Phát hiện các bệnh về phổi dựa trên hình ảnh y tế sử dụng mô hình học sâu

Nhóm 4

Trường Đại học Khoa học Tự nhiên - ĐHQGHN

Một số khái niệm về sinh học được sử dụng

Trước khi tìm hiểu kỹ hơn về bài báo cáo, ta cần biết một số định nghĩa sau:

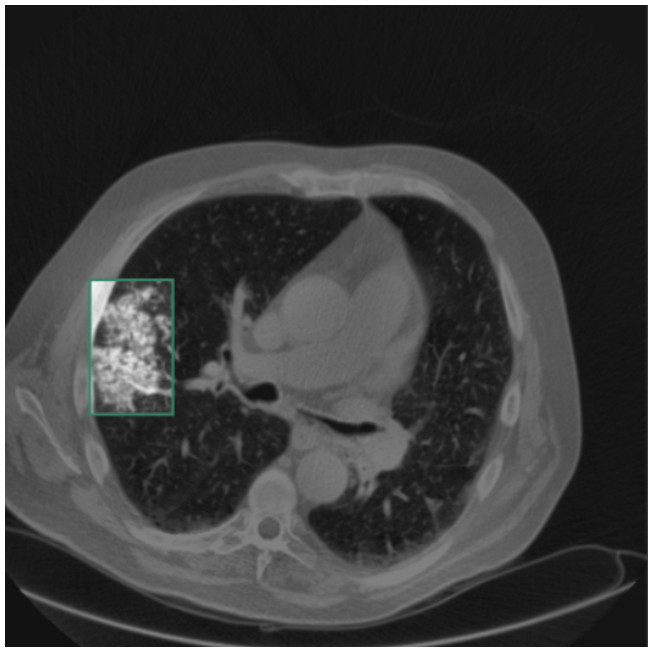
1. **Phổi** là cơ quan trao đổi khí của động vật, con người.
2. **X-Quang(X-Ray)** là một dạng bức xạ điện từ,có năng lượng cao, cho phép chúng đi qua hầu hết các vật thể, bao gồm cả cơ thể con người.
3. **Mờ vùng khí phế nang (Airspace opacities)** là một thuật ngữ mô tả tình trạng nhu mô phổi bị làm đầy bởi các chất liệu có khả năng hấp thụ tia X
4. **Tràn dịch màng phổi(Pleural effusion)** là tình trạng tích tụ quá mức của dịch trong khoang màng phổi.
5. **Ung thư phổi(Lung Cancer)** là một loại ung thư bắt đầu khi các tế bào bất thường phát triển không kiểm soát trong phổi.

Đặc điểm trên ảnh X-Quang của các dị thường phổi đã giới thiệu

Bốn dị thường phổi nói trên, bao gồm xuất huyết, ung thư, tràn dịch màng phổi và tràn khí màng phổi. Mỗi loại dị thường có những dấu hiệu riêng biệt, từ sự thay đổi về mật độ đến cấu trúc và hình dạng

Đặc điểm trên ảnh X-Quang của các dị thường phổi đã giới thiệu

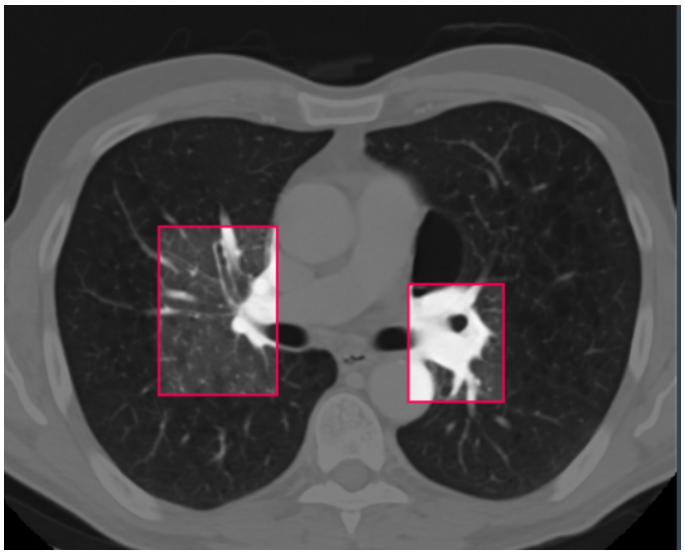
Xuất huyết: các dấu hiệu trên CT thường bao gồm các tổn thương lan tỏa hình mảng ở vùng quanh cửa phổi hai bên, phản ánh sự viêm nhiễm hoặc xuất huyết ở khu vực trung tâm của phổi.



Hình 1: Phổi có xuất huyết

Đặc điểm trên ảnh X-Quang của các dị thường phổi đã giới thiệu

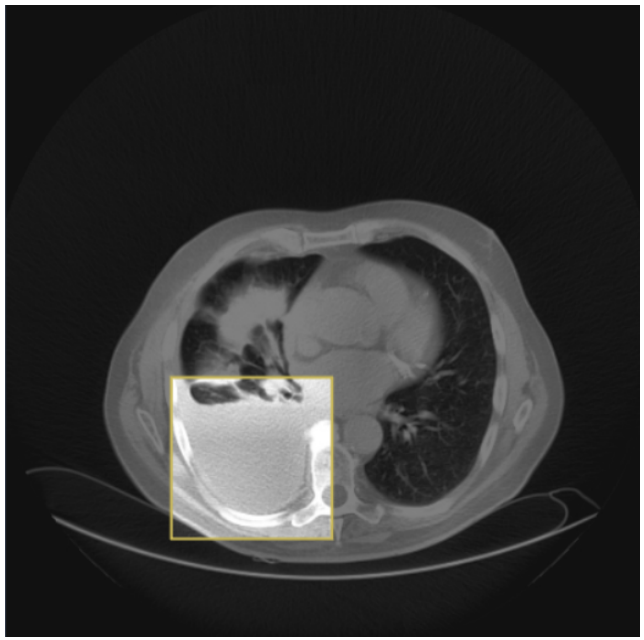
Ung thư phổi: các nốt hoặc khối có bờ không đều, kèm theo tăng sinh mạch máu xung quanh.



Hình 2: Ung thư phổi

Đặc điểm trên ảnh X-Quang của các dị thường phổi đã giới thiệu

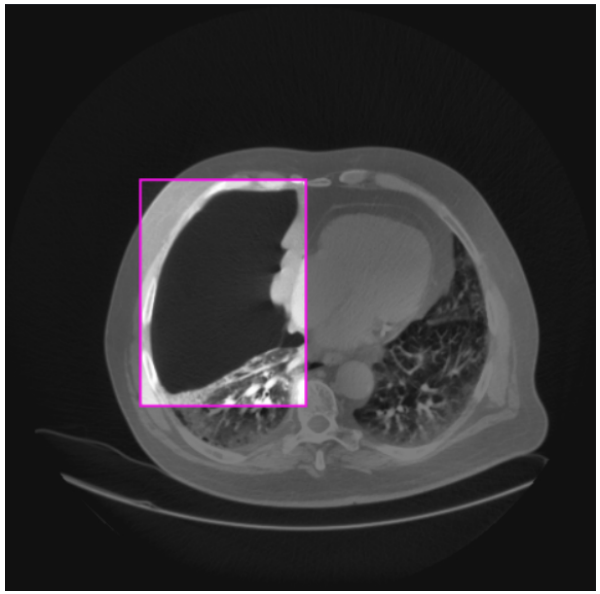
Tràn dịch màng phổi: Trên ảnh X-Quang, tràn dịch màng phổi xuất hiện như một vùng mờ ở đáy phổi với bờ trên cong, thường thấy ở góc sườn hoành.



Hình 3: Tràn dịch màng phổi

Đặc điểm trên ảnh X-Quang của các dị thường phổi đã giới thiệu

Tràn khí màng phổi: Tràn khí màng phổi trên ảnh CT thường dễ phát hiện ở cửa sổ phổi, với đặc trưng là khoảng không khí giữa màng phổi tạng và màng phổi thành, thiếu cấu trúc mạch máu. Phần khí này thường nằm ở góc trên ảnh, chèn ép khung xương sườn và tế bào phế nang.



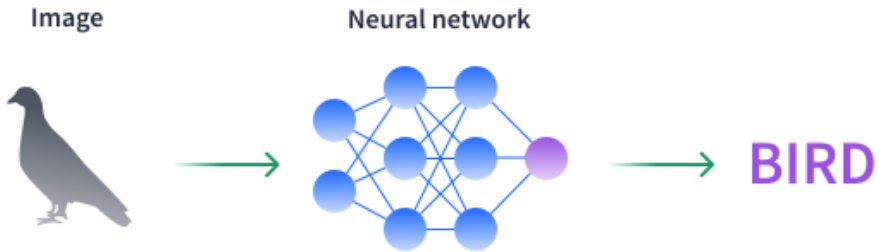
Hình 4: Tràn khí màng phổi

Giới thiệu bài toán chẩn đoán các bệnh về phổi thông qua ảnh X-quang

- Mục tiêu của bài toán là từ hình ảnh X-Quang của người khám, xác định được căn bệnh về phổi mà người đó mắc và xác định vị trí của bệnh trong phổi
- Để đạt được mục tiêu này, các kỹ thuật học sâu cần phải được sử dụng. Các kỹ thuật này có khả năng phát hiện các dấu hiệu bất thường trên ảnh X-Quang.
- Bài báo cáo này sẽ trình bày quy trình xây dựng mô hình từ việc tiền xử lý dữ liệu đếnlựa chọn mô hình đảm bảo hiệu quả cao trong thực nghiệm. Bên cạnh đó, việc đánh giá và tối ưu hóa hiệu suất của mô hình cũng sẽ được đề cập.

Học sâu

- **Học sâu** (Deep Learning) là một phương pháp học máy cho phép máy tự động học các biểu diễn đặc trưng "cấp cao"của dữ liệu.
- Thuật toán học sâu sử dụng mạng nơ ron nhân tạo, một hệ thống tính toán có khả năng học các đặc trưng cấp cao từ dữ liệu bằng cách tăng độ sâu (tức là số tầng) trong mạng. Mỗi tế bào trong một mạng nơ-ron được gọi là một nơ-ron.



Hình 5: Mô hình học sâu

Học sâu

Một mạng nơ ron nhân tạo gồm 4 phần:

- **1) Lớp đầu vào(Input Layer)**: Đây là nơi các quan sát huấn luyện được đưa vào thông qua các biến độc lập.
- **2) Các lớp ẩn(Hidden Layers)**: Đây là các lớp trung gian giữa lớp đầu vào và lớp đầu ra, nơi mạng nơ-ron học về các mối quan hệ và tương tác của các biến được đưa vào lớp đầu vào.
- **3) Lớp đầu ra(Output layer)**: Đây là lớp nơi kết quả cuối cùng được trích xuất sau tất cả các quá trình xử lý diễn ra trong các lớp ẩn.
- **4) Nút(Node)**: Một nút, còn được gọi là một nơ-ron, trong mạng nơ-ron là một đơn vị tính toán nhận vào một hoặc nhiều giá trị đầu vào và tạo ra một giá trị đầu ra.

Hàm Kích Hoạt - Activation fuction

Hàm bước nhị phân (Binary Step Function)

- Công thức:

$$f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

- Là bộ phân loại dựa trên ngưỡng, tức là liệu nơ-ron có được kích hoạt hay không dựa trên giá trị từ phép biến đổi tuyến tính.

Hàm tuyến tính (Linear Function)

- Công thức:

$$f(x) = ax$$

Hàm Kích Hoạt - Activation fuction

hàm kích hoạt Sigmoid (Sigmoid Activation Function)

- Công thức:

$$\text{Sigmoid}(x) = \sigma(x) = \frac{1}{1 + e^{-x}}$$

- Giá trị đầu ra nằm trong khoảng từ 0 đến 1, giúp chuẩn hóa dữ liệu và thích hợp cho các bài toán phân loại nhị phân.

Tanh (Hyperbolic Tangent)

- Công thức:

$$Tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

- Đầu ra của hàm Tanh nằm trong khoảng từ -1 đến 1.

Hàm Kích Hoạt - Activation fuction

Hàm kích hoạt ReLU (ReLU Activation Function)

- Công thức:

$$ReLU(x) = \max(0, x)$$

- Nếu đầu vào là số âm, hàm sẽ trả về 0; nếu đầu vào là số dương, hàm trả về chính giá trị đó.

Leaky ReLU

- Công thức:

$$f(x) = \begin{cases} 0.01x, & x < 0 \\ x, & x \geq 0 \end{cases}$$

- Là một biến thể của ReLU, cho phép gradient tồn tại với các giá trị âm (với hệ số nhỏ, thường là 0.01).

Hàm Kích Hoạt - Activation fuction

ELU

- Công thức:

$$f(x) = \begin{cases} x, & x \geq 0 \\ \alpha(e^x - 1), & x < 0 \end{cases}$$

Swish

- Công thức:

$$\text{Swish}(x) = x \cdot \text{Sigmoid}(x) = x \cdot \frac{1}{1 + e^{-x}}$$

- Là hàm phi tuyến, kết hợp giữa Sigmoid và ReLU..

Hàm Kích Hoạt - Activation fuction

Softmax

- Công thức:

$$Softmax(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

- Softmax chuyển đổi đầu ra thành xác suất, thường dùng ở lớp đầu ra cho bài toán phân loại nhiều lớp.

Mạng Nơ-ron tích chập(CNN)

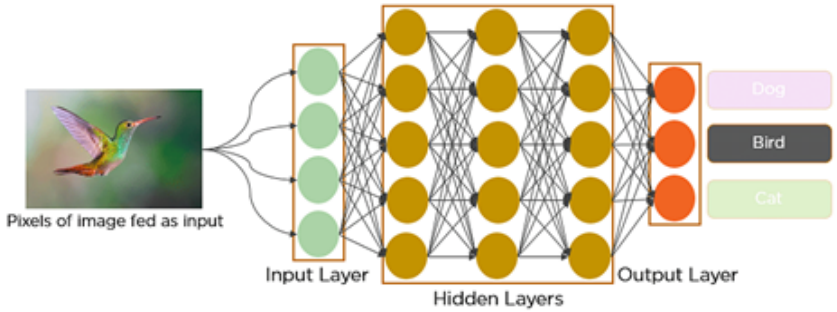
Mạng Nơ-ron tích chập (Convolutional Neural Network) hay còn được gọi là CNN, là một trong những mô hình Deep Learning cực kỳ tiên tiến cho phép xây dựng những hệ thống có độ chính xác cao và thông minh. Nhờ khả năng đó, CNN có rất nhiều ứng dụng, đặc biệt là những bài toán cần nhận dạng vật thể (object) để phát hiện và phân loại các đối tượng trong hình ảnh.

Các lớp trong mạng CNN

- Lớp tích chập
- Lớp ReLU - Rectified Linear Unit
- Pooling layer – Lớp gộp
- Lớp Fully Connected

Mạng Nơ-ron tích chập(CNN)

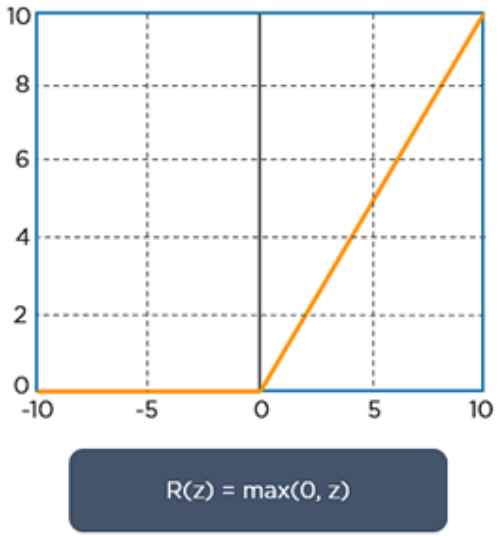
Lớp tích chập là lớp đầu tiên thực hiện việc trích xuất đặc trưng từ hình ảnh. Lớp này có các tham số bao gồm một tập hợp các bộ lọc có thể học được. Các bộ lọc này thường có kích thước nhỏ, thường là 3x3 hoặc 5x5 ở hai chiều đầu tiên, và độ sâu tương đương với độ sâu của đầu vào. Khi các bộ lọc này di chuyển dọc và ngang trên hình ảnh, chúng tạo ra một bản đồ đặc trưng (Feature Map) chứa các đặc điểm được trích xuất từ hình ảnh đầu vào.



Hình 6: Phép toán tích chập

Mạng Nơ-ron tích chập(CNN)

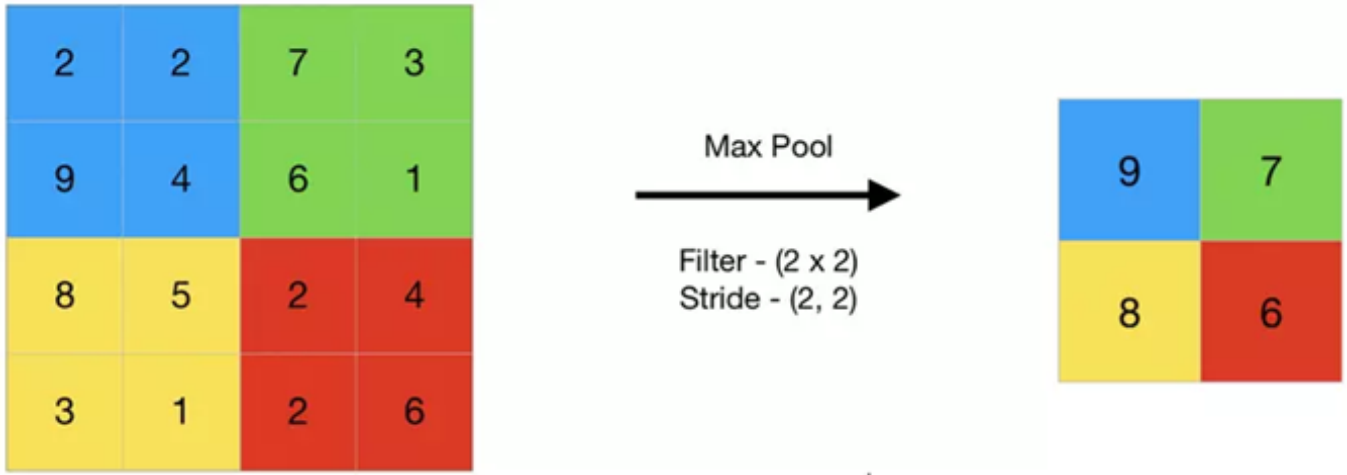
Lớp ReLU - Rectified Linear Unit thực hiện một phép biến đổi phi tuyến đơn giản nhưng vô cùng hiệu quả, bằng cách áp dụng một hàm kích hoạt theo từng phần tử, trong đó mọi giá trị âm đều được thay thế bằng 0, trong khi các giá trị dương vẫn được giữ nguyên. Điều này giúp đưa tính phi tuyến vào mạng, giúp mô hình có khả năng học các mối quan hệ phức tạp hơn từ dữ liệu.



Hình 7: Đồ thị hàm ReLU

Mạng Nơ-ron tích chập(CNN)

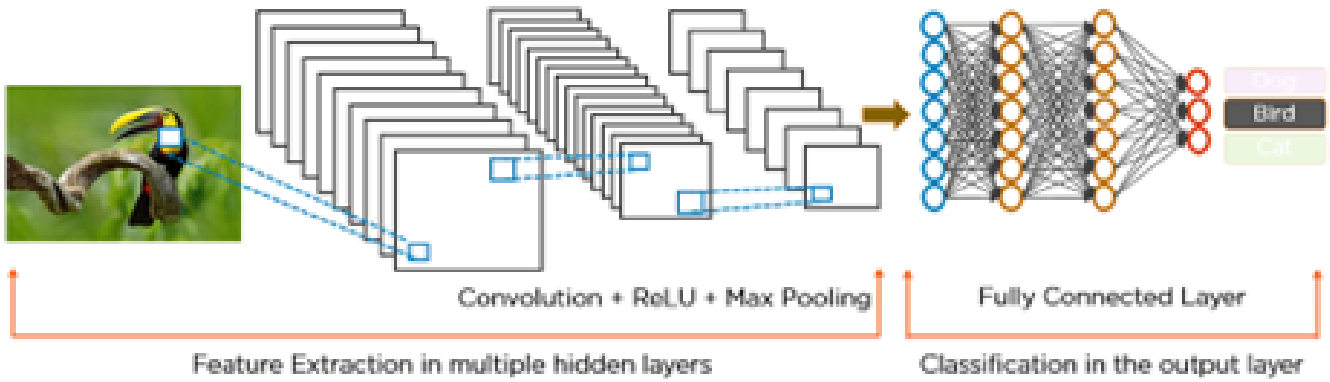
Pooling layer thường được dùng giữa các convolutional layer, để giảm kích thước dữ liệu nhưng vẫn giữ được các thuộc tính quan trọng. Kích thước dữ liệu giảm giúp giảm việc tính toán trong model, giảm độ phức tạp của mô hình và tránh overfitting.Các pooling có thể có nhiều loại khác nhau: **Max Pooling**, **Average Pooling**, **Sum Pooling**. Phổ biến là Max Pooling và Average Pooling.



Hình 8: Pooling layer

Mạng Nơ-ron tích chập(CNN)

Lớp Fully Connected sau khi ảnh được truyền qua nhiều convolutional layer và pooling layer thì model đã học được tương đối các đặc điểm của ảnh thì tensor của output của layer cuối cùng sẽ được là phẳng thành vector và đưa vào một lớp được kết nối như một mạng nơ-ron.



Hình 9: Fully Connected Layer

Cấu trúc của mạng CNN

- Mạng CNN là một trong những tập hợp của lớp Convolution được chồng lên nhau.
- Đặc điểm mô hình CNN có 2 khía cạnh cần phải đặc biệt lưu ý là tính bất biến và tính kết hợp. Với các loại chuyển dịch, co giãn và quay, người ta sẽ sử dụng pooli layer và làm bất biến những tính chất này.
- Pooling layer giúp tạo nên tính bất biến đối với phép dịch chuyển, phép co giãn và phép quay. Dựa trên cơ chế convolution, một mô hình sẽ liên kết được các layer với nhau.
- Với cơ chế này, layer tiếp theo sẽ là kết quả được tạo ra từ convolution thuộc layer kế trước. Mỗi nơ-ron sinh ra ở lớp tiếp theo từ kết quả filter sẽ áp đặt lên vùng ảnh cục bộ của nơ-ron tương ứng trước đó.

Convolutional Neural Network Training

Đào tạo Mạng nơ-ron tích chập (CNN) bao gồm hướng dẫn mô hình nhận dạng các mẫu trong dữ liệu thông qua quy trình học từng bước.

Chuẩn bị dữ liệu - Data Preparation

Chức năng mất mát - Loss Function

- **Mean Squared Error (MSE)** loss function là tổng bình phương các hiệu số giữa các mục trong vectơ dự đoán y và vectơ thực tế y-hat.

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{i=0}^{N}(y_i - \hat{y}_i)^2$$

Convolutional Neural Network Training

- **Cross-Entropy Loss Function** là hàm mất mát có thể đo lỗi giữa xác suất dự đoán và nhãn biểu diễn lớp thực tế

$$\mathcal{L}(\theta) = -\sum_{i=0}^{N} \hat{y}_i \cdot \log(y_i)$$

- **Mean Absolute Percentage Error (MAPE)** để đo hiệu suất của mạng nơ-ron trong các tác vụ dự báo nhu cầu.

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=0}^{N} \frac{|y_i - \hat{y}_i|}{\hat{y}_i}$$

Trình tối ưu hóa - Optimizer

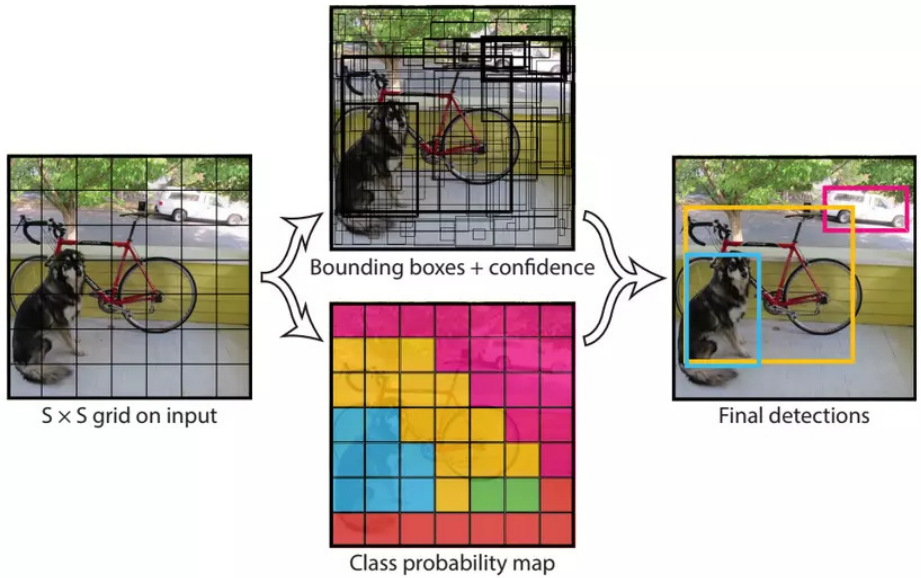
Truyền ngược - Backpropagation

Tổng quan về YOLO

- YOLO là thuật toán học sâu có giám sát để phát hiện vật thể trong ảnh hoặc video.
- Chỉ cần một mạng nơ-ron duy nhất để nhận diện và gán nhãn các đối tượng.
- Đầu vào gồm:
  - Ảnh hoặc video.
  - Tệp nhãn: Lưu các chỉ số như khung vật thể (bounding box) và tên đối tượng.

Cách thức hoạt động của YOLO

- YOLO chuẩn hóa ảnh đầu vào thành hình vuông bằng cách thêm viền (padding).
- Ảnh được chia thành lưới kích thước $S \times S$ (ví dụ: 13x13, 19x19,...).
- Mỗi ô lưới chịu trách nhiệm phát hiện các đối tượng có tâm nằm trong ô.



Bounding Boxes và Các Dự Đoán

- YOLO dự đoán các thông tin:
  - Tọa độ trung tâm (x, y).
  - Chiều rộng (w) và chiều cao (h).
  - Xác suất đối tượng p_{obj}.
  - Xác suất thuộc lớp c_i (ví dụ: người, xe, chó,...).
- **Confidence score:**

$$\text{confidence score} = p_{obj} \times \max(c_i)$$

Ví dụ: Nếu:

- $p_{obj} = 0.8$, $\max(c_i) = c_{human} = 0.7$.

Kết quả: Confidence score = $0.8 \times 0.7 = 0.72$.

Non-Max Suppression (NMS)

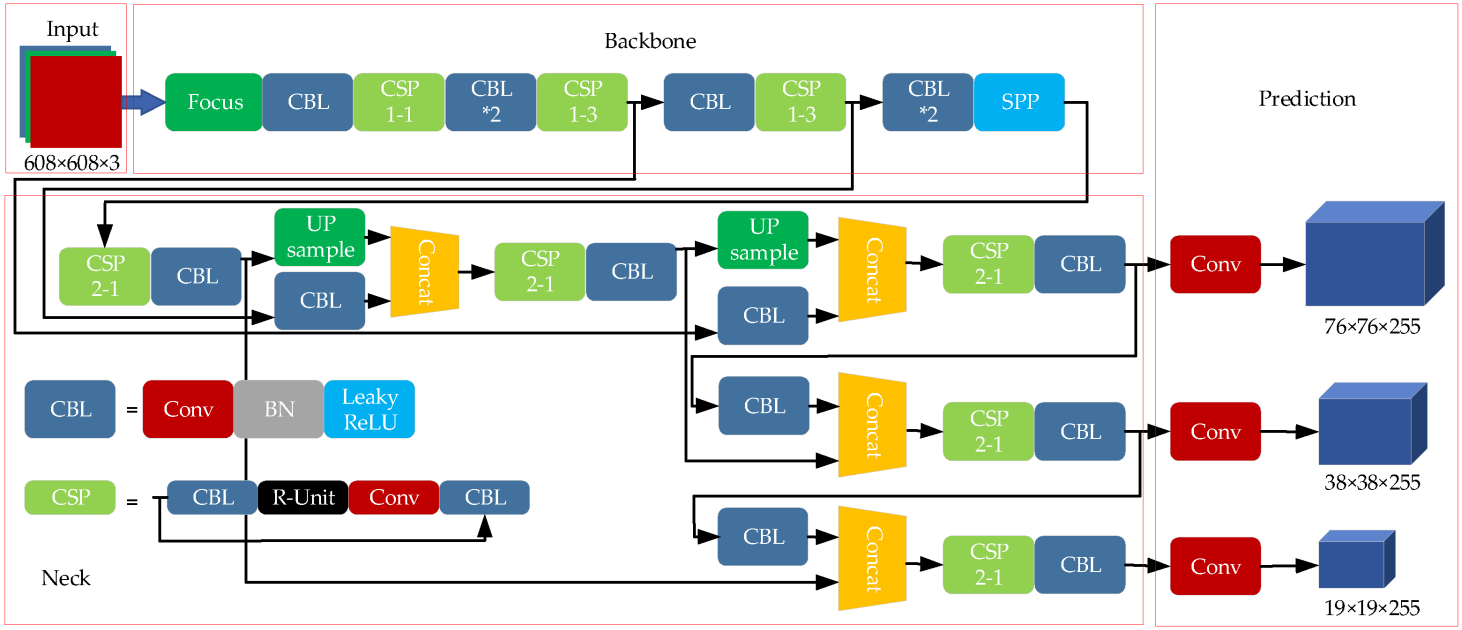
- Loại bỏ các bounding boxes dư thừa, giữ lại hộp đáng tin cậy nhất.
- Chọn hộp có confidence score cao nhất và loại các hộp có IoU lớn hơn ngưỡng.
- IoU (Intersection over Union):

$$\text{IoU(A,B)} = \frac{\text{Diện tích phần giao}}{\text{Diện tích phần hợp}}$$

Yolov5

- YOLOv5 (You Only Look Once version 5): Phiên bản thứ 5 của YOLO.
- Ưu điểm so với các phiên bản trước: Tốc độ nhanh hơn, hiệu quả cao hơn nhờ cải tiến kiến trúc và kỹ thuật huấn luyện.
- Ngôn ngữ và framework: Phát triển bằng Python và PyTorch.
- Lợi ích:
  - Dễ sử dụng, tích hợp với các thư viện học sâu phổ biến.
  - Dễ thử nghiệm và tối ưu hóa cho các ứng dụng thực tế.

Kiến trúc của YOLOv5



- **Backbone:**
  - Chịu trách nhiệm trích xuất đặc trưng từ ảnh đầu vào.
  - **Các thành phần chính:**
    - **C3 block:** Học đặc trưng ở nhiều cấp độ với các tham số: số lượng block trong khối, kích thước Kernel, Bước nhảy(stride).
    - **SPP (Spatial Pyramid Pooling):** Xử lý đặc trưng không gian ở nhiều tỷ lệ, cải thiện khả năng nhận diện.

Kiến trúc YOLOv5

- **Neck:**
  - Kết hợp và chuẩn bị các đặc trưng từ Backbone cho Head.
  - **Chức năng chính:**
    - **Upsample:** Tăng kích thước đặc trưng, hỗ trợ tổng hợp thông tin.
    - **C3 block và CONV layers:** Tinh chỉnh đặc trưng, tối ưu hóa nhận diện cả đối tượng nhỏ và lớn.
- **Head:**
  - Phần chịu trách nhiệm dự đoán cuối cùng, bao gồm:
    - **Bounding box:** Dự đoán vị trí và kích thước các hộp chứa vật thể.
    - **Confidence score:** Độ chắc chắn của dự đoán.
    - **Class probabilities:** Phân loại các vật thể trong ảnh.
  - Tích hợp đầu ra từ Neck để tạo ra các dự đoán chính xác.
  - Dựa vào **anchor box** và các thông tin đặc trưng để cải thiện hiệu suất nhận diện.

Hàm mất mát

Công thức tổng quát:

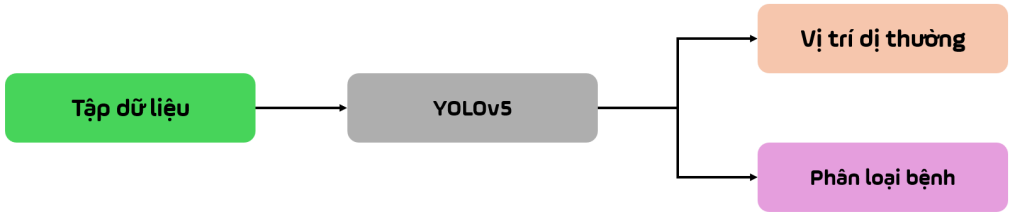
$$L = \lambda_{box} L_{box} + \lambda_{cls} L_{cls} + \lambda_{obj} L_{obj}$$

trong đó:

- L_{box}: là hàm mất mát khung bao
- L_{cls}: là hàm mất mát phân loại
- L_{obj}: là hàm mất mát tin cậy
- λ_{box}, λ_{cls}, λ_{obj}: là các hệ số để điều chỉnh tầm quan trọng của từng thành phần trong quá trình huấn luyện.

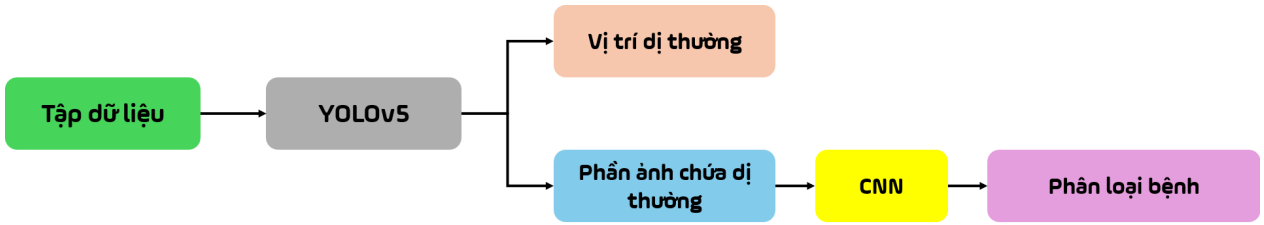
Áp dụng các mô hình trên cho bài toán dự đoán và phân loại bệnh phổi

- **Mô hình 1**: Sử dụng mô hình YOLO cho việc phát hiện và phân loại bệnh:



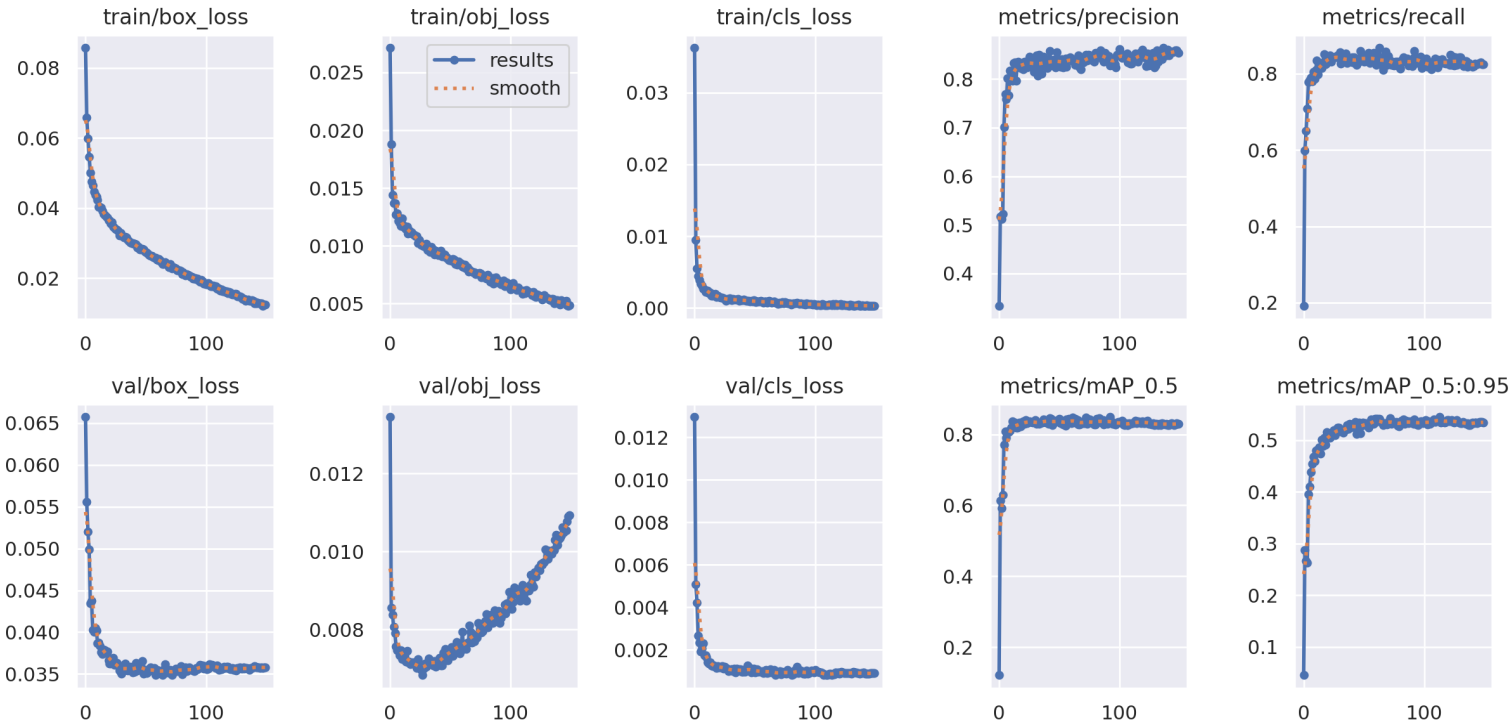
- **Mô hình 2**: Sử dụng CNN trong việc phân loại bệnh:

- **Mô hình 3**: Sử dụng kết hợp YOLO và CNN:



Kết quả thực nghiệm với YOLOv5

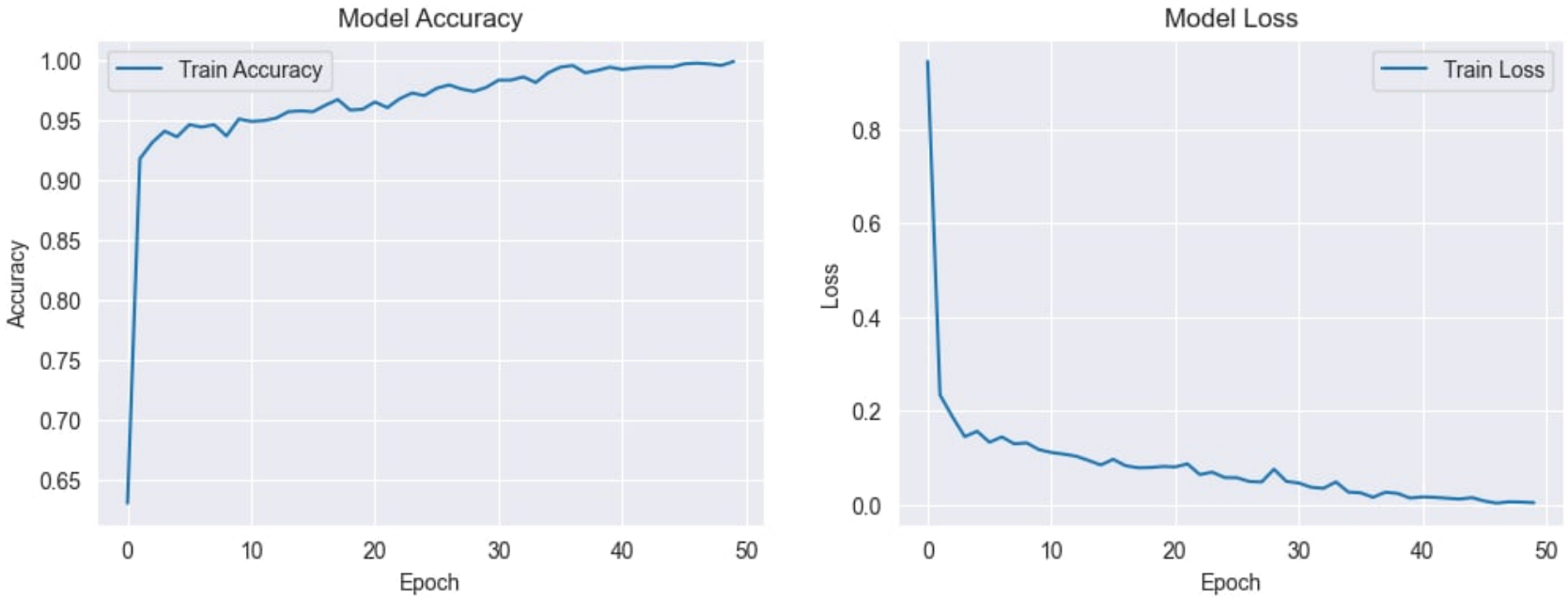
Mô hình được huấn luyện 150 epochs, trong tổng thời gian 2 giờ 54 phút, tương đương 1,16 phút 1 epoch cho ra kết quả sau:



Hình 10: Biểu đồ kết quả huấn luyện của YOLOv5

Kết quả thực nghiệm với mô hình CNN

Mô hình được huấn luyện 100 epochs, trong tổng thời gian 1118.63 giây (≈ 18 phút 39 giây), tương đương với khoảng 11 giây cho mỗi epoch cho ra kết quả sau:



Hình 11: Biểu đồ độ tăng trưởng độ chính xác và mất mát khi huấn luyện của CNN

Kết quả thực nghiệm với mô hình CNN

Báo cáo phân loại điều tra:

	precision	**recall**	**f1-score**	**support**
Class 0	0.74	0.83	0.78	63
Class 1	0.84	0.91	0.87	179
Class 2	0.78	0.84	0.81	133
Class 3	1.00	0.65	0.79	92
accuracy			0.83	467
macro avg	0.84	0.81	0.81	467
weighted avg	0.84	0.83	0.83	467

Bảng 1: Báo cáo phân loại trên tập kiểm tra của CNN

Kết quả thực nghiệm với mô hình YOLOv5 kết hợp CNN

Mô hình được sử dụng lại kết quả huấn luyện của YOLOv5 và sử dụng lại mô hình CNN để xử lí phân loại bounding box, với thời gian kiểm tra là 503,61 giây ($\approx$ 8 phút 23 giây) cho 561 ảnh:

	precision	**recall**	**f1-score**	**support**
Class 0	0.79	0.81	0.79	77
Class 1	0.86	0.88	0.87	1207
Class 2	0.81	0.86	0.83	168
Class 3	0.95	0.90	0.92	150
accuracy			0.85	602
macro avg	0.84	0.84	0.84	602
weighted avg	0.85	0.85	0.85	602

Bảng 2: Báo cáo phân loại trên tập kiểm tra của mô hình CNN kết hợp YOLOv5

So sánh mô hình

Chỉ số	CNN	YOLOv5	CNN + YOLOv5
Độ chính xác (Accuracy)	82.87%	84%	85%
mAP50	Không áp dụng	84%	85%
mAP50-95	Không áp dụng	54%	55%
Precision (Lớp 0)	0.74	0.78	0.79
Precision (Lớp 1)	0.84	0.85	0.86
Precision (Lớp 2)	0.78	0.80	0.81
Precision (Lớp 3)	1.00	0.96	0.95
Recall (Lớp 0)	0.83	0.79	0.81
Recall (Lớp 1)	0.91	0.87	0.88
Recall (Lớp 2)	0.84	0.85	0.86
Recall (Lớp 3)	0.65	0.88	0.90
F1-Score (Lớp 0)	0.78	0.78	0.79
F1-Score (Lớp 1)	0.87	0.86	0.87
F1-Score (Lớp 2)	0.81	0.82	0.83
F1-Score (Lớp 3)	0.79	0.97	0.92

Hình 12: So sánh chỉ số đánh giá của ba mô hình phát hiện và phân loại

Thanks for listening!!!