ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

KHOA TOÁN TIN HỌC

BÁO CÁO MÔN HỌC PHÂN TÍCH XỬ LÍ ẢNH

ĐỀ TÀI

Phát hiện các tổn thương cột sống trên ảnh X-quang bằng phương pháp kết hợp mô hình phân loại và phát hiện

GV hướng dẫn: Huỳnh Thanh Sơn

Nhóm sinh viên thực hiện:

Họ và tên	*MSSV*
Phạm Ngọc Hào	23110146
Nguyễn Đặng Văn Cảnh	23110135

Hồ Chí Minh, 2026

Mục lục

1 Mở đầu

Các bệnh lý cột sống hiện đang là một trong những vấn đề sức khỏe phổ biến nhất trên toàn cầu, gây ảnh hưởng không nhỏ đến khả năng vận động và chất lượng cột sống của bệnh nhân. Trong quy trình chẩn đoán lâm sàng, chụp X-quang là phương pháp chẩn đoán hình ảnh cơ bản, chi phí thấp và được sử dụng rộng rãi để phát hiện các tổn thương. Tuy nhiên, việc chẩn đoán những tổn thương cột sống thông qua sự phân tích và đánh giá ảnh X-quang đòi hỏi bác sĩ chẩn đoán hình ảnh có kinh nghiệm dày dặn và tính chuyên môn cao. Một số thương tổn thường có kích thước nhỏ, độ tương phản thấp và dễ bị che khuất bởi các cấu trúc giải phẫu phức tạp. Ngoài ra, áp lực quá tải tại các cơ sở y tế hiện nay có thể dẫn đến tình trạng mệt mỏi cho bác sĩ, làm tăng nguy cơ sai sót hoặc bỏ sót tổn thương.

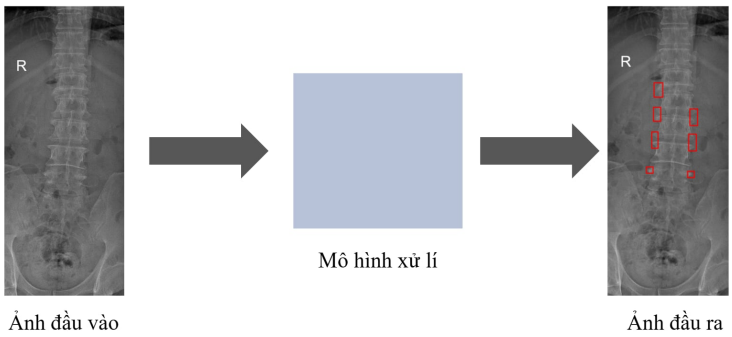
Xuất phát từ thực tế trên, đồ án này tập trung xây dựng một quy trình hoàn chỉnh có khả năng tự động phát hiện và khoanh vùng các tổn thương trên ảnh X-quang cột sống. Cụ thể, nghiên cứu tập trung giải quyết các bài toán sau. Trước hết, đồ án xây dựng mô hình phân loại bằng mô hình DenseNet-201 [1] để đánh giá tổng quan xác suất mắc bệnh của ảnh đầu vào, đóng vai trò như một bộ lọc mức độ rủi ro. Tiếp theo, xây dựng mô hình phát hiện tổn thương bằng mô hình Sparse R-CNN [2] để xác định các tổn thương nhỏ và khó phát hiện. Cuối cùng, đồ án đề xuất cơ chế hợp nhất quyết định bằng cách sử dụng một thuật toán kết hợp giữa kết quả phân loại và phát hiện. Cơ chế này sử dụng độ tin cậy của mô hình phân loại để điều chỉnh ngưỡng nhạy của mô hình phát hiện, nhằm giảm thiểu tỉ lệ báo động giả trong khi vẫn đảm bảo không bỏ sót các ca bệnh quan trọng.

Để thực hiện các mục tiêu trên, nhóm đã sử dụng và đánh giá trên bộ dữ liệu VinDR-SpineXR [3], bao gồm các ảnh X-quang cột sống được gán nhãn bởi các bác sĩ chẩn đoán hình ảnh nhiều kinh nghiệm, nhằm đảm bảo tính khách quan cho quá trình huấn luyện và kiểm thử mô hình.

2 Giới thiệu

2.1 Phát biểu bài toán

Trong đồ án này, bài toán được đặt ra là xây dựng một hệ thống tự động có khả năng phát hiện và khoanh vùng các tổn thương trên ảnh X-quang cột sống. Với đầu vào là ảnh X-quang cột sống, mô hình cần dự đoán vị trí các vùng bất thường và biểu diễn chúng thông qua các hộp giới hạn trên ảnh đầu vào (xem hình 1).

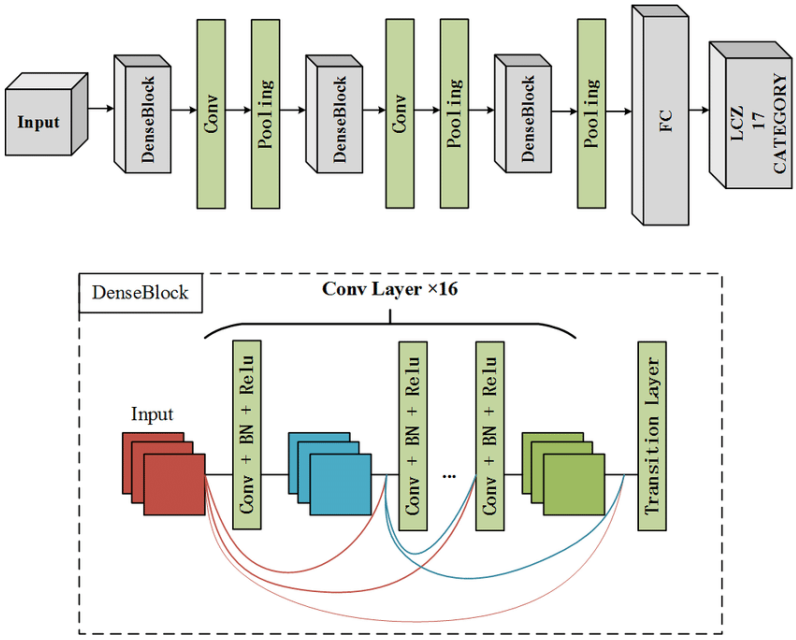


Hình 1: Minh họa ảnh đầu vào và ảnh đầu ra

Bài toán được xem là một bài toán phát hiện đối tượng trong lĩnh vực thị giác máy tính y sinh, trong đó mỗi tổn thương được mô hình hóa dưới dạng một đối

tượng cần được định vị chính xác. Mục tiêu của hệ thống là hỗ trợ bác sĩ trong quá trình sàng lọc và chẩn đoán, giảm thiểu nguy cơ bỏ sót tổn thương và nâng cao độ tin cậy trong phân tích ảnh X-quang.

2.2 Mô hình phân loại DenseNet



Hình 2: Sơ đồ kiến trúc DenseNet với các khối Dense Block

Trong bài toán phân loại ảnh X-quang cột sống, kiến trúc DenseNet-201 [1] có khả năng khai thác hiệu quả các đặc trưng hình ảnh. DenseNet là một mạng nơ-ron tích chập có chứa các khối Dense Block. Trong một khối DenseNet, đầu ra của mỗi lớp sau được nối trực tiếp với tất cả các lớp phía trước (xem hình 2). Cụ thể, đầu ra của lớp thứ ℓ trong một khối Dense Block được xác định bởi:

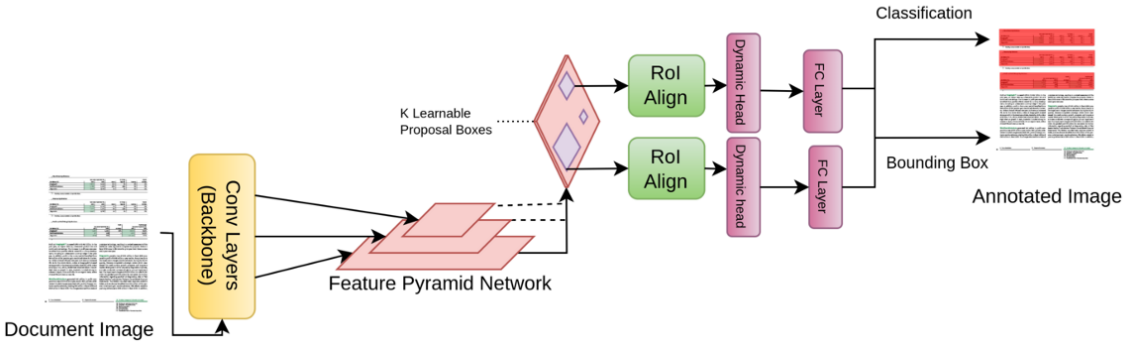
$$x_\ell = H_\ell([x_0, x_1, \ldots, x_{\ell-1}])$$

trong đó $[\cdot]$ là phép nối các đặc trưng của mỗi lớp và $H_\ell(\cdot)$ là tổ hợp các phép Batch Normalization, ReLU và phép tích chập. Cơ chế trên giúp cải thiện khả năng lan truyền ngược, tăng cường tái sử dụng đặc trưng và hạn chế mất mát thông tin ở các lớp sâu của mạng, với dữ liệu ảnh y tế có tổn thương nhỏ và khó quan sát.

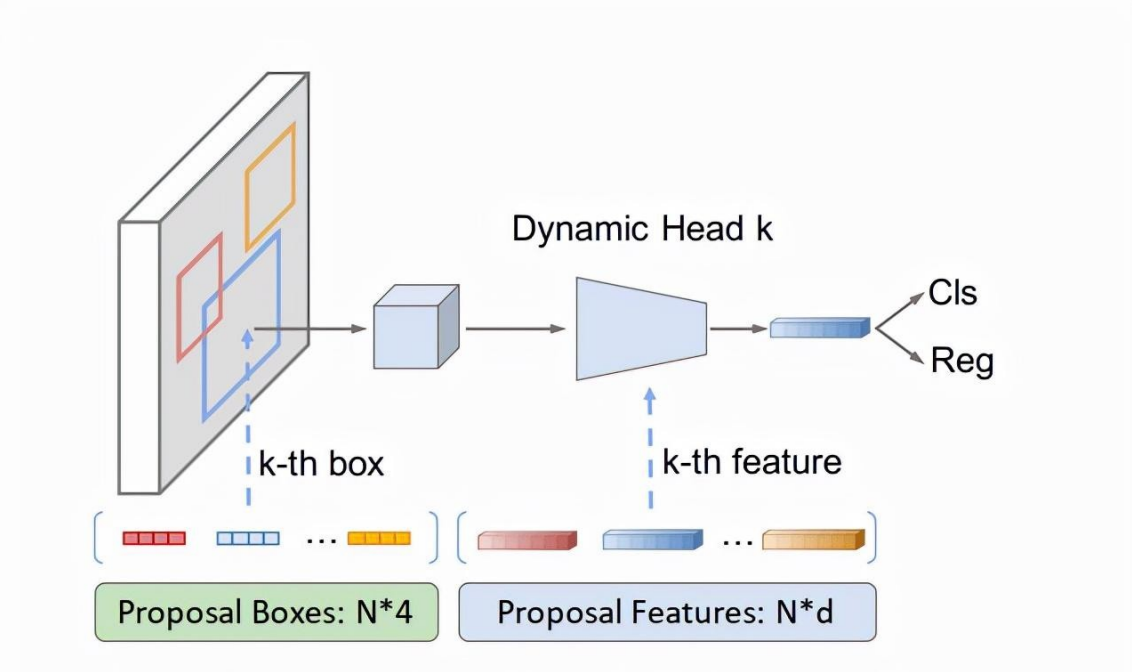
2.3 Mô hình phát hiện Sparse R-CNN

Sparse R-CNN [2] là mô hình mạng nơ-ron sâu được đề xuất cho bài toán phát hiện đối tượng trong ảnh. Khác với các mô hình truyền thống, Sparse R-CNN sử dụng tập nhỏ các đề xuất được học trực tiếp và tinh chỉnh lặp lại trong quá trình huấn luyện. Cấu trúc của mô hình được minh họa trong hình 3, 4. Trước hết, ảnh đầu vào được đưa qua mạng xương sống (thường là ResNet) để trích xuất các đặc trưng ban đầu. Sau đó, mạng FPN được sử dụng để tạo ra các đặc trưng đa tỷ lệ, giúp mô hình phát hiện hiệu quả các đối tượng có kích khác nhau. Tiếp theo, mô hình khởi tạo một tập hợp cố định các đề xuất (thường từ 100 đến 300 đề xuất), mỗi đề xuất gồm các thông tin về hộp giới hạn và đặc trưng tương ứng. Các đề xuất này được tinh chỉnh qua nhiều lần lặp, trong đó đặc trưng trích xuất lại các hộp hiện tại và được xử lý bởi các khối Dynamic Head nhằm học cách kết hợp thông tin một

cách thích ứng. Cuối cùng, các đặc trưng đã được tinh chỉnh được đưa qua các lớp mạng nơ-ron để thực hiện phân loại đối tượng và hồi quy vị trí hộp giới hạn, tạo ra kết quả phát hiện cuối cùng.



Hình 3: Sơ đồ kiến trúc tổng thể của mô hình Sparse R-CNN



Hình 4: Minh họa Dynamic Head trong mô hình Sparse R-CNN. Cụ thể, phần này nhận đầu vào bao gồm hình ảnh đầu, một tập hợp hộp đề xuất và đặc trưng tương ứng của hộp. Trong đó, tập hợp các đề xuất là tham số có thể học. Mạng xương sống ở phía trước trích xuất các đặc trưng của các hộp đề đề xuất và các đặc trưng tương ứng rồi đưa qua Dynamic Head các đặc trưng của vật thể, và cuối cùng cho ra các kết quả phân lớp và phát hiện.

2.4 Các độ đo đánh giá mô hình

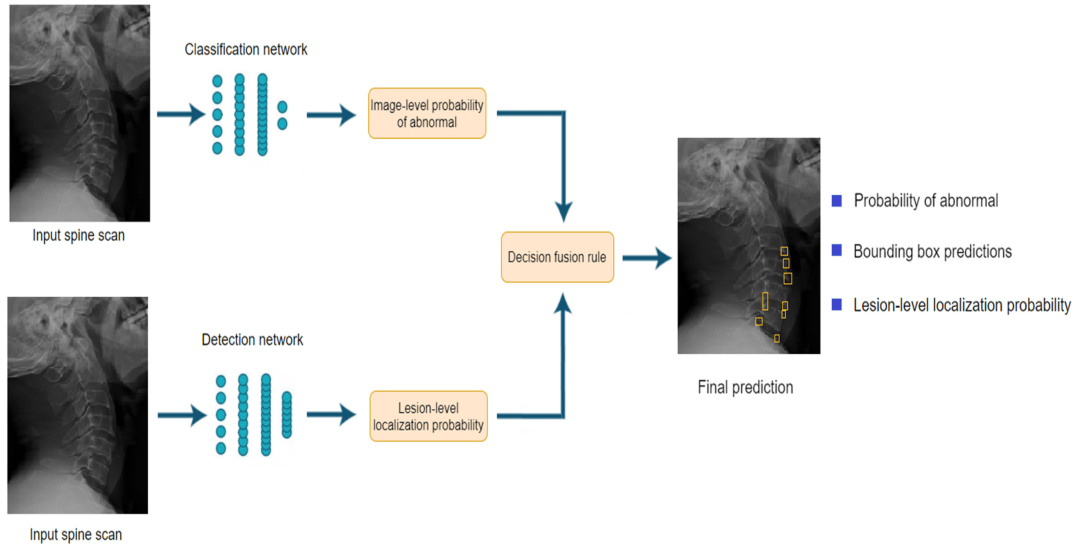
Để đánh giá hiệu quả của hệ thống, đồ án sử dụng các độ đo tiêu chuẩn được cộng đồng nghiên cứu tin dùng, nhằm đảm bảo tính khách quan cho từng bài toán. Do dữ liệu X-quang y tế thường mất cân bằng giữa số lượng mẫu bình thường và mẫu bệnh, Accuracy không phản ánh đầy đủ hiệu năng mô hình. Giai đoạn phân loại tập trung vào các chỉ số AUROC, phản ánh khả năng phân biệt của mô hình giữa lớp bệnh và lớp không bệnh trên mọi ngưỡng quyết định và các chỉ số bổ trợ như

F1-score, độ nhạy, độ đặc hiệu giúp đánh giá chi tiết hơn khả năng phát hiện đúng ca bệnh và giảm báo động giả. Đối với bài toán phát hiện tổn thương, đồ án sử dụng chỉ số AP và mAP@0.5 [4] làm độ đo chính. Chỉ số này là trung bình của Precision trên tất cả các lớp bệnh lý. Chỉ số mAP@0.5 cung cấp đánh giá tổng quan về khả năng mô hình định vị và phân loại chính xác các tổn thương so với nhãn thực tế của bác sĩ, đồng thời cho phép so sánh hiệu năng giữa các mô hình khác nhau.

3 Quy trình đề xuất

3.1 Tổng quan về quy trình

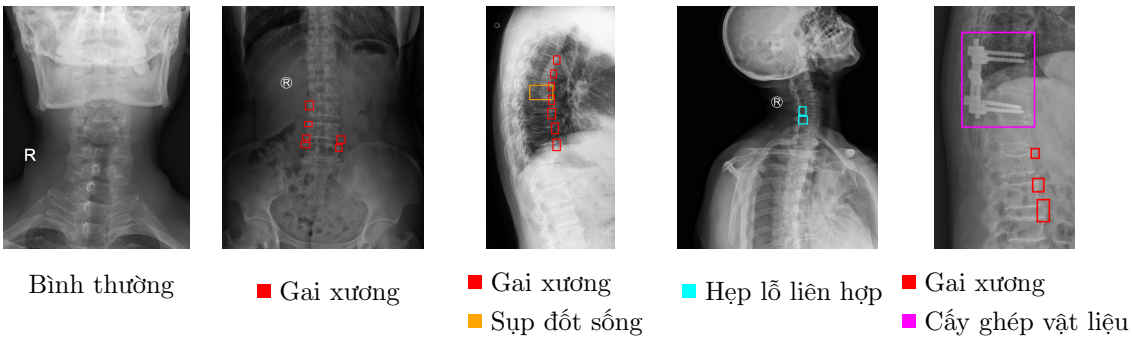
Quy trình của bài toán (xem hình 5) gồm hai phần chính:(1) một mô hình phân lớp được dùng để nhận ảnh X-quang là một ảnh đầu vào và dự đoán xem ảnh có thể là bình thường hoặc bất thường; (2) một mô hình phát hiện nhận cùng một hình ảnh và dự đoán phân vùng những chỗ bất thường. Để tối ưu hiệu suất của mô hình phát hiện, đồ án đề xuất một cơ chế kết hợp quyết định để kết hợp kết quả của hai mô hình.



Hình 5: Tổng quan về đồ án phân loại bất thường và định vị tổn thương. (trích nguồn từ: [5])

3.2 Tập dữ liệu

Đồ án thu thập và sử dụng bộ dữ liệu VinDR-SpineXR [3]. Đây là một trong những bộ dữ liệu mở lớn nhất hiện nay về ảnh X-quang cột sống, được công bố bởiVinBigData. Dữ liệu bao gồm các ảnh X-quang cột sống được chụp bởi hai tư thế chính: nhìn thẳng và nhìn nghiêng. Toàn bộ dữ liệu gốc được lưu trữ dưới dạng y tế chuẩn DICOM (Digital Imaging and Communications in Medicine). Bộ dữ liệu có tổng cộng $10,466$ ảnh X-quang (gồm $8,389$ ảnh huấn luyện và $2,077$ ảnh kiểm tra) từ khoảng $5,000$ bệnh nhân. Các tổn thương được chia thành nhiều loại bệnh lý khác nhau, mỗi hình ảnh trong bộ dữ liệu đã được gán nhãn thủ công bởi các bác sĩ chẩn đoán hình ảnh giàu kinh nghiệm.



Hình 6: Ví dụ về các ảnh X-quang với chú thích của các bác sĩ chẩn đoán, với ảnh bất thường được đánh dấu bằng các hộp giới hạn.

3.3 Tiền xử lí dữ liệu

Bộ dữ liệu gốc gồm các ảnh định dạng DICOM với dung lượng lớn và các dải giá trị phức tạp. Đồ án thực hiện thực hiện tiền xử lí dữ liệu để phù hợp với các mô hình đã đề cập. Quá trình chuyển đổi ảnh định dạng DICOM sang ảnh mức xám diễn ra qua các bước sau. Đầu tiên, ưu tiên sử dụng bảng tra cứu giá trị nếu có, giúp hiện đúng cửa sổ quan sát giải phẫu xương. Trong trường hợp không áp dụng được bước trên, ảnh sẽ được xử lí bằng kỹ thuật cân bằng lược đồ xám để tăng cường độ tương phản tổng thể của ảnh. Cuối cùng, thực hiện đảo ngược màu đối với các ảnh có thuộc tính quang học MONOCHROME1 về chuẩn MONOCHROME2 thông dụng trong chẩn đoán hình ảnh. Sau quá trình này, ảnh được chuẩn hóa về khoảng giá trị $[0, 255]$ và lưu dưới định dạng PNG.

3.4 Phương pháp huấn luyện

Tập dữ liệu huấn luyện bao gồm 8,389 ảnh (4,260 ảnh bình thường và 4,129 ảnh bất thường) được sử dụng để tối ưu hóa tham số cho hai mô hình học sâu. Quy trình huấn luyện được thiết lập riêng biệt cho từng tác vụ bằng hai mô hình phân loại và phát hiện.

3.4.1 Huấn luyện mô hình phân loại

Mô hình DenseNet-201 [1] được sử dụng để phân loại nhị phân (Bình thường/Bất thường) cho toàn bộ ảnh X-quang đầu vào. Dữ liệu ở giai đoạn này để chia thành hai phần là tập huấn huấn luyện gồm 7130 ảnh và tập kiểm định gồm 1259 ảnh. Ảnh đầu vào được thay đổi kích thước về 224×224 pixels để phù hợp với kiến trúc mạng. Các giá trị pixel được chuẩn hóa, sau đó dữ liệu được tăng cường bằng các thao tác sau: Xoay, lật ngang, trượt, biến dạng, cân bằng lược đồ màu, thay đổi độ sáng và tương phản. Mô hình được huấn luyện sử dụng hàm mất mát Binary Cross-Entropy và thuật toán tối ưu SGD. Trong quá trình kiểm định mô hình, một ảnh được xem là bất thường nếu xác suất của bất thường lớn hơn một ngưỡng tối ưu. Thực tế, đồ án sẽ xác định ngưỡng tối ưu cho quá trình phân lớp bằng cách tối ưu chỉ số Youden's Index [6]: $J(c) = q(c) + r(c) - 1$, trong đó độ nhạy q và độ đặc hiệu r là các hàm theo giá trị ngưỡng c.

3.4.2 Huấn luyện mô hình phát hiện

Đối với giai đoạn định vị tổn thương, đồ án sẽ định vị 7 loại bệnh quan trọng: gai xương, hẹp khe đĩa đệm, vật liệu cấy ghép phẫu thuật, hẹp lỗ liên hợp, trượt đốt sống

thắt lưng, sụp đốt sống, và các tổn thương khác. Do thiếu hụt số mẫu dương tính, các bệnh còn lại được xem là "các tổn thương khác". Mô hình Sparse R-CNN [2] được triển khai trong phần này. Mạng được huấn luyện để xác định tổn thương cột sống sử dụng thuật toán tối ưu AdamW. Trong giai đoạn học, hàm mất mát hồi quy hộp giới hạn và hàm mất mát phân loại ở cấp vùng được tối thiểu hóa đồng thời.

3.5 Cơ chế kết hợp quyết định

Mặc dù mô hình phát hiện Sparse R-CNN [2] có khả năng định vị chính xác tổn thương với độ chính xác cao, nhưng trong thực tế các mô hình này thường gặp vấn đề tỉ lệ dương tính giả cao, tức là nhận diện nhầm các nhiễu ảnh hoặc cấu trúc xương bình thường thành bệnh lý. Ngược lại, mô hình phân loại DenseNet-201 [1] có khả năng khai thác thông tin ngữ cảnh toàn cục tốt hơn, đặc biệt trong việc nhận diện trạng thái bình thường của ảnh. Do đó, phần này đề xuất một cơ chế quyết định nhằm tận dụng ưu điểm của cả hai mô hình. Trước khi thực hiện kết hợp, ngưỡng quyết định c^* cho mô hình phân loại cần được tối ưu hóa trên tập kiểm định bằng chỉ số Youden's Index [6] nhằm cân bằng giữa độ nhạy và độ đặc hiệu. Với ảnh đầu vào $\mathbf{I}$, đầu ra của giai đoạn phân lớp biểu diễn xác suất của ảnh được xem là bất thường, được ký hiệu là $P(\text{abnormal}|\mathbf{I})$. Để cực đại hóa hiệu quả của bộ phát hiện, đồ án đề xuất một cơ chế quyết định kết hợp hai mô hình phân loại và phát hiện. Cho bất kỳ $\mathbf{I}$ xác suất dự đoán $P(\text{abnormal}|\mathbf{I}) \geq c^*$, toàn bộ kết quả dự đoán của bộ phát hiện đều được giữ lại. Với trường hợp $P(\text{abnormal}|\mathbf{I}) < c^*$, chỉ những hộp giới hạn dự đoán có độ tin cậy lớn hơn 0.5 đều được giữ lại.

4 Thực nghiệm và kết quả

4.1 Cài đặt môi trường và thực thi chi tiết

Đồ án được triển khai bằng ngôn ngữ Python, sử dụng các thư viện học sâu và xử lý ảnh phổ biến trong quá trình huấn luyện cũng như đánh giá mô hình. Tất cả các giai đoạn đều được triển khai và huấn luyện bằng thư viện PyTorch (phiên bản 2.1) trên hệ thống máy với NVIDIA RTX4050 GPU. Đối với giai đoạn huấn luyện mô hình phân lớp ảnh, tất cả ảnh huấn luyện được chuyển về ảnh 224×224 pixels và được chuẩn hóa theo phân phối trung bình và độ lệch chuẩn của tập dữ liệu ImageNet. Mạng DensetNet-201 [1] được huấn luyện từ đầu đến cuối sử dụng thuật toán tối ưu với 23595 lần lặp, sử dụng batch sizes 32 và tốc độ học 0.001. Ở quá trình huấn luyện mô hình phát hiện ảnh, toàn bộ ảnh huấn luyện được chuyển về ảnh 640 pixels, sau đó gom nhóm vào các batch nhỏ gồm 4 ảnh. Mô hình Sparse R-CNN [2] được khởi tạo trọng số huấn luyện sẵn, được huấn luyện trên $50,000$ vòng lặp bằng thuật toán tối ưu AdamW, với tốc độ học giảm 10 lần tại vòng lặp thứ $40,000$ và $45,000$.

4.2 Các chỉ số đánh giá

Phần này báo cáo hiệu suất của hai mô hình đã đề cập ở trên. Hiệu suất mô hình phân loại được đánh giá bằng các chỉ số AUROC, độ nhạy, độ đặc hiệu và F1-score. Khoảng tin cậy 95% (95% confidence interval, CI) được ước lượng bằng phương pháp bootstrap với 1000 lần lặp cho mỗi chỉ số. Trong mỗi lần bootstrap, một tập dữ liệu mới được tạo ra bằng cách lấy mẫu có hoàn lại từ tập gốc [7]. Đối với bài toán phát hiện, mô hình được đánh giá bằng chỉ số mAP@0.5 [4] theo chuẩn PASCAL VOC. Với mỗi lớp, trung bình precision (AP) được tính là trung bình của 101 giá trị precision ứng với recall từ 0 đến 1, và mAP@0.5 [4] là trung bình AP trên tất cả

các lớp tổn thương.

4.3 Kết quả thực nghiệm

Phần này sẽ trình bày về phần hiệu suất riêng của hai mô hình DenseNet-201 [1] và Sparse R-CNN [2] đã huấn luyện.

4.3.1 Hiệu suất mô hình phân lớp

Thực hiện đánh giá trên tập kiểm tra gồm 2,077 ảnh, tại ngưỡng tối ưu 0.7568 mô hình DenseNet-201 [1] cho kết quả AUROC 87.07% (95% CI 85.65, 88.58), độ nhạy 73.88% (95% CI 71.23, 76.60%), độ đặc hiệu 82.74% (95% CI 80.40, 84.91) và F1-score 76.85 (95% CI 74.82, 78.87%). Bảng 1 hiển thị kết quả hiệu suất của mô hình phân lớp trên tập kiểm tra.

Bảng 1: Hiệu suất phân loại của mô hình DenseNet-201 trên tập kiểm tra (phần trăm, CI 95%).

Mô hình	AUROC	Độ nhạy	Độ đặc hiệu	F1-score
DenseNet-201 [1]	86.98 (85.40, 88.54)[(*)]	78.06 (75.56, 80.57)	80.67 (78.30, 82.85)	78.59 (76.73, 80.52)

[(*)] CI: khoảng tin cậy 95% được ước lượng bằng phương pháp bootstrap với 1 000 lần lặp.

4.3.2 Hiệu suất của mô hình phát hiện

Trung bình Precision (AP) của mỗi tổn thương được phát hiện bởi mô hình Sparse R-CNN được trình bày trong bảng 2. Kết quả cho thấy mô hình Sparse R-CNN đạt hiệu năng cao nhất đối với tổn thương vật liệu cấy ghép phẫu thuật (TT3), trong khi hiệu năng thấp nhất ở nhóm các tổn thương khác (TT7).

Bảng 2: Hiệu suất của mô hình phát hiện trên tập kiểm tra

Mô hình	TT1[(*)]	TT2	TT3	TT4	TT5	TT6	TT7	mAP@0.5
Sparse R-CNN [2]	8.56	4.44	33.46	6.49	8.10	15.92	0.75	24.05

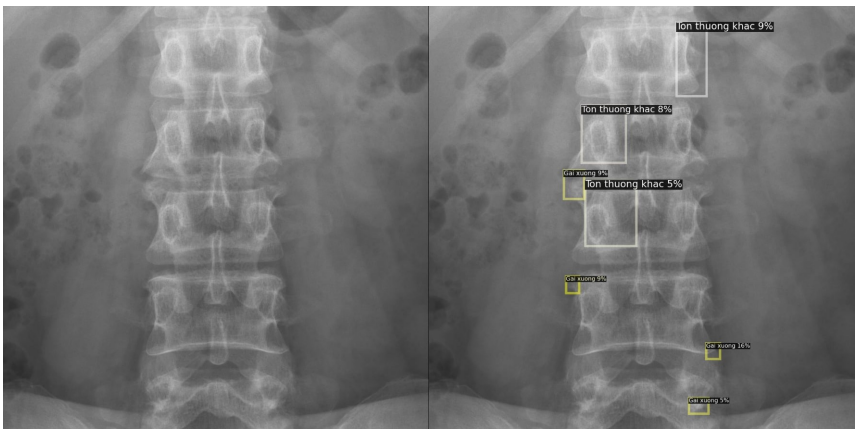
[(*)] TT1, TT2, TT3, TT4, TT5, TT6, TT7 đại diện cho các loại tổn thương gai xương, hẹp khe đĩa đệm, vật liệu cấy ghép phẫu thuật, hẹp lỗ liên hợp, trượt đốt sống thắt lưng, sụp đốt sống, và các tổn thương khác.

4.3.3 Hiệu suất của mô hình kết hợp

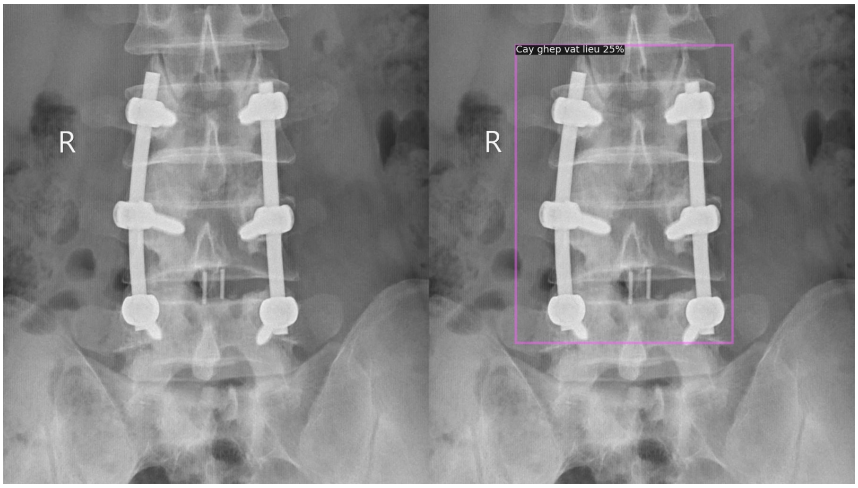
Hiệu suất của mô hình kết hợp so với mô hình phát hiện được trình bày ở bảng 3. Kết quả cho thấy mô hình kết hợp giữa hai mạng phân lớp và mạng phát hiện thông qua cơ chế quyết định đầu ra được đề xuất ở trên cho hiệu suất tốt hơn ở việc phát hiện toàn bộ các loại tổn thương, ngoại trừ TT4 và TT5.

Bảng 3: Hiệu suất của mô hình kết hợp so với mô hình phát hiện trên tập dữ liệu kiểm tra

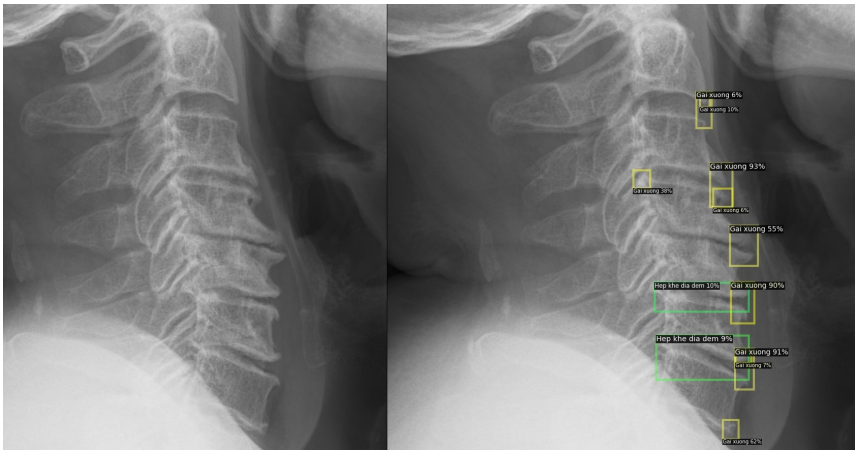
Mô hình	TT1	TT2	TT3	TT4	TT5	TT6	TT7	mAP@0.5
Sparse R-CNN [2]	8.562	4.444	33.466	6.499	8.106	15.928	0.750	24.051
Mô hình kết hợp	8.610	5.303	33.636	6.314	7.613	16.247	0.809	24.273



(a) Thoái hóa thắt lưng



(b) Cấy ghép vật liệu



(c) Thoái hóa cổ

Hình 7: Trực quan hóa kết quả phát hiện tổn thương trên tập kiểm tra. Mỗi cặp hình bao ảnh ảnh đầu vào (bên trái) và ảnh kết quả với nhãn lớp được đánh dấu bằng các hộp giới hạn cùng với độ tin cậy. (a) Phát hiện gai xương và tổn thương khác vùng thắt lưng. (b) Phát hiện vật liệu cấy ghép. (c) Phát hiện đa tổn thương (gai xương, hẹp khe đĩa đệm) vùng cổ.

5 Kết luận

Đồ án này đã nghiên cứu và xây dựng hệ thống phân loại ảnh bất thường và phát hiện các tổn thương trên ảnh X-quang cột sống bằng cách kết hợp hai mô hình DenseNet [1] và Sparse R-CNN [2]. Đồng thời, đồ án cũng đã đề xuất ra cơ chế kết hợp hai mô hình dựa trên ngưỡng tối ưu Youden's Index [6] để tối đa hóa hiệu suất của cả cả hệ thống. Bên cạnh những kết quả trên. Tuy nhiên, trong suốt quá trình thực hiện đồ án vẫn còn nhiều hạn chế về nhiều phương diện. Trước hết là về mặt phần cứng, mặc dù đồ án đã huấn luyện mô hình phân loại ở mức khá tốt và cho kết quả AUROC là 87.07% (95% CI 85.65, 88.58, trong khi ở mô hình phát hiện lại không đáp ứng đủ về mặt cấu hình máy. Cụ thể, với việc gom thành các nhóm nhỏ gồm 4 ảnh cho mỗi lần huấn luyện là khá ít. Mặc dù mô hình được khởi tạo với trọng số được tiền huấn luyện trên tập có sẵn, đồ án này vẫn được xem là huấn luyện từ đầu. Hơn thế nữa, bộ dữ liệu VinDR-SpineXR [3] gặp vấn đề mất cân về mặt dữ liệu giữa các tổn thương, các loại tổn thương hiếm với các đặc trưng khó chiếm số lượng ít, khiến mô hình khó học được đặc trưng phân biệt, dẫn đến hiệu suất phát hiện thấp hơn so với các lớp có đặc trưng thị giác rõ ràng. Một vấn đề khác là dữ liệu y tế trước khi đưa vào mô hình vẫn chưa xử lý bằng các phương pháp chuẩn theo y tế. Đó là lý do tại sao mô hình phát hiện cho ra kết quả mAP@0.5 là 24.051. Mặc khác, việc sử dụng mô hình kết hợp hai mạng phân lớp và mạng phát hiện thông qua cơ chế quyết định đầu ra phần nào cải thiện được mô hình phát hiện thuần túy. Trong các nghiên cứu tương lai, nhóm đề xuất việc phát triển và mở rộng bộ dữ liệu về cả số lượng lẫn quy mô bao gồm các tổn thương hiếm. Ngoài ra, có thể xem xét việc cải tiến kiến trúc mô hình phát hiện thông qua việc sử dụng các mạng xương sống hiện đại hơn kết hợp cơ chế học đa tỉ lệ nhằm tăng khả năng nhận diện các tổn thương nhỏ. Về mặt ứng dụng, hệ thống được đề xuất có tiềm năng hỗ trợ bác sĩ trong việc sàng lọc ban đầu các bất thường trên ảnh X-quang cột sống, giúp giảm tải khối lượng công việc và tăng độ chính xác trong chẩn đoán. Tóm lại, mặc dù vẫn còn tồn tại những hạn chế nhất định, đồ án đã phát triển ra hệ thống hỗ trợ chẩn đoán tự động trên ảnh X-quang cột sống. Trong tương lai, với sự cải thiện về dữ liệu và cả phần cứng, tài nguyên tính toán và phương pháp học sâu, hướng tiếp cận này sẽ mang lại giá trị thiết thực trong thực hành lâm sàng.

Tài liệu tham khảo

[1] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[2] P. Sun, R. Zhang, Y. Jiang, T. Kong, C. Xu, W. Zhan, H. Li, and P. Luo, "Sparse r-cnn: End-to-end object detection with learnable proposals," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[3] H. H. Pham, H. Nguyen Trung, and H. Q. Nguyen, "Vindr-spinexr: A large annotated medical image dataset for spinal lesions detection and classification from radiographs (version 1.0.0)." PhysioNet, 2021. Dataset accessed via PhysioNet.

[4] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.

[5] H. T. Nguyen, H. H. Pham, N. T. Nguyen, H. Q. Nguyen, T. Q. Huynh, M. Dao, and V. Vu, "Vindr-spinexr: A deep learning framework for spinal lesions detection and classification from radiographs," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021* (M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, eds.), (Cham), pp. 291–301, Springer International Publishing, 2021.

[6] W. J. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.

[7] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, vol. 57 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, 1993.