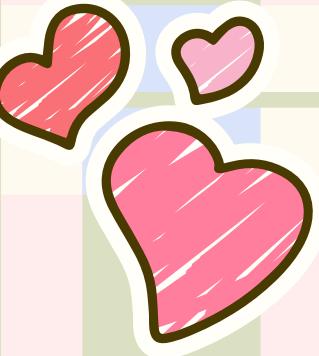
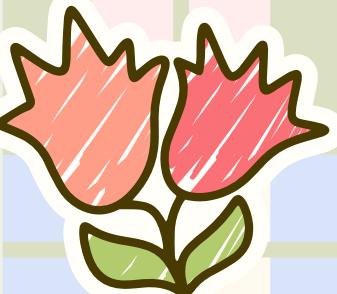




# PHÁT HIỆN VIỄM PHỔI TỪ ẢNH X-QUANG

Nhóm 19

# THÀNH VIÊN NHÓM



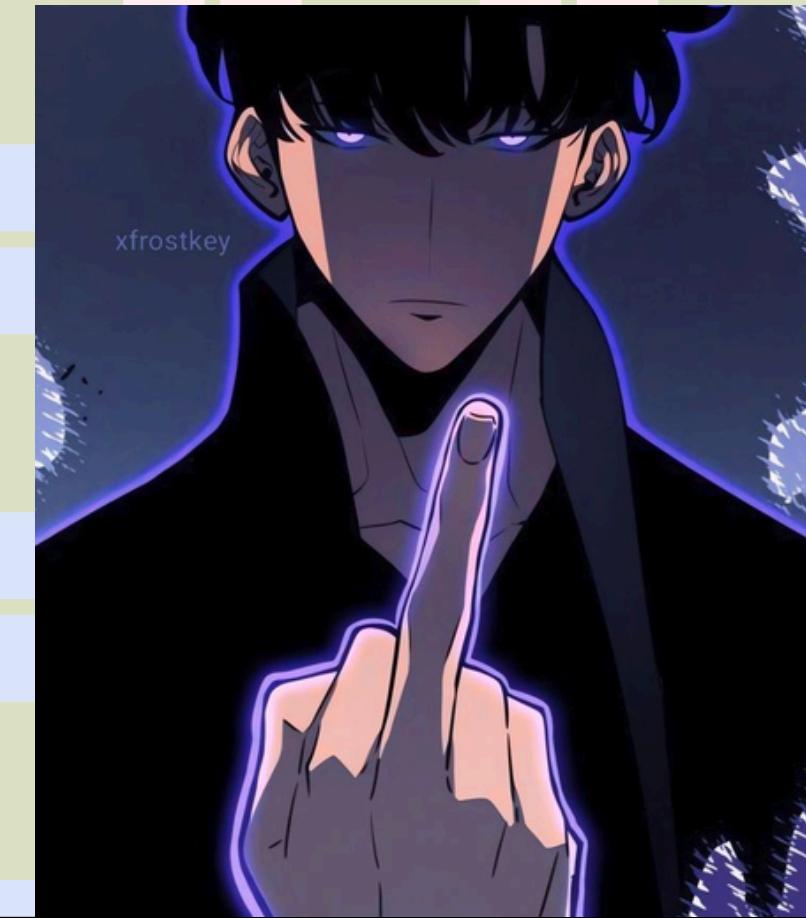
NGUYỄN HỮU AN



NGUYỄN TIỀN ĐẮT



NGUYỄN VĂN DŨNG



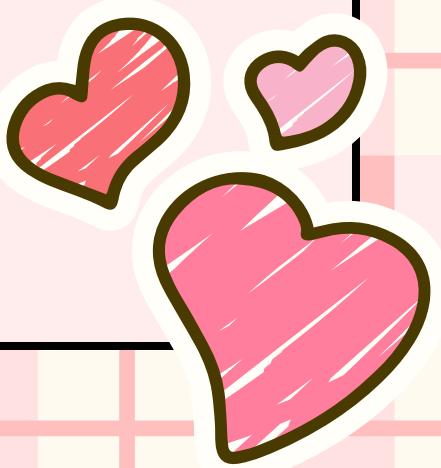
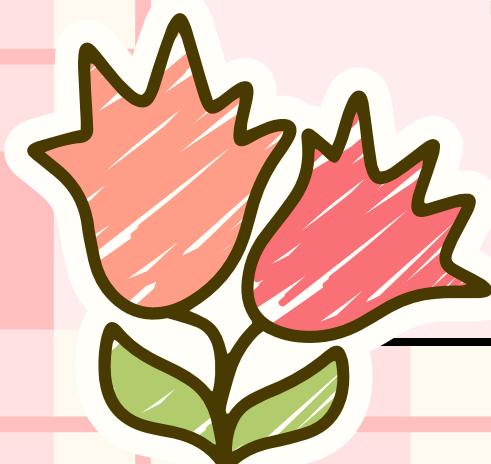
VŨ ĐỨC QUÝ

# THỰC TRẠNG

- Viêm phổi là nguyên nhân tử vong hàng đầu ở trẻ em dưới 5 tuổi (WHO).
- Đặc điểm ảnh X-quang khó phân tích: độ tương phản thấp, ranh giới mờ.
- Bệnh viện tuyến dưới thiết hụt bác sĩ có chuyên môn.
- Số lượng bệnh nhân lớn gây quá tải cho bác sĩ.
- Áp lực công việc và sự mệt mỏi gây ra sự sai lệch trong chuẩn đoán hình ảnh.

# LÝ DO CHỌN ĐỀ TÀI

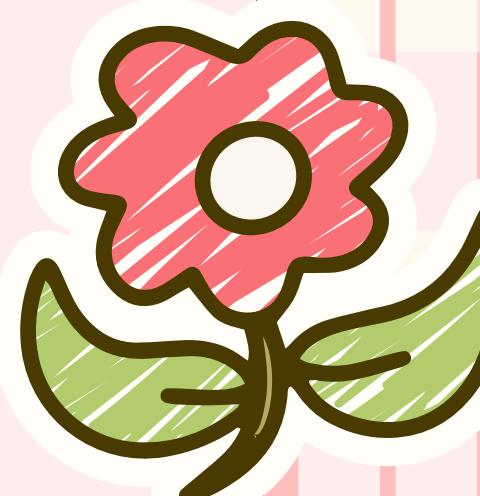
- Sự cần thiết của một mô hình AI giúp chuẩn đoán hình ảnh chính xác.
- Hỗ trợ bác sĩ trong quá trình khám chữa bệnh lâm sàng.
- Xử lý lượng lớn ảnh X-Quang giảm tải khối lượng công việc thời điểm cao điểm.
- Bước nhỏ đưa công nghệ AI tiến gần với thực tiễn đời sống.



# LÝ DO CHỌN ĐỀ TÀI

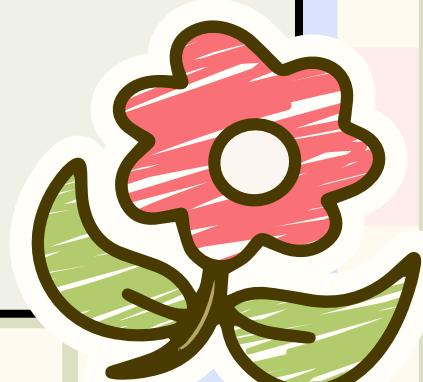
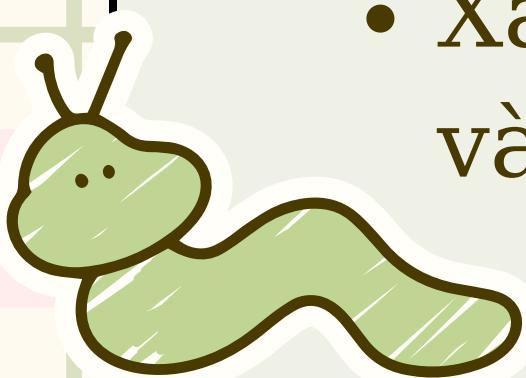
- Sự cần thiết của một mô hình AI giúp chuẩn đoán hình ảnh chính xác.
- Hỗ trợ bác sĩ trong quá trình khám chữa bệnh lâm sàng.
- Xử lý lượng lớn ảnh X-Quang giảm tải khối lượng công việc thời điểm cao điểm.
- Bước nhỏ đưa công nghệ AI tiến gần với thực tiễn đời sống.

**Mục tiêu:** Xây dựng hệ thống hỗ trợ chẩn đoán dựa trên Học sâu để sàng lọc nhanh và chính xác.

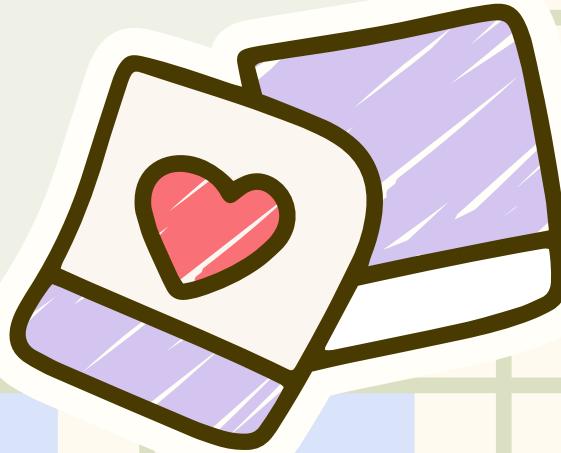




- Nghiêm cứu cơ sở lý thuyết để tìm ra giải pháp phù hợp nhất với bài toán.
- Tiên xử lý bộ dữ liệu ảnh X-Quang ngực được công bố trên Kaggle.
- Thiết kế và cài đặt lại kiến trúc mạng DenseNet để tối ưu hóa cho dữ liệu ảnh y tế.
- Kết quả mong đợi: Độ chính xác  $> 95\%$ , Recall  $> 0.95$ , AUC  $> 0.99$ .
- Xây dựng App Web kiểm thử khả năng thực tiễn của mô hình và hướng phát triển trong tương lai.

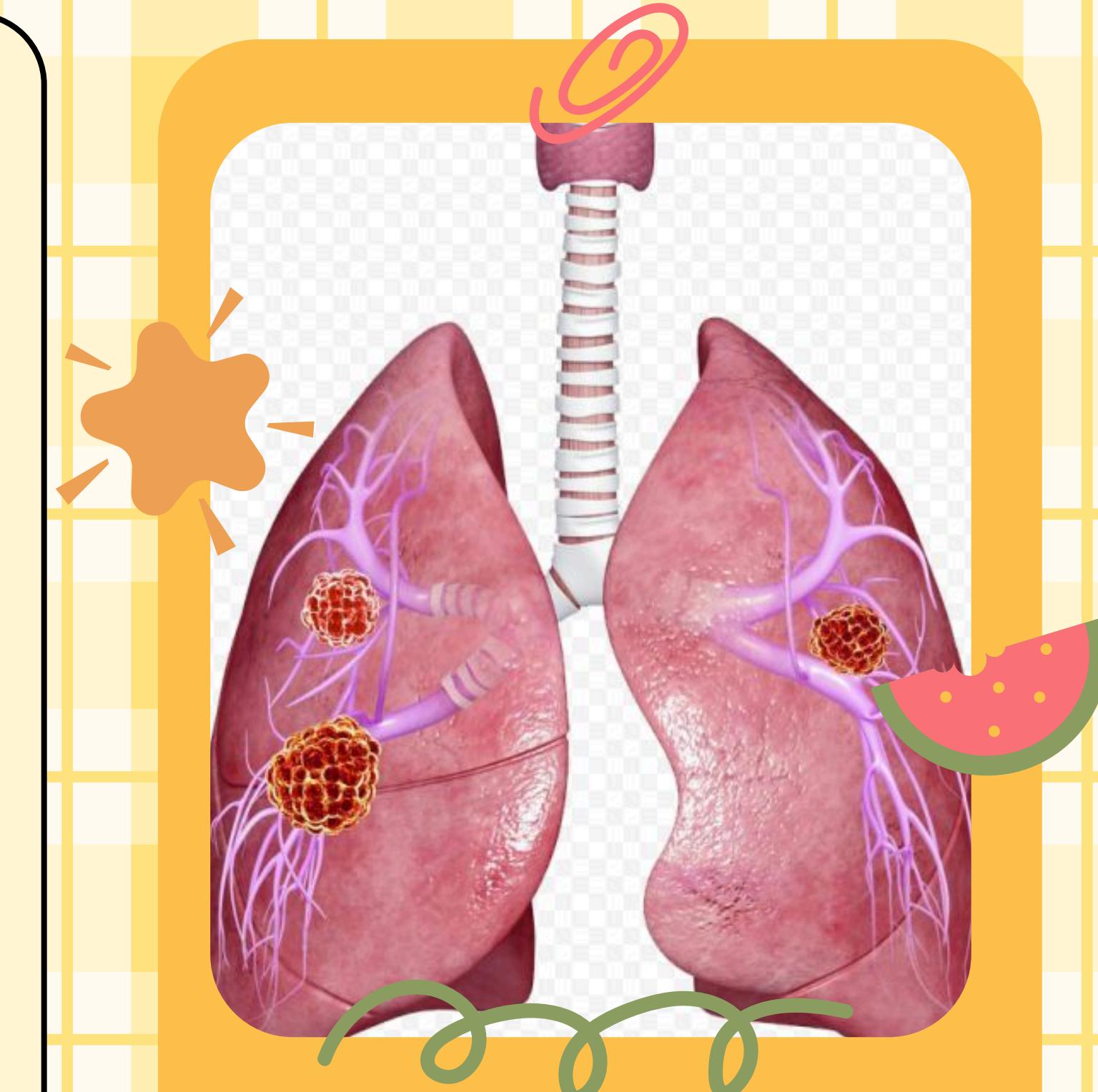


# CƠ SỞ LÝ THUYẾT

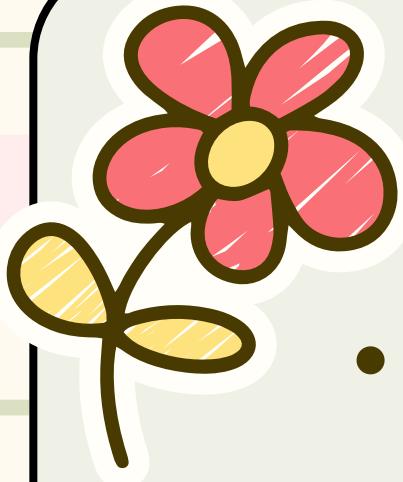


# TỔNG QUAN VỀ VIÊM PHỔI

- Tình trạng nhiễm trùng làm túi khí phổi chứa đầy dịch hoặc mủ, gây cảm trở hô hấp.
- Tác nhân: Vi khuẩn, Virus, Nấm
- Phương pháp chẩn đoán: X-quang lồng ngực

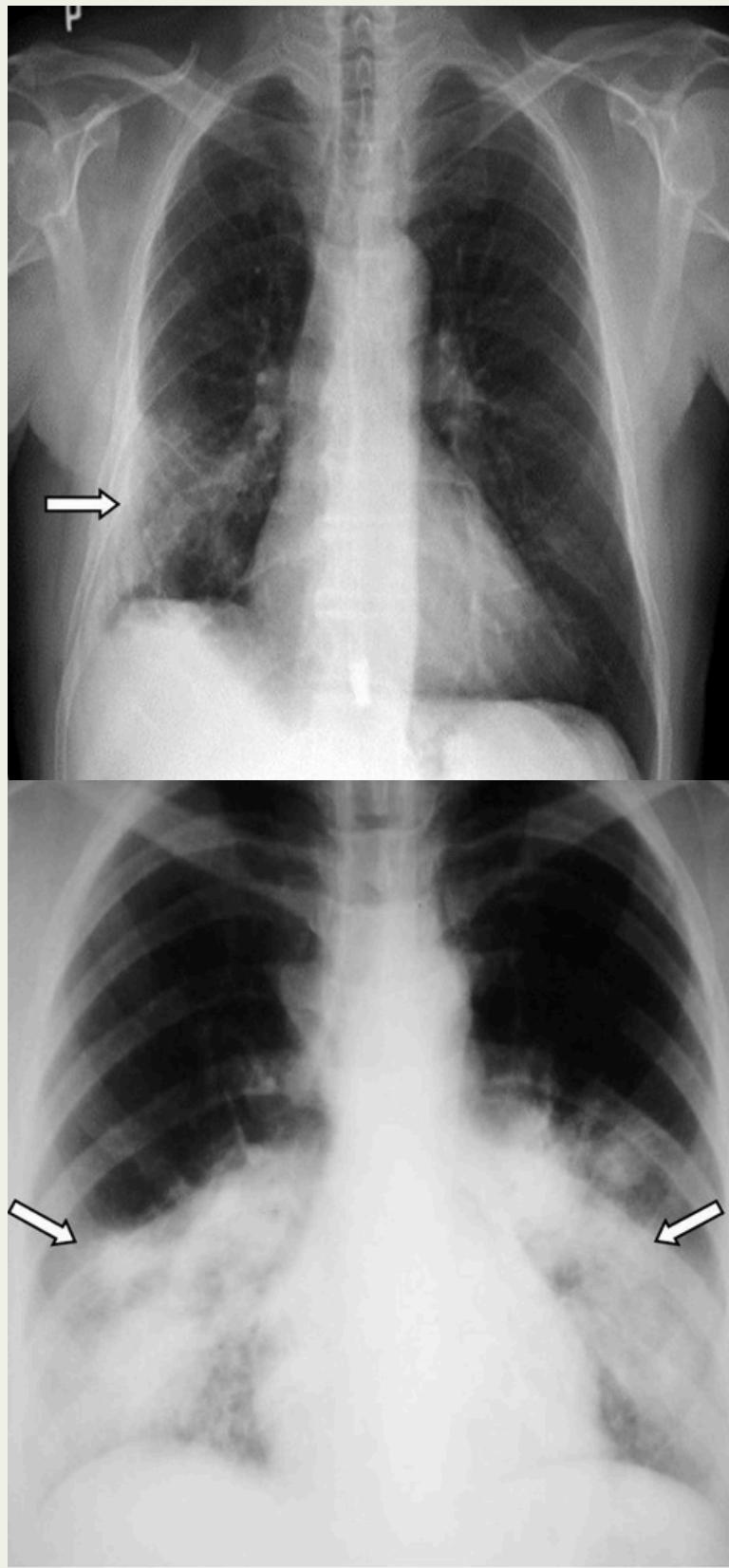


# THÁCH THỨC TRONG CHẨN ĐOÁN HÌNH ẢNH



- Vùng tổn thương (nốt mờ, đám thâm nhiễm) có độ tương phản thấp, ranh giới không rõ, dễ bị xương sườn/tim che khuất.
- Yếu tố con người gây ra sự sai sót trong chẩn đoán hình ảnh.
- Sự gia tăng bệnh nhân trong thời điểm đại dịch.

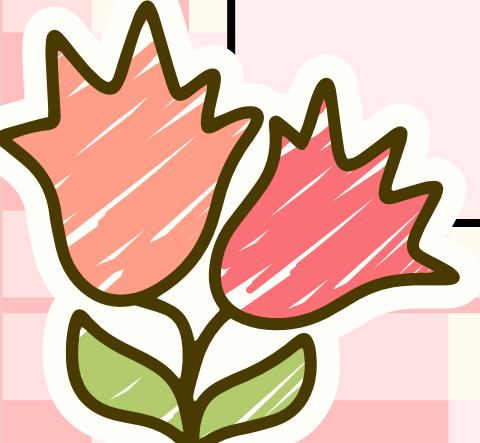
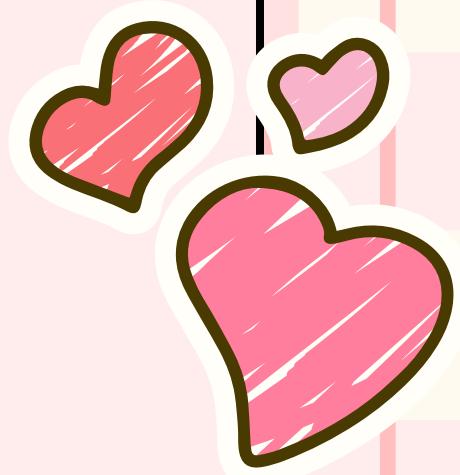
**Giải pháp:** Ứng dụng mô hình học sâu để hỗ trợ "đọc phim" tự động, giảm tải áp lực y tế.





## HỌC SÂU VÀ MẠNG NƠ-RON TÍCH CHẬP



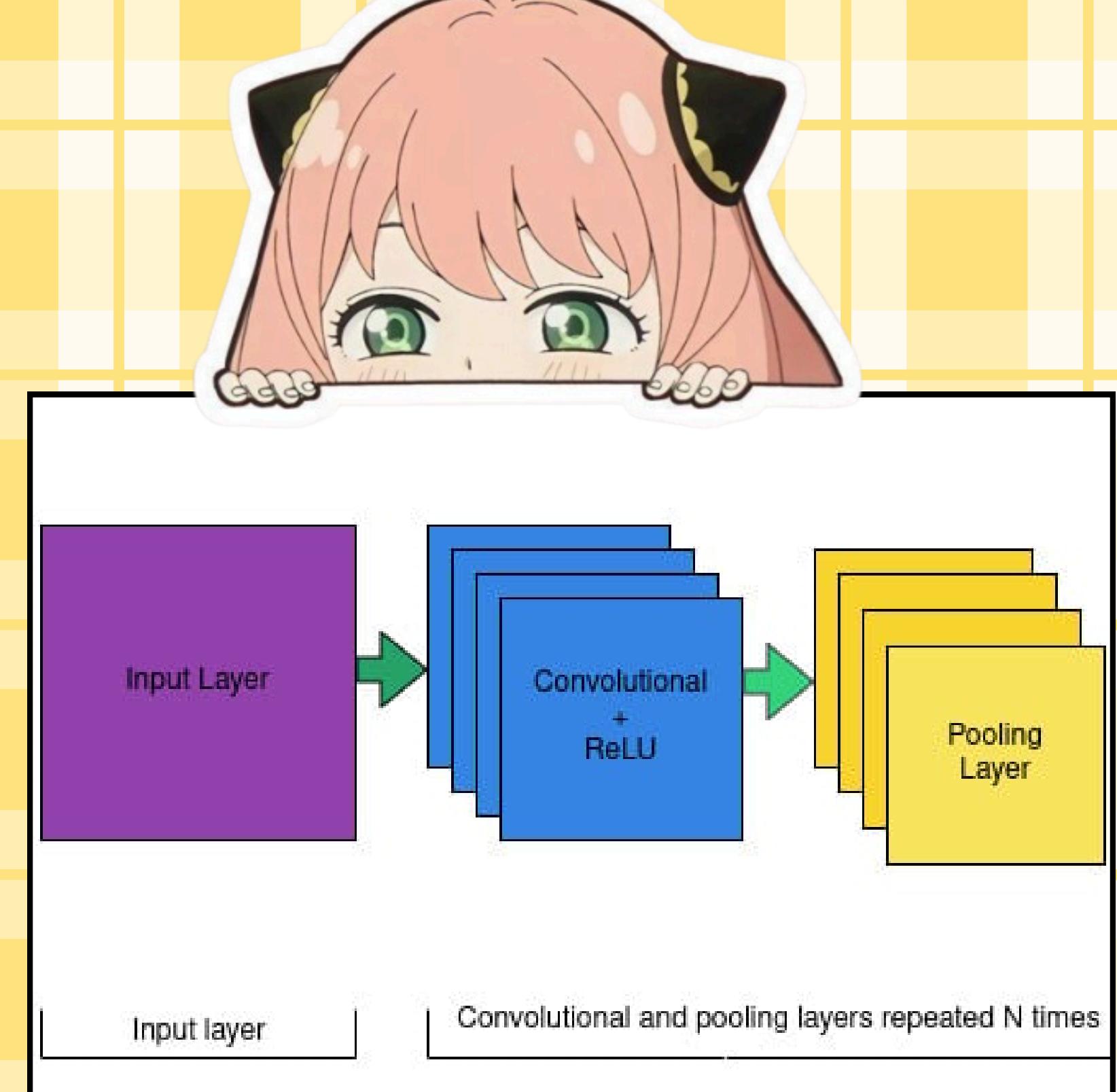
- Học sâu có khả năng tự động trích xuất đặc trưng bằng việc sử dụng nhiều lớp ẩn xếp chồng lên nhau
  - Học các hàm phi tuyến tính phức tạp, giúp nhận diện chính xác các đặc điểm mờ nhạt tốt hơn mắt thường hoặc thuật toán cổ điển.
  - Học sâu đã chứng minh được hiệu quả vượt trội trong việc phát hiện bệnh lý qua hình ảnh
- 
- 

# HỌC SÂU VÀ MẠNG NƠ-RON TÍCH CHẬP

- CNN là "tiêu chuẩn vàng" trong thị giác máy tính, ưu việt hơn mạng nơ-ron truyền thống nhờ cơ chế chia sẻ trọng số.
- Khả năng tự động trích xuất đặc trưng từ thấp (cạnh, góc) đến cao (hình dạng tổn thương) mà không cần con người can thiệp thủ công.
- Điều này giúp mô hình xử lý hiệu quả dữ liệu dạng lưới (pixel ảnh) mà vẫn tối ưu được tài nguyên tính toán.

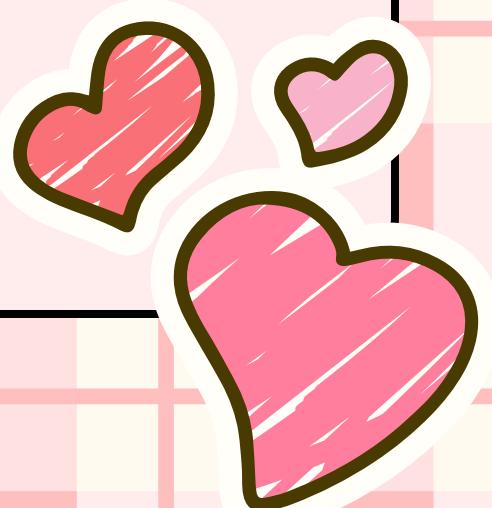
## THÀNH PHẦN CỐT LÕI CỦA CNN

- Lớp Tích chập: Dùng bộ lọc quét qua ảnh để tạo bản đồ đặc trưng.
- Hàm kích hoạt: Giúp mô hình học được các yếu tố phi tuyến tính, giảm chi phí tính toán.
- Lớp Gộp: Giảm kích thước không gian, giữ lại đặc trưng quan trọng nhất



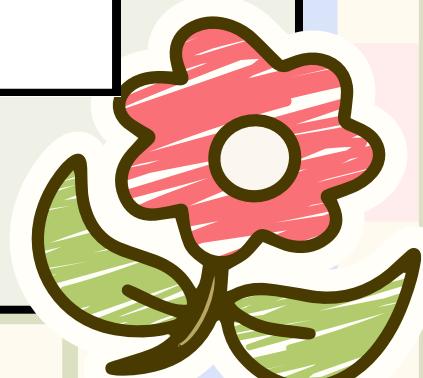
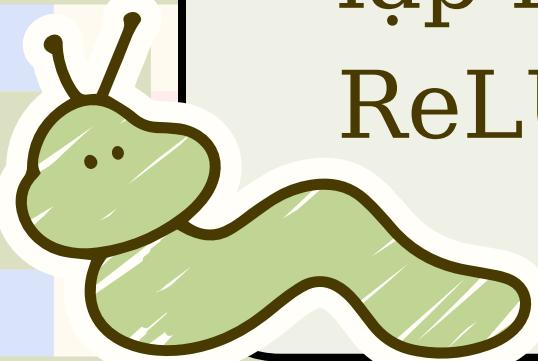
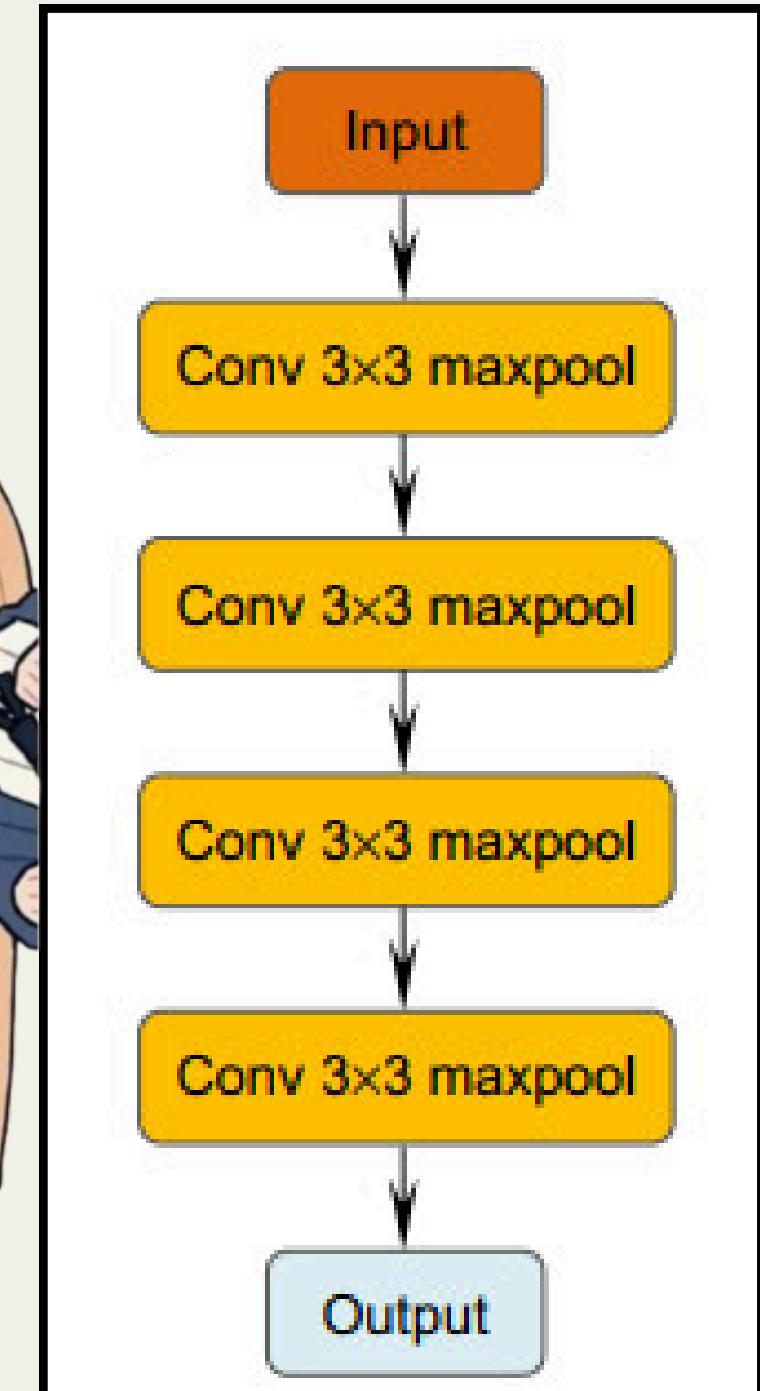


# MỘT SỐ KIẾN TRÚC MẠNG CNN



# VGGNET

- Được phát triển bởi Nhóm Hình học Thị giác của Đại học Oxford.
- Sử dụng các bộ lọc tích chập nhỏ kích thước  $3 \times 3$  thay vì các bộ lọc lớn hơn
- Cấu trúc rất đồng nhất, chủ yếu bao gồm các khối lớp tích chập  $3 \times 3$  được lặp lại, sau đó là các lớp kích hoạt ReLU và các lớp gộp cực đại



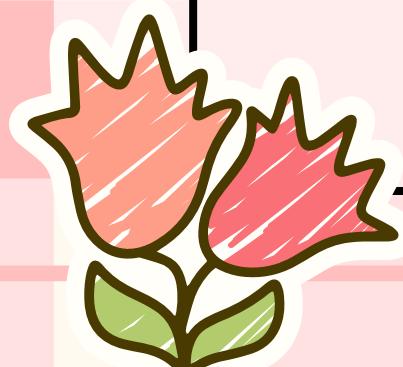
# ƯU ĐIỂM VÀ HẠN CHẾ

## ƯU ĐIỂM

- Khả năng trích xuất các đặc trưng hình ảnh rất hiệu quả.
- Xếp chồng nhiều lớp tích chập kích thước nhỏ ( $3 \times 3$ ) liên tiếp tương đương với các bộ lọc lớn.
- Tăng tính phi tuyến tính nhờ sử dụng nhiều hàm kích hoạt hơn.

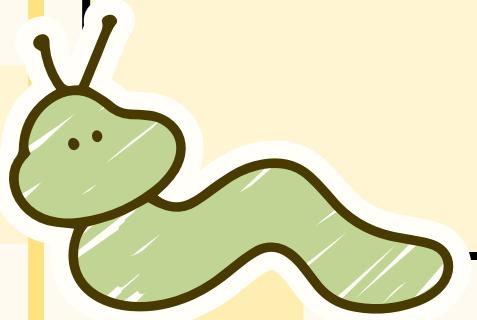
## HẠN CHẾ

- Số lượng tham số khổng lồ: 138 triệu tham số.
- Tiêu tốn nhiều tài nguyên tính toán và mất nhiều thời gian để huấn luyện.
- Gặp phải vấn đề "biến mất đạo hàm" khi mạng quá sâu.



# RESNET

- Giới thiệu vào năm 2015.
- Thiết kế để làm việc với hàng trăm lớp.
- Giải quyết hiện tượng “Biến mất đạo hàm” (Gradient Vanishing).
- Sử dụng kết nối “tắt” đồng nhất để xuyên qua một hay nhiều lớp.



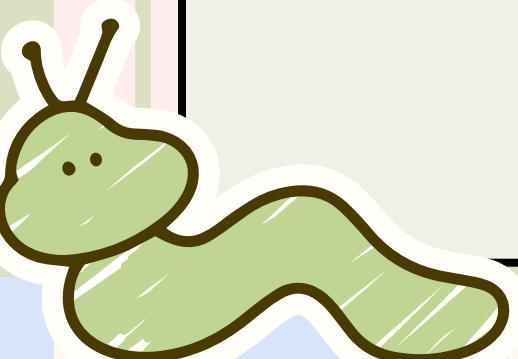
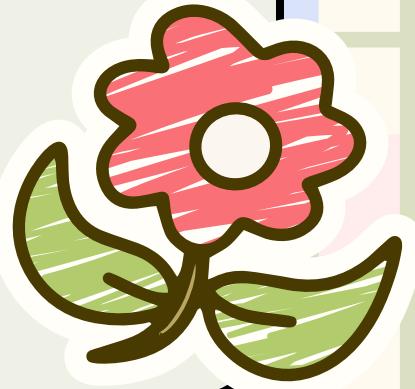
# ƯU ĐIỂM VÀ HẠN CHẾ

## ƯU ĐIỂM

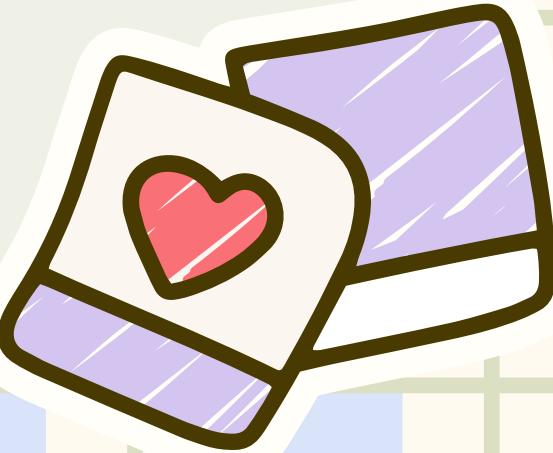
- Giải quyết vấn đề "Biến mất đạo hàm" (Vanishing Gradient).
- Huấn luyện được mạng cực sâu lên đến 152 lớp.
- Nhẹ hơn đáng kể về số lượng tham số, giảm tải tài nguyên tính toán.

## HẠN CHẾ

- Cơ chế cộng trong các kết nối tắt có thể vô tình làm mất đi một số đặc trưng ban đầu.
- Hiệu quả tham số chưa cao.

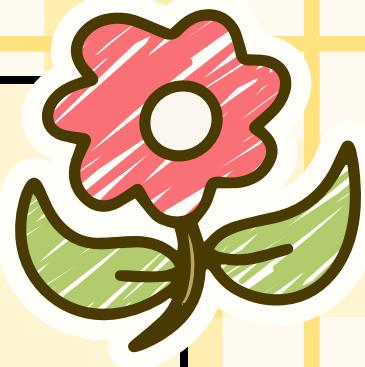
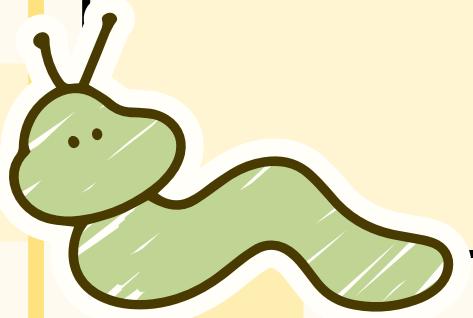


# DENSENET



# DENSENET

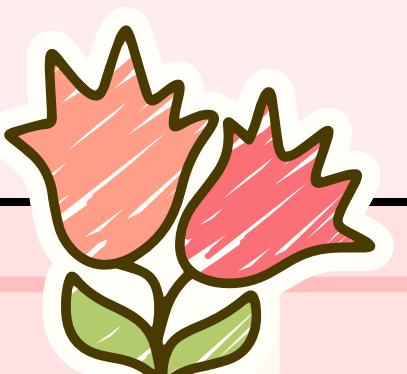
- DenseNet (Densely Connected Convolutional Networks) là một bước đột phá trong kiến trúc mạng nơ-ron sâu.
- Được giới thiệu bởi Gao Huang và cộng sự tại hội nghị CVPR 2017.
- Mạng gồm 2 thành phần chính: Khối dày đặc và Lớp truyền tiếp.



# CƠ CHẾ HOẠT ĐỘNG

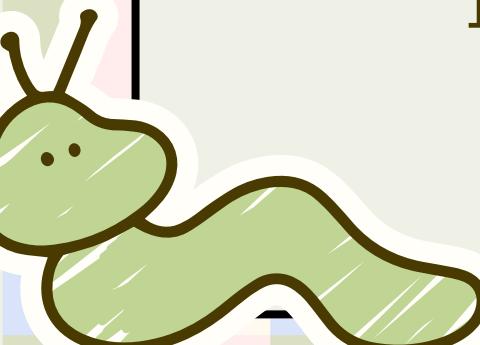
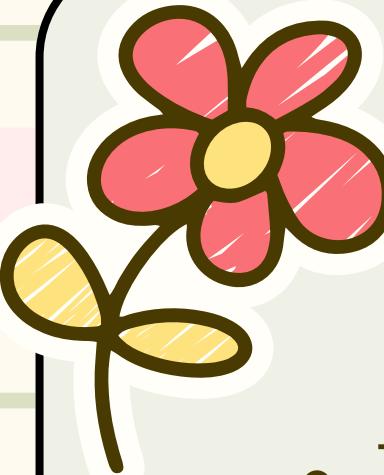


- Khác với VGGNet nơi lớp thứ L chỉ nhận đầu vào từ lớp L-1, hay ResNet sử dụng phép cộng để gộp đặc trưng
- DenseNet sử dụng phép Ghép nối theo chiều sâu nhận đầu vào từ tất cả các lớp trước đó
- Công thức:  $x_l = H_l([x_0, x_1, \dots, x_{l-1}])$
- Lớp cuối cùng có thể học trực tiếp các đặc trưng từ lớp đầu tiên cũng như các đặc trưng trùu tượng ở giữa, tạo ra một "bộ nhớ chung" cho toàn mạng.



# KHỐI DÀY ĐẶC

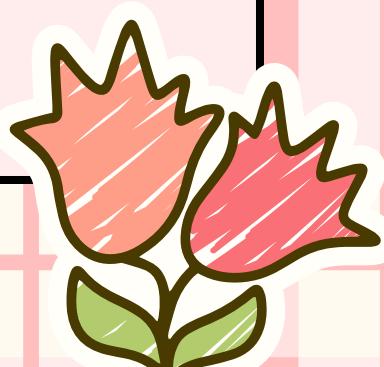
- Module chính nơi kích thước ảnh được giữ nguyên để thực hiện ghép nối.
- Tăng độ sâu của mạng và tái sử dụng đặc trưng.
- Các lớp trong một Dense Block tuân theo cấu trúc Pre-activation bao gồm: **Batch Normalization -> ReLU -> Convolution**.
- Growth Rate ( $k$ ) là siêu tham số kiểm soát độ rộng của mạng. Mỗi lớp sẽ tạo ra thêm  $k$  bản đồ đặc trưng mới và ghép vào luồng dữ liệu chính<sup>10</sup>.



# LỚP CHUYỂN TIẾP



- Việc ghép nối liên tục làm số lượng kênh tăng lên rất nhanh và kích thước ảnh cần được thu nhỏ dần.
- DenseNet sử dụng các lớp Transition nằm giữa các Dense Block để "nén" mô hình.
- Convolution 1x1: Giảm số lượng kênh đi một nửa để giảm tham số.
- Average Pooling 2x2: Giảm kích thước không gian của ảnh đi một nửa .



# ƯU ĐIỂM VÀ HẠN CHẾ

## ƯU ĐIỂM

- Giải quyết vấn đề "Biến mất đạo hàm" (Vanishing Gradient).
- Tăng cường lan truyền và tái sử dụng đặc trưng
- Yêu cầu ít tham số hơn đáng kể để đạt được cùng mức độ chính xác.
- Nhẹ hơn và tốn ít bộ nhớ lưu trữ mô hình

## HẠN CHẾ

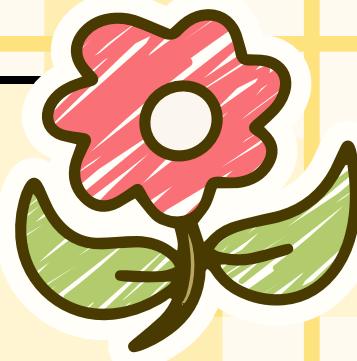
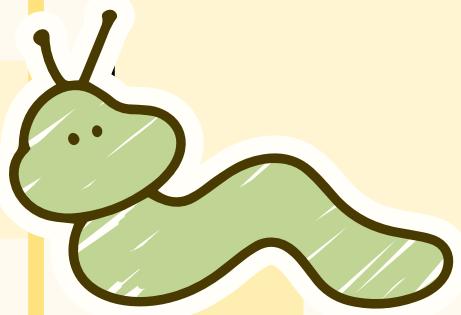
- Tiêu tốn tài nguyên bộ nhớ khi huấn luyện
- Giới hạn độ phân giải đầu vào do hệ quả của tốn bộ nhớ.
- Thời gian huấn luyện có thể lâu hơn do khối lượng phép tính lớn.



# HÀM MẤT MÁT

- Hàm mất mát sử dụng là hàm Cross-Entropy Loss
- Trừng phạt rất nặng các dự đoán sai mà mô hình lại có độ "tin cậy" cao
- Thúc đẩy gradient thay đổi mạnh mẽ để sửa lỗi nhanh chóng trong giai đoạn đầu huấn luyện.

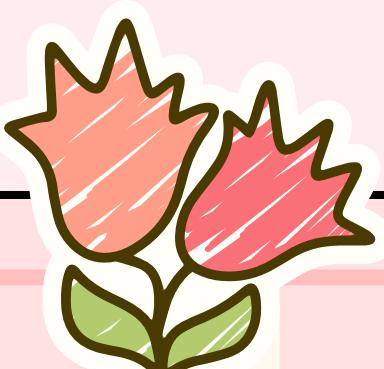
$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)]$$



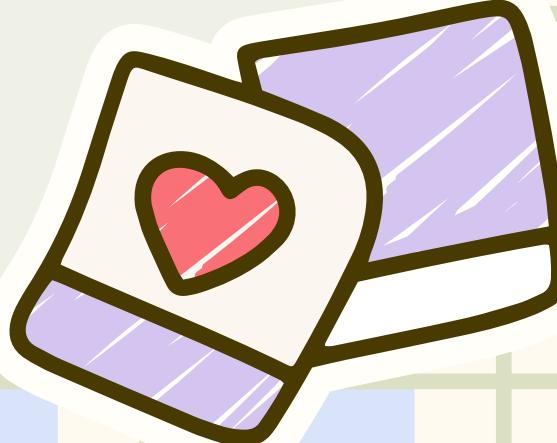
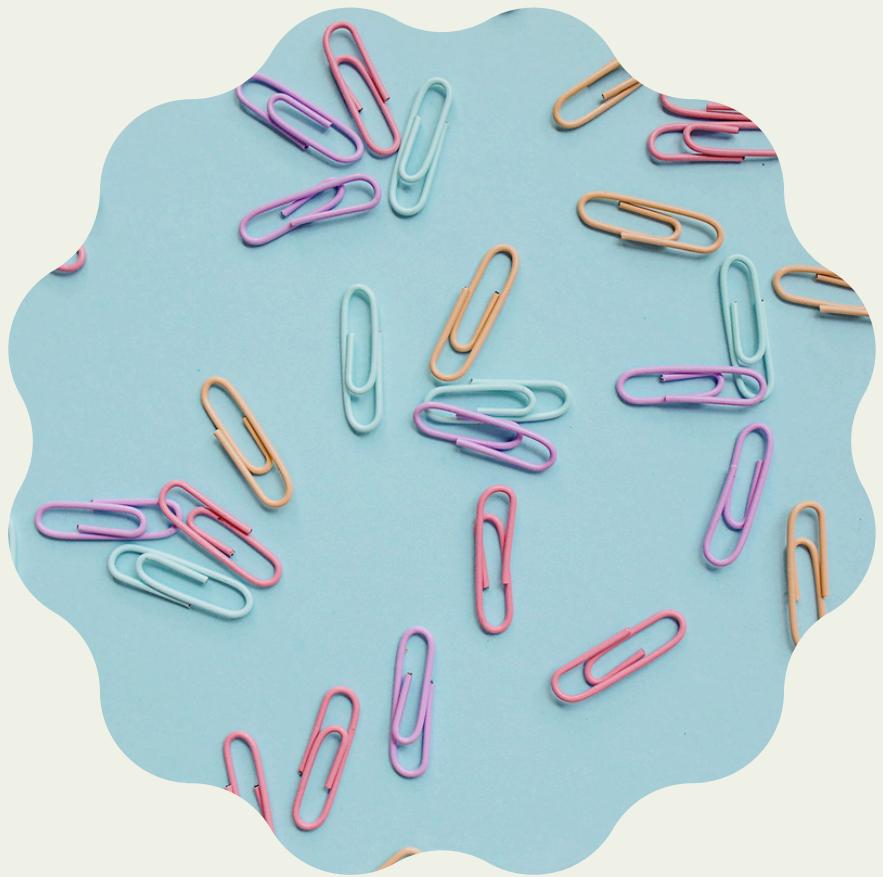
# THUẬT TOÁN TỐI ƯU HÓA



- Thuật toán Adam (Adaptive Moment Estimation) thay vì SGD truyền thống.
- Adam là sự kết hợp ưu điểm của hai thuật toán mở rộng của Gradient Descent: **Momentum** và **RMSProp**
- **Momentum**: Giúp vượt qua các điểm cực tiểu địa phương bằng đà chuyển động.
- **RMSProp**: Tự động điều chỉnh tốc độ học cho từng tham số riêng biệt.



# XÂY DỰNG MÔ HÌNH VÀ TRIỂN KHAI



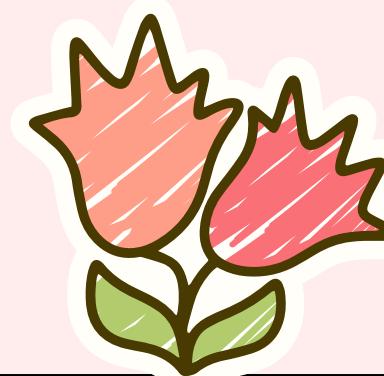
# MÔ TẢ DỮ LIỆU

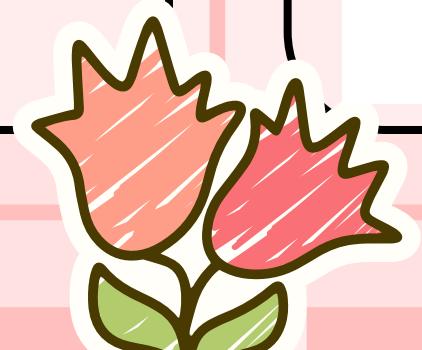
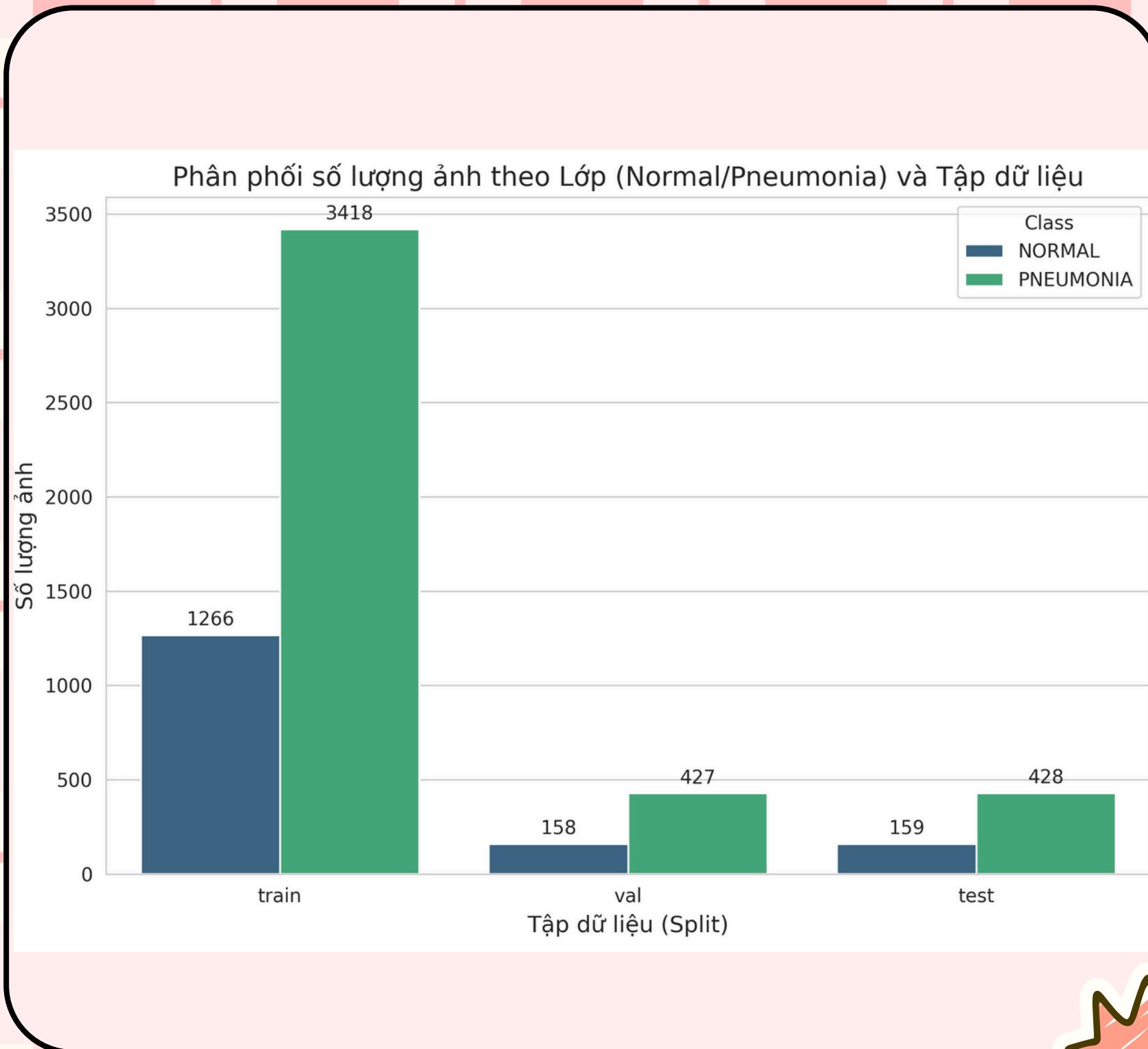
- Bộ dữ liệu chuẩn hóa "Chest X-Ray Images (Pneumonia)" công bố trên Kaggle
- Bộ dữ liệu này có tổng cộng 5,856 ảnh X-quang.
- Sự mất cân bằng lớp đáng kể, với số lượng ca bệnh Viêm phổi chiếm đa số.
- Tổng số ảnh Normal: 1,583 ảnh, tổng số ảnh Pneumonia: 4,273 ảnh.
- Các tập huấn luyện, tập xác thực, tập kiểm thử không đồng đều về số lượng ảnh.
- Giữ nguyên tỷ lệ tập dữ liệu để huấn luyện mô hình và đánh giá, kết quả trả ra có thể bị sai lệch.

# CẤU TRÚC BỘ DỮ LIỆU

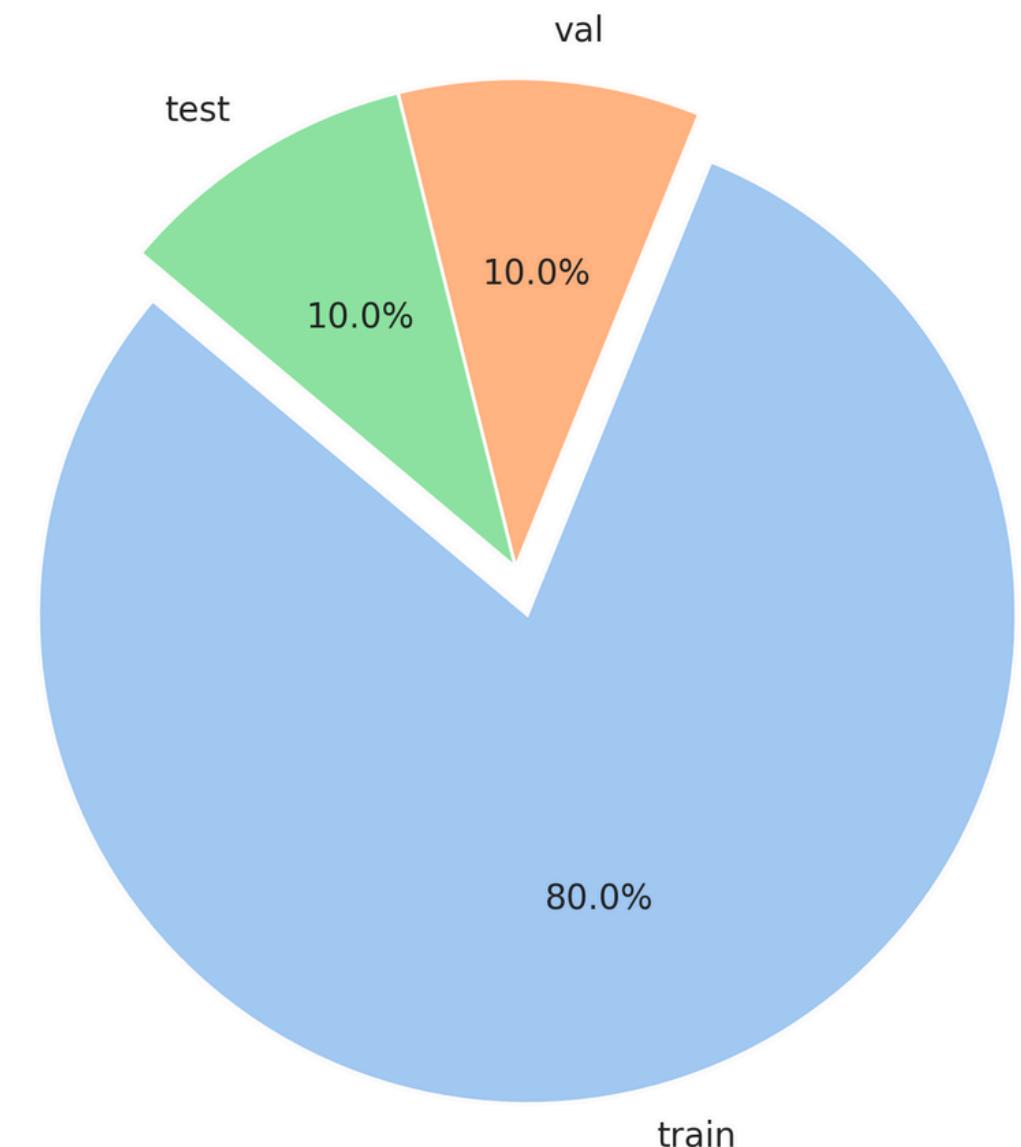


- Dữ liệu được gộp lại và chia ngẫu nhiên theo tỷ lệ 80% Train- 10% Validation- 10% Test
- **Tập Huấn luyện:** 4,684 ảnh (3,418 Pneumonia / 1,266 Normal).
- **Tập Xác thực:** 585 ảnh (427 Pneumonia / 158 Normal).
- **Tập Kiểm thử:** 587 ảnh (428 Pneumonia / 159 Normal).

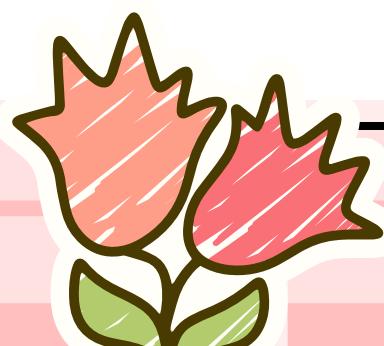
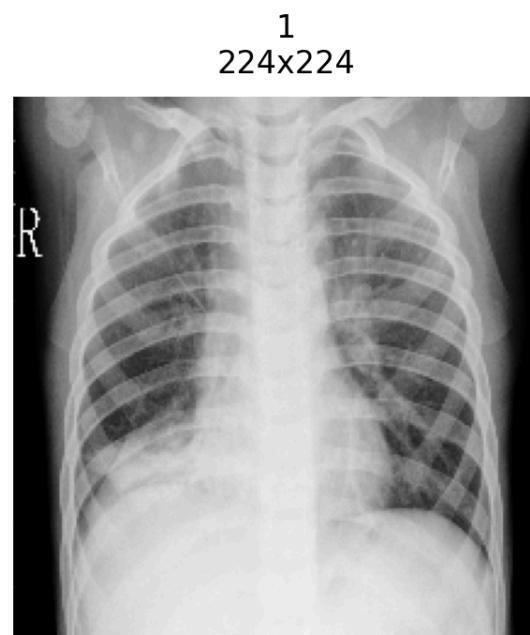
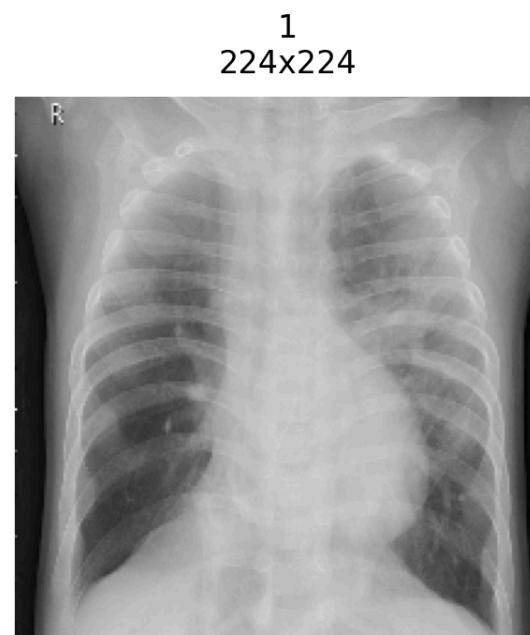
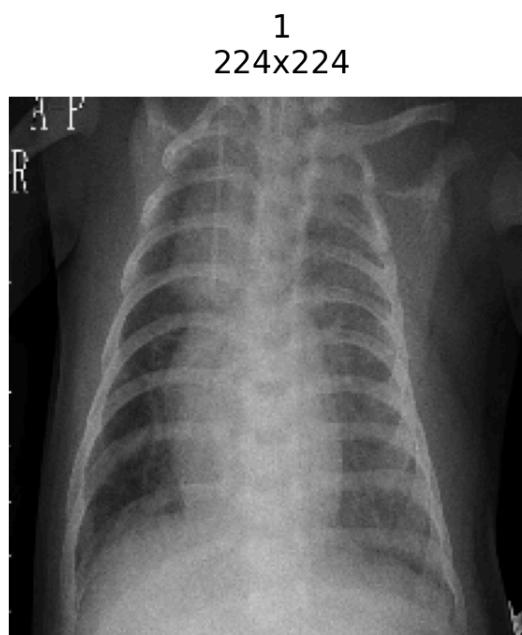
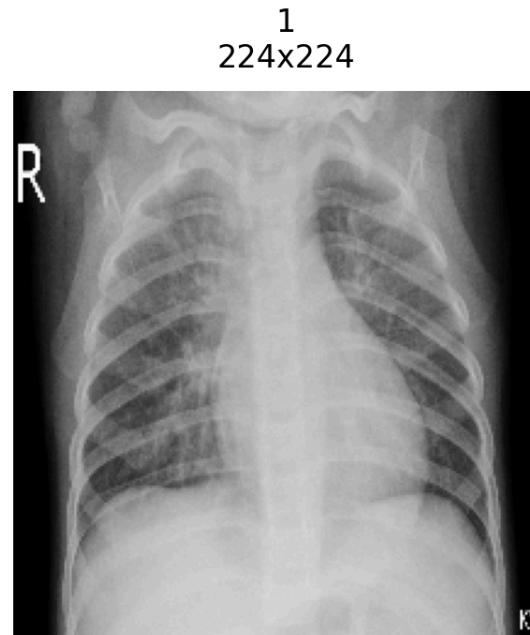
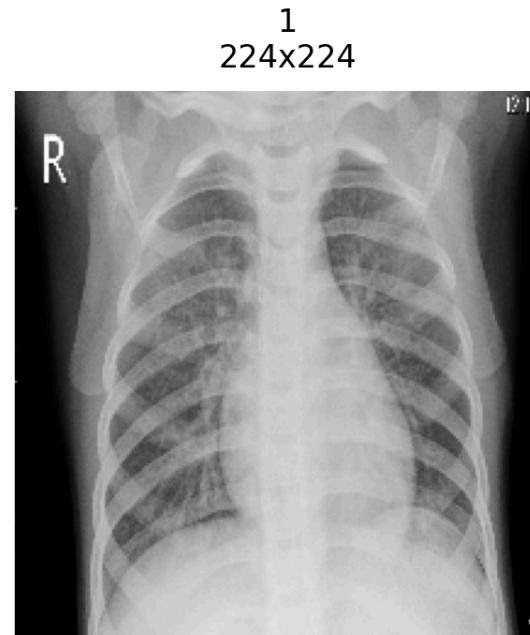
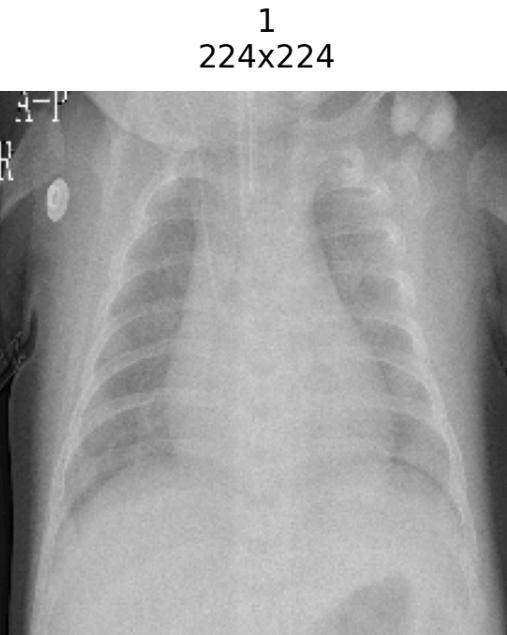




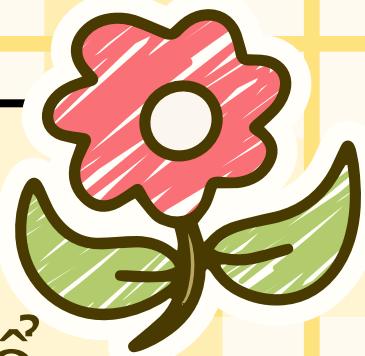
Tỷ lệ phân chia dữ liệu (Train - Val - Test)



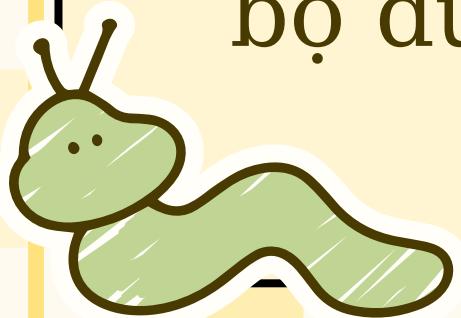
# MỘT SỐ ẢNH TỪ TẬP DỮ LIỆU



# BIẾN ĐỔI VÀ CHUẨN HÓA ẢNH



- Tất cả ảnh được thay đổi kích thước về cố định 224x224 pixels để phù hợp với đầu vào tiêu chuẩn của kiến trúc DenseNet.
- Chuyển sang định dạng ảnh xám (Grayscale) thay vì ảnh màu RGB, giúp giảm độ phức tạp tính toán.
- Dữ liệu được chuyển sang dạng Tensor, đồng thời thay đổi phạm vi giá trị pixel từ [0, 255] về [0.0, 1.0] để thuận lợi cho tính toán trên GPU.
- Các ảnh đầu vào được chuẩn hóa theo phân phối chuẩn dựa trên bộ dữ liệu ImageNet để tăng tốc độ hội tụ của mô hình.

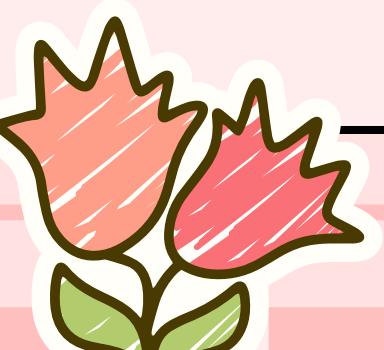


# XỬ LÝ MẤT CÂN BẰNG DỮ LIỆU

- Số lượng ảnh Viêm phổi gấp gần 3 lần ảnh Bình thường. Nhóm sử dụng kỹ thuật **Lấy mẫu ngẫu nhiên có trọng số**.
- Tính trọng số: Trọng số của mỗi lớp được tính nghịch đảo với số lượng mẫu của lớp đó
- Trong quá trình tạo Batch huấn luyện, hệ thống sẽ bốc mẫu dựa trên xác suất tỉ lệ thuận với trọng số.
- Đảm bảo tỉ lệ giữa Normal và Pneumonia trong mỗi Batch đưa vào mô hình xấp xỉ 50:50.

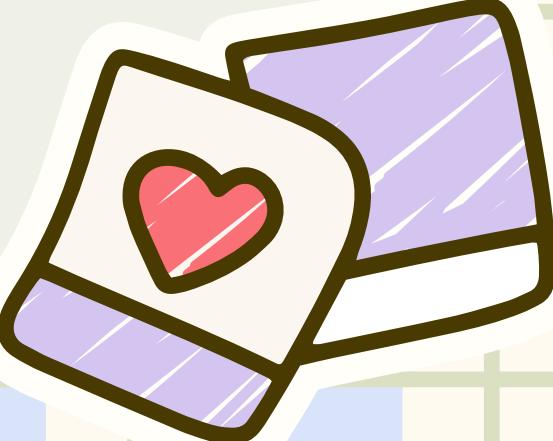
# TỔ CHỨC DỮ LIỆU

- Dữ liệu sau xử lý được đóng gói vào các DataLoader của PyTorch để nạp vào mô hình.
- **Train Loader:** Sử dụng batch size = 32 và áp dụng Weighted Sampler để cân bằng dữ liệu.
- **Validation & Test Loaders:** Sử dụng batch size = 32 nhưng không dùng sampler để đánh giá khách quan hiệu năng mô hình.

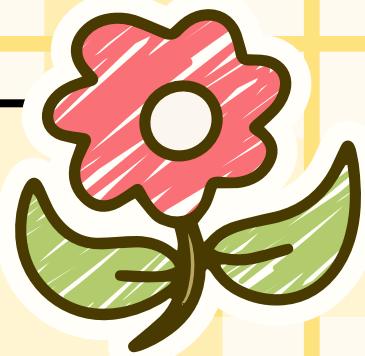


# XÂY DỰNG MÔ HÌNH

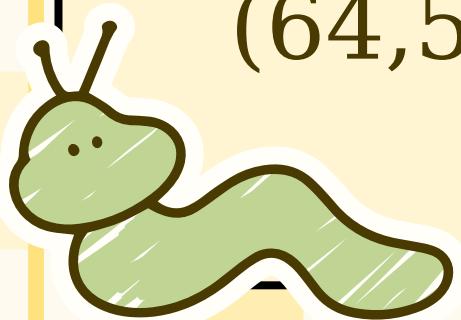
- Tối ưu lượng tài nguyên tính toán để có thể chạy được trên nhiều thiết bị từ thấp đến cao
- Mô hình phải có khả năng tính toán với thông lượng lớn và độ chính xác cao.
- Chọn mô hình mạng DenseNet là mô hình nền tảng để giải quyết bài toán đặt ra.
- Tự cài đặt và thiến kế kiến trúc DenseNet
- Giúp tối ưu hóa mô hình cho dữ liệu ảnh đen trắng, giảm thiểu dữ thừa tham số và cho phép kiểm soát hoàn toàn luồng xử lý dữ liệu.



# KHỐI ĐẦU VÀO



- Chịu trách nhiệm giảm kích thước không gian của ảnh gốc và trích xuất các đặc trưng.
- **Đầu vào:** Ảnh Tensor kích thước (Batch, 1, 224, 224)
- **Convolution**  $7 \times 7$  với stride = 2 và padding = 3 giúp giảm kích thước ảnh xuống còn một nửa và tăng độ sâu lên 64 kênh.
- **MaxPooling**  $3 \times 3$  với stride=2: tiếp tục giảm kích thước ảnh xuống còn một nửa nữa thông qua lớp gộp cực đại
- **Kết quả:** Đầu ra của khối này là một Feature Map có kích thước (64, 56, 56) là đầu vào cho chuỗi Dense Blocks phía sau.

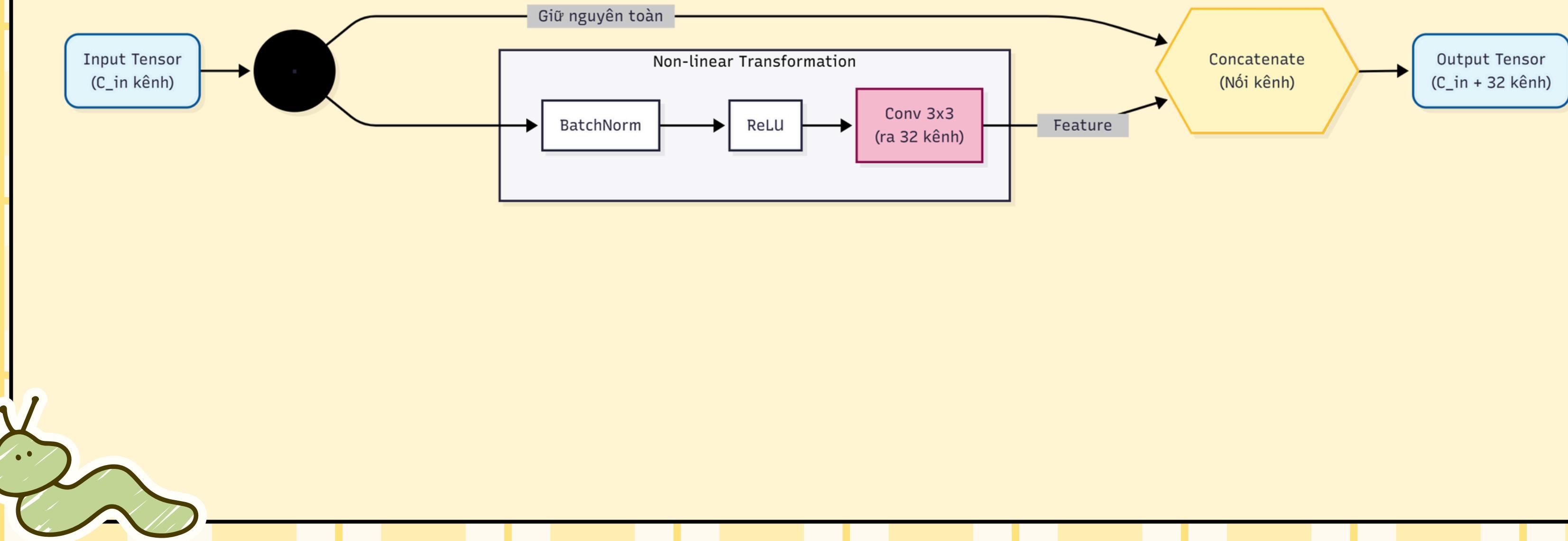


# KHỐI DÀY ĐẶC

- Trái tim của kiến trúc, nơi diễn ra quá trình "Tái sử dụng đặc trưng".
- Bao gồm 4 Dense Block với số lượng lớp lần lượt là [4,4,8,8].
- **Pre-activation:** Batch Normalization → ReLU → Convolution.  
Thứ tự này giúp luồng gradient truyền ngược ổn định hơn.
- **Convolution 3×3:** Thực hiện tích chập để tạo ra k đặc trưng mới
- **Cơ chế ghép nối:** Đầu ra của lớp này không đè lên đầu vào mà  
được ghép nối vào phía sau theo chiều sâu.

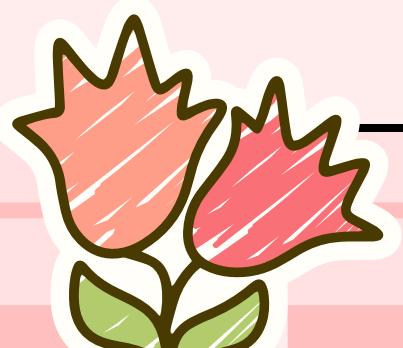


# KHỐI DÀY ĐẶC

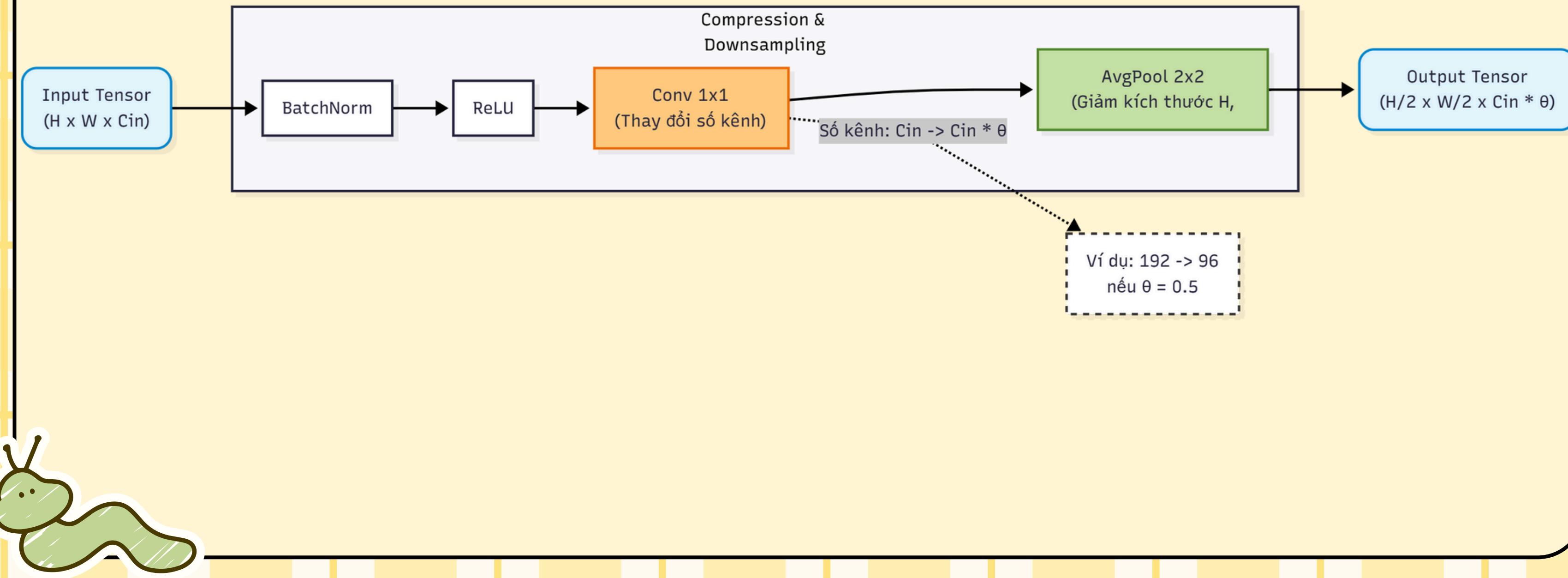
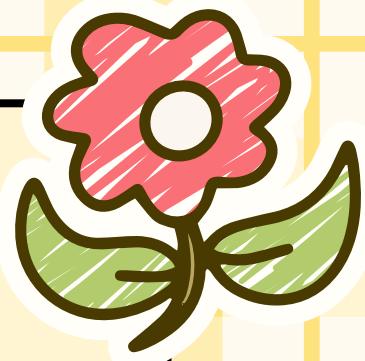


# LỚP TRUYỀN TIẾP

- Cơ chế ghép nối liên tục làm số lượng kênh tăng lên rất nhanh gây bùng nổ tham số.
- Các lớp Transition được đặt giữa các Dense Block để thực hiện nhiệm vụ "nén" mô hình.
- **Conv  $1 \times 1$** : Đóng vai trò giảm số lượng kênh đi một nửa
- **Average Pooling  $2 \times 2$** : Giảm kích thước không gian của ảnh đi một nửa



# LỚP TRUYỀN TIẾP



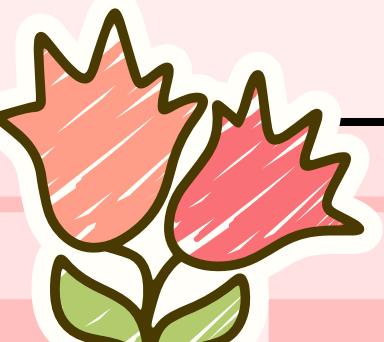
# LỚP PHẦN LOẠI

- Sau khi ảnh đi qua 4 khối Dense Block, kích thước ảnh chỉ còn lại  $7 \times 7$  nhưng chứa đựng thông tin đặc trưng rất sâu.
- Batch Norm & ReLU: Chuẩn hóa đặc trưng lần cuối.
- Biến đổi Feature Map từ kích thước  $(C, 7, 7)$  thành vector  $(C, 1, 1)$ .
- Flatten: Dát phẳng thành vector 1 chiều.
- Fully Connected: Lớp kết nối đầy đủ duy nhất trong mạng, ánh xạ số lượng đặc trưng về 2 giá trị tương ứng với lớp Normal và Pneumonia.

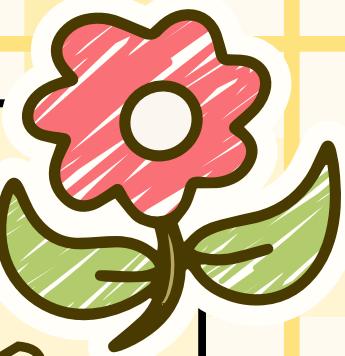
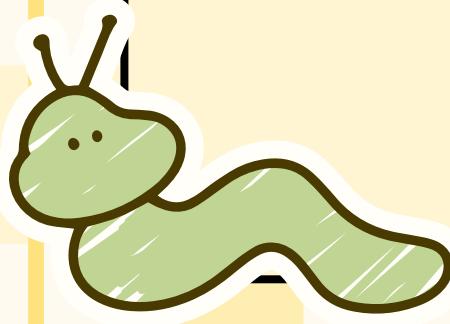


# TĂNG CƯỜNG DỮ LIỆU

- Một pipeline tăng cường dữ liệu mạnh mẽ được áp dụng chỉ riêng cho tập huấn luyện.
- Xoay ảnh ngẫu nhiên trong khoảng  $\pm 10$  độ để mô phỏng sự sai lệch tư thế chụp.
- Lật ảnh trái/phải với xác suất 50%.
- Thay đổi nhẹ độ sáng và độ tương phản  $\pm 10\%$  để mô hình thích nghi với các điều kiện phơi sáng khác nhau.
- Đưa pixel về phân phối chuẩn với  $\text{mean}=0.5$ ,  $\text{std}=0.5$ .

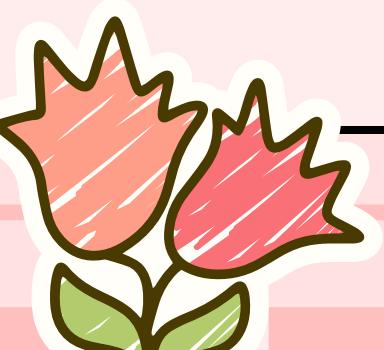


# THỰC NGHIỆM VÀ ĐÁNH GIÁ



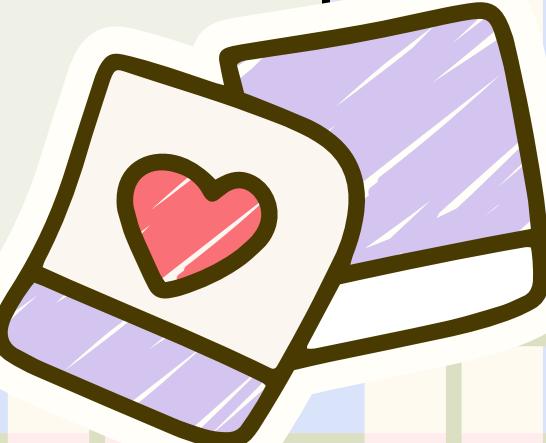
# THIẾT LẬP MÔI TRƯỜNG

- Toàn bộ kiến trúc mô hình và dữ liệu được chuyển sang hoạt động trên GPU thông qua cơ chế của PyTorch.
- **Nền tảng:** Google Colab
- **GPU:** NVIDIA Tesla T4 (VRAM 16GB)
- **RAM:** 12GB
- **Framework:** PyTorch kết hợp với thư viện torchvision để xử lý dữ liệu ảnh, Scikit-learn để tính toán các chỉ số đánh giá và Matplotlib dùng để trực quan hóa

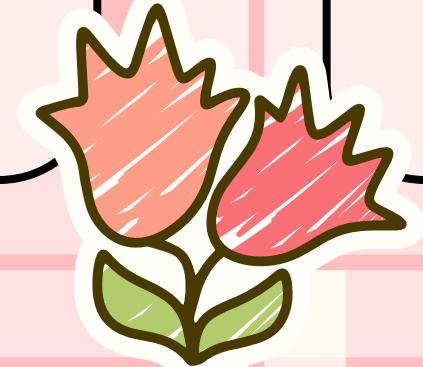
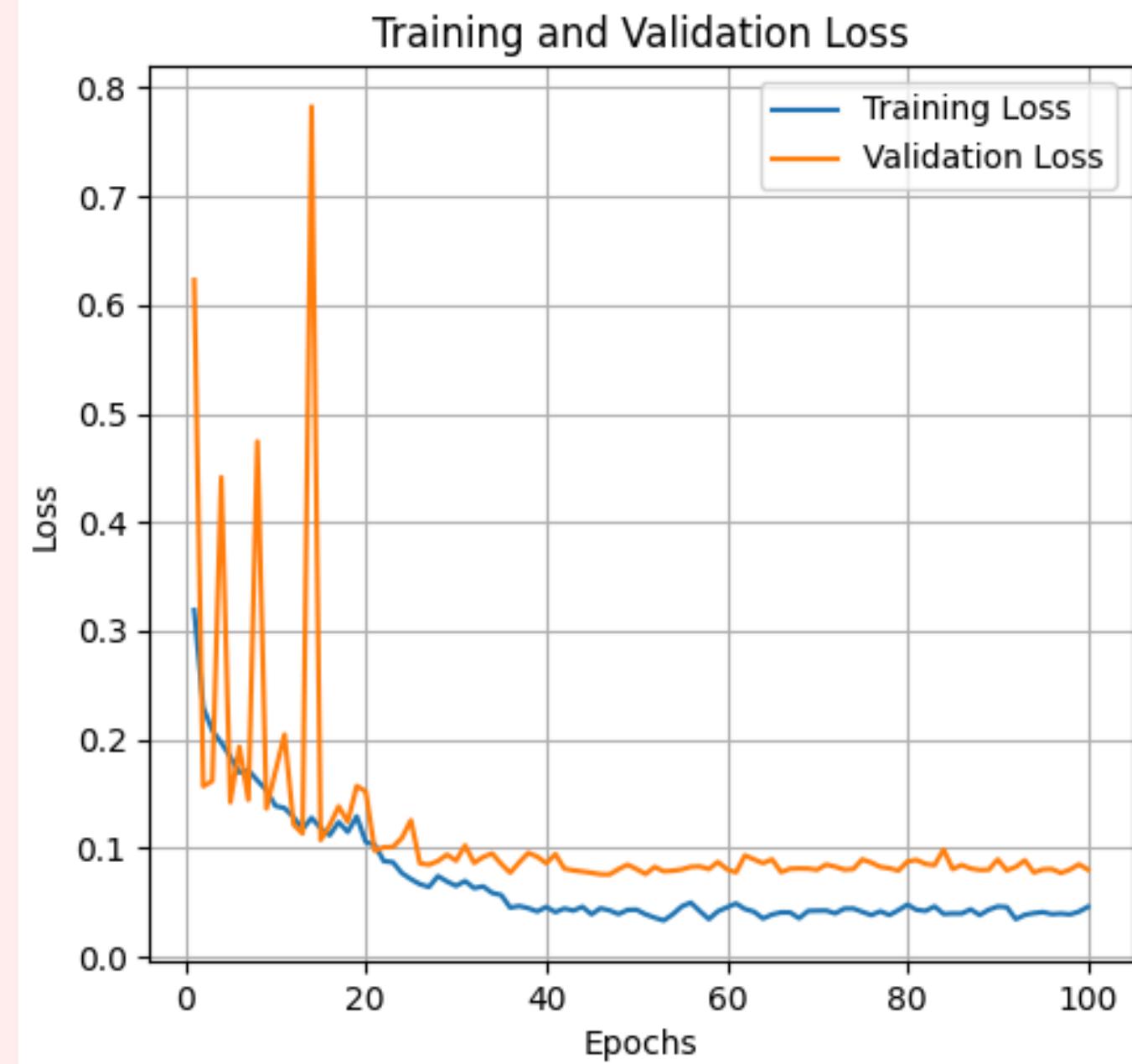
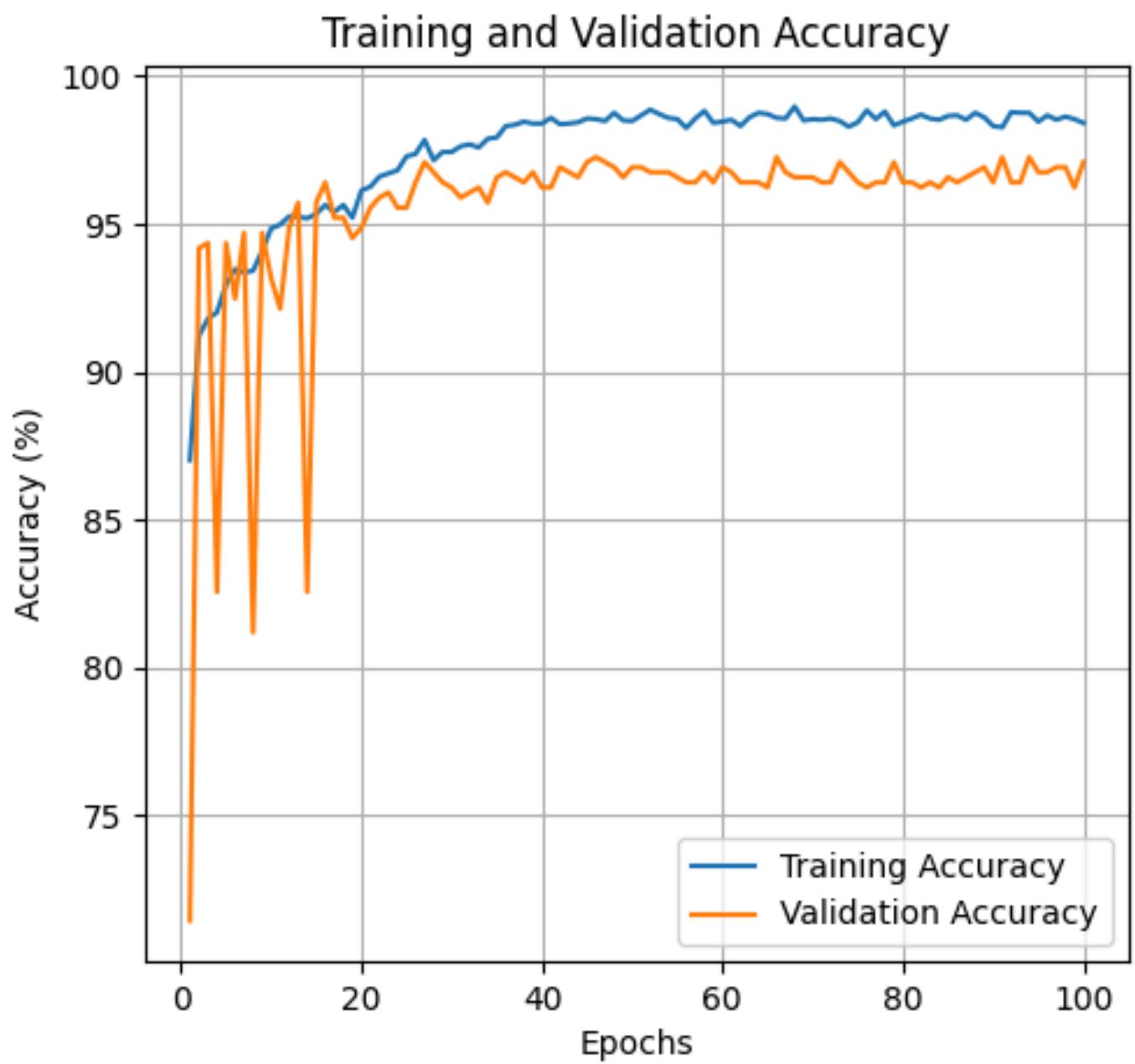


# SIÊU THAM SỐ VÀ TỐI ƯU HÓA

- Thuật toán tối ưu hóa: Adam
- Tốc độ học: 0.001
- Kích thước Batch: 32
- Số vòng lặp (Epochs): 100
- Kích thước đầu vào: 224 x 224
- Hàm mất mát: Cross Entropy
- Suy giảm trọng số: 1e-4



# QUÁ TRÌNH HUẤN LUYỆN



# KẾT QUẢ

- Accuracy (97.27%)
- Precision (97.29%)
- Recall (97.27%)
- F1-Score (97.28%)

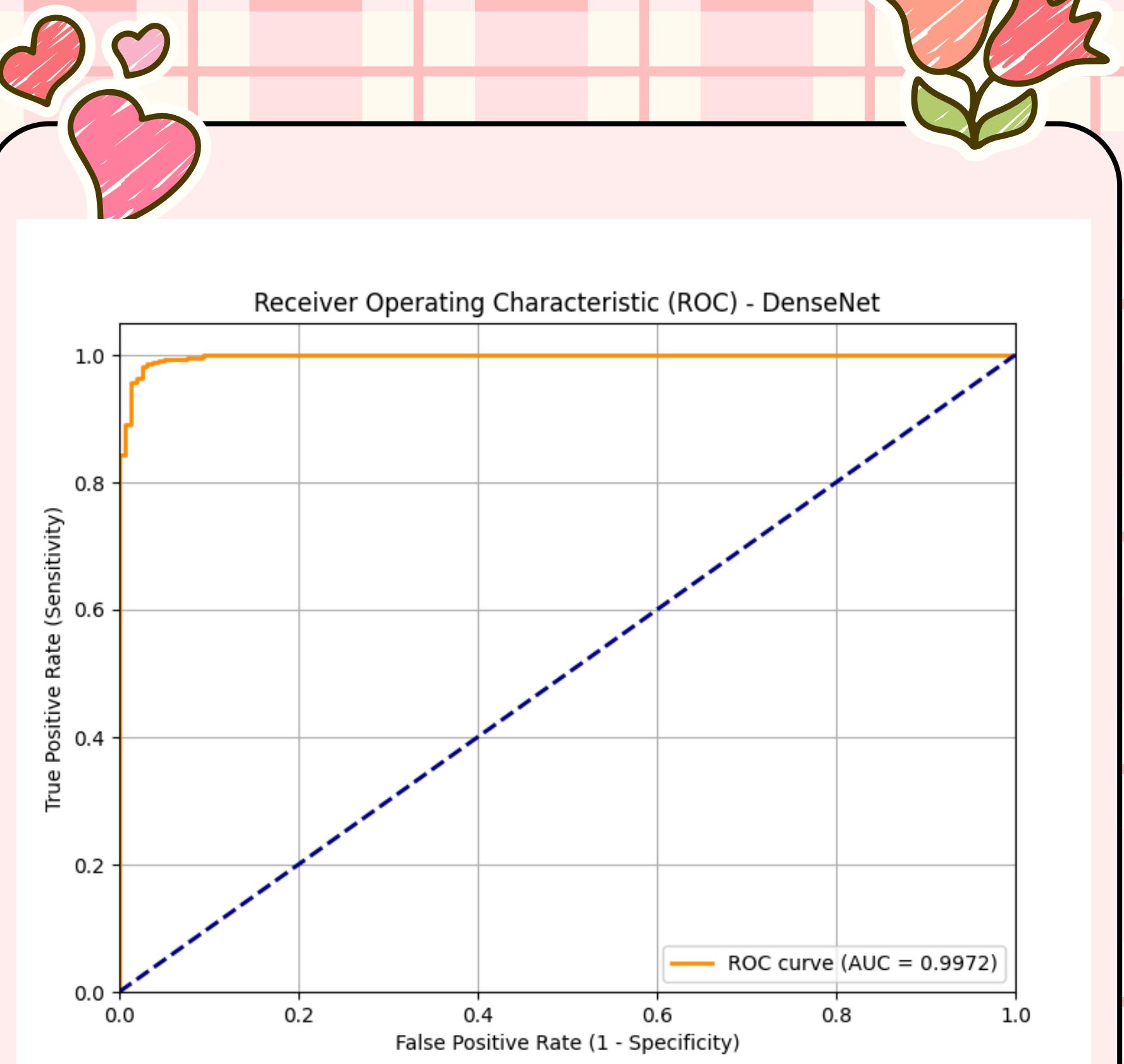
Confusion Matrix - Pneumonia Detection

		NORMAL	PNEUMONIA
AL	NORMAL	154	5
	PNEUMONIA	7	421

0.9574 | Precision: 0.9575 | Recall: 0.9574 | F1-Score: 0.9575 | AUC:

# KẾT QUẢ

- AUC (0.9956)
- Chỉ số diện tích dưới đường cong ROC tiệm cận 1
- Khẳng định khả năng phân tách tuyệt vời giữa hai lớp bệnh và thường



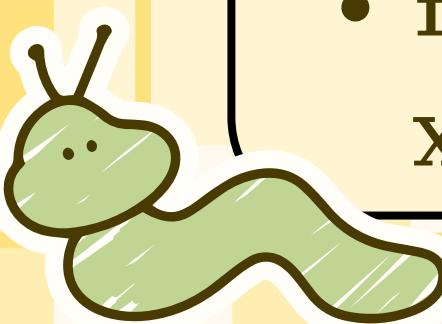
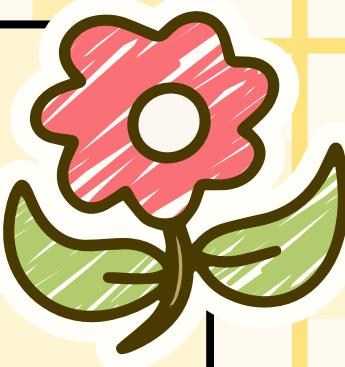
# SO SÁNH VỚI VGGNET VÀ RESNET

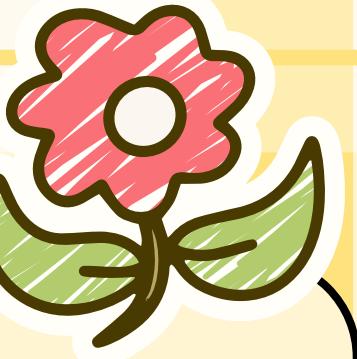
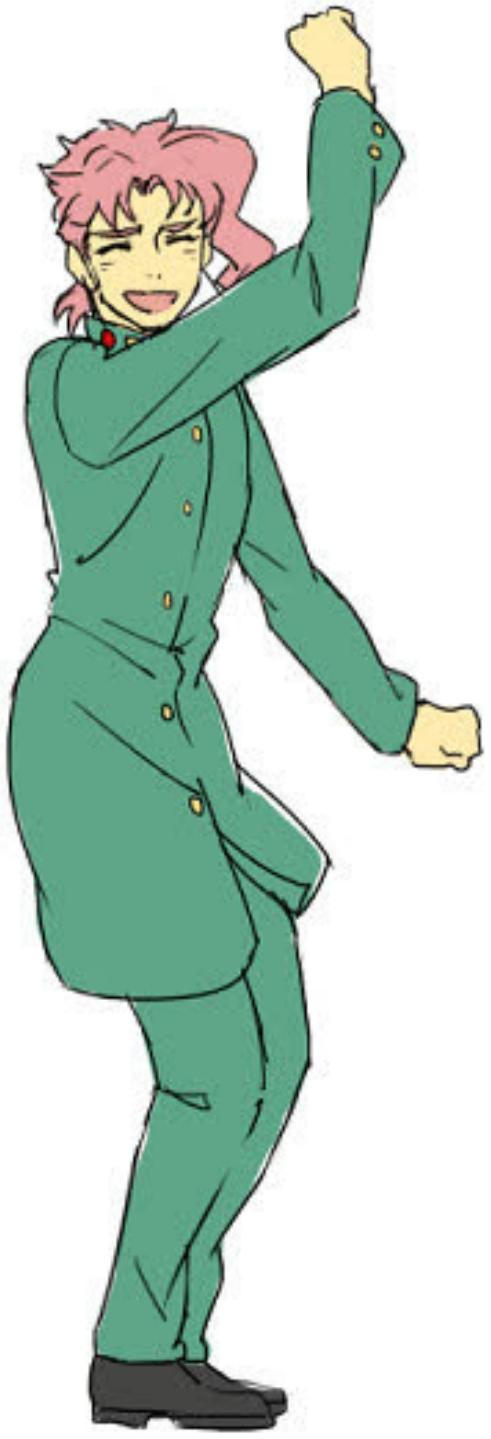
Mô hình	Accuracy	Precision	Recall	F1-Score	AUC
VGGNet	96.59%	96.69%	96.59%	96.62%	0.9943
ResNet	<b>97.61%</b>	<b>97.61%</b>	<b>97.61%</b>	<b>97.60%</b>	0.9953
DenseNet (Đề xuất)	97.27%	97.29%	97.27%	97.28%	<b>0.9956</b>

So sánh hiệu năng giữa mô hình DenseNet và các mô hình khác trên tập kiểm thử

# HƯƠNG PHÁT TRIỂN

- Tăng độ phân giải đầu vào từ 224 x 224 lên 512 x 512, v.v.
- Áp dụng kiến trúc mới, nghiên cứu chuyển sang các mô hình Vision Transformer
- Phân loại đa lớp: phân biệt cụ thể nguyên nhân: Vi khuẩn, Virus (COVID-19), hay Nấm.
- Định vị và khoanh vùng chính xác vị trí và tính toán diện tích vùng viêm.
- Đánh giá mô hình trên dữ liệu từ nhiều bệnh viện khác nhau để đảm bảo độ chính xác.
- Đóng gói mô hình thành Web App/Mobile App cho chẩn đoán từ xa.





THANK YOU

