

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ



BÁO CÁO CUỐI KỲ

MÔN: KHOA HỌC NHẬN THỨC

Đề tài: Phát triển các phương pháp học sâu cho phân đoạn tự động khôi phục não từ ảnh cộng hưởng từ

Nhóm sinh viên thực hiện:

Nguyễn Mạnh Tường – 22027536

Mai Quốc Hiếu – 22027521

Hoàng Việt Anh – 22027519

Lớp: K67E-RE

Giảng viên hướng dẫn:

TS Nguyễn Thế Hoàng Anh

CN Nguyễn Minh Kiên

Mã học phần: RBE3046 69

Hà Nội, tháng 12 năm 2025

Mục lục

Lời cảm ơn	1
Mục lục	1
1 Mở đầu bối cảnh khoa học nhận thức	1
1.1 Đặt vấn đề	1
1.2 Vai trò của học sâu trong bài toán phân đoạn	2
1.3 Mục tiêu của báo cáo	2
2 Cơ sở lý thuyết: U não, ảnh MRI và học sâu phân đoạn	4
2.1 Đặc điểm khói u não và các thành phần cấu trúc	4
2.2 Ảnh cộng hưởng từ (MRI) và các chuỗi trong BRaTS	5
2.3 Học sâu và phân đoạn ngữ nghĩa	6
2.3.1 Phân đoạn ngữ nghĩa là gì?	6
2.3.2 Mạng nơ-ron tích chập (CNN)	7
2.3.3 Kiến trúc U-Net 2D	7
2.4 Liên hệ với khoa học nhận thức	8
3 Dữ liệu BRaTS, tiền xử lý và giải quyết mất cân bằng lớp	9
3.1 Mô tả dữ liệu và bài toán phân đoạn	9
3.2 Đặc điểm dữ liệu và thách thức	9
3.3 Tiền xử lý và chuẩn hoá dữ liệu	10
3.3.1 Lọc và chọn lát cắt – chú ý không gian thủ công	10
3.3.2 Chuẩn hoá kích thước và cường độ	11
3.3.3 Chuẩn hoá nhãn	11
3.4 Giải quyết mất cân bằng lớp: từ Accuracy đến Dice Score	12
3.4.1 Tại sao Accuracy là không đủ?	12
3.4.2 Dice Similarity Coefficient (DSC)	12
4 Mô hình U-NET 2D và chiến lược huấn luyện	13
4.1 Lý do chọn U-Net 2D thay vì 3D	13
4.2 Cấu trúc chi tiết của U-Net 2D	14
4.3 Chiến lược chia tập dữ liệu	15
4.4 Hàm mất mát và bộ tối ưu trong mô hình baseline	16
4.5 Hệ thống chỉ số đánh giá	16
5 Kết quả thực nghiệm và phân tích	17
5.1 Kết quả huấn luyện với CCE – Quan sát từ đường học	17
5.2 Kết quả định lượng trên tập Test (mô hình CCE)	17
5.3 Dánh giá định tính qua hình ảnh	18
5.4 Cải tiến hàm mất mát: Dice Loss + CCE	19
5.4.1 Động cơ	19
5.4.2 Kết quả sau cải tiến	20

6 Thảo luận khoa học nhận thức,hạn chế, hướng phát triển và kết luận	21
6.1 Học sâu như một mô hình nhận thức thị giác	21
6.2 Mất cân bằng lớp như một dạng bias nhận thức	21
6.3 Overfitting và ranh giới giữa ghi nhớ – khái quát	22
6.4 Vai trò trong hệ thống nhận thức mở rộng	22
6.5 Hạn chế	23
6.6 Hướng phát triển	24
6.7 Kết luận chung	25

Danh sách hình

1.1.1 <i>Hình ảnh minh họa</i>	2
2.2.1 <i>Bộ dữ liệu tiêu chuẩn cho phân đoạn u não</i>	6
2.3.1 <i>Minh họa nhãn phân đoạn 4 lớp: Background, ED, ET, NCR</i>	7
3.2.1 (a) <i>Mất cân bằng lớp toàn cục với Background áp đảo; (b) Trong các lớp u cũng tồn tại nhóm voxel khó ($p<0.5$), thường tập trung ở vùng rìa/khó phân biệt.</i>	10
4.2.1 <i>Kiến trúc U-Net 2D</i>	14
5.1.1 <i>Training curves: Loss/Accuracy/Dice/Mean IoU (Train vs Val)</i>	17
5.3.1 <i>Dánh giá định tính</i>	19
5.4.1 <i>Ảnh CCE được chip trên một lát cắt</i>	19
5.4.2 <i>Ảnh so sánh CCE vs (Dice+CCE) trên cùng 1 lát</i>	20

Chương 1

Mở đầu bối cảnh khoa học nhận thức

1.1 Đặt vấn đề

Trong khoa học nhận thức, một câu hỏi trung tâm luôn được đặt ra là: hệ thống trí tuệ – dù là con người hay máy móc – tiếp nhận, biểu diễn và xử lý thông tin phức tạp như thế nào để hỗ trợ ra quyết định?

Câu hỏi này không chỉ mang tính lý thuyết, mà gắn chặt với thực hành đời sống, đặc biệt trong lĩnh vực y khoa – nơi mỗi quyết định đều có thể ảnh hưởng trực tiếp đến tính mạng bệnh nhân. Ở đó, thông tin cảm giác (ảnh y khoa, tín hiệu sinh lý, âm thanh tim phổi...) phải được xử lý nhanh chóng, chính xác và ổn định.

Các hệ thống chẩn đoán y khoa là ví dụ điển hình: từng bước trong quy trình – từ quan sát ảnh, phát hiện bất thường, so sánh với kinh nghiệm, cho đến suy luận chẩn đoán – đều là những hoạt động nhận thức có cấu trúc.

Khối u não là một trong những dạng bệnh lý thần kinh nguy hiểm nhất, gắn liền trực tiếp với cấu trúc và chức năng của hệ thần kinh trung ương. Việc phát hiện sớm, đánh giá đúng vị trí, kích thước, mức độ lan rộng của khối u là yếu tố then chốt quyết định hiệu quả điều trị.

Trong thực hành lâm sàng, cộng hưởng từ (Magnetic Resonance Imaging – MRI) là công cụ chủ lực giúp bác sĩ “nhìn thấy” các cấu trúc bên trong não mà không cần can thiệp xâm lấn. Tuy nhiên, để đi từ ảnh MRI đến quyết định lâm sàng là một quá trình phức tạp, đòi hỏi sự phối hợp giữa tri giác, trí nhớ, suy luận và kinh nghiệm.

Tuy nhiên, việc đọc ảnh và phân vùng khối u trên từng lát cắt MRI vẫn chủ yếu dựa vào thao tác thủ công: bác sĩ phải quan sát hàng trăm lát cắt, sau đó dùng phần mềm để vẽ ranh giới khối u. Công việc này:

- **Tốn rất nhiều thời gian và công sức:**

Một ca chụp MRI có thể gồm hàng trăm lát cắt. Việc khoanh từng vùng u trên ảnh đòi hỏi tập trung cao độ và kéo dài trong thời gian dài, dẫn đến mệt mỏi và giảm hiệu suất.

- **Dễ bị ảnh hưởng bởi mệt mỏi, chủ quan của người đọc phim:**

Khi mệt mỏi, bác sĩ có thể bỏ sót các vùng u nhỏ hoặc ranh giới mờ. Hai bác sĩ khác nhau đôi khi có thể vẽ ranh giới khác nhau cho cùng một khối u.

- **Khó đảm bảo tính nhất quán giữa các bác sĩ (inter-rater variability):**

Những khác biệt trong kinh nghiệm, thói quen, tiêu chuẩn đánh giá khiến kết quả phân đoạn không hoàn toàn đồng nhất. Điều này ảnh hưởng trực tiếp đến theo dõi điều trị và so sánh kết quả giữa các cơ sở y tế.

Từ góc nhìn khoa học nhận thức, bài toán này đặt hệ thống nhận thức (con người hoặc máy) vào tình huống tải nhận thức cao: lượng thông tin lớn, yêu cầu chính xác cao, thời gian hạn chế, và sai sót có thể gây hậu quả nghiêm trọng. Do đó, việc hỗ trợ bác sĩ thông qua các công cụ tính toán thông minh trở thành nhu cầu tất yếu.



Hình 1.1.1: *Hình ảnh minh họa*

1.2 Vai trò của học sâu trong bài toán phân đoạn

Sự phát triển mạnh mẽ của học sâu (deep learning), đặc biệt là mạng nơ-ron tích chập (Convolutional Neural Network – CNN) và các kiến trúc phân đoạn như U-Net, đã mở ra khả năng xây dựng các hệ thống “thị giác nhân tạo” có khả năng:

- **Tự động học đặc trưng hình ảnh từ dữ liệu thô:**

Thay vì con người phải thiết kế thủ công các đặc trưng (biên, texture, mức xám...), mô hình học sâu tự trích xuất những đặc trưng phù hợp trực tiếp từ dữ liệu. Điều này tương tự cách não người dần học cách nhận diện khuôn mặt hoặc vật thể qua kinh nghiệm.

- **Thực hiện phân vùng khối u với độ chính xác cao:**

Mô hình không chỉ nói “có u hay không”, mà còn vẽ được ranh giới chính xác ở mức pixel, giúp bác sĩ định lượng thể tích u, đánh giá tiến triển điều trị và lên kế hoạch phẫu thuật/xạ trị.

- **Đóng vai trò như một “trợ lý nhận thức” cho bác sĩ:**

Hệ thống có thể xử lý nhanh hàng trăm lát cắt, đánh dấu những vùng nghi ngờ, giúp bác sĩ tập trung vào việc kiểm tra và ra quyết định, thay vì làm những thao tác lặp lại nặng về thị giác.

Trong bối cảnh môn Khoa học nhận thức, mô hình học sâu không chỉ đơn thuần là công cụ kỹ thuật, mà còn là một mô hình tính toán mô phỏng cách hệ thống trí tuệ xử lý thông tin: có các tầng trừu tượng, cơ chế học từ lỗi, khả năng khái quát hóa, và chịu ảnh hưởng mạnh bởi dữ liệu đầu vào – tương tự các quá trình diễn ra trong hệ thần kinh sinh học.

1.3 Mục tiêu của báo cáo

Báo cáo này tập trung vào việc:

- Phát triển và đánh giá một mô hình học sâu U-Net 2D cho bài toán phân đoạn tự động khói u não trên bộ dữ liệu BRaTS.
Mục tiêu không chỉ là đạt chỉ số cao, mà còn hiểu mô hình học gì, sai ở đâu và vì sao.
- **Phân tích chi tiết toàn bộ quy trình:**
bao gồm chuẩn bị dữ liệu, tiền xử lý, thiết kế mô hình, lựa chọn hàm mất mát, chiến lược huấn luyện và các chỉ số đánh giá — cho thấy mỗi bước đều ảnh hưởng đến hành vi nhận thức của mô hình.
- **Liên hệ với các khái niệm cốt lõi trong khoa học nhận thức:**
như học, ghi nhớ, khái quát hóa, bias, cách định nghĩa phần thưởng và hệ nhận thức mở rộng, qua đó xem AI như một phần trong hệ thống nhận thức lai người-máy.

Chương 2

Cơ sở lý thuyết: U não, ảnh MRI và học sâu phân đoạn

2.1 Đặc điểm khối u não và các thành phần cấu trúc

Trong thực hành thần kinh – ung bướu, khối u não thường không phải chỉ là một “khối đặc” đồng nhất. Thay vào đó, mỗi khối u bao gồm nhiều vùng với đặc tính sinh học khác nhau, phản ánh các giai đoạn tiến triển của bệnh. Việc hiểu rõ từng thành phần giúp bác sĩ không chỉ phát hiện u, mà còn suy luận về mức độ ác tính, tốc độ phát triển và khả năng đáp ứng điều trị.

Necrotic Core / Non-Enhancing Tumor (NCR/NET)

Đây là vùng lõi hoại tử hoặc mô u không bắt thuốc tương phản. Tế bào tại đây thường đã chết hoặc hoạt động trao đổi chất rất thấp. Vùng này thường nằm sâu bên trong khối u.

Tại sao vùng này quan trọng?

- Nó cho thấy khối u đã phát triển đến mức nguồn dinh dưỡng không đủ cung cấp, dẫn tới chết tế bào.
- Lõi hoại tử thường liên quan đến u ác tính cao, do tốc độ tăng sinh vượt quá khả năng cung cấp máu.
- Khi phân đoạn chính xác, bác sĩ có thể so sánh trước – sau điều trị để xem liệu lõi hoại tử có tăng lên (dấu hiệu đáp ứng) hay không.

Ví dụ: trong nhiều trường hợp xạ trị, lõi hoại tử tăng lên chứng tỏ mô u đang bị phá huỷ.

Edema (ED)

Vùng phù nề bao quanh khối u, chứa nhiều dịch. Đây không phải là u, mà là phản ứng của mô não lành trước áp lực và sự xâm lấn.

Tại sao cần phân biệt rõ phù nề?

- Phù làm tăng áp lực nội sọ, gây đau đầu, nôn, giảm ý thức — nguy hiểm về mặt triệu chứng.
- Phù có thể làm bác sĩ “đánh giá nhầm” kích thước u nếu không phân đoạn đúng.
- Trong phẫu thuật, bác sĩ chỉ muốn cắt bỏ phần u, không phải toàn bộ vùng phù.

Ví dụ: Người bệnh có thể có vùng phù rất rộng, nhưng phần u thật sự nhỏ hơn nhiều.

Enhancing Tumor (ET)

Đây là vùng bắt thuốc tương phản mạnh sau tiêm Gadolinium — biểu hiện mô u còn hoạt động, tăng sinh mạch.

Vai trò lâm sàng:

- Giúp xác định vùng u “hoạt động tích cực nhất” để sinh thiết.

- Là chỉ dấu quan trọng để đánh giá tiến triển bệnh.
- Dùng để lập kế hoạch xạ trị chính xác (nhắm vào vùng hoạt động cao).

Ví dụ: Nếu ET tăng kích thước theo thời gian, bác sĩ nghi ngờ u đang tiến triển, dù lõi hoại tử có thể vẫn giữ nguyên.

Việc phân biệt rõ các cấu trúc giúp:

- Dánh giá thể tích từng vùng để lên kế hoạch xạ trị/phẫu thuật chi tiết và an toàn hơn.
- Theo dõi đáp ứng: ví dụ, lõi hoại tử tăng nhưng ET giảm là dấu hiệu điều trị đang hiệu quả.

2.2 Ảnh cộng hưởng từ (MRI) và các chuỗi trong BRaTS

MRI không chỉ tạo ra “một ảnh duy nhất”. Mỗi lần chụp tạo ra nhiều chuỗi ảnh khác nhau, mỗi chuỗi nhấn mạnh một đặc tính mô khác nhau. Có thể xem mỗi chuỗi như một “giác quan riêng”, bổ sung thông tin cho nhau.

Bộ dữ liệu BRaTS sử dụng 4 chuỗi chính:

- **T1 (Native):**

Nhấn mạnh cấu trúc mô. Giúp phân biệt rõ chất trắng – chất xám – dịch não tủy. Tốt để quan sát cấu trúc giải phẫu tổng thể, nhưng không phải lúc nào cũng làm nổi khói u.

- **T1ce (Post-contrast T1-weighted):**

Là T1 sau tiêm thuốc tương phản Gadolinium. Các vùng u tăng sinh mạch (ET) sẽ sáng rõ. Đây là chuỗi nổi bật nhất khi cần phát hiện u đang hoạt động.

- **T2-weighted (T2):**

Làm nổi bật vùng chứa dịch, ví dụ phèn nề. Vùng có nhiều nước sáng lên. Chuỗi này rất hữu ích khi cần quan sát mức độ lan rộng của phèn.

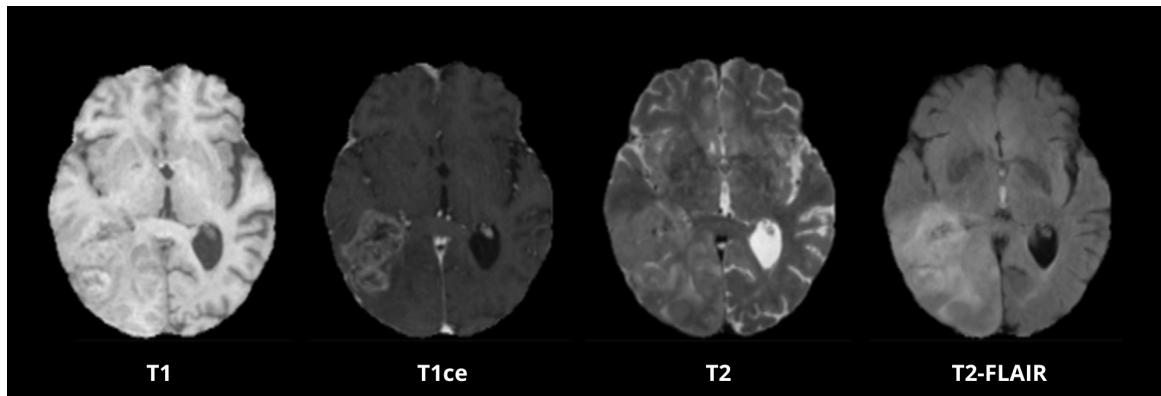
- **T2-FLAIR:**

Khử tín hiệu dịch tự do (như dịch não tủy), vì thế các tổn thương gần não thắt và đường giữa rõ hơn. Đặc biệt hiệu quả trong việc phát hiện tổn thương chất trắng và phèn nề lan tỏa.

Từ góc nhìn khoa học nhận thức, hệ thống (bác sĩ hoặc mô hình) phải:

- nhận nhiều nguồn cảm giác khác nhau,
- hợp nhất chúng,
- chuyển thành một biểu diễn chung để hiểu đúng tình trạng bệnh.

Điều này giống như việc con người kết hợp thị giác – thính giác – xúc giác để hiểu môi trường xung quanh.



Hình 2.2.1: Bộ dữ liệu tiêu chuẩn cho phân đoạn u não

2.3 Học sâu và phân đoạn ngữ nghĩa

2.3.1 Phân đoạn ngữ nghĩa là gì?

Trong phân đoạn ngữ nghĩa, nhiệm vụ không chỉ là phát hiện “có u hay không”, mà là gán nhãn cho **từng pixel/voxel**:

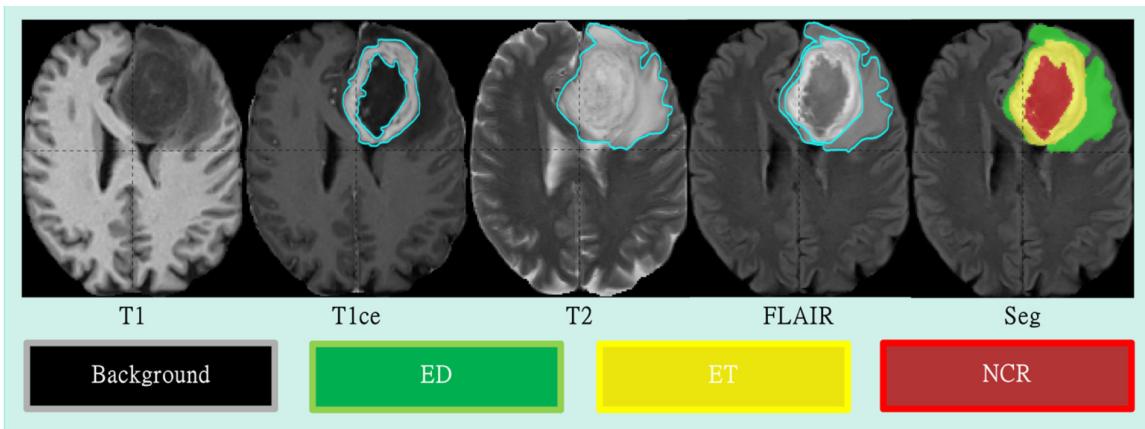
- 0 – Background
- 1 – Edema
- 2 – Enhancing Tumor
- 3 – Necrotic / Non-Enhancing Tumor

Tại sao việc này khó?

- Ranh giới u có thể mờ, lấn vào mô lành.
- Từng lớp u lại có đặc tính hình ảnh khác nhau.
- Các khối u nhỏ dễ bị bỏ sót.

So với phân loại ảnh hay phát hiện bounding box, phân đoạn yêu cầu:

- Hiểu chính xác hình dạng, độ rộng và phạm vi của từng tổn thương.
- Khôi phục ranh giới chi tiết ở mức pixel, vì chỉ cần lệch vài milimet cũng ảnh hưởng đến điều trị.



Hình 2.3.1: *Minh họa nhãn phân đoạn 4 lớp: Background, ED, ET, NCR*

2.3.2 Mạng nơ-ron tích chập (CNN)

CNN là nền tảng của các hệ thống thị giác máy tính hiện đại. Mỗi tầng CNN có vai trò riêng:

- Các bộ lọc nhỏ trượt trên ảnh để phát hiện cạnh, góc, đường cong.
- Pooling giảm kích thước nhưng giữ đặc trưng quan trọng, giống như bộ não bỏ qua chi tiết nhỏ nhưng giữ cấu trúc chính.
- Tầng sâu hơn học các đặc trưng trừu tượng như hình khối, mô hình lặp lại và ngữ cảnh.

Nhờ đó, CNN có khả năng dần dần xây dựng biểu diễn từ đơn giản đến phức tạp — giống cách hệ thị giác sinh học hoạt động.

2.3.3 Kiến trúc U-Net 2D

U-Net được thiết kế riêng cho ảnh y sinh, nơi cần phân đoạn chính xác.

Encoder (contracting path)

Encoder nén thông tin lại bằng chuỗi Conv + ReLU + MaxPool:

- Mỗi bước giảm kích thước không gian nhưng tăng số kênh.
- Mục tiêu là học các đặc trưng trừu tượng dần, từ biên tới hình dạng tổng thể.

Decoder (expanding path)

Decoder làm điều ngược lại: khôi phục kích thước ảnh ban đầu.

Quan trọng nhất là:

- Up-convolution phỏng to feature map.
- Concatenate với feature map đã lưu từ Encoder thông qua skip-connection.

Skip-connection giúp giữ lại:

- thông tin chi tiết (biên, đường viền),
- kết hợp với ngữ cảnh trừu tượng từ tầng sâu.

Lớp đầu ra

Conv 1×1 ánh xạ về số lớp (4), sau đó softmax sinh ra xác suất cho từng lớp tại mỗi pixel.

2.4 Liên hệ với khoa học nhận thức

Về bản chất, U-Net mô phỏng:

- xử lý đa tầng giống hệ thị giác sinh học,
- kết hợp giữa bottom-up (dữ liệu lên) và top-down (bối cảnh điều chỉnh),
- học từ sai lệch dự đoán — tương tự brain predictive coding.

Nhờ đó, U-Net không chỉ là một mô hình kỹ thuật, mà còn là công cụ để nghiên cứu cách hệ thống nhận thức xử lý thông tin thị giác phức tạp.

Chương 3

Dữ liệu BRaTS, tiền xử lý và giải quyết mất cân bằng lớp

3.1 Mô tả dữ liệu và bài toán phân đoạn

Nhóm sử dụng bộ dữ liệu BRaTS tiêu chuẩn, một trong những bộ dữ liệu phổ biến nhất cho nghiên cứu phân đoạn u não. Điểm mạnh của bộ dữ liệu này là đã được chuẩn hóa tương đối tốt, có nhãn chuyên gia, và bao gồm nhiều dạng u khác nhau.

Trong bộ dữ liệu:

- **Mỗi bệnh nhân có bộ ảnh MRI gồm 4 chuỗi T1, T1ce, T2, FLAIR.**

Nhờ đó, mô hình không chỉ thấy u từ một góc nhìn duy nhất, mà có khả năng kết hợp thông tin từ nhiều đặc tính mô khác nhau.

- **Mask phân đoạn đi kèm được gán nhãn bởi chuyên gia lâm sàng.**

Đây là “chuẩn vàng” (ground truth) mà mô hình phải học theo, giúp đảm bảo kết quả đánh giá có ý nghĩa y khoa.

- **Mục tiêu chính:**

mô hình cần phân biệt chính xác ba vùng u (NCR/NET, ED, ET) và vùng nền Background trên từng lát cắt 2D — đồng nghĩa với việc phải hiểu cấu trúc u chứ không chỉ phát hiện có u hay không.

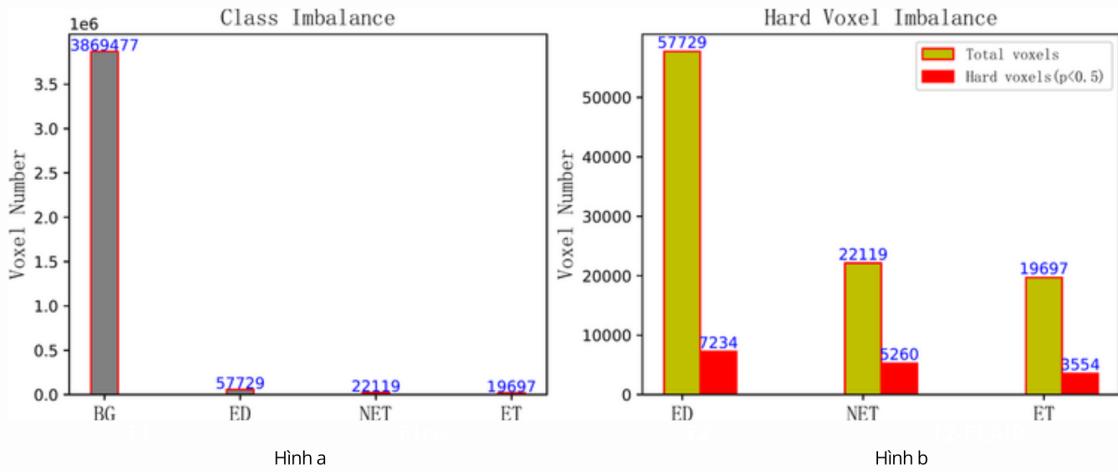
Từ góc nhìn khoa học nhận thức, ta đang yêu cầu mô hình xây dựng một **biểu diễn nội tại** đủ mạnh. Khi đưa vào một “chuỗi quan sát cảm giác” (4 ảnh cùng lát), hệ thống phải tự học được:

- mô nào là bình thường,
- mô nào là phì,
- mô nào là lồi hoại tử,
- mô nào là khối u đang hoạt động.

Điều này tương đương với cách nào người học nhận diện các kiểu mô qua kinh nghiệm và phản hồi.

3.2 Đặc điểm dữ liệu và thách thức

Khi phân tích thống kê, ta nhận thấy một vấn đề rất lớn:



Hình 3.2.1: (a) Mất cân bằng lớp toàn cục với Background áp đảo; (b) Trong các lớp u cũng tồn tại nhóm voxel khó ($p < 0.5$), thường tập trung ở vùng rìa/khở phân biệt.

- **Vùng Background chiếm hơn 98% tổng voxel.**

Nghĩa là hầu hết ảnh chỉ toàn mô não lành hoặc nền đen. Nếu chỉ tối ưu theo Accuracy, mô hình chỉ cần “đoán hết là nền” là cũng đã gần như đúng.

- **Các vùng u chỉ chiếm khoảng 1–2%.**

Thậm chí ở nhiều lát, khối u nhỏ đến mức chỉ vài trăm pixel. Điều này khiến mô hình rất dễ bỏ sót nếu không có chiến lược phù hợp.

Không chỉ mất cân bằng về số lượng, còn có thêm sự không đồng đều về không gian:

- **Các lát gần đỉnh đầu hoặc cổ gần như không có u.**
- **U tập trung chủ yếu ở vùng giữa não.**

Do đó, nếu mô hình không được hướng dẫn tốt, nó sẽ:

- học thói quen “coi mọi thứ là bình thường”,
- bỏ qua các chi tiết nhỏ nhưng cực kỳ quan trọng.

Ở đây tồn tại một sự tương đồng thú vị với nhận thức con người: nếu người học hiếm khi gặp tình huống hiếm (như u rất nhỏ), họ thường có xu hướng đánh giá thấp rủi ro của nó. Điều này cũng chính là một dạng **thiên lệch nhận thức**.

3.3 Tiền xử lý và chuẩn hóa dữ liệu

3.3.1 Lọc và chọn lát cắt – chú ý không gian thủ công

Trong thực tế, không phải lát MRI nào cũng có thông tin giá trị. Vì vậy, nhóm chọn khoảng 100 lát cắt quan trọng cho mỗi bệnh nhân (từ lát 22–122):

- **Các lát đầu và cuối chủ yếu là nền đen, ít cấu trúc não.**

Nếu đưa vào đào tạo, mô hình chỉ học thêm thói quen “dánh nền”.

- **Loại bỏ những lát này giúp:**

giảm số lượng mẫu vô ích, giảm thời gian huấn luyện và tập trung hơn vào vùng có khả năng chứa u.

Có thể xem đây như một dạng **spatial attention do con người áp đặt**: ta hướng ánh nhìn của mô hình về những nơi đáng quan tâm, giống cách bác sĩ bỏ qua vùng cổ để tập trung vào bán cầu não.

3.3.2 Chuẩn hoá kích thước và cường độ

Chuẩn hoá kích thước:

- Tất cả lát cắt được resize về 128×128 .
- Vẫn giữ được hình dạng tổng thể của u.
- Cho phép tăng batch size và huấn luyện nhanh hơn.

Chuẩn hoá cường độ:

Giá trị pixel của từng ảnh được chia cho giá trị lớn nhất:

$$X' = \frac{X}{\max(X)}$$

Ý nghĩa:

- đưa toàn bộ ảnh về khoảng $[0, 1]$,
- tránh trường hợp một chuỗi có cường độ quá cao làm “lấn át” các chuỗi khác,
- ổn định gradient, giúp mô hình học ổn định hơn.

3.3.3 Chuẩn hoá nhãn

Trong BRaTS, nhãn có thể theo nhiều quy ước khác nhau. Do đó, nhóm chuẩn hoá về cùng một format:

- 0 – Background
- 1 – Edema
- 2 – Enhancing Tumor
- 3 – Necrotic / Non-Enhancing Tumor

Việc này đảm bảo:

- mỗi kênh đầu ra của mô hình tương ứng đúng với một loại tổn thương,
- khi đánh giá, ta hiểu được “mô hình sai ở đâu và sai với lớp nào”.

Nếu bước này làm sai, bác sĩ sẽ gần như không thể diễn giải được mô hình.

3.4 Giải quyết mất cân bằng lớp: từ Accuracy đến Dice Score

3.4.1 Tại sao Accuracy là không đủ?

Nếu mô hình dự đoán toàn bộ ảnh là nền, Accuracy vẫn có thể đạt 98–99%. Nhưng rõ ràng:

- mô hình gần như không phát hiện u,
- bệnh nhân thật sự không được hưởng lợi gì.

Từ góc nhìn nhận thức, điều này giống việc đặt ra một tiêu chí đánh giá sai. Khi phần thưởng chỉ tập trung vào “trường hợp bình thường”, mô hình sẽ không có động lực học những tình huống hiếm nhưng quan trọng.

3.4.2 Dice Similarity Coefficient (DSC)

Dice đo mức độ chồng lấp giữa mask dự đoán và mask chuẩn:

$$D(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

Ý nghĩa:

- Dice cao chỉ khi mô hình vừa ít bỏ sót u,
- vừa không dự đoán nhầm quá nhiều.

Do vậy, Dice phản ánh đúng hơn chất lượng phân đoạn trong bối cảnh y khoa, nơi **mỗi pixel sai đều có thể ảnh hưởng đến điều trị**.

Chương 4

Mô hình U-NET 2D và chiến lược huấn luyện

4.1 Lý do chọn U-Net 2D thay vì 3D

Nhóm lựa chọn U-Net 2D trên từng lát cắt thay vì U-Net 3D trên toàn khối thể tích không chỉ vì lý do kỹ thuật, mà còn vì cân nhắc giữa hiệu quả, tính khả thi và phạm vi của môn học.

Giới hạn phần cứng

U-Net 3D sử dụng kernel 3 chiều và phải xử lý đồng thời nhiều lát cắt liên tiếp. Điều này khiến:

- nhu cầu bộ nhớ GPU tăng đột biến;
- batch size bị giảm rất nhỏ, khó ổn định;
- thời gian huấn luyện kéo dài lên nhiều lần.

Trong điều kiện máy tính sinh viên, điều này gần như không khả thi nếu không có GPU mạnh.

Ngược lại:

- U-Net 2D xử lý từng lát ảnh riêng lẻ;
- bộ nhớ tiêu tốn thấp hơn rất nhiều;
- cho phép thử nghiệm nhiều cấu hình khác nhau.

Ví dụ: thay vì phải chờ 12–24 giờ cho một mô hình 3D, mô hình 2D có thể hoàn thành nhiều lượt chạy chỉ trong vài giờ, giúp chúng em có nhiều cơ hội so sánh, chỉnh sửa và tối ưu.

Thời gian huấn luyện

Mô hình 3D có số tham số lớn hơn rất nhiều so với 2D. Điều này dẫn đến:

- huấn luyện chậm,
- dễ overfitting,
- khó tinh chỉnh hyper-parameters.

Trong khuôn khổ một môn học, mô hình 2D giúp:

- thử nhiều hàm mất mát khác nhau,
- đánh giá ảnh hưởng của preprocessing,
- hiểu rõ hành vi mô hình thay vì chỉ chạy một lần duy nhất.

Tính chuẩn hoá

U-Net 2D là baseline kinh điển trong phân đoạn y sinh. Việc dùng baseline này:

- giúp dễ so sánh với nghiên cứu trước;
- tạo nền tảng kiến thức vững chắc;
- làm cơ sở để sau này nâng cấp sang 3D hoặc hybrid.

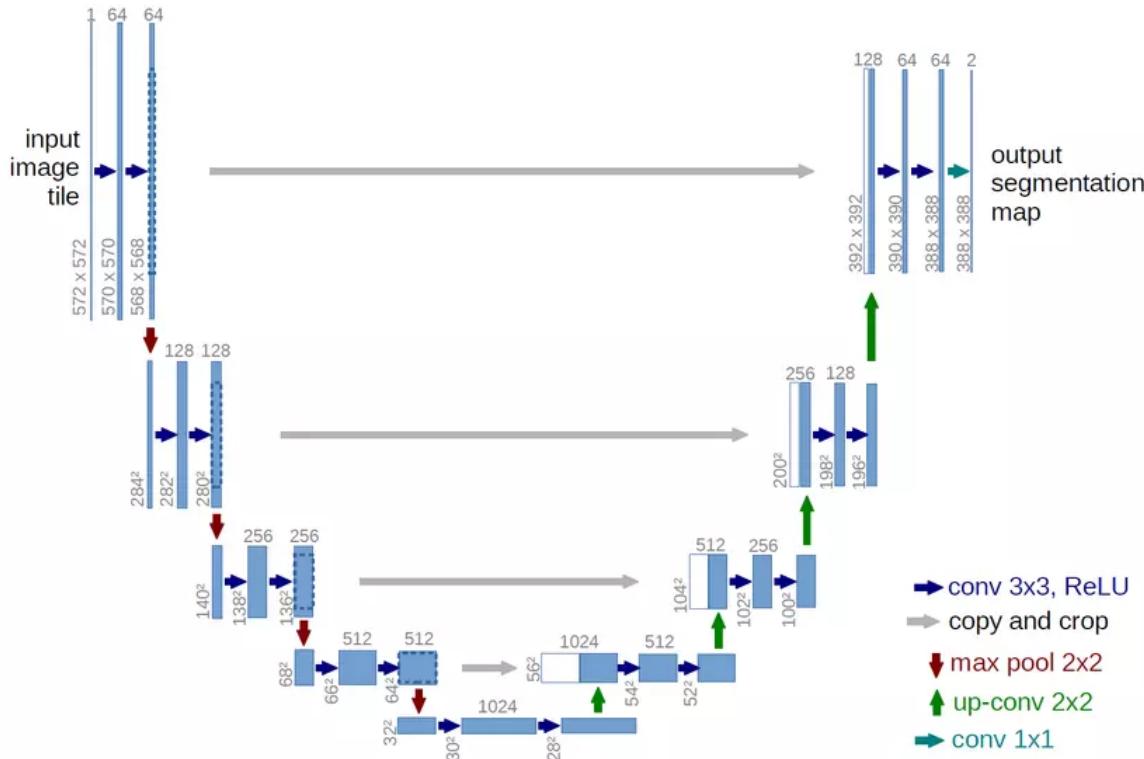
Tuy nhiên, hạn chế là:

mô hình không tận dụng được mối liên hệ giữa các lát theo trục z.

Trong khi đó, bác sĩ luôn xem liên tiếp nhiều lát để hình dung 3D. Đây là điểm yếu quan trọng và được chúng em phân tích kỹ trong phần hạn chế và hướng phát triển.

4.2 Cấu trúc chi tiết của U-Net 2D

U-Net 2D hoạt động theo triết lý: nén thông tin lại để trích xuất đặc trưng, rồi mở lại để tái tạo không gian.



Hình 4.2.1: Kiến trúc U-Net 2D

Encoder

Encoder đóng vai trò giống “mắt” và “vùng thị giác sơ cấp”. Nó trả lời câu hỏi:

Trong ảnh này có những đặc trưng gì quan trọng?

- **Block 1:** Conv $3 \times 3 \rightarrow$ ReLU \rightarrow Conv $3 \times 3 \rightarrow$ ReLU \rightarrow MaxPool 2×2 (64 kênh). Học cạnh, góc, biên, đường viền u.

- **Block 2:** tăng lên 128 kênh. Bắt đầu nhận ra cấu trúc lớn hơn như hình dạng cục bộ.
- **Block 3:** 256 kênh. Nhận diện các vùng tổn thương và bối cảnh xung quanh.
- **Block 4:** 512 kênh. Học biểu diễn trừu tượng hơn, giảm phụ thuộc chi tiết nhỏ.
- **Bottleneck (1024 kênh).**

Dây là nơi lưu trữ thông tin “khái quát” toàn ảnh.

Càng đi sâu, mô hình nhìn ít chi tiết hơn nhưng hiểu nhiều hơn về cấu trúc tổng thể.

Decoder

Decoder có nhiệm vụ:

khôi phục hình dạng ban đầu nhưng vẫn giữ hiểu biết đã học được.

- Up-convolution 2×2 từ 1024 → 512, nối với Block 4.
- Sau đó Conv 3×3 + ReLU để tinh chỉnh.
- Tiếp tục up-conv và concat với Block 3, 2, 1.

Điểm quan trọng:

skip connections

Nếu không có skip connections, khi nén xuống, mô hình sẽ mất ranh giới chi tiết. Nhờ nối tắt:

- feature map chi tiết từ encoder được giữ lại,
- kết hợp với đặc trưng trừu tượng từ tầng sâu,
- giúp vẽ ranh giới u mượt và sắc nét hơn.

Liên hệ nhận thức:

Não người cũng kết hợp tín hiệu từ giác quan (chi tiết) với tri thức đã có (bối cảnh).

4.3 Chiến lược chia tập dữ liệu

Chia dữ liệu không chỉ để “tránh gian lận”, mà còn là cách kiểm tra mô hình:

mô hình đang học thật hay chỉ học thuộc?

- **Train (80%):** dùng để cập nhật trọng số.
- **Validation (20% của train):** theo dõi overfitting, dùng để điều chỉnh hyper-parameters.
- **Test ($\approx 15\%$ bệnh nhân):** chỉ đánh giá *một lần duy nhất*.

Nếu mô hình tốt trên Train nhưng kém trên Test:

mô hình chỉ ghi nhớ dữ liệu, không khái quát.

Giống như học sinh học thuộc đáp án nhưng không hiểu bài.

4.4 Hàm mất mát và bộ tối ưu trong mô hình baseline

Categorical Cross-Entropy (CCE)

CCE đo sự khác biệt giữa:

- phân bố dự đoán,
- nhãn thật (ở dạng one-hot).

Nó giúp mô hình:

- tự tin hơn khi dự đoán đúng,
- bị phạt mạnh nếu sai hoàn toàn lớp.

Tuy nhiên, như đã phân tích ở chương trước, nếu không cẩn thận:

CCE sẽ ưu tiên lớp nền vì nó nhiều hơn.

Optimizer Adam

Adam được chọn vì:

- hội tụ nhanh,
- thích ứng bước nhảy theo từng tham số,
- ổn định trên nhiều loại mô hình CNN.

Learning rate ban đầu 0.001 là lựa chọn cân bằng giữa:

- đi nhanh nhưng không “quá đà”.

ReduceLROnPlateau

Nếu Validation Loss không giảm:

- learning rate được giảm tự động,
- giúp quá trình học “tinh chỉnh” mượt hơn.

Giống như khi học bài: ban đầu học nhanh, về sau cần chậm lại để nắm kỹ chi tiết.

4.5 Hệ thống chỉ số đánh giá

Không thể chỉ nhìn một chỉ số duy nhất.

- **Mean IoU:** mức độ chồng lấp theo từng lớp.
- **Precision:** mô hình có “đánh nhầm” nhiều không?
- **Sensitivity (Recall):** có bỏ sót u không? (cực kỳ quan trọng).
- **Specificity:** có phân loại nhầm mô lành không?

Nếu Recall thấp → nguy hiểm vì bỏ sót u. Nếu Precision thấp → bác sĩ mất thời gian sửa lỗi.
Do đó, bài toán giống như cân bằng:

thận trọng vs. liều lĩnh — y hệt quá trình ra quyết định của con người.

Chương 5

Kết quả thực nghiệm và phân tích

5.1 Kết quả huấn luyện với CCE – Quan sát từ đường học

Trong giai đoạn đầu, việc quan sát các đồ thị Accuracy, Loss, Dice và Mean IoU theo epoch đóng vai trò như một “cửa sổ” cho phép chúng ta theo dõi cách mô hình đang học. Đây không chỉ là những đường cong đơn thuần, mà phản ánh toàn bộ động lực của quá trình tối ưu.

Các đồ thị cho thấy:

- **Loss trên Train và Validation giảm khá đều, sau đó tiệm cận một giá trị ổn định.**

Điều này cho thấy mô hình đang học từ dữ liệu một cách hợp lý. Nếu Loss tăng hoặc dao động mạnh, đó thường là tín hiệu cảnh báo mô hình đang học không ổn định.

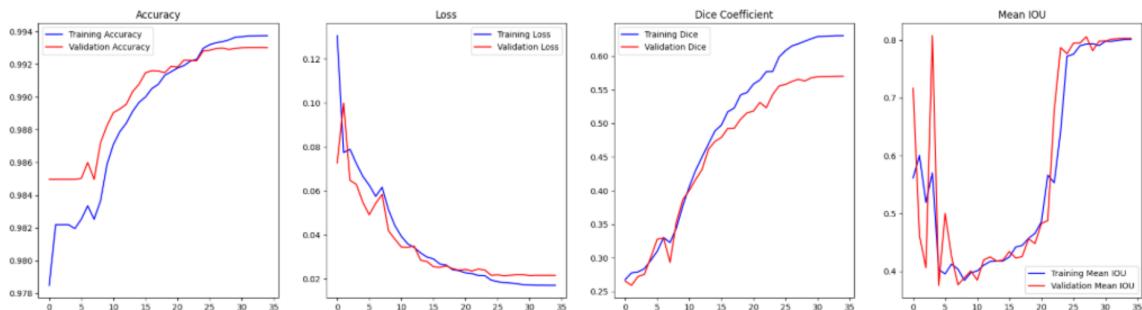
- **Accuracy và Dice tăng dần, tiến tới trạng thái gần bão hòa sau khoảng 30–35 epoch.**

Khi các chỉ số tăng đều, nghĩa là mô hình dần cải thiện khả năng nhận diện và phân vùng. Trạng thái “bão hòa” cho thấy mô hình đã gần đạt giới hạn với cấu trúc và dữ liệu hiện tại.

- **Đường Train và Validation không tách xa nhau, chỉ biểu hiện overfitting nhẹ.**

Nếu hai đường chênh lệch lớn, mô hình có thể chỉ ghi nhớ dữ liệu Train. Ở đây, sự khác biệt nhỏ cho thấy khả năng khai quát vẫn được duy trì.

Tổng hợp lại, mô hình không rơi vào tình trạng “học vẹt” mà đã nắm được cấu trúc chung của dữ liệu. Điều này rất quan trọng, vì nó chứng minh rằng quá trình tiền xử lý, lựa chọn kiến trúc và chiến lược huấn luyện đang đi đúng hướng.



Hình 5.1.1: Training curves: Loss/Accuracy/Dice/Mean IoU (Train vs Val)

5.2 Kết quả định lượng trên tập Test (mô hình CCE)

Khi đánh giá trên tập Test độc lập, mô hình đạt:

- $\text{Loss} \approx 0.019$: mô hình ít sai số tổng thể.
- $\text{Accuracy} \approx 99.37\%$: phần lớn pixel được phân loại đúng.
- $\text{Mean IoU} \approx 0.80$: mức độ chồng lấp tương đối tốt giữa mask dự đoán và mask chuẩn.

- Dice trung bình $\approx 0.59\text{--}0.60$: chất lượng phân đoạn ở mức trung bình-khá.
- Sensitivity ≈ 0.993 : hầu như không bỏ sót u.
- Specificity ≈ 0.998 : rất hiếm khi nhầm mô lành thành u.

Ở đây có một điểm quan trọng:

Accuracy rất cao không có nghĩa mô hình phân đoạn tốt.

Chính Dice và IoU mới phản ánh thực chất bài toán. Phân tích theo từng lớp giúp nhìn sâu hơn:

Edema (ED)

Dice đạt khoảng 0.70–0.73 — cao nhất trong các lớp.

Lý do:

- vùng phù thường lớn, lan tỏa nhưng vẫn có ranh giới khá rõ;
- tín hiệu sáng rõ trên T2 và FLAIR;
- dễ nhận diện vì tạo ra sự tương phản rõ rệt.

Ví dụ: nhiều ca bệnh có vùng phù rộng bao quanh khói u — mô hình gần như luôn nhận đúng vùng này.

Enhancing Tumor (ET)

Dice trên 0.60 cho thấy mô hình nhận diện tương đối tốt vùng tăng mạch.

- ET nổi bật trên T1ce sau tiêm;
- kích thước trung bình, ranh giới thường khá rõ;
- đóng vai trò quan trọng trong đánh giá ác tính.

Necrotic Core (NCR)

Dice chỉ khoảng 0.47–0.50.

- vùng nhỏ, nhiều chỗ ranh giới mờ;
- tín hiệu có thể tương tự nền;
- dễ bị mô hình “bỏ qua”.

Điều này tương đồng với con người: chúng ta thường tập trung vào các vùng lớn, nổi bật, trong khi bỏ lỡ những chi tiết nhỏ nhưng quan trọng.

5.3 Đánh giá định tính qua hình ảnh

Đánh giá chỉ số chưa đủ — phải nhìn trực tiếp mask dự đoán.

Các hình minh họa cho thấy:

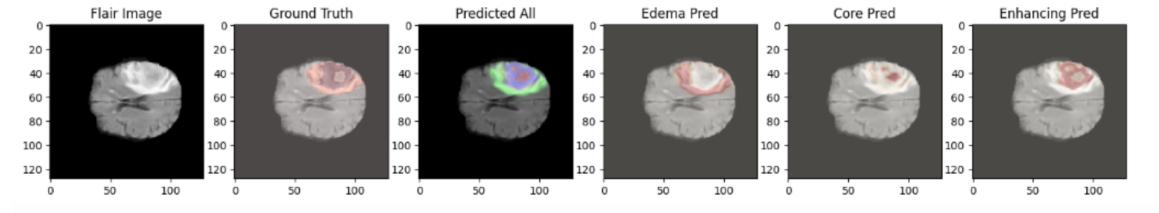
- Mask dự đoán bám khá sát nhãn bác sĩ.
- Các lớp ED, ET, NCR được phân biệt bằng màu sắc rõ ràng.

- Cấu trúc lõi–viền–phù được tái hiện tương đối chính xác.

Tuy nhiên:

- một số vùng biên còn gồ ghề;
- đoạn phù lan tỏa đôi khi bị “xé nhỏ”;
- ranh giới mịn chưa đạt như nhân chuyên gia.

Như vậy, mô hình đã tạo được “bản đồ chú ý không gian”, nhưng vẫn cần điều chỉnh để sát hơn với nhu cầu lâm sàng.



Hình 5.3.1: *Dánh giá định tính*

5.4 Cải tiến hàm mất mát: Dice Loss + CCE

5.4.1 Động cơ

CCE có một nhược điểm lớn:

- lớp nền chiếm đa số → ảnh hưởng mạnh đến loss;
- các lớp u hiếm gặp phần rất nhỏ vào tối ưu.

Hệ quả:

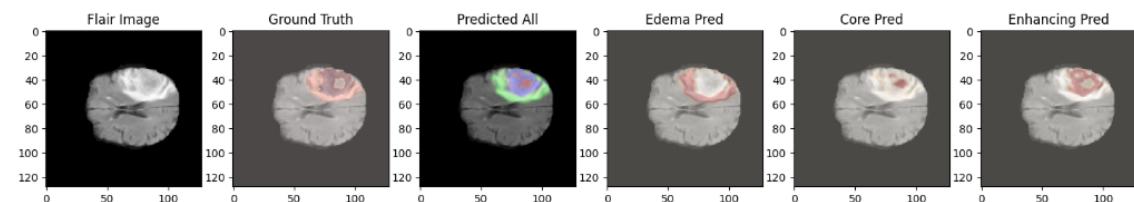
mô hình có xu hướng coi mọi thứ là nền.

Để khắc phục, nhóm bổ sung Dice Loss:

$$L = \text{CCE} + \lambda(1 - \text{Dice})$$

Dice Loss buộc mô hình:

- tối ưu trực tiếp sự chồng lấp vùng u,
- quan tâm hơn đến các lớp nhỏ,
- giảm bỏ sót các vùng quan trọng.



Hình 5.4.1: *Ảnh CCE được chụp trên một lát cắt*

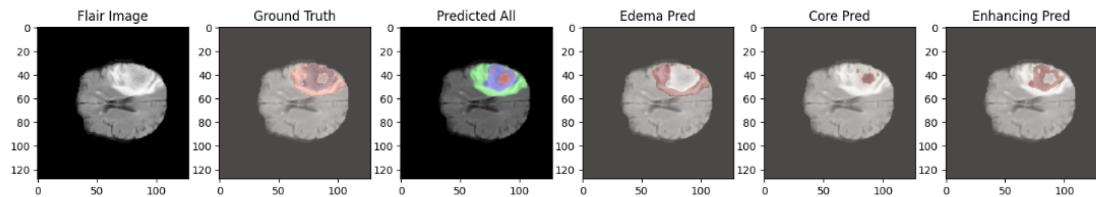
5.4.2 Kết quả sau cài tiến

Nhờ Dice + CCE:

- Dice tổng thể tăng lên > 0.73 ;
- Dice Necrotic tăng đáng kể: 0.57–0.58;
- Dice Edema và ET đều tăng quanh 0.69–0.70;
- Accuracy, Sensitivity, Specificity vẫn duy trì rất cao.

Về mặt trực quan:

- biên phân đoạn mượt hơn,
- mask u nhô không bị “vỡ vụn”,
- ranh giới khối u rõ và liên tục hơn.



Hình 5.4.2: Ảnh so sánh CCE vs (Dice+CCE) trên cùng 1 lát

Chương 6

Thảo luận khoa học nhận thức, hạn chế, hướng phát triển và kết luận

6.1 Học sâu như một mô hình nhận thức thị giác

Mạng U-Net 2D không chỉ là một công cụ kỹ thuật cho phân đoạn ảnh, mà còn có thể được xem như một mô hình tính toán mô phỏng cách hệ thống tri giác thị giác của con người hoạt động. Nhìn dưới lăng kính khoa học nhận thức, mỗi thành phần trong kiến trúc U-Net đều có một “hình ảnh tương ứng” với các cơ chế xử lý thông tin trong não.

- **Các lớp gần đầu vào giống các vùng thị giác sơ cấp (V1):**

Ở những lớp đầu tiên, receptive field của các neuron rất nhỏ, chúng phản ứng với những đặc trưng rất cơ bản như cạnh, đường thẳng, góc, hoặc sự thay đổi đột ngột về cường độ. Điều này tương tự như vùng V1 ở não, nơi neuron thị giác đáp ứng với các hướng cạnh và vị trí cụ thể trên võng mạc. U-Net 2D học được những đặc trưng bậc thấp này trực tiếp từ ảnh MRI, tạo nền tảng cho việc nhận diện các cấu trúc phức tạp hơn phía sau.

- **Các tầng sâu hơn học các cấu trúc trừu tượng hơn:**

Khi đi sâu vào mạng, receptive field của neuron lớn dần, cho phép chúng “nhìn” được vùng rộng hơn của ảnh. Tại đây, mạng không chỉ nhận diện cạnh hay góc, mà bắt đầu học được các pattern phức tạp hơn như hình dạng tổng thể của khối u, mối quan hệ giữa các chuỗi MRI (T1, T1ce, T2, FLAIR) và bối cảnh xung quanh u. Điều này giống với các vùng thị giác cao hơn (V2, V4, IT) nơi não không chỉ thấy “đường thẳng” mà thấy “vật thể”.

- **Skip connections như cơ chế top-down feedback:**

Skip connections cho phép thông tin từ các tầng sâu (mang tính trừu tượng, ngữ cảnh toàn cục) quay trở lại kết hợp với thông tin chi tiết từ các tầng nông. Điều này rất gần với khái niệm feedback trong hệ thị giác sinh học, nơi các vùng cao cấp gửi tín hiệu trở lại vùng thấp hơn để điều chỉnh ý nghĩa của những gì đang được nhìn thấy. Ví dụ, khi biết rằng “ở vùng này có u”, hệ thống sẽ diễn giải những tín hiệu mơ hồ xung quanh theo hướng “có khả năng là u”.

Quá trình huấn luyện mạng bằng gradient descent và hàm mất mát phản ánh cơ chế cập nhật synapse dựa trên lỗi dự đoán (prediction error) – một ý tưởng trung tâm trong lý thuyết *predictive coding*. Theo đó, não luôn cố gắng dự đoán đầu vào cảm giác, so sánh với thực tế và điều chỉnh mô hình nội tại để giảm sai số. Mạng U-Net, khi được huấn luyện để tối thiểu hóa loss, thực chất cũng đang thực hiện một quá trình “học từ sai lầm” tương tự.

6.2 Mất cân bằng lớp như một dạng bias nhận thức

Một trong những đặc điểm nổi bật của dữ liệu BRaTS là sự mất cân bằng lớp rất lớn: phần lớn voxel thuộc về nền hoặc mô não lành, trong khi voxel thuộc về u chỉ chiếm một tỷ lệ nhỏ. Tình huống này không chỉ là một vấn đề kỹ thuật, mà còn là một ví dụ tiêu biểu cho cách mô trường học tập có thể tạo ra *thiên lệch* trong hệ thống nhận thức – dù là sinh học hay nhân tạo.

- **Thiên lệch “mọi thứ đều lành” do kinh nghiệm lệch:**

Nếu hệ thống (mô hình hoặc con người) liên tục gặp các trường hợp “bình thường” và rất hiếm khi được trải nghiệm các trường hợp “bất thường nhưng quan trọng” như u nhổ, nó dễ hình thành một dạng bias: mặc định xem mọi thứ là bình thường và bỏ qua dấu hiệu hiếm. Trong bối cảnh y khoa, điều này tương đương với việc bác sĩ không được tiếp xúc đủ với ca bệnh hiếm, nên dễ bỏ sót khi gặp thực tế.

- **Tối ưu theo Accuracy càng làm lệch mạnh hơn:**

Nếu chỉ sử dụng Accuracy làm tiêu chí tối ưu, mô hình nhanh chóng học chiến lược “an toàn nhưng vô dụng”: dự đoán gần như tất cả pixel là nền. Accuracy vẫn cao vì nền chiếm đa số, nhưng năng lực phát hiện u lại rất kém. Đây là một ví dụ rõ ràng cho việc chọn sai thước đo dẫn đến hình thành chiến lược học sai.

Việc đưa Dice Loss vào hàm mất mát có thể được hiểu như một sự **tái thiết kế quy tắc phần thưởng**. Bay giờ, mô hình chỉ được “khen thưởng” khi nó chồng lấp tốt vùng u – tức là khi xử lý tốt chính những phần hiếm nhưng quan trọng. Điều này rất gần với cách các mô hình trong khoa học nhận thức mô tả vai trò của cấu trúc phần thưởng trong việc định hình chiến lược học của con người và động vật.

6.3 Overfitting và ranh giới giữa ghi nhớ – khái quát

Trong quá trình huấn luyện, khi số epoch tăng quá cao, mô hình bắt đầu có dấu hiệu overfitting:

- **Loss trên Train tiếp tục giảm, nhưng trên Validation không giảm đáng kể, thậm chí dao động:**

Điều này là dấu hiệu điển hình cho thấy mô hình đang “thuộc lòng” những chi tiết rất riêng của từng mẫu trong tập Train – kể cả nhiễu – nhưng không cải thiện khả năng xử lý dữ liệu mới. Trong một số trường hợp, hiệu năng trên Validation thậm chí còn xấu đi.

Từ góc độ khoa học nhận thức, đây chính là ranh giới giữa hai kiểu “học”:

- **Học khái niệm (concept learning):**

Hệ thống hình thành được cấu trúc tri thức trừu tượng, có thể áp dụng cho nhiều trường hợp khác nhau. Ví dụ, hiểu “khối u” là gì ở mức khái quát, chứ không chỉ nhớ từng mẫu cụ thể.

- **Học thuộc ví dụ (rote memorization):**

Hệ thống ghi nhớ chi tiết bề mặt của từng trường hợp riêng lẻ mà không nắm được quy luật sâu. Điều này giống như học sinh học thuộc đáp án mà không hiểu lý thuyết, dẫn đến không làm được bài mới chỉ thay đổi chút dữ liệu.

Các kỹ thuật như regularization, data augmentation, early stopping đóng vai trò giống như những “cơ chế quy tắc hóa” trong trí óc: chúng buộc hệ thống phải tìm những quy luật ổn định, ít phụ thuộc vào chi tiết ngẫu nhiên, từ đó tăng khả năng khái quát lên những tình huống mới.

6.4 Vai trò trong hệ thống nhận thức mở rộng

Mô hình phân đoạn u não không nhằm thay thế bác sĩ, mà được thiết kế để hoạt động như một **thành phần trong hệ thống nhận thức mở rộng** (extended cognition) – nơi trí tuệ con người và công cụ tính toán bổ trợ lẫn nhau.

- **Tự động hóa các tác vụ thị giác lặp lại, tiêu tốn nhiều tài nguyên nhận thức:**

Dánh dấu vùng u trên hàng trăm lát cắt là việc rất tốn thời gian và dễ gây mệt mỏi. Khi mô hình đảm nhận phần “nhìn” và “tố” này, bác sĩ có thể tiết kiệm đáng kể công sức.

- **Bác sĩ vẫn giữ vai trò trung tâm trong suy luận và ra quyết định:**

Mô hình chỉ cung cấp đề xuất (gợi ý phân đoạn), chứ không tự động ra chẩn đoán cuối cùng. Bác sĩ là người xem xét ảnh, đối chiếu với lâm sàng, lịch sử bệnh, kết quả xét nghiệm khác để quyết định phác đồ.

- **Tăng cường chứ không thay thế năng lực nhận thức con người:**

Khi được tích hợp vào phần mềm trực quan, mô hình trở thành một “cộng tác viên thị giác tốc độ cao”, giúp bác sĩ tập trung vào các tầng nhận thức cao hơn như lập luận, đối thoại với bệnh nhân, giải thích nguy cơ và lựa chọn điều trị.

Cách nhìn này rất phù hợp với quan điểm hiện đại trong khoa học nhận thức: trí tuệ không chỉ nằm trong bộ não riêng lẻ, mà được phân bố trong toàn bộ hệ thống người–công cụ–môi trường.

6.5 Hạn chế

Từ các kết quả thực nghiệm, có thể chỉ ra một số hạn chế chính của hệ thống hiện tại. Những hạn chế này vừa là điểm yếu kỹ thuật, vừa gợi ý cho các câu hỏi thú vị trong khoa học nhận thức về giới hạn của mô hình học sâu khi mô phỏng tri giác của con người.

Overfitting

Nếu huấn luyện quá lâu hoặc không áp dụng đủ kỹ thuật regularization, khoảng cách giữa hiệu năng trên Train và Validation có xu hướng tăng. Điều này đồng nghĩa với nguy cơ:

- mô hình hoạt động rất tốt trên dữ liệu quen thuộc;
- nhưng suy giảm đáng kể trên dữ liệu mới, chẳng hạn từ bệnh viện khác hoặc máy MRI khác.

Độ mịn đường biên

Ranh giới giữa vùng u và mô lành, nhất là ở vùng phì nè, chưa mượt như mask do bác sĩ vẽ tay. Vẫn còn:

- các pixel nhiễu, lỗ hổng nhỏ trong vùng u;
- đường biên “răng cưa”, không phản ánh đúng hình dạng sinh lý.

Điều này cho thấy mô hình vẫn chưa nắm bắt đầy đủ các ràng buộc hình học và cấu trúc liên tục của mô não.

Khó khăn với vùng Necrotic nhỏ

Mặc dù Dice cho vùng Necrotic đã được cải thiện khi dùng Dice Loss kết hợp CCE, nhưng vẫn thấp hơn đáng kể so với Edema:

- vùng này nhỏ, ít voxel, khó học;
- tín hiệu hình ảnh có thể mờ hoặc giống mô lành;
- dễ bị “át” bởi lớp nền trong quá trình tối ưu.

Giới hạn của U-Net 2D

Việc sử dụng U-Net 2D đồng nghĩa với việc mỗi lát được xử lý gần như độc lập, trong khi khối u là một cấu trúc 3D liên tục:

- mô hình không “thấy” trực tiếp hình dạng 3D của khối u;
- không khai thác được thông tin liên lát – vốn là điều bác sĩ luôn sử dụng khi lướt qua nhiều lát liên tiếp.

Độ tổng quát hóa theo thiết bị/đơn vị

Dữ liệu BRaTS đã được chuẩn hoá rất tốt, nhưng đó không phải là bức tranh đầy đủ của thực tế. Khi triển khai ra môi trường lâm sàng đa dạng hơn:

- khác biệt về loại máy, tham số chụp, nhiễu, độ phân giải;
- khác biệt về dân số bệnh nhân;
- có thể làm giảm hiệu năng nếu không có thêm bước thích nghi (adaptation).

6.6 Hướng phát triển

Để khắc phục các hạn chế đã nêu và tiến gần hơn tới một hệ thống vừa mạnh về kỹ thuật, vừa phù hợp với thực tiễn nhận thức và lâm sàng, có thể xem xét một số hướng phát triển sau.

Tối ưu hơn cho vùng Necrotic và các lớp hiém

- **Loss có trọng số:** tăng trọng số cho các lớp hiém, đặc biệt là Necrotic, để chúng đóng vai trò lớn hơn trong hàm mất mát.
- **Kết hợp Focal Loss, Tversky Loss:** các hàm mất mát này được thiết kế đặc biệt cho dữ liệu mất cân bằng, phạt mạnh hơn những lỗi trên lớp hiém và vùng nhỏ.

Tăng cường regularization và dữ liệu

- **Data augmentation mạnh:** xoay, dịch, zoom, lật, jitter cường độ, thêm nhiễu... giúp mô hình học được những đặc trưng ổn định, ít phụ thuộc vào vị trí hoặc cường độ tuyệt đối.
- **Dropout, weight decay:** buộc mô hình không phụ thuộc quá nhiều vào một số neuron, từ đó giảm nguy cơ overfitting.
- **Semi-supervised learning:** tận dụng các ảnh chưa gán nhãn (nếu có) để mô hình học thêm được cấu trúc dữ liệu, giống như con người học không chỉ từ những ví dụ có nhãn rõ ràng mà còn từ tiếp xúc thường xuyên với môi trường.

Khai thác thông tin 3D

- **U-Net 3D:** xử lý trực tiếp toàn bộ volume 3D, cho phép mô hình học hình dạng khối u theo cả ba chiều.
- **Mô hình 2.5D:** dùng chồng nhiều lát kề nhau làm đầu vào cho một mạng 2D, vừa tận dụng được thông tin trục z, vừa không đòi hỏi tài nguyên như 3D.

Cơ chế chú ý và học đa chuỗi thông minh hơn

- **Channel attention:** cho phép mô hình tự quyết định chuỗi MRI nào quan trọng hơn tại mỗi vị trí. Ví dụ, khi tìm ET thì T1ce được “ưu tiên nhìn” hơn.
- **Spatial attention:** mô hình học cách tập trung vào những vùng không gian có khả năng chứa u, tương tự cơ chế chú ý không gian của con người khi đọc phim.

Tích hợp vào quy trình lâm sàng thực tế

- **Xây dựng giao diện tương tác:** nơi bác sĩ có thể xem ảnh gốc, mask dự đoán, chỉnh sửa nếu cần, và lưu lại thông tin để cải thiện mô hình.
- **Đánh giá hiệu quả thực tế:** không chỉ nhìn vào chỉ số Dice, mà còn đo lường thời gian tiết kiệm được, số ca được phát hiện sớm hơn, mức độ hài lòng của bác sĩ khi sử dụng hệ thống.

6.7 Kết luận chung

Báo cáo đã trình bày toàn bộ quá trình:

- xây dựng và huấn luyện mô hình U-Net 2D trên dữ liệu BRaTS để phân đoạn khối u não;
- phân tích chi tiết các bước tiền xử lý, thiết kế mô hình, hàm mất mát, chiến lược huấn luyện và hệ thống chỉ số đánh giá;
- làm rõ những khó khăn do mất cân bằng lớp, nguy cơ overfitting và bias khi chọn tiêu chí đánh giá;
- kết nối các kết quả kỹ thuật với các khái niệm cốt lõi của khoa học nhận thức như học biểu diễn, thiên lệch nhận thức, ranh giới giữa học thuộc và khái quát, và khái niệm hệ thống nhận thức mở rộng.

Về mặt kỹ thuật, mô hình đạt:

- Accuracy $\sim 99.37\%$, Mean IoU ≈ 0.80 ;
- Sensitivity và Specificity đều trên 99%;
- Dice cho vùng Edema khoảng 0.7, các vùng Enhancing và Necrotic đạt mức chấp nhận được và cải thiện rõ rệt khi kết hợp Dice Loss với CCE.

Về mặt khoa học nhận thức, hệ thống là một minh họa sống động cho việc:

- một mô hình học sâu có thể học được biểu diễn nội bộ từ dữ liệu cảm giác phức tạp như MRI đa chuỗi;
- cách phân bố dữ liệu và thiết kế tiêu chí đánh giá định hình hành vi và chiến lược học của hệ thống;
- các mô hình AI có thể đóng vai trò là thành phần của một hệ nhận thức lai người-máy, hỗ trợ và tăng cường năng lực của con người thay vì thay thế hoàn toàn.

Những kết quả này cho thấy học sâu là một công cụ đầy hứa hẹn cho bài toán phân đoạn u não, đồng thời mở ra nhiều câu hỏi thú vị cho khoa học nhận thức về việc làm thế nào để thiết kế, đánh giá và tích hợp các mô hình trí tuệ nhân tạo vào hệ sinh thái nhận thức rộng lớn của con người trong thực hành lâm sàng.