

ĐẠI HỌC BÁCH KHOA HÀ NỘI
TRƯỜNG ĐIỆN – ĐIỆN TỬ

NHẬP MÔN TRÍ TUỆ NHÂN TẠO
BÀI TOÁN PHÂN DÙNG DỰA TRÊN ẢNH
SIÊU ÂM BUỒNG TRÚNG

BÙI HOÀNG SƠN

Son.BH232278M@sis.hust.edu.vn

Thạc sỹ: Khoa học máy tính

Môi trường cảm thụ Đa phương tiện và Tương tác

Giảng viên hướng dẫn: TS. Trần Thế Hùng

HÀ NỘI, THÁNG 07 NĂM 2024

MỤC LỤC

DANH MỤC HÌNH ẢNH.....	ii
DANH MỤC BẢNG BIỂU	ii
TÓM TẮT	iii
1. CHƯƠNG 1: TỔNG QUAN – CƠ SỞ LÝ THUYẾT.....	1
1.1. Tổng quan	1
1.2. Bài toán phân vùng ảnh siêu âm buồng trứng.....	1
1.3. Neural Network	2
1.3.1. Convolutions Neural Networks.....	3
1.3.2. Các thành phần trong mạng CNNs	4
1.3.3. Deconvolution Neural Network	5
1.3.4. Học chuyển giao (Transfer Learning) với Pretrained Model	6
1.3.5. Receptive Field	7
1.4. Bài toán phân vùng ảnh	7
1.4.1. Khái niệm phân vùng ảnh và ứng dụng.....	7
1.4.2. Ứng dụng trong y tế	8
1.4.3. Các phép phân vùng phổ biến.....	9
1.5. Một số phương pháp phân vùng ảnh.....	11
1.5.1. Unet	11
1.5.2. TransUNet	11
2. CHƯƠNG 2. MÔ HÌNH PHÂN VÙNG KHỐI UBT.....	13
2.1. Tổng quan khung đề xuất mô hình.....	13
2.2. Chi tiết mô hình	13
2.3. Cơ chế Attention Gate.....	15
2.4. Khối SPPF (Spatial Pyramid Pooling Fast)	16
2.4.1. Khối SPP (Spatial Pyramid Pooling)	16
2.4.2. SPPF	16
2.5. Hàm mất mát	18
3. CHƯƠNG 3: KẾT QUẢ THỬ NGHIỆM ĐẠT ĐƯỢC	19
3.1. Cơ sở dữ liệu	19
PHẦN KẾT LUẬN.....	24
1. Kết quả đạt được	24
2. Hướng phát triển.....	24
DANH MỤC TÀI LIỆU THAM KHẢO.....	25

DANH MỤC HÌNH ẢNH

DANH MỤC BẢNG BIỂU

TÓM TẮT

Buồng trứng là một thành phần của hệ thống sinh sản nữ để sản xuất trứng, nhưng u nang buồng trứng có thể dẫn đến các biến chứng như xoắn, vô sinh và ung thư, do đó cần phải chẩn đoán nhanh chóng. Hình ảnh siêu âm thường được sử dụng để phát hiện u nang buồng trứng, nhưng việc phân chia các vùng u nang từ mô xung quanh đặt ra một thách thức do mô hình phức tạp. Hình ảnh siêu âm bao gồm các vấn đề như nhiễu lốm đốm, bóng, độ tương phản thấp và ranh giới mờ khiến nhiệm vụ trở nên phức tạp hơn. Mặt khác, khối u buồng trứng có rất nhiều loại, mỗi loại lại có hình dạng, kích thước và vị trí khác nhau nên càng làm trầm trọng thêm những thách thức này. Trong dự án này, đề xuất một mô hình dựa trên kiến trúc được đưa ra vào năm 2018 đó là Attention Unet với một số thay đổi về phần Encoder sử dụng VGG16 và thêm khối SPPF giải quyết vấn đề đa dạng về hình dáng và kích thước làm để tăng độ chính xác của mô hình cho việc phân vùng khối u buồng trứng qua ảnh siêu âm. Và cũng đưa ra một hàm mất mát mới kết hợp từ các hàm mất mát sẵn có giải quyết sự mất cân bằng dữ liệu và có hiệu năng tốt hơn so với các hàm mất mát cũ. Thông qua bộ dữ liệu OTU2D thì kết quả đạt được cho việc phân vùng qua ảnh siêu âm là 80,75% IOU và 89.14% DSC.

CHƯƠNG 1: TỔNG QUAN – CƠ SỞ LÝ THUYẾT

1.1. Tổng quan

Ung thư buồng trứng là một căn bệnh nguy hiểm nhất đối với phụ nữ hiện nay. Khối u buồng trứng còn gọi là u buồng trứng là một tình trạng y tế mà khối u ác tính hoặc lành tính phát triển trong buồng trứng. Khối u buồng trứng có thể xuất hiện ở mọi độ tuổi, nhưng nó thường xuất hiện ở phụ nữ trung niên hoặc phụ nữ cao tuổi. Có hai loại khối u chính. Thứ nhất là khối u ác tính hay còn gọi là ung thư buồng trứng. U buồng trứng ác tính có thể lây lan sang các cơ quan và mạch máu lân cận gây nguy hiểm cho tính mạng con người. Thứ hai là u lành tính. Nó gây ra các triệu chứng, và các vấn đề sức khỏe khác nhau. Hiện nay, ung thư buồng trứng ngày càng gia tăng. Việc phát hiện sớm khối u buồng trứng sẽ làm giảm tỉ lệ tử vong. Để xác định loại và kích thước khối u buồng trứng, các bác sĩ thường chẩn đoán qua ảnh siêu âm, hình ảnh chụp cộng hưởng từ (MRI) và xét nghiệm máu. Phát hiện khối u buồng trứng thông qua ảnh siêu âm là một phương pháp thông dụng và hiệu quả trong lĩnh vực y học. Hình ảnh siêu âm được ưa chuộng trong việc khám lâm sàng vì ưu điểm của nó là thời gian thực, không xâm lấn và không bức xạ. Nó đã trở thành một phương thức hình ảnh được phổ biến nhất trong thực hành lâm sàng. Hình ảnh siêu âm có thể giúp xác định có khối u buồng trứng hay không, đánh giá được kích thước, vị trí, và tính chất của nó. Quá trình phát hiện khối u buồng trứng qua ảnh siêu âm có thể cung cấp nhiều thông tin quan trọng quyết định về việc chẩn đoán và điều trị. Nếu phát hiện khối u buồng trứng, bệnh nhân sẽ được hướng dẫn tiếp theo bởi bác sĩ để phân loại u và quyết định lộ trình điều trị. Tuy nhiên do chất lượng hình ảnh kém có xuất hiện nhiều đốm và bóng nên việc giải thích hình ảnh siêu âm đòi hỏi phải có kỹ năng cao và có nhiều năm kinh nghiệm trong ngành. Do nhu cầu khám lâm sàng ngày càng tăng nên nhiều phương pháp phân tích, đánh giá hình ảnh tự động tiên tiến ra đời.

1.2. Bài toán phân vùng ảnh siêu âm buồng trứng

Ngày nay với sự phát triển không ngừng của trí tuệ nhân tạo ở mọi lĩnh vực trong cuộc sống, việc áp dụng trí tuệ nhân tạo vào lĩnh vực y tế sẽ giúp đỡ cho các y bác sĩ rất nhiều. Với sự gia tăng của căn bệnh ung thư buồng trứng trên thế giới. Thì việc phát hiện sớm khối u buồng trứng sẽ làm giảm tỉ lệ tử vong. Việc phát hiện sớm sẽ giúp bệnh nhân và các bác sĩ đưa ra phương pháp điều trị hiệu quả. Điều này có ý nghĩa cho việc bảo vệ buồng trứng. Vì vậy hiện nay nhiều cách tiếp cận được đưa ra để chẩn đoán và phát hiện khối u buồng trứng từ ảnh siêu âm thông qua các mô hình học sâu để giải quyết các vấn đề về u buồng trứng. Việc lựa chọn

hàm mất mát phù hợp sẽ đạt được hiệu quả cao trong việc huấn luyện mô hình. Từ đó đạt được kết quả cao giúp các bác sĩ trong việc chẩn đoán và phát hiện khối u buồng trứng. Một trong những mô hình phổ biến được áp dụng phổ biến hiện nay cho việc phân đoạn ảnh y tế là mô hình có kiến trúc U-net. Sử dụng các mô hình có kiến trúc U-net và so sánh các hàm mất mát để cho thấy được sự hiệu quả khi phân vùng khối u buồng trứng. Hình 1-1 là mô tả đầu vào và đầu ra của bài toán phân vùng khối u buồng trứng qua ảnh siêu âm. Với đầu vào là ảnh siêu âm 2D có khối u buồng trứng và đầu ra là ảnh nhị phân. Ảnh đầu ra phải xác định được điểm ảnh nào thuộc khối u và điểm ảnh nào không thuộc khối u

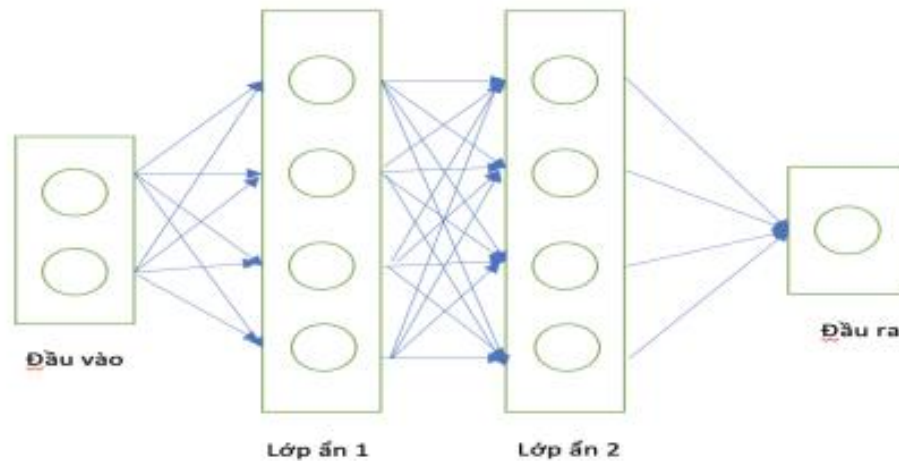


Hình 1-1: Minh họa đầu vào và đầu ra phân vùng khối u buồng trứng từ ảnh siêu âm

1.3. Neural Network

Học sâu (Deep learning) là một phần của lĩnh vực trí tuệ nhân tạo là một nhánh của Machine Learning tập trung vào việc xây dựng nơ-ron nhân tạo, cụ thể là mạng nơ-ron sâu, để thực hiện các nhiệm vụ thông minh như con người. Học sâu là một nhánh quan trọng của trí tuệ nhân tạo mang lại những lợi ích to lớn những năm gần đây ở nhiều lĩnh vực khác nhau như thị giác máy tính, xử lý ngôn ngữ tự nhiên, xử lý tiếng nói, điều khiển tự động và nhiều ứng dụng khác. Deep learning là một phương thức của Machine Learning, trong đó mạng nơ-ron nhân tạo được sử dụng để mô phỏng bộ não của con người. Mạng nơ-ron bao gồm nhiều lớp khác nhau, và số lượng lớp càng nhiều thì mạng càng sâu. Mỗi lớp chứa các nút mạng liên kết với các lớp khác. Mỗi liên kết giữa các nút có một trọng số tương ứng, trọng số này ảnh hưởng đến mạng nơ-ron theo mức độ của nó. Hình 1-2 là một hệ thống tính toán mô phỏng lại bộ não của con người gồm các lớp liên kết và nhiều kết nối với nhau. Mỗi nút trong mạng nơ-ron có một hàm kích hoạt, có nhiệm vụ chuyển đổi đầu ra từ nút đó. Dữ liệu được đưa vào mạng nơ-ron sẽ đi qua từng lớp và cuối cùng trả về kết quả đầu ra, được gọi output layer. Trong quá trình huấn luyện mô hình mạng nơ-ron, các trọng số được trao đổi và nhiệm vụ của mô hình là tìm ra được bộ trọng số phù sao cho phán đoán là tốt nhất. Các hệ thống Deep learning

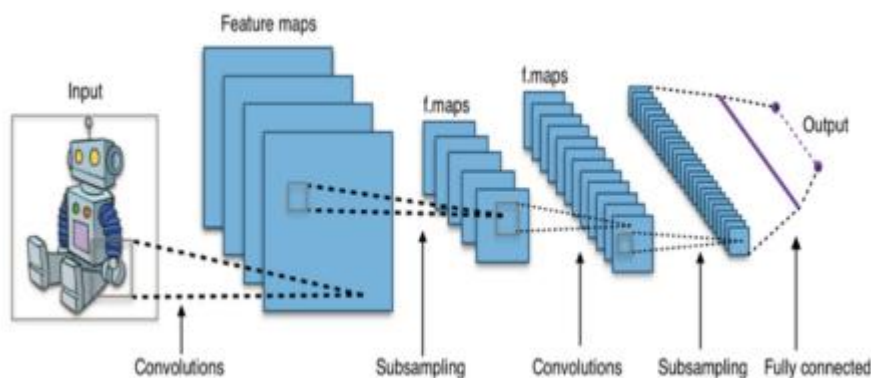
yêu cầu phần cứng rất mạnh để có thể xử lý được lượng dữ liệu lớn và thực hiện nhiều phép toán phức tạp. Nhiều mô hình Deeplearning có thể mất rất nhiều thời gian để triển khai trên những phần cứng tiên tiến nhất hiện nay. Đó cũng là một thách thức lớn của Deeplearning



Hình 1-2: Cấu trúc mạng neural

1.3.1. Convolutions Neural Networks

Mạng Convolution Neural Networks (CNN) là một mạng rất phổ biến trong lĩnh vực AI và Machine Learning. Nó có tạo ra những ứng dụng AI thông minh và có độ chính xác cao được sử dụng rộng rãi ở các công ty lớn trên thế giới. Đặc biệt CNN rất quan trọng trong việc xử lý hình ảnh, phân tích hình ảnh đóng vai trò quan trọng trong các nhiệm vụ về thị giác máy tính. Mạng thần kinh tích chập có một kiến trúc chuyên biệt dành cho việc xử lý hình ảnh. lớp tích chập (convolutional layer), lớp gộp (pooling layer), và lớp kết nối đầy đủ (fully connected layer) để học và biểu diễn các đặc trưng trong hình ảnh. Một hàm kích hoạt (activation function) (thường là ReLU) được áp dụng tới đầu ra của các lớp tích chập. Kết hợp với các dạng lớp khác nhau để giảm kích thước của đầu vào. Các lớp kết nối đầy đủ ở cuối có nhiệm vụ phân loại đầu ra thành các nhãn khác nhau.

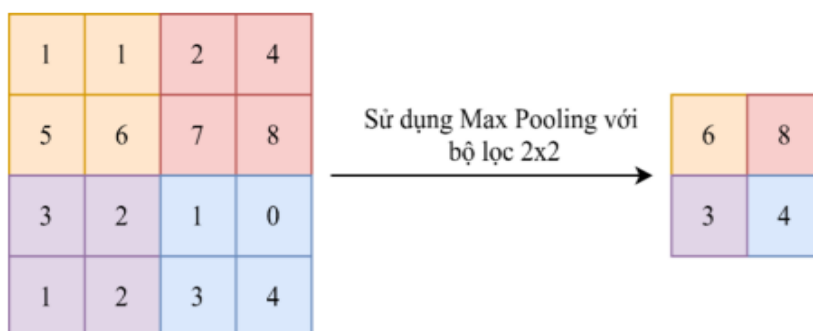


Hình 1-3: Convolutional Neural Network

1.3.2. Các thành phần trong mạng CNNs

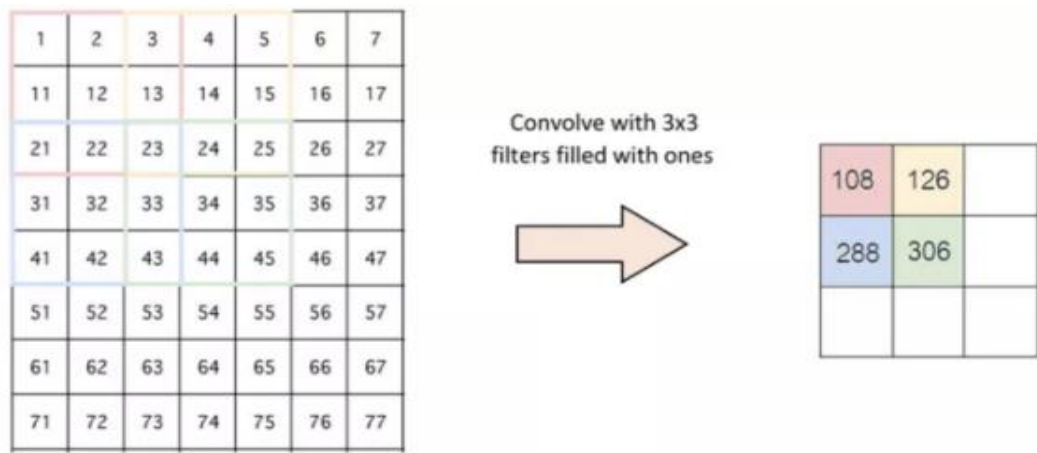
Lớp tích chập (Convolution layer): tích chập là một phép toán thực hiện đối với 2 ma trận đầu vào, kết quả đầu ra là một ma trận thứ 3. Một trong hai ma trận đầu vào của chúng ta ở đây chính là dữ liệu bức ảnh, còn ma trận thứ 2 là ma trận mà chúng ta sẽ phải tự định nghĩa, chúng còn được gọi là các bộ lọc. Tích chập là lớp đầu tiên để trích xuất các tính năng từ hình ảnh đầu vào. Tích chập duy trì mối quan hệ giữa các pixel bằng cách tìm hiểu các tính năng hình ảnh bằng cách sử dụng các ô vuông nhỏ của dữ liệu đầu vào.

Lớp gộp (Pooling layer): Lớp pooling làm giảm tham số khi hình ảnh quá lớn. Không gian pooling còn được gọi là lấy mẫu con hoặc lấy mẫu xuống làm giảm kích thước của mỗi map nhưng vẫn giữ được thông tin quan trọng. Các pooling có thể có nhiều loại như là maxpooling, sumpooling, average pooling



Hình 1-4: Minh họa lớp MaxPooling

Bước nhảy (stride): stride là ma trận số pixel thay đổi ma trận đầu vào. Khi stride bằng 1 thì ta di chuyển ma trận kernel 1 pixel. Khi tride bằng 2 ta di chuyển ma trận kernel đi 2 pixel và tiếp tục như vậy. Dưới đây là hình ảnh có stride bằng 2.



Hình 1-5: Minh họa bước nhảy (Stride)

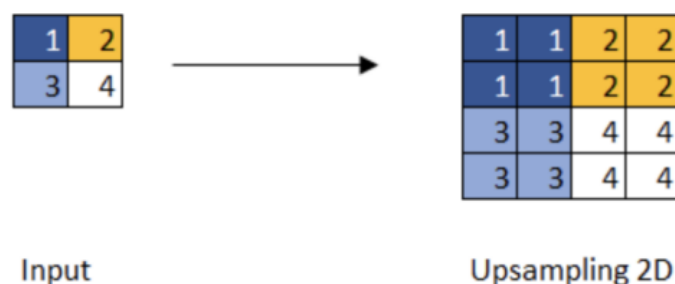
Lớp kết nối đầy đủ (Fully Connected layer): Tầng kết nối đầy đủ nhận đầu vào là các dữ liệu đã được làm phẳng, mà mỗi đầu vào đó được kết nối với tất cả các neuron. Trong đó mạng CNN thường được, các tầng kết nối đầy đủ thường được tìm ở cuối mạng và được dùng để tối ưu hoá mục tiêu của mạng ví dụ như độ chính xác của các lớp.

Lớp kích hoạt (Activation layer): Là một phần quan trọng trong kiến trúc của mạng neuron. Mỗi lớp kích hoạt nhận đầu vào từ lớp trước đó và thực hiện phép biến đổi theo hàm kích hoạt để tạo ra đầu ra của nó. Hàm kích hoạt giúp mô hình được các biểu diễn phi tuyến của dữ liệu, giúp nó có thể học được những đặc trưng phức tạp và mối quan hệ phi tuyến giữa các đặc trưng. Một số hàm kích hoạt thường được sử dụng như sigmoid, tanh, Relu,... Nếu như không có hàm kích hoạt thì mạng neuron sẽ biến thành một mạng khổng lồ với khối lượng tính toán phức tạp.

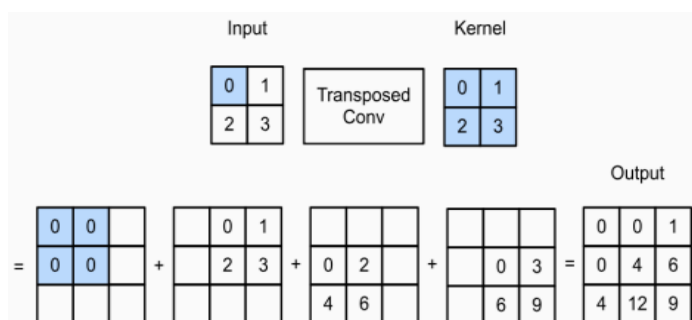
1.3.3. Deconvolution Neural Network

Mạng giải chấp được sử dụng khi đầu vào có kích thước ta muốn đầu ra có kích thước lớn hơn. Ngược lại với CNN, khi đầu vào có kích thước lớn hơn thì qua mạng CNN sẽ có kích thước đầu ra được thu nhỏ. Nó còn được gọi là tích chấp lên (Deconvolution) hoặc tích chấp chuyển vị (Transpose convolution).

- **Upsampling 2d layer:** Chức năng chính của lớp này là tăng kích thước không gian của đầu ra. Một số phương pháp sử dụng upsampling là Max-unpooling, phương pháp nội suy binnear, Bed of nails,...
- **Tích chấp chuyển vị (Transposed convolution):** Dùng trong mạng giải chấp để mapping lại các đặc trưng đã trích xuất ở lớp tích chấp để tạo ra các pixel ở đầu ra.



Hình 1-6: Minh họa UpSampling



Hình 1-7: Hoạt động của Transposed Convolution

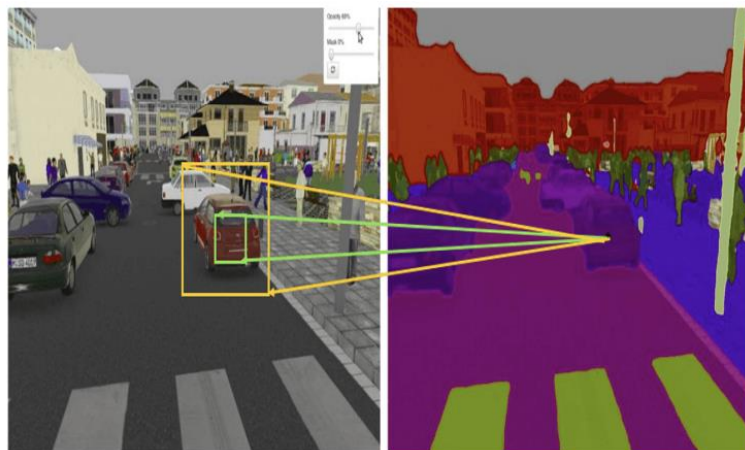
1.3.4. Học chuyển giao (Transfer Learning) với Pretrained Model

Transfer learning là một phương pháp học máy (Machine Learning) mà mô hình được huấn luyện trên tập dữ liệu có sẵn. Sau đó được chuyển giao và học nhiệm vụ mới. Phương pháp này được sử dụng để cải thiện hiệu suất mô hình và đảm bảo đúng đầu ra trên nhiệm vụ mới. Transfer learning được sử dụng rộng rãi trong nhiều lĩnh vực như thị giác máy tính, xử lý ngôn ngữ tự nhiên, và âm thanh. Như vậy với transfer learning thì chúng ta có thể xây dựng mô hình trong thời gian ngắn. Dưới đây là một số tác dụng chính của transfer learning:

- Transfer learning đặc biệt hiệu quả với các nhiệm vụ có tập dữ liệu nhỏ, bằng cách sử dụng các thông tin đã học được ở trên các tập dữ liệu lớn hơn để khái quát hoá các đặc trưng, đặc điểm của tập dữ liệu mới.
- Khi sử dụng transfer learning giúp chúng ta tiết kiệm được thời gian đào và chi phí tính toán của mô hình so với việc chúng ta huấn luyện lại mô hình từ ban đầu.
- Transfer learning còn giải quyết vấn đề mất cân bằng dữ liệu. Nếu tập dữ liệu có sự mất cân bằng giữa các lớp thì việc sử dụng transfer learning có hiệu quả cao vì nó được học các đặc trưng từ các tập dữ liệu lớn

1.3.5. Receptive Field

Receptive Field trong deep learning được định nghĩa là kích thước của một vùng trong không gian đầu vào được nhìn thấy bởi các pixel đầu ra qua một kernel. Không giống như mạng kết nối đầy đủ khi mà các node trong lớp phụ thuộc vào toàn bộ đầu vào của mạng thì đối với CNN nó chỉ phụ thuộc vào một vùng đầu vào. Vùng này được gọi là Receptive Field. Đối với bài toán phân vùng ảnh là việc dự đoán xem từng pixel tương ứng ở đầu vào thuộc lớp nào của ảnh đầu vào. Mà điều kiện lý tưởng nhất cho điều này là tại mỗi pixel đầu ra nó có một receptive field thật lớn. Điều này giúp các mô hình không bị bỏ qua những chi tiết quan trọng khi học. Tuy vậy thì không phải pixel nào trong receptive field đều có độ quan trọng như nhau đối với neuron mà nó ảnh hưởng tới. Các pixel gần trung tâm của receptive field sẽ đóng góp nhiều vai trò trong việc tính toán neuron tham chiếu hơn là các pixel ở rìa receptive field. Chúng ta cần receptive field đủ lớn để mà có thể nhận dạng vật thể lớn, có thể bắt được các đặc điểm tốt của vật thể cần dự đoán (ở hình 1.8 thì receptive field màu vàng sẽ tốt hơn receptive field màu xanh)



Hình 1-8: Minh họa mô tả Receptive Field

1.4. Bài toán phân vùng ảnh

1.4.1. Khái niệm phân vùng ảnh và ứng dụng

Phân vùng hình ảnh (image segmentation) là một trong những lĩnh vực chính của thị giác máy tính, được hỗ trợ bởi một lượng lớn nghiên cứu liên quan đến cả thuật toán xử lý hình ảnh và kỹ thuật học tập. Phân vùng hình ảnh đứng đằng sau các ứng dụng nổi bật như Robotics, Hình ảnh y tế, Xe tự lái và Phân tích video thông minh. Lĩnh vực này cũng được biết đến bởi lịch sử nghiên cứu, phát triển lâu dài, với các công trình đầu tiên ra mắt vào đầu năm 1970-1972. Mục đích của phân vùng hình ảnh là nhóm các vùng hoặc phân vùng một hình ảnh theo các nhãn lớp tương ứng. Tác vụ này tương đương với việc nhóm các điểm ảnh. Ngoài phân loại,

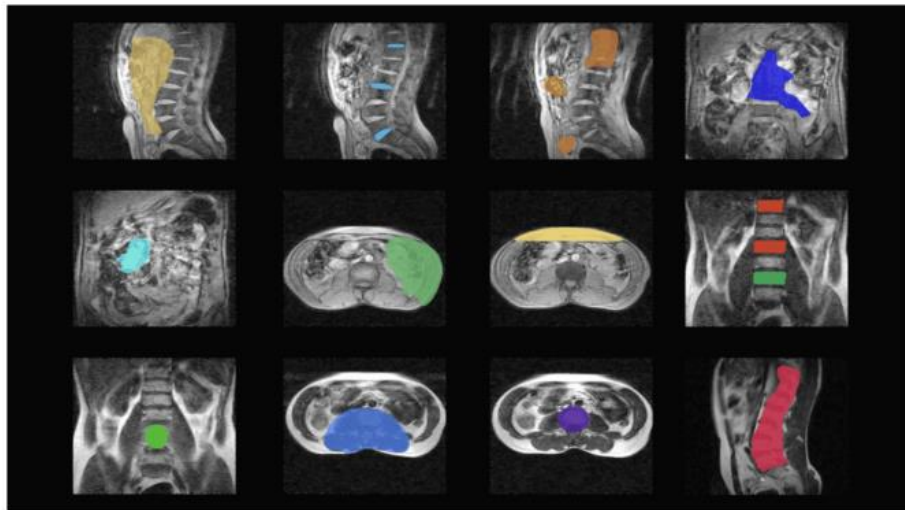
phân vùng hình ảnh cũng yêu cầu khoanh vùng (xác định vị trí chính xác của một đối tượng bằng cách xác định ranh giới của chúng). Do đó, có thể coi đây là bài toán mở rộng của phân loại hình ảnh (image classification). Các loại phân vùng hình ảnh

- **Semantic segmentation:** Semantic segmentation đề cập đến việc phân loại các điểm ảnh trong một hình ảnh thành các lớp ngữ nghĩa. Các điểm ảnh thuộc về một lớp cụ thể được phân loại một cách đơn giản vào lớp đó mà không cần xem xét thông tin hoặc bối cảnh nào khác.
- **Instance segmentation** việc xác định và tách các đối tượng riêng lẻ trong một hình ảnh, bao gồm cả việc phát hiện ranh giới của từng đối tượng và gán nhãn duy nhất cho từng đối tượng.
- **Panoptic segmentation:** Panoptic segmentation, tác vụ phân đoạn được phát triển gần đây nhất, là sự kết hợp giữa semantic segmentation và instance segmentation, trong đó ranh giới của mỗi đối tượng trong ảnh được tách biệt và danh tính của đối tượng được dự đoán.

1.4.2. Ứng dụng trong y tế

Thuật toán image segmentation có thể hỗ trợ các bác sĩ chẩn đoán những khối u từ ảnh x-quang, ảnh siêu âm. Ưu điểm segmentation không chỉ cho ta biết về vị trí của khối u mà còn cho ta biết về đặc điểm, kích thước của khối u

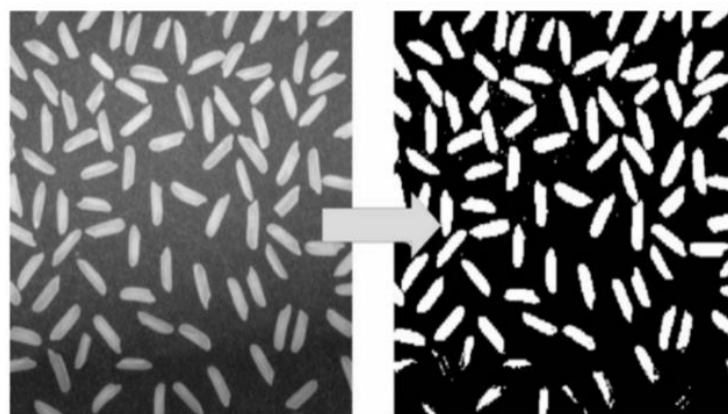
- **Nhận diện khối u:** Nhận diện khối u là một trong những ứng dụng phổ biến trong phân vùng ảnh y tế. Với sự giúp đỡ của phân vùng, bác sĩ có thể xác định được kích thước và vị trí khối u. Điều này rất quan trọng trong kế hoạch điều trị ung thư. Phân vùng ảnh y tế hoạt động bằng cách tách khối u ra khỏi các mô khỏe mạnh. Điều này cho phép bác sĩ hiểu rõ hơn về kích thước và hình dạng của khối u. Độ chính xác của phân vùng ảnh đối với việc nhận diện khối u cao hơn các phương pháp truyền thống.
- **Chẩn đoán bệnh:** Phân vùng hình ảnh để chẩn đoán nhiều loại bệnh, bao gồm bệnh tim mạch, bệnh phổi và bệnh xương. Ví dụ, trong bệnh tim mạch, phân vùng hình ảnh là để đo kích thước và hình dạng của tim và các buồng của nó. Thông tin này có thể được sử dụng để chẩn đoán các tình trạng như bệnh cơ tim và suy tim. Trong bệnh phổi, phân vùng hình ảnh có thể được sử dụng để đo thể tích của phổi và phát hiện những bất thường như khối u và nốt sần. Trong bệnh về xương, phân vùng hình ảnh có thể được sử dụng để đo mật độ xương và phát hiện gãy xương.



Hình 1-9: Ứng dụng ảnh y tế trong phân đoạn tổn thương

1.4.3. Các phép phân vùng phổ biến

- **Phân vùng dựa trên ngưỡng:** Phân vùng ngưỡng ảnh là một dạng phân vùng ảnh đơn giản, giúp tạo ra một hình ảnh nhị phân hoặc nhiều màu dựa trên việc đặt giá trị ngưỡng theo cường độ điểm ảnh của hình ảnh gốc. Mỗi giá trị điểm ảnh nằm trong khoảng 0-255, màu đen là 0, màu trắng là 255. Nếu giá trị điểm ảnh nhỏ hơn ngưỡng thì giá trị điểm ảnh được gán bằng 0 còn lớn hơn giá trị ngưỡng được gán giá trị 255. Do đó, hình ảnh được chuyển đổi thành ảnh nhị phân, dẫn đến quá trình thường được gọi là nhị phân hóa. Ngưỡng hình ảnh rất hữu ích trong trường hợp sự khác biệt về giá trị điểm ảnh giữa hai lớp mục tiêu là rất cao và dễ dàng chọn giá trị trung bình làm ngưỡng.



Hình 1-10: Phân vùng dựa trên ngưỡng

- **Phân vùng dựa theo cạnh:** Cạnh trong ảnh đánh dấu những vị trí hình ảnh không liên tục về mức xám, màu sắc, kết cấu, v.v. Khi di chuyển từ vùng này sang vùng khác, mức xám có thể thay đổi. Vì vậy, nếu tìm thấy sự gián đoạn đó, ta có thể tìm thấy cạnh. Thực tế, có nhiều toán tử phát hiện cạnh,

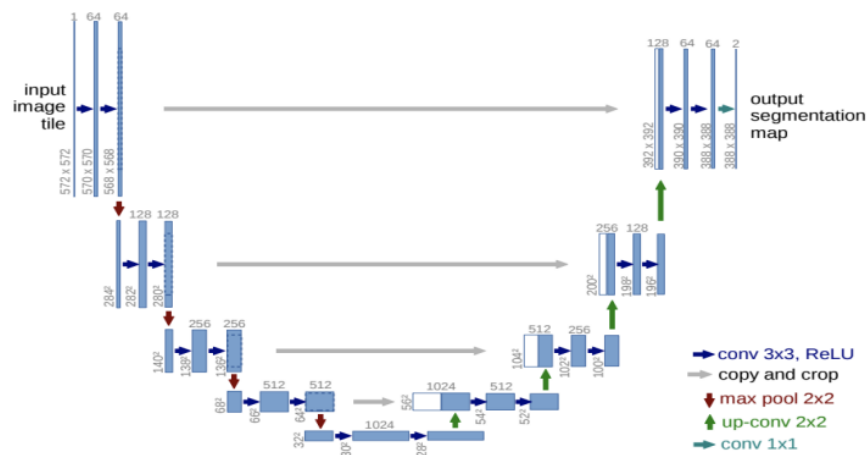
nhưng hình ảnh thu được là kết quả phân vùng trung gian, và không nên nhầm lẫn với hình ảnh được phân vùng cuối cùng. Để ra được kết quả cuối, cần thực hiện một số bước bổ sung bao gồm: kết hợp các phân vùng cạnh thu được làm một, để giảm số lượng phân vùng và có được một đường viền liền mạch của đối tượng. Cạnh tìm được thường có đặc trưng về "hướng" và "Độ lớn". Một số toán tử phát hiện cạnh cung cấp cả hai yếu tố này, chẳng hạn như Sobel edge operator, canny edge detector, Kirsch edge operator, Prewitt edge operator, Robert's edge operator, . . .

- ***Phân vùng dựa trên khu vực***: Một vùng có thể được phân loại là một nhóm các điểm ảnh kết nối với nhau và có các thuộc tính tương đồng về cường độ, màu sắc, v.v. Trong kiểu phân vùng này, có một số quy tắc được định sẵn mà pixel phải tuân theo để đảm bảo có thể phân loại thành các vùng điểm ảnh tương tự. Phương pháp phân vùng dựa trên khu vực được ưu tiên hơn phương pháp phân vùng dựa trên cạnh trong trường hợp ảnh bị nhiễu. Có hai nhóm kĩ thuật chính trong việc phân vùng dựa trên khu vực:
- ***Phát triển khu vực (Region growing method)***: với một số điểm ảnh làm điểm ảnh hạt giống và sau đó kiểm tra các điểm ảnh liền kề. Nếu các điểm ảnh liền kề tuân theo các quy tắc được xác định trước, thì điểm ảnh đó sẽ được thêm vào vùng của điểm ảnh gốc và quá trình sẽ tiếp tục cho đến khi không còn điểm tương đồng nào. Phương pháp này thực hiện theo cách tiếp cận từ dưới lên. Trong trường hợp khu vực đang phát triển, quy tắc ưu tiên có thể được đặt làm ngưỡng.
- ***Phân tách và hợp nhất khu vực (Region splitting and merging method)***: Đầu tiên ảnh được chụp thành một vùng. Nếu không thỏa mãn theo quy tắc được xác định trước, vùng sẽ được chia thành 4 phần và tiếp tục áp dụng quy tắc. Công việc lặp đi lặp lại cho đến khi không chia nhỏ thêm được nữa. Đối với hợp nhất khu vực, mỗi điểm ảnh được coi là một vùng riêng lẻ. Ta chọn một vùng làm vùng hạt giống để kiểm tra tính tương đồng của các vùng lân cận dựa trên quy tắc được định trước. Nếu giống nhau, chúng sẽ được hợp nhất thành một vùng duy nhất và cứ tiếp tục như vậy cho đến khi xây dựng các vùng được phân đoạn của toàn bộ hình ảnh
- ***Phân vùng dựa trên kỹ thuật phân cụm***: là một loại thuật toán học máy không giám sát, được sử dụng phổ biến trong phân vùng ảnh. Một trong những thuật toán Clustering thường được ứng dụng cho tác vụ phân vùng ảnh là KMeans Clustering. Loại phân cụm này có thể được sử dụng để tạo các phân đoạn trong một hình ảnh có màu

1.5. Một số phương pháp phân vùng ảnh

Cùng với sự phát triển của khoa học hiện đại ngày nay. Có rất nhiều phương pháp được đưa ra cho việc phân vùng ảnh y tế mang lại độ hiệu quả cao. Một cách tiếp cận cho việc phân vùng ảnh y tế được sử dụng rộng rãi hiện nay là học sâu. Thông qua các mô hình học sâu hiện đại ngày nay đã mang lại độ chính xác cao và hiệu quả trong việc phân đoạn ảnh y tế. Dưới đây là một số mô hình hiện đại đã được đề xuất cho việc phân vùng ảnh y tế. Các phương pháp này đều có thể áp dụng vào việc phân vùng cho khối u buồng trứng qua ảnh siêu âm.

1.5.1. Unet



Hình 1-11: Kiến trúc mạng UNet

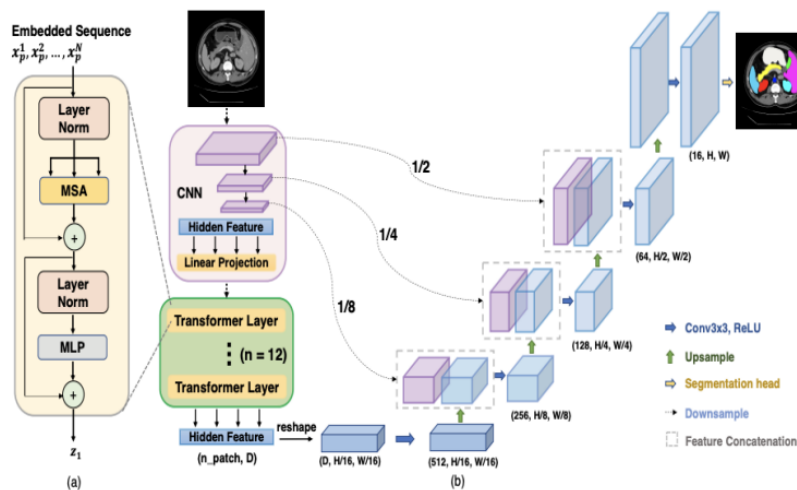
Unet được giới thiệu bởi Ronneberger và các cộng sự phát triển năm 2015 nhằm phân vùng các cấu trúc não thần kinh trong não người. Kiến trúc Unet có hình dạng chữ U bao gồm có hai phần: phần encoder nhiệm vụ mã hoá ảnh đầu vào và làm giảm kích thước của ảnh đầu vào thông qua các lớp tích chập và maxpooling để trích xuất những đặc trưng quan trọng của ảnh. Phần decoder gồm các lớp đối xứng với các lớp ở phần encoder. Quá trình áp dụng upsampling giúp kích thước các lớp tăng dần lên để tạo ra một ảnh phân vùng có kích thước như ảnh đầu vào. Ngoài ra Unet sử dụng các skip connection từ phần encoder sang phần decoder nhằm giữ lại nhiều đặc trưng trọng. Từ đó, Unet trở thành một kiến trúc phổ biến trong phân vùng ảnh đặc biệt là xử lý ảnh y tế.

1.5.2. TransUNet

TransUnet là một kiến trúc được đề xuất bởi Chen vào năm 2021, dựa trên kiến trúc của Transformer để phân vùng ảnh y tế. TransUnet được xây dựng trên kiến trúc của Unet có phần encoder là sự kết hợp của CNN và Transformer nhằm giải

quyết vấn đề của mạng tích chập truyền thống về xử lí ảnh đầu vào có kích thước lớn.

TransUnet bao gồm mô-đun Transformer, cơ chế attention và mạng nơ ron. Cơ chế attention nắm bắt các thông tin, sự phụ thuộc giữa các vị trí trong hình ảnh, thiết lập thông tin theo ngữ cảnh tổng thể biểu diễn trong đối tượng. Điều này TransUnet nắm bắt thông tin ngữ nghĩa trong hình ảnh, khái quát hình ảnh được tốt hơn. Đặc biệt TransUnet sử dụng CNN ở đầu tiên để trích xuất đặc trưng của ảnh đầu vào. Sau đó các đặc trưng này được chia nhỏ sau đó đưa vào khối có 12 lớp transformer. Sự kết hợp giữa CNN và transformer làm cho mô hình học được các thông tin cục bộ của ảnh đầu vào, hiệu quả hơn so với việc chỉ dùng các lớp transformer làm encoder. Phần decoder là sự kết hợp của lớp unsampling và các đặc trưng ở CNN để làm giàu thêm thông tin ngữ nghĩa. Ở bước cuối cùng là khôi phục các đặc trưng để thành bản đồ phân đoạn. Do TransUnet sử dụng 12 lớp transformer xếp chồng lên nhau nên số lượng tham số tăng làm cho mô hình huấn luyện lâu.



Hình 1-12: Kiến trúc mạng TransUnet

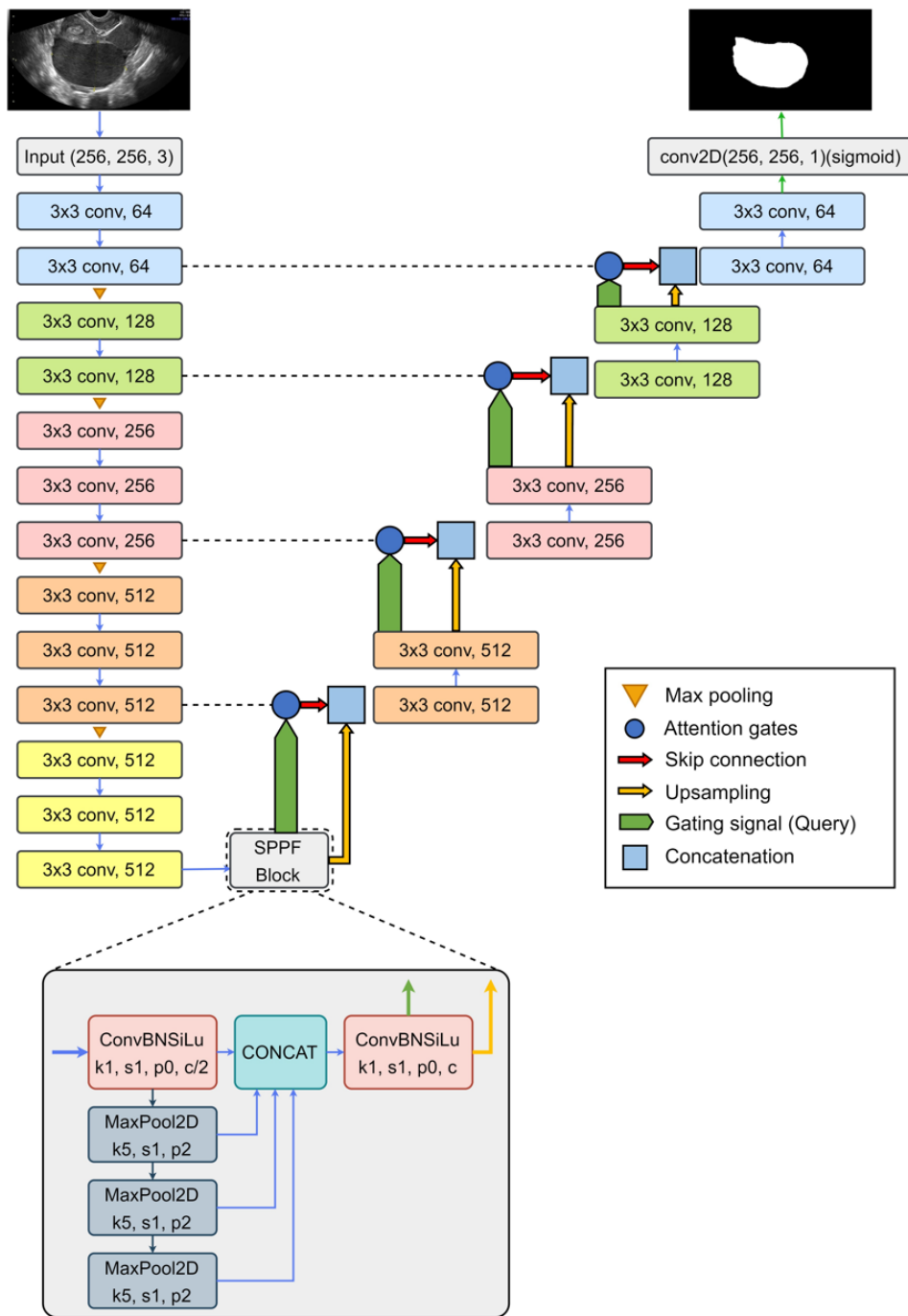
CHƯƠNG 2. MÔ HÌNH PHÂN VÙNG KHỐI UBT

2.1. Tổng quan khung đề xuất mô hình

Dựa kiến trúc của mô hình Attention Unet được công bố năm 2018 đã cho thấy được sự hiệu quả hơn mô hình Unet. Với mô hình Attention Unet tác giả Ozan Oktay và các cộng sự đề xuất cơ chế Attention Gate (AG) có khả năng tạo ra các trọng số attention cho mỗi đặc trưng ở vị trí không gian ảnh. Điều này giúp mô hình giúp mô hình quyết định các vùng quan trọng cần được tập trung hơn trong quá trình học và làm giảm nhiễu cho mô hình. Attention Unet giúp cải thiện thông tin từ phần encoder sang phần decoder làm cho thông tin ít bị mất mát, đặc biệt là các vùng biên giữa các cấu trúc. Phần encoder được thay thế bằng pretrained mô hình VGG16 để tăng độ hiệu quả trong quá trình học trong nhiệm vụ phân vùng khối u buồng trứng từ ảnh siêu âm. Ngoài ra mô hình được thêm một khối SPPF để truyền thông tin quan trọng từ encoder sang phần decoder được mô tả như hình 2-1, giúp cải thiện hiệu suất của mô hình. Các phần của mô hình sẽ giải quyết các vấn đề, thách thức mà liên quan tới việc phân vùng khối u buồng trứng qua ảnh siêu âm mà đã đề cập.

2.2. Chi tiết mô hình

Phần Encoder sử dụng mô hình pretrained VGG16 gồm các lớp như tích chập, Maxpooling và hàm kích hoạt Relu. Mô hình pretrained được sử dụng các lớp đầu tiên của mô hình và bỏ đi các lớp kết nối đầy đủ. Do được huấn luyện trên tập ImageNet có nhiều loại dữ liệu khác nhau nên mô hình học được nhiều đặc trưng đa dạng và tổng quát. Điều này giúp chúng hiệu quả trong việc học các tập dữ liệu mới. Với backbone VGG16 cho kết quả tốt hơn so với khối backbone của Attention Unet đã được chứng minh. Việc sử dụng mô hình pretrained giúp tiết kiệm thời gian và chi phí tính toán so với việc huấn luyện mô hình lại từ đầu. Việc huấn luyện lại từ đầu có thể gây tốn kém và đòi hỏi chi phí tính lớn. Ngoài ra việc sử dụng các mô hình pretrained ở các lớp đầu tiên của mô hình làm giảm rủi ro overfitting, đặc biệt với các tập dữ liệu nhỏ.



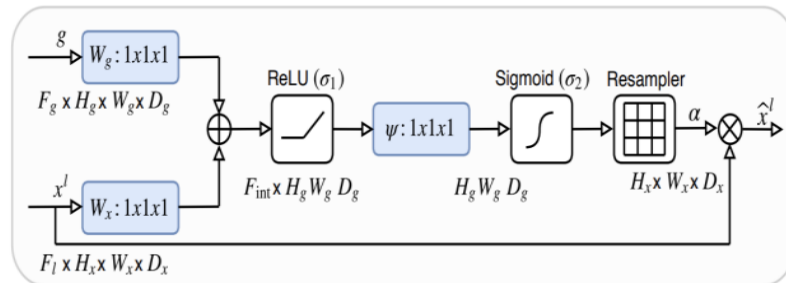
Hình 2-1: Kiến trúc mô hình

Phần decoder là sự kết hợp của các lớp tích chập, unsampling, batchnorm và hàm kích hoạt Relu gồm có 4 khối tương ứng với chiều sâu của phần encoder. Việc có lớp batchnorm khiến cho quá trình huấn luyện ổn định hơn, mô hình hội tụ nhanh hơn. Sử dụng batchnorm khiến các ma trận được khởi tạo dễ dàng hơn, giúp mô hình tránh các các vấn đề gradient vanishing và overfitting . Mô hình còn sử dụng skip connection giữa encoder và decoder thông qua các attention gate làm cho mô hình học được nhiều thông tin ngữ nghĩa, giữ được thông tin từ phần encoder sang

phần decoder. Phần dencoder làm tái tạo lại các đặc trưng của ảnh để đưa ra ảnh nhị phân ở đầu ra.

2.3. Cơ chế Attention Gate

Attention Gate thường được sử dụng trong việc xử lý ngôn ngữ tự nhiên. Vì đã cho thấy được hiệu quả trong xử lý ngôn ngữ tự nhiên nên cấu trúc Attention Gate được đưa vào kiến trúc Unet để cải thiện sự biểu diễn của mô hình. Hình 2.2 thể hiện cấu trúc của Attention Gate. Để nắm bắt được Receptive Field đủ lớn và thông tin ngữ nghĩa, ngữ cảnh phần encoder đã trích xuất đặc trưng giảm dần tỉ lệ chiều dài, chiều rộng của ảnh và tăng chiều sâu. Tuy nhiên vẫn còn khó khăn để duy trì làm giảm dự đoán false-positive đối với những đối tượng có kích thước nhỏ, thay đổi. Nhằm cải thiện độ chính xác thì Attention Gate được tích hợp vào kiến trúc Unet. Attention Gate nhận hai đầu vào chính : một đầu vào x là một đặc trưng từ nhánh encoder và một đầu vào g từ nhánh decoder gọi là Gating Signal. Sau đó cả hai đầu vào đi qua lớp tích chập để trích xuất đặc trưng và được cộng lại với nhau. Sau đó đi qua hàm kích hoạt ReLU tiếp tục thông qua một lớp tích chập và hàm sigmoid để tạo ra hệ số attention. Hệ số này thể hiện mức độ quan trọng của của hai đặc trưng đầu vào. Đầu ra được kết hợp bởi hệ số attention và đầu vào của phần encoder.

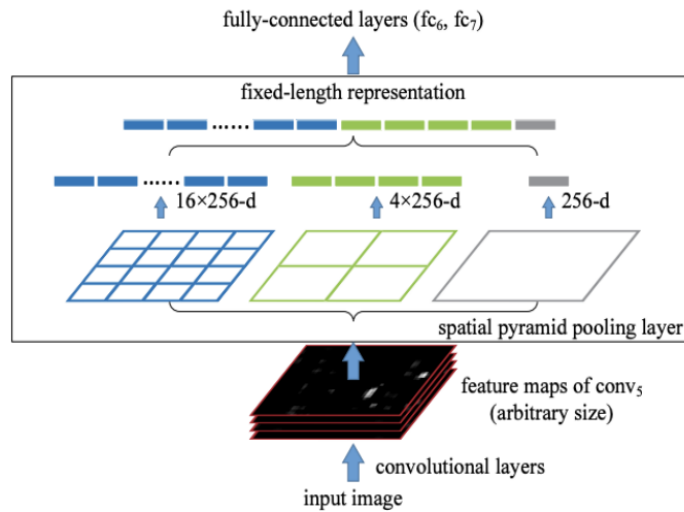


Hình 2-2: Cấu trúc Attention Gate

Khi thêm attention gate vào kiến trúc Unet sẽ giúp mô hình tập trung vào vùng quan trọng của ảnh, cải thiện khả năng biểu diễn giúp mô hình học cách tổ chức thông tin linh hoạt và hiệu quả hơn. Attention gate giúp cho mô hình hiểu được những đặc trưng phức tạp của dữ liệu đầu vào. Ngoài ra còn làm giảm ảnh hưởng nhiễu giúp mô hình tập trung vào thông tin quan trọng theo như bài báo Attention Unet đã phân tích.

2.4. Khối SPPF (Spatial Pyramid Pooling Fast)

2.4.1. Khối SPP (Spatial Pyramid Pooling)



Hình 2-3: Kiến trúc lớp SPP (Spatial Pyramid Pooling)

Hình 2.3 là lớp SPP (Spatial Pyramid Pooling) được công bố năm 2014 để phục vụ cho việc nhận diện vật thể. SPP cũng đã chứng minh được hiệu năng biểu diễn trong các nhiệm vụ của thị giác máy tính như nhận diện đối tượng và phân loại. Cách hoạt động của SPP với đầu ra là lớp kết nối đầy đủ là mỗi channel trong phần trích xuất đặc trưng của ảnh có kích thước ($H \times W \times 256$) sẽ được Maxpooling với các kernel khác nhau sẽ tạo ra các feature map. Sau đó các feature map được làm phẳng thành các vector và được đưa vào lớp kết nối đầy đủ (fully connected) để nhận diện vật thể. Thông thường các lớp kết nối đầy đủ có kích thước đầu vào cần được giữ nguyên. Nhưng với sự xuất hiện của SPP thì kích thước đầu vào sẽ có kích thước đa dạng hơn, một vector cố định vẫn được sinh ra sau khi đi qua SPP. Hơn nữa SPP sử dụng các kernel có kích thước khác nhau làm tăng Receptive Field. SPP không chỉ giúp giữ lại thông tin từ các vùng không gian quan trọng của hình ảnh, mà còn giảm bớt số lượng tham số cần học, giảm độ phức tạp của mô hình và tăng cường khả năng nhận biết đối với hình ảnh có kích thước khác nhau. SPP có thể tích hợp vào CNN để nâng cao hiệu suất của mô hình.

2.4.2. SPPF

Đầu vào của SPPF có kích thước là $W \times H \times C$ sẽ đi vào khối có lớp tích chập, Batchnorm, hàm kích hoạt SiLU có kernel là 1, Stride = 1, padding = 0, filters = $C/2$. Sau đó lần lượt đi vào ba lớp Maxpooling2D có kernel=5, stride=1, padding=2. Các đầu ra được nối lại với nhau sau đi tiếp tục đi qua khối tích chập, batchnorm, SiLU để cho ra đầu ra của khối có kích thước giống như kích thước đầu vào. Với hàm kích hoạt SiLU được biểu diễn như sau:

$$SiLu(x) = x \cdot \sigma(x)$$

Trong đó : $\sigma(x)$ là hàm sigmoid và x là đầu vào của hàm

SPPF (hình 2.4) được đưa ra trong kiến trúc của yolov5 giống như SPP nhưng nhanh hơn gấp đôi so với SPP. Thay vì sử dụng các lớp Maxpooling song song thì SPPF sử dụng các lớp maxpooling tuần tự. Hơn nữa thay vì sử dụng các kernel size khác nhau thì SPPF sử dụng tất cả các kernel bằng 5. Điều này cho phép SPPF có thời gian thực thi ngắn hơn. Bằng cách này SPPF không chỉ làm giảm số lượng tham số của mô hình mà còn làm giảm độ phức tạp tính toán và cải thiện hiệu quả tính toán của mô hình. Việc sử dụng các lớp maxpooling tuần tự làm tăng tính bất biến của mô hình. Maxpooling chọn giá trị lớn nhất của mỗi vùng giúp giữ lại những đặc trưng quan trọng và làm cho mô hình ít nhạy cảm hơn đối với sự thay đổi của đối tượng trong ảnh. SPPF có thể nắm bắt các đặc điểm ở quy mô toàn cầu, bán cục bộ và cục bộ, cho phép mô hình hiểu rõ hơn về cấu trúc không gian và mối quan hệ theo ngữ cảnh của các vật thể ở các quy mô, không gian khác nhau. Việc tích hợp khối SPPF vào mô hình attention Unet sẽ giúp mô hình học được nhiều đặc trưng hơn của vật thể và thích ứng một cách tốt hơn với các đối tượng ở tỉ lệ khác nhau và cải thiện hơn trước những thay đổi về tỉ lệ của đối tượng. Ngoài ra SPPF sẽ là cầu nối quan trọng để truyền thông tin từ phần encoder sang phần decoder mà sẽ làm giảm đi sự mất mát thông tin của đối tượng.

Nhận xét

Mô hình đề xuất Attention-Unet với pretrained VGG16 và khối SPPF có những điểm mạnh nhất định trong việc phân vùng khối u buồng trứng qua ảnh siêu âm. Với vấn đề có các loại u khác nhau, với các kích thước khác nhau của ung thư buồng trứng dẫn đến sự mất cân bằng dữ liệu. Sử dụng pretrained mô hình VGG16 mang lại hiệu quả tốt cho việc phân vùng. Vì mô hình được những đặc tổng quát khi pretrained mô hình được huấn luyện trên tập dữ liệu lớn. Với sự tích hợp của cơ chế attention và khối SPPF làm cho mô hình tập trung vào những vùng ảnh quan trọng và loại bỏ đi những vấn đề của ảnh siêu âm như nhiễu, đốm, độ tương phản thấp, bóng. Ngoài ra, cơ chế attention làm giảm mất mát thông tin từ phần encoder sang decoder để tái tạo đặc trưng, điều này rất quan trọng với bài toán phân vùng. Bằng cách ghép các đặc trưng ở tỉ lệ khác nhau, SPPF cho phép nắm bắt được các đặc điểm ở nhiều mức độ. Điều này giúp mô hình học được những đặc trưng quan trọng của khối u ở nhiều kích thước khác nhau.

2.5. Hàm mất mát

Focal loss được giới thiệu vào năm 2017 là hàm mất mát đặc biệt để giải quyết vấn đề của mô hình khi đối mặt với vấn đề mất cân bằng dữ liệu giữa các lớp trong nhiệm vụ phân loại. Đặc biệt Focal loss hạn chế ảnh hưởng đến các mẫu dễ học và tập trung vào các mẫu khó học.

$$FocalLoss = -\alpha \cdot (1 - p)^{\gamma} \cdot \log(p)$$

SSIM loss (Structural Similarity Index) : SSIM được biết đến là một phương thức đánh giá sự tương đồng giữa hai ảnh. SSIM đo lường ba yếu tố chính đó là độ tương phản, độ đồng nhất và độ giả định. Về sau nhiều bài báo đã sử dụng SSIM là hàm mất mát cho bài toán phân vùng.

$$SSIMLoss(X, Y) = 1 - \frac{(2\mu_X\mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)}$$

IOULoss hay thường được gọi là Jacard Loss là một hàm mất mát được dùng phổ biến cho bài toán phân vùng ảnh. Jacard đo lường sự tương đồng của kết quả dự đoán và mục tiêu thực tế

$$JacardLoss = 1 - \frac{y * \hat{y}}{(y + \hat{y} - y * \hat{y})}$$

JointLoss: Dựa các hàm mất mát đã có, đề xuất một hàm mất mát mới có độ hiệu quả cao hơn so với các hàm mất mát đã được nêu phía trên. JointLoss là hàm mất mát được kết từ 3 hàm mất mát khác là Focal Loss, SSIM loss, Jacard loss với FocalLoss giải quyết vấn đề mất cân bằng dữ liệu, SSIM loss đo sự tương đồng về cấu trúc. JointLoss có công thức dưới đây:

$$JointLoss = \omega_1 \cdot FocalLoss + \omega_2 \cdot JacardLoss + \omega_3 \cdot SSIMLoss$$

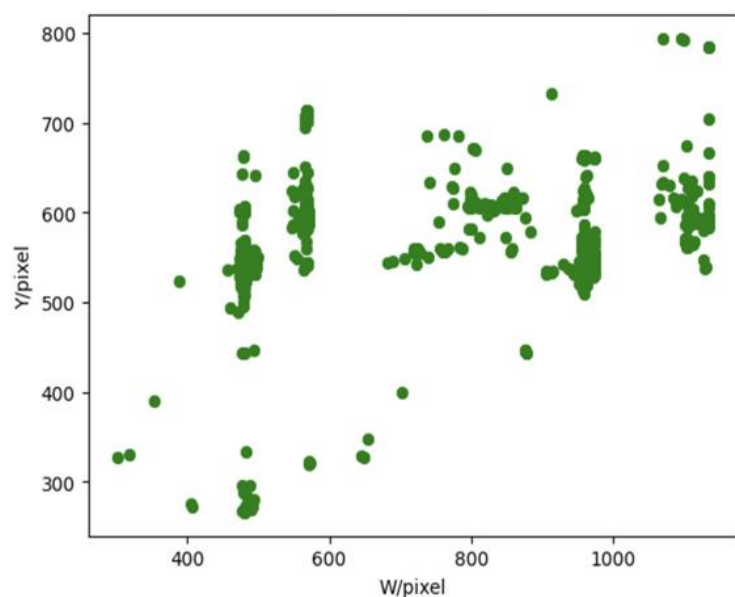
3. CHƯƠNG 3: KẾT QUẢ THỬ NGHIỆM ĐẠT ĐƯỢC

3.1. Cơ sở dữ liệu

Cơ sở dữ liệu được sử dụng trong bài toán là tập OTU2D ở trong tập dữ liệu MultiModality Ovarian Tumor Ultrasound (MMOTU) được đưa ra bởi bệnh viện Beijing Shijitan, Capital Medical University được lựa chọn từ 294 bệnh nhân. Tập dữ liệu MMOTU bao gồm bộ dữ liệu OTU2D bao gồm 1469 ảnh siêu âm 2D và bộ dữ liệu OTU3D gồm 170 ảnh siêu âm 3D có 8 loại u buồng trứng khác nhau. Số ảnh của các loại u trong tập dữ liệu không cân bằng. Điều này chủ yếu do một số loại u phổ biến hơn và một số loại u hiếm gặp hơn nên có sự chênh lệch về số lượng. Chiều rộng và chiều cao của ảnh trong tập OTU2D nằm trong khoảng từ 302-1135 và 226-794 pixel như hình 3.1. là hình ảnh gốc và nhãn của khối u buồng trứng theo từng lớp tương ứng với thứ tự của bảng 3.1. Ở trong tập dữ liệu OTU2D việc gán nhãn dữ liệu khối u buồng trứng được thực hiện bởi 27 chuyên gia. Mỗi ảnh được gán nhãn bởi một chuyên gia sau đó được kiểm tra bởi một chuyên gia khác để đảm bảo cho chất lượng của việc gán nhãn. Trong quá trình gán nhãn các chuyên gia tham khảo các báo về bệnh lý nên việc gán nhãn trở nên chính xác và thuyết phục hơn siêu âm trong tập OTU2D và OTU3D của tập MMOTU. Qua đó ta thấy được sự mất cân bằng giữa các lớp của bộ dữ liệu.

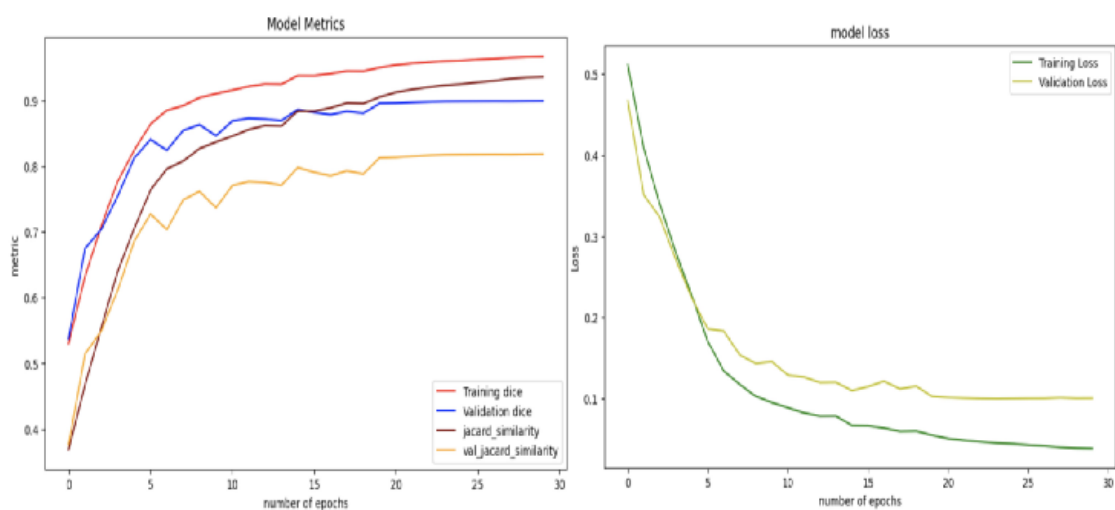
Bảng 3-1: Số lượng ảnh có trong mỗi lớp của CSDL MMOTU

Lớp	Số lượng ảnh	
	OTU_2D	OTU_3D
Chocolate cyst	336	35
Serous cystadenoma	219	40
Teratoma	336	32
Theca cell tumor	88	31
Simple cysts	66	18
Ovary normal	267	0
Mucinous cystadenoma	104	5
High-grade serous cystadenoma	53	9

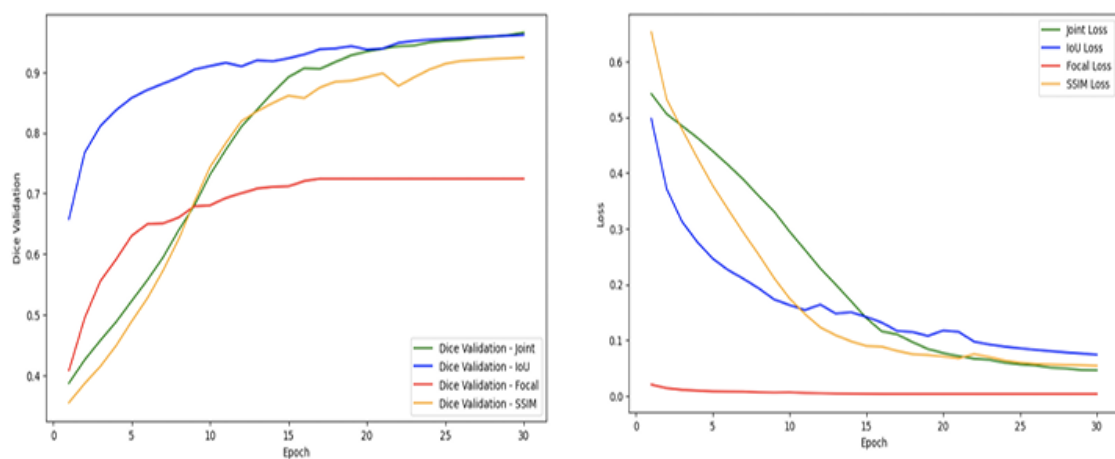


Hình 3-1: Phân số kích thước ảnh trong bộ CSDL OTU2D

3.2. Kết quả thử nghiệm



Hình 3-2: Đồ thị độ đo Dice Coefficient và Loss khi traning



Hình 3-3: Đồ thị trực quan Dice_Coeff và Loss khi Validation

Sau khi huấn luyện các mô hình để so sánh với mô hình đề xuất trên tập dữ liệu OTU2D thì kết quả đạt được được biểu diễn như bảng 3.2 thông qua các phương thức đánh giá DSC(%), IOU(%), Recall(%), Precision(%) và hàm mất mát mà em đã đề xuất jointloss. Qua đó ta thấy được sự hiệu quả khi áp dụng pretrained VGG16 vào Attention Unet với nhiệm vụ phân vùng ảnh khối u buồng trứng qua ảnh siêu âm. Khi thay đổi phần encoder bằng pretrained VGG16 thì DSC tăng gần 5(%), IoU tăng 7.33(%) so với mô hình Attention Unet. Cuối cùng là mô hình em đề xuất khi kết hợp Attention Unet với pretrained VGG16 và khối SPPF cho thấy được hiệu quả cao đối với nhiệm vụ được đặt ra. Khi thêm khối SPPF thì DSC tăng 1.38(%) và IOU tăng 2,2(%). Từ bảng so sánh ta thấy được hiệu quả của mô hình đề xuất mang lại khi so sánh sự hiệu quả phân vùng khối u buồng với các mô hình hiện đại khác.

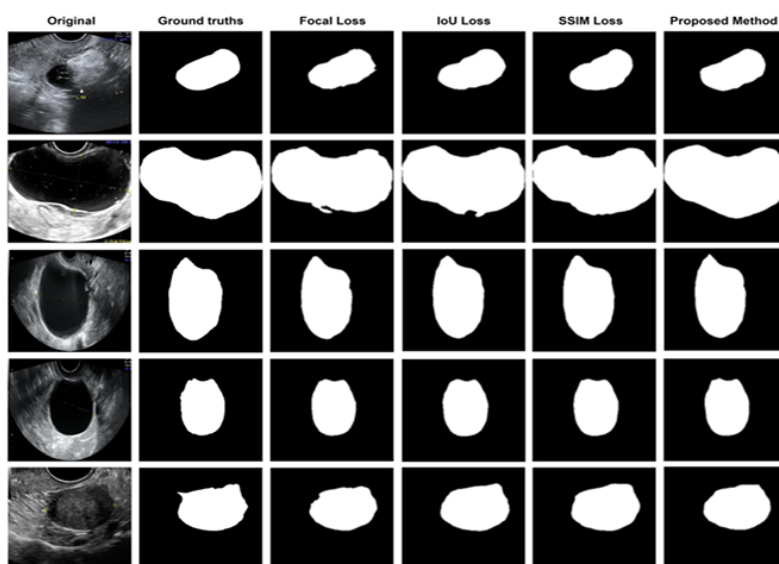
Bảng 3-2: So sánh thông số giữa các mô hình

Model	DSC	IoU	Recall	Precision
MobileNetV2 UNet [19]	80.32	67.56	77.39	85.78
DANet [33]	87.75	78.50	87.90	88.50
Attention UNet [1]	82.97	71.22	80.99	85.89
VGG16_Attention_Unet	87.76	78.55	88.55	88.83
Proposed Method	89.14	80.75	88.50	89.87

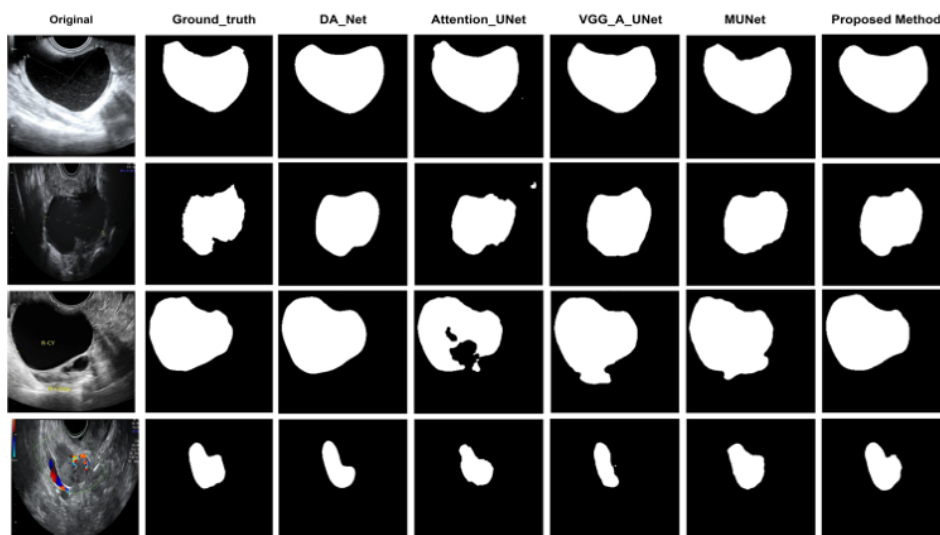
Qua bảng 3.3 đã cho thấy rằng FocalLoss đạt được kết quả tốt nhất về Precision. Tuy nhiên FocalLoss sẽ bỏ lỡ nhiều phân đoạn của khối u vì có Recall khá thấp. Trái với FocalLoss, các hàm mất mát SSIMLoss, IOULoss lại có Recall cao hơn nhưng nó sẽ tạo ra các false-positive nhiều hơn, điều đó giải thích tại sao Precision thấp hơn. Bởi vậy sự kết hợp của FocalLoss, IOULoss, SSIMLoss sẽ nâng cao hiệu quả cho việc phân vùng. Ta thấy được hiệu quả của Joint Loss được cải thiện so với các hàm mất mát khác hay được sử dụng trong việc phân vùng ảnh. Với DSC là 89.14(%) tăng 1.23(%) so với IoULoss và gần 2(%) so với DiceLoss, tăng 1% so với SSIM loss, IoULoss và DiceLoss là hai hàm mất mát hay được sử dụng trong phân vùng ảnh. Với IoU của hàm mất mát đề xuất là 80.75(%) cao hơn IOULoss gần 2(%), gần 3(%) so với DiceLoss và gần 2% so với SSIMLoss . ReCall đạt được là 88.5(%) trong khi đó Precision là 89.87(%) kém 1% so với FocalLoss. Qua đó cho thấy được hiệu quả của hàm mất mát mà em đề xuất cho mô hình Attention Unet với pretrained VGG16 và khối SPPF

Bảng 3-3: Kiểm thử mô hình đề xuất với từng hàm mất mát

Model	Loss	DSC	IoU	Recall	Precision
Model Proposed	BCE	86.68	76.86	87.50	88.57
	Dice	87.27	77.76	88.19	87.63
	Focal	72.43	57.03	84.62	90.98
	SSIM	86.8	77.03	88.40	87.33
	IoU	88.01	78.89	89.20	87.28
	Joint	89.14	80.75	88.50	89.87



Hình 3-4: Kết quả đánh giá trên từng hàm mất mát



Hình 3-5: Kết quả kiểm thử với các mô hình

Hình 3.4 là biểu diễn kết quả phân vùng của mô hình em đề xuất với các hàm mất mát khác nhau. Qua kết quả được biểu diễn ta thấy đường bo của hàm mất mát đề xuất JointLoss giống nhãn (Label) hơn so với các hàm mất mát khác thường dùng như FocalLoss, IOULoss (Jaccard Loss) và SSIM Loss. Điều này chứng tỏ với

JointLoss thì mô hình có thể bắt được như đặc trưng quan trọng và các đặc trưng ở viền bo của khối u buồng trứng.

Hình 3.5 là kết quả phân vùng khối u buồng trứng trên các mô hình khác nhau với hàm mất mát là JointLoss. So với Attention Unet thì khi sử dụng pretrained mô hình VGG16 sẽ cho kết quả tốt hơn. Tuy nhiên các điểm ảnh mà có sử dụng pretrained VGG16 nhận nhầm nhiều hơn còn Attention Unet nhận nhầm ít hơn. Danet một mô hình ít được sử dụng trong phân vùng ảnh y tế cũng mang lại một kết quả tốt. Đối với mô hình đề xuất khi kết hợp khối SPPF và pretrained VGG16 cho thấy được kết quả tốt hơn so với các mô hình khác, đường bo của khối u cũng giống với nhãn (ground truth) hơn các mô hình khác.

PHẦN KẾT LUẬN

1. Kết quả đạt được

Với mục tiêu phân vùng khối u buồng trứng qua ảnh siêu âm sử dụng trí tuệ nhân tạo. Em đã phân tích một số mô hình ứng dụng trong phân vùng ảnh y tế cũng như phân vùng ảnh khối u buồng trứng. Sau đó, đề xuất ra mô hình phân vùng u buồng trứng từ ảnh siêu âm dựa trên sự kết hợp của Attention Unet với pretrained VGG16 và khối SPPF cho việc phân vùng khối u buồng trứng qua ảnh siêu âm. Mô hình đề xuất có thể học được những đặc trưng phức tạp và quan trọng để từ đó nâng cao được độ chính xác. Với cơ chế attention và khối SPPF tập trung vào những vùng ảnh quan trọng và làm giảm nhiễu của ảnh siêu âm. Ngoài ra khối SPPF còn thể hiện hiệu quả đối với những khối u hay đổi về kích thước và hình dạng. Ngoài ra em còn đề xuất một hàm mất mát JointLoss là sự kết hợp của các hàm mất mát sẵn là IOULoss (Jaccard Loss), SSIMLoss, FocalLoss để nâng cao hiệu suất của mô hình giải quyết vấn đề mất cân bằng dữ liệu. JointLoss đã cho thấy được sự hiệu quả khi so sánh với các hàm mất mát phổ biến được sử dụng trong phân vùng ảnh như Dice Loss, IOU Loss, Focal Loss, BCE Loss. Kết quả thử nghiệm trên tập dữ liệu OTU2D cho thấy mô hình đề xuất cho kết quả tốt hơn so với các mô hình khác như Attention Unet, DAnet, MU-net. Trên tập dữ liệu OTU2D, mô hình đề xuất đạt kết quả lần lượt là 89.14, 80.75, 88.5 và 89.87 trên các độ đo DSC, IoU, Recall và Precision.

2. Hướng phát triển

Hoàn thành việc phân vùng cho khối u buồng trứng qua ảnh siêu âm. Kết quả phân vùng đã được so sánh với các mô hình khác trên cùng tập dữ liệu và đạt được những kết quả tích cực. Tuy nhiên nó vẫn còn nhiều khó khăn và thử thách. Dữ liệu ảnh về khối u đang còn ít, sự nhiễu, đốm, bóng của ảnh siêu âm xung quanh các khối u, sự đa dạng về kích thước và hình dáng của các loại u buồng trứng. Đó là những thử thách mà cần đưa ra hướng giải quyết để nâng cao thêm độ chính xác. Dưới đây là một số đề xuất để phát triển đề tài:

- Nghiên cứu, tìm hiểu về bước tiền xử lý ảnh u buồng trứng của ảnh siêu âm khi đưa vào huấn luyện mô hình để nâng cao kết quả phân vùng.
- Tối ưu hoá và cải thiện mô hình hiện tại để nâng cao hiệu suất phân vùng.

Hướng nghiên cứu này có thể mở rộng kiến thức để ứng dụng vào việc phân vùng, phân loại ảnh siêu âm trong việc đánh giá và phân loại các khối u.

DANH MỤC TÀI LIỆU THAM KHẢO

- [1] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, “Attention u-net: Learning where to look for the pancreas,” arXiv preprint arXiv:1804.03999, 2018.
- [2] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” 2015.
- [3] D. Saha, A. Mandal, and R. Ghosh, “Mu net: Ovarian follicle segmentation using modified u-net architecture,” International Journal of Engineering and Advanced Technology (IJEAT), vol. 11, no. 4, April 2022.
- [4] Q.-V. T. T.-L. L. T.-H. T. H. V. D.-H. V. H.-T. D. V.-T. N. C.-M. P. P.-T. N. Thi Huong Nguyen, Thi-Loan Pham, “Systematic evaluation of loss functions for ovarian tumors segmentation from ultrasound images,” in International Conference on Health Science and Technology (ICHST), 2023.
- [5] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 9, pp. 1904–1916, 2015.
- [6] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” arXiv preprint arXiv:1708.02002, 2017.
- [7] S. Zhao, B. Wu, W. Chu, Y. Hu, and D. Cai, “Correlation maximized structural similarity loss for semantic segmentation,” arXiv preprint arXiv:1910.08711, 2019.
- [8] Q. Zhao, S. Lyu, W. Bai, L. Cai, B. Liu, M. Wu, X. Sang, M. Yang, and L. Chen, “A multi-modality ovarian tumor ultrasound image dataset for unsupervised cross domain semantic segmentation,” arXiv preprint arXiv:2207.06799, 2022.