|  |  |
| --- | --- |
|  | TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN  KHOA CÔNG NGHỆ THÔNG TIN |

ĐỀ CƯƠNG KHOÁ LUẬN TỐT NGHIỆP

xác định tế bào ung thư tuyến tiền liệt từ hình ảnh bệnh lý SỬ DỤNG MẠNG PHÂN TÍCH KIM TỰ THÁP

*(Identyfy Prostate Cancer from Pathology Images using Pyramid Scene Parsing Network)*

* 1. Thông tin chung

**Người hướng dẫn:**

* PGS.TS. Lê Hoàng Thái.

**[Nhóm]Sinh viên thực hiện:**

Nguyễn Thái Bảo (MSSV : 20127448).

Trần Phú (MSSV: )

**Loại đề tài:** Nghiên cứu.

**Thời gian thực hiện:** Từ *tháng 09/2023* đến *tháng 03/2024.*

* 1. Nội dung chi tiết
     1. Giới thiệu đề tài

Ung thư tuyến tiền liệt (PCa) là căn bệnh thứ sáu phổ biến nhất và gây tử vong thứ hai nhiều nhất ở nam giới trên toàn thế giới. Có nhiều kỹ thuật khác nhau để phát hiện và đánh giá PCa. Tuy nhiên, việc kiểm tra siêu vi kính của mẫu mô sinh thiết đã được nhuộm màu bởi các bác sĩ bệnh lý là phương pháp chính xác nhất. Dựa trên các mẫu histological có thể quan sát được, từng vùng của mô được gán một Gleason grade từ 1 đến 5. Điểm Gleason cuối cùng được báo cáo là tổng của hai mẫu rõ nét nhất và mẫu thứ hai rõ nét nhất; ví dụ, một mẫu mô có mẫu rõ nét nhất là Gleason grade 4 và mẫu thứ hai rõ nét nhất là Gleason grade 3 sẽ có điểm Gleason là 4+3.

Việc đánh giá Gleason của ung thư tuyến tiền liệt thường được thực hiện thông qua kiểm tra bằng mắt (bằng kính hiển vi) mẫu mô tuyến tiền liệt bởi các bác sĩ bệnh lý chuyên gia. Tuy nhiên, đây là một nhiệm vụ tốn thời gian và có sự biến động lớn giữa các bác sĩ, vì vậy các phương pháp tự động hỗ trợ máy tính có tiềm năng cải thiện tốc độ, độ chính xác và khả năng tái lập của kết quả.

* + 1. Mục tiêu đề tài

Với sự phát triển của lĩnh vực học sâu, rất nhiều mô hình học sâu áp dụng cho bài toán phân đoạn ảnh đã suất hiện. Trong vài năm trở lại đây, các mô hình đã được áp dụng nhiều hơn trong lĩnh vực y học để thực hiện các tác vụ phân đoạn các tế bào bất thường trong hình ảnh bệnh lý. Mặc dù độ chính xác của các mô hình ngày càng được cải thiện nhưng việc đưa vào thực tiễn vẫn còn gặp nhiều thách thức to lớn.

Với đề tài “xác định tế bào ung thư tuyến tiền liệt từ hình ảnh bệnh lý” đã đặt mục tiêu quan trọng là phát triển và ứng dụng các giải pháp tự động hóa trong việc đưa ra dự đoán về các phân vùng bị ung thư từ hình ảnh bệnh lý. Thách thức phản ảnh sự cần thiết của việc sử dụng trí tuệ nhân tạo để cải thiện quá trình chuẩn đoán và dự đoán Pca, một căn bệnh quan trọng và phổ biến ảnh hưởng đến nam giới trên toàn thế giới.

Mục tiêu cuối cùng là giúp cải thiện tốc độ, độ chính xác và tính đáng tin cậy của phân vùng ung thư, từ đó tạo điều kiện thuận lợi cho các bác sĩ trong việc đưa ra quyết định điều trị và dự đoán tiến triển của bệnh cho bệnh nhân.

* + 1. Phạm vi của đề tài

Nội dung nghiên cứu chính của đề tài là tập trung nghiên cứu mô hình học sâu đã đạt kết quả khá tốt gần đây là mô hình PSPNet [1]. Đầu tiên, nhóm sẽ tìm hiểu lý thuyết, thực nghiệm, phân tích các ưu điểm và khuyết điểm của mô hình. Từ đó nhóm sẽ nghiên cứu và đề xuất một số phương pháp để cải tiến mô hình.

Quá trình trên thực hiện dựa trên tập dữ liệu Gleason là: Gleason 2019 Challenge [2].

* + 1. Cách tiếp cận dự kiến
       1. Bài toán

Bài toán xác định phân vùng tế bào ung thư tuyến tiền liệt từ hình ảnh bệnh lý có đầu vào là các hình ảnh bệnh lý (Histopathology) về các tế bào và đầu ra của bài toán là phân vùng các điểm ung thư trên tế bào với các giai đoạn. Để thực hiện bài toán này, mô hình học sâu sẽ phải phân loại từng điểm ảnh của hình ảnh đầu vào thuộc vào phân loại các lớp là các giai đoạn ung thư (Gleason grade) gồm các lớp “Begin”, “Gleason grade 3”, “Gleason grade 4”, “Gleason grade 5”. Hình 1 là ví dụ của đầu vào và đầu ra của bài toán.

A close up of a picture

Description automatically generated with medium confidence*Hình 1: Hình bên trái là hình ảnh tế bào được đưa vào mô hình để dự đoán. Hình bên phải là đầu ra của mô hình với các phân vùng bị ung thư với các giai đoạn*

Các độ đo thường dùng cho bài toán liên quan đến phân đoạn ảnh bao gồm: Accuracy, Precision, Recall, IoU (Intersection over Union), Dice, v.v… . Trong đề tài này, nhóm quyết định sử dụng độ đo mean IoU (trung bình điểm IoU của các ảnh đầu vào) và độ đo Dice (được biết đến là F1 Score) để đánh giá kết quả thực nghiệm của mô hình. Công thức của 2 độ đo được thể hiện ở dưới hình 2.

A diagram of a mathematical equation

Description automatically generated

*Hình 2: Công thức của độ đo Iou và Dice được sử dụng trong việc đánh kết quả đầu ra của bài toán*

* + - 1. Các công trình liên quan

Nhiều phương pháp ứng dụng mạng học sâu đã được nghiên cứu và phát triển để thực hiện bài toán phân đoạn ảnh bệnh lý vào những năm gần đây. Đầu tiên có thể kể đến U-Net [3]. Về cơ bản, U-Net [3] sử dụng kiến trúc mã hóa – giải mã (encoder – decoder), bao gồm một đường thu hẹp để trích xuất đặc trưng và một đường mở rộng với một lớp phân loại, để thực hiện phân loại nhị phân đối với từng điểm ảnh trong một bản đồ đặc trưng đã được lấy mẫu. Bên cạnh đó có FCN [4], là một kiến trúc được sử dụng chủ yếu để phân đoạn theo ngữ nghĩa. Chúng chỉ sử dụng các lớp được kết nối cục bộ, chẳng hạn như tích chập, gộp và lấy mẫu lại. DeepLab [5], là một kiến trúc áp dụng một cách rất linh hoạt tích chập Atrous thay vì các phương pháp trước đó là áp dụng Transposed Convolution. Bên cạnh đó tác giả cũng áp dụng phương pháp Conditional Random Field để tinh chỉnh kết quả dự báo chuẩn xác hơn.

Mặc dù có rất nhiều mô hình đã được đề xuất và đạt kết quả khá tốt cho bài toán phân đoạn hình ảnh bệnh lý nhưng vẫn chưa có nhiều thực nghiệm để kiểm chứng độ hiệu quả của những mô hình đó cho bài toán phân đoạn tế bào ung thư nhóm sẽ thực hiện cải tiến dựa trên mô hình PSPNet [1].

* + - 1. Mô hình

A diagram of a diagram

Description automatically generatedPSPNet [1] là một mô hình phân đoạn ngữ nghĩa sử dụng mô-đun phân tích cú pháp kim tự tháp để khai thác thông tin ngữ cảnh toàn cầu bằng cách tổng hợp ngữ cảnh dựa trên các khu vực khác nhau. Sử dụng mạng CNN đã được huấn luyện từ trước(Resnet [6]) để đưa ra được bản đồ đặc trưng (features map) có kích thước bằng 1/8 so với ảnh gốc. Sau đó đi qua một mô-đun tổng hợp kim tự tháp với 4 cấp độ khác nhau sau đó được hợp nhất lại với nhau và thực hiện việc upsampling để đưa về kích thước bằng với kích thước ảnh gốc.

*Hình 3: Mô hình PSPNet* [1]

A diagram of a model architecture

Description automatically generated

*Hình 4: Kiến trúc mạng Resnet* [6]

Mục tiêu chính của mô-đun tổng hợp kim tự tháp (pyramid pooling module) là nâng cao khả năng phân đoạn các vùng khác nhau trong hình ảnh với các kích thước vùng quan sát khác nhau. PPM hoạt động bằng cách chia hình ảnh đầu vào thành các phạm vi hoặc cửa sổ con với các kích thước khác nhau, sau đó áp dụng một phép gom nhóm (pooling operation) trong từng cửa sổ con để trích xuất thông tin cụ thể từ mỗi phạm vi. Sau đó, các thông tin từ các cửa sổ con này được tổng hợp lại để tạo ra một biểu đồ của hình ảnh ở nhiều tỷ lệ kích thước khác nhau. Điều này giúp mô hình có khả năng phân đoạn các vật thể ở các kích thước và tỷ lệ khác nhau trong hình ảnh.

* + - 1. Hướng dẫn cải tiến đề xuất
         1. Đề xuất một số cơ chế chú ý để cải thiện thông tin tích chập.
* CBAM: Convolutional Block Attention Module

A diagram of a block diagram

Description automatically generatedCBAM [7] bao gồm 2 phần là Channel Attention và Spatial Attention. Input feature maps sẽ được thực hiện Channel Attention trước, sau đó thực hiện tiếp Spatial Attention (Hình 5).

*Hình 5: Khối CBAM* [7]

Channel Attention (Hình 6). Gồm phần Squeeze (thu) và phần Excitation (phóng). Phần Squeeze có trách nhiệm thu gom thông tin toàn cục (global information) sử dụng Global Average Pooling (GAP) và GMP (Global Max Pooling). Phần Excitation có trách nhiệm tạo attention trên chiều channel sử dụng lớp Fully Connected (FC) với activation function. Tác giả của CBAM có nói rằng GMP cũng thu thập thông tin quan trọng của vật thể nhưng theo một khía cạnh khác. Sau đó, features thu được thì GMP và GAP đều được cho đi qua chung phần Excitation (phóng).

A diagram of a channel attention model

Description automatically generated with medium confidence

*Hình 6: Channel Attention trong CBAM*

Spatial Attention (Hình 7). Cách thực hiện Spatial Attention của CBAM rất đơn giản, tương tự như Channel Attention. Thay vì thực hiện GAP và GMP theo chiều channel, ta thực hiện GAP và GMP theo chiều spatial của feature maps, tạo ra 2 feature maps tổng hợp với số chiều channel là 1. Hai feature maps này sau đó được nối lại với nhau và đi qua một 7×7 Conv kèm Sigmoid với số filter là 1 để tạo ra attention. Đặc biệt là phải dùng một Conv có kernel size to (chẳng hạn như 7x7) để có thể tính toán được vùng thông tin rộng hơn và tổng quát hơn.

A diagram of a system

Description automatically generated

*Hình 7: Spatial Attention trong CBAM*

* Dual Attention Network

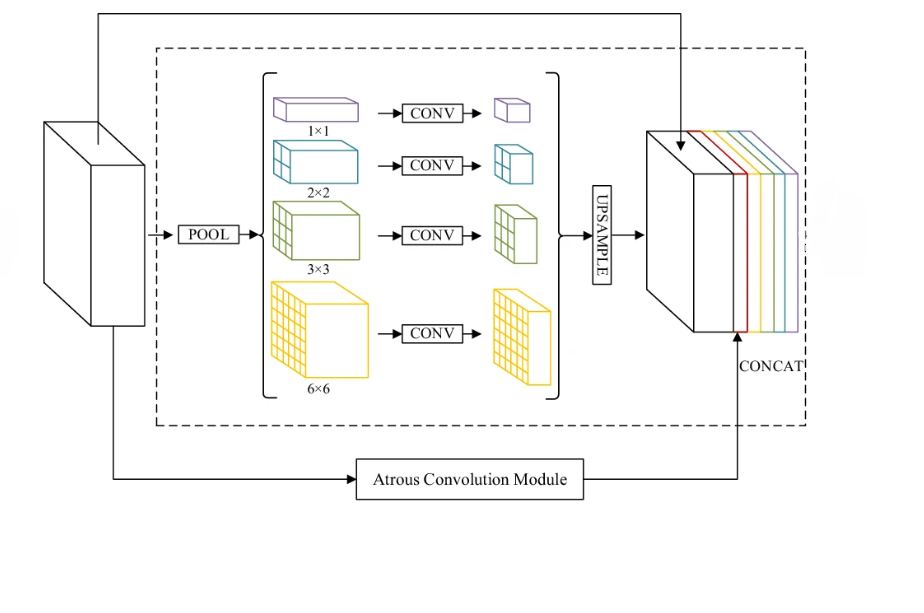
A diagram of a network

Description automatically generatedDual attention netword [8] được đề xuất để tích hợp các đặc trưng cục bộ với sự phụ thuộc toàn cục của chúng. DAN bao gồm 2 thành phần chính là Positional Attention Module (PAM) và Channel Attention Module (CAM). Trong đó, PAM tổng hợp có chọn lọc những đặc trưng tại từng vị trí bằng tổng trọng số của đặc trưng ở tất cả vị trí. Những đặc trưng tương tự sẽ liên quan đến nhau bất kể khoảng cách giữa chúng. Trong khi đó, CAM nhấn mạnh có chọn lọc sự phụ thuộc giữa các kênh bằng cách tích hợp những đặc trưng liên quan giữa chúng. Đầu ra của hai khối này sẽ được đi qua một hàm được gọi là Sum fusion để tổng hợp sự biểu diễn các đặc trưng, góp phần nâng cao hiệu quả phân đoạn. (Hình 8).

*Hình 8: Kiến trúc mô hình Dual Attention Network* [8]

* + - * 1. *Thêm vào một nhánh mô-đun tích chập astrous (Astrous convolution module) vào mô-đun kim tự tháp (Pyramid pooling module) nhằm cải thiện việc thu thập thông tin của mô hình.*

Việc thêm một nhánh astrous convolution [9] vào mô-đun kim tự tháp nhằm trích xuất các đặc điểm chung từ các đặc điểm nông trong quá trình lan truyền qua mạng cnn, đồng thời mở rộng cùng receptive fields. Và đầu ra của nhánh astrous convolution là một vùng thông tin nông thu thập từ ảnh đầu vào. Đầu ra của mô-đum kim tự tháp được kết hợp với đầu ra của nhánh astrous convolution để mạng có thể giữ được thông tin toàn cục vào cục bộ.



*Hình 9: Mô-đun kim tự tháp (Pyramid pooling module) kết hợp với astrous convolution*

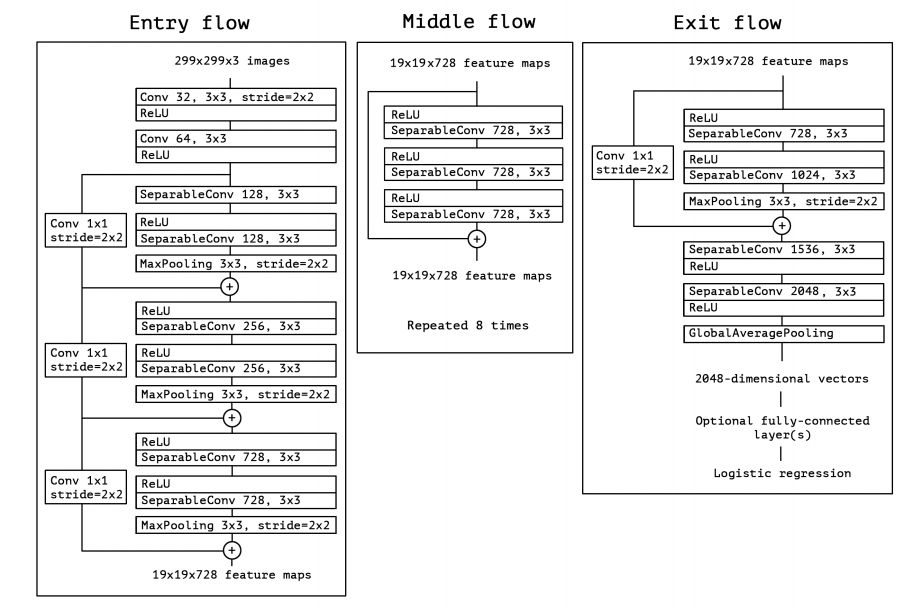
* + - * 1. *Sử dụng mô hình đã được huấn luyện từ trước như Xception , … để nâng cao hiệu quả huấn luyện.*

Xception [10] là một kiến trúc mạng tích chập sâu bao gồm các kết cấu có thể phân tách theo chiều sâu. Nó được phát triển bởi các nhà nghiên cứu của Google. Google đã trình bày cách diễn giải các mô-đun Inception trong mạng nơ ron tích chập như một bước trung gian giữa phép tích chập thông thường và phép toán tích chập có thể phân tách theo chiều sâu (một phép tích chập theo chiều sâu, sau đó là một phép tích chập theo điểm).

Xception có cấu trúc dựa trên hai điểm chính:

* Tích chập có thể phân tách theo chiều sâu (Depthwise Separable Convolution).
* Các kết nối tương tự như resitdual của resnet được gọi là phím tắt giữa các khối (Shortcuts between Convolution blocks).

Xception gồm 3 khối lớn chính là Entry flow, Middle flow, Exit flow.



*Hình 10: kiến trúc Xception* [10]

Việc nhóm sự dụng kiến trúc đã huấn luyện xception vì theo kết quả đánh giá của tác giả cho thấy Xception [10] sử dụng kiến trúc tích chập phân tách theo chiều sâu (Convolutional depthwise separable) thay vì convolution thông thường như Resnet [6]. Kiến trúc này giúp mô hình có thể học được các đặc trưng ở mức độ chi tiết cao hơn và giảm lượng tham số cần thiết.

* + - * 1. *Kĩ thuật test time augmentation (TTA).*

Kĩ thuật Test time agumentation [11] là một kĩ thuật được dùng trong việc đánh giá mô hình trong quá trình huấn luyện, việc thực hiện các biến đổi phù hợp đối với tập dữ liệu kiểm tra nhằm cải thiện hiệu suất dự đoán tổng thể. Cụ thể trong TTA [11], kỹ thuật tăng cường dữ liệu sẽ được áp dụng trên từng ảnh của tập kiểm tra, nhiều ảnh tăng cường mới sẽ được tạo ra. Mô hình sẽ thực hiện dự đoán riêng trên các ảnh tăng cường này và trả về kết quả dự đoán trung bình trên các cách biến đổi ảnh.

Nhóm sẽ áp dụng kĩ thuật này theo hai hướng là lật dọc (vertical flip) và lật ngang (horizontal flip) cho quá trình đánh giá mô hình.

* + 1. Kết quả dự kiến của đề tài

Kết quả dự kiến của đề tài bao gôm:

* Hiểu được tổng quan bài toán phân đoạn từ hình ảnh bệnh lý và một số mô hình tiêu biểu trong lĩnh vực này.
* Nắm được cơ chế hoạt động, điểm mạnh và yếu của mô hình PSPNet. Nhóm sẽ tiến hành huấn luyện mô hình lại từ đầu trên tập dữ liệu đã nêu trên
* Kết hợp các phương pháp, kĩ thuật cải tiến đã nêu trên và tiến hành đánh giá tường hình thước cải tiến. Sau đó lập bảng đánh giá so sánh kết quả đạt được để đưa ra được phương pháp cải tiến tối ưu nhất.
* Cuối cùng là so sánh kết quả của nhóm với các mô hình khác.
  + 1. Kế hoạch thực hiện

|  |  |  |
| --- | --- | --- |
| **THỜI GIAN** | **CÔNG VIỆC** | **PHÂN CÔNG** |
| Tháng 8/2023 | 1. Nghiên cứu các khái niệm và lý thuyết cho bài toán phân đoạn ảnh bệnh lý nói chung và bài toán phân đoạn tế bào ung thư nói riêng.  2. Tìm hiểu về các công trình nghiên cứu đã được đề xuất để giải quyết bài toán. | Tất cả thành viên |
| Tháng 9/2023 | 1. Thảo luận và chọn ra một mô hình tiêu biểu để nghiên cứu sâu.  2. Nghiên cứu chi tiết về bài báo của mô hình đã chọn.  3. Xử lý các tập dữ liệu phù hợp cho việc thực nghiệm của mô hình. | Tất cả thành viên |
| Tháng 10/2023 | 1. Đọc hiểu mã nguồn PSPNet của bài báo gốc  2. Tiến hành cài đặt trên máy cá nhân đẻ thực nghiệm | Thái Bảo |
| Tháng 11/2023 | 1. Từ kết quả thực nghiệm, phân tích cơ chế hoạt động, các tính chất, điểm mạnh và điểm yếu của mô hình.  2. Đề xuất hướng cải tiến cho mô hình | Tất cả thành viên |
| Tháng 12/2023 | 1. Tiến hành cài đặt các hướng cải tiến dựa trên mô hình gốc.  2. Huấn luyện mô hình cải tiến dựa trên tập dữ liệu đã xử lý .  3. Lập bảng so sánh đánh giá độ hiệu quả, thời gian huấn luyện.  4. Chọn ra mô hình cải tiến tối ưu nhất và hiệu quả nhất để thực hiện huấn luyện và đánh giá chi tiết nhất. | Tất cả thành viên |
| Tháng 2-3/2024 | 1. Viết báo cáo khóa luận  2. Chuẩn bị powerpoint thuyết trình | Tất cả thành viên |

Tài liệu tham khảo

|  |  |
| --- | --- |
| [1] | J. S. X. Q. X. W. J. J. Hengshuang Zhao, Pyramid Scene Parsing Network, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017. |
| [2] | Danielle Walker. Miccai automatic prostate gleason grading challenge, 2019. |
| [3] | O. RONNEBERGER, P. FISCHER and T. BROX, U-net: Convolutional networks for biomedical image segmentation, International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015. p. 234-241.. |
| [4] | E. S. a. T. D. J. Long, Fully convolutional networks for semantic segmentation, In CVPR, 2015. |
| [5] | G. P. I. K. K. M. a. A. L. Y. L. Chen, Semantic image segmentation with deep convolutional nets and fully connected, crfs. arXiv:1412.7062, 2014.. |
| [6] | X. Z. S. R. a. J. S. K. He, Deep residual learning for image recognition., In CVPR, 2016. |
| [7] | J. P. J.-Y. L. I. S. K. Sanghyun Woo, CBAM: Convolutional Block Attention Module, In CVPR, [v2] 2018. |
| [8] | J. e. a. FU, Dual attention network for scene segmentation., In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019. p. 3146-3154.. |
| [9] | "A Primer on Atrous(Dilated) and Depth-wise Separable Convolutions," [Online]. Available: https://towardsdatascience.com/a-primer-on-atrous-convolutions-and-depth-wise-separable-convolutions-443b106919f5. |
| [10] | F. Chollet., Xception: Deep Learning with Depthwise Separable Convolutions., In CVPR [v3] 2017.. |
| [11] | M. KIMURA, Understanding Test-Time Augmentation., In International Conference on Neural Information Processing. Springer, Cham, 2021. p. 558-569.F. |