

Trường Đại học Khoa học Tự nhiên
Khoa Công Nghệ Thông Tin



Truy vấn thông tin thị giác – Project

Nguyễn Gia Bảo Danh — 21127589

Phạm Bùi Tuấn Anh — 21127579

Huỳnh Hiệp Phát — 21127130

Giảng viên

Phạm Minh Hoàng

Nguyễn Trọng Việt

Võ Hoài Việt

TP Hồ Chí Minh – Tháng 7 - 20

Contents

I.	Thông tin sinh viên.....	3
1.	Phân công công việc.....	3
2.	Bảng đóng góp.....	3
II.	Các công trình nghiên cứu liên quan.....	4
1.	Medical Image Retrieval Using Pretrained Embeddings	4
a.	Giới thiệu.....	4
b.	Các đóng góp chính	4
c.	Phương pháp	4
d.	Kết quả thực nghiệm	4
e.	Kết luận	5
2.	Medical Image Retrieval using Deep Convolutional Neural Network.....	5
a.	Giới thiệu.....	5
b.	Phương pháp	5
c.	Kết quả thực nghiệm	6
d.	Kết luận	6
3.	X-MIR: EXplainable Medical Image Retrieval	7
a.	Giới thiệu	7
b.	Các đóng góp chính.....	7
c.	Phương pháp.....	7
d.	Kết quả thực nghiệm.....	8
e.	Kết luận.....	8
4.	Medical Image Retrieval Based on Convolutional Neural Network and Supervised Hashing.....	8
a.	Giới thiệu.....	8
b.	Các đóng góp chính	8
c.	Phương pháp	9
d.	Kết quả thực nghiệm	9
e.	Kết luận	9
III.	Nội dung báo cáo.....	10
1.	Giới thiệu.....	10
2.	Bộ dữ liệu.....	10
3.	Phương pháp	11
a.	Tạo tập dữ liệu điểm chuẩn 3D-MRI	11
b.	Retrieval Method	14
IV.	Thực nghiệm.....	16
V.	Giải pháp nâng cao hiệu quả	17

II. Các công trình nghiên cứu liên quan

1. Medical Image Retrieval Using Pretrained Embeddings

a. Giới thiệu

+ Hệ thống truy xuất hình ảnh dựa trên nội dung (CBIR) trong lĩnh vực hình ảnh y tế gặp nhiều thách thức do sự đa dạng của các kỹ thuật và định dạng hình ảnh. Việc phát triển các hệ thống truy xuất hiệu quả rất quan trọng trong việc thúc đẩy nghiên cứu y tế và cải tiến các công cụ chẩn đoán. Nghiên cứu này đánh giá tính khả thi của việc sử dụng bốn mô hình được huấn luyện trước để truy xuất hình ảnh y tế ở các cấp độ khác nhau.

b. Các đóng góp chính

- + Đánh giá hiệu suất của các mô hình DINOv1, DINOv2, DreamSim, và Swin Transformer (cùng với ResNet50) trong nhiệm vụ truy xuất hình ảnh y tế 3D.
- + So sánh hai phương pháp lập chỉ mục cơ sở dữ liệu vector: LSH và HNSW.
- + Đề xuất phương pháp tổng hợp thông tin từ các lát cắt 2D để truy xuất hình ảnh 3D, và so sánh các chiến lược lấy mẫu khác nhau.

c. Phương pháp

+ Nghiên cứu sử dụng cơ sở dữ liệu MSD Challenge với các ảnh y tế 3D và 4D. Các mô hình DINOv1, DINOv2, và DreamSim được huấn luyện tự giám sát trên tập dữ liệu ImageNet. Swin Transformer và ResNet50 được huấn luyện trên tập dữ liệu RadImageNet. Các ảnh được tiền xử lý để phù hợp với yêu cầu đầu vào của các mô hình, và các chiến lược lấy mẫu khác nhau được áp dụng để cải thiện hiệu suất truy xuất.

d. Kết quả thực nghiệm

+ Các thử nghiệm cho thấy các mô hình được huấn luyện trước có thể truy xuất hình ảnh y tế với độ chính xác cao. Các mô hình DINOv1 và DINOv2 đạt hiệu suất tốt nhất ở cấp độ truy xuất hình ảnh 3D, với độ chính xác gần như hoàn hảo. Các phương pháp lập chỉ mục HNSW và LSH đều cho thấy khả năng xử lý hiệu quả các yêu cầu truy xuất với tốc độ nhanh và độ chính xác cao.

e. Kết luận

+ Nghiên cứu chứng minh rằng việc sử dụng các mô hình học sâu được huấn luyện trước có thể cải thiện đáng kể hiệu suất truy xuất hình ảnh y tế mà không cần huấn luyện thêm. Việc so sánh các phương pháp lập chỉ mục và chiến lược lấy mẫu cung cấp cái nhìn sâu sắc về các lựa chọn tối ưu cho các ứng dụng truy xuất hình ảnh y tế trong thực tế. Các kết quả này mở ra khả năng ứng dụng rộng rãi các mô hình học sâu trong các hệ thống CBIR cho y tế, đồng thời đề xuất hướng phát triển tiếp theo để cải thiện hiệu suất và độ chính xác.

2. Medical Image Retrieval using Deep Convolutional Neural Network

a. Giới thiệu

Bài báo đề xuất một khung làm việc dựa trên "deep learning" cho việc truy xuất hình ảnh y tế dựa trên nội dung ("content-based medical image retrieval" - CBMIR) sử dụng một mạng nơ-ron tích chập sâu ("deep convolutional neural network" - DCNN). Cách tiếp cận này nhằm cải thiện hiệu suất truy xuất bằng cách sử dụng các đặc trưng học được từ hình ảnh y tế để thu hẹp khoảng cách ngữ nghĩa giữa các đặc trưng hình ảnh cấp thấp và các khái niệm y tế cấp cao.

b. Phương pháp

Tổng quan khung làm việc:

- Hệ thống đề xuất sử dụng DCNN cho cả phân loại và trích xuất đặc trưng.
- Mô hình DCNN có ba loại lớp: lớp tích chập ("convolutional layers"), lớp gộp ("pooling layers"), và lớp kết nối đầy đủ ("fully connected layers").
- Mạng được huấn luyện để phân loại hình ảnh thành 24 lớp y tế dựa trên cơ quan cơ thể.

Giai đoạn 1: Phân loại:

- DCNN được huấn luyện trên một tập dữ liệu hình ảnh y tế, được chia thành các lớp.
- Mạng học các đặc trưng phân cấp sâu trực tiếp từ hình ảnh, giúp phân loại hiệu quả.

Giai đoạn 2: Trích xuất đặc trưng cho CBMIR:

- Các đặc trưng được trích xuất từ ba lớp kết nối đầy đủ cuối cùng (FCL1, FCL2, và FCL3) của DCNN đã huấn luyện.
- Một cơ sở dữ liệu đặc trưng được tạo ra bằng cách sử dụng các đặc trưng này.
- Để truy xuất, hình ảnh truy vấn được đưa qua mạng, và các đặc trưng của nó được so sánh với các đặc trưng trong cơ sở dữ liệu sử dụng khoảng cách Euclid ("Euclidean distance").

c. Kết quả thực nghiệm

Tập dữ liệu:

- Tập dữ liệu chứa 7200 hình ảnh từ 24 lớp, với các hình ảnh được thay đổi kích thước thành 256×256 pixel.
- Tập dữ liệu được chia thành 5040 hình ảnh huấn luyện và 2160 hình ảnh thử nghiệm.

Hiệu suất phân loại:

- DCNN đạt độ chính xác phân loại trung bình 99.77%.
- Các chỉ số đánh giá bao gồm độ chính xác ("precision"), độ hồi đáp ("recall"), và "F1 measure".

Hiệu suất truy xuất:

- "Mean Average Precision" (mAP) là 0.53 không có dự đoán lớp và 0.69 với dự đoán lớp.
- Kết quả truy xuất tốt hơn khi sử dụng dự đoán lớp, vì các hình ảnh không liên quan từ các lớp khác bị loại bỏ.

So sánh với các hệ thống tiên tiến:

- Phương pháp đề xuất được so sánh với các hệ thống khác bằng các chỉ số như mAP.
- Mặc dù một số phương pháp đạt mAP cao hơn, hệ thống đề xuất hỗ trợ dữ liệu đa phương tiện, cung cấp phạm vi ứng dụng rộng hơn.

d. Kết luận

+ Khung làm việc được đề xuất cho CBMIR sử dụng DCNN cho thấy sự cải thiện đáng kể trong hiệu suất truy xuất bằng cách sử dụng dự đoán lớp để lọc kết quả tìm kiếm. Hệ thống đạt độ chính xác phân loại cao và kết quả truy

xuất đầy hứa hẹn, với các cải tiến tiềm năng trong tương lai bao gồm các tập dữ liệu lớn hơn và thích ứng cho các ứng dụng thể tích 3D.

3. X-MIR: EXplainable Medical Image Retrieval

a. Giới thiệu

+ Mục tiêu: Bài báo này nhằm giải quyết vấn đề truy xuất hình ảnh y tế có thể giải thích được, đặc biệt là trong bối cảnh đại dịch COVID-19 và các bệnh về da. Việc truy xuất hình ảnh y tế có thể giải thích được là rất quan trọng để các bác sĩ có thể hiểu và tin tưởng vào các quyết định của hệ thống học máy.

+ Vấn đề: Các hệ thống học máy hiện tại thường được xem như “hộp đen”, nghĩa là chúng đưa ra quyết định mà không cung cấp lý do rõ ràng. Điều này gây khó khăn cho việc chấp nhận và sử dụng các hệ thống này trong y tế.

b. Các đóng góp chính

+ Phát triển: Bài báo phát triển một tiêu chuẩn cho truy xuất hình ảnh trên hai bộ dữ liệu y tế công khai, giúp đánh giá và so sánh các phương pháp khác nhau.

+ Áp dụng: Sử dụng bản đồ saliency dựa trên sự tương đồng để xác định các vùng quan trọng trong hình ảnh y tế, giúp giải thích quyết định của mô hình.

+ Đánh giá: Đề xuất các chỉ số nhân quả để đánh giá hiệu quả của các thuật toán saliency khác nhau, đảm bảo rằng các bản đồ saliency thực sự phản ánh các yếu tố quan trọng trong hình ảnh.

c. Phương pháp

+ Học sâu: Sử dụng các mô hình học sâu để học không gian nhúng cho hình ảnh y tế. Không gian nhúng này giúp phân tách các lớp khác nhau trong dữ liệu, làm cho việc truy xuất hình ảnh trở nên chính xác hơn.

+ Bản đồ saliency: Sử dụng ba loại bản đồ saliency dựa trên sự tương đồng:

+ Dựa trên che khuất: Xác định các vùng quan trọng bằng cách che khuất từng phần của hình ảnh và quan sát sự thay đổi trong kết quả.

+ Dựa trên sự chú ý: Sử dụng các mô hình chú ý để xác định các vùng quan trọng trong hình ảnh.

+ Dựa trên kích hoạt: Sử dụng các lớp kích hoạt trong mô hình học sâu để xác định các vùng quan trọng.

d. Kết quả thực nghiệm

+ Hiệu suất: Các mô hình học sâu cho thấy khả năng phân tách các lớp khác nhau trong không gian nhúng, giúp cải thiện độ chính xác của việc truy xuất hình ảnh.

+ Đánh giá: Các bản đồ saliency cho thấy các vùng quan trọng của hình ảnh được sử dụng để tính toán sự tương đồng, giúp giải thích quyết định của mô hình một cách rõ ràng hơn.

e. Kết luận

+ Tiềm năng: Các bản đồ saliency có thể giúp các bác sĩ hiểu rõ hơn về quyết định của mô hình học máy, từ đó tăng cường sự tin tưởng và chấp nhận các hệ thống này trong y tế. Điều này có thể dẫn đến việc cải thiện hiệu suất truy xuất hình ảnh y tế và hỗ trợ tốt hơn cho việc chẩn đoán và điều trị bệnh.

4. Medical Image Retrieval Based on Convolutional Neural Network and Supervised Hashing

a. Giới thiệu

Bài báo này tập trung vào vấn đề truy xuất ảnh y tế, một lĩnh vực quan trọng trong y học hiện đại nhằm hỗ trợ các bác sĩ và chuyên gia y tế tìm kiếm và so sánh các hình ảnh y tế tương tự. Việc truy xuất hình ảnh nhanh và chính xác có thể hỗ trợ chẩn đoán và điều trị bệnh hiệu quả hơn. Phương pháp được đề xuất kết hợp các mạng nơ-ron tích chập (CNN) và kỹ thuật băm có giám sát để cải thiện hiệu suất truy xuất hình ảnh y tế.

b. Các đóng góp chính

+ Tích hợp CNN và mã hóa băm: Phương pháp này làm việc này sử dụng CNN để trích xuất các đặc trưng sâu từ hình ảnh y tế. Sau đó, các đặc trưng này được mã hóa thành các mã nhị phân nhỏ gọn, giúp quá trình truy xuất nhanh hơn và giảm dung lượng lưu trữ.

+ Kiến trúc mạng Siamese: Mạng Siamese được sử dụng để học các đặc trưng của hình ảnh và làm cho các hình ảnh tương tự có các đặc trưng gần nhau trong không gian đặc trưng. Mạng này gồm hai nhánh CNN chia sẻ trọng số, nhận đầu vào là các cặp hình ảnh và sử dụng hàm mất mát tương phản để tối ưu hóa.

+ Hàm mất mát mới: Một hàm mất mát được thiết kế để khuyến khích các mã nhị phân dễ phân biệt hơn. Hàm mất mát này bao gồm một thuật ngữ điều chuẩn giúp đầu ra của CNN gần với các giá trị nhị phân mong muốn, cải thiện hiệu suất của quá trình nhị phân hóa.

+ Tập trung vào học có giám sát: Phương pháp này đặc biệt hữu ích trong các tình huống học có giám sát, nơi các nhãn dữ liệu có sẵn. Khả năng học sâu của CNN được khai thác tối đa để tối ưu hóa các mã nhị phân, làm tăng độ chính xác của việc truy xuất hình ảnh.

c. Phương pháp

+ Cấu trúc mạng: Mạng Siamese nhận đầu vào là các cặp hình ảnh và học cách ánh xạ chúng vào không gian mã nhị phân k-bit. Các mã này được thiết kế để các hình ảnh tương tự có các mã nhị phân gần nhau, giúp truy xuất hiệu quả hơn.

+ Tối ưu hóa hàm mất mát: Hàm mất mát tương phản được sử dụng để đảm bảo rằng các hình ảnh tương tự có đầu ra gần nhau, trong khi các hình ảnh không tương tự có đầu ra cách xa nhau. Điều này giúp tạo ra các mã nhị phân có tính phân biệt cao.

d. Kết quả thực nghiệm

+ Tập dữ liệu: Các thử nghiệm được thực hiện trên các tập dữ liệu hình ảnh y tế tiêu chuẩn để đánh giá hiệu quả của phương pháp đề xuất. Các hình ảnh này bao gồm các loại hình ảnh y tế khác nhau như MRI, CT, và X-quang.

+ Đánh giá hiệu suất: Phương pháp được đánh giá qua các chỉ số như độ chính xác, độ nhạy, và thời gian truy xuất. Kết quả cho thấy phương pháp CNN kết hợp với băm có giám sát vượt trội so với các phương pháp truyền thống về cả độ chính xác và tốc độ.

+ So sánh với các phương pháp khác: Bài báo cũng so sánh phương pháp đề xuất với một số phương pháp truy xuất hình ảnh y tế hiện có. Phương pháp mới cho thấy hiệu suất vượt trội trong hầu hết các thử nghiệm, khẳng định tính ưu việt của việc kết hợp CNN và kỹ thuật băm.

e. Kết luận

+ Phương pháp truy xuất hình ảnh y tế dựa trên CNN và băm có giám sát trong kiến trúc mạng Siamese cung cấp một giải pháp mạnh mẽ và hiệu quả. Việc tích hợp CNN để trích xuất các đặc trưng sâu và mã hóa chúng thành các mã nhị phân nhỏ gọn giúp cải thiện đáng kể hiệu suất truy xuất hình ảnh y tế. Kết quả thực nghiệm cho thấy phương pháp này không chỉ tăng độ chính xác mà còn giảm thời gian truy xuất, giúp các chuyên gia y tế có thể chẩn đoán và điều trị bệnh nhanh chóng và chính xác hơn.

+ Khung làm việc này có tiềm năng ứng dụng rộng rãi trong nhiều lĩnh vực y tế khác nhau, từ chẩn đoán bệnh đến nghiên cứu y khoa, giúp cải thiện chất lượng chăm sóc sức khỏe và nâng cao hiệu quả làm việc của các bác sĩ và chuyên gia y tế.

III. Nội dung báo cáo.

1. Giới thiệu

Chúng em nhận thấy rằng việc sử dụng hình ảnh y tế ngày càng tăng trong các cơ sở chăm sóc sức khỏe đặt ra một thách thức đáng kể do khối lượng công việc ngày càng tăng đối với các bác sĩ X quang. Tuy nhiên, nó cũng mang lại cơ hội nâng cao kết quả chăm sóc sức khỏe nếu được tận dụng hiệu quả. Truy xuất hình ảnh 3D có tiềm năng giảm khối lượng công việc của bác sĩ X quang bằng cách cho phép các bác sĩ lâm sàng tìm kiếm hiệu quả thông qua các trường hợp chẩn đoán tương tự hoặc có liên quan khác, dẫn đến chẩn đoán nhanh hơn và chính xác hơn.

Tuy nhiên, lĩnh vực truy xuất hình ảnh y tế 3D vẫn đang nổi lên, thiếu các tiêu chuẩn đánh giá được thiết lập, bộ dữ liệu toàn diện và nghiên cứu kỹ lưỡng. Nghiên cứu này sẽ cố gắng thu hẹp khoảng cách này bằng cách giới thiệu một tiêu chuẩn mới cho Truy xuất hình ảnh y tế 3D (3D-MIR) bao gồm bốn giải pháp khác nhau được chụp bằng chụp cắt lớp vi tính.

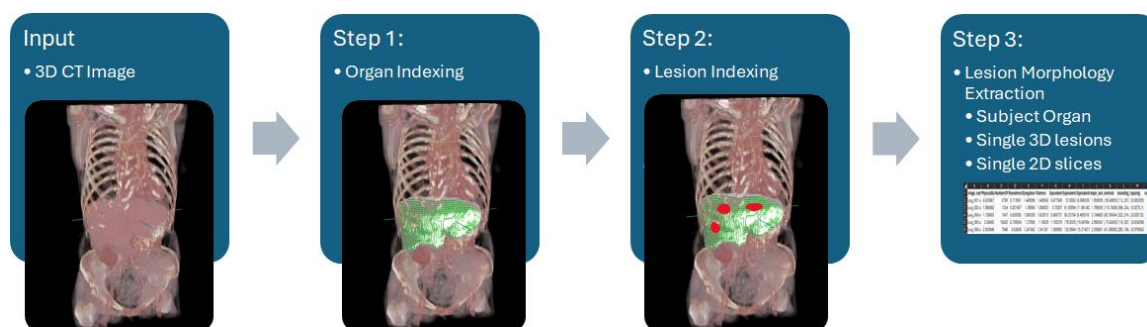
Sử dụng điểm chuẩn này, bài nghiên cứu đã tìm ra một bộ chiến lược tìm kiếm đa dạng sử dụng các lát cắt 2D tổng hợp, khối lượng 3D và nhúng đa phương thức từ các mô hình nền tảng đa phương thức phổ biến làm truy vấn. Đánh giá định lượng và định tính của từng phương pháp đã được cung cấp cùng với một cuộc thảo luận chuyên sâu, nhằm mang lại cái nhìn sâu sắc cho nghiên cứu trong tương lai.

2. Bộ dữ liệu

Dữ liệu được lấy từ tập dữ liệu phân đoạn hình ảnh y tế trong thử thách Phân đoạn Y khoa (The Medical Segmentation Decathlon - MSD). Dữ liệu bao gồm tổng cộng 2.633 hình ảnh ba chiều từ nhiều bộ phận giải phẫu được quan tâm bằng nhiều phương thức và nhiều nguồn khác nhau. Cụ thể dữ liệu chứa các cơ quan sau: Não, Tim, Gan, Hô hấp, Tuyến tiền liệt, Phổi, Tuyến tụy, Mạch gan, Lá lách và Đại tràng. Trong bài nghiên cứu đặc biệt tập trung vào hệ CT ngực từ MSD, bao gồm các điểm tổn thương 3D ở 4 cơ quan chính của ngực và bụng: gan, đại tràng, tuyến tụy và phổi.

3. Phương pháp

a. Tạo tập dữ liệu điểm chuẩn 3D-MRI



Hình 1 Quá trình này bắt đầu với toàn bộ khối CT đầu vào. Trong bước 1 ở ảnh trên, phân đoạn cơ quan được áp dụng cho 104 cơ quan khác nhau, nhưng trong ví dụ này, chương trình chỉ hiển thị gan, được làm nổi bật bằng màu xanh lá cây. Bước 2 liên quan đến việc xác định và lập chỉ mục từng tổn thương (hiển thị bằng màu đỏ) trên các cơ quan khác nhau. Cuối cùng, Bước 3 tiến hành phân tích hình thái của các tổn thương, tạo ra các chỉ số cụ thể cho mỗi đối tượng, cơ quan và tổn thương

Trong phần này, chúng em sẽ trình bày chi tiết về quá trình phát triển bộ dữ liệu 3D-MIR đa cơ quan. Mục tiêu của bước này là sử dụng các bộ dữ liệu công khai có chú thích tổn thương 3D để xác định và định lượng các tổn thương, và cung cấp mô tả hình thái chi tiết về các tổn thương này trong bối cảnh các cơ quan liên quan của chúng. Điều quan trọng cần lưu ý là mục tiêu của bước này không phải là tạo ra các báo cáo chẩn đoán lâm sàng tiêu chuẩn hoặc phân đoạn; thay vào đó, bước này tập trung vào việc tạo ra các mô tả chi tiết nhấn mạnh các đặc điểm hình thái của các tổn thương để hỗ trợ đo lường sự liên quan giữa truy vấn và tìm kiếm trong các thiết lập truy xuất.

Bộ dữ liệu 3D-MIR được xây dựng bằng hai nguồn chính: (i) bộ dữ liệu Medical Segmentation Decathlon (MSD) , cung cấp 10 nhóm công khai. Nghiên cứu đặc biệt tập trung vào nhóm CT ngực từ MSD, có chú thích tổn thương 3D ở bốn cơ quan chính của ngực và bụng bao gồm: gan, ruột kết, tụy và phổi. Để thực hiện quá trình đánh giá của chúng tôi, chúng tôi đã chọn làm việc với các khối từ tập huấn luyện, vì các khối từ tập kiểm tra không có nhãn. Phân phối các cơ quan có nhãn trong các khối được chọn như sau: gan (131 khối), phổi (64 khối), ruột kết (126 khối), như vậy có tổng cộng 281 khối chứa tổn thương. Hình 1 hiển thị các bước chính để xây dựng bộ dữ liệu sử dụng ví dụ về gan. (ii) ở bước này sử dụng mô hình phân đoạn ngữ nghĩa đa cơ quan toàn diện, TotalSegmentator dựa trên các kiến trúc giống U-Net . Tạo tập dữ liệu gồm 3 bước như sau:

Bước 1: Lập Chỉ Mục Cơ Quan

Đầu tiên chương trình áp dụng TotalSegmentator cho mỗi khối CT, tạo ra 104 khối CT cắt và các nhãn tương ứng cho mỗi đối tượng. Các khối cắt này

được xác định bởi các hộp giới hạn 3D, mỗi hộp bao gồm cơ quan quan tâm. Phân đoạn ban đầu này đạt được hai mục tiêu chính: nó định vị các tổn thương trong một bản đồ giải phẫu toàn diện và hợp lý hóa quá trình lập chỉ mục cho các truy vấn cụ thể về cơ quan, tăng cường hiệu quả và độ chính xác của truy xuất dữ liệu. Tại Hình 1, Bước 1 hiển thị gan đã được phân đoạn màu xanh lá cây.

Bước 2: Lập Chỉ Mục Tổn Thương

Trong bước này, bắt đầu phân tích khối lượng nhãn 3D liên quan đến mỗi lần quét CT để xác định các tổn thương. Vì các tổn thương 3D được biểu diễn dưới dạng khối lượng nhị phân, phân biệt cấu trúc tổn thương khỏi giải phẫu xung quanh, bằng cách sử dụng phân tích kết nối một thành phần. Phương pháp này hiệu quả trong việc tách rời các tổn thương riêng lẻ, cho phép gán các định danh duy nhất cho từng tổn thương và đối tượng tương ứng. Sau đó, bắt đầu tiến hành định lượng sự chồng lấn không gian của từng tổn thương với 104 cơ quan được phân đoạn bởi TotalSegmentator. Việc định lượng này rất quan trọng vì nó cho phép chúng ta lập bản đồ chính xác từng tổn thương với cơ quan tương ứng và phân bố tổn thương trong cơ quan đó. Việc lập bản đồ như vậy rất quan trọng trong việc tạo ra một bộ dữ liệu mạnh mẽ và chính xác, tạo điều kiện cho các phân tích và truy vấn cụ thể về cơ quan. Hình 2, Bước 2 hiển thị tổn thương riêng lẻ được định vị màu đỏ trong gan.

Bước 3: Trích Xuất Hình Thái Tổn Thương

Trong bước này sẽ tập trung vào việc trích xuất các đặc điểm hình thái của từng tổn thương được xác định trong các bước trước bằng cách phân tích hình học 3D và 2D của các tổn thương được định vị trong mỗi cơ quan. Hơn nữa, chúng tôi trích xuất các chỉ số hình thái tương ứng với:

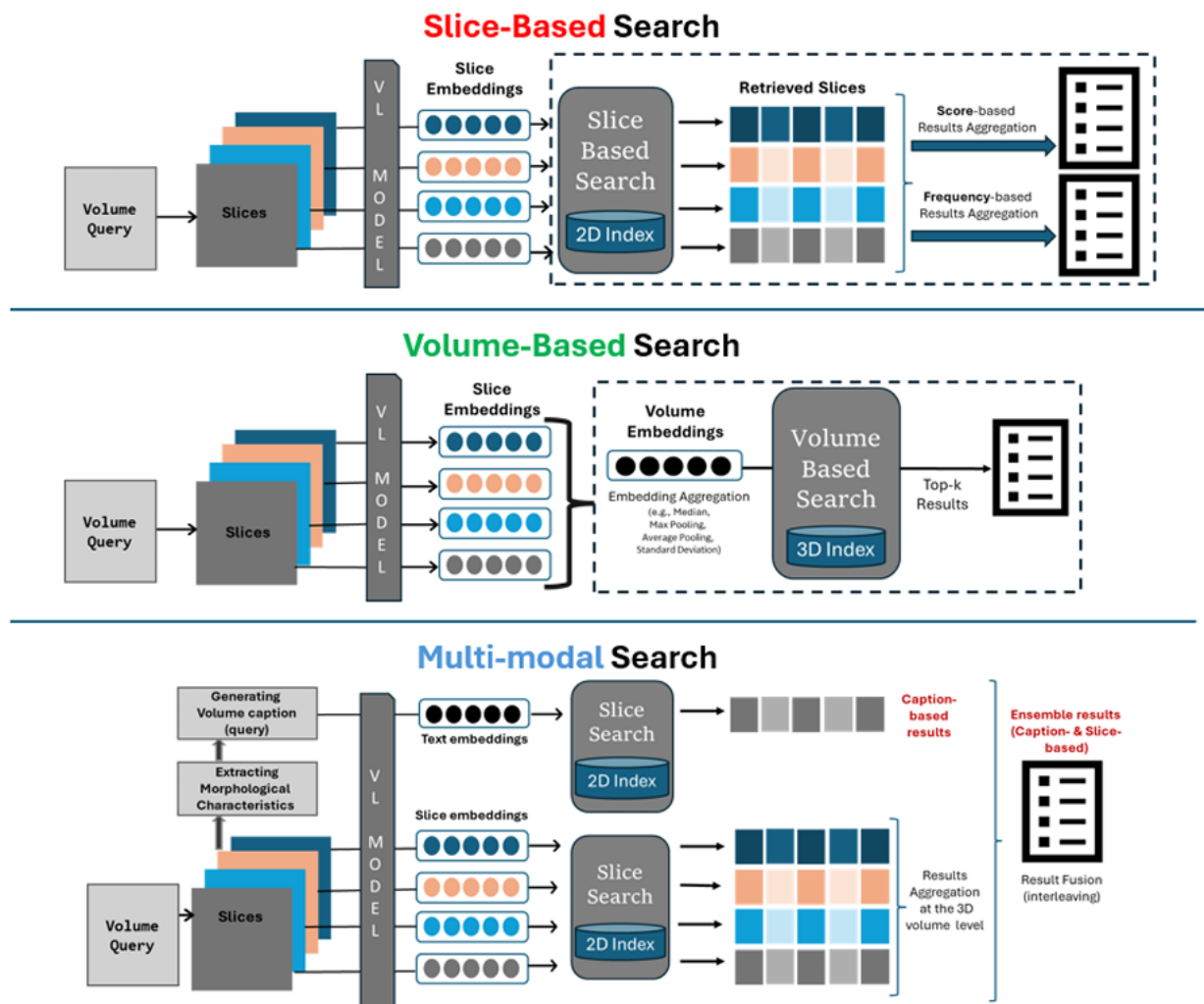
a. Đối tượng - Cơ quan: Các đối tượng được nhóm dựa trên các đặc điểm hình thái của tổn thương của họ, theo các hướng dẫn do Ủy ban chung về ung thư của Hoa Kỳ về hệ thống phân loại Tumor, Node, Metastasis (TNM). Việc nhóm này tuân theo các khuyến nghị của hệ thống TNM về việc phân loại các giai đoạn ung thư và cung cấp một khuôn khổ tiêu chuẩn hóa để liên kết hình thái tổn thương với giai đoạn ung thư. Chúng tôi tập trung chủ yếu vào hình thái và phân bố tổn thương, thay vì phân giai đoạn ung thư. Cụ thể, chúng ta xem xét số lượng, chiều dài và thể tích của các tổn thương để xác định ba nhóm:

- **Nhóm Tổn Thương 1:** các đối tượng có một tổn thương nhỏ hơn 2 cm.
- **Nhóm Tổn Thương 2:** các đối tượng có một tổn thương lớn hơn 2 cm hoặc nhiều tổn thương, nhưng không cái nào vượt quá 5 cm.

- **Nhóm Tổn Thương 3:** các đối tượng có một hoặc nhiều tổn thương, mỗi tổn thương lớn hơn 5 cm.
- b. Tổn thương 3D riêng lẻ:** Đối với mỗi tổn thương riêng biệt, chúng ta tính toán các mô tả hình thái 3D toàn diện. Các mô tả này rất quan trọng trong việc hiểu các đặc điểm của tổn thương, bao gồm: 1) Hình thái tổn thương, đặc biệt tập trung vào các hình dạng đối xứng, 2) thể tích của tổn thương trong không gian thực tế, kích thước của tổn thương, được biểu diễn qua các kích thước của một ellipsoid 3D phù hợp. Để trích xuất các chỉ số này, chúng tôi đã sử dụng bộ công cụ Insight Segmentation and Registration Toolkit (ITK).
- c. Các lát cắt 2D riêng lẻ:** Để hoàn thành phân tích hình thái 3D, định lượng các lát cắt 2D liên quan đến tổn thương. Trong mỗi lát cắt được chọn, chúng tôi đo diện tích tổng cộng của các tổn thương, đếm số lượng tổn thương riêng lẻ và phân tích hình thái của chúng, tập trung vào các đặc điểm như độ tròn, cho phép truy xuất dựa trên lát cắt.

Bộ dữ liệu 3D-MIR được thiết kế có chủ đích để tạo điều kiện cho cả các truy vấn rộng (chẳng hạn như phân biệt giữa các cơ quan có và không có tổn thương) và các truy vấn chi tiết (như số lượng và kích thước tổn thương trong mỗi cơ quan). Nó có thể thích ứng để tích hợp với các bộ dữ liệu công khai hiện có mà đã bao gồm các nhãn tổn thương. Chúng tôi dự định công khai bộ dữ liệu này cùng với phiên bản sẵn sàng của bài báo này.

b. Retrieval Method



Nghiên cứu dựa vào mô hình ngôn ngữ hình ảnh BiomedCLIP để tạo ra các nhúng(embedding) cho hình ảnh 2D và chú thích của chúng. BiomedCLIP là mô hình SOTA trong truy xuất đa phương thức y sinh, phân loại hình ảnh và trả lời câu hỏi trực quan, vượt trội hơn các mô hình như CLIP(open ai), MedCLIP và PubMedCLIP, cũng như các mô hình SOTA cụ thể trong lĩnh vực chẩn đoán hình ảnh như BioViL trong các nhiệm vụ chẩn đoán hình ảnh (ví dụ: phát hiện viêm phổi RSNA).

Data:

BiomedCLIP được huấn luyện trước trên PMC-15M, một bộ dữ liệu gồm 15 triệu cặp hình ảnh-văn bản y sinh từ hơn ba triệu bài báo. BiomedCLIP cũng được cải tiến với các điều chỉnh cụ thể cho xử lý ngôn ngữ hình ảnh y sinh. Các hình ảnh input bao gồm các lát cắt 2D được trích xuất từ các tập CT scan, với các giá trị voxel được biểu diễn trong các đơn vị Hounsfield (HU).

Để xử lý, nghiên cứu đã chuẩn hóa các tập 3D này từ phạm vi ban đầu của $[-1000, 1000]$ HU đến thang đo giá trị 8-bit trong khoảng $[0, 255]$. Sau đó, từ các giá trị chuẩn hóa này, tạo ra một hình ảnh đơn sắc.

B1 Truy xuất dựa trên lát cắt

Phương pháp tìm kiếm này dựa vào các lát cắt 2D của các tập 3D. Mỗi lát cắt 2D được mã hóa bằng cách sử dụng các nhúng từ model. Tất cả các lát cắt 2D của tất cả các tập được lập chỉ mục trong cơ sở dữ liệu vector, sử dụng thư viện Faiss.

Trong bước tìm kiếm, đối với một tập 3D có input là V , truy xuất dựa trên lát cắt bao gồm việc sử dụng tất cả các lát cắt 2D của V làm truy vấn (được ký hiệu là QV), sau đó tìm kiếm trong cơ sở dữ liệu vector để có được 20 lát cắt tương tự nhất cho mỗi lát cắt trong QV dựa trên Euclidean distance.

Các lát cắt 2D được truy xuất sau đó được tổng hợp lại với các tập 3D gốc của chúng để trả về k hình ảnh giống nhất. ba phương pháp tổng hợp kết quả lát cắt khác nhau đã được so sánh: tần suất, điểm tương đồng tối đa và tổng điểm tương đồng của các lát cắt được truy xuất $R(QV)$ cho mỗi hình ảnh 3D có các lát cắt trong kết quả truy vấn.

Ký hiệu các lát cắt được truy xuất từ QV là $R(QV)$, và tập hợp các lát cắt được truy vấn thuộc cùng một tập V_i là $RS(V_i)$, ba phương pháp tổng hợp dựa trên các phương pháp tính điểm tập gốc như sau:

Trong mỗi phương pháp tổng hợp, các tập gốc được chấm điểm sau đó lần lượt được xếp hạng theo các công thức tương ứng trong các phương trình 1, 2 và 3. Điểm SimScore dựa trên khoảng cách Euclidean giữa các nhúng của lát cắt truy vấn và các nhúng của một lát cắt được truy xuất.

B2 Truy xuất dựa trên tập

Phương pháp tìm kiếm dựa trên tập dựa vào việc tổng hợp các nhúng của các lát cắt 2D để tạo ra một vector nhúng đại diện cho toàn bộ tập 3D. bốn phương pháp tổng hợp nhúng khác nhau đã được so sánh: trung vị, max pooling, average pooling và độ lệch chuẩn. Mỗi phương pháp tổng hợp Mi đã được nêu trên sẽ được sử dụng để tạo ra một cơ sở dữ liệu/vector chỉ mục riêng biệt (chỉ mục 3D). Trong bước tìm kiếm, chúng ta sẽ tạo các nhúng của ảnh cần truy vấn V theo 1 trong 4 phương pháp nêu trên và tìm kiếm các tập/vector tương ứng nhất trong chỉ mục 3D tương ứng cơ sở dữ liệu bên trên.

IV. Thực nghiệm

Input Data: [Medical Segmentation Decathlon](#).

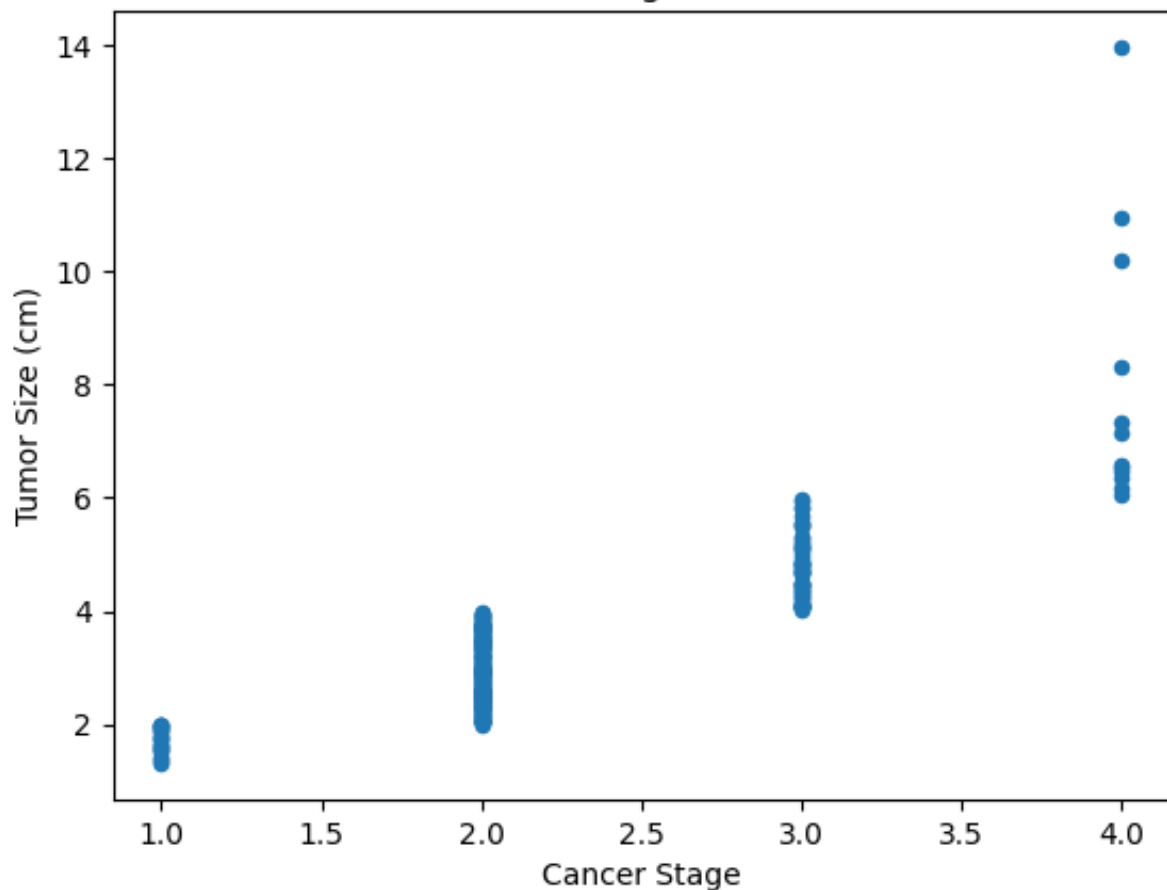
Bước 1: Organ Segmentation

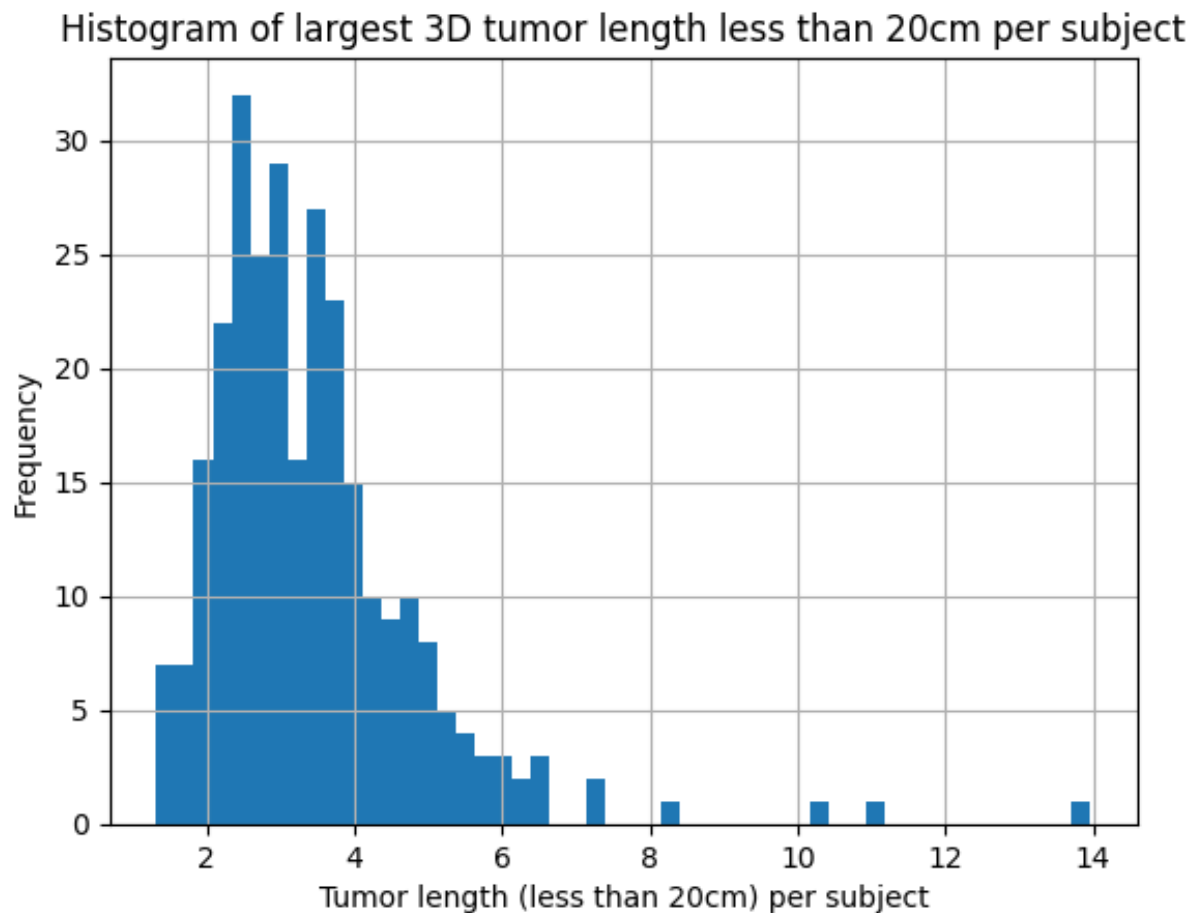
- chạy file [run_totalsegmentator.py](#) để thực hiện phân đoạn nội tạng(Organ Segmentation)

Bước 2: Image Processing and Quantification

- Tùy vào cơ quan mà chúng ta chọn để chạy file định lượng cho cơ quan tương ứng
- Sau khi chạy file ta có được một số kết quả như số lượng khối u của các kích cỡ khác nhau, 1 số biểu đồ để trực quan hoá các mối quan hệ của cơ quan nội tạng và bệnh

Distribution of cancer stage vs tumor size for Liver





Bước 3: Embeddings Generation (BiomedCLIP)

- Tải và xử lý trước hình ảnh chụp CT từ dataset.
- Đọc và chuyển đổi các bản quét CT từ định dạng NIfTI thành một mảng có nhiều mảng.
- Đối với mỗi CT scan: load CT scan -> Generate embeddings -> lưu embeddings vào .pkl file.

Bước 4: Retrieval Methods & Evaluation

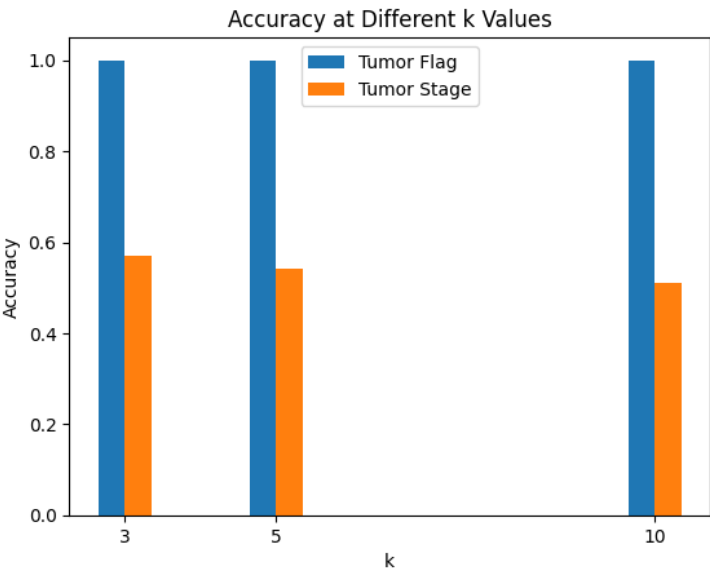
- Có 3 phương pháp truy vấn Truy xuất dựa trên lát cắt, Truy vấn dựa trên tập, Truy vấn đa phương tiện(hiện nay chúng em chưa thể chạy được bước cuối do 1 số lỗi)

V. Giải pháp nâng cao hiệu quả

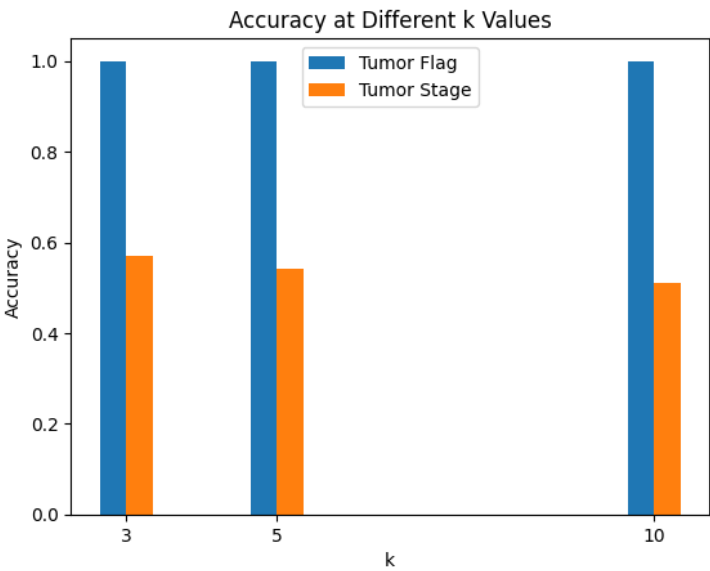
Trong phần cải tiến, nhóm em đã bổ sung thêm điều kiện " lesion group" vào quá trình truy vấn nhằm cải thiện độ chính xác trong việc phân loại và đánh giá giai đoạn khối u. Đoạn mã mới so sánh cả giai đoạn ung thư và nhóm tổn thương tuyến tụy (pancreas_lesion_group) giữa hình ảnh truy vấn và kết quả truy xuất, giúp cải thiện khả năng nhận diện sự tương đồng giữa các mẫu.

Nếu cả giai đoạn ung thư và nhóm tổn thương tuyến tụy của hình ảnh truy vấn và kết quả khớp nhau, kết quả sẽ được đánh dấu là tương tự. Điều này giúp tăng cường độ chính xác của phương pháp trong việc phân loại chi tiết, đặc biệt trong các trường hợp yêu cầu đánh giá chính xác hơn về giai đoạn và nhóm tổn thương của khối u.

Trước khi cải tiến:



Sau khi cải tiến:



Kết quả	k = 3	k = 5	K = 10
Trước cải tiến			
Tumor Flag	1.000	1.000	1.000
Tumor Stage	0.4881	0.5214	0.5409
Sau cải tiến			
Tumor Flag	1.0000	1.0000	1.0000

Tumor Stage	0.5714	0.5429	0.5104
-------------	--------	--------	--------

Nhận xét:

- Độ chính xác cho việc phát hiện khối u đạt mức tối đa ở tất cả các truy vấn ($k = 3, 5, 10$) cho thấy sự hiệu quả và độ chính xác trong việc phát hiện khối u. Độ chính xác sau khi cải thiện vẫn ở mức độ tối đa.
- Trước khi cải thiện, việc phân biệt các giai đoạn của khối u vẫn chưa quá chính xác với giá trị k_3 đạt 0.4881, k_5 đạt 0.5214 và k_{10} đạt 0.5409 cho thấy hiệu suất ko ổn định của phương pháp. Sau cải thiện, độ chính xác được cải thiện đáng kể, đặt biệt là ở k_3 là 0.5713.

VI. Kết luận

Đây là nghiên cứu giới thiệu điểm chuẩn đầu tiên cho Truy xuất hình ảnh y tế 3D (3D-MIR), bao gồm bốn loại cơ quan được chụp bằng phương pháp chụp cắt lớp vi tính (CT scan), và đánh giá một loạt các phương pháp tìm kiếm dựa trên các mô hình nền tảng đa phương tiện tiên tiến nhất. Những phương pháp này bao gồm các truy vấn dựa trên lát cắt 2D tổng hợp, tập 3D và nhúng đa phương tiện. Các phát hiện của nghiên cứu cho thấy trong khi các mô hình nền tảng hiện tại thành thạo trong việc xác định các chi tiết ngữ nghĩa thô như sự hiện diện của tổn thương, chúng cũng đồng thời gặp khó khăn với thông tin chi tiết như việc nhóm các tổn thương theo kích thước. Do các mô hình nền tảng hiện tại được thiết kế cho đầu vào 2D, nghiên cứu của chúng tôi tập trung vào việc điều tra hai cách tiếp cận để kết hợp thông tin 2D vào ngữ cảnh 3D cho việc tìm kiếm hình ảnh 3D: dựa trên lát cắt và dựa trên tập. Kết quả thử nghiệm của nghiên cứu gợi ý rằng các phương pháp dựa trên tập có thể hiệu quả hơn cho các phân loại rộng, trong khi các phương pháp dựa trên lát cắt có thể là lựa chọn hiệu quả hơn để nắm bắt các chi tiết nhỏ. Cuối cùng, kết quả của nghiên cứu chứng minh rằng vẫn còn khoảng trống đáng kể để cải thiện hiệu suất hơn nữa trên tất cả các nhiệm vụ.

P/s: Trong lúc thực hiện tìm hiểu nghiên cứu, tui em đã chọn một paper đang được thực hiện nên trong quá trình thực hiện và thuyết trình, các dữ liệu của dự án được upload sau khi thực hiện thuyết trình nên tui em đã làm lại phần thuyết trình của mình nhiều lần. Mong thầy thông cảm.

VII. Tài liệu tham khảo

1. [Medical Image Retrieval Using Pretrained Embeddings](#)
2. [3D-MIR: A BENCHMARK & EMPIRICAL STUDY ON 3D MEDICAL IMAGE RETRIEVAL IN RADIOLOGY](#)
3. [Medical Image Retrieval using Deep Convolutional Neural Network](#)
4. [X-MIR: EXplainable Medical Image Retrieval](#)
5. [Medical Image Retrieval Based on Convolutional Neural Network and Supervised Hashing](#)