

THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo:

<https://youtu.be/gRA6sV6ymoU>

- Link slides:

<https://github.com/thinguyen243/CS2205.FEB2025/blob/main/Slide.pdf>

- Họ và Tên: Nguyễn Ngọc Lan Thi

- MSHV: 240101077



- Lớp: CS2205.FEB2025

- Tự đánh giá (điểm tổng kết môn): 9/10

- Số buổi vắng: 0

- Số câu hỏi QT cá nhân: 3

- Link Github:

<https://github.com/thinguyen243/CS2205.FEB2025/>

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

PHÂN ĐOẠN ẢNH Y KHOA TỰ ĐỘNG DỰA TRÊN YOLO-NAS VÀ MedSAM2
KẾT HỢP NÉN MÔ HÌNH SAU HUẤN LUYỆN

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

AUTOMATED MEDICAL IMAGE SEGMENTATION VIA YOLO-NAS AND
MEDSAM2 WITH POST-TRAINING COMPRESSION

TÓM TẮT

Đề tài nghiên cứu này phát triển hệ thống phân đoạn ảnh y khoa tự động kết hợp YOLO-NAS và MedSAM2. Vấn đề của MedSAM2 là cần prompt thủ công, hạn chế tính tự động. Giải pháp là sử dụng YOLO-NAS để phát hiện tổn thương, sinh bounding box, tạo prompt tự động cho MedSAM2, loại bỏ tác vụ thủ công và giảm sai lệch. Ngoài ra, kỹ thuật nén mô hình hậu huấn luyện được áp dụng để giảm kích thước mô hình mà vẫn đảm bảo hiệu năng. Hệ thống được huấn luyện và đánh giá trên 20.000 dữ liệu ảnh y khoa và 1.000 videos y khoa từ các dataset NIH ChestX-ray14, BraTS và Medical Segmentation Decathlon. Kết quả sẽ được so sánh với các mô hình như FastSAM, MobileSAM và nnU-Net về độ chính xác, kích thước mô hình và tốc độ xử lý.

GIỚI THIỆU

Phân đoạn ảnh y khoa là bước quan trọng trong phân tích hình ảnh y tế, hỗ trợ chẩn đoán và điều trị lâm sàng. MedSAM2 – một biến thể mạnh của Segment Anything Model – cho kết quả phân đoạn chính xác cao, nhưng vẫn cần prompt thủ công, gây hạn chế trong tính tự động, hiệu quả và độ tin cậy. Ngoài ra, kích thước lớn của các mô hình phân đoạn hiện nay là rào cản lớn khi triển khai trên thiết bị có tài nguyên hạn chế (như thiết bị di động hoặc hệ thống tại cơ sở y tế nhỏ).

Để khắc phục các hạn chế này, nghiên cứu đề xuất tích hợp YOLO-NAS – một mô hình phát hiện đối tượng nhẹ, hiệu quả – nhằm sinh tự động bounding box, đóng vai

trò như prompt đầu vào cho MedSAM2. Điều này giúp loại bỏ sự phụ thuộc vào thao tác thủ công và giảm nguy cơ sai lệch do prompt không chính xác. Đồng thời, nghiên cứu áp dụng kỹ thuật Quantization để nén mô hình sau huấn luyện, giảm kích thước và tài nguyên tính toán, trong khi vẫn giữ được độ chính xác chấp nhận được.

Nghiên cứu tập trung trả lời ba câu hỏi chính:

- Có thể kết hợp YOLO-NAS và MedSAM2 để tạo pipeline phân đoạn ảnh y khoa hoàn toàn tự động, thay thế cho prompt thủ công không?
- Độ chính xác của bounding box do YOLO-NAS sinh ra có tương đương hoặc gần với các prompt thủ công trong phân đoạn không?
- Việc áp dụng Quantization ảnh hưởng thế nào đến độ chính xác và tốc độ của hệ thống phân đoạn?

Hệ thống sẽ được huấn luyện và đánh giá trên các tập dữ liệu chuẩn như NIH ChestX-ray14, BraTS, và Medical Segmentation Decathlon, với đánh giá dựa trên các chỉ số như IoU, Dice, tốc độ xử lý, kích thước mô hình, và được so sánh với các mô hình hiện có như FastSAM, MobileSAM và nnU-Net. Kỳ vọng kết quả sẽ chứng minh tính hiệu quả và khả thi cao của giải pháp đề xuất trong bối cảnh triển khai thực tế.

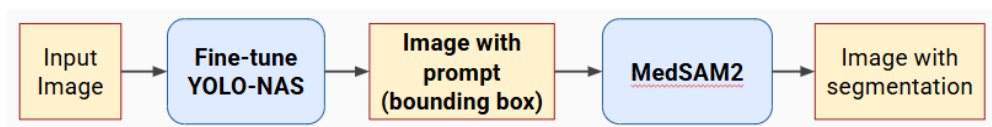
MỤC TIÊU

Thiết kế và xây dựng pipeline phân đoạn ảnh y khoa tự động bằng cách kết hợp YOLO-NAS để sinh prompt cho MedSAM2.

Đánh giá hiệu quả và độ chính xác của bounding box sinh tự động từ YOLO-NAS trên tập dữ liệu có gán nhãn thực.

Ứng dụng kỹ thuật Quantization nhằm giảm kích thước mô hình và phân tích ảnh hưởng đến hiệu năng tổng thể của hệ thống.

NỘI DUNG VÀ PHƯƠNG PHÁP



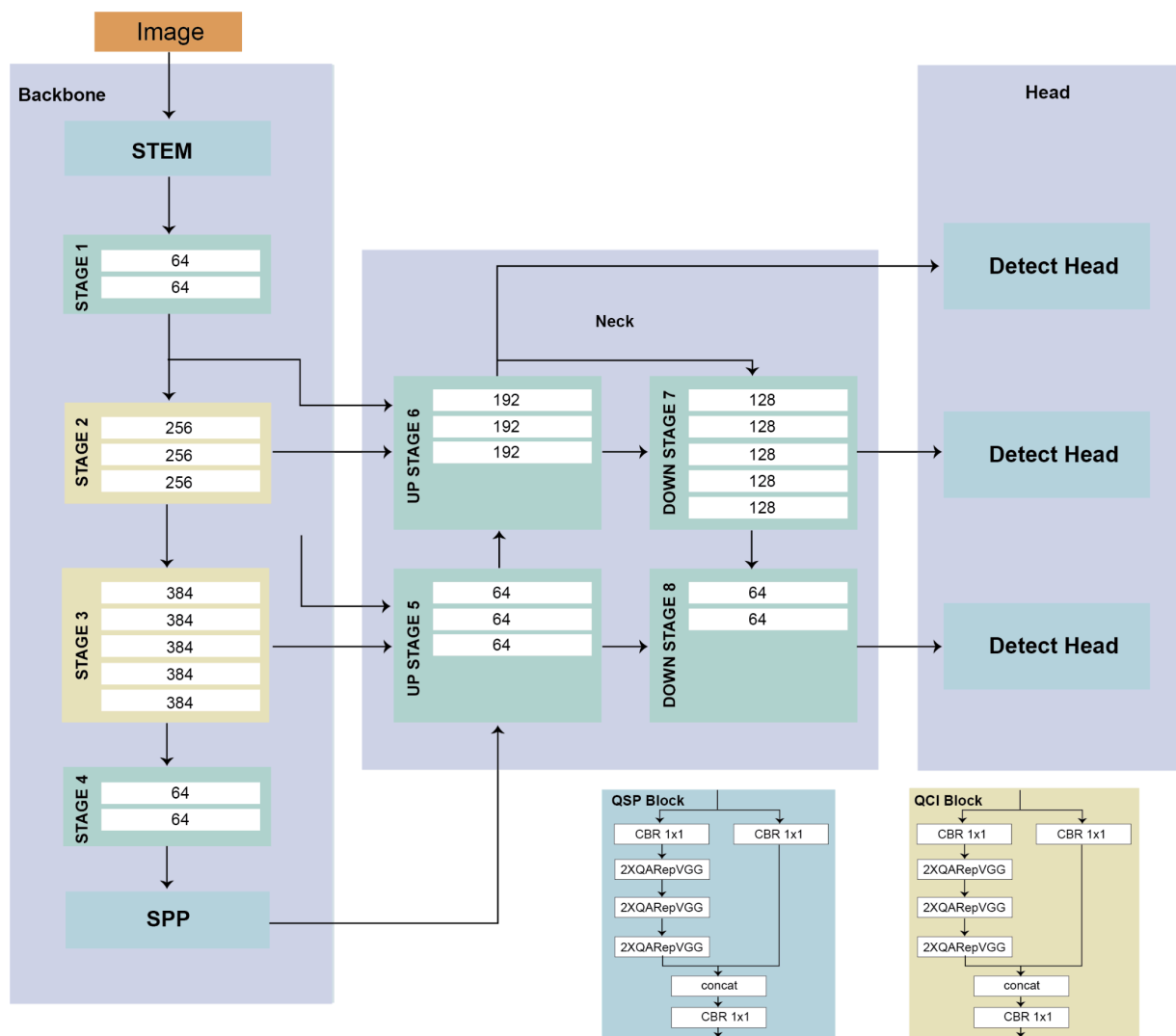
Hình 1: Pipeline phân đoạn ảnh y khoa tự động

1. Thu thập và chuẩn bị dữ liệu:

Khoảng 20.000 ảnh và 1.000 video y khoa từ các nguồn đáng tin cậy như NIH ChestX-ray14, BraTS và Medical Segmentation Decathlon sẽ được thu thập, làm sạch và chuẩn hóa về kích thước 640x640 pixel. Tập dữ liệu sẽ được chia theo 70% train, 20% validate và 10% test.

2. Huấn luyện mô hình phát hiện (YOLO-NAS):

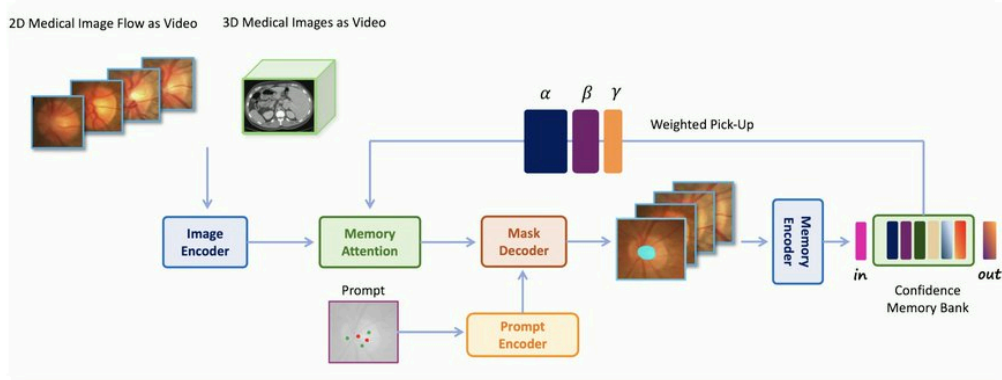
Sau đó, mô hình YOLO-NAS được huấn luyện trên các ảnh đã chuẩn hóa để phát hiện các vùng nghi ngờ bệnh lý thông qua bounding boxes. Việc huấn luyện được thực hiện trên GPU RTX 3080, 10GB VRAM với 100 epoch, batch size 4 và learning rate $1e-3$. Kết quả cho ra là danh sách bounding box $[x, y, width, height]$ và độ tin cậy (confidence).



Hình 2. Kiến trúc mô hình YOLO-NAS

3. Phân đoạn với MedSAM2 sử dụng bounding box tự sinh:

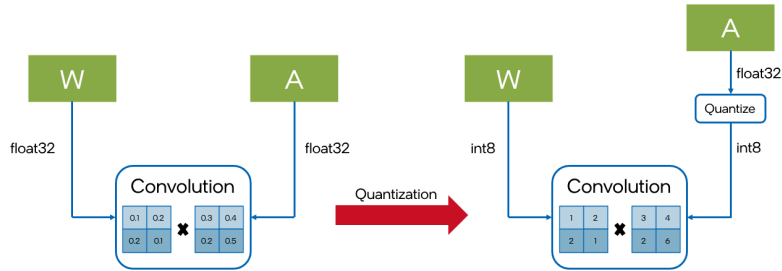
Giai đoạn tiếp theo, từ các bounding boxes đầu ra của YOLO-NAS, mô hình MedSAM2 (SAM2 đã tinh chỉnh với dữ liệu y khoa) sẽ sử dụng thông tin này như prompt để thực hiện phân đoạn chi tiết đối tượng bệnh lý, so sánh với prompt gốc để đánh giá mức độ tương thích. Đầu ra là mask phân đoạn với cùng kích thước ảnh đầu vào (640x640).



Hình 3. Kiến trúc mô hình MedSAM2

4. Giảm kích thước mô hình:

Cả hai mô hình sẽ được nén bởi Post-Training Quantization (INT8) để giảm kích thước, giảm tiêu thụ tài nguyên mà vẫn giữ độ chính xác cao.



Hình 4. Kiến trúc mô hình Post-Training Quantization

5. Thử nghiệm và đánh giá:

- Đo các chỉ số: IoU, Dice, tốc độ xử lý trung bình mỗi ảnh, kích thước mô hình.

$$dice(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad IOU(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

- So sánh với các pipeline khác: FastSAM, MobileSAM, nnU-Net.

KẾT QUẢ MONG ĐỢI

Độ chính xác cao: trên 75% theo IoU và trên 85% theo Dice Score

Tốc độ suy luận nhanh: dưới 500ms mỗi ảnh

Tối ưu bộ nhớ tốt: dưới 1.5GB RAM, dưới 100MB bộ nhớ cho mô hình

Bảng so sánh mong đợi so với các mô hình khác:

Mô hình	IoU/Dice	Kích thước	Prompt thủ công	Tốc độ xử lý
YOLO-NAS + MedSAM2 + quantization	>0.75 / >0.85	≤ 100MB	Không	< 500ms
FastSAM	0.65 / 0.72	Khoảng 120MB	Không	Trung bình
MobileSAM	0.60 / 0.68	Khoảng 90MB	Không	Nhanh
nnU-Net	0.78 / 0.83	> 200MB	Có	Chậm

TÀI LIỆU THAM KHẢO

[1]. Jiayuan Zhu et al.: Medical SAM 2: Segment medical images as video via Segment Anything Model 2. arXiv:2408.00874, 2024

[2]. Juan Terven, Diana Cordova-Esparza: A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. arXiv:2304.00501, 2023

[3]. Isensee, F. et al.: nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation. MICCAI, 2024.

[4]. Victor Ion Butoi et al.: UniverSeg: Universal Medical Image Segmentation. ICCV, 2023, pp. 21438-21451

[5]. R. Azad et al.: Medical Image Segmentation Review: The Success of U-Net. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 46, no. 12, 2024, pp. 10076-10095

[6]. Chengtao Lv, Hong Chen, Jinyang Guo, Yifu Ding, Xianglong Liu: PTQ4SAM: Post-Training Quantization for Segment Anything. CVPR, 2024, pp. 15941-15951