

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH
KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN GIỮA KÌ
THỊ GIÁC MÁY TÍNH VÀ ỨNG DỤNG**

**PHÁT HIỆN KHỐI U NÃO SỬ DỤNG CÁC PHƯƠNG
PHÁP HỌC SÂU**

Người thực hiện: **HỒ VÕ HOÀNG DUY – 20087541**

Lớp : 420301436401

Khoá : 16

Người hướng dẫn: **KS. LƯU GIANG NAM**

TS. HỒ ĐẮC QUÁN

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH
KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN GIỮA KÌ
THỊ GIÁC MÁY TÍNH VÀ ỨNG DỤNG**

**PHÁT HIỆN KHỐI U NÃO SỬ DỤNG CÁC PHƯƠNG
PHÁP HỌC SÂU**

Người thực hiện: **HỒ VÕ HOÀNG DUY- 20087541**

Lớp : **420301436401**

Khoá : **16**

Người hướng dẫn: **KS. LƯU GIANG NAM**

TS. HỒ ĐẮC QUÁN

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024

LỜI CẢM ƠN

Tôi xin gửi lời cảm ơn chân thành nhất đến quý thầy KS. Lưu Giang Nam và TS. Hồ Đắc Quán – Giảng viên đã tận tình hướng dẫn cho chúng tôi trong suốt quá trình học tập môn thị giác máy tính và ứng dụng đã tạo điều kiện cho tôi làm Đồ án này. Dưới sự hướng dẫn của các Thầy tôi được tiếp cận với những kiến thức chuyên môn sâu, rộng trong lĩnh vực thị giác máy tính. Qua đó có thể hoàn thành Đồ án này một cách thuận lợi. Một lần nữa chúng tôi xin gửi lời cảm ơn sâu sắc đến các Thầy. Chúc Thầy thật nhiều sức khỏe, thành công trong công việc và trong cuộc sống. Xin chân thành cảm ơn Thầy.

ĐỒ ÁN ĐƯỢC HOÀN THÀNH TẠI TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TP HỒ CHÍ MINH

Tôi xin cam đoan đây là sản phẩm đồ án của riêng tôi / chúng tôi và được sự hướng dẫn của KS. Lưu Giang Nam và TS. Hồ Đắc Quán. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong đồ án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình. Trường đại học Công nghiệp TP Hồ Chí Minh không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

TP. Hồ Chí Minh, ngày tháng năm

Tác giả

(ký tên và ghi rõ họ tên)

Hồ Võ Hoàng Duy

PHẦN ĐÁNH GIÁ CỦA GIẢNG VIÊN

Tp. Hồ Chí Minh, ngày tháng năm
(kí và ghi họ tên)

TÓM TẮT

Ung thư não (khối u não) là một căn bệnh nguy hiểm, có tỷ lệ tử vong rất cao khi người bệnh mắc phải. Việc phát hiện sớm khối u não đóng vai trò quan trọng trong việc điều trị và nâng cao tỷ lệ sống cho bệnh nhân. Với sự phát triển mạnh mẽ của trí tuệ nhân tạo (AI) máy tính có thể giúp con người chuẩn đoán những khối u não một cách nhanh chóng và có độ chính xác cao hơn. Trong đồ án này đề xuất sử dụng hai mô hình học sâu (Deep Learning) tiên tiến là Faster R-CNN và YOLOv9 để phát hiện khối u trong hình ảnh chụp. Faster R-CNN là một mô hình hiệu quả trong việc phát hiện đối tượng từ hình ảnh. Mô hình gồm 2 giai đoạn tạo ra vùng đề xuất (region proposals) có thể chứa các khối u và phân loại các vùng đề xuất và dự đoán chính xác khối u. YOLOv9 là mô hình phát hiện đối tượng được phát triển trong thời gian gần đây mô hình này có tốc độ xử lý nhanh và độ chính xác cao. Kết quả thực nghiệm cho thấy YOLOv9 đạt kết quả cao, và chính xác hơn so với Faster R-CNN. Đồ án này chứng minh tính hiệu quả của hai mô hình học sâu trong việc phát hiện khối u trong não. Việc sử dụng các mô hình này giúp nâng cao hiệu quả chuẩn đoán, phát hiện và điều trị khối u não.

MỤC LỤC

LỜI CẢM ƠN.....	i
PHẦN ĐÁNH GIÁ CỦA GIẢNG VIÊN.....	iii
TÓM TẮT.....	iv
1. Giới thiệu	6
2. Phân tích yêu cầu bài toán	6
2.1. Yêu cầu bài toán	6
2.2. Các phương pháp giải quyết bài toán	6
3. Phương pháp giải quyết bài toán	7
3.1. Faster R-CNN.....	8
3.1.1. Kiến trúc.....	9
3.1.2. Cách thức hoạt động	10
3.2. Mô hình YOLO	10
3.2.1. Giới thiệu về YOLO	10
3.2.2. Lịch sử phát triển của YOLO.....	12
4. Thực nghiệm.....	16
4.1. Dữ liệu	16
4.2. Phân tích, xử lý dữ liệu.....	16
4.3. Phương pháp đánh giá	20
4.3.1. mAP@50.....	20
4.3.2. mAP@50-95	20
4.4. Kết quả đạt được.....	21
5. Kết luận.....	24
Tài liệu tham khảo	25

PHÁT HIỆN KHỐI U NÃO SỬ DỤNG PHƯƠNG PHÁP HỌC SÂU

1. Giới thiệu

Phát hiện khối u não sớm không chỉ là yếu tố quan trọng để nhanh chóng đưa ra các phương pháp điều trị kịp thời và hiệu quả mà còn giảm nguy cơ tử vong tăng cơ hội phục hồi cho bệnh nhân. Khối u não có thể gây ra các biến chứng nghiêm trọng cho bệnh nhân như đau đầu, mất trí nhớ, co giật và làm suy giảm các chức năng cơ thể. Do đó việc phát hiện sớm và điều trị khối u não có vai trò quan trọng trong việc cải thiện chất lượng cuộc sống và tăng khả năng sống sót cho bệnh nhân. Việc phát hiện khối u não trong hình ảnh y tế là một thách thức đối với các chuyên gia y tế do khối u có thể xuất hiện nhiều vị trí trong não và có kích thước, hình dạng khác nhau đồng thời bị che khuất bởi những cấu trúc khác nhau trong não. Điều này làm cho quá trình phát hiện trở nên phức tạp và đòi hỏi sự chính xác cao từ các phương pháp hỗ trợ. Đồ án này nhằm mục đích nghiên cứu và ứng dụng các phương pháp học sâu tiên tiến vào việc phát hiện và xác định vị trí khối u não trong hình ảnh y tế. Bằng các sử dụng mô hình phát hiện đối tượng là Faster R-CNN và YOLOv9. Tôi hy vọng có thể xây dựng một hệ thống phát hiện khối u não chính xác và đáng tin cậy có thể áp dụng trong thực tế từ đó có thể cải thiện khả năng chuẩn đoán và điều trị cho bệnh nhân mắc phải căn bệnh này giúp họ tăng khả năng sống và sớm trở lại cuộc sống bình thường.

2. Phân tích yêu cầu bài toán

2.1. Yêu cầu bài toán

Bài toán phát hiện khối u não sử dụng phương pháp học sâu là một bài toán quan trọng giúp hỗ trợ các y bác sĩ trong việc chuẩn đoán và điều trị kịp thời cho bệnh nhân. Bài toán này có các yêu cầu sau: Đầu vào là hình ảnh y tế của não. Đầu ra sẽ là bounding box phát hiện khối u trong não, khối u này có thể là lành tính hoặc ác tính. Có thể sử dụng các phương pháp mô hình tiên tiến trong việc phát hiện đối tượng để giải quyết bài toán này. Bài toán cũng yêu cầu về độ chính xác cao, tốc độ xử lý nhanh và có thể ứng dụng trong thực tế.

2.2. Các phương pháp giải quyết bài toán

Đối với bài toán phát hiện đối tượng nói chung và bài toán phát hiện khối u não nói riêng, có nhiều phương pháp học sâu đã được đề xuất trước đây có thể giải quyết vấn đề này. Những mô hình học sâu tiêu biểu trong việc phát hiện đối tượng:

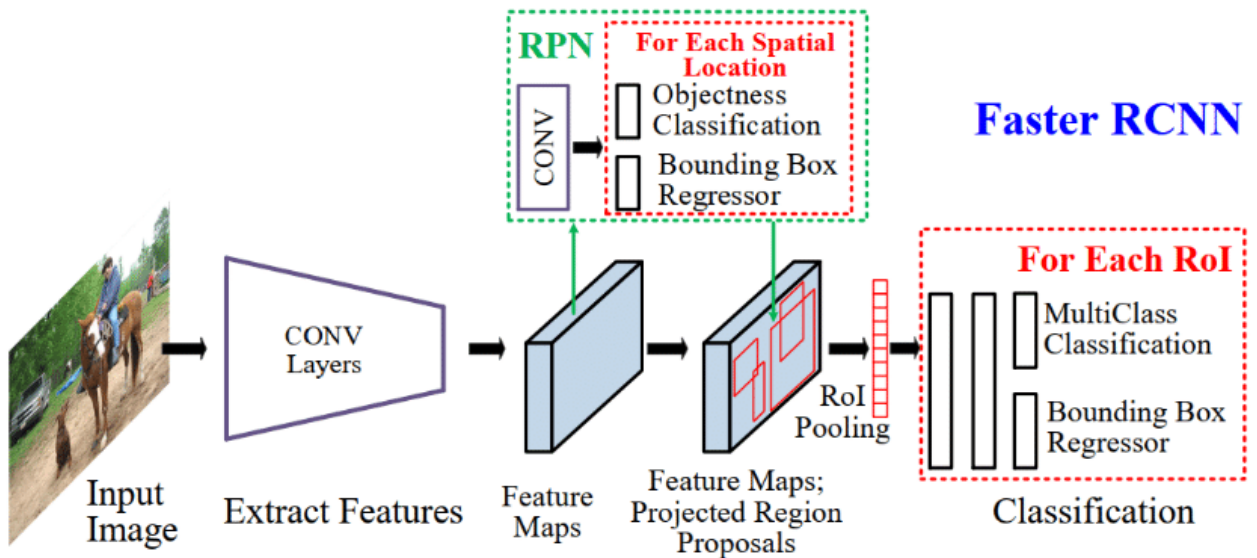
- **R-CNN (Region-based Convolutional Neural Network) [1]** là một trong những phương pháp tiên tiến trong việc phát hiện đối tượng. Được giới thiệu bởi Ross Girshick và cộng sự, 2014. R-CNN đã mở ra hướng tiếp cận mới cho bài toán phát hiện đối tượng. R-CNN đã mang lại những tiến bộ đáng kể trong việc phát hiện đối tượng so với các phương pháp truyền thống trước đây. Tuy nhiên R-CNN cũng có một số điểm hạn chế bao gồm tốc độ huấn luyện và dự đoán chậm do phải xử lý mỗi vùng ứng cử một cách độc lập. Để cải thiện những hạn chế trên thì các biến thể Fast R-CNN và Faster R-CNN đã được phát triển, mang lại sự tiện lợi và hiệu quả hơn trong thực tế.
- **Fast R-CNN [2]** được giới thiệu bởi Ross Girshick, 2015. Đây là mô hình cải tiến những điểm hạn chế còn tồn tại trước đó của R-CNN. Fast R-CNN giải quyết những hạn chế của R-CNN bằng cách kết hợp việc trích xuất đặc trưng và phát hiện vị trí đối tượng vào một mạng neural duy nhất. Sự kết hợp giữa hiệu suất và tốc độ của Fast R-CNN đã làm cho nó trở thành một trong những phương pháp phổ biến và mạnh mẽ trong lĩnh vực phát hiện đối tượng.
- **Faster R-CNN [3]** là một phương pháp tiên tiến trong lĩnh vực thị giác máy tính được giới thiệu Shaoqing Ren và cộng sự, 2015. Mô hình đạt được sự đánh giá cao nhờ khả năng kết hợp giữa độ chính xác cao và tốc độ xử lý nhanh trong việc định vị và xác định và phát hiện vị trí các đối tượng.
- **YOLO (You Only Lock Once) [4]** là một lĩnh vực tiên tiến trong phát hiện đối tượng được giới thiệu bởi Joseph Redmon và cộng sự, 2016. YOLO là một mô hình phát hiện và phân loại đối tượng trên hình ảnh với đặc điểm nổi bật là tốc độ cao và khả năng phát hiện đối tượng thời gian thực. Tính linh hoạt, tốc độ và khả năng phát hiện thời gian thực là những yếu tố đã làm cho YOLO trở thành một trong những phương pháp phổ biến nhất và mạnh mẽ nhất trong lĩnh vực thị giác máy tính.

3. Phương pháp giải quyết bài toán

Để giải quyết bài toán phát hiện khối u não trên hình ảnh y tế, tôi đề xuất sử dụng hai phương pháp là mô hình Faster R-CNN và YOLOv9. Mỗi mô hình sẽ được áp dụng độc lập và đưa ra đánh giá về hiệu suất của từng phương pháp trong việc phát hiện và định vị khối u não trên hình ảnh y tế. Faster R-CNN sẽ được sử dụng với kiến trúc được tinh chỉnh để phát hiện và định vị khối u não trong các hình ảnh y tế. Mô hình Faster R-CNN sẽ được huấn luyện trên một tập dữ liệu chứa các hình ảnh y tế có chứa khối u não, nhằm học được các đặc trưng đặc biệt của khối u và cách để phát hiện chúng một cách chính xác. Trong khi đó, YOLOv9 cung cấp một cách tiếp cận khác với việc phát hiện đối tượng trong thời gian thực. Với kiến trúc được cải thiện và tối ưu hóa, YOLOv9 có khả năng phát hiện và định vị các khối u không chỉ nhanh chóng mà còn chính xác. Mô hình YOLOv9 sẽ được huấn luyện trên cùng một tập dữ liệu giống như Faster R-CNN để so sánh hiệu suất của hai phương pháp. Bằng cách này, chúng tôi hy vọng có thể đánh giá và so sánh hiệu suất của Faster R-CNN và YOLOv9 trong việc phát hiện khối u não trên hình ảnh y tế. Kết quả từ đồ án này sẽ cung cấp thông tin quan trọng về sự hiệu quả của các phương pháp thị giác máy tính trong việc hỗ trợ chẩn đoán và điều trị bệnh nhân mắc các vấn đề liên quan đến khối u não.

3.1. Faster R-CNN

Faster R-CNN (Faster Region-CNN) là một trong những mô hình nổi tiếng và mạnh mẽ trong lĩnh vực nhận diện vật thể và phát hiện vị trí của chúng trong ảnh. Được giới thiệu bởi Shaoqing Ren, Kaiming He, Ross Girshick và Jian Sun của Microsoft Research, Faster R-CNN kết hợp cả hai giai đoạn của phương pháp R-CNN (Region-based Convolutional Neural Network) và Fast R-CNN để tạo ra một mô hình phát hiện vật thể có khả năng chạy nhanh hơn.



Hình 1. Kiến trúc Faster R-CNN

3.1.1. Kiến trúc

a. Backbone Network:

- Faster R-CNN sử dụng một mạng CNN (Convolutional Neural Network) như VGG, ResNet, hoặc một mạng CNN tiên tiến hơn để trích xuất các đặc trưng từ ảnh đầu vào. Mạng này thường được tiền huấn luyện trên một tập dữ liệu lớn như ImageNet để học các đặc trưng phổ quát.

b. Region Proposal Network (RPN):

- RPN là phần quan trọng của Faster R-CNN, được sử dụng để đề xuất các vùng quan trọng (regions of interest - RoIs) trong ảnh đầu vào.
- RPN là một mạng CNN nhỏ được đặt trên top của feature map được tạo ra bởi backbone network. Nó tạo ra một tập hợp các vùng đề xuất cũng như xác suất đó là vùng chứa vật thể.
- Mỗi vùng đề xuất được đặc trưng bằng một vector và sau đó được đưa vào mô hình RoI pooling.

c. Region of Interest (RoI) Pooling:

- RoI Pooling là một phần quan trọng của Faster R-CNN, cho phép mô hình xử lý các vùng đề xuất với kích thước không đồng nhất thành các tensor có kích thước cố định để đưa vào mạng phân loại vật thể.
- Nó chia các vùng đề xuất thành các ô lưới cố định và thực hiện pooling trên mỗi ô lưới đó để tạo ra các feature map có kích thước cố định.

d. Classifier và Bounding Box Regressor:

- Các feature map từ RoI pooling được đưa vào các mạng fully connected layers để phân loại vật thể và dự đoán các bounding box (hình chữ nhật bao quanh vật thể).
- Classifier dự đoán xác suất của mỗi lớp vật thể trong từng vùng đề xuất.
- Bounding box regressor dự đoán các tham số (độ dài, chiều rộng, tọa độ) của bounding box tương ứng với vật thể trong mỗi vùng đề xuất.

Với kiến trúc này, Faster R-CNN có khả năng phát hiện vật thể trong ảnh một cách chính xác và nhanh chóng, đặc biệt phù hợp cho các ứng dụng thời gian thực và yêu cầu độ chính xác cao trong nhận diện vật thể.

3.1.2. Cách thức hoạt động

- Trích xuất đặc trưng: Đầu tiên, ảnh được đưa qua mạng CNN để trích xuất các đặc trưng.
- Tạo vùng đề xuất (Region Proposals): RPN tạo ra một số lượng lớn các vùng đề xuất dựa trên các feature map.
- RoI Pooling: Các vùng đề xuất được chuyển qua RoI Pooling để chuyển đổi chúng thành các feature map có kích thước cố định.
- Phân loại và dự đoán bounding box: Các feature map từ RoI Pooling được sử dụng để phân loại vật thể và dự đoán bounding box tương ứng.
- Non-maximum Suppression (NMS): Để loại bỏ các bounding box trùng lặp và giữ lại các bounding box tốt nhất, Faster R-CNN sử dụng thuật toán NMS.

3.2. Mô hình YOLO

3.2.1. Giới thiệu về YOLO

YOLO, viết tắt của "You Only Look Once", là một trong những phương pháp tiên tiến nhất và phổ biến nhất trong lĩnh vực nhận diện vật thể và phát hiện vị trí của chúng trong ảnh và video. YOLO nổi bật với khả năng phát hiện và phân loại nhanh chóng và chính xác trên ảnh thời gian thực.

- YOLO sử dụng một hàm mất mát tổng quát (composite loss function) kết hợp giữa hàm mất mát về dự đoán bounding box (bounding box regression loss) và hàm mất mát về phân loại lớp vật thể (classification loss).
- Hàm mất mát này giúp cải thiện độ chính xác của mô hình bằng cách đồng thời tối ưu hóa cả việc dự đoán vị trí và loại của vật thể.

f. Anchor Boxes:

- YOLO sử dụng các anchor boxes để dự đoán bounding box.
- Anchor boxes giúp mô hình dự đoán bounding box cho các vật thể ở nhiều kích thước và tỷ lệ khác nhau, từ đó cải thiện khả năng nhận diện vật thể của mô hình.

Mô hình YOLO có nhiều đặc điểm đặc biệt giúp nó trở thành một công cụ mạnh mẽ trong lĩnh vực nhận diện vật thể và phát hiện vị trí, đặc biệt là trong các ứng dụng yêu cầu tính toán nhanh chóng và độ chính xác cao.

3.2.2. Lịch sử phát triển của YOLO

YOLOv1 (2015):

- YOLO được giới thiệu lần đầu tiên bởi Joseph Redmon, Santosh Divvala, Ross Girshick và Ali Farhadi trong bài báo khoa học có tựa đề "You Only Look Once: Unified, Real-Time Object Detection" vào năm 2015.
- YOLOv1 áp dụng một mạng neural network một lớp để dự đoán các bounding box và xác suất của các lớp vật thể trực tiếp từ toàn bộ ảnh.

YOLOv2 (2016) [5]:

- Phiên bản cải tiến YOLOv2 được công bố vào năm 2016 bởi Joseph Redmon và Ali Farhadi.
- YOLOv2 cải thiện nhiều vấn đề của YOLOv1 bằng cách sử dụng các kỹ thuật như batch normalization, multi-scale training, anchor boxes và các cải tiến khác để tăng cường hiệu suất và chính xác của mô hình.

YOLOv3 (2018) [6]:

- YOLOv3, được công bố vào năm 2018 bởi Joseph Redmon và các cộng sự, tiếp tục cải thiện hiệu suất và độ chính xác của mô hình YOLO.

- YOLOv3 sử dụng một kiến trúc mạng neural network sâu hơn với nhiều lớp hơn và các kỹ thuật như skip connections để cải thiện khả năng nhận diện và phát hiện vật thể.

YOLOv4 [7]:

- YOLOv4, được giới thiệu vào năm 2020 bởi Alexey Bochkovskiy, Chien-Yao Wang và các cộng sự, là phiên bản cải tiến nhất và mạnh mẽ nhất của YOLO đến thời điểm đó.
- YOLOv4 sử dụng nhiều kỹ thuật mới như CSPDarknet53, PANet, SPP, SAM, và các kỹ thuật tăng cường khác để cải thiện hiệu suất và độ chính xác của mô hình.

YOLOv5 (2020) [8]:

- YOLOv5, phát triển bởi Glenn Jocher và đội ngũ của Ultralytics, được giới thiệu vào năm 2020.
- YOLOv5 là một phiên bản độc lập được phát triển dựa trên YOLOv4 nhưng không có sự tham gia của Joseph Redmon. Phiên bản này được xây dựng trên cơ sở của PyTorch và có nhiều cải tiến về hiệu suất và dễ dàng triển khai.

YOLOv6 (2022) [9]:

- YOLO v6 được đề xuất vào năm 2022 bởi Li và cộng sự như một cải tiến so với các phiên bản trước. Một trong những điểm khác biệt chính giữa YOLO v5 và YOLO v6 là kiến trúc CNN được sử dụng. YOLO v6 sử dụng một biến thể của kiến trúc EfficientNet có tên là EfficientNet-L2. Đó là một kiến trúc hiệu quả hơn so với EfficientDet được sử dụng trong YOLO v5, với ít tham số hơn và hiệu quả tính toán cao hơn. Nó có thể đạt được kết quả tiên tiến trên các điểm chuẩn phát hiện đối tượng khác nhau.
- YOLO v6 cũng giới thiệu một phương pháp mới để tạo các anchor box, được gọi là “dense anchor boxes”.

YOLOv7 [10]:

- YOLO v7 có một số cải tiến so với các phiên bản trước. Một trong những cải tiến chính là việc sử dụng các anchor box.
- Các anchor box là một tập hợp các hộp được xác định trước với các tỷ lệ khung hình khác nhau được sử dụng để phát hiện các đối tượng có hình dạng

khác nhau. YOLO v7 sử dụng chín anchor box, cho phép YOLO phát hiện phạm vi hình dạng và kích thước đối tượng rộng hơn so với các phiên bản trước, do đó giúp giảm số lượng xác định sai.

- Một cải tiến quan trọng trong YOLO v7 là việc sử dụng một loss function mới gọi là “focal loss”. Các phiên bản trước của YOLO đã sử dụng cross-entropy loss function tiêu chuẩn, được biết là kém hiệu quả hơn trong việc phát hiện các đối tượng nhỏ. Focal loss giải quyết vấn đề này bằng cách giảm trọng số mất mát cho các ví dụ được phân loại tốt và tập trung vào các ví dụ khó—các đối tượng khó phát hiện.
- YOLO v7 cũng có độ phân giải cao hơn so với các phiên bản trước. Nó xử lý hình ảnh ở độ phân giải 608 x 608 pixel, cao hơn độ phân giải 416 x 416 được sử dụng trong YOLO v3. Độ phân giải cao hơn này cho phép YOLO v7 phát hiện các đối tượng nhỏ hơn và có độ chính xác tổng thể cao hơn.

YOLOv8 [11]:

- YOLOv8 là lần lặp lại mới nhất trong YOLO loạt máy dò đối tượng thời gian thực, cung cấp hiệu suất tiên tiến về độ chính xác và tốc độ.
- Xây dựng dựa trên những tiên bộ của trước đó YOLO Phiên bản YOLOv8 Giới thiệu các tính năng mới và tối ưu hóa làm cho nó trở thành một lựa chọn lý tưởng cho các tác vụ phát hiện đối tượng khác nhau trong một loạt các ứng dụng.
- Kiến trúc xương sống và cỗ nâng cao: YOLOv8 sử dụng kiến trúc xương sống và cỗ hiện đại, dẫn đến hiệu suất trích xuất tính năng và phát hiện đối tượng được cải thiện.
- Chia tách không neo Ultralytics Đầu: YOLOv8 áp dụng phân chia không neo Ultralytics đầu, góp phần vào độ chính xác tốt hơn và quy trình phát hiện hiệu quả hơn so với các phương pháp tiếp cận dựa trên neo.
- Tối ưu hóa sự đánh đổi độ chính xác-tốc độ: Tập trung vào việc duy trì sự cân bằng tối ưu giữa độ chính xác và tốc độ, YOLOv8 phù hợp cho các tác vụ phát hiện đối tượng theo thời gian thực trong các lĩnh vực ứng dụng đa dạng.
- Nhiều mô hình Pre-training: YOLOv8 Cung cấp một loạt các mô hình được đào tạo trước để phục vụ cho các nhiệm vụ và yêu cầu hiệu suất khác nhau,

giúp bạn dễ dàng tìm thấy mô hình phù hợp cho trường hợp sử dụng cụ thể của mình.

YOLOv9 [12]:

- YOLOv9 đánh dấu một bước tiến đáng kể trong việc phát hiện đối tượng thời gian thực, giới thiệu các kỹ thuật đột phá như Thông tin chuyển màu có thể lập trình (PGI) và Mạng tổng hợp lớp hiệu quả tổng quát (GELAN). Mô hình này thể hiện những cải tiến đáng kể về hiệu quả, độ chính xác và khả năng thích ứng, thiết lập các tiêu chuẩn mới trên bộ dữ liệu MS COCO. Dự án YOLOv9, trong khi được phát triển bởi một nhóm mã nguồn mở riêng biệt, được xây dựng dựa trên cơ sở mã mạnh mẽ được cung cấp bởi Ultralytics YOLOv5, thể hiện tinh thần hợp tác của cộng đồng nghiên cứu AI.
- Trong nhiệm vụ phát hiện đối tượng thời gian thực tối ưu, YOLOv9 nổi bật với cách tiếp cận sáng tạo để vượt qua những thách thức mất thông tin vốn có trong các mạng nơ-ron sâu. Bằng cách tích hợp PGI và kiến trúc GELAN linh hoạt, YOLOv9 không chỉ nâng cao năng lực học tập của mô hình mà còn đảm bảo lưu giữ thông tin quan trọng trong suốt quá trình phát hiện, do đó đạt được độ chính xác và hiệu suất vượt trội.

Trong đồ án này, tôi sử dụng các mô hình như Faster R-CNN, YOLOv5, YOLOv8 và YOLOv9 là một phần quan trọng để xây dựng một hệ thống phát hiện khối u não hiệu quả. Mỗi mô hình mang lại những ưu điểm và hạn chế riêng, từ đó tạo ra những khía cạnh đa dạng trong quá trình đánh giá và lựa chọn mô hình tốt nhất cho bài toán này.

Faster R-CNN là một trong những mô hình phát hiện vật thể hàng đầu, được biết đến với khả năng đồng thời phát hiện và phân loại vật thể nhanh chóng. Tuy nhiên, việc áp dụng Faster R-CNN cho việc phát hiện khối u não có thể đối mặt với thách thức về hiệu suất và độ chính xác, đặc biệt là khi đối diện với các khối u có kích thước nhỏ hoặc có sự mất mát thông tin nghiêm trọng.

Trong khi đó, YOLO (You Only Look Once) là một loạt các mô hình phát hiện vật thể được biết đến với tốc độ cao và khả năng phát hiện đồng thời trên toàn bức ảnh. YOLOv5, YOLOv8 và YOLOv9 đều mang lại cải tiến về hiệu suất và độ chính xác

so với các phiên bản trước đó. Tuy nhiên, sự đánh đổi giữa tốc độ và độ chính xác vẫn là một yếu tố cần xem xét khi sử dụng YOLO cho bài toán phát hiện khối u não, đặc biệt là với các khối u nhỏ hoặc có biên độ màu sắc và độ tương phản thấp.

Qua quá trình đánh giá và so sánh giữa các mô hình, chúng tôi sẽ có cái nhìn rõ ràng hơn về ưu và nhược điểm của mỗi mô hình trong bối cảnh cụ thể của bài toán phát hiện khối u trong não. Sự lựa chọn cuối cùng về mô hình phụ thuộc vào nhiều yếu tố như yêu cầu về tốc độ, độ chính xác, kích thước của khối u và điều kiện cụ thể của dữ liệu.

4. Thực nghiệm

4.1. Dữ liệu

Tập dữ liệu được sử dụng trong đề án này là **Brain tumor object detection datasets[14]** được chia sẻ từ nền tảng kaggle. Đây là tập dữ liệu về hình ảnh chụp MRI của não người. Được gán nhãn các khối u bằng tay. Được sử dụng trong bài toán nhận diện, phát hiện khối u trong não. Tập dữ liệu được phân thành hai nhãn là 0 và 1, 0 đối với khối u ác tính (negative) và 1 đối với khối u lành tính (positive).

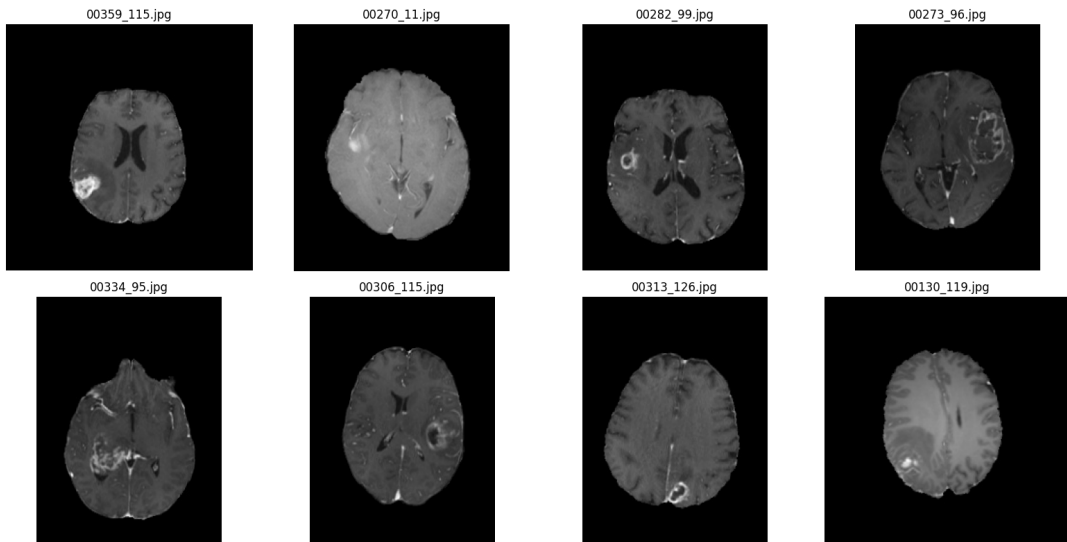
Trong đề án này tỷ lệ dữ liệu của tôi được chia như sau:

Bảng 1. Bộ dữ liệu

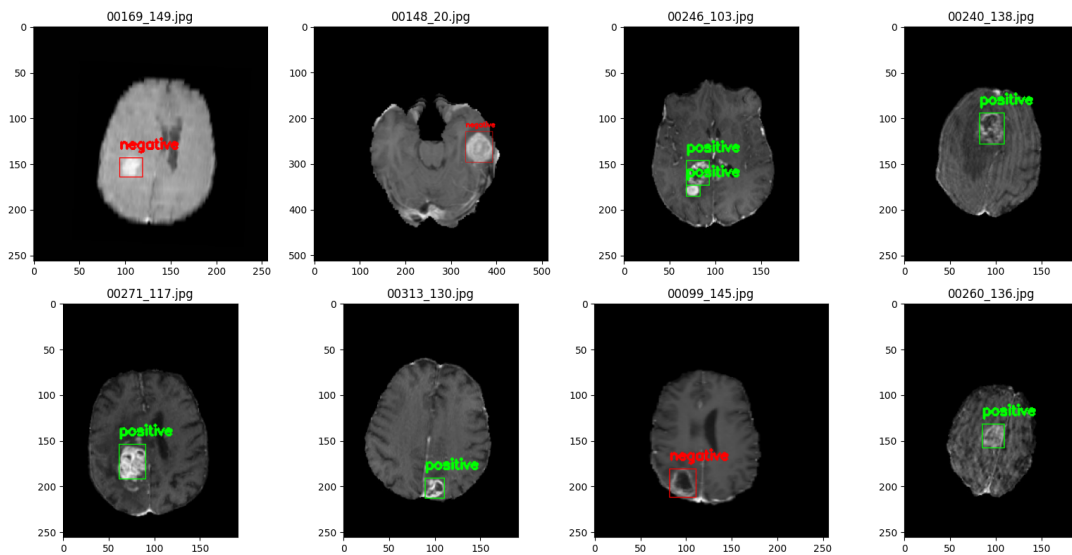
Dữ liệu	Số lượng
Train	237 ảnh
Test	75 ảnh
Val	59 ảnh

4.2. Phân tích, xử lý dữ liệu

Hình 3, 4 là một số hình ảnh chụp MRI não trong tập dữ. **Hình 4** là các hình ảnh đã được gán nhãn bằng tay trên makessen.ai. Các nhãn sẽ bao gồm ‘negative’ và ‘positive’.

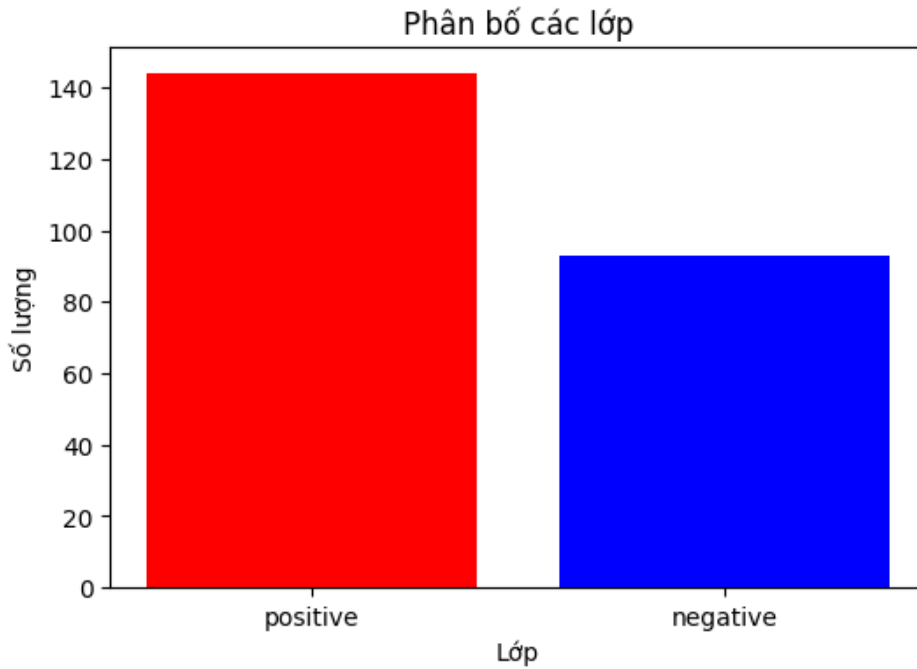


Hình 3. Một số hình ảnh MRI não trong tập dữ liệu.



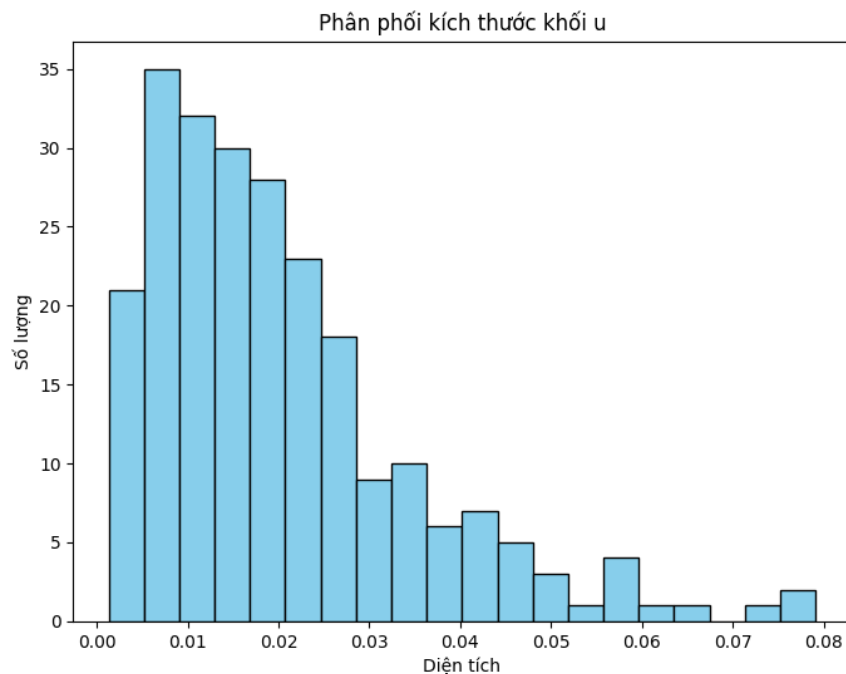
Hình 4. Hình ảnh MRI não đã được gán nhãn.

Hình 5 là biểu thể hiện sự phân bố của 2 lớp trong tập dữ liệu. Quan sát biểu đồ có thể thấy lớp positive có số lượng nhiều hơn hẳn so với lớp negative. Sự phân bố không đồng đều này có thể dẫn đến một số kết quả không tốt trong quá trình huấn luyện mô hình.

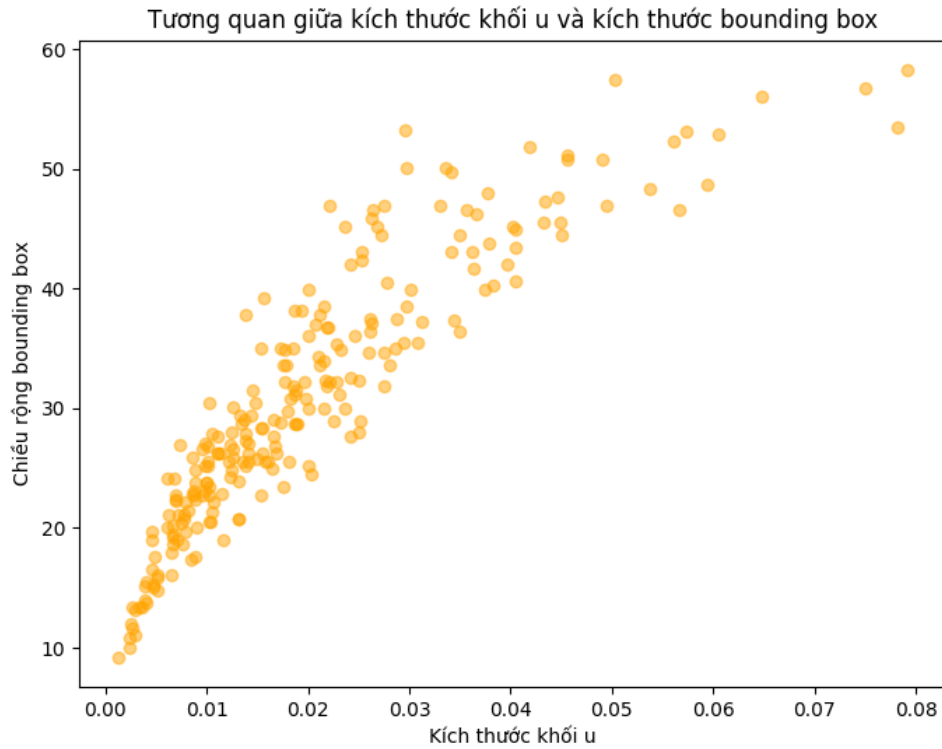


Hình 5. Biểu đồ phân bố của các lớp trong dữ liệu.

Hình 6 là một biểu đồ hình cột thể hiện số lượng khối u theo diện tích Phần lớn các khối u có diện tích nhỏ (dưới 0,04). Có một số ít khối u có diện tích lớn (trên 0,06). Biểu đồ này cung cấp một cái nhìn tổng quan về kích thước các khối u não trong tập dữ liệu

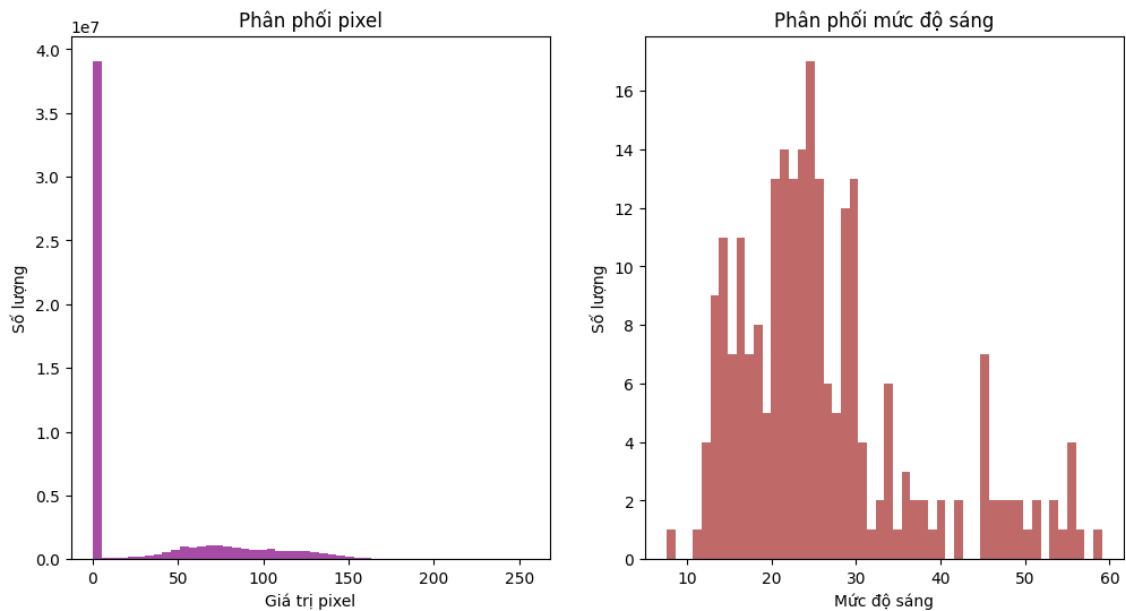


Hình 6. Biểu đồ phân phối kích thước khối u



Hình 7. Biểu đồ tương quan giữa kích thước khối u và bounding box

Hình 7 là biểu đồ thể hiện sự tương quan tuyến tính giữa kích thước khối u và bounding box. Nếu kích thước khối u càng lớn thì kích thước bounding box càng lớn và ngược lại.



Hình 8. Biểu đồ phân phối pixel và mức độ sáng của ảnh MRI não trong tập dữ liệu.

Hình 8 là biểu đồ một bộ đôi histogram, mỗi histogram thể hiện một phần của dữ liệu ảnh, giúp phân tích phân phối của các đặc tính khác nhau. Cả hai histogram đều cung cấp cái nhìn tổng quan về đặc tính của dữ liệu ảnh trong tập dữ liệu đào tạo, giúp hiểu rõ hơn về phân phối pixel và mức độ sáng, từ đó có thể đưa ra các quyết định hoặc điều chỉnh phù hợp khi tiến hành xử lý ảnh.

4.3. Phương pháp đánh giá

4.3.1. mAP@50

mAP@50 (mean Average Precision at 50) là một phương pháp đo lường hiệu suất trong bài toán nhận diện vật thể (object detection), nơi một ngưỡng xác suất được áp dụng để quyết định liệu một dự đoán có được coi là chính xác hay không.

Trong mAP@50, số 50 thường đề cập đến số lượng đề xuất tối đa cho mỗi hình ảnh. Cụ thể, một mô hình object detection thường sẽ tạo ra một danh sách các dự đoán về vị trí và loại của các vật thể trong mỗi hình ảnh. mAP@50 sẽ tính toán trung bình của Average Precision (AP) cho mỗi lớp vật thể khi chỉ xem xét 50 dự đoán có xác suất cao nhất cho mỗi hình ảnh.

Điểm mạnh của mAP@50 là nó cung cấp một cái nhìn cụ thể về hiệu suất của một mô hình object detection trong các tình huống thực tế, nơi việc giới hạn số lượng dự đoán là cần thiết để giảm thiểu độ phức tạp tính toán và tăng tốc độ xử lý.

Tuy nhiên, một điểm yếu của mAP@50 là việc chọn ngưỡng xác suất cố định (trong trường hợp này là 50) có thể không phản ánh đúng mức độ chính xác của mô hình trên mọi tình huống. Điều này có thể dẫn đến sự bỏ qua các dự đoán có xác suất thấp nhưng thực tế là đúng, hoặc sự lạc quan về hiệu suất khi chỉ xem xét các dự đoán có xác suất cao.

mAP@50 là một phương pháp đo lường hiệu suất hữu ích trong bài toán nhận diện vật thể, nhưng cần phải được sử dụng cẩn thận và kết hợp với các phương pháp đo lường khác để có cái nhìn toàn diện về hiệu suất của một mô hình object detection.

4.3.2. mAP@50-95

mAP@50-95 (mean Average Precision from 50 to 95) là một phương pháp đo lường hiệu suất trong bài toán nhận diện vật thể (object detection), tập trung vào việc tính toán trung bình của Average Precision (AP) trong khoảng từ 50 đến 95 ngưỡng xác suất khác nhau.

Trong $mAP@50-95$, ngưỡng xác suất thường được thay đổi từ 50 đến 95 với bước nhảy cố định (ví dụ: mỗi 5 điểm), và cho mỗi ngưỡng xác suất này, AP được tính toán. Cuối cùng, $mAP@50-95$ là trung bình cộng của tất cả các giá trị AP này.

Sự lựa chọn của phạm vi từ 50 đến 95 cho phép đánh giá hiệu suất của mô hình object detection ở nhiều ngưỡng xác suất khác nhau. Việc sử dụng một phạm vi rộng như vậy có thể giúp đánh giá được mức độ chính xác và độ phủ của mô hình trên các vật thể với các ngưỡng xác suất khác nhau.

Một số điểm cần lưu ý khi sử dụng $mAP@50-95$ là:

Đa dạng ngưỡng xác suất: Bằng cách tính toán AP từ 50 đến 95, $mAP@50-95$ cung cấp cái nhìn tổng quan về hiệu suất của mô hình trên nhiều mức độ tin cậy khác nhau.

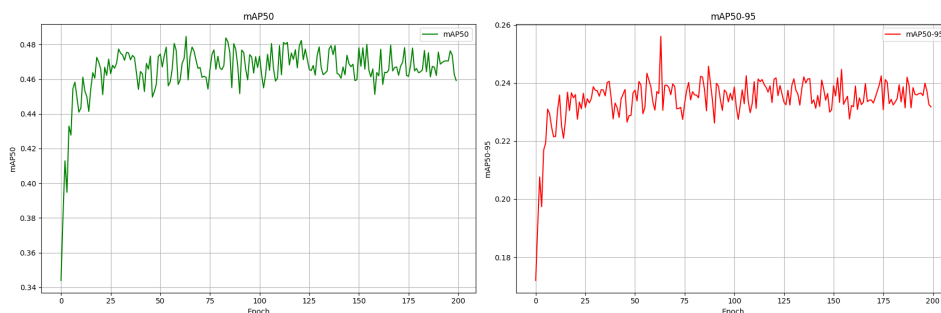
Độ tin cậy cao: Mô hình với $mAP@50-95$ cao có xu hướng phát hiện và xác định vị trí của các vật thể chính xác và đáng tin cậy ở nhiều ngưỡng xác suất.

Đánh giá toàn diện: Một mô hình có $mAP@50-95$ tốt thường cho thấy khả năng phát hiện và xác định vị trí của các vật thể đối với một loạt các ngưỡng xác suất, từ thấp đến cao.

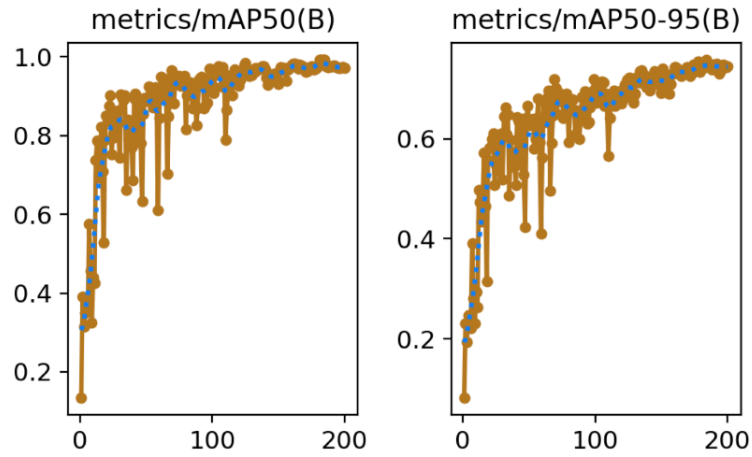
$mAP@50-95$ là một phương pháp đo lường hiệu suất quan trọng và toàn diện trong bài toán nhận diện vật thể, giúp đánh giá và so sánh hiệu suất của các mô hình object detection trên nhiều ngưỡng xác suất khác nhau.

4.4. Kết quả đạt được

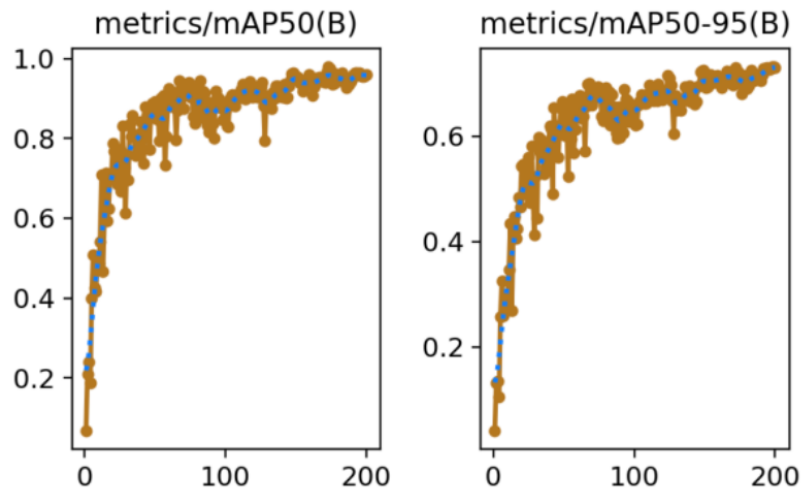
Sau khi thực hiện huấn luyện các mô hình trên tập dữ liệu, tôi đạt được các kết quả sau. Các biểu đồ $mAP@50$, $mAP@50-95$ của các mô hình:



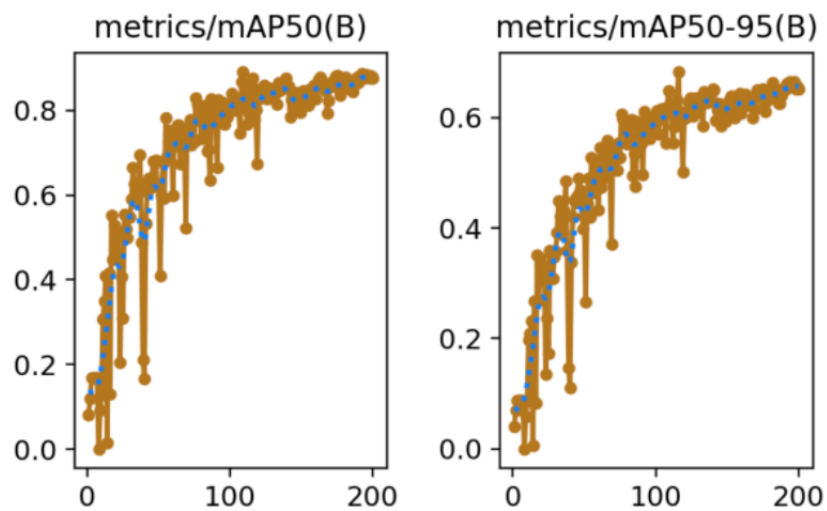
Hình 9. Biểu đồ $mAP@50$, $mAP@50-95$ của Faster R-CNN



Hình 10. Biểu đồ mAP@50, mAP@50-95 của YOLOv5.



Hình 11. Biểu mAP@50, mAP@50-95 của YOLOv8

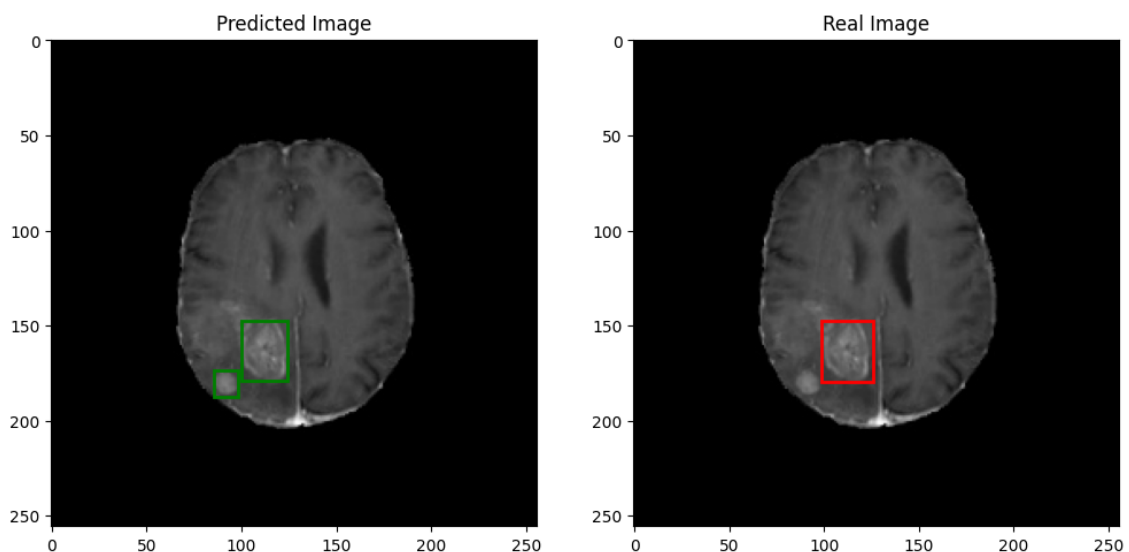


Hình 12. Biểu mAP@50, mAP@50-95 của YOLOv9

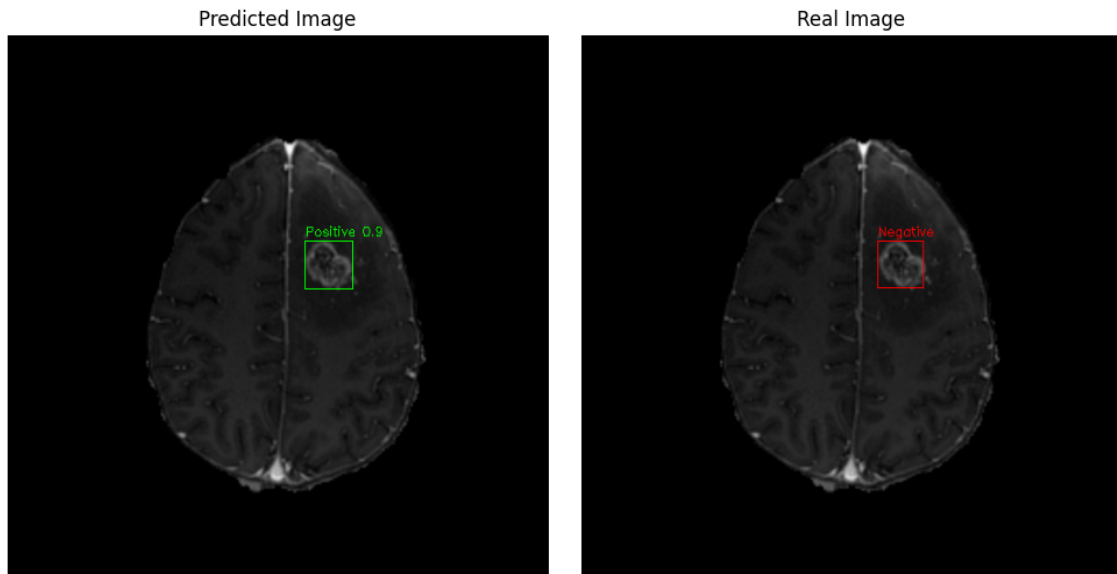
Bảng 2. Kết quả các mô hình dựa trên các độ đo mAP@50 và mAP@50-95

	mAP@50	mAP@50-95
Faster R-CNN	0.459	0.231
YOLOv5	0.895	0.66
YOLOv8	0.891	0.657
YOLOv9	0.81	0.58

Các hình 9, 10, 11, 12 và Bảng 2 có thể thấy kết quả của các mô hình Faster R-CNN, YOLOv5, YOLOv8 và YOLOv9 trên các độ đo mAP@50 và mAP@50-95. Ta có thể thấy mô hình YOLO cho ra kết quả và tốc độ xử lý vượt trội hoàn toàn so với Faster R-CNN. Để giải thích cho điều này thì YOLO sử dụng kiến trúc mạng đơn giản hơn Faster RCNN. YOLO chỉ sử dụng một mạng nơ-ron tích chập để dự đoán cả vị trí và lớp của các đối tượng, trong khi Faster RCNN sử dụng hai mạng riêng biệt: mạng đầu tiên để đề xuất vùng (region proposal) và mạng thứ hai để phân loại các vùng đề xuất. Kiến trúc đơn giản của YOLO giúp tăng tốc độ xử lý đáng kể. YOLO sử dụng phương pháp dự đoán dựa trên lưới (grid-based prediction) để dự đoán vị trí của các đối tượng. Phương pháp này chia hình ảnh thành một lưới các ô và dự đoán vị trí của các đối tượng trong mỗi ô. Faster RCNN sử dụng phương pháp dự đoán dựa trên vùng (region-based prediction) để dự đoán vị trí của các đối tượng. Phương pháp này sử dụng các vùng đề xuất được tạo bởi mạng đầu tiên để dự đoán vị trí của các đối tượng. Phương pháp dự đoán dựa trên lưới của YOLO giúp tăng tốc độ xử lý vì nó không cần phải tạo các vùng đề xuất.



Hình 13. Kết quả dự đoán của Faster R-CNN.



Hình 14. Kết quả dự đoán của YOLOv9.

5. Kết luận

Trong đồ án này, tôi đã trình bày về phương pháp phát hiện khối u não sử dụng các phương pháp học sâu, tập trung vào hai phương pháp phổ biến là YOLO (You Only Look Once) và Faster R-CNN. Bằng cách sử dụng các mô hình học sâu và các kỹ thuật tiên tiến trong xử lý ảnh y khoa, đồ án này nhấn mạnh sự quan trọng của việc áp dụng trí tuệ nhân tạo vào lĩnh vực y học.

Phương pháp YOLO đã được chứng minh là hiệu quả trong việc phát hiện khối u não với tốc độ xử lý nhanh và độ chính xác cao. Kiến trúc mạng đơn giản và quy trình dự đoán một lần của YOLO đã giúp giảm thiểu thời gian tính toán, phù hợp với các ứng dụng y học yêu cầu phản hồi nhanh chóng.

Kết quả của đồ án này là sự minh họa cho sự tiến bộ đáng kể mà trí tuệ nhân tạo đem lại cho lĩnh vực y học, đặc biệt là trong việc phát hiện và chẩn đoán bệnh lý. Các phương pháp học sâu như YOLO và Faster R-CNN không chỉ mang lại hiệu suất cao trong phát hiện khối u não mà còn mở ra tiềm năng lớn trong việc phát triển các công nghệ hỗ trợ y học trong tương lai.

Tài liệu tham khảo

1. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).
2. Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).
3. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28.
4. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
5. Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7263-7271).
6. Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.
7. Bochkovski, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.
8. Couturier, R., Noura, H. N., Salman, O., & Sider, A. (2021). A deep learning object detection method for an efficient clusters initialization. arXiv preprint arXiv:2104.13634.
9. Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., ... & Wei, X. (2022). YOLOv6: A single-stage object detection framework for industrial applications. arXiv preprint arXiv:2209.02976.
10. Wang, C. Y., Bochkovski, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 7464-7475).
11. Reis, D., Kupec, J., Hong, J., & Daoudi, A. (2023). Real-time flying object detection with YOLOv8. arXiv preprint arXiv:2305.09972.

12. Wang, C. Y., Yeh, I. H., & Liao, H. Y. M. (2024). YOLOv9: Learning what you want to learn using programmable gradient information. arXiv Prepr. arXiv preprint arXiv:2402.13616.