



Phân đoạn u não dựa trên dữ liệu ảnh MRI

Lê Mạnh Cương¹ Nguyễn Tuấn Anh² Phạm Quý Đô³

Khoa Toán - Cơ - Tin học
Trường Đại học Khoa học Tự Nhiên - ĐHQGHN

Ngày 7 tháng 1 năm 2026

Nội dung

1 Giới thiệu bài toán

2 Dữ liệu

- Tiền xử lý
- Thống kê & phân tích dữ liệu

3 Phương pháp

- Quy trình tổng quan
- UNet
- UNet++
- VNet
- UNETR
- Swin UNet3D

4 Thực nghiệm

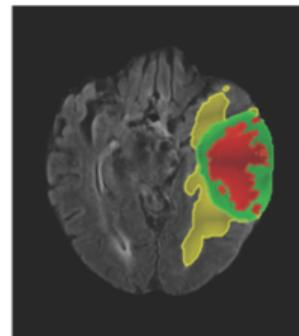
- Các độ đo đánh giá
- Các thử nghiệm thực hiện

5 Kết quả

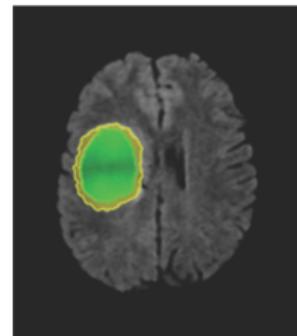
Giới thiệu bài toán phân đoạn u não trên ảnh MRI

- **U não** là sự tăng sinh bất thường của tế bào trong hộp sọ, gây ảnh hưởng trực tiếp đến hệ thần kinh trung ương.
- **Glioma** là nhóm u ác tính phổ biến nhất:
 - **LGG:** tiến triển chậm, thời gian sống trung vị ≈ 11.6 năm
 - **HGG:** tiến triển nhanh, thời gian sống trung vị ≈ 15 tháng
- Ý nghĩa lâm sàng: phát hiện sớm – đo kích thước – xác định biên u – lập kế hoạch điều trị.
- Phân đoạn thủ công: tốn nhiều thời gian, công sức \rightarrow nhu cầu về các phương pháp phân đoạn u não tự động.

High Grade Glioma



Low Grade Glioma



Ảnh MRI đa phương thức và các phân vùng khối u

Ưu điểm của MRI:

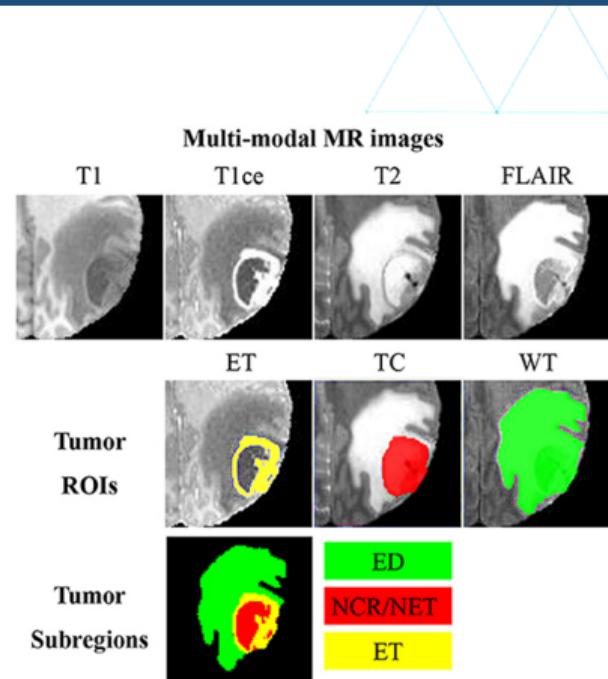
- Không bức xạ ion hóa
- Độ tương phản mô mềm cao
- Có nhiều chuỗi xung → phản ánh đặc tính sinh học khác nhau.

Các chuỗi thường dùng:

- T1, T1ce (tiêm tương phản)
- T2, Flair

Các phân vùng u trong ảnh MRI:

- ET (Enhancing Tumor): tăng tín hiệu mạnh trên T1ce
- NCR/NET: mô hoại tử hoặc không tăng quang
- ED (Edema): vùng phù quanh u, tăng tín hiệu trên Flair.



Bài toán phân đoạn u não từ ảnh MRI

- **Đầu vào:** Ảnh MRI não dạng 2D hoặc 3D, đa chuỗi (T1, T1CE, T2, FLAIR)
- **Đầu ra:** Bản đồ phân đoạn gồm ET, TC, WT
- **Mục tiêu:**
 - Xác định chính xác ranh giới u
 - Giảm phụ thuộc vào vẽ nhãn thủ công (tốn thời gian & mang tính chủ quan)
 - Hỗ trợ lập kế hoạch điều trị và tiên lượng bệnh

Nội dung

1 Giới thiệu bài toán

2 Dữ liệu

- Tiền xử lý
- Thống kê & phân tích dữ liệu

3 Phương pháp

- Quy trình tổng quan
- UNet
- UNet++
- VNet
- UNETR
- Swin UNet3D

4 Thực nghiệm

- Các độ đo đánh giá
- Các thử nghiệm thực hiện

5 Kết quả

Bộ dữ liệu

❑ BraTS2020:

- Bộ dữ liệu chuẩn cho bài toán phân đoạn u não trên MRI đa chuỗi.
- Bao gồm ảnh MRI trước phẫu thuật của bệnh nhân glioma từ nhiều trung tâm lâm sàng

❑ Quy mô dữ liệu:

- 369 ca có nhãn, 125 ca không nhãn
- Mỗi ca chứa 4 chuỗi MRI
- Với tập có nhãn đi kèm bản đồ phân đoạn được gán nhãn thủ công bởi chuyên gia

❑ Cấu trúc nhãn: 0 – Background, 1 -- NCR/NET, 2 — ED, 4 — ET

❑ Các bước tiền xử lý đã có trong dataset:

- Loại bỏ hộp sọ, chuẩn hóa không gian
- Resample về độ phân giải $1 \times 1 \times 1 \text{ mm}^3$; chuẩn hóa kích thước về $240 \times 240 \times 155$ voxel

Tất cả dữ liệu lưu dưới định dạng NIfTI (.nii.gz).

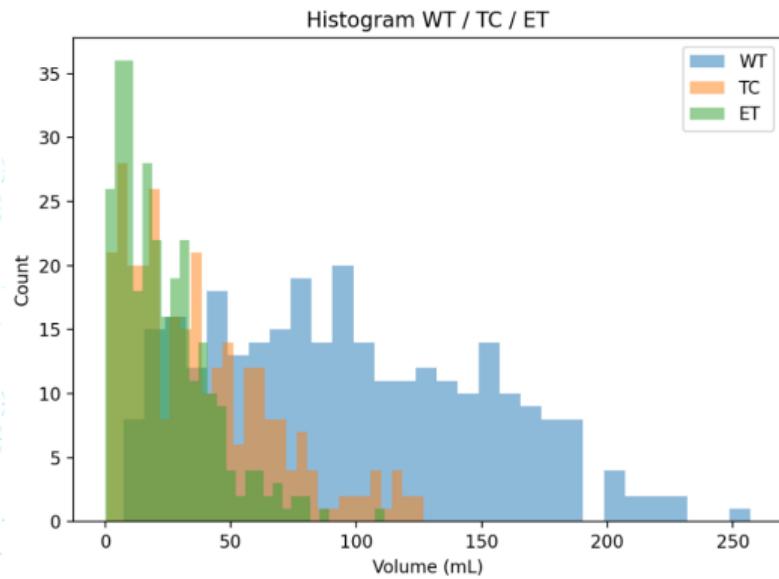
Tiền xử lý dữ liệu 2D & tách lát cắt

- ① Xác định bounding box 2D vuông chứa toàn bộ vùng não trên mặt phẳng (x, y)
- ② Chuẩn hóa cường độ bằng percentile
Tính 1-percentile và 99-percentile \rightarrow cắt ngưỡng \rightarrow chuẩn hóa về $[0, 1]$.
- ③ Cắt thành các lát cắt \rightarrow crop \rightarrow resize về 256×256 .
Ảnh MRI dùng interpolation tuyến tính, mask dùng nearest neighbor.
- ④ Ánh xạ nhãn 4 \rightarrow 3

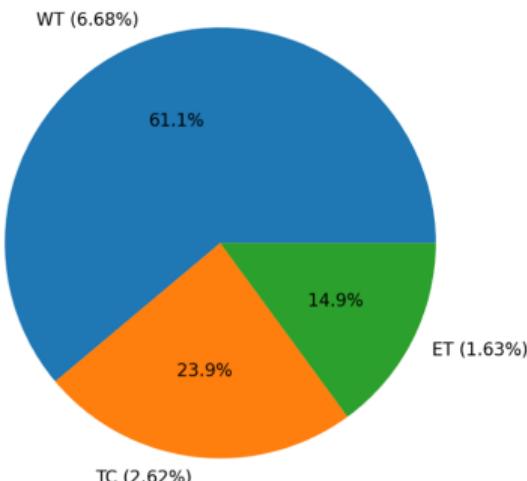
Tiền xử lý dữ liệu 3D

- ① Crop thể tích MRI bằng bounding box đã xác định trước đó, giữ nguyên chiều sâu.
- ② Chuẩn hóa cường độ voxel theo Z-score
- ③ Ánh xạ nhãn $4 \rightarrow 3$

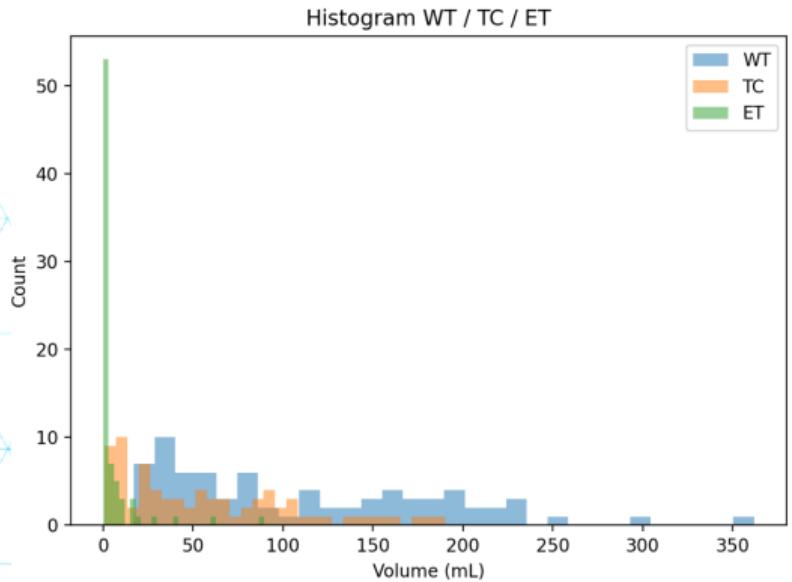
Phân bố thể tích các vùng u của nhóm HGG



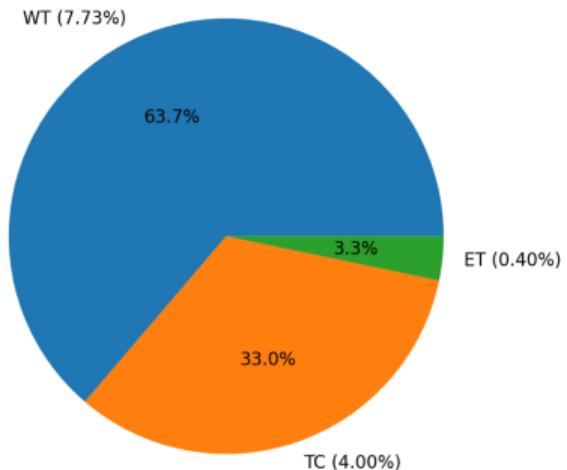
Mean % of WT / TC / ET relative to brain volume (overlapping sets)



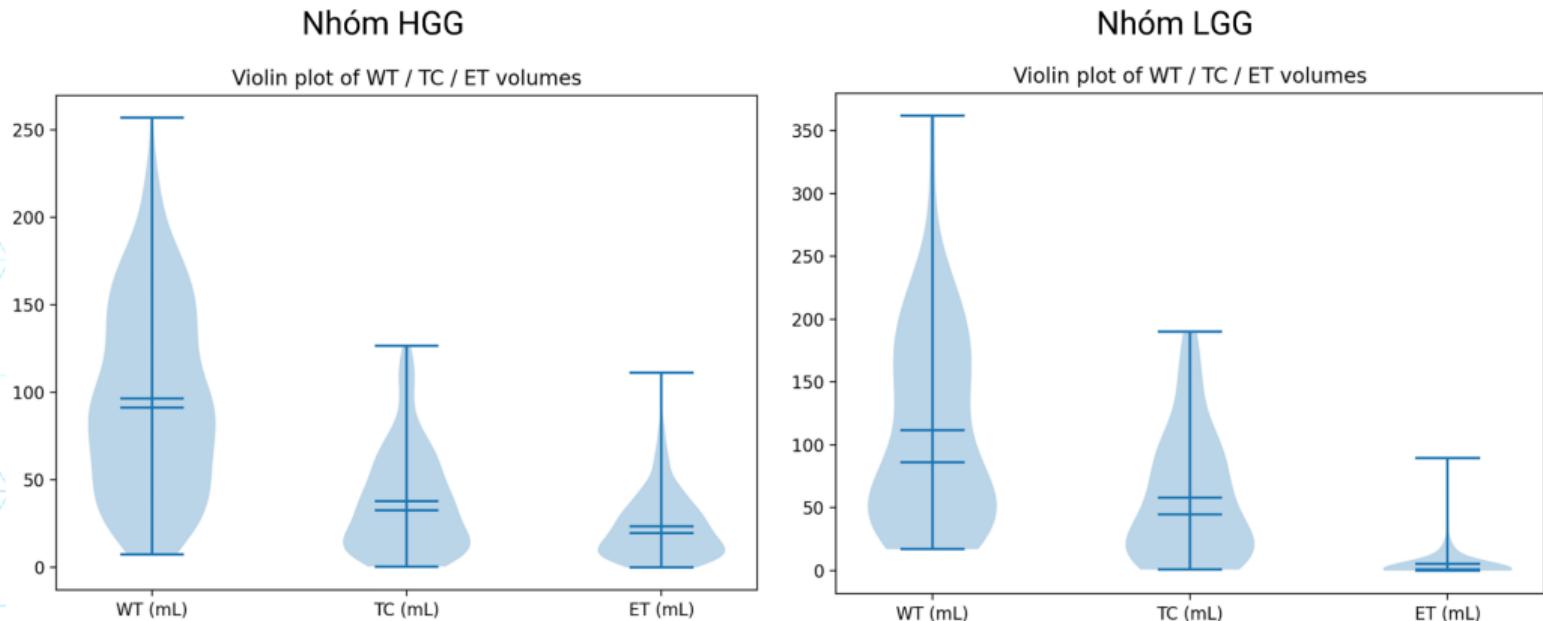
Phân bố thể tích các vùng u của nhóm LGG



Mean % of WT / TC / ET relative to brain volume (overlapping sets)



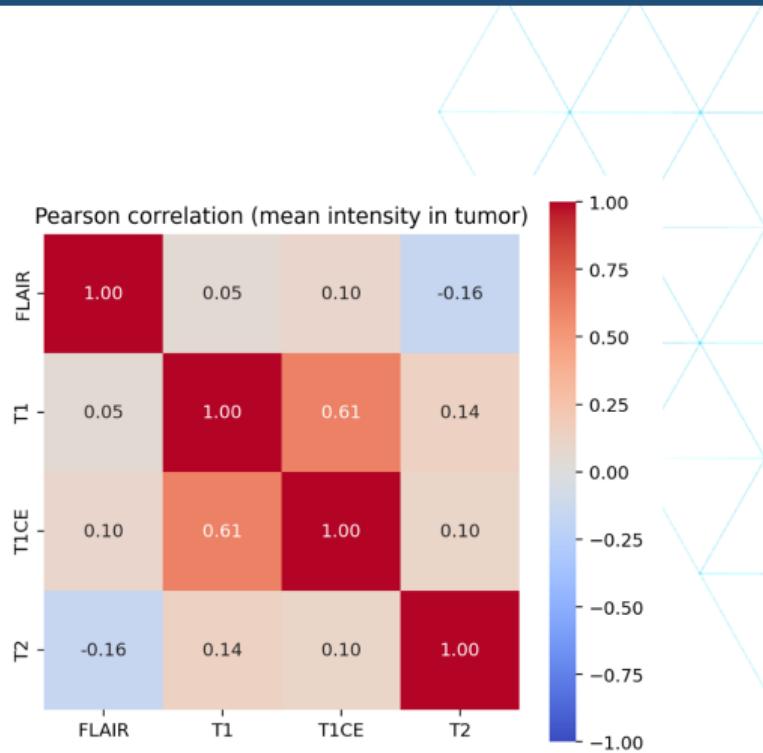
So sánh 2 nhóm HGG vs LGG



⇒ Nhận xét

Tương quan giữa các modality trong vùng khối u

- T1 và T1ce có tương quan mạnh ($r = 0.61$)
 - T1ce là phiên bản T1 có tiêm gadolinium → tăng tín hiệu ở vùng tăng sinh ET
 - Mức tương quan cao nhưng không trùng lặp hoàn toàn
- Flair có tương quan thấp với các chuỗi còn lại
Phản ánh đặc tính sinh lý khác biệt:
 - Flair nhạy với vùng phù quanh u ED
 - T1ce, T1 nhạy với cấu trúc/tăng sinh mạch→ Flair bổ sung quan trọng cho mô tả ranh giới toàn bộ khối u WT.



Phân chia dữ liệu

Tập	Tỷ lệ	Số ca chụp
Tập train	70%	258
Tập validation	15%	55
Tập test	15%	56

Phân tầng dự trên 3 yếu tố:

- Độ ác tính: HGG, LGG
- Sự hiện diện của vùng ET
- Kích thước khối u: phân nhóm theo quantile.

⇒ Giúp duy trì phân bố đặc trưng đồng đều

Nội dung

1 Giới thiệu bài toán

2 Dữ liệu

- Tiền xử lý
- Thông kê & phân tích dữ liệu

3 Phương pháp

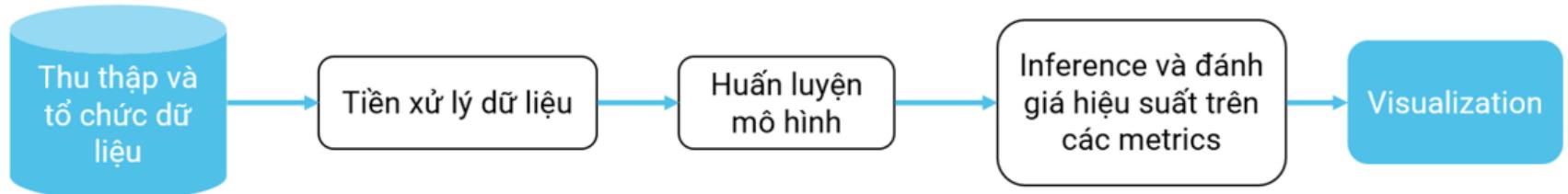
- Quy trình tổng quan
- UNet
- UNet++
- VNet
- UNETR
- Swin UNet3D

4 Thực nghiệm

- Các độ đo đánh giá
- Các thử nghiệm thực hiện

5 Kết quả

Quy trình tổng quan



UNet

Các mô hình phân đoạn hiện đại thường dựa trên kiến trúc encoder-decoder, tiêu biểu là UNet.

- Nhánh encoder: trích xuất đặc trưng + giảm dần độ phân giải không gian
- Nhánh decoder: khôi phục lại độ phân giải ban đầu.

Skip connection

Kết nối đặc trưng có độ phân giải cao, mức độ trừu tượng thấp từ encoder sang các đặc trưng có mức độ trừu tượng cao hơn từ decoder.



Skip connection nguyên bản của UNet: nối trực tiếp feature từ encoder sang decoder dù 2 loại feature này có mức độ trừu tượng rất khác nhau → mô hình học khó hơn.

UNet++ (2018) ra đời giải quyết vấn đề đó bằng việc:

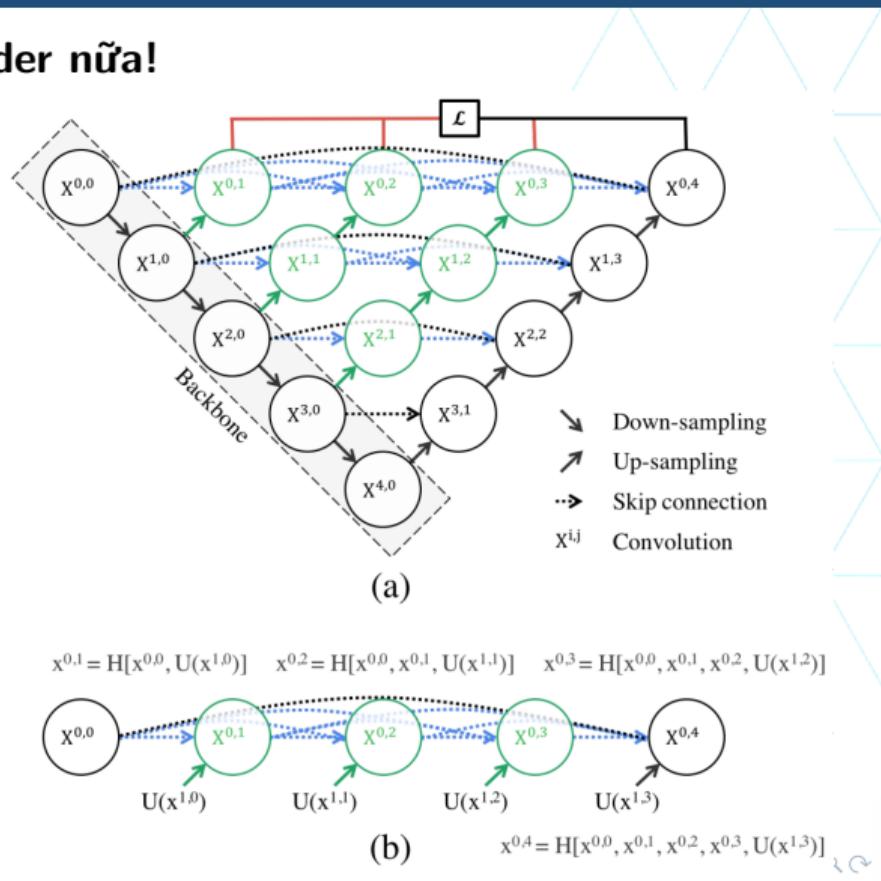
- Thiết kế lại các skip connections
- Deep supervision
- Cho phép pruning linh hoạt.

Skip connections được thiết kế trong UNet++

UNet++ không nối thẳng encoder với decoder nữa!

- Mỗi đường skip sẽ trải qua một chuỗi các khối **dense convolution**: nối tất cả các đầu vào trước đó + feature được upsample từ tầng sâu hơn.
⇒ Nâng dần mức ngữ nghĩa của feature encoder để tiệm cận decoder trước khi ghép vào nhau → *mạng học dễ dàng hơn, kết quả phân đoạn chi tiết hơn.*

Mỗi node skip càng sâu thì càng giàu thông tin và tiệm cận decoder.



Deep Supervision

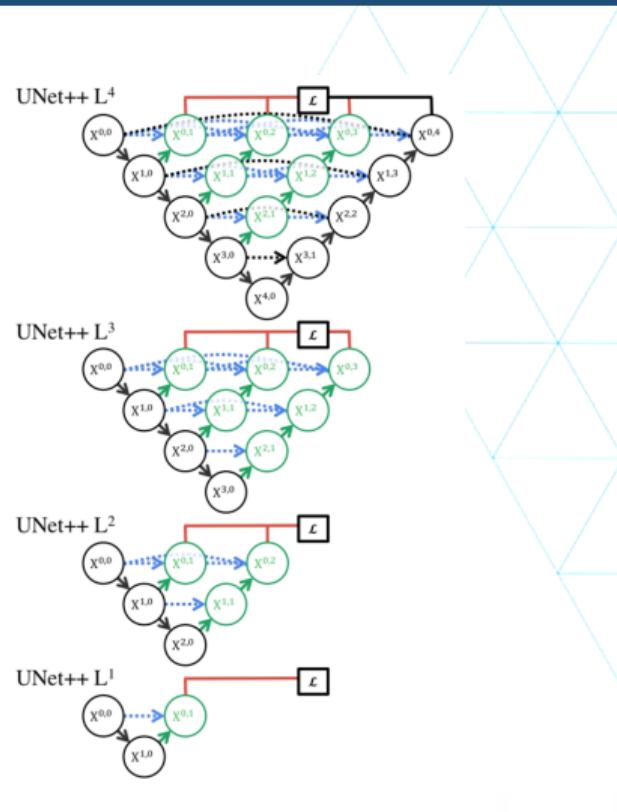
Mục đích của Deep supervision:

- Giúp gradient lan truyền trực tiếp vào các tầng giữa → huấn luyện nhanh và ổn định hơn
- Cho phép lựa chọn nhiều đầu ra phân đoạn tương ứng với các độ sâu khác nhau

2 chế độ:

- **Accurate mode:** Dùng tất cả các nhánh phân đoạn, tổng hợp hoặc lấy trung bình → kết quả chính xác nhất.
- **Fast mode:**
 - Chỉ chọn 1 nhánh L^1, L^2 hoặc L^3 để inference
 - Các nhánh còn lại có thể cắt bỏ (prune)

→ Cắt ngắn mô hình để giảm thời gian inference mà không phải huấn luyện lại từ đầu.



VNet (2016) được thiết kế như một mạng fully convolution 3D dành riêng cho bài toán phân đoạn thể tích trong ảnh y khoa.

- VNet học trực tiếp trên toàn bộ thể tích MRI và sinh ra mask phân đoạn 3D cùng kích thước → cho phép mô hình nắm bắt trọn vẹn ngữ cảnh không gian theo ba chiều.
- Kiến trúc được xây dựng theo dạng encoder-decoder 3D với các khối residual và cơ chế downsample, upsample sử dụng tích chập 3D.

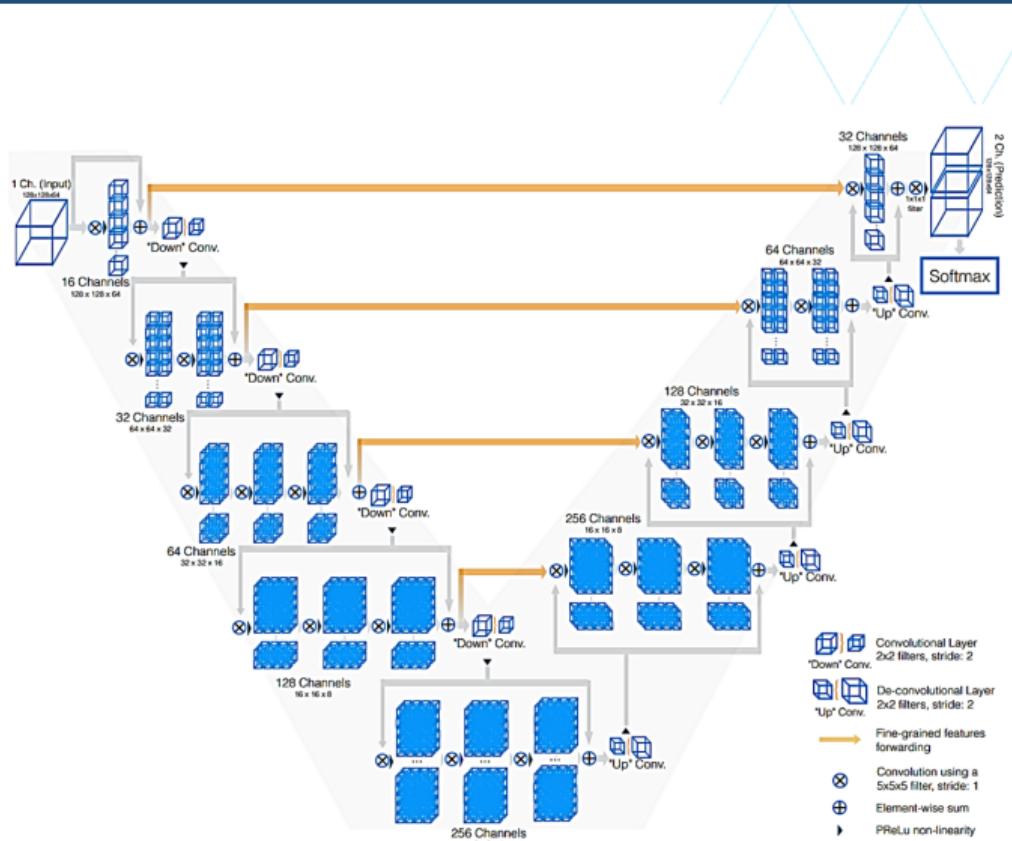
VNet

Mỗi stage của VNet:

- Sử dụng 1-3 lớp tích chập 3D với kernel $5 \times 5 \times 5$ kết hợp với hàm kích hoạt **PReLU** thay vì ReLU.
- Được xây dựng theo dạng **residual block**.

Lợi ích của residual:

- Giúp mạng sâu dễ huấn luyện hơn
- Tránh gradient vanishing
- Tăng tốc độ hội tụ và cải thiện chất lượng phân đoạn.

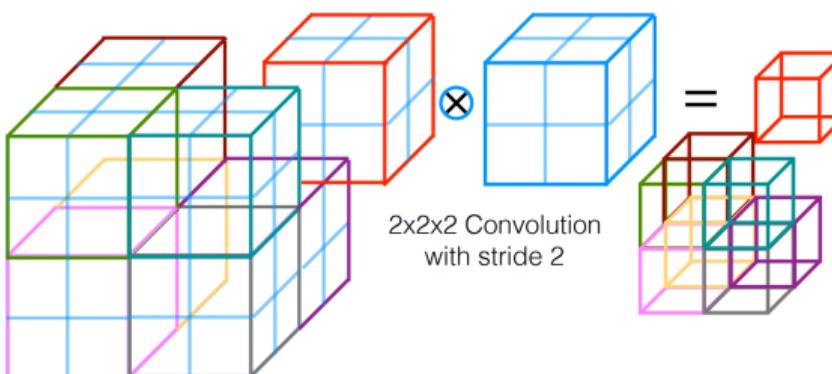


Cơ chế Downsample bằng tích chập 3D

Thay vì sử dụng max-pooling, VNet giảm độ phân giải bằng:

- 1 lớp convolution 3D với kernel $2 \times 2 \times 2$, stride = 2.
- Giảm một nửa kích thước không gian
- Tăng gấp đôi số kênh

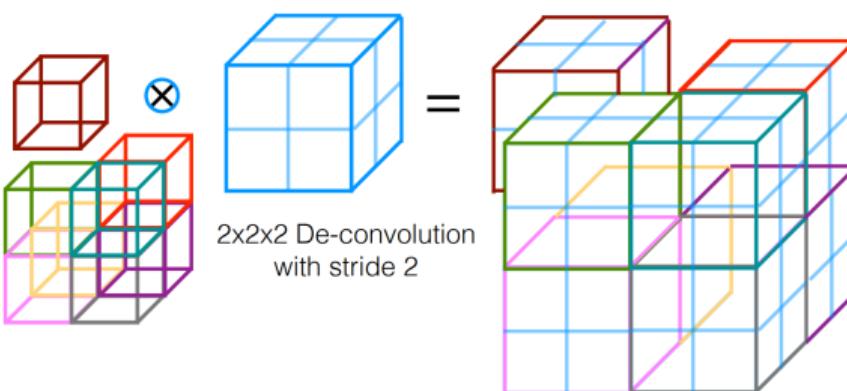
⇒ **Ưu điểm so với pooling:** ...



Cơ chế Upsample bằng deconvolution 3D

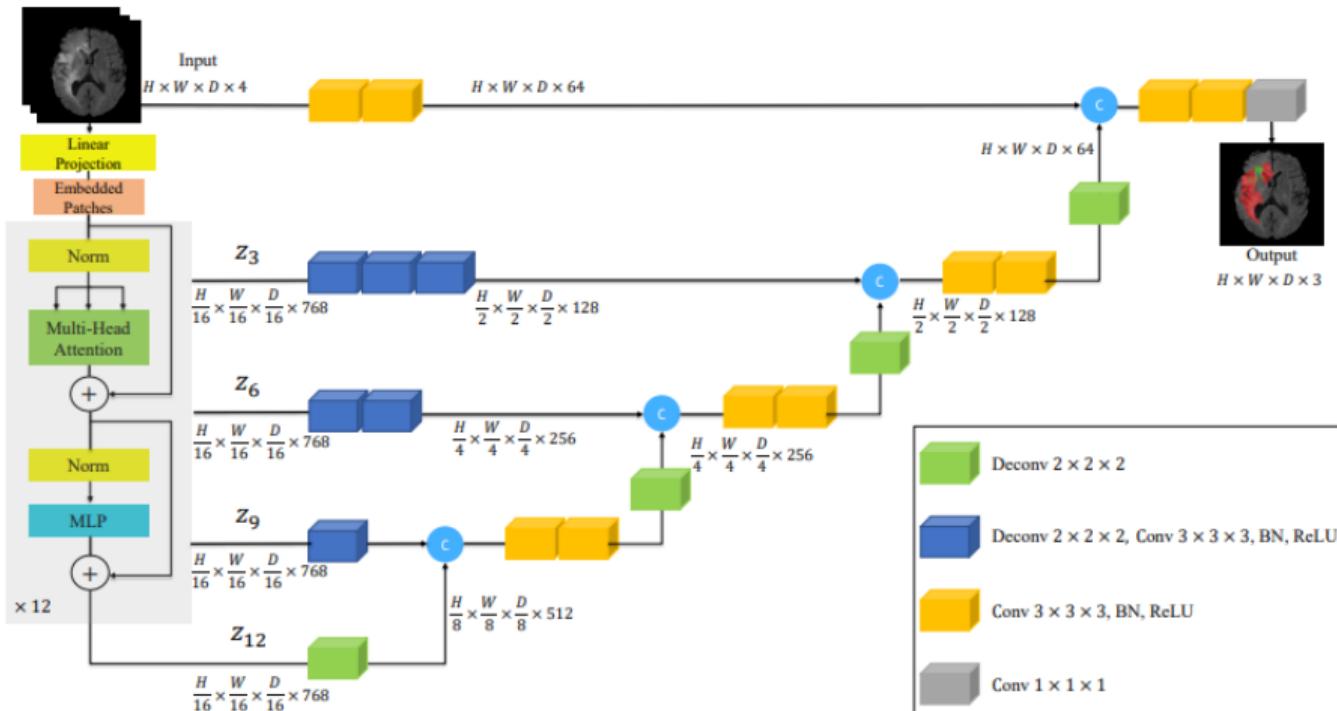
Nhánh decoder của VNet khôi phục kích thước không gian bằng:

- 3D transposed convolution kernel $2 \times 2 \times 2$, stride = 2
- Tăng độ phân giải lên gấp đôi
- Giảm số kênh đi một nửa.



UNETR (2021) là kiến trúc lai giữa Transformer và CNN dành cho phân đoạn ảnh y tế 3D. Để khắc phục hạn chế về vùng tiếp nhận cục bộ của các mạng tích chập truyền thống, UNETR sử dụng encoder Transformer để mã hóa ảnh dưới dạng chuỗi patch 3D, giúp học hiệu quả các phụ thuộc dài hạn và ngữ cảnh toàn cục. Các biểu diễn này được kết hợp với một decoder CNN thông qua hệ thống skip connection đa độ phân giải, giúp mô hình vừa nắm bắt thông tin diện rộng, vừa duy trì được các chi tiết không cục bộ để dự đoán kết quả phân đoạn chính xác.

Kiến trúc UNETR



Đầu vào và Embedding

- UNETR sử dụng kiến trúc encoder-decoder dạng thu hẹp-mở rộng, lấy cảm hứng từ U-Net.
- Encoder dựa trên Transformer nhằm học các đặc trưng toàn cục từ dữ liệu ảnh y tế 3D.
- Decoder thực hiện khôi phục độ phân giải và tạo bản đồ phân đoạn, trong đó hai nhánh được kết nối với nhau thông qua skip connection.

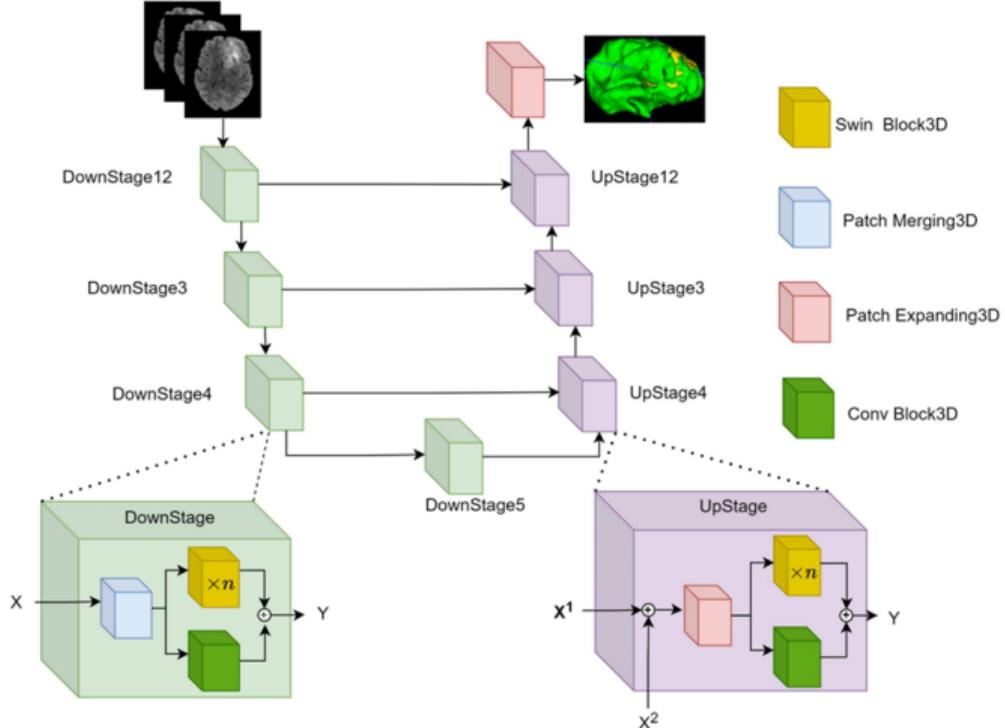
Encoder Transformer & Skip Connection

- Encoder bao gồm nhiều khôi Transformer với cơ chế multi-head self-attention và các lớp MLP.
- Layer normalization và residual connection được sử dụng nhằm ổn định và cải thiện quá trình học.
- Các đặc trưng trích xuất từ các tầng trung gian được reshape thành tensor 3D và được truyền sang decoder thông qua skip connection.

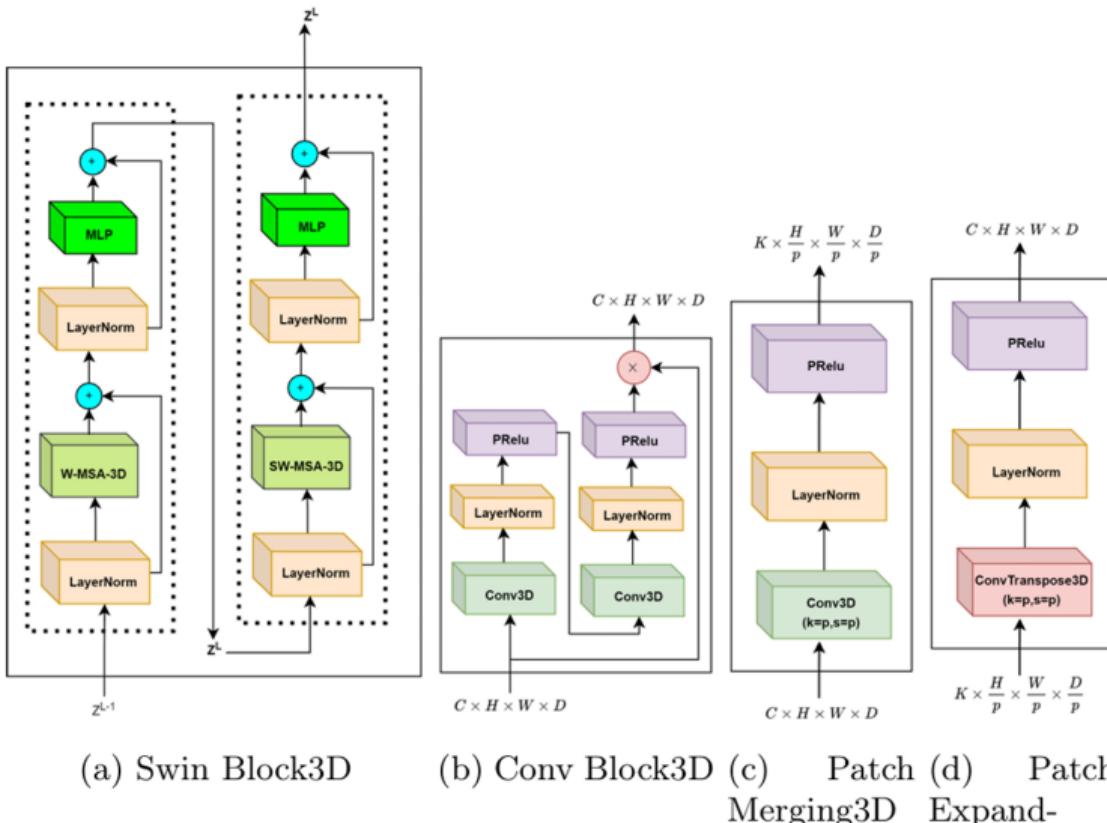
Decoder và Đầu ra

- Decoder bắt đầu từ nút thắt cổ chai (*bottleneck*), tương ứng với đầu ra Transformer cuối cùng.
- Các đặc trưng được upsampling bằng các lớp deconvolution và được kết hợp với các đặc trưng từ encoder thông qua skip connection.
- Cuối cùng, một lớp tích chập $1 \times 1 \times 1$ kết hợp với hàm softmax tạo ra bản đồ phân đoạn theo từng voxel.

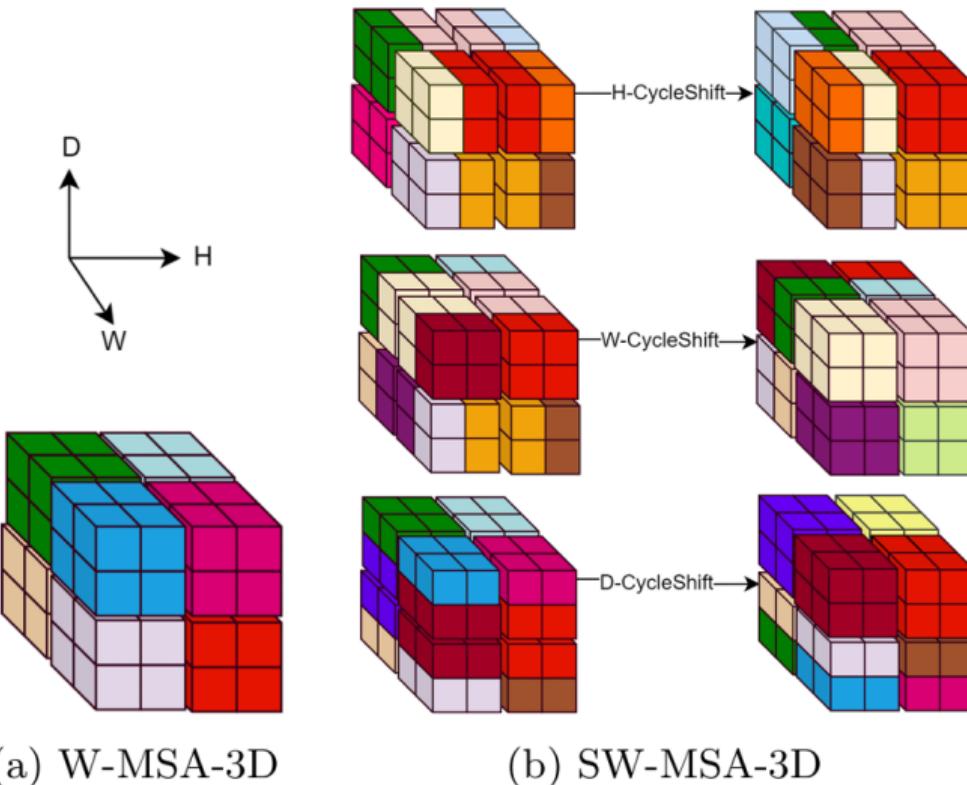
Swin UNet3D



Các module cốt lõi của Swin SwinUNet3D



Window attention 3D



Nội dung

1 Giới thiệu bài toán

2 Dữ liệu

- Tiền xử lý
- Thống kê & phân tích dữ liệu

3 Phương pháp

- Quy trình tổng quan
- UNet
- UNet++
- VNet
- UNETR
- Swin UNet3D

4 Thực nghiệm

- Các độ đo đánh giá
- Các thử nghiệm thực hiện

5 Kết quả

Các độ đo đánh giá

Đánh giá chất lượng phân đoạn trên các vùng WT, TC, ET thông qua:

- **Dice, IoU**

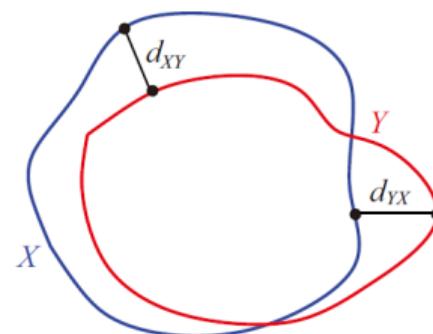
Đánh giá độ trùng khớp giữa vùng dự đoán P và vùng nhãn thực G .

- **Average Surface Distance (ASD)**

Đo khoảng cách trung bình giữa bề mặt dự đoán S_P và bề mặt nhãn thực S_G

- **HD95**

Đo mức sai lệch biên lớn nhất ở 95% điểm biên.



Thử nghiệm

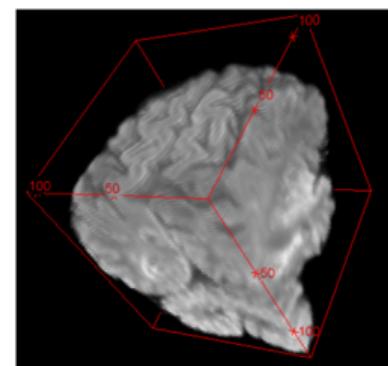
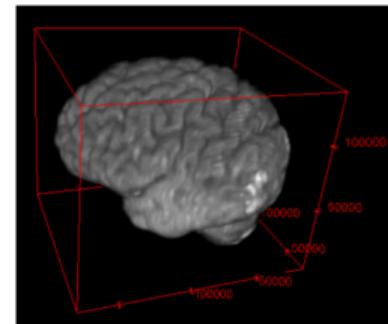
Giữa kì:

- ① Huấn luyện UNet trên các lát cắt 2D
- ② Huấn luyện Unet++ trên các lát cắt 2D.
- ③ Huấn luyện VNet/VNet multi-head/VNet multi-encoder trên toàn bộ thể tích 3D/ các patch 3D.

Cuối kì:

- ① Huấn luyện Swin UNet3D trên các patch 3D.

Toàn bộ thể tích và 1 patch của ca chụp 101

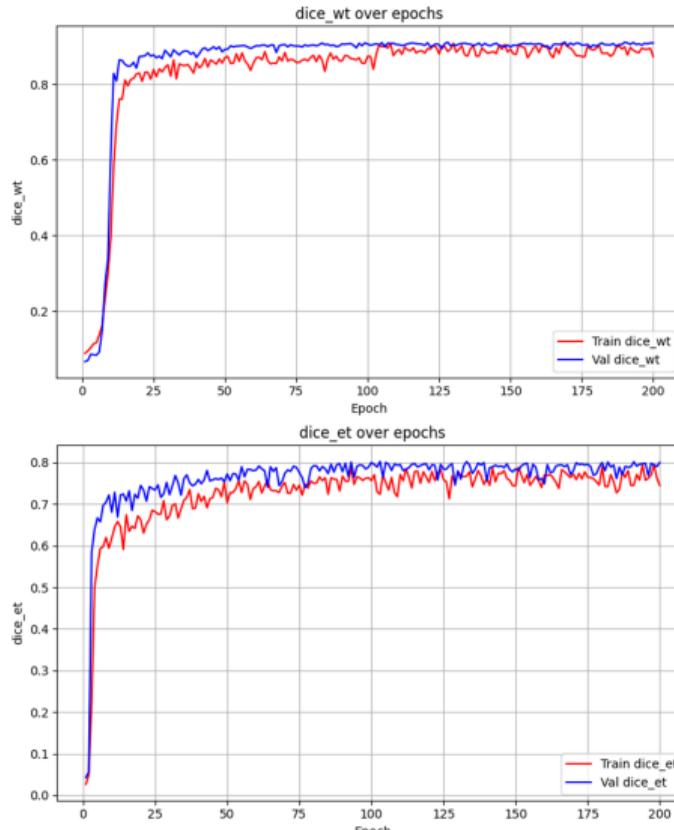
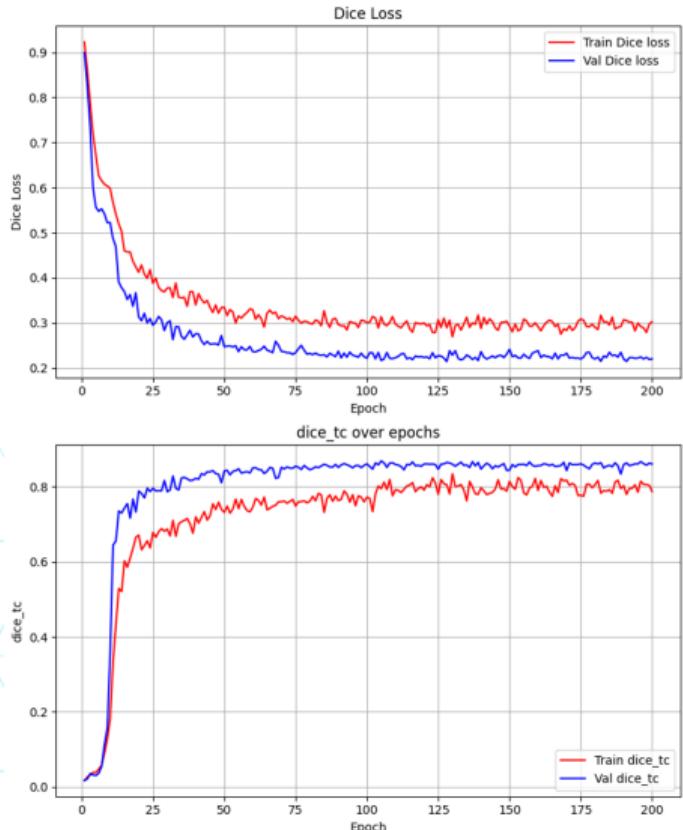


Quá trình inference

Một số thử nghiệm mô hình theo hướng patch-based khi inference sẽ sử dụng sliding window trên toàn bộ thể tích 3D để tạo ra dự đoán cuối cùng:

- Patch: $128 \times 128 \times 128$
- Stride: $20 \times 20 \times 20$
- Các vùng chồng lấn được hợp nhất bằng trung bình softmax (probability average)
- Dự đoán cuối cùng lấy argmax theo kênh.

Huấn luyện VNet trên các patch 3D sử dụng DiceLoss



Nội dung

1 Giới thiệu bài toán

2 Dữ liệu

- Tiền xử lý
- Thống kê & phân tích dữ liệu

3 Phương pháp

- Quy trình tổng quan
- UNet
- UNet++
- VNet
- UNETR
- Swin UNet3D

4 Thực nghiệm

- Các độ đo đánh giá
- Các thử nghiệm thực hiện

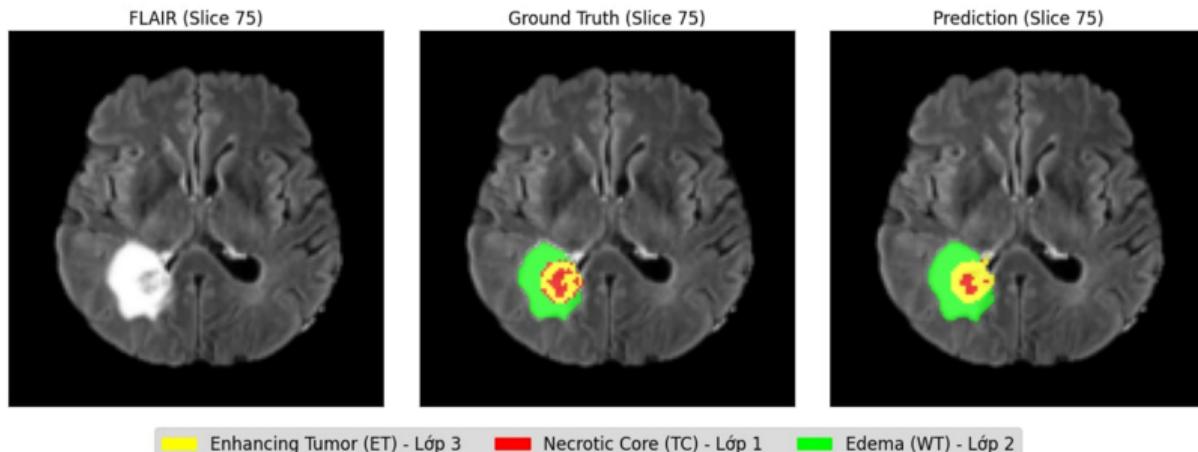
5 Kết quả

Kết quả UNet

ROI	Dice	IoU	ASD (mm)	HD95 (mm)
WT	0.8612	0.7725	3.6368	19.0979
TC	0.7549	0.6570	4.6777	12.5280
ET	0.7148	0.5992	9.1472	18.2997

Kết quả UNet++

ROI	Dice	IoU	ASD	HD95
WT	0.8954	0.8158	1.5725	5.0218
TC	0.8196	0.7297	2.0064	6.2421
ET	0.7360	0.6278	1.6917	5.0079



Hình 1: Kết quả phân đoạn của UNet++ trên lát cắt thứ 75 của ca chụp 011

Kết quả VNet

- VNet huấn luyện trên toàn bộ thể tích 3D:

ROI	Dice	IoU	ASD	HD95
WT	0.8501	0.7706	1.5178	5.8878
TC	0.7949	0.7068	2.2944	7.5172
ET	0.7265	0.6219	1.7098	5.1759

- VNet huấn luyện trên các patch 3D - DiceCELoss

Region	Dice	IoU	ASD	HD95
WT	0.8977	0.8236	3.1626	10.1072
TC	0.8353	0.7464	3.5748	9.3785
ET	0.7519	0.6556	3.4380	7.8940

Kết quả VNet

- VNet huấn luyện trên các patch 3D - DiceLoss

ROI	Dice	IoU	ASD	HD95
WT	0.9024	0.8236	2.0219	6.9727
TC	0.8633	0.7664	1.6688	5.5709
ET	0.7617	0.6650	1.6565	5.3642

- VNet Multi-head huấn luyện trên các patch 3D

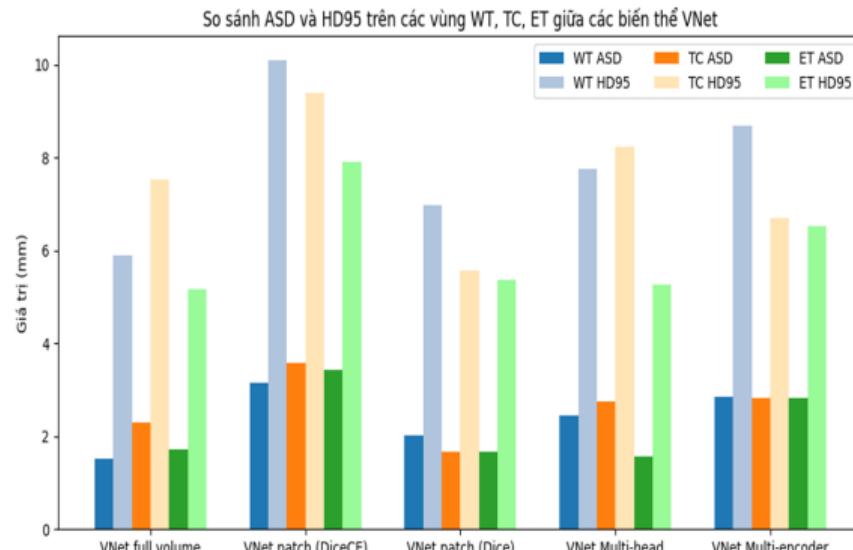
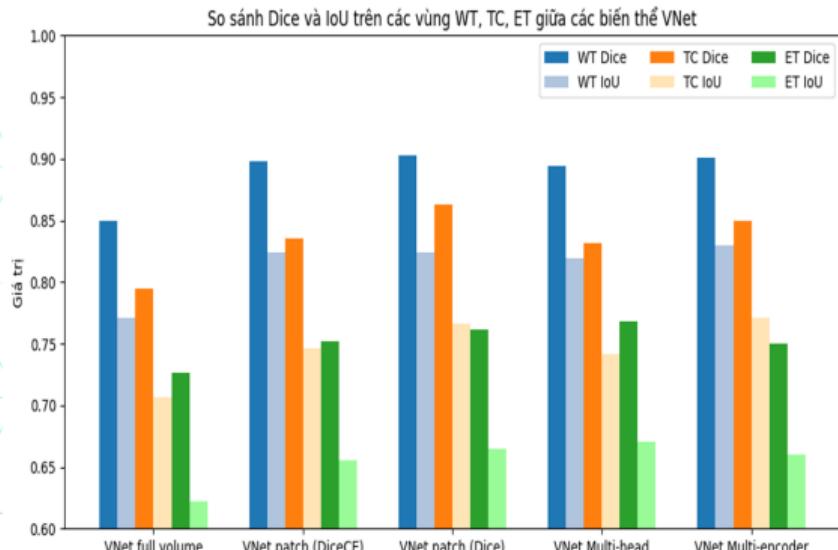
ROI	Dice	IoU	ASD	HD95
WT	0.8944	0.8195	2.4352	7.7570
TC	0.8315	0.7416	2.7558	8.2220
ET	0.7680	0.6703	1.5727	5.2660

Kết quả VNet

- **VNet Multi-encoder huấn luyện trên các patch 3D**

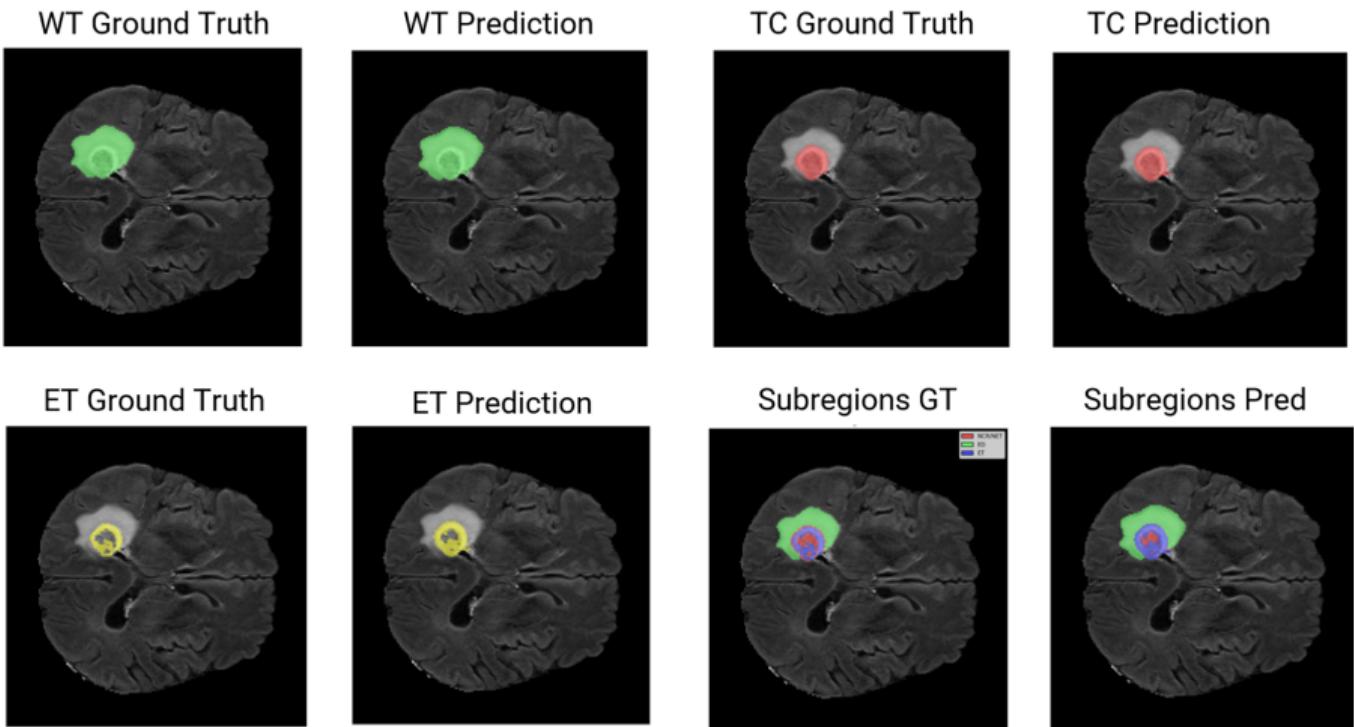
ROI	Dice	IoU	ASD	HD95
WT	0.9011	0.8300	2.8607	8.6980
TC	0.8499	0.7706	2.8147	6.6950
ET	0.7498	0.6599	2.8145	6.5197

Kết quả VNet



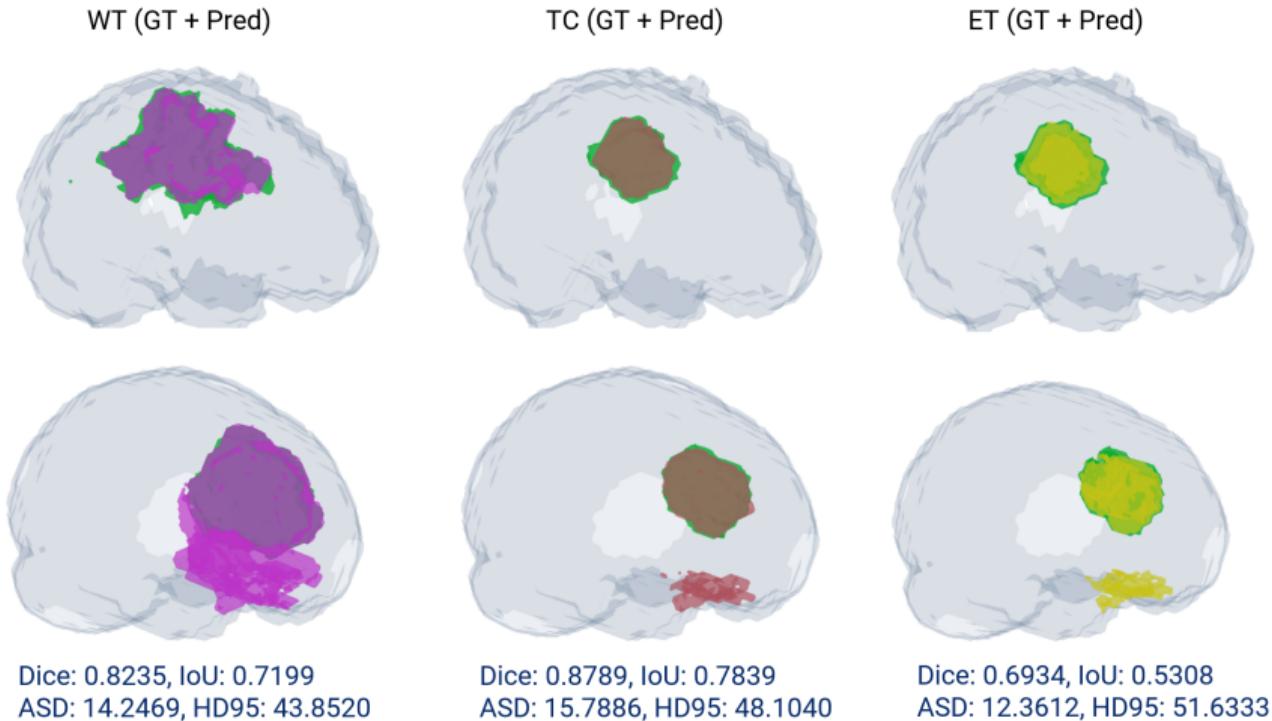
⇒ VNet huấn luyện trên các patch 3D sử dụng **DiceLoss** đạt hiệu suất cao nhất trên tất cả thử nghiệm đã thực hiện.

Kết quả VNet



Hình 2: Kết quả phân đoạn của VNet trên lát cắt thứ 75 của ca chụp 011

Kết quả VNet

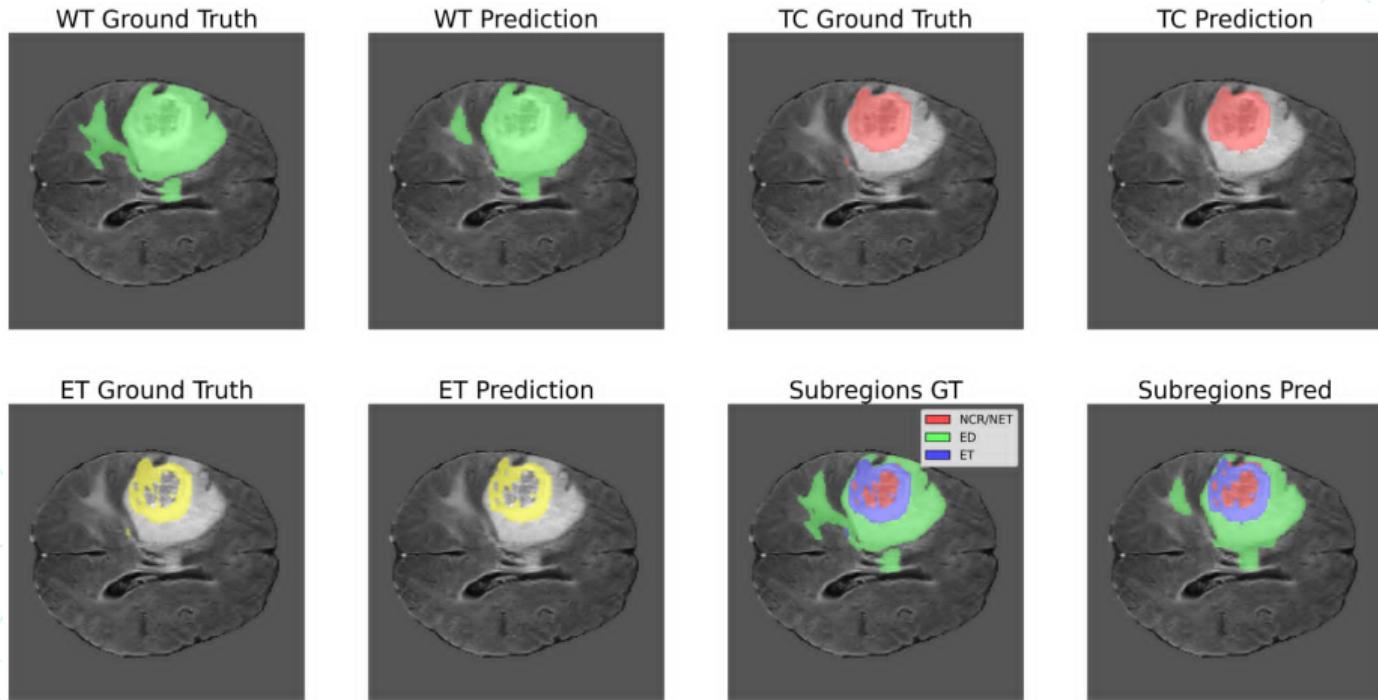


Hình 3: Dựng mesh 3D bằng marching cubes kết quả phân đoạn trên ca chụp 053 và 193

Kết quả UNETR

ROI	Dice	IoU	ASD	HD95
WT	0.8729	0.7887	3.4713	10.5030
TC	0.7819	0.6777	5.2696	12.7303
ET	0.6869	0.5788	4.2527	12.4837

Kết quả UNETR

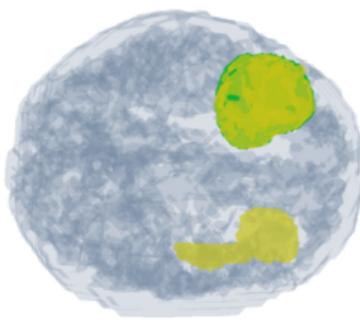
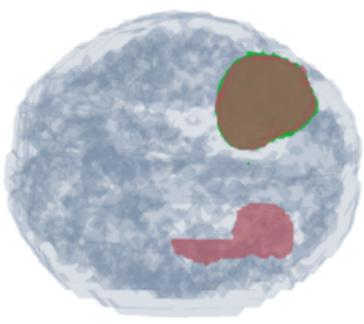
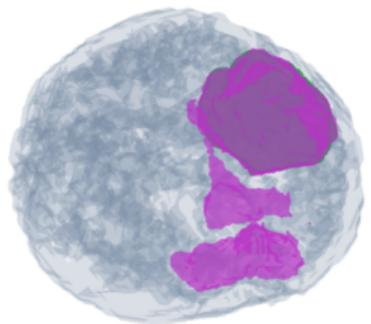
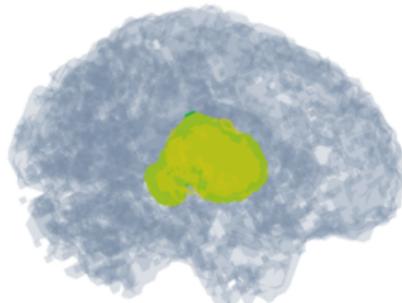
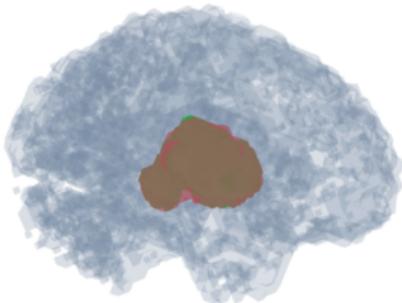
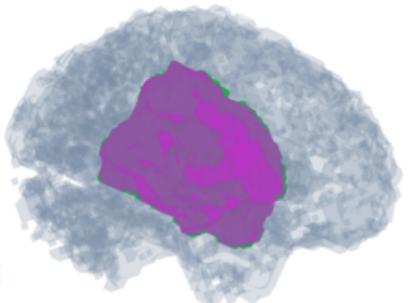


Hình 4: Kết quả phân đoạn của UNETR trên lát cắt thứ 88 của ca chụp 91

Kết quả Swin UNet3D

ROI	Dice	IoU	ASD	HD95
WT	0.9108	0.8351	1.0635	5.5856
TC	0.8446	0.7639	1.5692	5.3643
ET	0.8084	0.7091	1.1882	4.2647

Kết quả Swin UNet3D



So sánh các phương pháp trên ROI WT

Phương pháp	Dice ↑	IoU ↑	ASD ↓	HD95 ↓
UNet	0.8612	0.7725	3.6368	19.0979
UNet++	0.8954	0.8158	1.5725	5.0218
VNet (3D)	0.9024	0.8236	2.0219	6.9727
UNETR	0.8729	0.7887	3.4713	10.5030
Swin UNet3D	0.9108	0.8351	1.6035	5.5856

So sánh các phương pháp trên ROI TC

Phương pháp	Dice ↑	IoU ↑	ASD ↓	HD95 ↓
UNet	0.7549	0.6570	4.6777	12.5280
UNet++	0.8196	0.7297	2.0064	6.2421
VNet (3D)	0.8633	0.7664	1.6688	5.5709
UNETR	0.7819	0.6777	5.2696	12.7303
Swin UNet3D	<i>0.8446</i>	0.7639	1.5692	5.3643

So sánh các phương pháp trên ROI ET

Phương pháp	Dice ↑	IoU ↑	ASD ↓	HD95 ↓
UNet	0.7148	0.5992	9.1472	18.2997
UNet++	0.7360	0.6278	1.6917	5.0079
VNet (3D)	0.7617	0.6650	1.6565	5.3642
UNETR (3D)	0.6869	0.5788	4.2527	12.4837
Swin UNet3D	0.8084	0.7091	1.1882	4.2647