Phát triển hàm mất mát cho mạng TransUnet trong phân vùng ảnh MRI khối u não

Trần Thị Thảo*, Phạm Văn Trường, Nguyễn Hữu Thắng

Trường Đại học Bách khoa Hà Nội.
*Email: thao.tranthi@hust.edu.vn.
Nhận bài ngày 28/02/2022; Hoàn thiện ngày 21/3/2022; Chấp nhận đăng ngày 10/4/2022.
DOI: https://doi.org/10.54939/1859-1043.j.mst.78.2022.28-38

TÓM TẮT

Phân vùng ảnh khối u não trong chụp cộng hưởng từ MRI (Magnetic resonance imaging) rất hữu ích cho việc chẩn đoán, dự đoán tốc độ phát triển, đo thể tích khối u và lập phác đồ điều trị khối u não. Tuy vậy, việc phân vùng ảnh MRI khối u não thực tế gặp khó khăn do sự đa dạng của các khối u não về kích thước, hình dạng, vị trí và sự không đồng nhất của chúng. Trong bài báo này, chúng tôi đề xuất một hướng tiếp cận cho phân vùng ảnh MRI khối u não dùng mạng nơ-ron học sâu, với hàm mất mát dựa trên hàm Tversky. Cụ thể, chúng tôi đề xuất sử dụng mô hình TransUnet, một mô hình được giới thiệu gần đây dựa trên kiến trúc Transformer và U-Net để huấn luyện và kiểm thử dữ liệu. Đồng thời chúng tôi cũng đề xuất một hàm mất mát mới dùng để huấn luyện mạng nơ-ron qua đó có thể giải quyết những khó khăn vừa nêu trong phân vùng ảnh MRI khối u não. Phương pháp đề xuất đã được kiểm chứng trên tập dữ liệu Brain LGG Segmentation, kết quả cho thấy mạng TransUNet với hàm mất mát đề xuất hoạt động tốt, với các chỉ tiêu đánh giá có kết quả cao hơn so với một số phương pháp khác.

Từ khóa: Mạng nơ-ron học sâu; Mô hình TransUnet; Phân vùng ảnh MRI khối u não; Hàm Tversky.

1. ĐẶT VẤN ĐỀ

U não là một khối tăng trưởng của các tế bào bất thường trong não, và là một trong những bệnh gây tử vong hàng đầu trong các bệnh nhân về ung thư [1]. U não có thể chia thành u não lành tính và u não ác tính. Chẩn đoán hình ảnh là phương pháp thường được sử dụng để chẩn đoán u não. Các kỹ thuật chẩn đoán hình ảnh có thể kể đến là phương pháp chụp cắt lớp CT(computerized tomography), phương pháp cộng hưởng từ hạt nhân (MRI). Cho dù khối u não là lành tính, ác tính hay di căn, tất cả đều có khả năng đe dọa nghiêm trọng tới tính mạng bởi não được bao bọc trong hộp xương sọ, nó không thể mở rộng để có chỗ cho khối lượng ngày càng lớn của khối u. Kết quả là, khối u chèn ép và chiếm chỗ của các mô não khỏe mạnh. Việc xác định chính xác kích thước cũng như vị trí khối u có thể giúp cho các bác sĩ trong quá trình chẩn đoán, điều trị bệnh.

Việc xác định khối u nếu thực hiện thủ công cần phải có bác sĩ chẩn đoán hình ảnh sử dụng thông tin thu bởi ảnh MRI cùng với những kiến thức về giải phẫu và sinh lý có được qua nhiều năm nghiên cứu và thực nghiệm trên lĩnh vực y khoa. Quá trình này đòi hỏi bác sĩ phải xem xét nhiều hình ảnh từng lát cắt, chẩn đoán khối u và vẽ thủ công các vùng khối u một cách cẩn thận. Ngoài việc tốn thời gian, việc phân vùng thủ công cũng phụ thuộc vào kinh nghiệm và kiến thức y khoa của bác sĩ chẩn đoán hình ảnh. Do đó, kết quả phân vùng có thể khác nhau giữa những lần thực hiện và giữa ý kiến chuyên khoa của các bác sĩ khác [2].

Thực tế trong ảnh MRI của khối u, có rất nhiều thách thức. Thách thức đầu tiên có thể kể đến đó là sự đa dạng về vị trí, kích thước và hình khối. Kích thước và vị trí của các khối u là khác nhau ở mỗi bệnh nhân. Tuy nhiên, ngay cả ở cùng một bệnh nhân, kích thước của khối u cũng khác nhau trên từng lát cắt ảnh. Sự phức tạp này có thể dễ dàng nhận thấy khi quan sát ảnh MRI 3D. Khi chia từng lát cắt của ảnh 3D thành từng lát cắt 2D, kích thước của khối u thay đổi trên từng lát cắt. Điều này khiến cho việc phân vùng khối u trở nên khó khăn hơn. Bên cạnh đó, diện tích khối u rất nhỏ so với kích thước não. Bên cạnh đó, một khó khăn nữa trong việc xác định

khối u một cách tự động đó là sự không nhất quán về hình dạng. Hình dạng của khối u ở mỗi bệnh nhân là khác nhau, không có một khuôn mẫu hay đặc điểm cụ thể nào.

Ngày nay, nhờ những tiến bộ của khoa học máy tính và học máy, việc phân vùng ảnh MRI khối u não một cách tự động và chính xác có thể hỗ trợ các bác sĩ ra quyết định chẩn đoán và giúp đỡ các bệnh nhân. Với những tập dữ liệu liên tục được đóng góp bởi các nhà khoa học, số lượng những ca bệnh dùng cho huấn luyện mô hình ngày càng gia tăng, giúp cho máy có thể học được kinh nghiệm từ rất nhiều chuyên gia. Khi số lượng mẫu ảnh gia tăng đến mức đủ lớn, kết quả của mô hình có thể ngang hoặc thậm chí chính xác hơn chẩn đoán của bác sĩ. Đã có nhiều phương pháp tiếp cận trong bài toán phân vùng ảnh u não như hướng sử dụng đường bao chủ động [3, 4], và các phương pháp học máy [5-8]. Với những thành tựu rực rỡ của kỹ thuật học sâu, nhất là mạng nơ-ron tích chập CNN (Convolutional Neural Network), đã giúp cho các bài toán phân vùng ảnh nói chung và phân vùng ảnh y tế nói riêng thu được rất nhiều thành tựu. Có thể kể đến một số mô hình phân vùng ảnh dùng học sâu nổi tiếng, đó là mạng tích chập toàn phần FCN (Fully Convolutional Network), [9] mạng U-Net [10], và gần đây là mô hình TransUnet [11]- một cấu trúc dựa trên Transformer vốn nổi tiếng trong xử lý ngôn ngữ tự nhiên. Các mô hình mạng kể trên và các mô hình phát triển dựa trên hai cấu trúc mạng này đã được ứng dụng nhiều trong lĩnh vực phân vùng ảnh [6, 12].

Bên cạnh việc phát triển các kiến trúc mới, hàm mất mát trong phân vùng ảnh cũng được các nhà nghiên cứu trong lĩnh vực thị giác máy tính đặc biệt chú trọng. Vì tính chất mất cân bằng dữ liệu là tính chất điển hình của dữ liệu y tế, các hàm mất mát cần phải tập trung giải quyết. Hàm mất mát cross-entropy (CE) [10], cross-entropy có trọng số được ứng dụng trong phân vùng cấu tạo của não dựa trên ảnh MRI [13], hay hàm mất mát Dice [14], hàm mất mát Tversky [15] là những ví dụ tiêu biểu cho phân vùng ảnh y tế. Nghiên cứu cho thấy sử dụng những hàm mất mát dạng này đem lại sự hiệu quả vượt trội. Hơn nữa, việc áp dụng hàm mất mát lại cực kỳ đơn giản, không tốn chi phí. Vì vậy, hiện nay, sử dụng các hàm mất mát giải quyết việc mất cân bằng dữ liệu được coi là mặc định trong xử lí ảnh y tế dùng kỹ thuật học sâu. Trong quá trình nghiên cứu về ảnh y tế cũng như học sâu, chúng tôi nhận thấy hướng phát triển các hàm mất mát trong kỹ thuật học sâu trong các ứng dụng phân vùng ảnh y tế nói chung và phân vùng khối u não còn có nhiều tiềm năng để phát triển. Trong bài báo này chúng tôi đề xuất một hàm mất mát mới, giải quyết việc mất cân bằng dữ liệu, góp phần giải quyết một số thách thức của phân vùng khối u não như là sự đa dạng của các khối u não về kích thước, hình dạng, vị trí và sự không đồng nhất của chúng. Đồng thời chúng tôi đề xuất sử dụng một mô hình mạng hiện đại dựa trên kiến trúc transformer, là TransUnet TransUnet [11]- và thiết lập các cấu hình so cho phù hợp và thích nghi với bài toán phân vùng ảnh u não.

Phần tiếp theo của bài báo gồm các phần sau: Phần 2 khái quát một số mô hình và hàm mất mát trong phân vùng ảnh nói chung dùng kỹ thuật học sâu; Phần 3 trình bày về kiến trúc TransUnet và hàm mất mát đề xuất; Một số kết quả thực nghiệm và đánh giá được giới thiệu trong Phần 4; Cuối cùng, phần 5 kết thúc với kết luận và hướng phát triển.

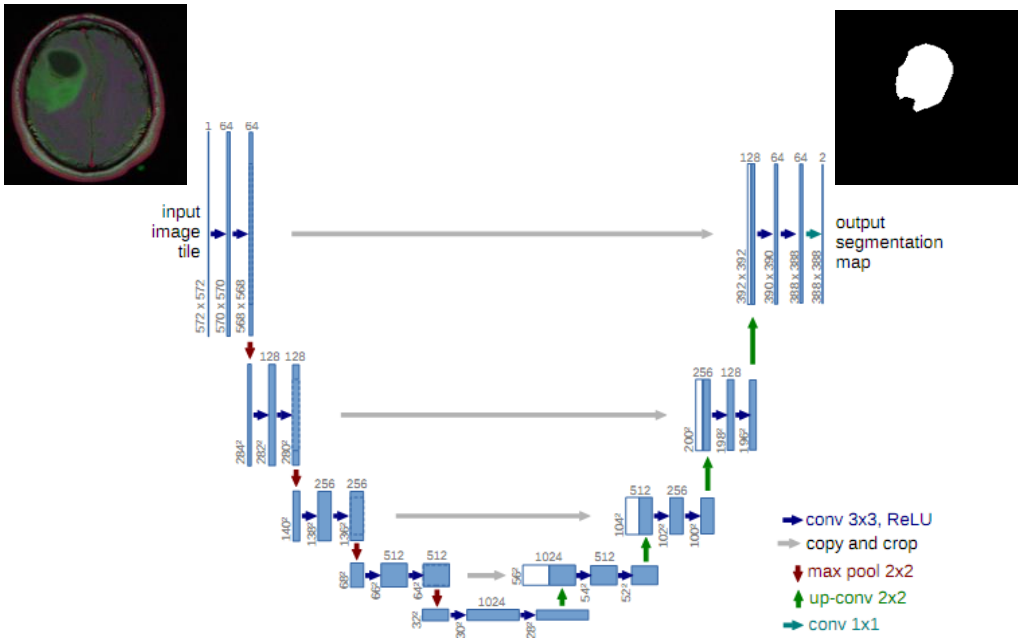
2. CÁC NGHIÊN CỨU LIÊN QUAN

2.1. Mô hình U-Net và các mô hình mở rộng của U-Net

Mô hình mạng U-Net [5], được giới thiệu bởi Olaf Ronnenberger và cộng sự, dành cho phân vùng ảnh y sinh vào năm 2015, là một mô hình được cải tiến và phát triển dựa trên mô hình mạng nơ-ron tích chập toàn phần đã được giới thiệu trước đó. Sở dĩ mô hình có tên là U-Net do cấu trúc đối xứng của mình, như minh họa ở hình 1. Phần mã hóa trong U-Net tương tự như các mạng nơ ron tích chập truyền thống, gồm các lớp tích chập và các lớp giảm chiều nhằm trích xuất những đặc trưng của ảnh.

Điều nổi bật của mạng U-Net chính là lớp giải mã. Tại lớp này, số lần mở rộng chiều sẽ tương

ứng với số lần giảm chiều ở các lớp trước đó. Trong một số mô hình phát triển dựa trên U-Net, thay vì dùng lớp mở rộng kết hợp với một lớp tích chập, ta có thể sử dụng các lớp tích chập đảo với nhiều cửa sổ trượt.



Hình 1. *Minh họa mô hình U-Net. Ví dụ trong phân vùng khối u não.*

Một mô hình mạng phát triển của U-Net có thể kể đến là Attention UNet [6] được giới thiệu vào năm 2018. Cơ chế tập trung được thêm vào giúp mạng gia tăng sự hiệu quả trong việc xác định chính xác phân vùng ảnh. Trong quá trình tăng mẫu ở phía decoder, nếu chỉ dùng các phép tích chập chuyển vị hoặc phép nội suy thì thông tin về mặt không gian được tái tạo là không đủ chính xác. Để khắc phục nhược điểm này, U-Net sử dụng kết nối tắt (skip connection) kết hợp thông tin không gian từ khối encoder với khối decoder. Tuy nhiên, điều này dẫn đến nhiều phần trích xuất đối tượng cấp thấp dư thừa, vì tính năng biểu diễn kém trong các lớp ban đầu. Sự tập trung mềm được thực hiện tại các kết nối tắt sẽ chủ động ngăn chặn các kích hoạt ở các vùng không liên quan, giảm số lượng các thông tin dư thừa được đưa qua mạng. Nhờ cơ chế này, mạng Attention-Unet học cách tập trung vào khu vực mong muốn khi quá trình đào tạo diễn ra. Tính khả vi của cổng tập trung mềm cho phép mạng huấn luyện được trong quá trình lan truyền ngược, có nghĩa là các hệ số tập trung trở nên tốt hơn trong việc làm nổi bật các vùng ảnh quan trọng. Các cổng tập trung là một cách đơn giản để cải thiện U-Net một cách hiệu quả trong nhiều bộ dữ liệu khác nhau mà không tốn kém đáng kể về chi phí tính toán.

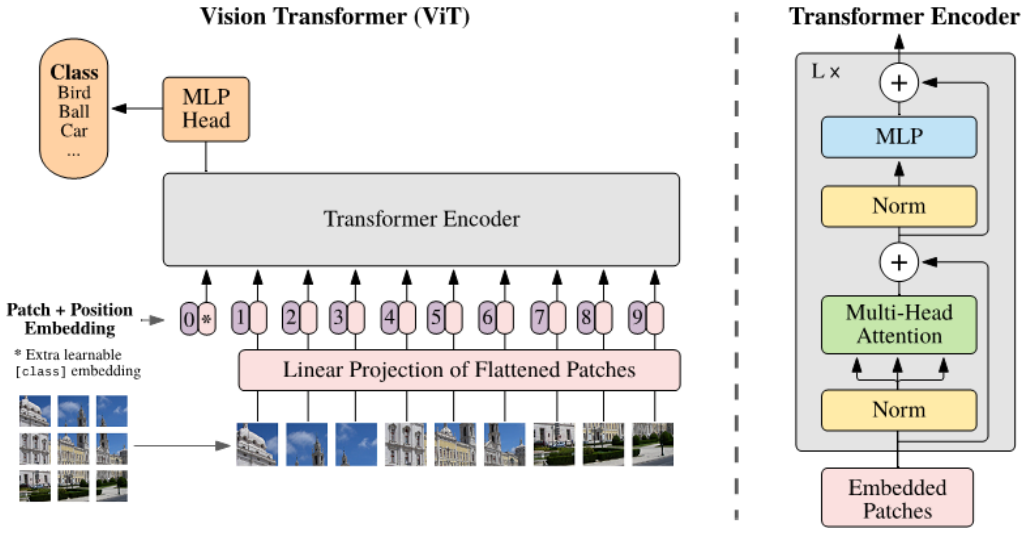
2.2. Transformers

Transformers, hay còn gọi là "người máy biến hình", được Vaswani và các cộng sự đề xuất [16] ban đầu cho bài toán xử lý ngôn ngữ tự nhiên NLP (Natural language processing). Khác với cơ chế xử lí chuỗi tuần tự, Transformer có thể xử lí đồng thời các vector trong chuỗi đầu vào, khiến các quá trình huấn luyện nhanh hơn rất nhiều. Việc tính toán song song có thể tận dụng được sức mạnh của các GPU, cho phép chúng ta có thể thử nghiệm nhanh chóng các ý tưởng, các kiến trúc mới dựa trên Transformer. Và thực tế cũng cho thấy, kiến trúc dựa trên Transformer đang là kiến trúc nổi tiếng nhất, thu hút nhiều sự tập trung nhất, không chỉ trong xử lí ngôn ngữ tự nhiên mà còn trong nhiều lĩnh vực khác. Không chỉ cải thiện về mặt tốc độ, Transformer còn có những cơ chế cải thiện sức mạnh của mô hình, độ chính xác và khả năng trích xuất thông tin một cách vượt trội. Giống như các kiến trúc chung trong các mạng nơ-ron xử

lí ngôn ngữ tự nhiên, Transformer cũng bao gồm hai khối chính: Encoder và Decoder.Khối Encoder có nhiệm vụ phân tích và trích xuất các đặc trưng từ chuỗi vector đầu vào, khối Decoder có chức năng xử lí các thông tin đặc trưng từ Encoder để tính toán vector đầu ra. Cơ chế chủ đạo tạo nên sự thành công của Transformer là cơ chế Tập trung nội bộ (Self-attention) và tập trung nội bộ nhiều đầu (Multi-head Self-attention). Một cơ chế tập trung có thể được mô tả như ánh xạ một query và một tập hợp các cặp key - value tới một kết quả đầu ra, trong đó query, key, value và kết quả đầu ra đều là các vector. Kết quả đầu ra được tính dưới dạng tổng có trọng số của các value, trong đó trọng số được gán cho mỗi value được tính bằng một hàm tương thích của query với key tương ứng.

2.3. Vision Transformer

Bên cạnh các thành tựu vượt trội trong lĩnh vực xử lý ngôn ngữ tự nhiên, Transformer đã và đang được áp dụng hiệu quả trong thị giác máy tính [11, 16]. Một trong các nghiên cứu điển hình đó là Vision Transformers (ViTs) giới thiệu bởi Dosovitskiy và các cộng sự [17].



Hình 2. *Minh họa kiến trúc ViT trong thị giác máy tính.*

Trong mô hình ViT, như minh họa ở hình 2, để áp dụng Transformer cho ảnh, các tác giả chia cắt bức ảnh thành các mảnh ghép. Giả sử kích thước ảnh gốc là $x \in R^{H\times W\times C}$, các mảnh ghép sẽ được cắt với kích thước là $x_p \in R^{N\times(P.P.C),}$ trong đó (H,W) lần lượt là độ cao và chiều rộng của bức ảnh gốc, C là số kênh của ảnh, còn (P,P) là kích thước của mảnh ghép. Như vậy, số mảnh ghép thu được sẽ là $N = HW/P^2$. Kỹ thuật này được gọi là patching. Sau khi chia cắt ảnh gốc, các mảnh ghép sẽ được duỗi thẳng thành một vector và được chiếu thành một vector D chiều qua một lớp kết nối đầy đủ. Kết quả của phép chiếu này được gọi là các mảnh ghép nhúng (patching embeddings). Đây là một phương hướng tiếp cận khá mới mẻ và độc đáo. Việc chia cắt bức ảnh thành nhiều phần, rồi lại duỗi thẳng chúng thành các vector, đã làm mất thông tin về không gian của bức ảnh. Tuy nhiên, kết quả mang lại thì lại khá bất ngờ. Phân chia mảnh ghép cũng chính là lí do mà tên của bài báo được đặt là "An image is worth 16 ×16 words", một bức ảnh có giá trị bằng 16x16 từ. Cũng giống như xử lí ngôn ngữ tự nhiên, các mảnh ghép trong Vision Transformer (ViT) cũng cần được mã hóa theo vị trí. ViT sử dụng mã hóa theo vị trí một chiều với các tham số có khả năng học được.

2.4. Các hàm mất mát thường dùng trong phân vùng ảnh

Trong các bài toán phân vùng ảnh dùng kỹ thuật học sâu, các hàm mất mát thường được sử dụng là hàm Dice loss và cross-entropy. Phần dưới đây mô tả sơ lược về các hàm mất mát này trong trường hợp phân loại nhị phân hay hai lớp: đối tượng quan tâm (ví dụ khối u não) và nền.

Giả sử gọi $y \in [0,1]$ là ảnh nhị phân- đầu ra dự đoán bởi mạng nơ-ron, $\hat{y} \in [0,1]$ tương ứng là nhãn-một ảnh nhị phân được khoanh vùng bởi bác sĩ hoặc chuyên gia (ground truth). Mục tiêu của mạng là sao cho ảnh đầu ra khớp với ground truth nhất trong quá trình huấn luyện. Giả sử các nhãn này có tổng số pixel là *N*.

Hàm mất mát Dice được định nghĩa theo công thức sau:

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} \hat{y}_i y_i}{\sum_{i=1}^{N} \hat{y}_i + \sum_{i=1}^{N} y_i} \quad (1)$$

Hàm mất mát binary cross entropy (BCE) được biểu diễn như sau:

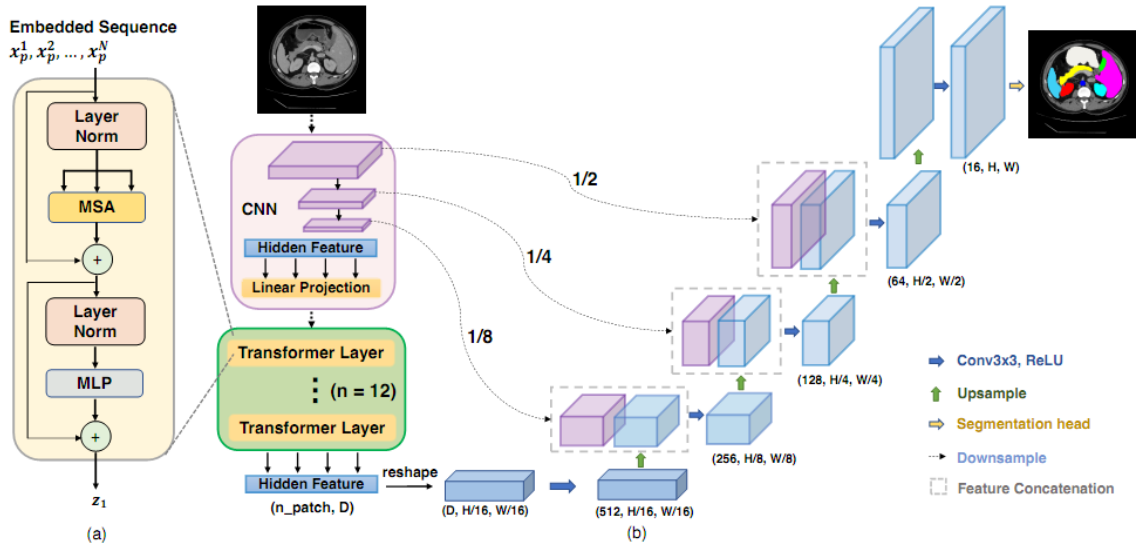
$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N} \hat{y}_i \log y_i + (1-\hat{y}_i)\log(1-y_i) \quad (2)$$

Thông thường trong đa phần các ứng dụng phân vùng ảnh, hai hàm mất mát trên thường được kết hợp với nhau, còn gọi là hàm mất mát Dice+BCE để tăng hiệu quả của mô hình mạng.

3. PHƯƠNG PHÁP THỰC HIỆN

3.1. Kiến trúc TransUnet

TransUnet [11] được công bố vào đầu năm 2021, lần đầu tiên kết hợp giữa ViT và mạng nơ-ron UNet. TransUNet không chỉ kế thừa ý tưởng của ViT mà còn phát triển để khắc phục điểm yếu của mạng. Các tác giả của TransUNet nhận thấy rằng, nếu sử dụng Vision Transformer làm bộ Encoder cho mạng phân vùng ảnh thì kết quả mang lại chưa đủ thuyết phục. Dựa trên phân tích ViT duỗi thẳng các vùng mảnh ghép thành vector một chiều làm mất thông tin về mặt không gian, và cũng không mang lại thông tin cho khối Decoder trong quá trình tăng mẫu, TransUNet đề xuất thêm vào các lớp tích chập trước khi đưa vào các khối tập trung nhiều đầu. Điều này đã giúp cho TransUNet tận dụng được cả ưu điểm của Transformer, và kỹ thuật Encode kinh điển bằng mạng tích chập trong phân vùng ảnh y tế.



Hình 3. *Minh họa kiến trúc mạng TransUnet cho một ứng dụng phân vùng ảnh y tế (a) lớp Transformer; (b) cấu trúc của TransUNet.*

Khối Encoder của TransUNet sử dụng nhiều tầng Transformer đặt xếp chồng lên nhau. Sự cải tiến của TransUNet thể hiện ở việc không duỗi thẳng các mảnh ghép nhúng ra thành một

vector, mà đưa chúng qua một mạng tích chập để trích xuất đặc trưng. Kỹ thuật này được gọi là Hybrid CNN - Transformer, được thể hiện trên hình 3. Chính các tầng tích chập được thêm vào này đã làm tăng độ chính xác của mạng. Tác giả của TransUNet đưa ra hai lí do cho thiết kế này: (i) Thứ nhất, việc sử dụng các tầng tích chập làm trích xuất đặc trưng giúp cho các vector nhúng giữ lại được thông tin về mặt không gian; (ii)Thứ hai, thông qua các tầng tích chập, đặc trưng được trích xuất với nhiều độ phân giải khác nhau, do đó, có thể hỗ trợ quá trình Decode (giống như trong U-Net).

Ở khối Decoder, TransUNet sử dụng một kiến trúc với tên gọi Cascaded Upsampler, tức là tăng kích thước của đặc trưng theo nhiều tầng để giải mã mặt nạ đầu ra về kích thước bằng với đầu vào. Sau khi đi qua khối Encoder, kích thước của đặc trưng có dạng $z_L \in (HW/P^2, D)$, được sắp xếp lại thành kích thước $H/P \times W/P \times D$. Mục đích của việc này là để đưa đặc trưng từ không gian 2 chiều về không gian 3 chiều để có thể thực hiện được các phép toán tích chập. Qua mỗi phép upsampling, kích thước của đặc trưng được tăng lên gấp đôi, và cuối cùng khôi phục lại được kích thước gốc. Chúng ta có thể thấy rằng, sử dụng Cascaded Upsampler cho phép tổng hợp đặc trưng ở các lớp độ phân giải khác nhau, giữ lại được cả đặc trưng ở tầng thấp và tầng cao. Đây cũng là kỹ thuật rất phổ biến trong mạng nơ-ron.

3.2. Hàm mất mát đề xuất

Như đã đề cập trong phần giới thiệu, trong bài toán phân vùng khối u não, mặc dù đã có rất nhiều nghiên cứu và phương pháp đề xuất, vẫn còn có những thách thức. Một trong số đó là sự mất cân bằng dữ liệu, kích thước của khối u thường nhỏ hơn nhiều so với nền xung quanh của bức ảnh, do tính chất đa dạng của hình thái, kích thước khối u. Để góp phần giải quyết vấn đề này, trong nghiên cứu này, chúng tôi đề xuất một giải pháp đơn giản nhưng không làm tăng chi phí tính toán hay thông số của mạng, đó là cải tiến hàm mất mát. Cụ thể, chúng tôi đề xuất hàm mất mát như sau:

$$L_{Proposed} = \lambda L_{\text{BCE}} + \gamma L_{Tversky} \tag{3}$$

Trong đó, $\lambda, \gamma \in [0, 1]$ là các siêu tham số biểu diễn sự cân bằng giữa hai thành phần hàm mất mát; L_{BCE} là hàm binary cross entropy như biểu diễn ở công thức (2); và L_{Tversky} là hàm mất mát dựa trên chỉ số Tversky như mô tả ở dưới đây:

Với $y \in [0,1]$ là đầu ra dự đoán bởi mạng nơ-ron, $\hat{y} \in [0,1]$ tương ứng là ground truth. Chỉ số Tversky (Tversky index) [18] được định nghĩa như sau:

$$T(\hat{y}, y, \alpha, \beta) = \frac{\sum_{i=1}^{N} \hat{y}_i y_i}{\sum_{i=1}^{N} \hat{y}_i y_i + \alpha \sum_{i=1}^{N} \sum_{i=1}^{N} \hat{y}_i (1 - y_i) + \beta \sum_{i=1}^{N} (1 - \hat{y}_i) y_i} \tag{4}$$

trong đó, $0 \le \alpha, \beta \le 1$ là các tham số của Tversky index với $\alpha + \beta = 1$. Công thức (4) có thể viết lại thành:

$$T(\hat{y}, y, \alpha) = \frac{\sum_{i=1}^{N} \hat{y}_i y_i}{\sum_{i=1}^{N} \hat{y}_i y_i + \alpha \sum_{i=1}^{N} \sum_{i=1}^{N} \hat{y}_i (1 - y_i) + (1 - \alpha) \sum_{i=1}^{N} (1 - \hat{y}_i) y_i} \tag{5}$$

Dựa trên Tversky index, hàm mất mát Tversky được viết như sau:

$$L_{\text{Tversky}} = 1 - T(\hat{y}, y, \alpha) = 1 - \frac{\sum_{i=1}^{N} \hat{y}_i y_i}{\sum_{i=1}^{N} \hat{y}_i y_i + \alpha \sum_{i=1}^{N} \sum_{i=1}^{N} \hat{y}_i (1 - y_i) + (1 - \alpha) \sum_{i=1}^{N} (1 - \hat{y}_i) y_i} \tag{6}$$

3.3. Chỉ số đánh giá kết quả

3.3.1. Chỉ số IoU (Intersection over Union)

Chỉ số IoU, hay còn gọi là hệ số tương đồng Jaccard, là một phép thống kê được sử dụng để đánh giá sự tương đồng giữa các tập mẫu. Phép đo nhấn mạnh sự giống nhau giữa các tập mẫu hữu hạn và được định nghĩa chính thức là kích thước của phần giao chia cho kích thước của phần hợp của các tập mẫu. Biểu diễn toán học của chỉ số được viết như sau:

$$J\left(A,B\right)=\frac{\left|A\cap B\right|}{\left|A\cup B\right|}=\frac{\left|A\cap B\right|}{\left|A\right|+\left|B\right|-\left|A\cap B\right|} \tag{7}$$

3.3.2. Chỉ số Dice - Dice similarity coefficient (DSC)

Hệ số tương đồng Dice, còn được gọi là hệ số Sørensen-Dice, là một công cụ thống kê đo mức độ tương đồng giữa hai tập dữ liệu. Hệ số này thường được sử dụng rộng rãi và được coi là phép đo chính trong đánh giá các thuật toán phân vùng xử lý hình ảnh. Hệ số DSC có thể được viết như sau:

$$DSC\left(A,B\right)=\frac{2\times\left|A\cap B\right|}{\left|A\right|+\left|B\right|} \tag{8}$$

4. THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

4.1. Dữ liệu và phân tích dữ liệu

Tập dữ liệu được dùng trong nghiên cứu này là tập LGG Segmentation Dataset. Bộ dữ liệu bao gồm 3929 ảnh MRI não của hơn 100 bệnh nhân, dưới nhiều lát cắt khác nhau. LGG Segmentation cung cấp nhãn cho các khối u, được chính các nhà nghiên cứu và chẩn đoán hình ảnh tại Đại học Duke thực hiện và phê duyệt. Tập dữ liệu đã được sử dụng cho nhiều nghiên cứu và cũng được đưa vào cuộc thi phân vùng ảnh trên một trang web nổi tiếng về học máy - Kaggle. Trong tổng số 3929 ảnh, có đến hơn 2500 ảnh được chẩn đoán dương tính và chỉ có khoảng 1500 ảnh là âm tính. Trên thực tế, tỉ lệ số người bị u não là rất nhỏ so với người khỏe mạnh. Việc tạo một phân phối dữ liệu như vậy hỗ trợ rất hiệu quả cho các mô hình học máy, khi mà phân lớp dương tính khó chẩn đoán hơn được bổ sung nhiều mẫu hơn.

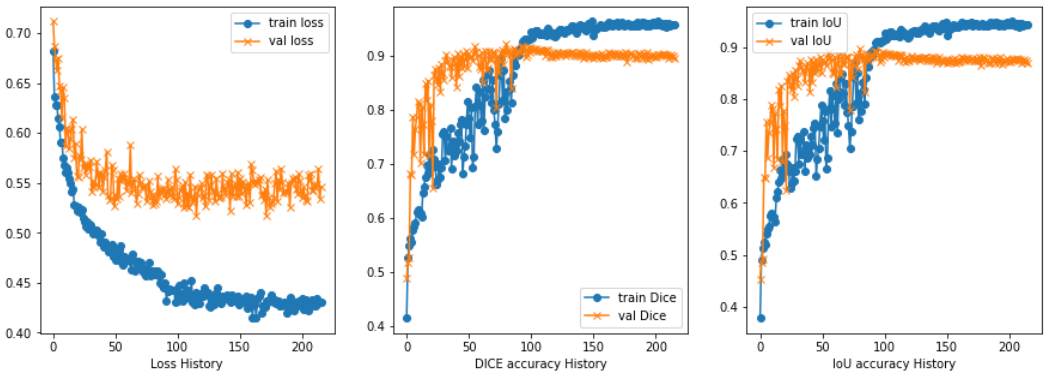
4.2. Huấn luyện và cài đặt tham số

Trong nghiên cứu này, chúng tôi cài đặt các tham số cho mô hình TransUNet để thích nghi cho bài toán phân vùng ảnh khối u não như sau: Kích thước ảnh đầu vào là 256 ×256; Kích thước mảnh ghép là 16 ×16; Số đầu tập trung là 12; số lớp Transformer là 12; số chiều của lớp ẩn là 768. Tỷ lệ dữ liệu huấn luyện, kiểm thử và kiểm tra được đặt như sau: tập huấn luyện: 70%; tập kiểm thử: 15%; và tập kiểm tra: 15%. Bên cạnh đó chúng tôi cũng sử dụng các kỹ thuật làm giàu dữ liệu như tăng giảm ngẫu nhiên độ sáng, tăng giảm ngẫu nhiên độ tương phản, thay đổi ngẫu nhiên tông màu.

Các tham số huấn luyện mô hình được cài đặt với hệ số học là 0.005;kích thước batch là 8, sử dụng kỹ thuật tối ưu Adam; hệ số điều chỉnh hệ số học đặt là 0.1; Các siêu tham số cho các hệ số hàm mất mát λ=0.3; γ=0.7; Quá trình huấn luyện và thử nghiệm được thực hiện máy tính với GeForce(R) GTX 2080, trên nền tảng ngôn ngữ Python với gói Pytorch.

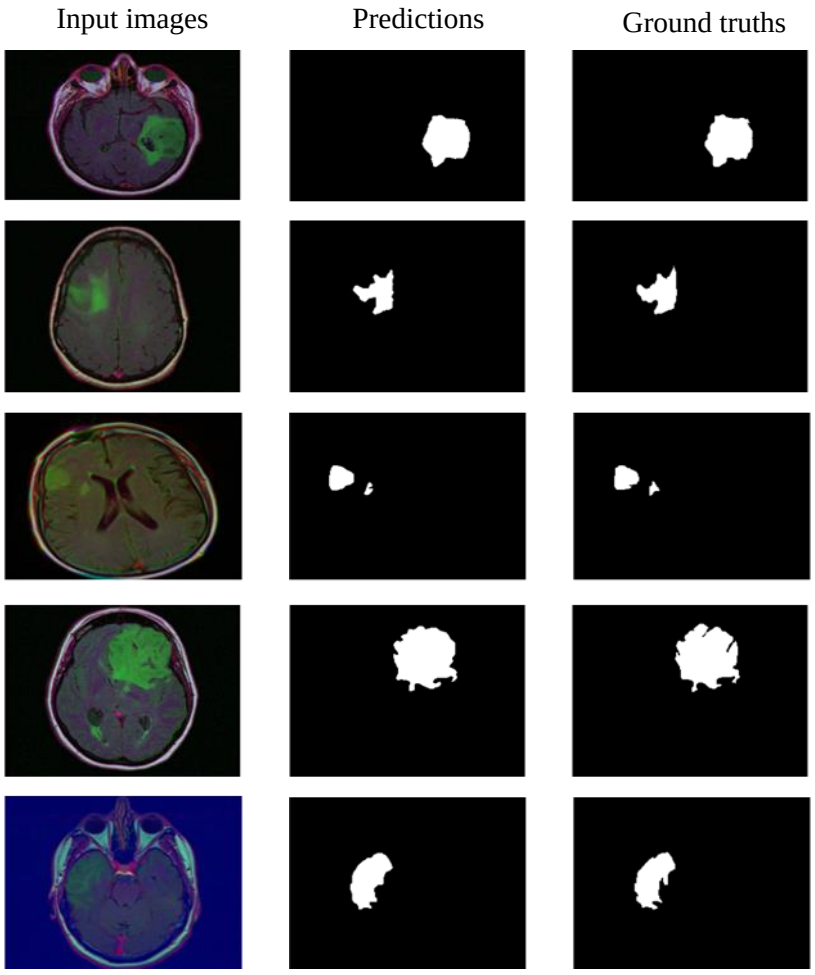
4.3. Kết quả của phương pháp đề xuất

Trước hết, để minh họa sự hội tụ của mô hình TransUnet sử dụng hàm mất mát đề xuất, chúng tôi vẽ đường cong học (learning curve) trên tập huấn luyện (train) và kiểm định (val) trên hình 4. Trên hình vẽ này ta thấy mô hình hội tụ sau khoảng 90 epoch. Các chỉ số ổn định với giá trị tương ứng lớn hơn 0.9 với thông số Dice, và 0.88 đối với thông số IoU; thể hiện hiệu năng của mô hình với hàm mất mát đề xuất.



(a) (b) (c)

Hình 4. *Đường cong học khi huấn luyện mô hình TransUnet dùng hàm mất mát đề xuất. (a) Giá trị mất mát loss; (b) Chỉ số Dice; (c) Chỉ số IoU.*



Hình 5. *Minh họa một số kết quả phân vùng ảnh (dự đoán) khối u não dùng phương pháp đề xuất. Cột 1: ảnh đầu vào (input images); Cột 2: Kết quả dự đoán (predictions); Cột 3: Kết quả chuẩn (ground truths).*

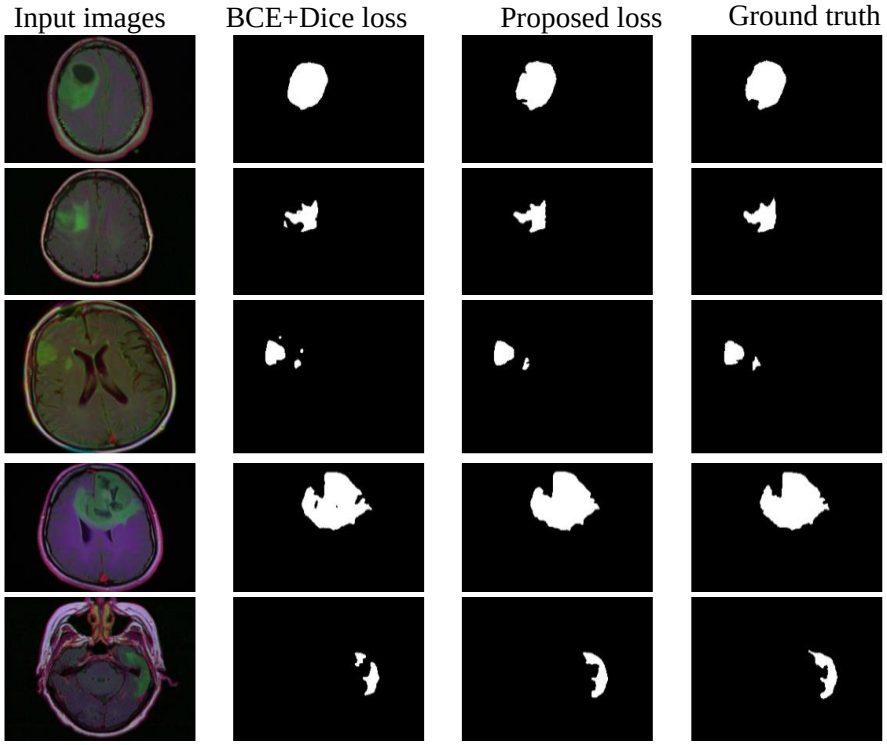
Một vài ảnh đại diện trên tập kiểm thử (test) được minh họa như hình 5. Trong đó, có thể thấy

là, ảnh ở hàng thứ 2, 3, và 5 có khối u khá nhỏ so với vùng nền- đây chính là một trong các thách thức của phân vùng ảnh khối u não. Như kết quả ở cột thứ hai trên hình này ta thấy, với kích thước cũng như các vị trí của khối u não khác nhau trong các bức ảnh, phương pháp đề xuất vẫn dự đoán được. Khi so sánh với hình ảnh chuẩn được xác định bằng tay bởi các chuyên gia (ground truths), ta thấy, kết quả dự đoán rất gần với các kết quả chuẩn này. Điều này thể hiện hiệu năng của phương pháp đề xuất cho bài toán phân vùng ảnh MRI khối u não.

4.4. Kết quả so sánh

4.4.1. So sánh với hàm mất mát BCE+Dice

Để đánh giá tính hiệu quả của hàm mất mát đề xuất, chúng tôi thực hiện huấn luyện mô hình TransUnet với hai hàm mất mát: BCE+Dice, và hàm mất mát BCE+Tversky đề xuất. Một số kết quả so sánh được biểu diễn trên hình 6. Trên hình này ta có thể thấy, hàm mất mát đề xuất cho kết quả dự đoán gần với ground truths hơn, thể hiện khả năng tốt hơn trong phân vùng khối u não.



Hình 6. *So sánh một số kết quả phân vùng ảnh khối u não bằng mạng TransUnet bằng hàm mất mát đề xuất với hàm mất mát BCE+Dice: Cột 1: ảnh đầu vào (input images); Cột 2: Kết quả dự đoán bằng hàm BCE+Dice; Cột 3: Kết quả dự đoán bằng hàm đề xuất; Cột 4: Khối u chuẩn khoanh bởi chuyên gia (ground truths).*

Bảng 1. *Giá trị theo chuẩn DSC (Dice similarity coefficient -DSC) và IoU (Intersection over Union) giữa kết quả dự đoán và ground truths trên tập test trên TransUnet khi huấn luyện bằng hàm BCE+Dice và hàm Tversky+BCE đề xuất.*

Loss function	**DSC**	**IoU**
BCE-Dice	0.90	0.87
Proposed	**0.91**	**0.88**

Để biểu diễn kết quả định lượng, chúng tôi đưa ra kết quả so sánh như trong bảng 1. Từ bảng này chúng ta thấy, với việc sử dụng hàm mất mát đề xuất, ta có thể tăng chỉ số DSC lên 0.01, từ

0.90 lên 0.91; và 0.87 lên 0.88 với chỉ số IoU. Điều này ta thấy tác dụng của hàm mất mát rất rõ ràng, nó không làm tăng số tham số huấn luyện của mạng, và phát sinh nhiều chi phí tính toán, trong khi có thể cải thiện đáng kể hiệu năng của bài toán phân vùng ảnh khối u não.

4.4.2. So sánh với các mô hình khác

Để minh họa ưu điểm của kiến trúc Transformer trong mô hình TransUnet, chúng tôi đã thực thi lại một số mô hình mạng, đó là UNet và Attention UNet. Kết quả so sánh gồm cả chỉ số DSC và IoU giữa các mô hình khác trên tập test được thể hiện trên bảng 2. Qua bảng này ta thấy, TransUnet cho kết quả cao hơn hai mô hình còn lại, đặc biệt là cả hai chỉ số đều cao hơn nhiều so với UNet.

Bảng 2. *So sánh các chỉ số DSC (Dice similarity coefficient) và IoU (Intersection over Union) trên tập test khi so sánh các mô hình Unet và Attention Unet với TransUnet khi huấn luyện bằng hàm mất mát đề xuất.*

Mô hình	DSC	IoU
UNet	0.84	0.73
Attention UNet	0.89	0.86
TransUnet	0.91	0.88

5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Trong bài báo này, chúng tôi đã đề xuất và áp dụng những kỹ thuật hiện đại nhất vào phân vùng ảnh y tế khối u não, sử dụng kiến trúc TransUNet. Chúng tôi cũng đề xuất một hàm mất mát phù hợp cho bài toán phân vùng khối u não, giúp giải quyết các thách thức của bài toán như sự khác nhau về kích thước, hình thái của khối u. Bài báo cũng so sánh mô hình đề xuất với một số mạng nơ-ron điển hình khác, từ đó đánh giá định tính cũng như định lượng của các chỉ tiêu phân vùng ảnh. Kết quả thực nghiệm trên một tập dữ liệu có nhãn gán bởi các chuyên gia cho kết quả tốt ở cả chỉ số Dice lẫn IoU. Điều này đã thể hiện ưu điểm của phương pháp đề xuất. Hàm mất mát này có thể mở rộng không chỉ trong phạm vi của phân vùng khối u não mà có thể ứng dụng trong các bài toán khác, nhất là khi có sự mất cân bằng dữ liệu, một trong các thách thức lớn của phân tích ảnh y tế.

Lời cảm ơn: *Nghiên cứu này được tài trợ bởi trường Đại học Bách khoa Hà Nội (HUST) trong đề tài mã số: T2021- PC-005.*

TÀI LIỆU THAM KHẢO

[1]. L. M. De Angelis, *"Brain Tumos,"* New England Journal of Medicine, vol. 344, pp. 114-123, 2001.
[2]. B. H. Menze, et al., *"The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) "* IEEE Trans Med Imaging, vol. 34, pp. 1993-2024, 2015.
[3]. J. Sachdeva, V. Kumar, I. Gupta, N. Khandelwal, and C. K. Ahuja, *"A novel content-based active contour model for brain tumor segmentation,"* Magnetic Resonance Imaging, vol. 30, pp. 694-715, 2012.
[4]. K. K. Shyu, V. T. Pham, T. T. Tran, and P. L. Lee, *"Unsupervised active contours driven by density distance and local fitting energy with applications to medical image segmentation,"* Mach. Vis. Appl., vol. 23, pp. 1159-1175, 2012.
[5]. M. Havaei, N. Guizard, H. Larochelle, and P. Jodoin, *"Deep Learning Trends for Focal Brain Pathology Segmentation in MRI,"* Machine Learning for Health Informatics pp. 125-148, 2016.
[6]. M. Buda, A. Saha, and M. A. Mazurowski, *"Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm,"* Comp. in Bio. and Med., vol. 109, pp. 218-225, 2019.
[7]. J. Zhang, J. Zeng, P. Qin, and L. Zhao, *"Brain tumor segmentation of multi-modality MR images via triple intersecting U-Nets,"* Neurocomputing, vol. 421, pp. 195-209, 2021.

[8]. Z. Liu, L. Chen, L. Tong, F. Zhou, Z. Jiang, and Q. Zhang, et al., *"Deep learning based brain tumor segmentation: A survey,"* arXiv:2007.09479, 2020, [online] p. Available: http://arxiv.org/abs/2007.09479, 2020.
[9]. J. Long, E. Shelhamer, and T. Darrell, *"Fully convolutional networks for semantic segmentation,"* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440, 2015.
[10]. O. Ronneberger, P. Fischer, and T. Brox, *"U-net: Convolutional networks for biomedical image segmentation,"* in Proceedings of the Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., 2015, pp. 234-241.
[11]. J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, et al., *"TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation,"* arXiv:2102.04306, 2021.
[12]. W. Chen, B. Liu, Peng. S., J. Sun, and X. Qiao, *"S3D-UNet: Separable 3D U-Net for Brain Tumor Segmentation,"* in Proceedings of the International MICCAI Brainlesion Workshop, 2018, pp. 358-368.
[13]. R. Mehta and J. Sivaswamy, *"M-net: A convolutional neural network for deep brain structure segmentation,"* in 2017 IEEE 14th International Symposium on Biomedical Imaging, 2017, pp. 18-21.
[14]. S. A. Taghanaki, Y. Zheng, S. K. Zhou, B. Georgescu, P. Sharma, D. Xu, et al., *"Combo loss: Handling input and output imbalance in multi-organ segmentation,"* Computerized Medical Imaging and Graphics, vol. 75, pp. 24-33, 2019.
[15]. S. S. M. Salehi, D. Erdogmus, and A. Gholipour, *"Tversky loss function for image segmentation using 3D fully convolutional deep networks,"* in Proceedings of the International Workshop on Machine Learning in Medical Imaging, 2017, pp. 379-387.
[16]. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., *"Attention is all you need,"* in Advances in neural information processing systems, 2017, pp. 5998–6008.
[17]. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al., *"An image is worth 16x16 words: Transformers for image recognition at scale,"* arXiv:2010.11929, 2020.
[18]. A. Tversky, *"Features of similarity,"* Psychol. Rev., vol. 84, p. 327, 1977.

ABSTRACT

Developing a loss function with TransUnet for brain tumor segmentation from MRI images

Segmentation of brain tumor in magnetic resonance images plays an important role in diagnosis and treatment planning for patients. However, brain tumor segmentation is a nontrivial task of the variations and differences in tumor sizes, topology, shapes, and the presence of intensity inhomogeneity. In this study, we proposed a new approach for brain tumor segmentation based on advances in deep neural networks. In particular, we propose using the TransUnet, a newly developed architecture based on Transformers and U-Net. In addition, we propose a new loss function to handle the size and shape variations of tumors. The approach is validated on the Brain LGG Segmentation. Experiments show performances of the proposed approach in comparison with other states of the arts.

Keywords: Deep neural networks; TransUnet; MRI Brain tumor segmentation; Tversky loss.