**VIỆN NGHIÊN CỨU DỮ LIỆU LỚN VINGROUP - VINBIGDATA**

**CHƯƠNG TRÌNH PHÁT TRIỂN NGUỒN NHÂN LỰC AI**



Deep Learning

**PHÁT HIỆN BẤT THƯỜNG TRONG ẢNH X-QUANG LỒNG NGỰC  
SỬ DỤNG UNET VÀ DAE**

Trần Bảo Sam 3658429

Nguyễn Hồng Phúc 3658420

Đỗ Quốc Cường 3658411

Nguyễn Thành Nhân 3658389

Nguyễn Quốc Anh

**Hà Nội, ngày 17 tháng 01 năm 2021**

# MỤC LỤC

[MỤC LỤC 2](#_Toc61809447)

[Chương I 3](#_Toc61809448)

[TỔNG QUAN 3](#_Toc61809449)

[Phát hiện bất thường trên dữ liệu ảnh y tế về phổi 3](#_Toc61809450)

[Một số phương pháp thường dùng 3](#_Toc61809451)

[Giải pháp 4](#_Toc61809452)

[Chương II 5](#_Toc61809453)

[DỮ LIỆU VÀ MÔ HÌNH 5](#_Toc61809454)

[Bộ dữ liệu VinBigData 5](#_Toc61809455)

[Phân khúc ảnh bằng Unet 6](#_Toc61809456)

[Mô hình DAE 10](#_Toc61809457)

[Chương III 13](#_Toc61809458)

[KẾT QUẢ THỰC NGHIỆM 13](#_Toc61809459)

[Kết quả Unet 13](#_Toc61809460)

[Kết quả của mô hình DAE 17](#_Toc61809461)

[TÀI LIỆU THAM KHẢO 22](#_Toc61809462)

# 

# Chương I

# TỔNG QUAN

## Phát hiện bất thường trên dữ liệu ảnh y tế về phổi

Các bệnh lý về phổi ngày càng là mối quan tâm lớn của cộng đồng. Ung thư phổi và các bệnh lý khác ngày càng vấn đề quan trọng cho các quốc gia đông dân và dân số đang trở nên già hóa như Trung Quốc và Việt Nam.

Ở Việt Nam, không khí ô nhiễm đã trở thành vấn đề đáng báo động. Hai thành phố lớn nhất Việt Nam thường xuyên nằm đầu danh sách xếp hạng các thành phố ô nhiễm nhất thế giới. Điều này tạo ra nguy cơ lớn cho cư dân sinh sống tại khu vực thành phố lớn và thúc đẩy nhu cầu khám chữa bệnh liên quan đến phổi.

Trong đại dịch Covid-19, sự bùng nổ các ca bệnh khiến các bệnh viện trở nên quá tải. Khi đó, các hệ thống nhận dạng và phân tích tự động dựa trên ảnh y tế đã mang lại hiệu quả to lớn.

Với những lý do trên, việc xây dựng các hệ thống nhận diện tự động đã trở thành một nhu cầu cấp thiết. Các mô hình này hỗ trợ hiệu quả cho các bác sĩ trong việc sàng lọc ban đầu để chuẩn bị cho quá trình phân tích sâu hơn của bác sĩ.

*Phát hiện bất thường trên dữ liệu ảnh y tế* là một trong những bài toán quan trọng nhằm tạo lập nền tảng cho các hệ thống trên.

## Một số phương pháp thường dùng

* *Object detection:* Giống như các bài toán phát hiện và khoanh vùng đối tượng trong CV, ta có thể áp dụng phương pháp này vào bài toán.
* *Các model Classify:* Phân loại ảnh x-ray theo loại bệnh …

*Hạn chế chung của các phương pháp trên:*

* Sự mất cân bằng dữ liệu giữa bất thường và bình thường.
* Có những bệnh hiếm gặp và ít dữ liệu.

*Hạn chế từng phương pháp:*

* *Object detection:*  Chỉ có thể nhận diện được những loại bệnh đã được train. Việc có thêm bệnh sẽ cần train lại mô hình tuy nhiên có thể giải quyết bằng phương pháp của bài toán nhận diện khuôn mặt. Tuy nhiên với những đối tượng chưa gắn nhãn thì phương pháp này chưa thể chủ động phát hiện ra những đối tượng đấy.
* *Classify:* Gặp vấn đề như trên và thiếu thuyết phục do không đưa ra được lý do cho việc tại sao 1 tấm x-ray lại có bệnh A, B, C, ….

## Giải pháp

Vì các vấn đề kể trên, chúng tôi đề xuất ra một quy trình như sau để giải quyết:

*Input:* Ảnh X-ray.

*Output:* Ảnh x-ray có khoanh vùng đối tượng bất thường và gắn nhãn cho chúng.

* B1: Dùng U-net để phân vùng phổi, lấy vùng dữ liệu phổi khỏi X-ray. Output của thuật toán này là Mask - Tạm gọi là Mask undfined (MU).
* B2: Dùng DAE để Khôi phục Mask - Tạm gọi là Mask normal (MN).
* B3: Giao MU và MN để detect các vùng bất thường lớn.
* B4: Dùng MN để cắt phần phần phổi ra khỏi ảnh tạ gọi là Lung- Undifined (LU).
* B5: Dùng DAE trê LU để khôi phục phần phổi trên thành phổi bình thường tạm gọi là Lung Normal (LN).
* B6: Giao LN, LU lại ta sẽ lấy ra được những điểm bất thường.
* B7: Đưa những điểm bất thường vô mô hình phân lớp (xây dựng theo mô hình nhận diện khuôn mặt) để classify.
* B8: Cập nhật những điểm dị thường chưa có class -> gắn nhãn để update mô hình hoặc thêm class.

*Scale down bài toán:* Để thử nghiệm phương pháp này, chúng tôi thử nghiệm mô hình để nhận diện bệnh Pneumothorax dựa trên ảnh X-ray.

# 

# Chương II

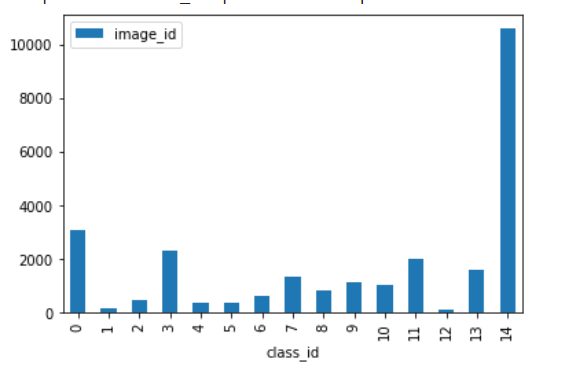
# DỮ LIỆU VÀ MÔ HÌNH

## Bộ dữ liệu VinBigData

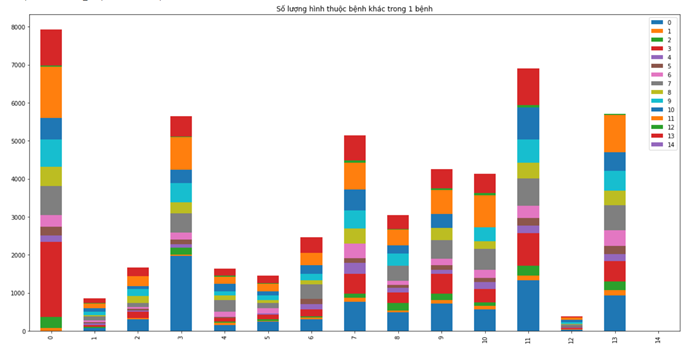
Sử dụng 1 phần của bộ dữ liệu “VinBigData Chest X-ray Abnormalities Detection" dùng trong cuộc thi “VinBigData Chest X-ray Abnormalities Detection" trên Kaggle (VinBigData, 2021).

**Đặc trưng:**

* Số lượng dữ liệu về phổi bình thường rất lớn, các bệnh khác có tỷ lệ ít một số còn nhỏ (Ảnh về số lượng ảnh X-ray mỗi bệnh):



* Có sự chồng chất nhiều bệnh trên một ảnh:



* Các điểm pixel của phổi phân bố có cấu trúc (Hình cắt ngang qua phổi ở tọa độ 60 và ảnh gốc của phổi):

|  |  |
| --- | --- |
|  |  |

## Phân khúc ảnh bằng Unet

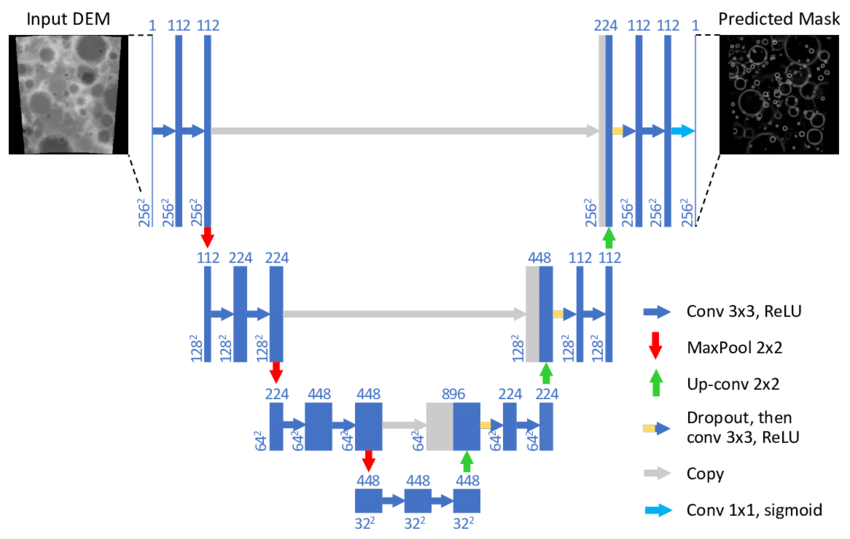
*Kiến trúc model:*

Đã có rất nhiều nghiên cứu về bài toán phân khúc ảnh (semantic segmentation), nhiều hơn về vấn đề phân khúc phổi. Một bài nghiên cứu nổi bật giới thiệu Unet dùng mạng tích chập để phân khúc ảnh y tế (Ronnerberger et al, 2015).

Mạng Unet bao gồm phần encoder và decoder. Phần encoder là mạng sâu tích chấp làm giảm kích cỡ biểu diễn của ảnh đầu vào. Mạng có thể có nhiều cấu trúc khác nhau, thừa kế từ các mạng nổi tiếng khác như VGG16, Resnet, hay Mobilenet.

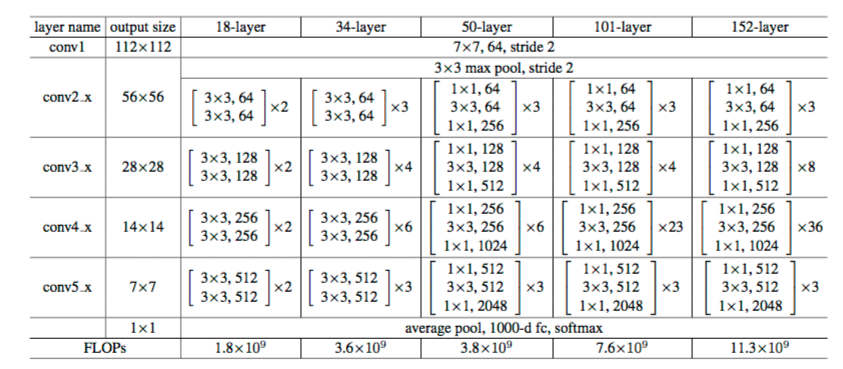
Điểm chung là encoder đều sử dụng nhiều lớp tích chập với bộ lọc (3,3) và Maxpool, có thể kết hợp xen kẽ với Batch Normalization và Dropout. Encoder được kết nối với một không gian biểu diễn (latent space), và ở bên kia không gian là decoder. Decoder là một mạng để tăng kích cỡ biểu diễn, có thể dùng lớp UpSampling hay TransposedConvolution, về kích cỡ giống đầu vào, tuy chỉ có một kênh (channel).

Unet có cầu trúc giống autoencoder, nhưng điều làm Unet đặc biệt là những kết nối nhảy cóc (skip connection) từ encoder đến decoder ở những lớp với output kích cỡ bằng nhau. Phần này của cấu trúc được thừa hưởng từ Resnet, với mục tiêu làm tăng khả năng nhận biết tương quan với độ sâu, mà không bị gradient explosion hay vanishing. Chúng ta coi bài toán như là phân loại từng pixel của ảnh. Do mục đích là của bài toán là tìm ra mask của phổi, output của mạng là một ma trận nhị phân có cùng kích cỡ với input, với pixel biểu diễn phổi có giá trị 1, còn lại là 0.

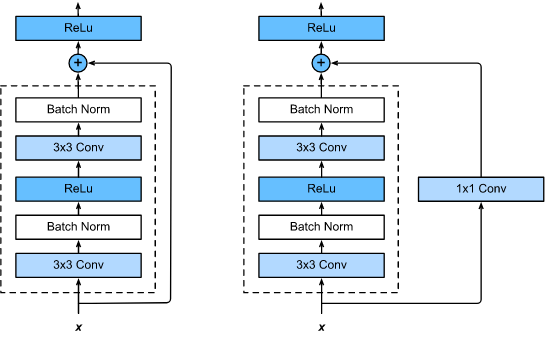


*Unet*

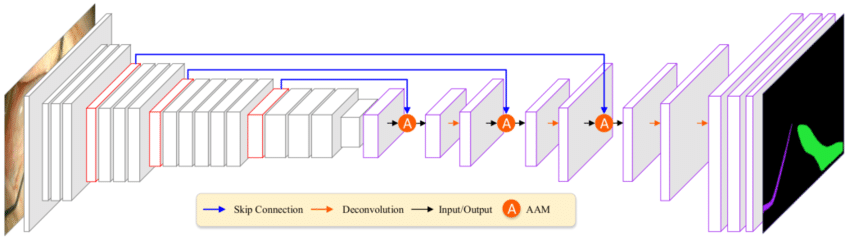
Để tăng khả năng của Unet, chúng tôi đã chọn viết Resnet18 hoặc 34 cho cấu trúc của encoder. Không chỉ có kết nối giữa encoder và decoder, encoder của Unet của chúng tôi bao gồm kết nối nhảy cóc mỗi 2 hộp convolution. Kết cấu encoder như sau:



Kết cấu Resnet, bao gồm 18, 34, 50, 101, 152 lớp



*Kết nối nhảy cóc - Skip connection từ Resnet*



*Cấu trúc Resnet*

*Thực nghiệm*

Nhóm cài đặt Unet với encoder là Resnet 18 hoặc 34. Decoder chúng tôi có lựa chọn dùng cả phương pháp UpSampling và TransposedConvolution. Dropout và BatchNormalization cũng được dùng nhằm tăng khả năng của model. Tuy vậy sau khi viết xong encoder, chúng tôi so sánh thấy model 34 chỉ có 11,2 triệu tham số, thiếu 200,000 tham số so với thống kê những nghiên cứu khác (Điều tương tự xảy ra với Resnet18).

*Training*

Trong 138 ảnh và mask phổi từ bộ Montgomery, chúng tôi chọn 110 ảnh để train, 28 ảnh để validate, và 566 ảnh ở bộ Shenzhen để test. (Lí do là chỉ train ít vậy, là do bộ Shenzhen chúng tôi tìm được muộn, nên với hạn chế về thời gian và tài nguyên, chúng tôi chỉ test trên bộ này. Thiết kế thí nghiệm cũng vô tình mô phỏng được thực tế, khi model được train và validate trên tập dữ liệu nhỏ, và deploy để infer trên dữ liệu lớn, chưa được model nhìn bao giờ).

Để nhanh chóng thử nghiệm nhiều mô hình khác nhau nhanh hơn, chúng tôi thử các parameter với một mini-Unet, một phiên bản nhẹ hơn của Unet cổ điển với 470,000 tham số. Dùng VM của Google Colab với GPU có RAM 12.75 GB và CPU có RAM 25 GB. Dùng với object load dữ liệu thừa hưởng model.Sequence của Keras, mỗi epoch train mất 1 phút 30 giây.

Nhóm đã thử nghiệm với cách cài đặt. Vấn đề chúng tôi quan sát được là model liên tục đoán với accuracy là 74,73%. Model liên tục đoán và dừng sớm. Đây là kết quả của việc model đoán toàn bộ bức ảnh là class 0 (Nền, không phải phổi). Để khắc phục, chúng tôi đã thử và thay đổi cài đặt như sau:

* Output như trong tờ báo là một lớp mask có 1 channel, với lớp cuối model là một lớp convolution có activation là sigmoid -> chúng tôi đổi thành lớp convolution với activation là softmax, output một mask one-hot với 2 channel.
* Loss của model chúng tôi đã thử cross entropy và dice -> chúng tôi phát hiện cross entropy hội tụ tốt hơn.
* Với Mini-Unet, do không hội tụ được, chúng tôi đã đổi từ Adam với learning rate 0.001 xuống dùng SGD và learning rate 0.00001.
* Nhóm chọn patience lớn hơn, với 50 epoch.
* Nhóm đã thử biến đổi (augment) ngẫu nhiên zoom [0.8, 1.2] (do có thể chụp phổi từ khoảng cách khác nhau), flip ngang (do hai phổi tương đối có hình giống nhau), và xoay ngẫu nhiên [-20,20] radian (do chụp có thể chéo). Tuy vậy, do điều kiện bộ valid và train chụp rất giống nhau (thậm chí cả bộ Shenzhen), augment sẽ làm model kém đi (Thực nghiệm cũng cho thấy vậy). chúng tôi hiện tại không dùng augmentation.
* Loss weight dựa vào tỉ lệ pixel [0.6, 1.9]. chúng tôi đã thử [1,5] nhưng không tốt bằng.

## Mô hình DAE

*Ý tưởng*

Hãy quan sát và so sánh dữ liệu:

* Phổi bình thường:

|  |  |
| --- | --- |
|  |  |

* Phổi bất thường:

|  |  |
| --- | --- |
|  |  |

* Giao dữ liệu:

|  |  |
| --- | --- |
|  |  |

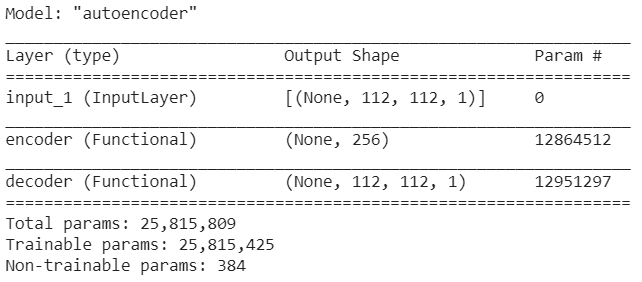
Như vậy một phổi bất thường, nếu như có thể khôi phục lại thì việc detect ra vấn đề trở nên dễ dàng. Góc nhìn khác, ta nhận thấy các điểm dị thường làm cho dữ liệu thay đổi phân bố đáng có của nó. Nếu vậy ta có thể coi nó như là một nhiễu. Nếu vậy mô hình DAE là một mô hình phù hợp cho việc khử nhiễu

*Xây dựng mô hình DAE*

Mô hình DAE được xây dựng dựa trên mô hình của Adrian Rosebrock (2020). Bài viết hướng dẫn xây dựng hàm tạo DAE một cách động. Dựa vào đó thông số của DAE trong bài viết này đã được thay đổi để phù hợp với bài toán. Ngoài ra hàm activation cuối cùng của DAE đã được thay đổi thành ReLu. Điều này được sử dụng để dữ liệu ra có cùng dạng (phân bố) với dữ liệu gốc. Nếu để activation là ‘Sigmoid’ thì kết quả cuối cùng sẽ bị phân tách 2 phía mà không liên tục. Toàn bộ activation của mô hình đều là ReLu.

|  |  |
| --- | --- |
| **Encoder model** | **Decoder model** |

**DAE**

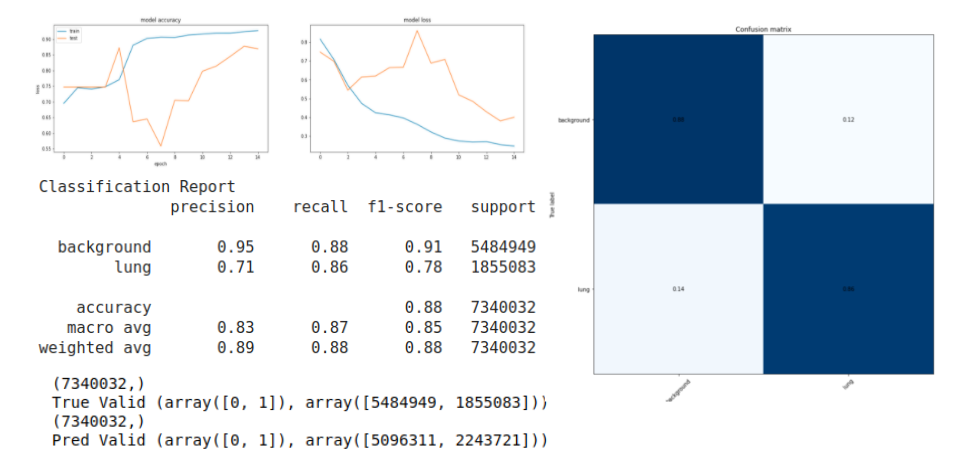


# Chương III

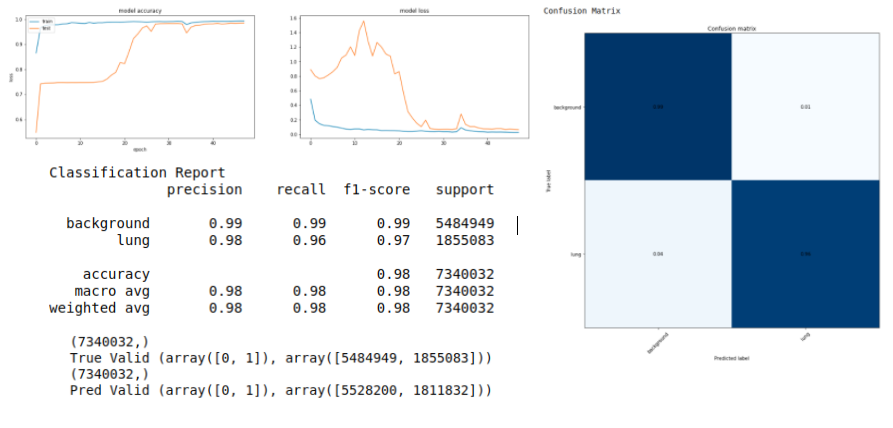
# KẾT QUẢ THỰC NGHIỆM

## Kết quả Unet

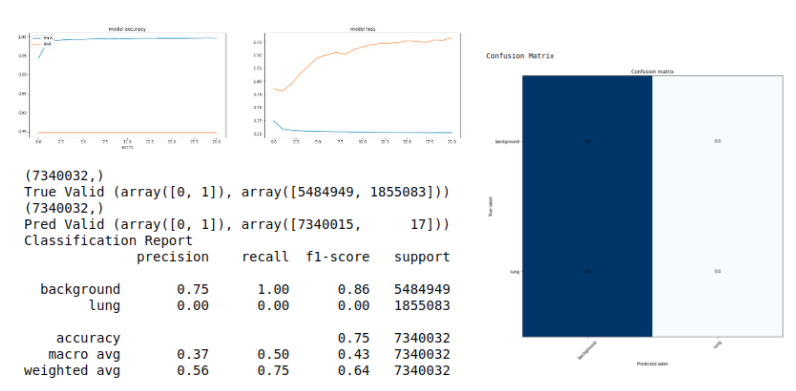
* Unet-mini: Sau khi một thời gian dài loss dao động mà accuracy không đổi, model thoát khỏi trạng thái và giảm rồi tăng dần tới 87%. Sau nhiều lần thí nghiệm, đây là kết quả tốt nhất với kết cầu mini-Unet.



* Dice/F1 score đạt được tốt nhất có thể là 88%, với đến 14% số pixel phổi bị đoán vào nền. Unet: chúng tôi thử Unet với adam với lr: 0.001. Lần này sau 19 epochs bị tắc ở 74% accuracy, model đã lên đến 98% accuracy cho tập valid.

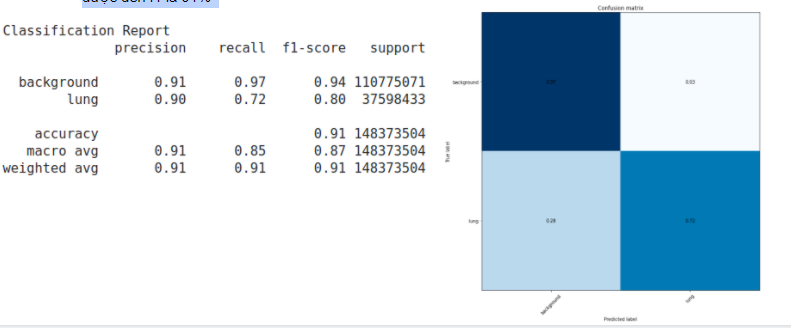


* Resnet50\_Unet: Do resnet50 có pretrained trên imagenet trong thư viện, nên chúng tôi chọn transfer learning để giúp hội tụ nhanh hơn và generalize tốt hơn, thay vì train lại Resnet34\_Unet từ đầu. Tuy vậy, model liên tục bị tắc ở 74% và không lên được sau 20 epochs.

Vì vậy chúng tôi chọn Unet cổ điển để giải quyết bài toán này.

*Đánh giá*

Để đánh giá model, chúng tôi test model trên 566 ảnh từ bộ Shenzhen. Kết quả F1-score đạt được đến f1 là 91%



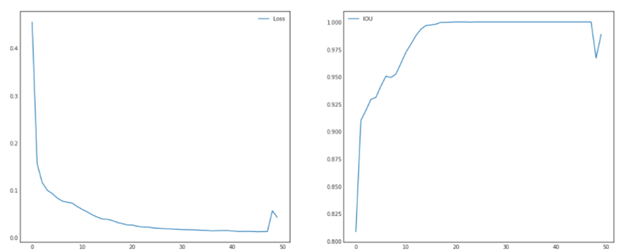
*Bước tiếp theo*

* Bước tiếp theo nên là thử tune các model khác, dùng để train một phần của bộ Shenzhen, để đảm bảo model có thể chạy tốt trên Shenzhen. Có thể sự khác biệt giữa 2 lần do nhiều lí do: máy chụp ảnh xray giữa 2 bệnh viễn khác nhau dẫn tới độ sáng khác nhau giữa 2 bản chụp, tỉ lệ kích cỡ phổi khác nhau, bộ Shenzhen có nhiều phổi có bệnh dị thường, làm bản chụp khác nhau,...
* Cross validate để đảm bảo kiến trúc model là tốt nhất, trên cả 2 bộ.
* Thu thập thêm dữ liệu để huấn luyện model.

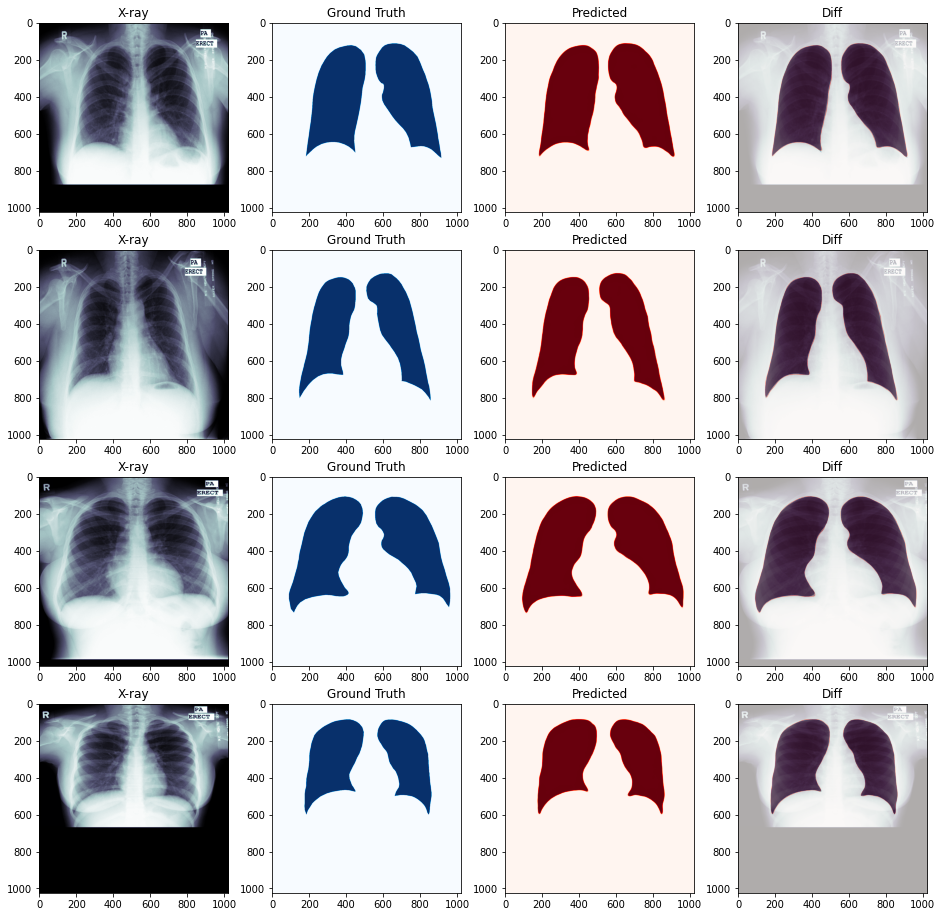
*Đánh giá Unet model trên tập dữ liệu Montgomery County X-ray Set*

Montgomery County X-ray Set: Hình ảnh X-ray trong Dataset này được thu thập từ chương trình kiểm soát bệnh lao của Bộ Y tế và Dịch vụ Nhân sinh của Hạt Montgomery, MD, Hoa Kỳ. Dataset này gồm 138 hình X-ray chụp sau - trước, trong đó 80 hình normal và 58 hình abnormal có biểu hiện của bệnh lao. Tất cả hình ảnh đều được khử nhận dạng và có sẵn ở định dạng DICOM.

Mạng Unet mà chúng tôi sử dụng cho việc phân vùng phổi có kiến trúc gồm 306 lớp (layers) và 25 839 494 tham số (parameters) được huấn luyện trên tập dữ liệu Shenzhen Hospital X-ray Set. Bởi vì một số hạn chế về phần cứng (Google Colab với GPU có RAM 12.75 GB và CPU có RAM 25 GB), trong quá trình huấn luyện dữ liệu, chúng tôi đã mắc phải các lỗi “Out of memory”. Để đảm bảo thời gian thực thi dự án này, chúng tôi quyết định sử dụng mô hình đã huấn luyện được thu thập từ Kaggle [1] để phân vùng dữ liệu x-ray hiện có. Kết quả huấn luyện mô hình thể hiện trong bảng sau:

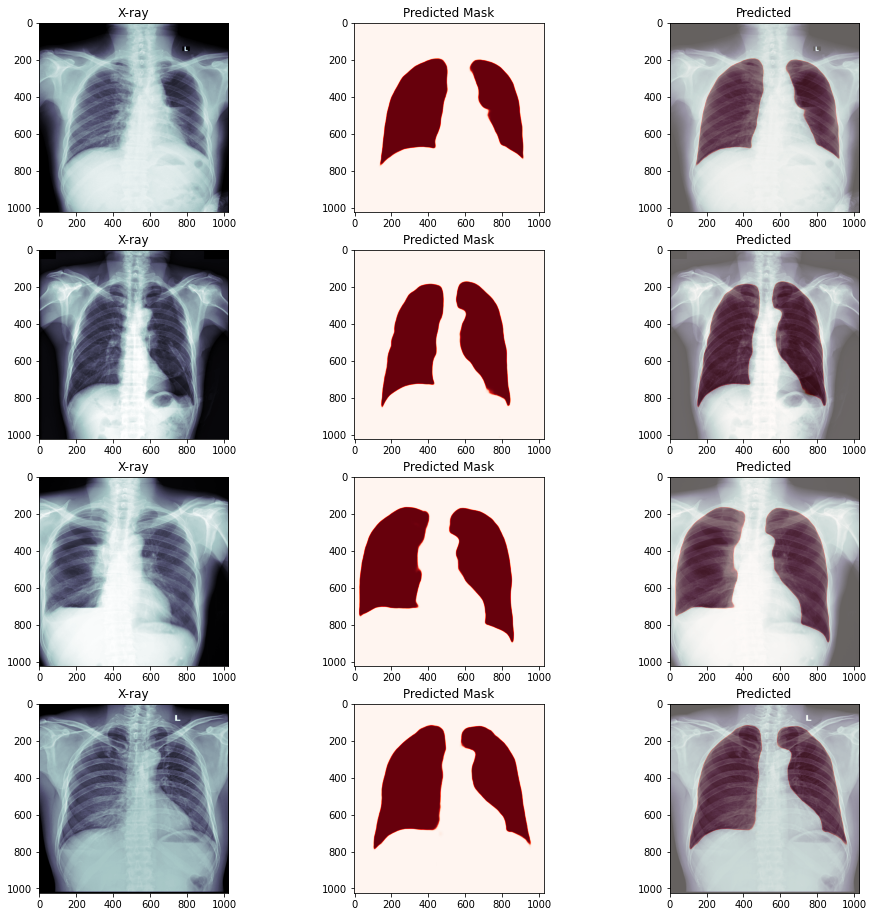


Chúng tôi đã sử dụng Unet model phân vùng các ảnh X-ray trong bộ dữ liệu Montgomery County X-ray Set. Sau đó, chúng tôi đã so sánh mask mà mô hình Unet dự đoán được với mask thực sự của bộ dữ liệu. Kết quả thu được với Intersection over Union Score (IoU) là 0.94 và f1-score là 0.98.



Phân vùng phổi trên bộ dữ liệu Montgomery County X-ray Set

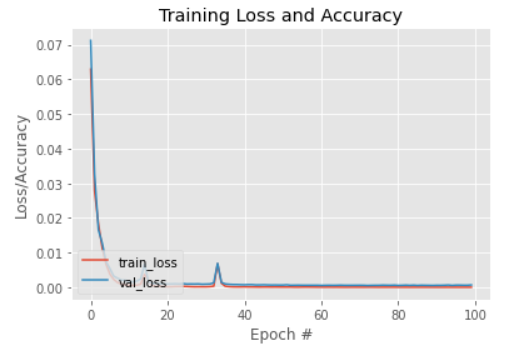
Vì kết quả của mô hình khá tốt khi kiểm thử trên tập dữ liệu Montgomery Country Xray Set với IoU Score khoảng 94%. Nên chúng tôi quyết định sử dụng mô hình Unet này phân vùng phổi trên tập dữ liệu “VinBigData Chest X-ray Abnormalities Detection".

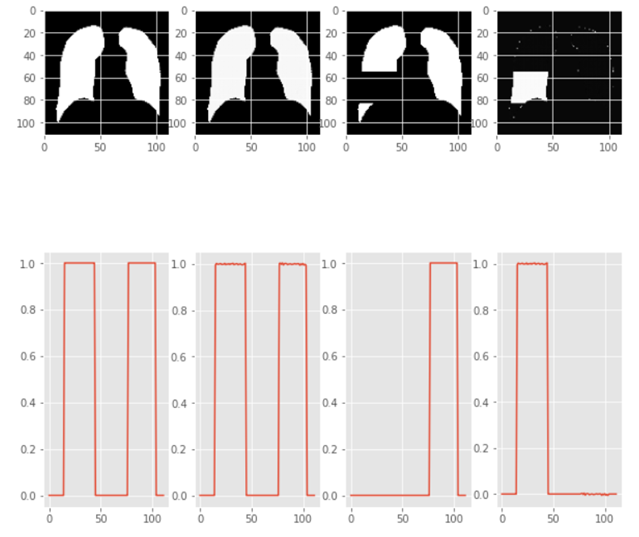


Phân vùng phổi trên bộ dữ liệu VinBigData Chest X-ray Abnormalities Detection

## Kết quả của mô hình DAE

*DAE trong khôi phục mask:*

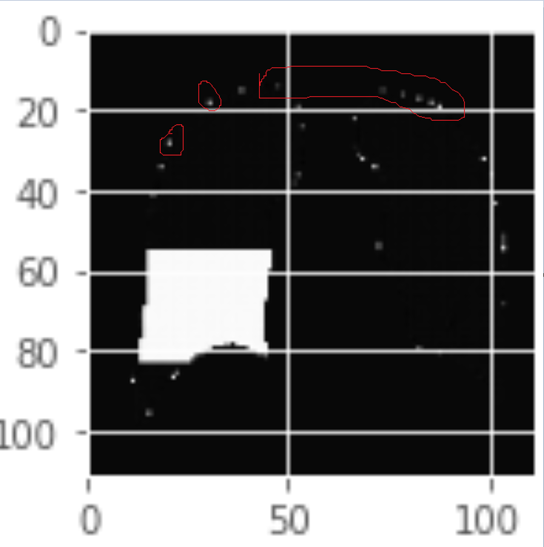




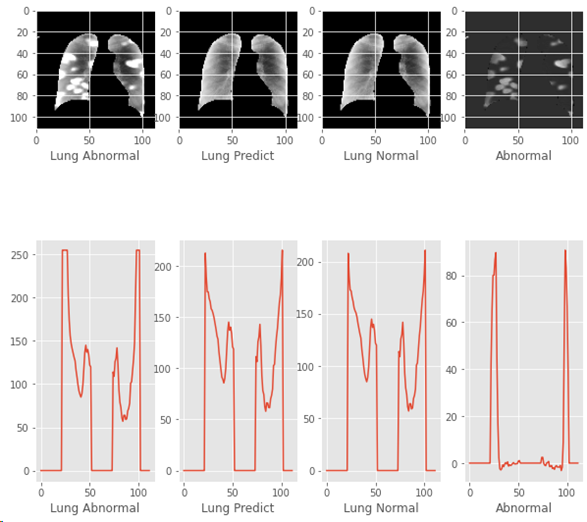
***Chú thích.*** *Từ trái sang phải lần lượt là*

* *Mask gốc - (MaskS) là ảnh mask bình thường, dùng tạo ra mask bất thường và để so sánh với Mask dự đoán.*
* *Mask dự đoán - (MaskP) được tạo ra bằng cách đưa mask bất thường vào mô hình.*
* *Mask Noise - (MaskN) được tạo ra bởi mask gốc. Đây là giả sử cho trường hợp Mask ko toàn vẹn.*
* *Mask loss - (MaskL) Được tìm bằng cách giao MaskP với MaskN. Đây là phần cần bổ sung cho MaskN để nó trở thành MaskS*

Kết quả của mô hình khá tốt. Đối với một ảnh đã khôi thì nó sẽ có một số phần vùng dư ra với kích thước nhỏ. Điều này là do độ chính xác của mô hình không phải 100% gây ra. Tuy nhiên những vùng này hoàn toàn có thể loại bỏ bằng cách so sánh nó với loss của mô hình. Một số điểm nhiễu từ MaskN (vùng khoanh đỏ và các vùng tương tự). Các điểm này hoàn toàn có thể loại bỏ bằng cách so sánh với loss của model.



*DAE trong khôi phục phổi*

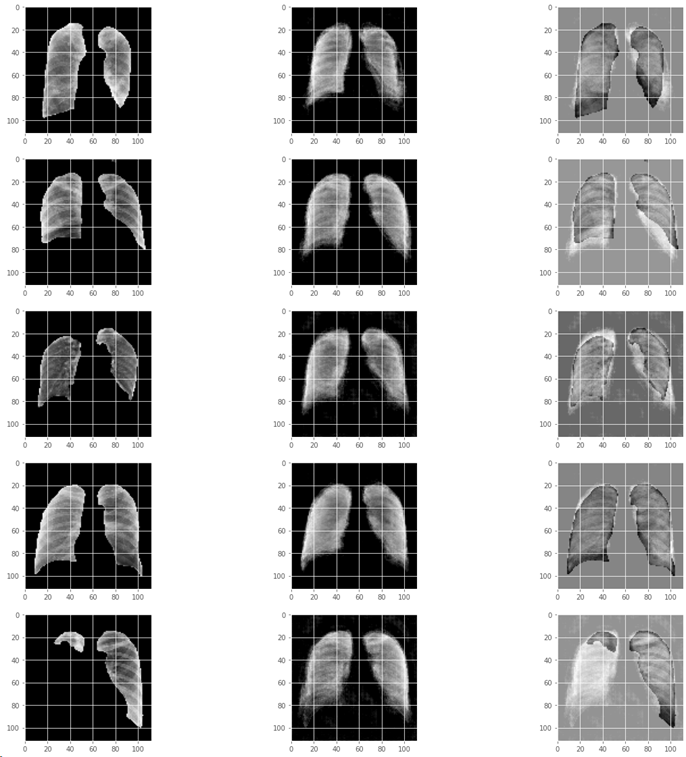


Giống với DAE trong mask, kết quả khả quan. Phần này tương tự mask. 2 model trên được train riêng biệt và sử dụng ứng với từng chức năng của nó:

Vấn đề và kiểm thử: Ở kiểm thử của bài viết, chỉ áp dụng thử nghiệm mô hình DAE – khôi phục phổi.

Kết quả kiểm thử mô hình với dữ liệu bệnh thực tế ta có thể thấy ở hình dưới. Kết quả không tốt một phần do dữ liệu bị giảm xuống còn (112,112) so với dữ liệu gốc mỗi chiều đều lớn hơn 1024. Dữ liệu sau giảm size bị mất mát rất nhiều dẫn đến mô hình không phân biệt được. Ngoài ra còn do kết quả của mô hình chưa được đồng bộ (cùng hệ quy chiếu) với ảnh gốc.

Vấn đề: Như vậy để mô hình được áp dụng tốt ta cần phải thêm layer và tăng kích thước ảnh đầu vào cho mô hình. Ngoài ra vì dữ liệu đầu ra có thể lớn hơn 255 so với ảnh gốc (lớn nhất là 255) nên khi chập hình lại còn tồn tại nhiều nhiễu. Ngoài ra cần xây dựng hàm để có thể boxed những vùng bất thường lại.



# TÀI LIỆU THAM KHẢO

Vingroup Big Data Institute (2021). *VinBigData Chest X-ray Abnormalities Detection* [online]. Available at:

<https://www.kaggle.com/c/vinbigdata-chest-xray-abnormalities-detection/overview>

Ronneberger, O., Fischer, P., Brox, T. (2015). *U-net: Convolutional networks for biomedical image segmentation.* In: Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2015. Volume 9351 of LNCS., Springer (2015) 234–241.

K. He, X. Zhang, S. Ren, and J. Sun (2015). *Deep residual learning for image recognition*. arXiv preprint arXiv:1512.03385, 2015.

Adrian Rosebrock (2020). *Denoising autoencoders with Keras, TensorFlow, and Deep Learning* [online]. Available at:

<https://www.pyimagesearch.com/2020/02/24/denoising-autoencoders-with-keras-tensorflow-and-deep-learning/>

**TRÍCH DẪN**

[1] Appy Patel, *Lung Segmentation UNet* [online]. Available at: https://www.kaggle.com/aaryapatel/lung-segmentation-unet