuẩn hóa BatchNormalization, sử dụng thêm 2 lớp tích chập 64 bộ lọc với kích thước kernel lần lượt là 3×3 và 1×1, và tiếp tục sử dụng hàm activation ReLU với hàm chuẩn hóa BatchNormalization sau 2 lớp tích chập. Sau khi đi qua khối A1 thì ảnh sẽ có kích thước đầu ra là (256, 256, 64) và tiếp tục chuyển qua khối A2 và A3. Trong 2 khối encoder A2 và A3 đều sử dụng residual block với bước nhảy (stride) là 2, trong residual block thực hiện hàm chuẩn hóa BatchNormalization và hàm activation ReLU, sử dụng 1 lớp tích chập với số lượng bộ lọc và stride tương ứng, rồi tiếp tục thực hiện chuẩn hóa BatchNormalization và hàm activation ReLU lần 2, sử dụng thêm 1 lớp tích chập với số lượng bộ lọc tương ứng và stride bằng 1, sau đó tạo ra kết nối tắt (shortcut connection) bằng 1 lớp tích chập Conv2D với filter và stride tương ứng trên đầu vào ban đầu, cuối cùng là cộng đầu ra của hai quá trình trên để tạo ra shortcut connection. Hai khối encoder A2 và A3 sử dụng residual block với số bộ lọc lần lượt là 128 và 256, sau khi đi qua hai khối thì ảnh sẽ có kích thước đầu ra tương ứng là (128, 128, 128) và (64, 64, 256). Vậy lớp đầu vào sau khi đi qua 3 khối encoder sẽ có kích thước là (64, 64, 256) rồi tiếp tục di chuyển qua khối bottleneck có kí hiệu là C. Trong khối bottleneck sẽ sử dụng residual block với 512 bộ lọc và stride bằng 2, khi lớp đầu vào đi qua bottleneck thì kích thước sẽ thay đổi là (32, 32 , 512) rồi tiếp tục di chuyển qua 3 khối decoder từ B3 đến B1. Bắt đầu với khối decoder B3 sẽ bao gồm 1 lớp UpSampling2D kích thước 2×2 để tăng kích thước không gian, 1 lớp Concatenate với đầu ra từ A3 và cuối cùng là sử dụng residual block với 256 bộ lọc và stride bằng 1. Sau khi đi qua khối B3 thì ảnh sẽ có đầu ra là (64, 64, 256) rồi di chuyển lần lượt qua 2 khối decoder B2 và B1. Trong 2 khối B2 và B1 chúng có cấu trúc tương tự như khối B3 nhưng chúng Concatenate với đầu ra tương ứng là A2, A1 và sử dụng residual block lần lượt với 128 và 64 bộ lọc. Kích thước của đầu ra sau khi đi qua B2 và B1 sẽ là (128, 128, 128) và (256, 256, 64) rồi cuối cùng là áp dụng một lớp đầu ra (output layers) gồm một lớp tích chập với 1 bộ lọc với kích thước kernel 1×1 và hàm kích hoạt sigmoid, kích thước ảnh đầu ra là (256, 256, 1). Như vậy thì lớp ảnh đầu vào sau khi đi qua mô hình ResUnet sẽ có kích thước là (256, 256, 1).

### ResUnet++

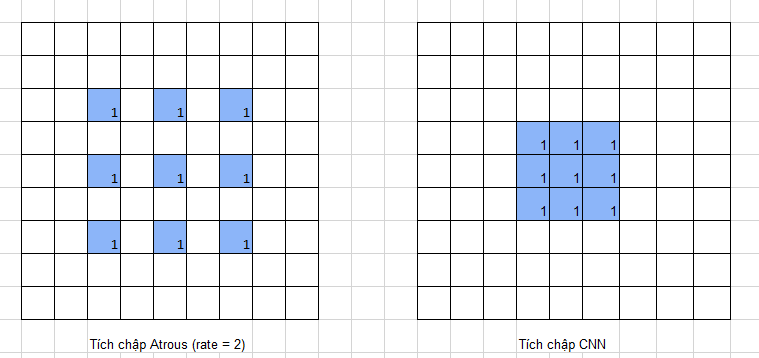
Mô hình ResUnet++ là mô hình được cải tiến từ ResUnet, mô hình này được áp dụng thêm một số kĩ thuật để cải thiện khả năng phân đoạn. ResUnet++ thêm một số kĩ thuật như squeeze và excitation blocks, attention blocks và Atrous Spatial Pyramidal Pooling (ASPP) [12]. Các khối squeeze và excitation blocks được thêm vào để cung cấp khả năng hiệu chỉnh kênh linh hoạt, nâng cao khả năng tập trung của mô hình vào các đặc trưng quan trọng. Squeeze và excitation blocks bao gồm có 3 phần: Squeeze, excitation và scaling như hình vẽ dưới đây:



Squeeze và excitation blocks [14]

Hàm squeeze được sử dụng để trích xuất thông tin từ mỗi kênh của feature map bằng các sử dụng Global Average Pooling (GAP) để nén toàn bộ thông tin từ mỗi kênh của feature map thành một giá trị duy nhất. Lớp “squeeze” trong hình 3.6 nén từ (B × H × W × C) thành (B × 1 × 1 × C) với B là kích thước batch, H là chiều cao, W là chiều rộng và C là số kênh của feature map. Tiếp theo là tới phần “Excitation”, trong phần này thì vector (B × 1 × 1 × C) được đưa qua các lớp fully connected (FC), ReLU và sigmoid để tạo ra trọng số cuối cùng. Vector trọng số mang ý nghĩa là tầm quan trọng của từng kênh, sau đó chúng sẽ qua phần “Scaling”. Trong scaling thực hiện việc vector trọng số cuối cùng sẽ nhân với feature map ban đầu để tạo ra bản đồ đặc trưng mới với các kênh đã được thiết lập lại trọng số, giúp mạng chú ý đến các đặc trưng quan trọng và bỏ qua các đặc trưng không quan trọng.

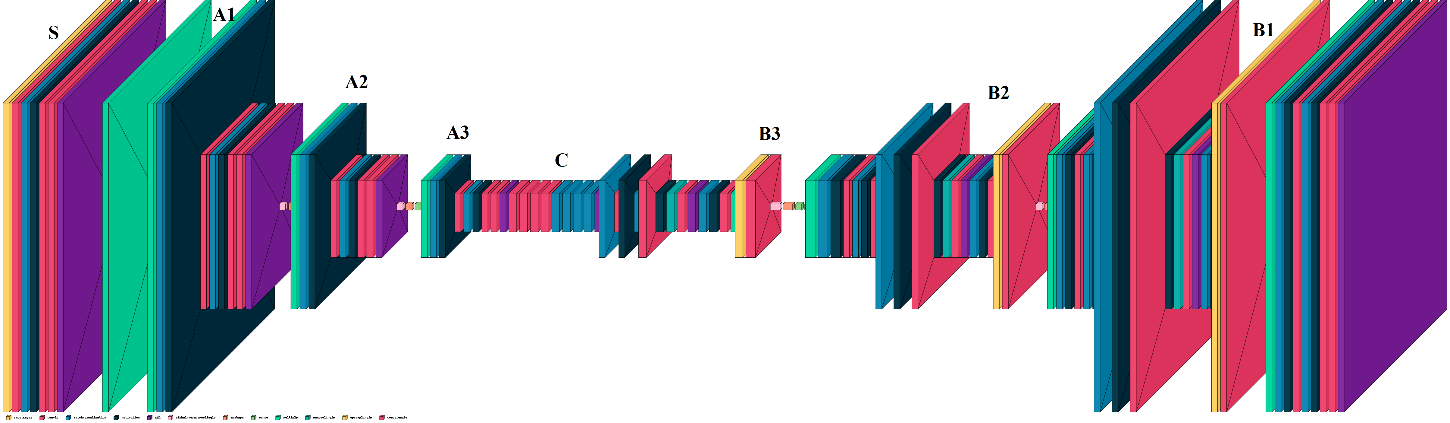
Khối ASPP (Atrous Spatial Pyramidal Pooling) là một kỹ thuật tiên tiến được sử dụng trong thị giác máy tính, mạng nơ-ron tích chập (CNN), để xử lý ảnh đầu vào với kích thước và tỉ lệ khung hình khác nhau giúp mô hình thu thập thông tin ngữ cảnh ở nhiều tỉ lệ khác nhau, từ đó cải thiện khả năng nhận diện và phân đoạn các đối tượng trong ảnh. Khối ASPP sử dụng Atrous Convolution để trích lọc đặc trưng, tích chập Atrous có những điểm mạnh hơn so với tích chập CNN đó là tầm nhìn (field of view) được mở rộng hơn, không làm giảm chiều của feature map quá sâu mà vẫn giữ được số lượng tham số và chi phí tính toán tương đương với tích chập CNN. Hình ảnh dưới đây sẽ so sánh giữa tích chập Atrous và tích chập CNN.



Tích chập Atrous và tích chập CNN

Từ hình 3.7 ở trên thì ta có thể thấy rằng tích chập Atrous mở rộng vùng nhận thức (receptive field) cách nhau một ô, dù bản chất vẫn là bộ lọc 3×3 nhưng hoạt động trên vùng 5×5. Trong khi đó tích chập CNN có kích thước receptive field (vùng nhận thức) trùng với kích thước bộ lọc. Kết quả đầu ra cho thấy tích chập Atrous đã tăng kích thước feature map tăng lên thành 5×5. Như vậy, tích chập Atrous có tác dụng là mở rộng vùng nhận thức (receptive field) giúp mô hình học được các đặc trưng ở phạm vi rộng hơn mà không làm giảm độ phân giải và hiệu quả tính toán. ASPP là tập hợp gồm nhiều tích chập Atrous với kích thước khác nhau được thực hiện đồng thời trên 1 feature map và vì nhiều tích chập Atrous với kích thước khác nhau chồng lên tạo ra một kim tự tháp bộ lọc Atrous. Khi tỉ lệ rate của Atrous tăng, phạm vi của bộ lọc trên feature map cũng mở rộng, giúp mô hình học được bối cảnh tổng quát tốt hơn. Ngược lại, các bộ lọc có kích thước nhỏ giúp mô hình học được bối cảnh cục bộ hiệu quả hơn.

Bên cạnh đó, mô hình ResUnet++ còn sử dụng thêm khối Residual block cho cả phần ecoder và decoder, kỹ thuật này giúp cho mô hình truyền thông tin giữa các lớp hiệu quả hơn với phương pháp là bỏ qua một hoặc nhiều lớp và giúp giải quyết vấn đề gradient biến mất, cho phép xây dựng các mạng sâu hơn. Khối attention được sử dụng trong phần decoder của ResUnet++, giúp mô hình tập trung vào các phần quan trọng của ảnh bằng thông tin từ skip connections trong phần encoder, từ đó cải thiện khả năng phân tích hình ảnh của mô hình. Hình 3.8 dưới đây sẽ là hình ảnh trực quan của mô hình ResUnet++ của nhóm.



Hình ảnh mô hình ResUnet++ [12]

Từ mô hình ResUnet++ đã được trực quan hóa như hình 3.8 ở trên thì mô hình gồm có 3 phần chính: encoder, bottleneck, decoder. Các khối encoder được đánh dấu là từ A1 đến A3, bottleneck là C và các khối decoder được đánh dấu từ B3 đến B1.

Lớp đầu vào của mô hình sẽ là một ảnh có kích thước (256, 256, 3), không giống như mô hình Unet và ResUnet, ResUnet++ bắt đầu với 1 khối Stem block, khối này được sử dụng để giảm kích thước không gian của ảnh đầu vào và trích xuất các đặc trưng quan trọng. Stem block bắt đầu với một lớp tích chập 3×3, theo sau đó là có hàm chuẩn hóa BatchNormalization và hàm activation ReLU. Tiếp theo, một lớp tích chập 3×3 khác được áp dụng, kèm theo đó là một shortcut connection gồm một lớp tích chập 1×1 và hàm chuẩn hóa BatchNormalization, tất cả các lớp tích chập trong khối stem block đều sử dụng 16 bộ lọc. Cuối cùng là một khối squeeze và excitation được sử dụng để cải thiện đặc trưng đầu ra. Trong khối squeeze và excitation bao gồm 1 lớp Global Average Pooling (GAP), tiếp theo là lớp Reshape, theo sau là 2 lớp Dense, lớp Dense thứ nhất sử dụng hàm activation ReLU và lớp Dense thứ hai sử dụng hàm activation Sigmoid, cuối cùng là 1 lớp Multiply. Sau khi đi qua khối stem block thì ảnh sẽ có kích thước là (256, 256, 16), rồi di chuyển qua khối encoder A1.

Trong encoder A1 sử dụng ResNet Block, khối này bắt đầu với 1 lớp chuẩn hóa BatchNormalization, hàm activation ReLU và một lớp tích chập 3×3, rồi chúng lặp lại 1 lần nữa. Kết hợp thêm 1 lớp shortcut connection bao gồm 1 lớp tích chập 1×1 với hàm chuẩn hóa BatchNormalization và cuối cùng là một khối squeeze và excitation cũng được áp dụng tương tự như Stem Block. Trong khối A1 sử dụng ResNet Block với 32 bộ lọc và áp dụng bước nhảy (stride) bằng 2, sau khi đi qua khối encoder A1 thì lớp ảnh đầu vào sẽ có kích thước là (128, 128, 32) rồi chuyển qua khối A2.

Trong khối encoder A2 cũng sử dụng ResNet Block tương tự như encoder A1 nhưng với 64 bộ lọc và bước nhảy bằng 2. Lớp ảnh đầu vào sau khi đi qua khối A2 sẽ có kích thước là (64, 64, 64), rồi tiếp tục di chuyển qua khối encoder A3. Trong khối encoder A3 cũng tương tự như khối A1 và A2 những với số lượng bộ lọc là 128 và bước nhảy bằng 2. Sau khi đi qua khối encoder A3, ảnh sẽ có kích thước là (32, 32, 128). Vậy lớp ảnh đầu vào sau khi đi qua các khối encoder có kích thước là (32, 32, 128), rồi đi tới lớp bottleneck C.

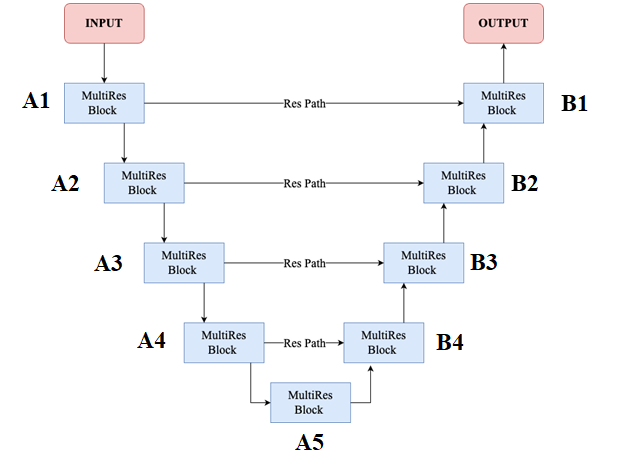
Trong lớp Bottleneck C là 1 khối ASPP với 256 bộ lọc, trong khối này bao gồm 4 lớp tích chập với tỉ lệ rate khác nhau lần lượt là 6, 12, 18 và 1, sau mỗi lớp tích chập là 1 lớp chuẩn hóa BatchNormalization kèm theo. Đầu ra của 4 lớp tích chập sẽ được kết hợp lại với nhau bằng cách dùng hàm Add và cuối cùng là áp dụng 1 lớp tích chập 1×1. Sau khi đi qua khối Bottleneck C thì ảnh đầu vào sẽ có kích thước là (32, 32, 256), rồi tiếp tục đi qua 3 khối decoder.

Khối decoder thứ nhất được đánh dấu là B3, trong B3 sử dụng 1 khối attention, 1 lớp UpSampling2D, 1 lớp Concatenate và 1 khối ResNet Block. Bắt đầu của khối attention là 1 lớp chuẩn hóa BatchNormalization, kết hợp với activation ReLU, 1 lớp tích chập 3×3 và MaxPooling 2x2. Chúng tiếp tục lặp lại theo thứ tự là lớp chuẩn hóa BatchNormalization, activation ReLU và 1 lớp tích chập 3×3. Theo sau khối attention là 1 lớp UpSampling2D 22. Sau đó chúng kết nối (concatenated) với encoder tương ứng thông qua skip connection. Cuối cùng là sử dụng 1 khối ResNet block với 128 bộ lọc và lớp ảnh đầu vào sau khi đi qua khối decoder B3 là (64, 64, 128) rồi di chuyển qua khối decoder B2 và B1.

Trong 2 khối decoder B2 và B1 cũng tương tự như khối B3, chỉ thay đổi là sử dụng ResNet Block với số bộ lọc lần lượt là 64 và 32. Lớp ảnh sau khi đi qua khối B2 là (128, 128, 64) và qua khối B1 sẽ là (256, 256, 32), rồi di chuyển qua lớp cuối cùng là lớp output. Trong lớp output sử dụng 1 khối ASPP 16 bộ lọc, sau đó áp dụng 1 lớp tích chập 1x1 với 1 bộ lọc và kết hợp với activation Sigmoid để ra kết quả cuối cùng. Lớp ảnh đầu vào sau khi đi qua lớp output sẽ là (256, 256, 1). Vậy từ lớp ảnh đầu vào có kích thước là (256, 256, 3) sau khi đi qua mô hình ResUnet++ sẽ có kích thước là (256, 256, 1).

### MultiResUnet

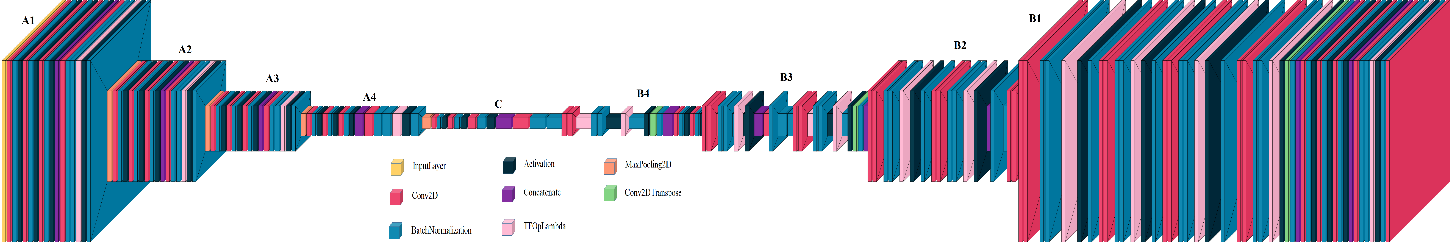
Mô hình MultiResUnet được phát triển dựa trên mô hình Unet, chúng sử dụng các khối MultiRes Block và Res Path như trong hình ảnh trực quan dưới đây.



MultiResUnet [6]

Như hình 3.9 ở trên thì ta thấy rằng MultiRes block được sử dụng để thay thế cho Convolution Block trong mô hình Unet. MultiRes block có một chuỗi 3 lớp tích chập nối tiếp nhau có kích thước là 3×3 và mỗi lớp tích chập đều có hàm chuẩn hoá BatchNormalization và activation ReLU. Cuối cùng là kết hợp đầu ra của cả ba lớp tích chập lại với nhau.

Res Path được sử dụng để thay cho skip connection trong Unet. Theo tác giả [6], tồn tại một khoảng cách giữa ngữ nghĩa các đặc trưng của encoder và decoder. Vì thế, sử dụng skip connection có thể không hiệu quả, nên tác giả đã sử dụng Res Path để thay thế cho chúng. Các đặc trưng của phần encoder được di chuyển qua Res Path, sau đó đầu ra của Res Path được kết hợp với các đặc trưng của phần decoder. Trong Res Path, ảnh đầu vào được chuyển qua một lớp tích chập 3×3, theo sau đó là một lớp chuẩn hoá BatchNormalization và activation ReLU. Bên cạnh đó, một nhánh khác của ảnh đầu vào được đưa qua lớp tích chập 1×1 và một lớp chuẩn hoá BatchNormalization. Sau đó, kết quả của hai nhánh sẽ được kết hợp lại rồi đưa qua hàm activation ReLU và lớp chuẩn hoá BatchNormalization. Quá trình này sẽ được lặp tuỳ theo giá trị tham số mà ta truyền vào, chúng sẽ thực hiện được nhiệm vụ là thu hẹp khoảng cách ngữ nghĩa giữa các đặc trưng của encoder và decoder, từ đó cả thiện được độ chính xác của mô hình. Để có thể hình dung về mô hình kỹ hơn thì nhóm thực hiện trực quan hoá chúng trong hình ảnh dưới đây



Hình ảnh mô hình MultiResUnet [6]

Mô hình MultiResUnet bao gồm có 4 khối encoder được đánh dấu từ A1 đến A4, Bottleneck là C và 4 khối decoder đánh dấu từ B4 đến B1. Đầu tiên là có một ảnh đầu vào có kích thước (256, 256, 3), sau đó đi qua các khối encoder. Trong các khối encoder bao gồm MultiRes Block là một chuỗi các lớp tích chập có kích thước khác nhau kết hợp với hàm chuẩn hóa BatchNormalization và activation ReLU như đã mô tả ở trên, tiếp theo là Res Path và cuối cùng là 1 lớp Max pooling 2×2. Trong 4 khối encoder đều có cấu trúc giống nhau chỉ thay đổi số lượng bộ lọc và tham số length, tham số length quyết định số lần quá trình lặp trong Res Path. Lớp ảnh đầu vào sẽ di chuyển vào khối encoder thứ nhất là A1, khối A1 sử dụng MultiRes Block với 32 bộ lọc và 4 lớp Res Path. Sau khi đi qua khối A1, ảnh sẽ có kích thước là (128, 128, 32) rồi di chuyển sang khối A2. Trong khối A2 áp dụng MultiRes Block với 64 bộ lọc và 3 lớp Res Path, lớp đầu ra sẽ có kích thước là (64, 64, 64), rồi di chuyển qua khối A3 và A4. Trong 2 khối A3 và A4 đều sử dụng MultiRes Block lần lượt với 128 và 256 bộ lọc, áp dụng lần lượt 2 lớp và 1 lớp cho phần Res Path. Ảnh sau khi đi qua khối A3 có kích thước là (32, 32, 128) và sau khi đi qua A4 có kích thước là (12, 16, 256). Vậy tóm lại, lớp ảnh đầu vào sau khi đi qua 4 khối encoder sẽ có kích thước là (16, 16, 256), rồi di chuyển qua khối bottleneck.

Trong khối bottleneck C sử dụng MultiRes Block với 512 bộ lọc và ảnh đầu ra có kích thước là (16, 16, 512) rồi tiếp tục di chuyển sang các khối decoder từ B4 đến B1.

Đầu tiên, mỗi khối decoder sử dụng 1 lớp tích chập chuyển vị (Transpose Convolution) 2 x 2 với bước nhảy (stride) là 2, sau đó chúng sẽ được kết hợp với đầu ra của phần Res Path tương ứng và cuối cùng là áp dụng một khối MultiRes Block với số lượng bộ lọc khác nhau tùy theo từng khối decoder. Sau khi đi qua khối bottleneck thì ảnh sẽ di chuyển qua khối decoder thứ nhất là khối B4. Trong khối B4 sử dụng MultiRes Block với 256 bộ lọc và kích thước đầu ra sẽ là (32, 32, 256), rồi di chuyển qua khối B3. Khối B3 áp dụng 128 filter cho MultiRes Block và lớp ảnh đầu ra có kích thước là (64, 64, 128), rồi di chuyển lần lượt qua khối B2 và B1. Trong 2 khối B3 và B4 áp dụng lần lượt 64 và 32 bộ lọc cho khối MultiRes Block và ảnh đầu ra sẽ lần lượt có kích thước là (128, 128, 64) và (256, 256, 32). Vậy ảnh đầu ra sau khi đi qua bồn khối decoder từ B4 đến B1 sẽ là (256, 256, 32). Đầu ra của mô hình là một ảnh có kích thước (256, 256, 1), khi áp dụng 1 lớp tích chập Conv2D 1×1 với 1 bộ lọc và hàm activation sigmoid. Vậy lớp ảnh đầu vào sau khi đi qua mô hình MultiResUnet sẽ có kích thước là (256, 256, 1).

## Phương pháp đánh giá

Để đánh giá các phương pháp đề xuất có kết quả tốt hay chưa tốt, chúng tôi sử dụng một số thang đo đánh giá dựa trên các mô hình nghiên cứu liên quan đã được các nhà nghiên cứu trên thế giới sử dụng như: Dice Score và Jaccard Index. Cụ thể các thang đo này được trình bày chi tiết như sau:

### Dice Score

Dice Score là một chỉ số dùng để đo lường mức độ tương đồng giữa hai tập hợp, đặc biệt là trong các bài toán phân đoạn ảnh. Dice Score được tính dựa trên sự so sánh giữa vùng dự đoán của mô hình và vùng thực tế. Công thức của Dice Score được biểu diễn như sau [15]:

𝐷 𝑖 𝑐 𝑒 =

Trong đó:

* + A là tập hợp các điểm ảnh thuộc vùng dự đoán.
  + B là tập hợp các điểm ảnh thuộc vùng thực tế (Ground Truth).
  + ∣ 𝐴 ∩ 𝐵 ∣ là số lượng phần tử chung giữa hai tập hợp.

Dice Score dao động từ 0 đến 1, với giá trị 1 biểu thị rằng hai tập hợp hoàn toàn trùng khớp. Một Dice Score cao cho thấy sự tương đồng lớn giữa vùng dự đoán của mô hình và vùng thực tế trên ảnh MRI. Điều này chỉ ra rằng mô hình có khả năng phân đoạn chính xác các khu vực quan trọng, chẳng hạn như các vùng não, với mức độ chính xác cao. Ngược lại, giá trị thấp của Dice Score có thể chỉ ra rằng mô hình có sự sai lệch đáng kể trong việc phân đoạn hoặc không nhận diện đúng các vùng cần thiết. Vì vậy, Dice Score là một công cụ quan trọng trong việc đánh giá và cải thiện hiệu quả phân đoạn ảnh..

### Jaccard Index

Jaccard Index hay còn gọi là Intersection over Union (IoU) dùng để đo lường sự tương đồng giữa hai tập hợp dữ liệu. Chỉ số này so sánh mức độ trùng khớp giữa vùng được phân đoạn bởi mô hình và vùng thực sự. Công thức tính toán Jaccard được biểu diễn như sau [16]:

J(A, B) =

Trong đó:

* A là tập hợp các điểm ảnh thuộc vùng được mô hình phân đoạn.
* B là tập hợp các điểm ảnh thuộc vùng thực tế (ground truth).
* |A ∩ B| là số lượng điểm ảnh chung giữa hai tập hợp, tức là số điểm ảnh mà mô hình dự đoán đúng.
* |A ∪ B| là tổng số điểm ảnh thuộc cả hai tập hợp (gồm cả những điểm ảnh được dự đoán đúng, sai và thiếu).

Jaccard Index cao cho thấy hai tập hợp dữ liệu (vùng phân đoạn của mô hình và vùng thực tế) có mức độ trùng khớp lớn. Cụ thể, chỉ số này tính toán tỷ lệ giữa số lượng điểm ảnh chung giữa hai tập hợp và tổng số điểm ảnh thuộc cả hai tập hợp. Khi giá trị Jaccard Index gần bằng 1, điều này chứng tỏ rằng mô hình đã phân đoạn chính xác và đầy đủ các vùng quan trọng, với sự trùng khớp cao giữa dự đoán và thực tế. Ngược lại, giá trị thấp cho thấy sự thiếu sót trong phân đoạn hoặc các lỗi trong nhận diện vùng, do đó làm giảm hiệu quả tổng thể của mô hình. Do đó, Jaccard Index là một công cụ hữu ích để đánh giá và cải thiện chất lượng của các phương pháp phân đoạn ảnh.

# THÍ NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ.

## Tập dữ liệu và cấu hình phần cứng.

Bộ dữ liệu được thu thập từ Figshare [17] chứa 233 bênh nhân, gồm có tất cả 2296 ảnh não MRI có định dạng là PNG với kích thước là 512×512 và tương ứng với từng ảnh là mask của chúng. Để đảm bảo quá trình huấn luyện, xác thực và kiểm thử mô hình diễn ra hiệu quả, bộ dữ liệu này đã được chia thành các tập sau:

* **Train**: 1377 ảnh MRI và 1377 ảnh mask. Tập này được sử dụng để huấn luyện mô hình, giúp mô hình học các đặc trưng và mối quan hệ giữa ảnh MRI và mask.
* **Valid**: 459 ảnh MRI và 459 ảnh mask. Tập này được sử dụng để xác thực mô hình trong quá trình huấn luyện, giúp theo dõi và điều chỉnh các siêu tham số nhằm tránh hiện tượng overfitting.
* **Test**: 459 ảnh MRI và 459 ảnh mask. Tập này được sử dụng để đánh giá hiệu suất cuối cùng của mô hình sau khi quá trình huấn luyện hoàn tất, đảm bảo mô hình có khả năng tổng quát hóa tốt trên dữ liệu mới.

Bảng 4.1 là trực quan hóa bộ dữ liệu mà nhóm sử dụng để huấn luyện các mô hình semantic segmentation:

Trực quan hóa bộ dữ liệu

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tên dữ liệu | Mẫu 1 | Mẫu 2 | Mẫu 3 | Mẫu 4 |
| MRI scans |  |  |  |  |
| Mask |  |  |  |  |

Nhóm sử dụng ngôn ngữ Python trên môi trường Google Colab để thực thi chương trình. Colab cấp phát bộ nhớ là 12.7 GB RAM, 15 GPU RAM, 78.2 GB ổ cứng. Việc sử dụng Google Colab có những lý do sau:

* **Tiện lợi:** Google Colab là một nền tảng trực tuyến, dễ dàng truy cập và sử dụng mà không cần cài đặt phần mềm phức tạp.
* **Miễn phí:** Colab cung cấp tài nguyên tính toán mạnh mẽ miễn phí, phù hợp cho các dự án nghiên cứu và thử nghiệm.
* **Tài nguyên GPU:** Colab cung cấp GPU mạnh mẽ, giúp tăng tốc quá trình huấn luyện mô hình Deep Learning, giảm thời gian xử lý so với việc sử dụng CPU.
* **Khả năng tích hợp:** Colab tích hợp tốt với các thư viện và công cụ phổ biến trong lĩnh vực Machine Learning và Deep Learning như TensorFlow, Keras, PyTorch, và OpenCV.

Ngoài ra, nhóm đã phát triển một giao diện web dựa trên **Streamlit** để người dùng có thể dễ dàng thực hiện các thử nghiệm dự đoán với mô hình đã được triển khai. Giao diện này giúp người dùng tương tác trực quan với mô hình, cho phép tải lên hình ảnh MRI và nhận về kết quả phân đoạn nhanh chóng và tiện lợi.

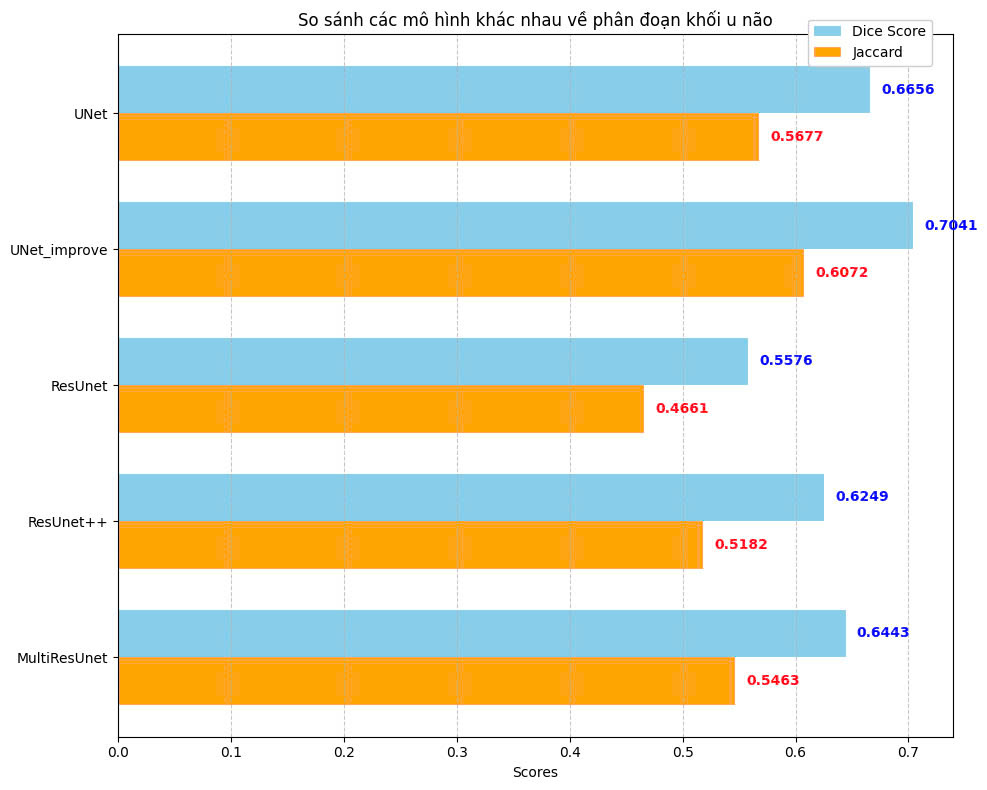
## Thí nghiệm và đánh giá kết quả.

Trong quá trình huấn luyện mô hình, nhóm sử dụng các siêu tham số như được trình bày trong Bảng 4.2. Learning rate được đặt là 0.001 để cân bằng giữa tốc độ học và độ ổn định. Batch size là 16 giúp cải thiện hiệu suất tính toán và duy trì độ chính xác. Thuật toán tối ưu Adam được chọn vì khả năng tăng tốc độ hội tụ và độ ổn định. Số epochs là 50, đảm bảo mô hình có đủ thời gian học mà không bị overfitting. Nhóm sử dụng dice loss để tối ưu hóa độ chính xác trong phân đoạn, cùng với Dice coefficient làm chỉ số đánh giá hiệu suất. Sự lựa chọn cẩn thận các siêu tham số này nhằm tối ưu hóa hiệu quả huấn luyện và khả năng tổng quát của mô hình.

Các tham số để huấn luyện mô hình.

|  |  |
| --- | --- |
| Tên tham số | Giá trị |
| Learning Rate | 0.001 |
| Batch Kích thước | 16 |
| Optimizer | Adam |
| Epochs | 50 |
| Loss Funtion | dice\_loss |
| Metrics | Dice coefficient |

Sau khi train xong các mô hình Semantic Segmentation mà nhóm đã đề xuất ở phần 3.2, nhóm sẽ sử dụng các phương pháp đánh giá đã mô tả ở phần 3.3 để đánh giá các mô hình bằng biểu đồ trực quan và số liệu cụ thể như bên dưới đây.



Dice Score và Jaccard của tất cả các mô hình.

Thông qua kết quả ở Hình 4.1, ta có thể thấy UNet gốc có Dice Score là 0.6656 và Jaccard Index là 0.5677. Điều này cho thấy UNet có khả năng phân đoạn tương đối tốt, tuy nhiên hiệu suất tổng thể vẫn còn hạn chế. Mặc dù mô hình có thể nhận diện đúng các vùng quan trọng, nhưng mức độ trùng khớp giữa vùng phân đoạn dự đoán và vùng thực tế vẫn chưa đạt được độ chính xác cao. Các giá trị này chỉ ra rằng UNet gốc có thể tạo ra một số dự đoán sai, dẫn đến độ chính xác tổng thể chưa hoàn thiện.

UNet cải tiến đạt được kết quả tốt nhất trong tất cả các mô hình với Dice Score là 0.7041 và Jaccard Index là 0.6072. Việc cải tiến mô hình bằng cách thêm cơ chế Attention Gate vào các Decoder Blocks và áp dụng GroupNormalization cùng Dropout đã giúp tăng cường khả năng của mô hình trong việc tập trung vào các vùng quan trọng và giảm thiểu nhiễu. Các cải tiến này không chỉ cải thiện độ chính xác mà còn giúp mô hình phân đoạn các vùng nhỏ hoặc phức tạp một cách hiệu quả hơn, do đó nâng cao cả Dice Score và Jaccard Index.

ResUnet có kết quả thấp hơn rõ rệt, với Dice Score chỉ đạt 0.5576 và Jaccard Index là 0.4661. Điều này cho thấy mô hình này gặp khó khăn trong việc phân đoạn chính xác các vùng quan trọng trên ảnh não. Mặc dù có thể nhận diện một số vùng, nhưng ResUnet dễ bỏ sót hoặc dự đoán sai các vùng cần phân đoạn, dẫn đến hiệu suất tổng thể không đạt yêu cầu.

ResUnet++ có sự cải thiện so với ResUnet, với Dice Score là 0.6249 và Jaccard Index là 0.5182. Những cải tiến này cho phép ResUnet++ nhận diện và phân đoạn chính xác hơn so với ResUnet gốc. Tuy nhiên, các chỉ số này vẫn thấp hơn so với UNet cải tiến, cho thấy hiệu suất tổng thể vẫn chưa thể đạt được mức tối ưu như UNet cải tiến.

MultiResUnet nổi bật với Dice Score là 0.6442 và Jaccard Index là 0.5462, cho thấy mô hình này có hiệu suất cao hơn ResUnet và ResUnet++, đặc biệt trong việc phân đoạn chính xác các vùng dương tính. Tuy nhiên, mô hình này vẫn chưa vượt qua được UNet cải tiến. Sự khác biệt này có thể xuất phát từ việc MultiResUnet mặc dù tránh được các lỗi dương tính giả, nhưng vẫn chưa hoàn thiện trong việc nhận diện tất cả các vùng cần phân đoạn, dẫn đến hiệu suất tổng thể thấp hơn so với UNet cải tiến.

Tóm lại, UNet cải tiến vượt trội hơn tất cả các mô hình khác nhờ các cải tiến về kiến trúc, giúp tăng cường khả năng phân đoạn chính xác và hiệu quả, đặc biệt trong việc nhận diện các vùng phức tạp hoặc nhỏ trên ảnh MRI não. Qua các chỉ số đánh giá, có thể thấy rằng mô hình UNet cải tiến đạt hiệu suất tốt nhất trong việc phân đoạn ảnh não.

Kết quả dự đoán của các mô hình.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| STT | MRI Scan | Ground Truth | Unet | Unet cải tiến | ResUnet | ResUnet++ | MultiResUnet |
| 1. |  |  |  |  |  |  |  |
| 2. |  |  |  |  |  |  |  |
| 3. |  |  |  |  |  |  |  |
| 4. |  |  |  |  |  |  |  |
| 5. |  |  |  |  |  |  |  |
| 6. |  |  |  |  |  |  |  |
| 7. |  |  |  |  |  |  |  |
| 8. |  |  |  |  |  |  |  |
| 9. |  |  |  |  |  |  |  |
| 10. |  |  |  |  |  |  |  |

Dựa trên các tiêu chí đánh giá, tất cả các mô hình đều có hiệu suất tốt trong bài toán phân đoạn khối u não từ tập dữ liệu test. Tuy nhiên, kết quả từ Hình 4.1 và Bảng 4.3 cho thấy UNet cải tiến là mô hình hiệu quả nhất. Hầu như mọi chỉ số của UNet cải tiến đều cao hơn so với các mô hình khác. Dựa trên các chỉ số chính như Dice Score và Jaccard Index, UNet cải tiến thể hiện độ chính xác và hiệu quả vượt trội trong việc phân đoạn các vùng khối u não.

Khi so sánh ngẫu nhiên 10 mẫu dữ liệu từ tập test giữa các mô hình, kết quả trong Bảng 4.4 cho thấy rằng các mô hình có một số lỗi như phân đoạn thừa, thiếu hoặc không phát hiện được khối u. Ví dụ, ở các trường hợp 3 và 4, MultiResUnet mắc lỗi khi phân đoạn thừa hoặc thiếu vùng khối u. Tổng thể trong 10 ví dụ này, UNet cải tiến có kết quả phân đoạn chính xác và đáng tin cậy nhất. Tuy nhiên, cũng có những tình huống mà UNet cải tiến gặp khó khăn, chẳng hạn như trong các trường hợp 1, 2 và 3, khi mô hình phân đoạn thiếu một số vùng, trong khi UNet gốc lại có kết quả tốt hơn.

Ở các trường hợp 8, 9 và 10, mặc dù đây là các mẫu đơn giản, nhưng các mô hình ResUnet, ResUnet++, và MultiResUnet lại dự đoán sai lệch khá xa so với kết quả thực tế. Trong khi đó, UNet cải tiến lại đưa ra kết quả chính xác hơn. Đặc biệt, trong các trường hợp 4, 5 và 6, dù có một chút phân đoạn thừa nhẹ, nhưng cả UNet gốc và UNet cải tiến vẫn đưa ra các kết quả phân đoạn tốt.

Lý do UNet cải tiến trả về kết quả tốt hơn các mô hình khác là do việc áp dụng cơ chế Attention Mechanism, giúp mô hình tập trung vào các thông tin quan trọng hơn trong quá trình phân đoạn, từ đó cải thiện độ chính xác. Ngoài ra, việc sử dụng GroupNormalization và Dropout cũng giúp giảm thiểu hiện tượng overfitting, tăng cường khả năng của mô hình trong việc xử lý các biến thể khác nhau của dữ liệu mới. Nhờ những cải tiến này, UNet cải tiến đã thể hiện sự vượt trội trong việc phân đoạn khối u não so với các mô hình khác.

# KẾT LUẬN

## Kết quả đạt được.

Trong nghiên cứu này, nhóm đã thực hiện các bước sau để đạt được mục tiêu đề ra:

* Tìm hiểu đặc trưng ảnh MRI u não: Đã nghiên cứu kỹ lưỡng về cấu trúc và đặc trưng của ảnh MRI để xác định các yếu tố cần thiết cho quá trình phân đoạn. Việc nắm rõ đặc trưng của ảnh MRI giúp nhóm hiểu rõ hơn về cách khối u não xuất hiện và ảnh hưởng đến các vùng não khác nhau, từ đó tạo cơ sở cho việc lựa chọn các mô hình và phương pháp phân đoạn phù hợp.
* Tìm hiểu các nghiên cứu liên quan và phân tích ưu nhược điểm của chúng: Nhóm đã tiến hành phân tích các phương pháp và mô hình hiện đại trong lĩnh vực phân đoạn ảnh não, từ đó lựa chọn và cải tiến mô hình phù hợp. Việc này giúp nhóm có cái nhìn tổng quan về các phương pháp hiện tại, đánh giá ưu và nhược điểm của chúng, từ đó đề xuất các cải tiến nhằm nâng cao hiệu quả phân đoạn.
* Tìm hiểu tập dữ liệu ảnh MRI u não: Đã phân tích tập dữ liệu ảnh MRI u não để tìm ra mô hình phù hợp. Nhóm đã nghiên cứu các bộ dữ liệu khác nhau, xác định các đặc điểm chính của chúng, và lựa chọn bộ dữ liệu thích hợp cho quá trình huấn luyện và kiểm tra mô hình. Điều này đảm bảo rằng mô hình được huấn luyện trên dữ liệu đa dạng và phong phú, nâng cao khả năng tổng quát hóa của mô hình.
* Đề xuất phương pháp phân đoạn ảnh MRI u não: Nhóm đã triển khai các mô hình phân đoạn ảnh như Unet, Unet cải tiến, ResUnet, ResUnet++, và MultiResUnet, đồng thời thực hiện các cải tiến phù hợp để nâng cao hiệu quả. Các mô hình này được lựa chọn và điều chỉnh dựa trên những hiểu biết thu được từ các bước trên, nhằm đảm bảo chúng hoạt động tốt nhất trên tập dữ liệu đã chọn.
* Thực nghiệm, đánh giá kết quả và so sánh với các phương pháp khác: Nhóm đã tiến hành các thí nghiệm trên tập dữ liệu thực tế, đánh giá hiệu quả của các mô hình bằng các chỉ số như Dice Score, và Jaccard. Kết quả cho thấy mô hình Unet cải tiến đạt hiệu suất tốt nhất với Dice Score và Jaccard cao nhất. Việc này chứng minh rằng các cải tiến được đề xuất thực sự có hiệu quả, và mô hình Unet cải tiến có thể được xem là một giải pháp tiềm năng cho bài toán phân đoạn ảnh MRI u não.

## Ưu và nhược điểm của phương pháp đề xuất.

Ưu điểm:

* Khả năng xử lý tự động: Mô hình có khả năng tự động xử lý và phân đoạn ảnh MRI não, giúp giảm thiểu sự can thiệp của con người và tăng độ chính xác trong chẩn đoán.
* Khả năng tổng quát hóa tốt: Mặc dù hiệu suất mô hình không quá cao (khoảng 70%), các cải tiến trên mô hình Unet giúp nó hoạt động tốt trên nhiều loại dữ liệu khác nhau, không chỉ trên tập dữ liệu huấn luyện. Điều này có nghĩa là mô hình có thể áp dụng cho các bộ dữ liệu mới mà không cần phải thay đổi cấu trúc đáng kể.
* Khả năng tùy chỉnh: Phương pháp đề xuất cho phép tùy chỉnh các tham số và cấu trúc của mô hình để phù hợp với các yêu cầu cụ thể của từng bộ dữ liệu và từng ứng dụng khác nhau. Điều này tạo điều kiện thuận lợi cho việc ứng dụng mô hình trong nhiều hoàn cảnh khác nhau.
* Hiệu quả trong việc phát hiện khối u: Mô hình có khả năng phát hiện và phân đoạn các khối u não trong ảnh MRI với mức độ chính xác chấp nhận được, cung cấp thông tin quan trọng hỗ trợ cho quá trình chẩn đoán và điều trị bệnh.

Nhược điểm:

* Hiệu suất không cao: Mô hình Unet cải tiến cho kết quả hiệu suất chỉ khoảng 70%, điều này có thể không đủ đáp ứng yêu cầu của một số ứng dụng lâm sàng đòi hỏi độ chính xác cao.
* Yêu cầu tài nguyên cao: Quá trình huấn luyện mô hình yêu cầu tài nguyên phần cứng lớn và thời gian huấn luyện dài, điều này có thể gây khó khăn trong các ứng dụng thực tế.
* Hạn chế về dữ liệu: Do sử dụng tập dữ liệu có sẵn, nên mô hình có thể chưa đạt được hiệu suất tối ưu trên các tập dữ liệu khác với chất lượng và tính đa dạng cao hơn.

## Hướng mở rộng tương lai

Để nâng cao hiệu quả và ứng dụng của mô hình Semantic Segmentation trong lĩnh vực y tế, nhóm đề xuất một số hướng mở rộng trong tương lai kèm theo các phương pháp khắc phục nhược điểm hiện tại:

* + Mở rộng tập dữ liệu: Thu thập thêm các bộ dữ liệu MRI não đa dạng và phong phú hơn, bao gồm cả dữ liệu từ nhiều nguồn khác nhau để đảm bảo tính đa dạng. Điều này sẽ giúp mô hình học được nhiều đặc trưng hơn, từ đó cải thiện hiệu suất phân đoạn.
  + Cải tiến mô hình: Tiếp tục nghiên cứu và cải tiến các mô hình phân đoạn hiện có. Khám phá các mô hình mới như Transformer-based models hoặc các kiến trúc mạng nơ-ron tiên tiến khác để cải thiện hiệu suất. Sử dụng kỹ thuật tăng cường dữ liệu (data augmentation) và tinh chỉnh các tham số của mô hình để tối ưu hóa kết quả.
  + Đánh giá hiệu suất: Thực hiện đánh giá hiệu suất xử lý của mô hình trong các tình huống thực tế, đảm bảo mô hình có thể hoạt động hiệu quả với các giới hạn về tài nguyên. Áp dụng các phương pháp giảm thiểu độ phức tạp tính toán như mô hình hóa nén (model compression) hoặc giảm bớt các tham số không cần thiết.

Với các hướng phát triển và phương pháp khắc phục này, nghiên cứu hi vọng sẽ góp phần nâng cao hiệu quả của các phương pháp phân đoạn ảnh MRI não, đồng thời mở ra nhiều cơ hội ứng dụng trong thực tiễn, đặc biệt là trong lĩnh vực y tế.

TÀI LIỆU THAM KHẢO

|  |  |
| --- | --- |
| [1] | A. A. S. R. M. M. Sophie De Beukelaer, "Case report: “Proust phenomenon” after right posterior cerebral artery occlusion," July 2023. |
| [2] | M. A. V. S. e. a. Smith D, "Radiopaedia.org," 23 Mar 2023. [Online]. Available: https://radiopaedia.org/articles/mri-brain-summary. |
| [3] | A. Y. W. W. C. T. D. T. A. F. W. G. a. A. W. Tsai, "A shape-based approach to the segmentation of medical imagery using level sets," IEEE Transactions on Medical Imaging, vol. 22, no. 2, pp. 137–154, 2003. |
| [4] | N. M. T. T. T. N. H. a. N. K. P. M. L. Le, "Using U-Net model to detect abnormal areas on brain MRI images," in *The National Conference on Fundamental and Applied IT Research (in Vietnamese) (FAIR)*, Nha Trang, 2020, pp. 656-662. |
| [5] | Z. F. R. Z. H. M. Xue Li, "Brain Tumor MRI Segmentation Method Based on Improved Res-UNet," in *Medicine and Education (ITME)*, IEEE Journal of Radio Frequency Identification ( Volume: 8), 02 January 2024, pp. 652 - 657. |
| [6] | N. &. R. M. S. Ibtehaz, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," in *Neural Networks*, 2020, pp. 121, 74-87. |
| [7] | P. F. a. T. B. O. Ronneberger, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Medical Image Computing and Computer Assisted Intervention (MICCAI), ser. LNCS, vol. 9351, Springer, 2015, p. 234– 241. |
| [8] | J. Johnson, "Build a ResUNet Model," [Online]. Available: https://discuss.pytorch.org/t/build-a-resunet-model/191335. |
| [9] | E. K. B. K. W. W. R. K. a. H.-W. M.-G. K. Held, "Markov random field segmentation of brain MRI images," IEEE Transactions on Medical Imaging, vol. 16, no. 6, pp. 878–886, 1997. |
| [10] | V. N. Y. T. D. Y. H. R. R. a. D. X. A. Hatamizadeh, "Swin UnetR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images," Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries" in Springer International Publishing, 2022, pp.272-284. |
| [11] | D. Soydaner, "Attention Mechanism in Neural Networks: Where it Comes and Where it Goes," 22 July 2021. |
| [12] | F. W. P. C. C. W. Foivos I. Diakogiannis, "ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data," 15 Jan 2020. |
| [13] | P. H. S. M. A. R. D. J. T. d. L. P. H. H. D. J. Debesh Jha, ResUNet++: An Advanced Architecture for Medical Image Segmentation, 2019. |
| [14] | L. S. S. A. G. S. E. W. Jie Hu, Squeeze-and-Excitation Networks, 2019. |
| [15] | Y. Varun, "Kaggle," 27 9 2020. [Online]. Available: https://www.kaggle.com/code/yerramvarun/understanding-dice-coefficient. |
| [16] | "VIBLO," [Online]. Available: https://viblo.asia/p/giai-thuat-jaccard-djeZ1P9GKWz. |
| [17] | J. Cheng, "Brain Tumor Dataset," 03 04 2017. [Online]. Available: https://figshare.com/articles/dataset/brain\_tumor\_dataset/1512427?file=7953679. |
| [18] | "MiAI," [Online]. Available: https://miai.vn/2020/06/16/oanh-gia-model-ai-theo-cach-mi-an-lien-chuong-2-precision-recall-va-f-score/. |
| [19] | M. R. S. N. T. a. J. L. Z. Zhou, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," Springer Verlag, 2018, p. 3–11. |
| [20] | Ö. A. A. L. S. &. R. Çiçek, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2016, pp. 424-432. |
| [21] | P. F. T. B. Olaf Ronneberger, "U-Net: Convolutional Networks for Biomedical," arxiv, 2015. |

PHỤ LỤC

Link dataset và code:

https://drive.google.com/drive/folders/1N5xCmsDk4gFsawL9N-7xoAdlp3Qyd36k?usp=sharing