**TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI**

**Viện Công Nghệ Thông Tin Và Truyền Thông**

**---🙠**🕮**🙢---**



**Bài tập Project II**

Đề tài: Chuẩn đoán bệnh nhân Covid từ hình ảnh chụp CT

***Giáo viên hướng dẫn: Vũ Văn Thiệu***

|  |  |
| --- | --- |
| ***Thành viên nhóm*** | ***MSSV*** |
| *Nguyễn Trung Kiên* | *20194089* |
| *Hoàng Ngọc Lâm* | *20194089* |

*Hà nội, ngày 14 tháng 8 năm 2022.*

Mục lục

[1. Giới thiệu 3](#_Toc111545377)

[2.     Tập dữ liệu 4](#_Toc111545378)

[3.     Đánh giá các mô hình 6](#_Toc111545379)

[3.1. Kết quả các mô hình: 6](#_Toc111545380)

[3.2. Mô hình mạng ResNet 7](#_Toc111545381)

[3.3. Mô hình mạng DenseNet 9](#_Toc111545382)

[4. Cải thiện mô hình 10](#_Toc111545383)

[4.1. Phương pháp 10](#_Toc111545384)

[4.2. Thử nghiệm 12](#_Toc111545385)

[5.     Tổng kết 14](#_Toc111545386)

[Tài liệu tham khảo: 14](#_Toc111545387)

# Giới thiệu

Bệnh do Coronavirus 2019 (COVID-19) là một bệnh truyền nhiễm đã gây ra khoảng 645500 ca tử vong trên toàn thế giới, trong số 595 triệu trường hợp bị nhiễm, tính đến ngày 1 tháng 8 năm 2022. Một trở ngại lớn trong việc kiểm soát sự lây lan của căn bệnh này là sự thiếu hụt các chuẩn đoán nhanh. Các xét nghiệm hiện tại chủ yếu dựa trên phản ứng chuỗi polymerase phiên mã ngược (RT-PCR). Trong thời gian cao điểm bùng phát COVID-19, các bộ xét nghiệm RT-PCR rất thiếu hụt. Do đó, nhiều trường hợp nghi ngờ không được xét nghiệm kịp thời và họ tiếp tục lây bệnh cho người khác một cách vô thức. Để giảm thiểu tình trạng thiếu bộ xét nghiệm RT-PCR, các bệnh viện đã và đang sử dụng các phương pháp chẩn đoán thay thế. Trong số đó, chụp cắt lớp vi tính (CT) đã được sử dụng để sàng lọc và chẩn đoán COVID-19.

Ví dụ, trong Quy trình chẩn đoán và điều trị bệnh viêm phổi do Coronavirus mới (Phiên bản thử nghiệm 5) 1 do Ủy ban Y tế Quốc gia và Cục Quản lý Nhà nước về Y học Cổ truyền Trung Quốc thực hiện, CT và các kỹ thuật chụp ảnh ngực khác được coi là một cách quan trọng để chẩn đoán COVID-19. Theo hướng dẫn này, nhiều bệnh viện ở Trung Quốc đã sử dụng phương pháp chụp cắt lớp vi tính CT để chẩn đoán COVID-19, phương pháp này đã được chứng minh là có hiệu quả. Để hiểu rõ hơn về vai trò của CT trong việc chẩn đoán COVID-19, theo ý kiến ​​của một bác sĩ X quang cao cấp tại Bệnh viện Đồng Tế, Vũ Hán, Trung Quốc, người đã chẩn đoán và điều trị cho bệnh nhân COVID-19 kể từ khi dịch bệnh này bùng phát ở Trung Quốc. Theo bác sĩ X quang này, trong thời gian bùng phát, chụp CT rất hữu ích để chẩn đoán COVID-19; ngoài thời gian bùng phát, chụp CT không hữu ích bằng. Lý do là như sau. Chụp CT có thể được sử dụng để đánh giá liệu bệnh nhân có bị nhiễm bệnh viêm phổi do vi rút hay không (COVID-19 là một loại viêm phổi do vi rút gây ra bởi vi rút SARS-CoV-2). Tuy nhiên, chụp CT không có khả năng xác định loại vi rút nào đang gây ra bệnh viêm phổi do vi rút: SARS-CoV-2 hoặc những loại khác. Nói một cách chính xác, chụp CT không thể được sử dụng để xác nhận liệu một bệnh nhân có bị nhiễm COVID-19 hay không. Tuy nhiên, trong thời gian bùng phát, hầu hết bệnh viêm phổi do vi rút là do SARS-CoV-2. Có nghĩa là, nếu một bệnh nhân được xác nhận là bị viêm phổi do virut theo kết quả chụp CT, thì viêm phổi do virut này rất có thể là COVID-19. Do thực tế này, CT được coi là hữu ích để chẩn đoán COVID-19 trong quá trình bùng phát của nó.

Trong thời gian bùng phát COVID-19, các bác sĩ X quang thường bận rộn, những người này có thể không có băng thông để đọc một số bản chụp CT kịp thời. Bên cạnh đó, đối với các bác sĩ X quang ở các vùng kém phát triển, chẳng hạn như các vùng nông thôn, họ có thể không được đào tạo bài bản để nhận ra COVID19 từ chụp CT, vì bệnh này tương đối mới. Để giải quyết những vấn đề này, một số công trình đã phát triển các phương pháp học sâu để sàng lọc COVID-19 từ hình ảnh chụp CT. Do lo ngại về quyền riêng tư, các bản chụp CT được sử dụng trong các tác phẩm này không được chia sẻ. Điều này cản trở rất nhiều đến việc nghiên cứu và phát triển các phương pháp AI tiên tiến hơn để sàng lọc COVID-19 chính xác hơn dựa trên ảnh chụp CT. Để giải quyết vấn đề này, nhóm đã thu thập tập dữ liệu COVID-CT chứa 349 hình ảnh CT dương tính với COVID-19 thuộc 216 bệnh nhân và 397 hình ảnh CT âm tính với COVID-19. Bộ dữ liệu được cung cấp với nguồn mở, để thúc đẩy thử nghiệm dựa trên CT của COVID-19. Từ 760 bản in trước medRxiv và bioRxiv về COVID-19, trích xuất các hình ảnh CT được báo cáo và chọn thủ công những hình ảnh có chứa các phát hiện lâm sàng về COVID-19 bằng cách đọc chú thích của những hình ảnh này.

# 2.     Tập dữ liệu

Sau khi thu thập tập dữ liệu này, nhóm nhận ra sự hạn chế về khả năng sử dụng của tập dữ liệu này. Đầu tiên, khi các hình ảnh CT gốc được đưa vào giấy, chất lượng của những hình ảnh này bị giảm sút, có thể làm cho các quyết định chẩn đoán kém chính xác hơn. Sự suy giảm chất lượng bao gồm: các giá trị đơn vị Hounsfield (HU) bị mất; số lượng bit trên mỗi pixel bị giảm xuống; độ phân giải của hình ảnh bị giảm. Thứ hai, bản chụp CT ban đầu chứa một chuỗi các lát cắt CT, nhưng khi đưa vào giấy tờ, chỉ một số lát cắt chính được chọn, điều này cũng có thể có tác động tiêu cực đến chẩn đoán. Theo bác sĩ X quang, những vấn đề được nêu ra trong những mối quan tâm này không ảnh hưởng đáng kể đến độ chính xác của việc ra quyết định chẩn đoán. Đầu tiên, bác sĩ X quang có kinh nghiệm có thể chẩn đoán chính xác từ hình ảnh CT chất lượng thấp. Ví dụ, với một bức ảnh chụp CT gốc bằng điện thoại thông minh, các bác sĩ X quang có kinh nghiệm có thể chẩn đoán chính xác chỉ bằng cách nhìn vào ảnh, mặc dù hình ảnh CT trong ảnh có chất lượng thấp hơn nhiều so với CT gốc. Tương tự như vậy, khoảng cách chất lượng giữa hình ảnh CT trong giấy tờ và CT gốc sẽ không ảnh hưởng nhiều đến độ chính xác của chẩn đoán. Thứ hai, mặc dù tốt hơn là đọc một chuỗi các lát cắt CT, nhưng đôi khi một lát CT đơn lẻ chứa đủ thông tin lâm sàng để đưa ra quyết định chính xác. Cuối cùng, nhóm tận dụng bộ dữ liệu COVID-CT thu thập được, hình ảnh CT gốc được thu thập ở nơi khác, mặt nạ tổn thương do bác sĩ X quang dán nhãn để huấn luyện mô hình chẩn đoán COVID-19, dựa trên học tập đa nhiệm vụ và học tập tự giám sát. Mô hình tốt nhất mà nhóm đạt được là DensetNet-169 với F1 là 0,80; AUC là 0,87 và độ chính xác là 0,78.

Trong phần này, nhóm mô tả tập dữ liệu COVID-CT được thu thập. Đầu tiên, bộ dữ liệu này thu thập 760 bản in trước về COVID-19 từ medRxiv2 và bioRxiv3, được đăng từ ngày 19 tháng 1 đến ngày 25 tháng 3 năm 2020. Hình ảnh CT được kết hợp với chú thích mô tả các phát hiện lâm sàng trong CT. Bộ dữ liệu đã sử dụng PyMuPDF4 để trích xuất thông tin cấu trúc cấp thấp của các tệp PDF của các bản in trước và định vị tất cả các hình được nhúng. Chất lượng (bao gồm độ phân giải, kích thước, v.v.) của các số liệu được bảo toàn tốt. Từ thông tin cấu trúc, nhóm cũng xác định các chú thích kết hợp với các số liệu. Với những số liệu và chú thích được trích xuất này, trước tiên chọn tất cả các hình ảnh CT theo cách thủ công. Sau đó, đối với mỗi hình ảnh CT, chú thích liên quan để đánh giá xem nó có dương tính với COVID-19 hay không. Nếu không thể đánh giá từ chú thích, đặt văn bản phân tích hình này trong bản in trước để đưa ra quyết định. Đối với mỗi hình ảnh CT, cũng thu thập thông tin meta được trích xuất từ ​​bài báo, chẳng hạn như tuổi bệnh nhân, giới tính, vị trí, bệnh sử, thời gian quét, mức độ nghiêm trọng của COVID-19 và báo cáo X quang.

Cuối cùng, bộ dữ liệu thu được 349 hình ảnh CT được dán nhãn là dương tính với COVID-19. Những hình ảnh CT này có kích thước khác nhau. Chiều cao tối thiểu, trung bình và tối đa là 153, 491 và 1853. Chiều rộng tối thiểu, trung bình và tối đa là 124, 383 và 1485. Những hình ảnh này là từ 216 trường hợp bệnh nhân. Hình 1 (bên phải) cho thấy một số ví dụ về hình ảnh CT COVID-19. Đối với những bệnh nhân được dán nhãn dương tính, 169 người trong số họ có thông tin về tuổi và 137 người trong số họ có thông tin về giới tính. Hình 2 (bên trái) cho thấy sự phân bố theo tuổi của bệnh nhân mắc COVID-19. Hình 2 (bên phải) cho thấy tỷ lệ giới tính của bệnh nhân mắc COVID-19. Bệnh nhân nam nhiều hơn bệnh nhân nữ, với số lượng lần lượt là 86 và 51. Bảng 1 so sánh tập dữ liệu COVID-CT với các tập dữ liệu khác. Bộ dữ liệu của chúng tôi có nhiều hình ảnh và bệnh nhân dương tính với COVID-19 hơn bộ sưu tập dữ liệu hình ảnh COVID-19 và cơ sở dữ liệu SIRM COVID-19, COVID-CS không có nguồn mở. Bộ dữ liệu phân đoạn CT COVID-19 có nhiều hình ảnh COVID-19 dương tính hơn, nhưng ít bệnh nhân hơn. Hình ảnh CT của cùng một bệnh nhân rất giống nhau về mặt thị giác. Do đó, sự đa dạng của hình ảnh trong tập dữ liệu lớn hơn so với trong tập dữ liệu phân đoạn CT COVID-19. Thu thập các hình ảnh CT âm tính COVID-19 làm bộ huấn luyện âm tính. Để phát triển các mô hình phân loại

nhị phân để chẩn đoán COVID-19, ngoài 349 ảnh CT COVID-19, bộ dữ liệu có bộ ảnh CT không COVID-19 làm âm tính các ví dụ huấn luyện. Nguồn của những hình ảnh này bao gồm:

• Cơ sở dữ liệu MedPix6, là cơ sở dữ liệu trực tuyến truy cập mở về các hình ảnh y tế, ca giảng dạy và các chủ đề lâm sàng. Nó chứa hơn 9000 chủ đề, 59000 hình ảnh từ 12000 trường hợp bệnh nhân.

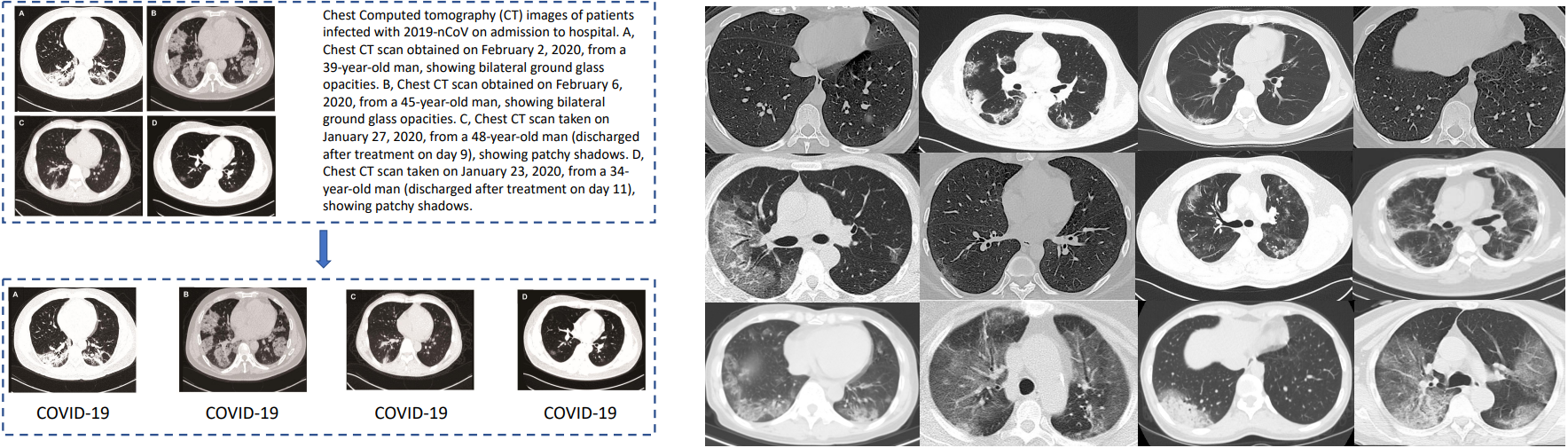
• Bộ dữ liệu LUNA7, chứa 888 bản chụp CT ung thư phổi từ 888 bệnh nhân.

• Trang web Radiopaedia8, chứa các hình ảnh X quang từ 36559 trường hợp bệnh nhân.

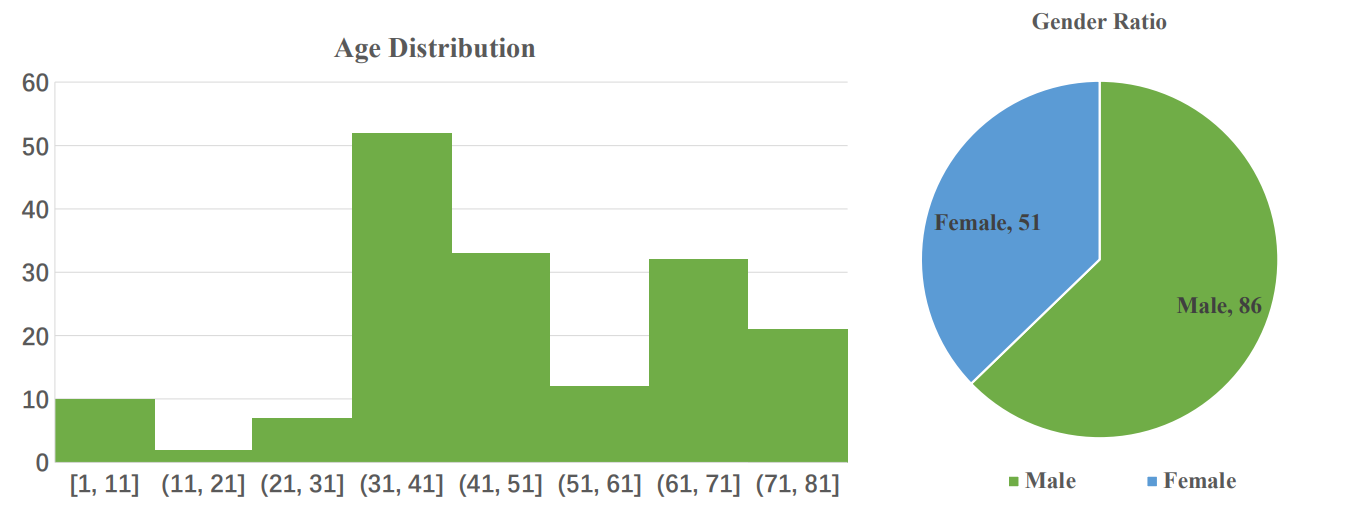
• PubMed Central (PMC) 9, là một kho lưu trữ toàn văn miễn phí các tài liệu tạp chí khoa học đời sống và y sinh. Một số giấy tờ có chứa hình ảnh CT.

*Bảng 1: So sánh COVID-CT với các bộ dữ liệu khác.*

|  |  |  |
| --- | --- | --- |
| Tập dữ liệu | Hình ảnh COVID | Bệnh nhân |
| COVID - CT | 349 | 216 |
| COVID-19 Image Data Collection (Cohen et al., 2020a) | 84 | 45 |
| SIRM COVID-19 Database | 100 | 60 |
| COVID-CS (Wu et al., 2020) | 68626 | 400 |
| COVID-19 CT Segmentation Dataset (Jun et al.) | 379 | 20 |



*Hình 1: (Bên trái) Đối với bất kỳ hình nào có chứa nhiều hình ảnh CT dưới dạng hình con, chúng tôi chia nó thành các CT riêng lẻ theo cách thủ công. (Bên phải) Ví dụ về hình ảnh CT dương tính với COVID-19.*



*Hình 2: (Trái) Phân bố tuổi của bệnh nhân COVID-19. (Phải) Tỷ lệ giới tính của bệnh nhân COVID-19.*

# 3.     Đánh giá các mô hình

## 3.1. Kết quả các mô hình:

Nhóm đã tìm hiểu các mô hình như DenseNet, ResNet, SimpleCNN, effecientNet, …

Train và test các mô hình với bộ dữ liệu COVID-19 CT Scan từ repository có phân phối như sau:

* Max CT scans per patient in COVID: 16.0 (patient 2)
* Average CT scans per patient in COVID: 1.6
* Min CT scans per patient in COVID: 1.0

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Type | NonCOVID-19 | COVID-19 | Total |
| Images distribution | Train | 234 | 191 | 425 |
| Val | 58 | 60 | 118 |
| Test | 105 | 98 | 203 |
| Patients distribution | Train | 105 | 130 | 235 |
| Val | 24 | 32 | 56 |
| Test | 42 | 54 | 96 |

**Train result (sau 30 epoch):**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mô hình | avg recall | avg precision | avg F1 score | avg accuracy | avg AUC |
| Dense121 | 0.7931 | 0.7931 | 0.7931 | 0.7966 | 0.8922 |
| Dense169 | 0.7586 | 0.9167 | 0.8302 | 0.8475 | 0.9149 |
| ResNet18 | 0.7586 | 0.7458 | 0.7521 | 0.7542 | 0.8739 |
| ResNet50 | 0.7586 | 0.8302 | 0.7928 | 0.8051 | 0.9040 |
| SimpleCNN | 0.7069 | 0.6508 | 0.6777 | 0.6695 | 0.7276 |
| effecientNet-b0 | 0.7241 | 0.8077 | 0.7636 | 0.7797 | 0.8868 |
| effecientNet-random | 0.6897 | 0.6154 | 0.6504 | 0.6356 | 0.6980 |

**Test result:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mô hình | recall | precision | F1 score | accuracy | AUC |
| Dense121 | 0.7429 | 0.7429 | 0.7429 | 0.7340 | 0.8152 |
| Dense169 | 0.8571 | 0.7500 | 0.8000 | 0.7783 | 0.8702 |
| ResNet18 | 0.8190 | 0.7227 | 0.7679 | 0.7483 | 0.8466 |
| ResNet50 | 0.8952 | 0.6861 | 0.7769 | 0.7340 | 0.8657 |
| SimpleCNN | 0.6571 | 0.6273 | 0.6419 | 0.6207 | 0.6561 |
| effecientNet-b0 | 0.7810 | 0.7961 | 0.7885 | 0.7833 | 0.8421 |
| effecientNet-random | 0.7333 | 0.6814 | 0.7064 | 0.6847 | 0.7493 |

Từ kết quả trên cho thấy mô hình ResNet50 và DenseNet169 cho kết quả tốt nhất. Dưới đây là trình bày về 2 mô hình đã cho kết quả tốt nhất này.

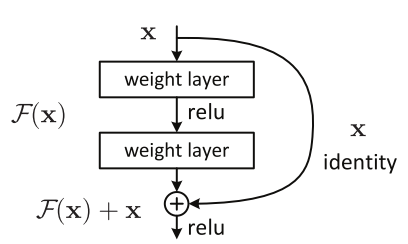
## 3.2. Mô hình mạng ResNet

ResNet (Residual Network) được giới thiệu đến công chúng vào năm 2015 và thậm chí đã giành được vị trí thứ 1 trong cuộc thi ILSVRC 2015 với tỉ lệ lỗi top 5 chỉ 3.57%. Không những thế nó còn đứng vị trí đầu tiên trong cuộc thi ILSVRC and COCO 2015 với ImageNet Detection, ImageNet localization, Coco detection và Coco segmentation.Hiện tại thì có rất nhiều biến thể của kiến trúc ResNet với số lớp khác nhau như ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-152,...Với tên là ResNet theo sau là một số chỉ kiến trúc ResNet với số lớp nhất định.

Mạng ResNet (R) là một mạng CNN được thiết kế để làm việc với hàng trăm hoặc hàng nghìn lớp chập. Một vấn đề xảy ra khi xây dựng mạng CNN với nhiều lớp chập sẽ xảy ra hiện tượng *Vanishing Gradient* dẫn tới quá trình học tập không tốt. Mạng ResNet ra đời cũng giải quyết vấn đề đó.

***Kiến trúc mạng ResNet***

Giải pháp mà ResNet đưa ra là sử dụng kết nối "tắt" đồng nhất để xuyên qua một hay nhiều lớp. Một khối như vậy được gọi là một Residual Block, như trong hình sau :



ResNet gần như tương tự với các mạng gồm có convolution, pooling, activation và fully-connected layer. Ảnh bên trên hiển thị khối dư được sử dụng trong mạng. Xuất hiện một mũi tên cong xuất phát từ đầu và kết thúc tại cuối khối dư. Hay nói cách khác là sẽ bổ sung Input X vào đầu ra của layer, hay chính là phép cộng mà ta thấy trong hình minh họa, việc này sẽ chống lại việc đạo hàm bằng 0, do vẫn còn cộng thêm X. Với H(x) là giá trị dự đoán, F(x) là giá trị thật (nhãn), chúng ta muốn H(x) bằng hoặc xấp xỉ F(x). Việc F(x) có được từ x như sau:

**X->weight1-> ReLU -> weight2**

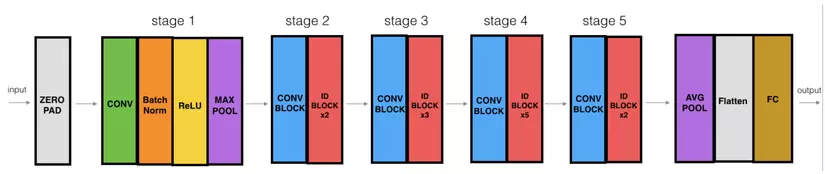
Giá trị H(x) có được bằng cách:

**F(x) + x -> ReLU**

Như chúng ta đã biết việc tăng số lượng các lớp trong mạng làm giảm độ chính xác, nhưng muốn có một kiến trúc mạng sâu hơn có thể hoạt động tốt.

***Xây dựng mạng ResNet-50***

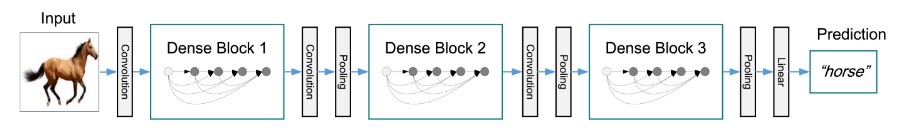
Hình dưới đây mô tả chi tiết kiến trúc mạng noron ResNet :

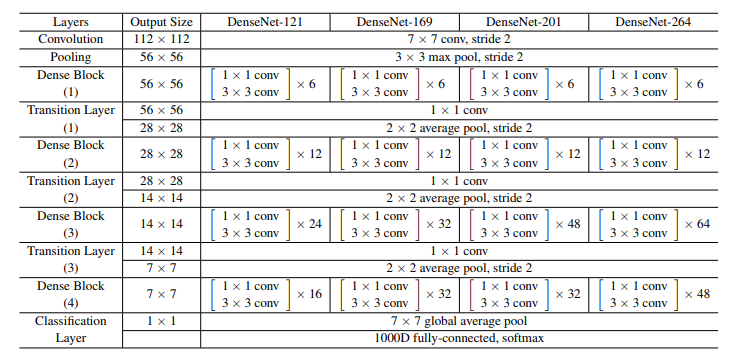


"ID BLOCK" trong hình trên là viết tắt của từ Identity block và ID BLOCK x3 nghĩa là có 3 khối Identity block chồng lên nhau. Nội dung hình trên như sau :

* Zero-padding : Input với (3,3)
* Stage 1 : Tích chập (Conv1) với 64 filters với shape(7,7), sử dụng stride (2,2). BatchNorm, MaxPooling (3,3).
* Stage 2 : Convolutiontal block sử dụng 3 filter với size 64x64x256, f=3, s=1. Có 2 Identity blocks với filter size 64x64x256, f=3.
* Stage 3 : Convolutional sử dụng 3 filter size 128x128x512, f=3,s=2. Có 3 Identity blocks với filter size 128x128x512, f=3.
* Stage 4 : Convolutional sử dụng 3 filter size 256x256x1024, f=3,s=2. Có 5 Identity blocks với filter size 256x256x1024, f=3.
* Stage 5 :Convolutional sử dụng 3 filter size 512x512x2048, f=3,s=2. Có 2 Identity blocks với filter size 512x512x2048, f=3.
* The 2D Average Pooling : sử dụng với kích thước (2,2).
* The Flatten.
* Fully Connected (Dense) : sử dụng softmax activation.

## 3.3. Mô hình mạng DenseNet

* DenseNet (Dense connected convolutional network) là một trong những network mới nhất cho visual object recognition. Nó cũng gần giống Resnet nhưng có một vài điểm khác biệt. Densenet có cấu trúc gồm các dense block và các transition layers. Được stack dense block- transition layers-dense block- transition layers như hình vẽ. Với CNN truyền thống nếu chúng ta có L layer thì sẽ có L connection, còn trong densenet sẽ có L(L+1)/2 connection.   
   *Densenet(2016).*
* Hãy tưởng tượng ban đầu ta có 1 image size (28,28,3). Đầu tiên ta khởi tạo feature layer bằng Conv tạo ra 1 layer size (28,28,24). Sau mỗi layer tiếp theo (Trong dense block ) nó sẽ tạo thêm K=12 feature giữa nguyên width và height. Khi đó output tiếp theo sẽ là (28,28,24 +12),(28,28,24 +12+12). Ở mỗi dense block sẽ có normalization, nonlinearity và dropout. Để giảm size và depth của feature thì transition layer được đặt giữa các dense block, nó gồm Conv kernel size =1, average pooling (2x2) với stride = 2 nó sẽ giảm output thành (14,14,48)

* Kiến trúc chi tiết:
* Một số ưu điểm của Densenet:
* Accuracy : Densenet training tham số ít hơn 1 nửa so với Resnet nhưng có same accuracy so trên ImageNet classification dataset.
* Overfitting : DenseNet resistance overfitting rất hiệu quả.
* Giảm được vanishing gradient.
* Sử dụng lại feature hiệu quả hơn

# 4. Cải thiện mô hình

COVID-CT chiết xuất từ ​​giấy rất hữu ích cho việc huấn luyện các mô hình chẩn đoán cho COVID-19 và mang lại hiệu suất tốt hơn nhiều so với chỉ sử dụng COVID-19 CT gốc. Nhưng độ chính xác còn thấp. Trong phần này, nhóm đã sử dụng các phương pháp cải thiện độ chính xác của chẩn đoán lên mức hữu ích hơn về mặt lâm sàng.

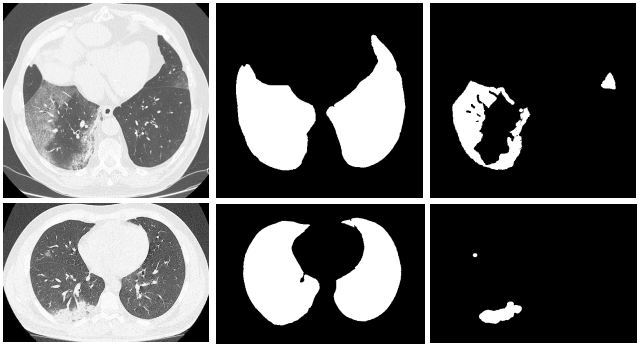
## 4.1. Phương pháp

Tổng số hình ảnh huấn luyện dương tính trong COVID-CT và COVID-Seg là 467. Huấn luyện mô hình học sâu trên một số lượng nhỏ hình ảnh như vậy sẽ dễ bị overfitting. Để giải quyết vấn đề này, nhóm đã tìm hiểu được hai cách: một là kết hợp thông tin bổ sung bao gồm mặt nạ phân đoạn vùng phổi và mặt nạ phân đoạn vùng tổn thương; cách khác là học cách biểu diễn hình ảnh tốt hơn.

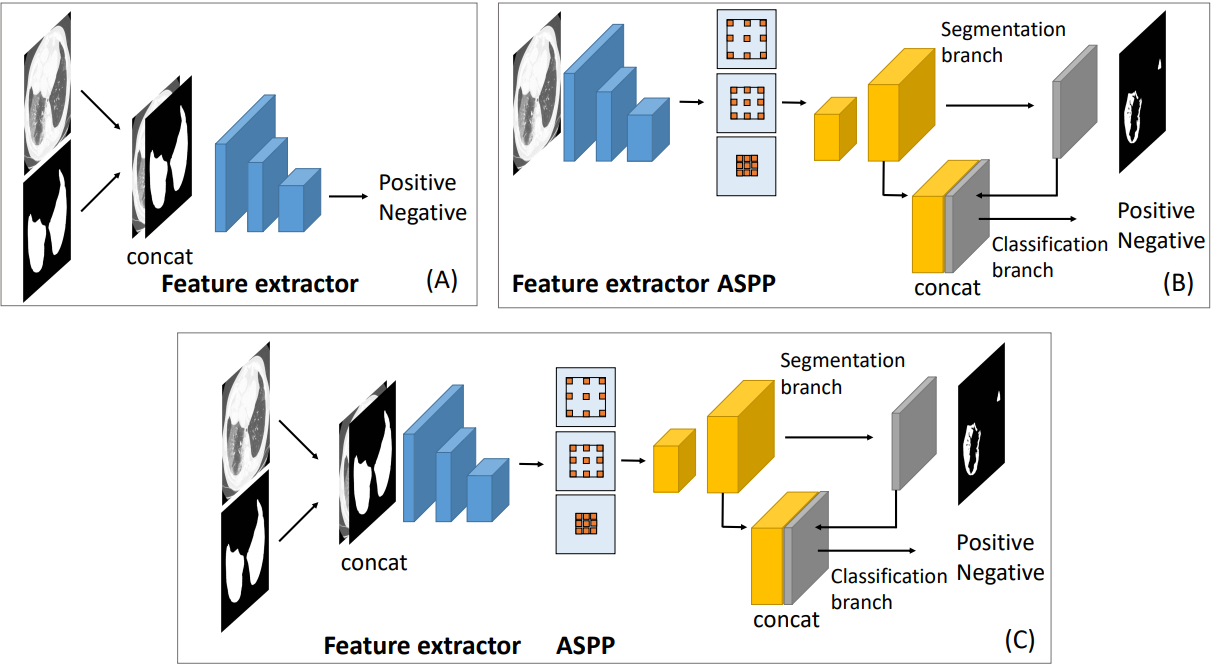
Mặt nạ phổi thông báo cho các mô hình chú ý hơn đến các vùng phổi có biểu hiện lâm sàng của COVID-19, thay vì chú ý đến các vùng nền không liên quan đến COVID-19. Trong khi mặt nạ phổi thu hẹp không gian tìm kiếm của COVID-19, chúng vẫn ở dạng hạt thô. Trong các vùng phổi, các vùng chứa COVID-19 chính xác chiếm một tỷ lệ nhỏ.

Mặt nạ tổn thương có thể xác định các khu vực như vậy và cung cấp hướng dẫn chi tiết cho mô hình liên quan đến khu vực cụ thể nào cần chú ý. Mặc dù mặt nạ thương tổn chính xác hơn, nhưng chúng khó lấy hơn. Chỉ các bác sĩ X quang có kinh nghiệm mới có thể cung cấp các nhãn như vậy trong khi mặt nạ phổi có thể được dán nhãn bởi những người không phải y tế.

Hình ảnh huấn luyện dương tính và âm tính trong COVID-CT và COVIDSeg đều được dán nhãn mặt nạ phổi. Chỉ những hình ảnh huấn luyện dương tính trong COVID-Seg mới được dán nhãn mặt nạ tổn thương. Hình 3 cho thấy một số ví dụ.



*Hình 3: Hình ảnh CT từ tập dữ liệu COVID-Seg (cột đầu tiên), mặt nạ phổi (cột thứ hai) và mặt nạ tổn thương (cột thứ ba).*



*Hình 4: Kiến trúc cho các mô hình được đề xuất của chúng tôi. (A) Kiến trúc để kết hợp mặt nạ phổi làm đầu vào. (B) Kiến trúc để kết hợp mặt nạ tổn thương làm đầu ra. (C) Kiến trúc để kết hợp cả mặt nạ phổi và mặt nạ tổn thương.*

Dựa trên những thực tế này, nhóm đã áp dụng các chiến lược khác nhau để sử dụng mặt nạ phổi và mặt nạ tổn thương. Mặt nạ phổi được kết hợp ở đầu vào của mô hình chẩn đoán. Sử dụng mặt nạ phổi được gắn nhãn trong COVID-CT và COVID-Seg, phát triển mô hình phân đoạn phổi dựa trên U-Net còn lại.. Với một hình ảnh CT (trong quá trình huấn luyện hoặc kiểm tra), sử dụng mô hình phân đoạn để phát hiện mặt nạ phổi, sau đó ghép nó với hình ảnh CT và đưa chúng vào mô hình chẩn đoán (như trong Hình 4 (A)). Các chi tiết của các thí nghiệm phân đoạn phổi được chuyển sang các chất bổ sung. Mặt nạ tổn thương được kết hợp ở đầu ra của mô hình chẩn đoán trong quá trình huấn luyện (như trong Hình 4 (B)). Họ giám sát mô hình để chú ý hơn đến các vùng có tổn thương. Hình 4 (C) cho thấy kiến ​​trúc để kết hợp đồng thời mặt nạ phổi và mặt nạ tổn thương.

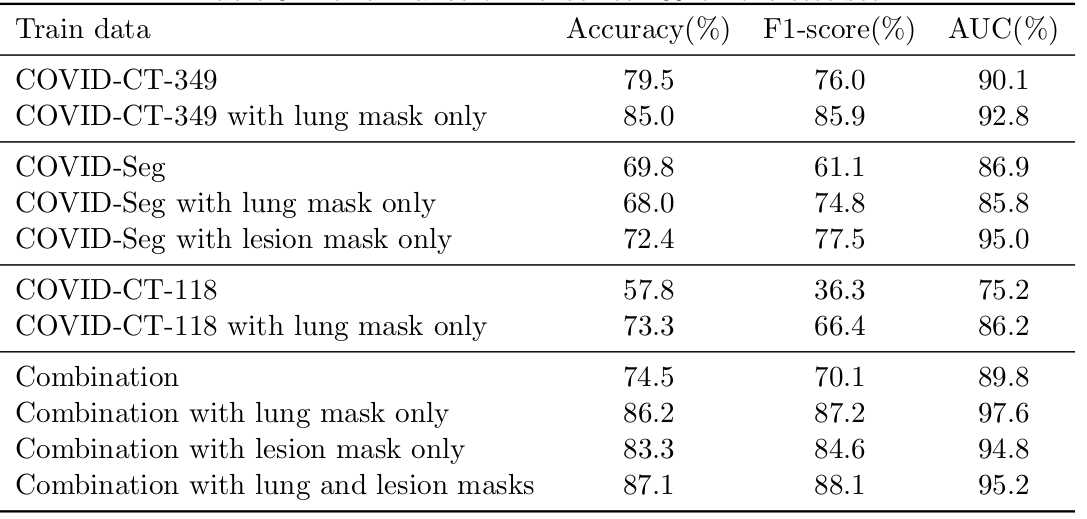
Trong quá trình huấn luyện, với một hình ảnh CT, mạng sẽ dự đoán nhãn lớp (cho dù hình ảnh này là dương tính với COVID-19 hay âm tính) và mặt nạ tổn thương, bằng cách giảm thiểu tổng của tổn thất phân loại và tổn thất phân đoạn tổn thương. Với một hình ảnh CT, trước tiên chúng tôi sử dụng mô hình phân đoạn phổi để thu được mặt nạ phổi. Sau đó, hình ảnh CT gốc cùng với hình ảnh mặt nạ phổi được ghép nối và đưa vào mạng trích xuất tính năng. Sau đó, các đối tượng địa lý trực quan được đưa vào một lớp tổng hợp kim tự tháp không gian khắc nghiệt (ASPP) (Chen và cộng sự, 2017) để trích xuất các bản đồ đối tượng địa lý dày đặc hơn và có độ phân giải cao hơn. Các bản đồ đặc điểm này được đưa vào hai nhánh: một nhánh phân đoạn tổn thương để dự đoán mặt nạ phân đoạn tổn thương và một nhánh phân loại cho nhãn lớp dự đoán. Mặt nạ phân đoạn tổn thương dự đoán được đưa vào nhánh phân loại như thông tin bổ sung. Một chức năng mất phân đoạn tổn thương được xác định giữa mặt nạ xác thực và mặt nạ dự đoán. Mất phân loại được xác định giữa nhãn xác thực và nhãn được dự đoán. Tổng của hai tổn thất này được giảm thiểu để học các tham số trọng số trong mạng. Mặt nạ thương tổn chỉ được yêu cầu trong quá trình huấn luyện. Trong quá trình thử nghiệm, dự đoán được đưa ra chỉ dựa trên hình ảnh thử nghiệm và mặt nạ phân đoạn phổi của chúng mà không cần mặt nạ tổn thương. Cách tiếp cận khác để cải thiện hiệu suất là học cách trình bày trực quan tốt hơn. Các trọng số trong DenseNet-169 và ResNet-50 được huấn luyện trước trên ImageNet bằng cách sử dụng học chuyển giao. Các nhãn lớp trong ImageNet được tận dụng trong quá trình huấn luyện trước. Vì các lớp ImageNet chủ yếu là phi y tế, các mô hình được ImageNet huấn luyện trước có thể thiên vị đối với các lớp miền chung này và khái quát hóa kém hơn trên phân loại COVID-19. Để giảm thiểu sự thiên vị này, chúng tôi sử dụng phương pháp học tự giám sát tương phản (CSSL) (He et al., 2019) để tinh chỉnh các mô hình ImageNet được huấn luyện trước trên hình ảnh CT huấn luyện trong nhiệm vụ chẩn đoán COVID-19. CSSL là một cách tiếp cận không có giám sát. Nó học các biểu diễn bằng cách giải quyết một nhiệm vụ phụ trợ. CSSL tạo ra các ví dụ bổ sung về CT, sau đó tìm hiểu mạng biểu diễn trực quan bằng cách giải quyết nhiệm vụ dự đoán rằng liệu hai bổ sung có bắt nguồn từ cùng một CT hay không. Trong quá trình huấn luyện CSSL, chỉ các hình ảnh CT đầu vào được sử dụng và không có nhãn nào được sử dụng. Sau khi huấn luyện CSSL, mạng được tinh chỉnh thêm bằng cách sử dụng cả CT đầu vào và nhãn đầu ra.

## 4.2. Thử nghiệm

Cài đặt thử nghiệm: Trong khuôn khổ học tập đa tác vụ, trọng số liên quan đến giảm phân đoạn và giảm phân loại đều được đặt thành 1. Phần còn lại của cài đặt siêu tham số giống như trong phần 3. Trong quá trình học tự giám sát, nhóm cài đặt siêu tham số tương tự như trong MoCo (He et al., 2019). Mỗi hình ảnh đầu vào được tăng cường với tính năng lật ngang ngẫu nhiên, cắt xén ngẫu nhiên với kích thước 0,2 trong diện tích, sự pha trộn màu ngẫu nhiên như độ sáng ngẫu nhiên với tỷ lệ 0,4, độ tương phản ngẫu nhiên 0,4, độ bão hòa ngẫu nhiên 0,4, màu sắc ngẫu nhiên 0,1, Gaussian và chuyển đổi thang xám ngẫu nhiên. Kích thước của bộ từ điển động được đặt là 512. Stochastic gradient descent (SGD) được sử dụng làm trình tối ưu hóa với kích thước minibatch là 128, trọng số giảm xuống 0,0001, momentum là 0,9 và tốc độ học ban đầu là 0,015.

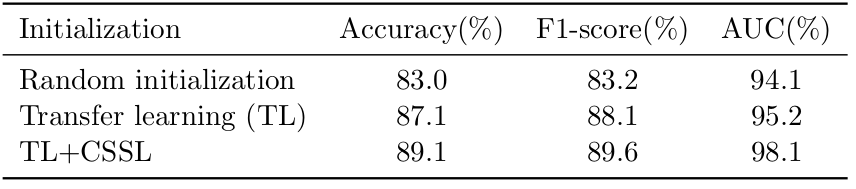
Kết quả Bảng 2 cho thấy hiệu suất của DenseNet-169 trên dữ liệu thử nghiệm. “Kết hợp” biểu thị tập dữ liệu huấn luyện kết hợp của COVID-CT-349 và COVID-Seg. DenseNet-169 được huấn luyện trước trên ImageNet thông qua học chuyển giao mà không cần tinh chỉnh bằng CSSL. Từ bảng này, nhóm đưa ra các nhận xét sau. Đầu tiên, kết hợp mặt nạ phổi có thể cải thiện hiệu suất rất nhiều. Điều này được thể hiện rõ khi so sánh (1) “COVID-CT-349 chỉ với mặt nạ phổi” và “COVID-CT-349”, (2) “COVID-CT-118 chỉ với mặt nạ phổi” và “COVID-CT- 118 ”, và (3)“ Chỉ kết hợp với mặt nạ phổi ”và“ Kết hợp cả hai ”. Trên COVID-Seg, việc kết hợp mặt nạ phổi cải thiện đáng kể F1, mặc dù độ chính xác và AUC bị giảm một chút. Thứ hai, kết hợp mặt nạ trị thương có thể cải thiện đáng kể hiệu suất. Điều này được thể hiện rõ ràng khi so sánh (1) “COVID-Seg chỉ với mặt nạ tổn thương” và “COVID-Seg” và (2) “Chỉ kết hợp với mặt nạ tổn thương” và “Kết hợp cả hai”. Thứ ba, trên tập dữ liệu kết hợp, sử dụng cả mặt nạ phổi và mặt nạ tổn thương đạt được hiệu quả tốt hơn so với chỉ sử dụng mặt nạ phổi hoặc chỉ sử dụng mặt nạ tổn thương. Nhìn chung, những kết quả này chứng minh tính hữu ích của mặt nạ phổi và mặt nạ tổn thương trong việc cải thiện hiệu suất chẩn đoán.

*Bảng 2: Hiệu suất của DenseNet-169 trên bộ kiểm tra.*



Bảng 3 cho thấy hiệu suất thử nghiệm của DenseNet-169 được huấn luyện trên tập dữ liệu kết hợp, theo các phương pháp huấn luyện trước khác nhau. Từ bảng này, chúng tôi đưa ra các nhận xét sau. Đầu tiên, CSSL + TL đạt được hiệu suất tốt hơn so với học chuyển giao (TL). Điều này chứng tỏ hiệu quả của CSSL trong việc cải thiện các biểu diễn đã học. Thứ hai, TL hoạt động tốt hơn so với khởi tạo ngẫu nhiên. Bằng cách kết hợp mặt nạ phổi và mặt nạ tổn thương thông qua học tập đa tác vụ và tận dụng quá trình huấn luyện trước CSSL, phương pháp của chúng tôi đạt được F1 là 0,90, AUC là 0,98 và độ chính xác là 0,89. Bác sĩ X quang cao cấp đã xác nhận rằng các mô hình có hiệu suất như vậy rất hữu ích về mặt lâm sàng.

*Bảng 3: So sánh các phương pháp tiền huấn luyện khác nhau.*



# 5.     Tổng kết

Nhóm đã thu thập bộ dữ liệu hình ảnh CT có sẵn công khai về COVID-19, để thúc đẩy sự phát triển của các phương pháp AI sử dụng CT để sàng lọc và kiểm tra bệnh nhân COVID-19. Bộ dữ liệu chứa 349 hình ảnh CT COVID-19 từ 216 bệnh nhân và 463 hình ảnh CT âm tính COVID-19 (được sử dụng làm ví dụ huấn luyện âm tính). Nhóm cũng thực hiện các nghiên cứu thử nghiệm để cải tiến thêm độ chính xác của tập dữ liệu này. Sử dụng tập dữ liệu này, nhóm tiếp cận dựa trên học tập đa nhiệm vụ và học tập tự giám sát tương phản, đạt được F1 là 0,80, AUC là 0,87 và độ chính xác là 0,78, trên một bộ thử nghiệm các hình ảnh CT gốc do bệnh viện tặng. Độ chính xác của mô hình vẫn cần tiếp tục được cải tiến nhưng vì số lượng hình ảnh vẫn còn hạn chế nên nhóm vẫn còn gặp khó khắn. Mô hình của nhóm vẫn còn nhiều hạn và thiếu sót, em rất mong nhận được sự góp ý từ thầy để trở nên hoàn thiện hơn.

### Tài liệu tham khảo:

- https://paperswithcode.com/paper/covid-ct-dataset-a-ct-scan-dataset-about

- https://pytorch.org/

- https://arxiv.org/pdf/2003.13865v3.pdf

- https://github.com/UCSD-AI4H/COVID-CT

- https://viblo.asia/p/gioi-thieu-mang-resnet-vyDZOa7R5wj

- https://dlapplications.github.io/2018-07-06-CNN/