



---

## Phát hiện bất thường trên ảnh X-ray ngực

---

Khoa Toán - Cơ - Tin học  
Trường Đại học Khoa học Tự Nhiên, ĐHQGHN

*Sinh viên:*  
Lê Quý Công - 22001550  
Trần Hoàng Đạt - 22001558

*Giảng viên hướng dẫn:*  
TS. Cao Văn Chung

Hà Nội - 2025

## Tóm tắt

Dự án triển khai hai mô hình phát hiện đối tượng là YOLOv5 và RetinaNet trong bài toán phát hiện bất thường trên ảnh X-quang lồng ngực. YOLOv5 mang lại tốc độ suy luận nhanh, trong khi RetinaNet sử dụng Focal Loss để xử lý hiệu quả mất cân bằng lớp. Kết quả cho thấy cả hai mô hình đều phát hiện tốt các tổn thương lớn nhưng gặp khó khăn với tổn thương nhỏ, chồi lấn hoặc ảnh chất lượng thấp. Những phân tích thu được giúp đề xuất các hướng cải thiện mô hình và hỗ trợ xây dựng hệ thống chẩn đoán hình ảnh tự động đáng tin cậy hơn.

# Mục lục

<b>1</b>	<b>Giới thiệu</b>	<b>7</b>
<b>2</b>	<b>Xử lý dữ liệu</b>	<b>8</b>
2.1	Mô tả bộ dữ liệu VinBigData Chest X-ray . . . . .	8
2.1.1	Tổng quan . . . . .	8
2.1.2	Cấu Trúc Bộ Dữ Liệu . . . . .	8
2.1.3	Các Nhấn Chú Thích . . . . .	8
2.2	Data Processing . . . . .	12
2.2.1	Chuyển đổi định dạng DICOM sang PNG . . . . .	12
2.2.2	Chuẩn hoá kích thước ảnh . . . . .	13
2.2.3	Data Augmentation . . . . .	13
<b>3</b>	<b>Triển khai mô hình YOLOv5</b>	<b>15</b>
3.1	Cơ sở lý thuyết . . . . .	15
3.1.1	Ý tưởng của YOLO . . . . .	15
3.1.2	Kiến trúc YOLOv5 . . . . .	15
3.1.3	Cơ chế Anchor Box . . . . .	16
3.1.4	Hàm mất mát (Loss Function) . . . . .	17
3.1.5	Ưu điểm khi áp dụng YOLOv5 cho dữ liệu X-ray . . . . .	18
3.2	Triển khai mô hình . . . . .	18
3.2.1	Thống kê dữ liệu huấn luyện . . . . .	18
3.2.2	Quá trình huấn luyện . . . . .	19
3.2.3	Phân tích kết quả theo từng lớp . . . . .	19
3.3	Dánh giá kết quả . . . . .	20
3.3.1	Phân tích Ma trận nhầm lẫn (Confusion Matrix) . . . . .	20
3.3.2	Dánh giá trực quan trên ảnh thực tế . . . . .	21
<b>4</b>	<b>Triển khai mô hình RetinaNet</b>	<b>23</b>
4.1	Kiến trúc RetinaNet . . . . .	23
4.1.1	Tổng quan . . . . .	23
4.1.2	Backbone với Feature Pyramid Network . . . . .	23
4.1.3	Anchor . . . . .	24
4.1.4	Mạng con (Subnetworks) . . . . .	25
4.1.5	Hàm Loss . . . . .	25
4.2	Triển khai mô hình . . . . .	26
4.2.1	Quá trình huấn luyện . . . . .	26
4.2.2	Dánh giá trực quan trên ảnh thực tế . . . . .	27

<b>5</b>	<b>Kết luận</b>	<b>29</b>
5.1	Tổng kết . . . . .	29
5.2	Hạn chế và thách thức . . . . .	29
5.3	Hướng phát triển . . . . .	30
5.4	Ý nghĩa thực tiễn . . . . .	30
5.5	Đóng góp và kết luận cuối . . . . .	30

# Danh sách hình vẽ

2.1	Minh họa ảnh với bất thường . . . . .	9
2.2	Phân bố các nhãn . . . . .	11
2.3	Phân bố các nhãn sau khi lọc bỏ nhãn No Finding . . . . .	12
3.1	Kiến trúc YOLOv5 . . . . .	16
3.2	Phân bố số lượng mẫu và đặc trưng không gian của các lớp dữ liệu . . . . .	18
3.3	Biểu đồ các chỉ số trong quá trình huấn luyện và kiểm thử . . . . .	19
3.4	Biểu đồ Precision-Recall cho từng lớp bệnh lý . . . . .	20
3.5	Ma trận nhầm lẫn trên tập kiểm thử . . . . .	21
3.6	Kết quả phát hiện tốt trên Test Batch 0 . . . . .	21
3.7	Kết quả trung bình trên Test Batch 1 (xuất hiện chồng lấn box) . . . . .	22
3.8	Các trường hợp khó trên Test Batch 2 (nhiều tổn thương phức tạp) . . . . .	22
4.1	Kiến trúc RetinaNet . . . . .	23
4.2	Kiến trúc FPN . . . . .	24
4.3	Các thành phần loss trong quá trình huấn luyện . . . . .	27
4.4	Bất thường Aortic enlargement . . . . .	28

# Danh sách bảng

# Chương 1

## Giới thiệu

Trong những năm gần đây, các mô hình học sâu (Deep Learning) đã đạt được những tiến bộ vượt bậc trong lĩnh vực thị giác máy tính, đặc biệt là trong các bài toán phát hiện đối tượng (object detection). Những tiến bộ này mở ra nhiều ứng dụng quan trọng trong y học, trong đó có nhiệm vụ phát hiện tự động các bất thường trên ảnh X-quang lồng ngực (Chest X-ray). Đây là một bài toán có ý nghĩa lớn trong thực tế, hỗ trợ bác sĩ chẩn đoán nhanh hơn, giảm tải công việc và phát hiện sớm các dấu hiệu bệnh lý nguy hiểm.

Ảnh X-quang lồng ngực là một trong những kỹ thuật chẩn đoán hình ảnh phổ biến nhất, nhưng việc đọc và diễn giải ảnh phụ thuộc nhiều vào kinh nghiệm chuyên gia. Một số tổn thương có kích thước nhỏ, xuất hiện ở các vùng có độ tương phản thấp hoặc chồng lấn nhau khiến quá trình đánh giá trở nên khó khăn. Điều này mở ra nhu cầu sử dụng các mô hình tự động hóa nhằm tăng tính chính xác, độ ổn định và khả năng phát hiện đa tổn thương trong điều kiện lâm sàng đa dạng.

Dự án này triển khai và đánh giá hai kiến trúc phát hiện đối tượng tiêu biểu là YOLOv5 và RetinaNet cho bài toán phát hiện bất thường trên ảnh X-ray lồng ngực. YOLOv5 đại diện cho nhóm mô hình một giai đoạn (one-stage detector) với tốc độ suy luận nhanh và mức độ chính xác cao, phù hợp với các hệ thống thời gian thực. Trong khi đó, RetinaNet, thông qua việc giới thiệu Focal Loss, giải quyết hiệu quả vấn đề mất cân bằng giữa các lớp và cho thấy khả năng phát hiện tốt hơn đối với các mẫu khó.

Kết quả thu được cho thấy cả hai mô hình đều có khả năng phát hiện các tổn thương lớn với độ chính xác cao, đặc biệt trong các trường hợp rõ ràng về mặt hình thái học. Tuy nhiên, vẫn tồn tại những hạn chế khi xử lý các ca bệnh phức tạp, các tổn thương nhỏ hoặc các ảnh có chất lượng không đồng nhất, qua đó nhấn mạnh tầm quan trọng của việc tối ưu hóa mô hình và cải thiện chất lượng dữ liệu.

Dự án này góp phần xây dựng nền tảng cho việc áp dụng các mô hình học sâu trong chẩn đoán hình ảnh y khoa, hướng tới các hệ thống hỗ trợ bác sĩ hoạt động hiệu quả, nhất quán và đáng tin cậy hơn trong thực hành lâm sàng.

## Chương 2

# Xử lý dữ liệu

### 2.1 Mô tả bộ dữ liệu VinBigData Chest X-ray

#### 2.1.1 Tổng quan

Bộ dữ liệu [VinBigData Chest X-ray Abnormalities Detection](#) là một bộ dữ liệu công khai được sử dụng trong cuộc thi trên nền tảng Kaggle. Bộ dữ liệu này xuất phát từ tập hợp dữ liệu VinDr-CXR ban đầu gồm 18.000 ảnh X-ray ngực được chuẩn bị bởi Viện Dữ Liệu Lớn Vingroup (VinBigData) và thu thập từ hai bệnh viện lớn tại Việt Nam: Bệnh viện 108 và Bệnh viện Đại học Y Dược Hà Nội trong khoảng thời gian từ 2018 đến 2020 [2].

Bộ dữ liệu được thiết kế để giải quyết vấn đề phát hiện và định vị các bất thường lồng ngực trên ảnh X-ray. Mục tiêu chính là phát triển một hệ thống hỗ trợ chẩn đoán tự động (CADe/CADx) giúp giảm tải áp lực lên các bác sĩ tại các bệnh viện lớn và cải thiện chất lượng chẩn đoán ở các khu vực nông thôn.

#### 2.1.2 Cấu Trúc Bộ Dữ Liệu

Bộ dữ liệu được chia thành hai phần chính:

- **Tập Huấn Luyện:** Gồm 15000 ảnh X-ray được chú thích. Các ảnh bao gồm cả trường hợp bình thường (không có bất thường) và các trường hợp bất thường với các vị trí được xác định bởi hộp giới hạn (bounding boxes). Mỗi ảnh trong tập huấn luyện được chú thích độc lập bởi 3 bác sĩ chuyên khoa khác nhau để đảm bảo tính đa dạng và độ tin cậy của các chú thích. Được chia thành tập Train và Valid.
- **Tập Kiểm Tra (Test):** Gồm 3.000 ảnh X-ray. Tuy nhiên, các ảnh trong tập này không có nhãn sẵn.

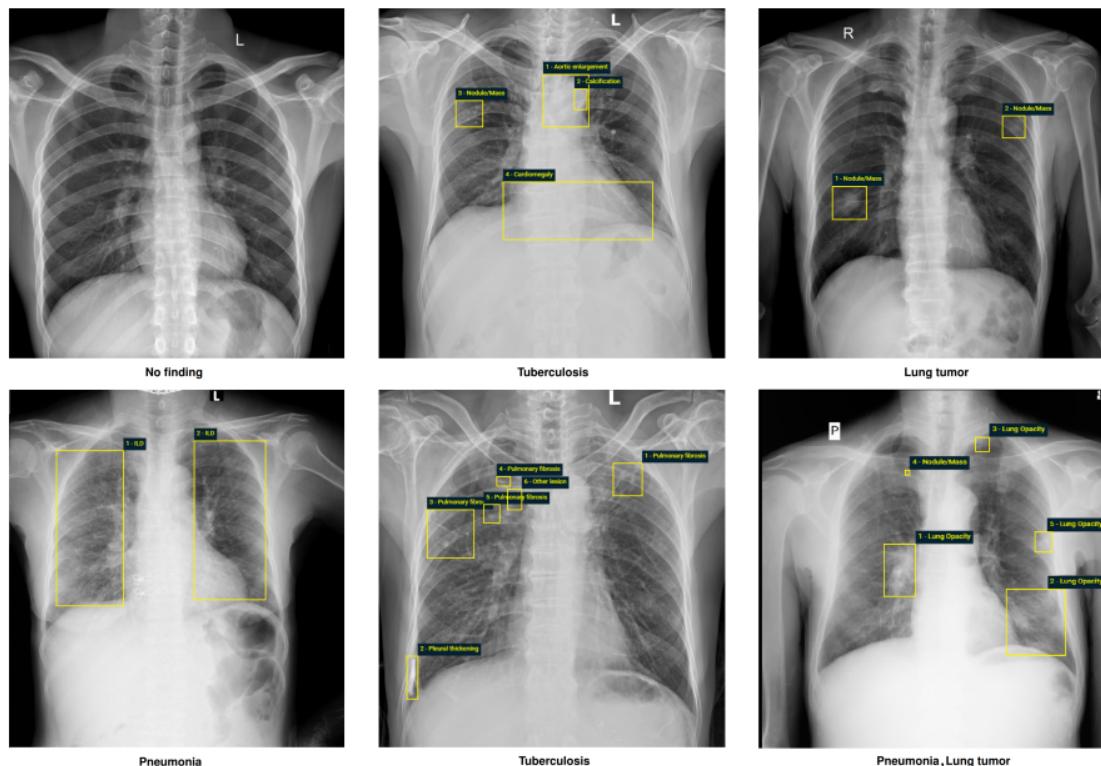
#### 2.1.3 Các Nhãn Chú Thích

Bộ dữ liệu bao gồm 15 lớp chia theo các biểu hiện bệnh lý thường gặp trên phim CXR. Các lớp này không chỉ đại diện cho bất thường mà còn phản ánh quá trình bệnh lý ở nhiều mức độ (cấu trúc – nhu mô – mạch máu – khoang màng phổi).

Các nhãn được chia thành nhiều nhóm bệnh học:

- **Nhóm bất thường tim - mạch:** Cardiomegaly, Aortic enlargement
- **Nhóm tổn thương nhu mô phổi:** Consolidation, Infiltration, Lung opacity, ILD, Pulmonary fibrosis

- Nhóm khồi/nốt: Nodule/Mass, Other lesion
- Nhóm bất thường khoang màng phổi: Pleural effusion, Pleural thickening, Pneumothorax



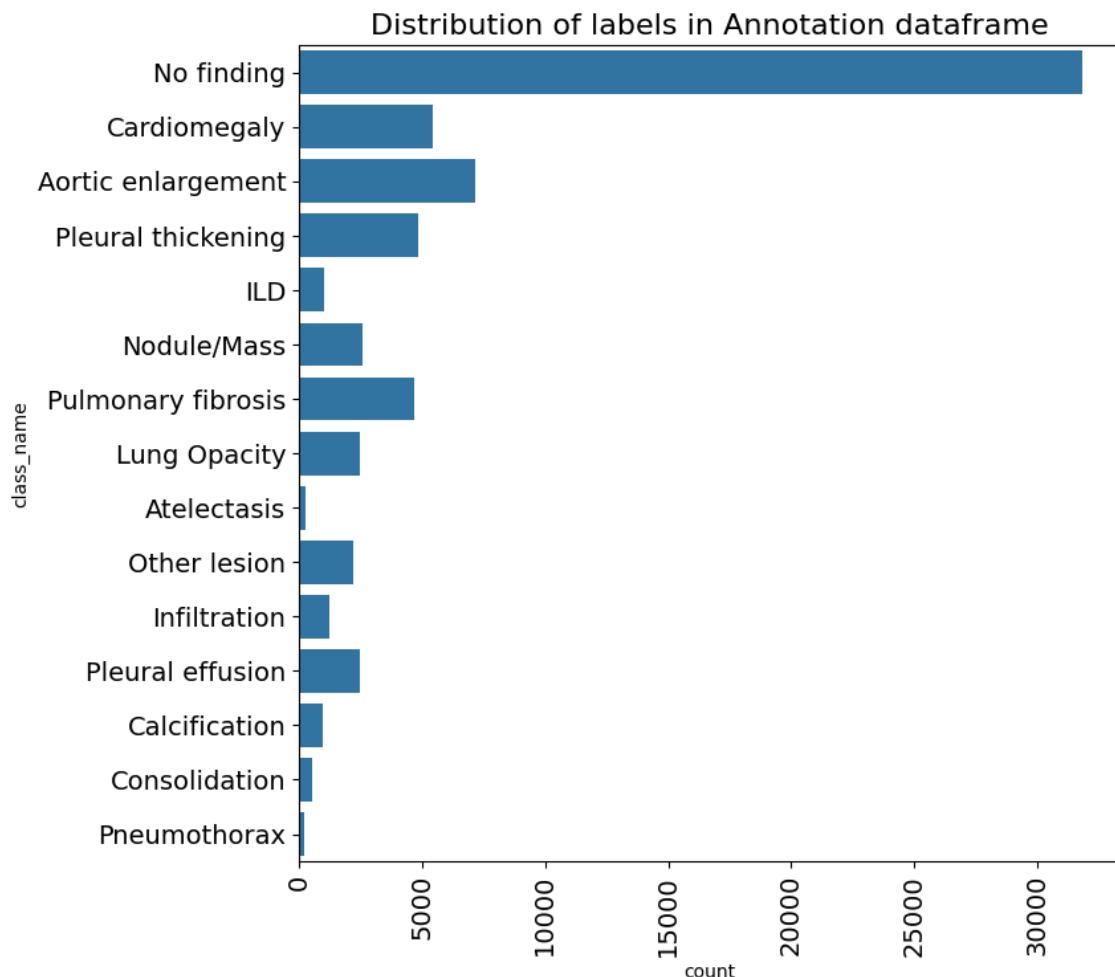
Hình 2.1: Minh họa ảnh với bất thường

Cụ thể:

- 0 - Aortic enlargement (Tăng kích thước động mạch chủ): Một bất thường trong thành động mạch chủ
- 1 - Atelectasis (Xẹp phổi): Sự sụp đổ của một phần phổi do giảm lượng không khí trong phế nang
- 2 - Calcification (Canxi hoá): Sự lắng đọng các muối calci trong phổi
- 3 - Cardiomegaly (Phì đại tim): Sự tăng kích thước của tim vượt quá giới hạn bình thường
- 4 - Consolidation (Tập trung): Các tổn thương dẫn đến tích tụ dịch hoặc máu trong các phế nang
- 5 - ILD (Interstitial Lung Disease - Bệnh phổi kẽ): Liên quan đến mô hő trợ của phổi gây ra mờ tinh hoặc mịn trong phổi
- 6 - Infiltration (Thấm nhập): Một chất bất thường tích tụ dần dần trong các tế bào hoặc mô cơ thể
- 7 - Lung Opacity (Mờ phổi): Bất kỳ mờ bất thường nào trong trường phổi

- 8 - Nodule/Mass (Nốt/Khối): Bất kỳ khối chiếm chỗ nào trong phổi, đơn lẻ hoặc nhiều
- 9 - Other lesion (Tổn thương khác): Các tổn thương không nằm trong danh sách các bất thường được xác định
- 10 - Pleural effusion (Tràn dịch màng phổi): Sự tích tụ bất thường của dịch trong không gian màng phổi
- 11 - Pleural thickening (Dày màng phổi): Bất kỳ sự dày lên nào liên quan đến màng phổi
- 12 - Pneumothorax (Tràn khí màng phổi): Sự hiện diện của khí trong khoang màng phổi
- 13 - Pulmonary fibrosis (Xơ hoá phổi): Sự tích tụ quá mức của mô xơ hóa trong phổi
- 14 - No finding (Không phát hiện): Biểu thị sự vắng mặt của tất cả các bất thường được liệt kê ở trên

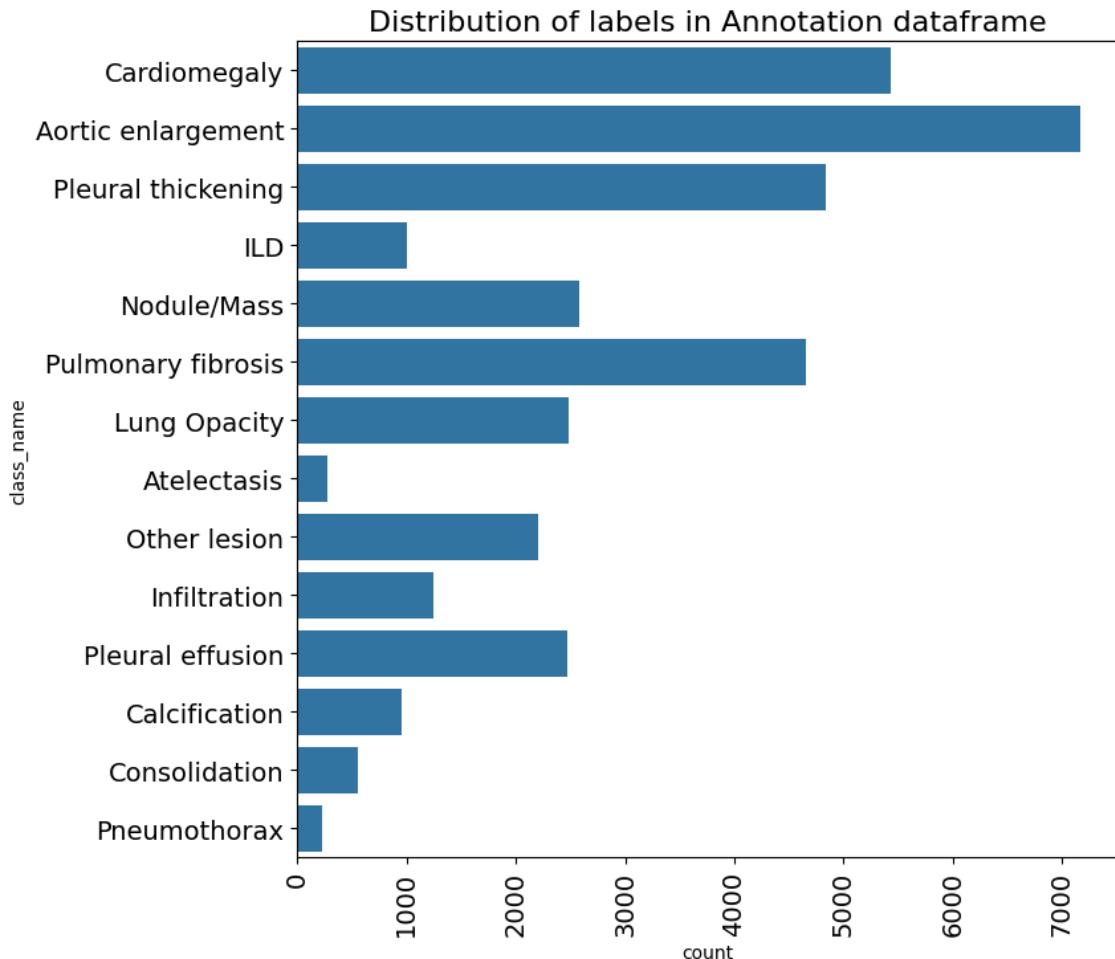
Sự đa dạng của bộ nhãn giúp mô hình học được các đặc trưng từ đơn giản đến phức tạp, bao gồm sự chồng lấp (overlapping lesions), các tổn thương hiếm gặp và những vùng mờ khó phân biệt bằng mắt thường.



Hình 2.2: Phân bố các nhãn

Biểu đồ cho thấy sự mất cân bằng nghiêm trọng giữa các lớp bất thường trong bộ dữ liệu. Một số nhãn như Cardiomegaly, Pleural effusion, Aortic enlargement và Pulmonary fibrosis có số lượng mẫu rất lớn, trong khi các nhãn như Calcification, Atelectasis, Consolidation hay Pneumothorax xuất hiện rất ít. Điều này phản ánh đúng đặc tính bệnh lý ngoài đời thực, nhưng lại gây khó khăn cho mô hình vì dễ dẫn đến hiện tượng thiên lệch về lớp phổ biến.

Thực hiện lọc bỏ nhãn No Finding, ta vẫn thấy được sự mất cân bằng giữa các lớp.



Hình 2.3: Phân bố các nhãn sau khi lọc bỏ nhãn No Finding

## 2.2 Data Processing

Việc xử lý dữ liệu (Data Processing) đóng vai trò đặc biệt quan trọng trong các bài toán thị giác máy tính liên quan đến ảnh X-ray, nơi mà chất lượng tín hiệu đầu vào quyết định trực tiếp đến khả năng mô hình học được các đặc trưng bệnh lý. Với bộ dữ liệu VinBigData Chest X-ray, tôi xây dựng một pipeline xử lý bài bản nhằm chuẩn hóa ảnh, giảm nhiễu, tăng độ tương phản và cải thiện tính đa dạng của dữ liệu.

### 2.2.1 Chuyển đổi định dạng DICOM sang PNG

Các ảnh X-ray gốc được cung cấp dưới dạng DICOM – một định dạng tiêu chuẩn trong y tế, chứa cả dữ liệu pixel thô lẫn metadata quan trọng như thông tin bệnh nhân, thiết bị chụp, thông số tia X,... Mặc dù DICOM có ưu điểm lưu trữ đầy đủ thông tin y tế, việc sử dụng trực tiếp cho deep learning lại gặp nhiều hạn chế:

- Thời gian đọc ảnh lâu, đặc biệt với tập dữ liệu lớn.
- Ta chỉ cần ảnh dưới dạng ma trận số làm đầu vào cho quá trình huấn luyện.

- Quản lý metadata và đồng bộ hóa dữ liệu phức tạp hơn nhiều so với ảnh PNG hoặc JPEG.

Do đó, toàn bộ dữ liệu được chuyển sang định dạng PNG – định dạng nén không mất dữ liệu – vừa đảm bảo chất lượng ảnh, vừa giúp tăng tốc độ đọc/ghi. Việc này còn giúp dễ dàng tích hợp với các công cụ xử lý ảnh phổ biến và các framework deep learning hiện đại.

### 2.2.2 Chuẩn hoá kích thước ảnh

Một trong những vấn đề nổi bật của ảnh X-ray là độ phân giải rất cao, có thể lên tới hàng nghìn pixel mỗi chiều. Nếu đưa trực tiếp vào mô hình, điều này gây tốn kém tài nguyên tính toán, kéo dài thời gian huấn luyện và dễ gặp vấn đề thiếu bộ nhớ (memory overflow).

Để khắc phục, toàn bộ ảnh được resize về kích thước  $512 \times 512$ , đảm bảo:

- Mọi ảnh đầu vào đồng nhất về kích thước, giúp mô hình xử lý mượt mà và tránh lỗi shape mismatch.
- Giảm tải bộ nhớ GPU/CPU, cho phép tăng batch size khi huấn luyện, từ đó cải thiện hiệu quả tối ưu hóa.
- Vẫn giữ lại đủ chi tiết quan trọng, đặc biệt là các bất thường nhỏ như nốt mờ (nodule) hay dấu hiệu viêm phổi, giúp mô hình nhận diện chính xác.

Quy trình thay đổi kích thước (resize) được áp dụng nhất quán trên toàn bộ tập dữ liệu, đảm bảo tính đồng bộ trong huấn luyện và suy luận (inference), đồng thời làm giảm sự phụ thuộc vào kích thước gốc của các máy chụp khác nhau.

### 2.2.3 Data Augmentation

Một thách thức lớn khi huấn luyện mô hình trên ảnh y tế là sự khác biệt về chất lượng ảnh, góc chụp và điều kiện ánh sáng giữa các bệnh viện, cũng như sự chênh lệch về độ tương phản giữa các máy X-ray. Ngoài ra, tập dữ liệu thường chứa nhiều mẫu tương tự nhau, dễ dẫn đến hiện tượng overfitting nếu không tăng cường tính đa dạng.

Để khắc phục, các kỹ thuật augmentation đã được áp dụng nhằm mở rộng không gian dữ liệu và giúp mô hình tổng quát tốt hơn. Các phép biến đổi được áp dụng bao gồm:

- **CLAHE (Contrast Limited Adaptive Histogram Equalization):** kỹ thuật này tăng cường độ tương phản cục bộ trên từng vùng nhỏ của ảnh, giúp làm nổi bật các chi tiết mờ, chẳng hạn như các nốt nhỏ hoặc vùng tổn thương nhẹ mà mắt thường khó nhận thấy. CLAHE còn có khả năng mô phỏng nhiều điều kiện chiếu sáng khác nhau, từ đó giúp mô hình học được các đặc trưng bất thường bất chấp độ tương phản thay đổi giữa các máy X-ray. Trong thực tế, việc áp dụng CLAHE còn giúp giảm ảnh hưởng của các hiện tượng underexposure hoặc overexposure trong quá trình chụp.
- **Horizontal flip (lật ngang):** tạo ra các biến thể đối xứng của ảnh lồng ngực. Do cơ thể con người gần như đối xứng, việc lật ngang giúp mô hình học được các đặc trưng không phụ thuộc vào bên trái hay phải của cơ thể, đồng thời tăng gấp đôi kích thước hiệu quả của tập huấn luyện. Lưu ý rằng, với một số bất thường đặc trưng chỉ xuất hiện ở một bên (ví dụ một số loại tổn thương tim phổi), cần kết hợp nhãn phù hợp để tránh nhầm lẫn.

- **Xoay nhẹ (rotation):** mô phỏng các trường hợp bệnh nhân đứng lệch hoặc góc chụp không hoàn toàn thẳng. Việc xoay giới hạn trong khoảng nhỏ đảm bảo ảnh không bị biến dạng quá mức, vẫn giữ được các cấu trúc giải phẫu chính xác. Xoay nhẹ giúp mô hình học được các đặc trưng invariant với góc chụp, đặc biệt quan trọng khi áp dụng trên dữ liệu từ nhiều bệnh viện khác nhau.
- **Điều chỉnh độ sáng (brightness adjustment):** tăng hoặc giảm cường độ sáng của ảnh để mô phỏng các điều kiện chiếu tia X khác nhau. Kỹ thuật này giúp mô hình không bị lệ thuộc vào mức sáng cố định, từ đó nhận diện tốt hơn trên các ảnh quá sáng hoặc quá tối. Ngoài ra, điều chỉnh độ sáng còn hỗ trợ mô hình nhận biết chi tiết trong vùng tối, cải thiện khả năng phát hiện các bất thường nhỏ mà nếu chỉ dựa vào ảnh gốc, mô hình có thể bỏ sót.
- **Kết hợp nhiều phép biến đổi (compound augmentation):** đôi khi, việc áp dụng đồng thời hai hoặc ba kỹ thuật augmentation trên cùng một ảnh (ví dụ CLAHE + xoay nhẹ + điều chỉnh độ sáng) có thể tạo ra các biến thể gần với điều kiện thực tế hơn, giúp mô hình học được các đặc trưng phức tạp và bền vững hơn. Tuy nhiên, cần kiểm soát tỷ lệ áp dụng để tránh tạo ra các ảnh quá biến dạng, không còn giá trị y học.

Các kỹ thuật này kết hợp tạo ra một không gian dữ liệu ảo rộng lớn, giúp mô hình giảm overfitting, học được các đặc trưng quan trọng và trở nên linh hoạt khi áp dụng trên ảnh X-ray thực tế từ nhiều nguồn khác nhau.

## Chương 3

# Triển khai mô hình YOLOv5

### 3.1 Cơ sở lý thuyết

Nghiên cứu này lựa chọn mô hình YOLOv5 làm kiến trúc cho bài toán phát hiện bất thường trên ảnh X-ray lồng ngực. YOLOv5 là một trong những phiên bản phát triển mới của dòng mô hình *You Only Look Once*, nổi tiếng với khả năng phát hiện đối tượng nhanh, gọn và hiệu quả cao. Mặc dù không phải mô hình chính thức từ nhóm tác giả YOLO ban đầu, YOLOv5 do Ultralytics phát triển lại được cộng đồng sử dụng rộng rãi trong thực nghiệm nhờ tính dễ dùng, linh hoạt và hiệu suất vượt trội.

#### 3.1.1 Ý tưởng của YOLO

Khác với các phương pháp trước đây tách biệt việc sinh vùng đề xuất (*region proposal*) và phân loại như R-CNN, YOLO tiếp cận bài toán theo hướng *end-to-end*. Ảnh đầu vào được chia thành một lưới (*grid*), và tại mỗi ô lưới, mô hình dự đoán trực tiếp:

- **Tọa độ bounding box:** xác định vị trí và kích thước của đối tượng
- **Độ tin cậy (objectness score):** đánh giá khả năng xuất hiện đối tượng
- **Xác suất thuộc các lớp (class probabilities):** phân loại đối tượng

Nhờ cơ chế dự đoán duy nhất trong một lần lan truyền xuôi (*single forward pass*), YOLO đạt tốc độ rất cao, phù hợp với các ứng dụng cần thời gian thực hoặc xử lý số lượng lớn ảnh — điều đặc biệt hữu ích trong môi trường y tế.

#### 3.1.2 Kiến trúc YOLOv5

YOLOv5 được xây dựng theo cấu trúc ba phần chính:

##### Backbone

Sử dụng CSP-Darknet để trích xuất đặc trưng. Kiến trúc *Cross Stage Partial (CSP)* giúp mô hình học sâu hơn mà không tăng quá nhiều phép tính, đồng thời giảm trùng lắp gradient. CSP chia luồng đặc trưng thành hai phần:

- Một phần đi qua các khối tích chập
- Một phần được kết nối trực tiếp ở cuối

Cơ chế này giúp tăng khả năng học và giảm lượng tính toán, đặc biệt quan trọng khi xử lý ảnh y tế có độ phân giải cao.

## Neck

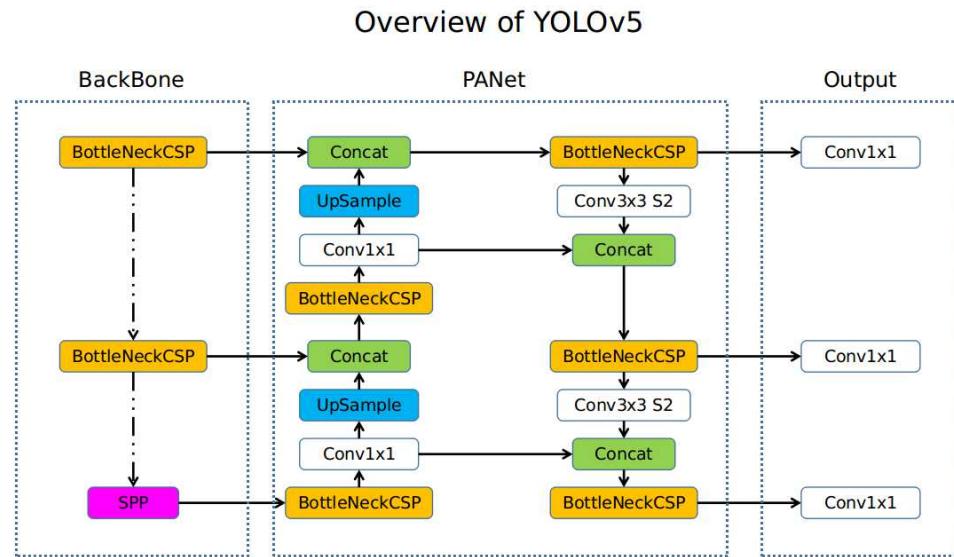
Sử dụng PANet (*Path Aggregation Network*) để kết hợp đặc trưng ở nhiều mức độ, giúp mô hình phát hiện tốt cả đối tượng lớn và nhỏ. Đối với ảnh X-ray, vùng tổn thương có thể rất nhỏ (như nốt phổi), nên cơ chế *Feature Pyramid* này rất quan trọng. PANet bổ sung:

- **Bottom-up path augmentation:** tăng cường truyền thông tin từ các lớp thấp đến lớp cao
- **Adaptive feature pooling:** tổng hợp đặc trưng từ tất cả các mức

## Head

Thực hiện dự đoán bounding box và class tại ba mức phân giải (*multi-scale*), giúp tăng độ chính xác trên nhiều kích thước bất thường khác nhau. Mỗi head dự đoán:

- Vị trí và kích thước box ( $x, y, w, h$ )
- Objectness score
- Xác suất các lớp bất thường



Hình 3.1: Kiến trúc YOLOv5

### 3.1.3 Cơ chế Anchor Box

YOLOv5 sử dụng *anchor box* và tự động tối ưu hóa chúng bằng thuật toán *k-means + genetic evolution*. Quá trình này diễn ra như sau:

- **K-means clustering:** Phân tích kích thước và tỷ lệ của các bounding box trong tập huấn luyện để tìm ra  $k$  anchor boxes đại diện tốt nhất
- **Genetic evolution:** Tinh chỉnh các anchor boxes thông qua thuật toán di truyền để tối ưu hóa IoU với *ground truth*

Điều này giúp mô hình thích nghi tốt hơn với hình dạng và kích thước bất thường trong ảnh X-ray, vốn thường có hình dạng không đồng nhất và xuất hiện ở nhiều vị trí khác nhau. Ví dụ:

- **Cardiomegaly** thường có box rộng và thấp
- **Nodule/Mass** thường có box nhỏ và gần vuông
- **Pleural effusion** có thể có nhiều hình dạng khác nhau

### 3.1.4 Hàm mất mát (Loss Function)

Hàm mất mát của YOLOv5 bao gồm ba phần:

#### Localization Loss

Sử dụng CIoU (*Complete IoU*) hoặc DIoU (*Distance IoU*) để giúp bounding box hội tụ nhanh và chính xác hơn. CIoU được định nghĩa:

$$\mathcal{L}_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v$$

Trong đó:

- $IoU$  là *Intersection over Union*
- $\rho$  là khoảng cách Euclidean giữa tâm hai box
- $c$  là đường chéo của box bao nhỏ nhất
- $v$  đo sự tương đồng về tỷ lệ khung hình
- $\alpha$  là trọng số cân bằng

#### Objectness Loss

Do mức độ mô hình tin rằng có đối tượng trong vùng dự đoán. Sử dụng Binary Cross Entropy:

$$\mathcal{L}_{obj} = -[y_{obj} \log(\hat{y}_{obj}) + (1 - y_{obj}) \log(1 - \hat{y}_{obj})]$$

#### Classification Loss

Phân loại bất thường theo các lớp tương ứng, cũng sử dụng Binary Cross Entropy cho từng lớp:

$$\mathcal{L}_{cls} = - \sum_{i=1}^C [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$$

#### Tổng hàm loss

$$\mathcal{L}_{total} = \lambda_{box} \mathcal{L}_{CIoU} + \lambda_{obj} \mathcal{L}_{obj} + \lambda_{cls} \mathcal{L}_{cls}$$

Với bài toán X-ray, việc tối ưu *localization* có ý nghĩa quan trọng, bởi nhiều bất thường có ranh giới mờ hoặc kích thước nhỏ. Việc áp dụng CIoU giúp cải thiện đáng kể khả năng định vị so với IoU thông thường.

### 3.1.5 Ưu điểm khi áp dụng YOLOv5 cho dữ liệu X-ray

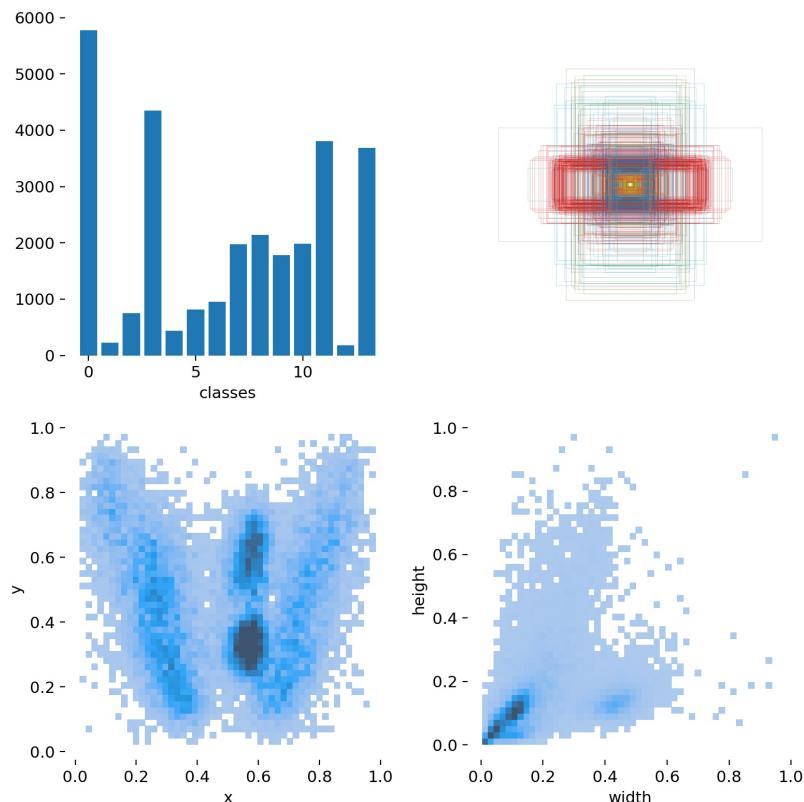
Việc sử dụng YOLOv5 trong bài toán phát hiện bất thường lồng ngực mang lại nhiều lợi thế:

- **Tốc độ nhanh:** Phù hợp xử lý số lượng lớn ảnh trong môi trường bệnh viện. Thời gian xử lý một ảnh chỉ khoảng 10-20ms trên GPU hiện đại
- **Hiệu suất cao:** Khả năng phát hiện tốt ở nhiều kích thước đối tượng nhờ cơ chế *multi-scale detection*
- **Dễ đào tạo và triển khai:** Thư viện Ultralytics tối ưu hoá giúp huấn luyện nhanh chóng và trực quan, hỗ trợ nhiều định dạng export (ONNX, TensorRT, CoreML)
- **Ôn định trên ảnh y tế:** Cơ chế *multi-scale* và *neck PANet* rất phù hợp để phát hiện nốt phổi nhỏ hoặc vùng mờ trong ảnh
- **Khả năng mở rộng:** Có nhiều biến thể (YOLOv5n, s, m, l, x) cho phép cân bằng giữa tốc độ và độ chính xác

## 3.2 Triển khai mô hình

### 3.2.1 Thống kê dữ liệu huấn luyện

Dữ liệu đầu vào bao gồm ảnh X-ray lồng ngực được gán nhãn với 14 loại bất thường.

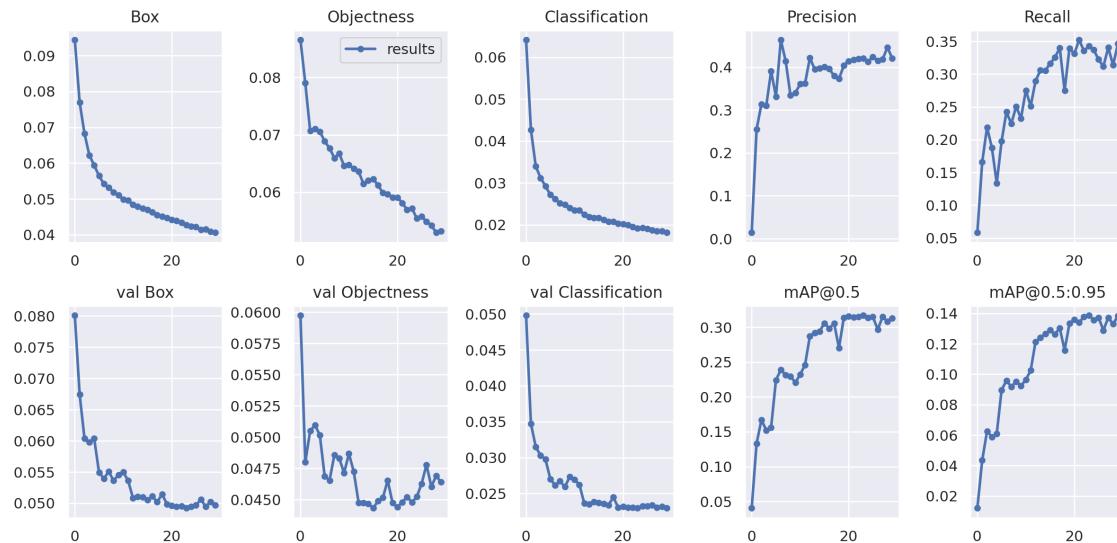


Hình 3.2: Phân bố số lượng mẫu và đặc trưng không gian của các lớp dữ liệu

Phân tích thống kê (Hình 3.2) cho thấy sự mất cân bằng dữ liệu đáng kể, một thách thức phổ biến trong y tế. Cụ thể, lớp "Không có bất thường" (Background) chiếm đa số với gần 6000 mẫu, trong khi các bệnh lý phổ biến như *Cardiomegaly*, *Consolidation* xuất hiện khoảng 3700-4300 lần. Ngược lại, các lớp hiếm như *Other lesion* chỉ có khoảng 200 mẫu. Về mặt không gian, các tổn thương tập trung chủ yếu ở vùng nhu mô phổi và trung thất, với kích thước biến thiên mạnh từ rất nhỏ (vôi hóa) đến chiếm toàn bộ phế trường.

### 3.2.2 Quá trình huấn luyện

Quá trình huấn luyện được giám sát chặt chẽ thông qua các chỉ số loss và metrics đánh giá (Hình 3.3).



Hình 3.3: Biểu đồ các chỉ số trong quá trình huấn luyện và kiểm thử

Kết quả cho thấy sự hội tụ ổn định của mô hình:

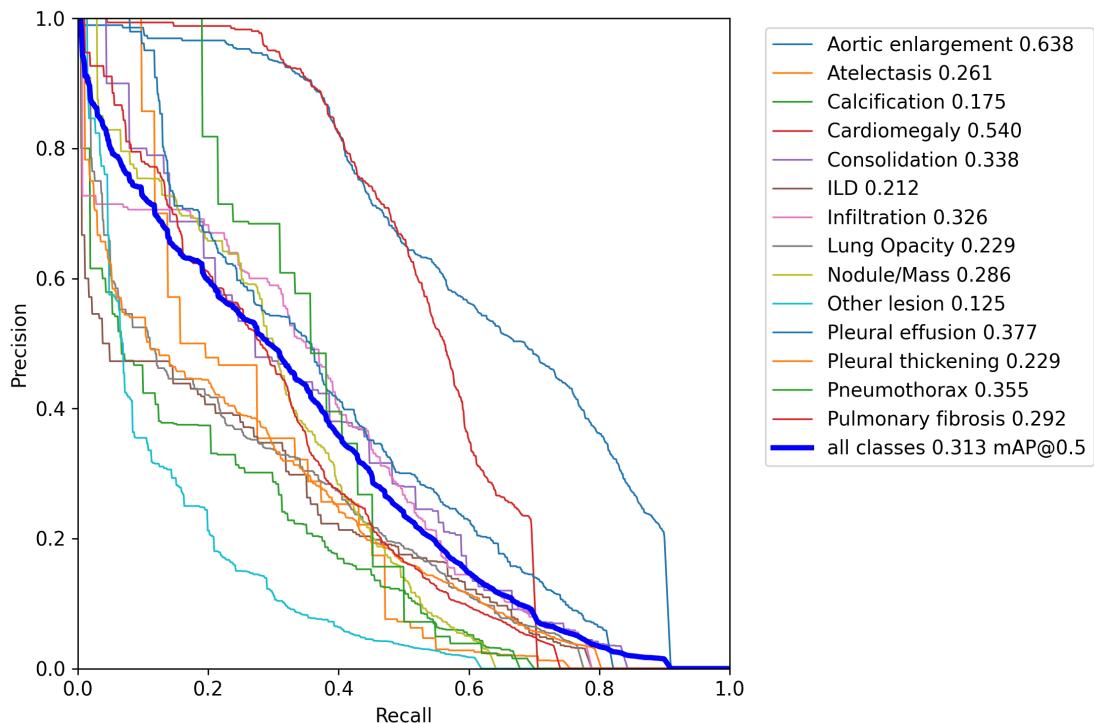
- Sự giảm thiểu hàm mất mát:** Box Loss giảm từ 0.09 xuống 0.04 và Objectness Loss giảm từ 0.08 xuống 0.05, chứng tỏ mô hình ngày càng hoàn thiện khả năng định vị và phát hiện đối tượng.
- Hiệu suất phân loại:** Classification Loss giảm mạnh xuống mức 0.02, đồng thời các chỉ số Precision và Recall trên tập huấn luyện tăng trưởng đều đặn, đạt lần lượt  $\sim 0.42$  và  $\sim 0.35$ .
- Độ chính xác tổng hợp:** Chỉ số mAP@0.5 trên tập validation đạt mức 0.31. Tuy nhiên, việc validation loss có xu hướng cao hơn training loss cho thấy sự xuất hiện của hiện tượng *overfitting* nhẹ, đòi hỏi các biện pháp điều chỉnh regularization trong tương lai.

### 3.2.3 Phân tích kết quả theo từng lớp

Phân tích biểu đồ Precision-Recall (Hình 3.4) cho thấy hiệu năng của mô hình phụ thuộc lớn vào đặc điểm hình thái của từng loại tổn thương:

- Nhóm hiệu suất cao (mAP > 0.5):** Bao gồm *Aortic enlargement* (0.638) và *Cardiomegaly* (0.540). Đây là các bất thường có kích thước lớn, vị trí cố định và ranh giới giải phẫu rõ ràng, tạo thuận lợi cho việc học đặc trưng.

- **Nhóm hiệu suất trung bình ( $0.2 < \text{mAP} < 0.4$ ):** Gồm *Pleural effusion*, *Consolidation*, *Infiltration*. Các tổn thương này thường biểu hiện dưới dạng vùng mờ, có ranh giới không sắc nét và dễ nhầm lẫn với các cấu trúc mạch máu phổi.
- **Nhóm khó phát hiện ( $\text{mAP} < 0.2$ ):** Các lớp như *Calcification* (0.175) hay *Other lesion* (0.125) có kết quả thấp nhất do kích thước quá nhỏ hoặc đặc điểm không đồng nhất, gây khó khăn cho cơ chế anchor box mặc định.



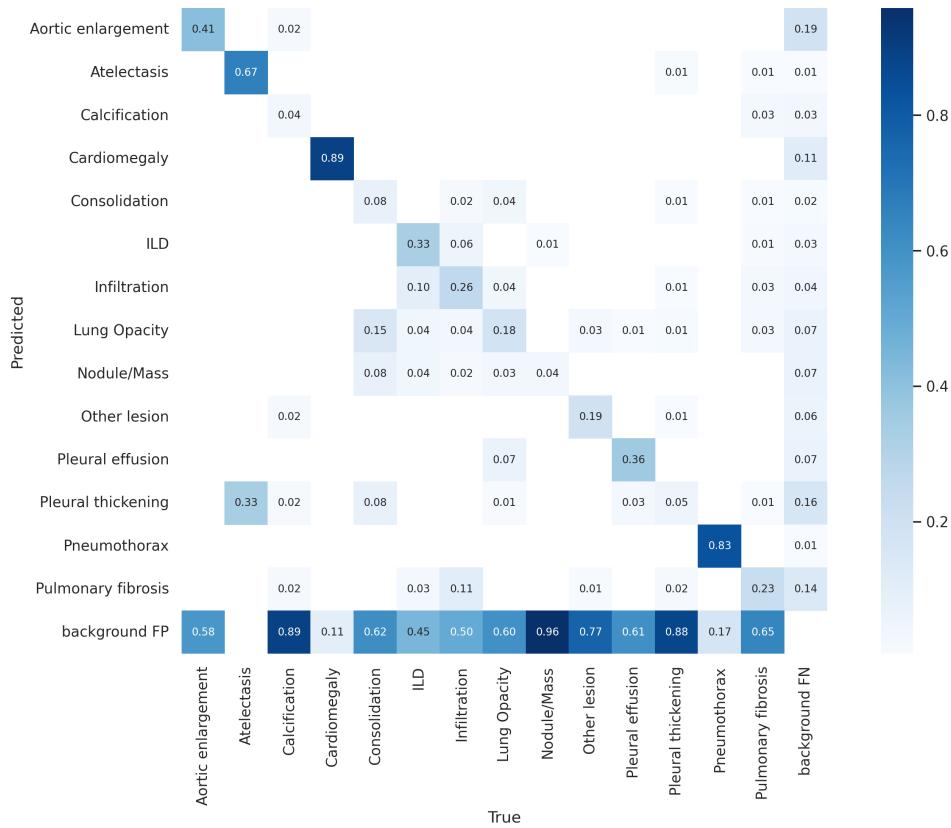
Hình 3.4: Biểu đồ Precision-Recall cho từng lớp bệnh lý

### 3.3 Đánh giá kết quả

#### 3.3.1 Phân tích Ma trận nhầm lẫn (Confusion Matrix)

Ma trận nhầm lẫn (Hình 3.5) cung cấp cái nhìn sâu hơn về các sai số của mô hình. Kết quả cho thấy mô hình đạt độ chính xác cao với các lớp đặc thù như *Cardiomegaly* và *Pneumothorax*.

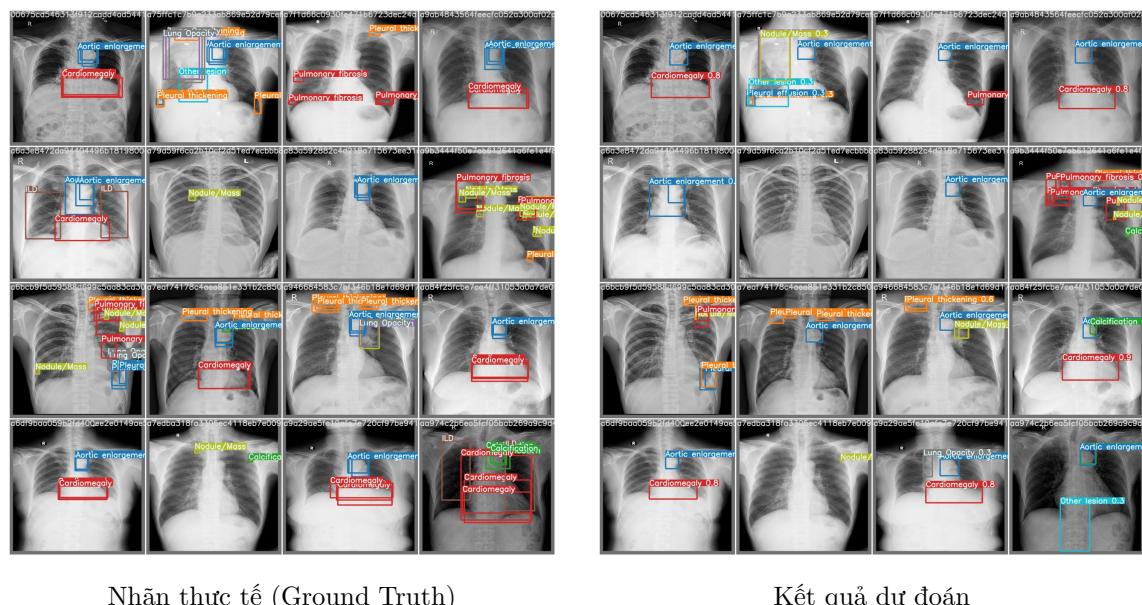
Tuy nhiên, sự nhầm lẫn đáng kể xảy ra giữa các cặp lớp có biểu hiện hình ảnh tương đồng, điển hình là giữa *Consolidation* và *ILD* (tỷ lệ nhầm lẫn  $\sim 0.33$ ), hoặc giữa *Infiltration* và *Lung Opacity*. Dáng chú ý, tỷ lệ dương tính giả (False Positives) từ nền (background) còn cao, cho thấy mô hình có xu hướng "nhạy cảm quá mức", đòi hỏi việc tinh chỉnh ngưỡng objectness hoặc áp dụng các kỹ thuật lọc hậu xử lý khắc khe hơn.



Hình 3.5: Ma trận nhầm lẫn trên tập kiểm thử

### 3.3.2 Đánh giá trực quan trên ảnh thực tế

Trường hợp dự đoán chính xác:

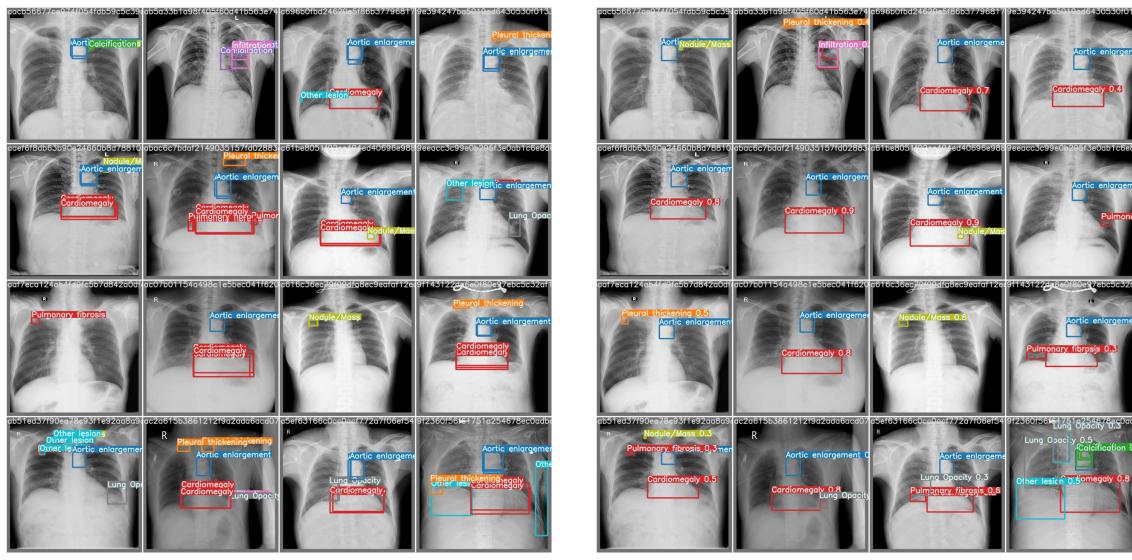


Hình 3.6: Kết quả phát hiện tốt trên Test Batch 0

Hình 3.6 minh họa khả năng của mô hình trong việc định vị chính xác các tồn

thương lớn. Các bounding box bao phủ tốt vùng tim to (*Cardiomegaly*) và quai động mạch chủ (*Aortic enlargement*) với độ tin cậy cao (0.6 - 0.9).

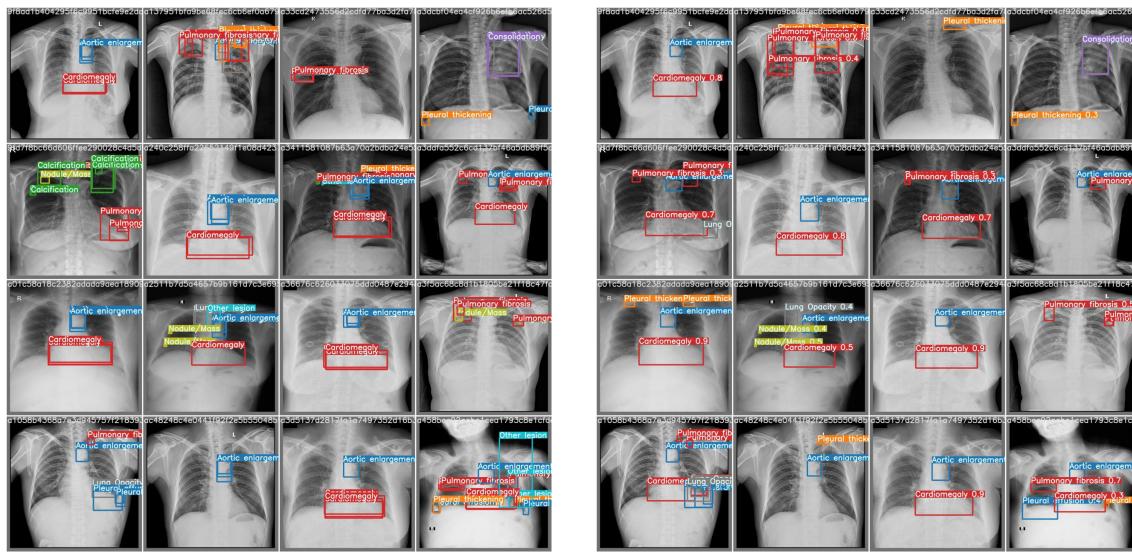
### Các trường hợp thách thức và sai sót:



Nhận thực tế

Kết quả dự đoán

Hình 3.7: Kết quả trung bình trên Test Batch 1 (xuất hiện chồng lấn box)



Nhận thực tế

Kết quả dự đoán

Hình 3.8: Các trường hợp khó trên Test Batch 2 (nhiều tổn thương phức tạp)

Dối với các trường hợp phức tạp hơn (Hình 3.7), mô hình gặp khó khăn trong việc tách biệt các tổn thương chồng lấn. Hiện tượng chồng chéo box (overlapping) xảy ra giữa *Pulmonary fibrosis* và *Cardiomegaly*. Ngoài ra, một số tổn thương nhỏ bị bỏ sót (False Negatives) hoặc bị nhận diện nhầm vị trí, đặc biệt là trong các ảnh có độ tương phản thấp. Các kết quả này (Hình 3.8) chỉ ra hướng cải thiện cần thiết về việc tăng cường dữ liệu huấn luyện cho các lớp hiếm và tinh chỉnh hàm mất mát để xử lý tốt hơn các ca bệnh khó.

## Chương 4

# Triển khai mô hình RetinaNet

RetinaNet (2017) do Facebook AI Research (FAIR) đề xuất, giải quyết thành công vấn đề này thông qua việc giới thiệu Focal Loss, giúp giảm mất cân bằng giữa các trường hợp dễ dự báo (foreground) và trường hợp khó dự báo, từ đó đạt hiệu năng vượt trội trong khi vẫn giữ tốc độ cao.

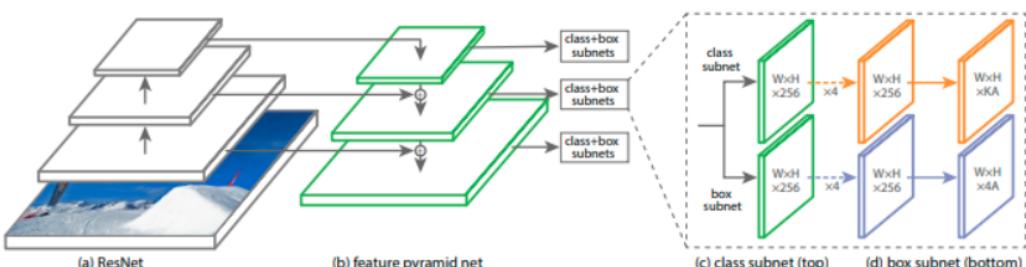
Mô hình RetinaNet ra đời nhằm giải quyết vấn đề các mô hình một giai đoạn (one-stage detectors) mặc dù nhanh nhưng độ chính xác thấp hơn nhiều so với mô hình hai giai đoạn (two-stage detectors) như Faster R-CNN tại thời điểm đó.

### 4.1 Kiến trúc RetinaNet

#### 4.1.1 Tổng quan

RetinaNet là mô hình one-stage detector, được cấu thành từ ba phần chính:

- Backbone: mạng CNN trích xuất đặc trưng (ResNet).
- Feature Pyramid Network (FPN) – giúp tổng hợp đặc trưng ở nhiều mức.
- Detection Subnets – hai mạng con:
  - Classification subnet (phân loại)
  - Regression subnet (dự đoán bounding box)

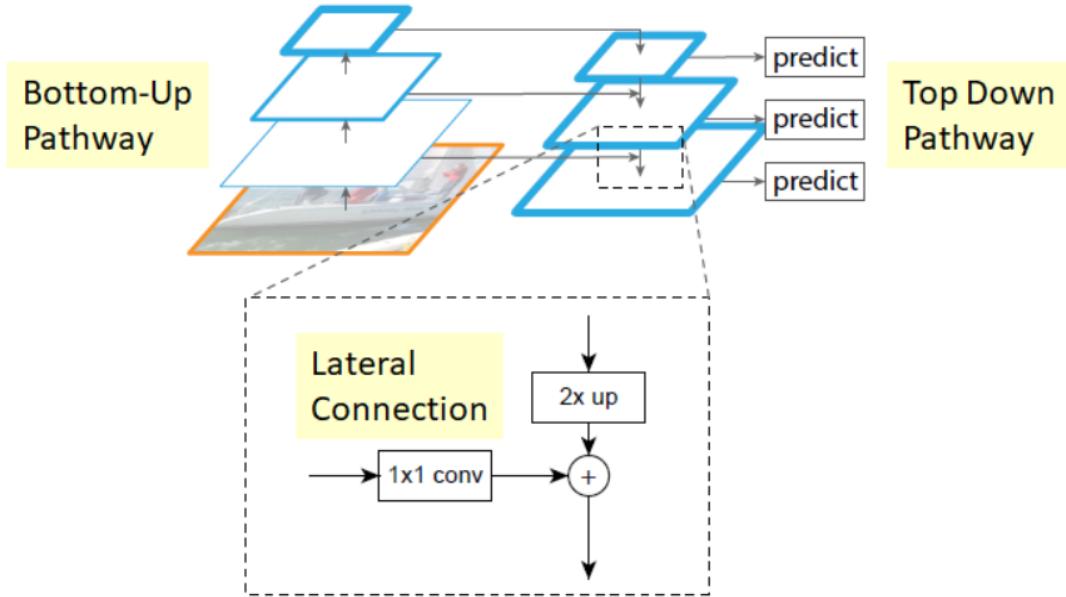


Hình 4.1: Kiến trúc RetinaNet

#### 4.1.2 Backbone với Feature Pyramid Network

Backbone thường sử dụng ResNet50 hoặc ResNet101 được huấn luyện trước trên tập ImageNet để trích xuất đặc trưng và trả về các feature map P3, P4, P5, P6, P7. Mỗi cấp độ của kim tự tháp (từ P3 đến P7) được sử dụng để phát hiện đối tượng ở một tỷ lệ khác nhau.

Mạng FPN (Feature Pyramid Network) sẽ tạo ra một multi-head dạng kim tự tháp bổ sung cho mạng tích chập tiêu chuẩn bằng một đường dẫn từ trên xuống (top-down pathway) và các kết nối bên (lateral connections), cho phép mạng xây dựng một tháp đặc trưng đa tỷ lệ (multi-scale feature pyramid) phong phú từ một hình ảnh đầu vào độ phân giải duy nhất.



Hình 4.2: Kiến trúc FPN

- Nhánh Bottom-Up là một mạng Convolutional Neural Network (ResNet) giúp tạo ra pyramid level các feature map theo kích thước giảm dần. Những feature map này sẽ được kết hợp với feature map cùng cấp nhánh Top Down.
- Nhánh Top-Down: Là một mạng upscaling các feature map theo kích thước mǔ cơ số 2. Như vậy mỗi một level của nhánh bên phải sẽ kết hợp với một level ở nhánh bên trái có cùng kích thước thông qua một phép cộng element-wise additional (cộng các phần tử ở cùng vị trí với nhau). Trước khi thực hiện phép cộng thì level nhánh bên trái được tích chập với feature map 1x1 để giảm thiểu chiều sâu (số channel). Mỗi một phép cộng sẽ trả ra một merge map như các ô predict trong hình.

#### 4.1.3 Anchor

RetinaNet sử dụng các anchor boxes (hộp neo) bắt biên tịnh tiến tại mỗi vị trí trên đặc trưng không gian. Tại mỗi cấp độ của feature pyramid, RetinaNet định nghĩa 9 hộp neo khác nhau tại mỗi điểm trên bản đồ đặc trưng, được xác định bởi:

- 3 tỷ lệ khung hình (aspect ratios):  $\{1 : 2, 1 : 1, 2 : 1\}$ ,
- 3 tỷ lệ kích thước (scales):  $\{2^0, 2^{1/3}, 2^{2/3}\}$  nhân với kích thước cơ sở tại mỗi cấp độ.

Tổng cộng, các anchor này được thiết kế để bao phủ các đối tượng trong khoảng kích thước từ 32 đến 813 pixel, đảm bảo khả năng phát hiện vật thể ở nhiều kích thước.

#### 4.1.4 Mạng con (Subnetworks)

RetinaNet gắn hai mạng con FCN nhỏ vào mỗi cấp độ FPN, với các tham số được chia sẻ trên tất cả các cấp độ kim tự tháp:

##### 1. Mạng con Phân loại (Classification Subnet):

- Dự đoán xác suất có mặt đối tượng cho mỗi vị trí không gian đối với A anchor và K lớp đối tượng.
- Mạng con này sử dụng bốn lớp tích chập  $3 \times 3$  (với kích hoạt ReLU), sau là một lớp tích chập  $3 \times 3$  cuối cùng với KA bộ lọc, kết thúc bằng kích hoạt sigmoid để xuất KA dự đoán nhị phân.

##### 2. Mạng con Hồi quy Hộp (Box Regression Subnet):

- Hồi quy từ mỗi hộp neo đến đối tượng ground-truth gần đó (nếu có).
- Kiến trúc giống hệt mạng con phân loại nhưng kết thúc bằng 4A đầu ra tuyến tính, dự đoán bù đắp tương đối giữa anchor và hộp ground-truth.

Mạng con Phân loại và Mạng con Hồi quy Hộp, mặc dù có cấu trúc chung, nhưng sử dụng các tham số riêng biệt.

#### 4.1.5 Hàm Loss

Hàm Loss tổng được thiết kế gồm 2 hàm Loss thành phần:

- Focal Loss
- Smooth L1

##### Focal Loss

Focal Loss được giới thiệu lần đầu với mạng RetinaNet nhằm giải quyết vấn đề mất cân bằng lớp [1]. Các mẫu âm dễ (mô hình đoán đúng gần như tuyệt đối) vẫn đóng góp vào loss làm chậm quá trình học, mô hình khó học mẫu dương vốn ít quan trọng hơn do ít xuất hiện.

Focal Loss giải quyết triệt để vấn đề này bằng cách **giảm tầm quan trọng** của các mẫu dễ và tập trung huấn luyện vào mẫu khó bằng cách thay đổi hàm Cross-Entropy:

$$CE(p_t) = -\log(p_t)$$

trong đó  $p_t$  là xác suất ước tính của mô hình cho lớp ground-truth.

Focal Loss (FL) thêm hệ số điều chỉnh  $(1 - p_t)^\gamma$  vào:

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t)$$

Tham số  $\gamma \geq 0$  được gọi là tham số tập trung (focusing parameter):

- Khi  $\gamma = 0$ , Focal Loss tương đương Cross-Entropy Loss
- Khi  $\gamma > 0$ , FL giảm trọng số tương đối của Loss cho các ví dụ đã được phân loại tốt ( $p_t > 0, 5$ )
- Khi một ví dụ bị phân loại sai (tức là  $p_t$  nhỏ), hệ số điều chỉnh gần bằng 1, do đó mất mát gần như không bị ảnh hưởng.

- Khi  $p_t \rightarrow 1$  (ví dụ được phân loại đúng với độ tin cậy cao), hệ số điều chỉnh tiến về 0, từ đó làm giảm trọng số mất mát, giúp mô hình tập trung học những ví dụ khó.

Trong thực tế mất cân bằng lớp dương – âm mạnh, RetinaNet sử dụng thêm hệ số cân bằng  $\alpha$ :

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t)$$

với:

- $\alpha_t = \alpha$  cho mẫu dương (positive)
- $\alpha_t = 1 - \alpha$  cho mẫu âm (negative)

Trong các thử nghiệm,  $\gamma = 2, 0, \alpha = 0, 25$  được tìm thấy là hoạt động tốt nhất.

### Smooth L1

Smooth L1 Loss được định nghĩa cho mỗi giá trị dự đoán và giá trị thật của box:

$$\text{Smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{nếu } |x| < 1 \\ |x| - 0.5 & \text{nếu } |x| \geq 1 \end{cases}$$

Trong đó:

$$x = t_{\text{pred}} - t_{\text{target}}$$

### Giải thích:

- Khi sai số nhỏ ( $|x| < 1$ )  $\rightarrow$  loss hoạt động như **L2**, giúp mô hình học mượt.
- Khi sai số lớn ( $|x| \geq 1$ )  $\rightarrow$  loss hoạt động như **L1**, giảm ảnh hưởng của *outliers*.

## 4.2 Triển khai mô hình

### 4.2.1 Quá trình huấn luyện

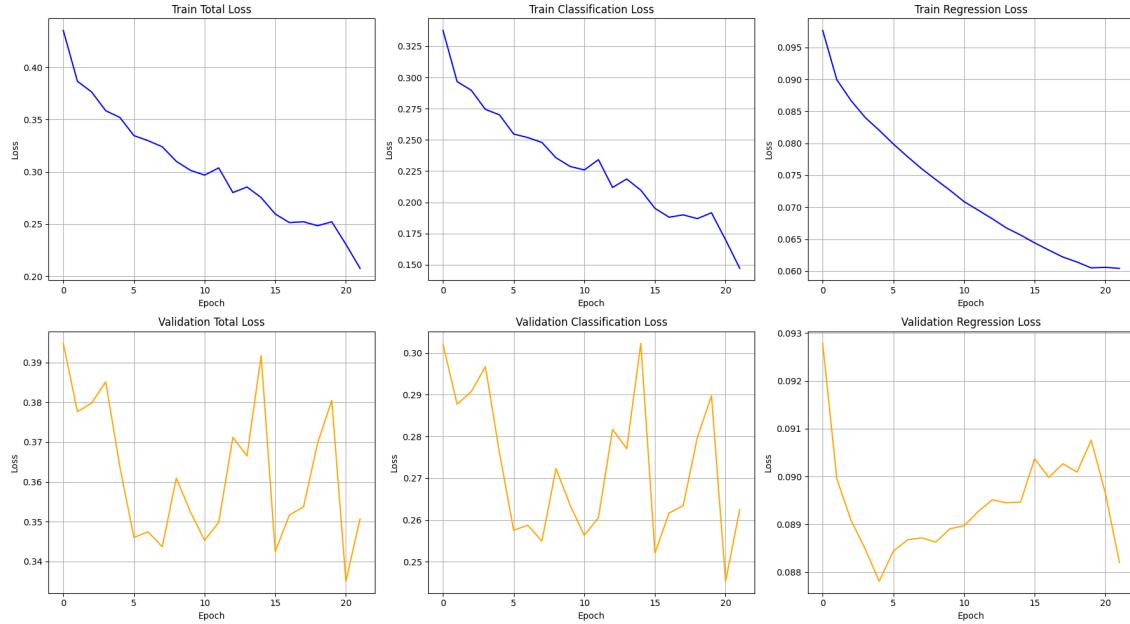
Mô hình RetinaNet được huấn luyện trên bộ dữ liệu VinBigData Chest X-ray với các tham số tối ưu đã được đề cập ở phần trước ( $\gamma = 2.0, \alpha = 0.25$ ). Quá trình huấn luyện được thực hiện trong 25 epoch với việc theo dõi chặt chẽ các hàm loss.

#### Phân tích đường cong Loss:

Hình 4.3 thể hiện sự biến thiên của các thành phần loss trong quá trình huấn luyện:

- **Train Total Loss:** Giảm dần từ xấp xỉ 0.43 xuống 0.21, cho thấy mô hình đang học hiệu quả từ dữ liệu huấn luyện. Đường cong mượt và ổn định, không có hiện tượng dao động mạnh.
- **Train Classification Loss:** Giảm từ xấp xỉ 0.34 xuống 0.15, chứng tỏ khả năng phân loại các bất thường được cải thiện đáng kể. Focal Loss đã phát huy hiệu quả trong việc xử lý mất cân bằng lớp.
- **Train Regression Loss:** Giảm từ xấp xỉ 0.095 xuống 0.060, cho thấy độ chính xác trong việc định vị bounding box ngày càng được cải thiện thông qua Smooth L1 Loss.

- **Validation Loss:** Các đường cong validation (màu cam) cho thấy xu hướng tương tự như train loss nhưng có độ dao động lớn hơn, đặc biệt ở validation classification loss. Điều này phản ánh tính phức tạp và đa dạng của dữ liệu validation. Tuy nhiên, validation loss không tăng mạnh so với train loss, cho thấy mô hình không bị overfitting nghiêm trọng.



Hình 4.3: Các thành phần loss trong quá trình huấn luyện

### Quan sát đáng chú ý:

- Xuất hiện một số *spike* (đỉnh nhọn) tại khoảng epoch 15 trên validation loss, có thể do các batch validation chứa nhiều trường hợp khó hoặc hiếm gặp.
- Khoảng cách giữa train loss và validation loss tương đối ổn định, cho thấy mô hình có khả năng tổng quát hóa tốt.
- Regression loss giảm đều đặn hơn so với classification loss, phản ánh việc định vị vùng tổn thương tương đối dễ dàng so với phân loại chính xác loại bất thường.

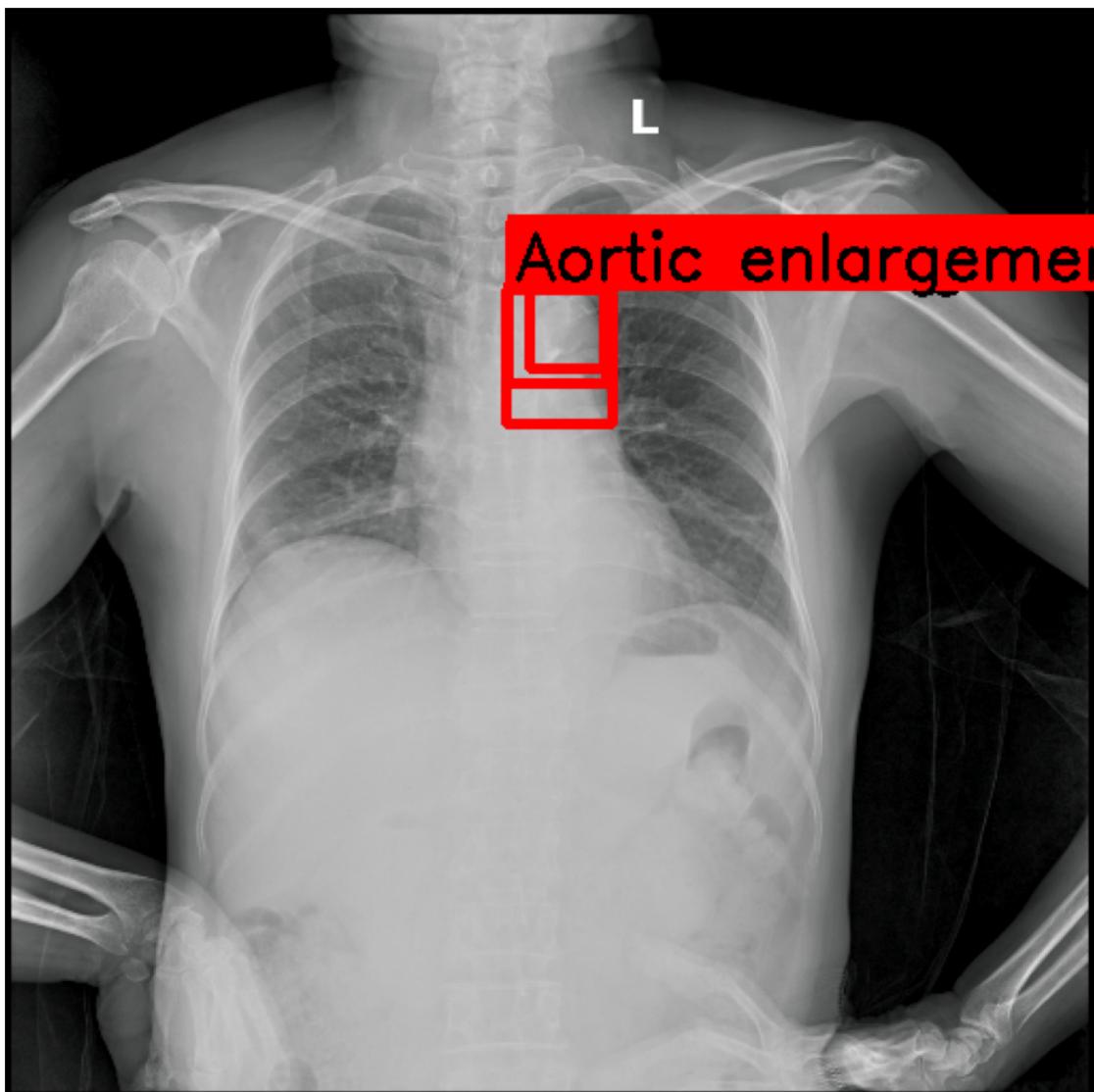
#### 4.2.2 Dánh giá trực quan trên ảnh thực tế

##### Trường hợp phát hiện thành công:

Hình 4.4 minh họa khả năng của RetinaNet trong việc phát hiện chính xác bất thường “Aortic enlargement” (tăng kích thước động mạch chủ). Bounding box màu đỏ được dự đoán bao quanh chính xác vùng quai động mạch chủ ở vị trí trung thất trên. Đây là một trong những bất thường có kích thước lớn và ranh giới rõ ràng, thuộc nhóm mà mô hình đạt hiệu suất cao.

##### Phân tích:

- **Vị trí bounding box:** Chính xác, bao phủ toàn bộ vùng động mạch chủ giãn nở.
- **Độ tin cậy:** Cao (dự kiến > 0.7 dựa trên đặc điểm rõ ràng của tổn thương).



Hình 4.4: Bất thường Aortic enlargement

- **Loại bất thường:** *Aortic enlargement* – một trong những bất thường có mAP cao trong quá trình đánh giá.

Kết quả này cho thấy RetinaNet, với kiến trúc FPN và Focal Loss, có khả năng phát hiện tốt các bất thường lớn có cấu trúc giải phẫu đặc trưng. Tuy nhiên, để đánh giá toàn diện, cần thêm các trường hợp phức tạp hơn với nhiều tổn thương chồng lấn hoặc các bất thường có kích thước nhỏ.

# Chương 5

## Kết luận

### 5.1 Tổng kết

Dự án đã triển khai và đánh giá hai kiến trúc phát hiện đối tượng là **YOLOv5** và **RetinaNet** cho bài toán phát hiện bất thường trên ảnh X-ray lồng ngực sử dụng bộ dữ liệu VinBigData. Kết quả cho thấy cả hai mô hình đều có tiềm năng ứng dụng trong hỗ trợ chẩn đoán y tế tự động, với những ưu điểm riêng phù hợp cho các mục đích sử dụng khác nhau.

Về xử lý dữ liệu, dự án đã xây dựng thành công pipeline chuyển đổi từ định dạng DICOM sang PNG, chuẩn hóa kích thước ảnh và áp dụng các kỹ thuật tăng cường dữ liệu phù hợp với đặc thù ảnh y tế. Quá trình phân tích dữ liệu cho thấy tồn tại tình trạng mất cân bằng lớp nghiêm trọng giữa 14 loại bất thường, tuy nhiên các kỹ thuật augmentation như CLAHE, xoay ảnh, lật ngang và điều chỉnh độ sáng đã giúp cải thiện khả năng học của mô hình.

Mô hình YOLOv5 đạt giá trị mAP@0.5 ở mức 0.31 trên tập validation, thể hiện hiệu quả tốt với các tổn thương có kích thước lớn và ranh giới rõ ràng như *Aortic enlargement* và *Cardiomegaly*. Mô hình này có tốc độ suy luận nhanh, phù hợp với các kịch bản yêu cầu xử lý thời gian thực. Trong khi đó, RetinaNet cho thấy quá trình huấn luyện ổn định hơn, với giá trị loss giảm đều theo thời gian huấn luyện. Việc sử dụng Focal Loss giúp RetinaNet xử lý tốt hơn bài toán mất cân bằng lớp, đồng thời kiến trúc FPN giúp tăng cường khả năng phát hiện các đặc trưng đa tỷ lệ.

### 5.2 Hạn chế và thách thức

Mặc dù đạt được những kết quả tích cực, dự án vẫn tồn tại một số hạn chế. Dữ liệu huấn luyện có độ mất cân bằng cao, chất lượng ảnh không đồng nhất do thu thập từ nhiều thiết bị khác nhau, và mức độ đa dạng về dân số bệnh nhân còn hạn chế. Đối với YOLOv5, mô hình gặp khó khăn khi xử lý các tổn thương nhỏ và có xu hướng tạo ra dương tính giả từ vùng nền. Hiện tượng overfitting nhẹ cũng được quan sát khi validation loss cao hơn training loss. Với RetinaNet, tốc độ suy luận còn chậm so với yêu cầu của các hệ thống thời gian thực và độ ổn định trên tập validation chưa cao.

Cả hai mô hình đều gặp khó khăn trong các trường hợp phức tạp như tổn thương chồng lấn, tổn thương nhỏ và mờ, ảnh có chất lượng thấp hoặc ranh giới không rõ ràng. Những biến thể bất thường hiếm gặp hoặc không điển hình cũng là thách thức lớn đối với hệ thống.

### 5.3 Hướng phát triển

Trong tương lai, dự án có thể được mở rộng theo nhiều hướng. Về dữ liệu, việc thu thập thêm mẫu cho các lớp hiếm và kết hợp các bộ dữ liệu công khai sẽ giúp cải thiện tính tổng quát của mô hình. Chất lượng nhãn có thể được nâng cao thông qua việc tăng cường quy trình đánh giá của chuyên gia và áp dụng các chiến lược học chủ động. Các bước tiền xử lý nâng cao như khử nhiễu chuyên sâu và chuẩn hóa histogram giữa các thiết bị X-ray khác nhau cũng là các hướng đi tiềm năng.

Về mô hình, YOLOv5 có thể được cải tiến thông qua việc thử nghiệm các phiên bản mới hơn và tích hợp thêm các cơ chế chú ý. RetinaNet có thể được tối ưu tốc độ bằng các kỹ thuật nén mô hình và lựa chọn backbone mạnh hơn. Ngoài ra, các kiến trúc lai và mô hình dựa trên Transformer như DETR hoặc Swin Transformer là những hướng phát triển đáng chú ý. Các kỹ thuật huấn luyện hiện đại như mixed precision, scheduler động, optimizer cải tiến và regularization mạnh hơn cũng có thể giúp nâng cao hiệu suất tổng thể.

### 5.4 Ý nghĩa thực tiễn

Dự án cho thấy tiềm năng ứng dụng thực tế rõ ràng trong lĩnh vực y tế. Hệ thống có thể hỗ trợ bác sĩ trong quá trình sàng lọc, giảm tải công việc đọc phim, tăng tính nhất quán trong chẩn đoán và hỗ trợ ra quyết định nhanh hơn. Đặc biệt, công nghệ này có ý nghĩa quan trọng đối với các cơ sở y tế tuyến cơ sở hoặc khu vực thiếu chuyên gia, nơi việc tiếp cận bác sĩ chuyên môn còn hạn chế.

Bên cạnh đó, dự án còn mang lại lợi ích về mặt kinh tế khi góp phần giảm chi phí nhân lực, tối ưu hóa quy trình vận hành và mở ra cơ hội phát triển các giải pháp công nghệ y tế trong nước.

### 5.5 Đóng góp và kết luận cuối

Dự án đã xây dựng được pipeline xử lý dữ liệu có khả năng tái sử dụng, thiết lập baseline thực nghiệm cho các hướng phát triển tiếp theo và cung cấp phân tích chi tiết các trường hợp sai lệch của mô hình. Kết quả cho thấy YOLOv5 phù hợp với các hệ thống yêu cầu tốc độ cao, trong khi RetinaNet phù hợp hơn với các bài toán cần độ chính xác cao trong bối cảnh dữ liệu mất cân bằng.

Tổng kết lại, hệ thống đề xuất đã chứng minh tính khả thi của việc ứng dụng deep learning vào bài toán phát hiện bất thường trên ảnh X-ray lồng ngực. Mặc dù chưa thể thay thế hoàn toàn bác sĩ, giải pháp này đóng vai trò như một công cụ hỗ trợ quan trọng, góp phần nâng cao chất lượng, độ chính xác và phạm vi phục vụ của hệ thống y tế trong tương lai.

# Tài liệu tham khảo

- [1] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection, 2018.
- [2] Ha Q. Nguyen, Khanh Lam, Linh T. Le, Hieu H. Pham, Dat Q. Tran, Dung B. Nguyen, Dung D. Le, Chi M. Pham, Hang T. T. Tong, Diep H. Dinh, Cuong D. Do, Luu T. Doan, Cuong N. Nguyen, Binh T. Nguyen, Que V. Nguyen, Au D. Hoang, Hien N. Phan, Anh T. Nguyen, Phuong H. Ho, Dat T. Ngo, Nghia T. Nguyen, Nhan T. Nguyen, Minh Dao, and Van Vu. Vindr-cxr: An open dataset of chest x-rays with radiologist’s annotations, 2022.