SIIM-FISABIO-RSNA COVID-19 Detection

120210211 이지환

***Abstract*— Five times more deadly than the flu, COVID-19 causes significant morbidity and mortality. Like other pneumonias, pulmonary infection with COVID-19 results in inflammation and fluid in the lungs. COVID-19 looks very similar to other viral and bacterial pneumonias on chest radiographs, which makes it difficult to diagnose. Computer vision model to detect and localize COVID-19 would help doctors provide a quick and confident diagnosis. As a result, patients could get the right treatment before the most severe effects of the virus take hold.**

Currently, COVID-19 can be diagnosed via polymerase chain reaction to detect genetic material from the virus or chest radiograph. However, it can take a few hours and sometimes days before the molecular test results are back. By contrast, chest radiographs can be obtained in minutes. While guidelines exist to help radiologists differentiate COVID-19 from other types of infection, their assessments vary. In addition, non-radiologists could be supported with better localization of the disease, such as with a visual bounding box.

Keywords*— *COVID-19, Image classification, localization, Object Detection

I. INTRODUCTION

COVID-19 is a major infectious disease that has long been around the world. In the early days of COVID-19, it took a long time to wait for the test results, but recently, technology has developed and positive negative results have been released a day later. Nevertheless, many researchers are conducting studies that allow treatments to make faster and more accurate diagnosis of COVID-19. In this competition, the goal is to detect COVID-19 through models of radiographs, so that deep learning models can perform well to help treatment.

As the leading healthcare organization in their field, the Society for Imaging Informatics in Medicine (SIIM)'s mission is to advance medical imaging informatics through education, research, and innovation. SIIM has partnered with the Foundation for the Promotion of Health and Biomedical Research of Valencia Region (FISABIO), Medical Imaging Databank of the Valencia Region (BIMCV) and the Radiological Society of North America (RSNA) for this competition.

In this competition, It’ll identify and localize COVID-19 abnormalities on chest radiographs. In particular, It'll categorize the radiographs as negative for pneumonia or typical, indeterminate, or atypical for COVID-19. Implemented model will work with imaging data and annotations from a group of radiologists.

If successful, It'll help radiologists diagnose the millions of COVID-19 patients more confidently and quickly. This will also enable doctors to see the extent of the disease and help them make decisions regarding treatment. Depending upon severity, affected patients may need hospitalization, admission into an intensive care unit, or supportive therapies like mechanical ventilation. As a result of better diagnosis, more patients will quickly receive the best care for their condition, which could mitigate the most severe effects of the virus.

A. DATASET OVERVIEW

- Files

train_study_level.csv - the train study-level metadata, with one row for each study, including correct labels.

train_image_level.csv - the train image-level metadata, with one row for each image, including both correct labels and any bounding boxes in a dictionary format. Some images in both test and train have multiple bounding boxes.

sample_submission.csv - a sample submission file containing all image- and study-level IDs.

(1) train study level.csv

id - unique study identifier

Negative for Pneumonia - 1 if the study is negative for pneumonia, 0 otherwise

Typical Appearance - 1 if the study has this appearance, 0 otherwise

Indeterminate Appearance - 1 if the study has this appearance, 0 otherwise

Atypical Appearance - 1 if the study has this appearance, 0 otherwise

(It is a one-hot encoding format.)

(2) train image level.csv

id - unique image identifier

boxes - bounding boxes in easily-readable dictionary format label - the correct prediction label for the provided bounding boxes

II. RELATED WORK

B. OBJECT DETECTION RELATED WORK

Faster R-CNN:Towards Real-Time Object Detection with Region Proposal Networks: Faster R-CNN further improves the region-based CNN baseline. Fast R-CNN uses selective search to propose RoI, which is slow and needs the same running time as the detection network. Faster R-CNN replaces it with a novel RPN (region proposal network) that is a fully convolutional network to efficiently predict region proposals with a wide range of scales and aspect ratios. RPN accelerates the generating speed of region proposals because it shares fully-image convolutional features and a common set of convolutional layers with the detection network. Furthermore, a novel method for different sized object detection is that multi-scale anchors are used as reference.

The anchors can greatly simplify the process of generating various sized region proposals with no need of multiple scales of input images or features. On the outputs (feature maps) of he last shared convolutional layer, sliding a fixed size window (3 × 3), the center point of each feature window is relative to a point of the original input image which is the center point of k (3 × 3) anchor boxes. The authors define anchor boxes have 3 different scales and 3 aspect ratios. The region proposal is parameterized relative to a reference anchor box. Then they measure the distance between predicted box and its corresponding ground truth box to optimize the location of the predicted box.

Experiments indicated that Faster R-CNN has greatly improved both precision and detection efficiency. On PASCAL VOC 2007 test set, Faster R-CNN achieved mAP of 69.9% as compared to Fast R-CNN of 66.9% with shared convolutional computations. As well, total running time of Faster R-CNN (198ms) was nearly 10 times lower than Fast R-CNN (1830ms) with the same VGG [26] backbone, and processing rate was 5fps vs. 0.5fps.

You Only Look Once:Unified,Real-Time Object Detection: YOLO: YOLO (you only look once) is a one-stage object detector proposed by Redmon et al. after Faster RCNN. The main contribution is real-time detection of full images and webcam. Firstly, it is due to this pipeline only predicts less than 100 bounding boxes per image while Fast R-CNN using selective search predicts 2000 region proposals per image. Secondly, YOLO frames detection as a regression problem, so a unified architecture can extract features from input images straightly to predict bounding boxes and class probabilities. YOLO network runs at 45 frames per second with no batch processing on a Titan X GPU as compared to Fast R-CNN at 0.5fps and Faster R-CNN at 7fps.

YOLO pipeline first divides the input image into an S × S grid, where a grid cell is responsible to detect the object whose center falls into. The confidence score is obtained by multiplying two parts, where P(object) denotes the probability of the box containing an object and IOU (intersection over union) shows how accurate the box containing that object. Each grid cell predicts B bounding boxes (x, y, w, h) and confidence scores for them and C-dimension conditional class probabilities for C categories. The feature extraction network contains 24 convolutional layers followed by 2 fully connected layers. When pre-training on ImageNet dataset, the authors use the first 20 convolutional layers and an average pooling layer followed by a fully connected layer. For detection, the whole network is used for better performance. In order to get finegrained visual information to improve detection precision, in detection stage double the input resolution of 224 × 224 in pre-training stage.

C. CLASSIFICATION RELATED WORK

EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks: Convolutional Neural Networks (ConvNets) are often built with a fixed resource budget and then scaled up for higher accuracy when more resources become available. This paper examines model scaling in detail and discovers that carefully balancing network depth, width, and resolution can enhance performance. Based on this finding, a new scaling method is introduced that uses a simple yet very effective compound coefficient to equally scale all depth/width/resolution dimensions. It shows how effective this method is at scaling up MobileNets and ResNets. Furthermore, neural architecture search is utilized to create a new baseline network and scale it up to create the family of models, called EfficientNets, which outperform earlier ConvNets in terms of accuracy and efficiency. In particular, EfficientNet-B7 achieves state-of-the-art 84.3% top-1 accuracy on ImageNet, while being 8.4x smaller and 6.1x faster on inference than the best existing ConvNet. With an order of magnitude fewer parameters, these EfficientNets also transfer well and attain state-of-the-art accuracy on CIFAR-100 (91.7%), Flowers (98.8%), and three other transfer learning datasets. It proves that a mobilesize EfficientNet model can be scaled up very successfully, surpassing state-of-the-art accuracy with an order of magnitude fewer parameters and FLOPS, on both ImageNet and five frequently used transfer learning datasets, due to this compound scaling method.

EfficientNets have shown a very good accuracy in multiple reallife problems like it was used for the automatic diagnosis of COVID-19, skin lesion classification, breast cancer detection, fruit recognition and classification of hematoxlin from images [7][8][9][10][11].

An Image Is Worth 16*16 Words:Transformers for Image Recognition At Scale: While the Transformer architecture has become the de-facto standard for natural language processing tasks, its applications to computer vision remain limited. In vision, attention is either applied in conjunction with convolutional networks, or used to replace certain components of convolutional networks while keeping their overall structure in place. It show that this reliance on CNNs is not necessary and a pure transformer applied directly to sequences of image patches can perform very well on image classification tasks. Vision Transformer (ViT) attains excellent results compared to state-of-the-art convolutional networks while requiring substantially fewer computational resources to train.

III. METHODOLOGY

In this section, the model architecture is discussed in detail including the dataset description, data augmentation and model architecture diagram. The whole architecture is shown by the following diagram.

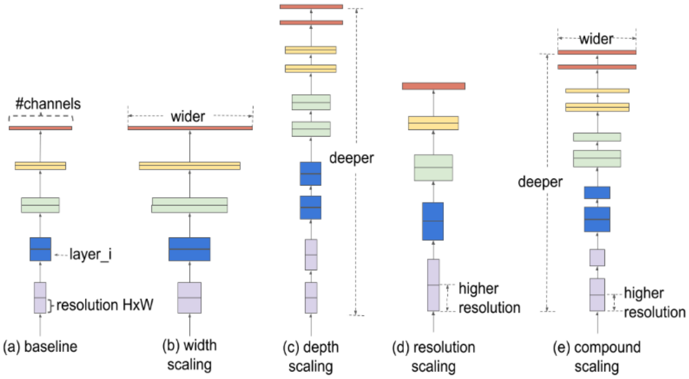

Figure 1: Architecture diagram of efficientnetb4

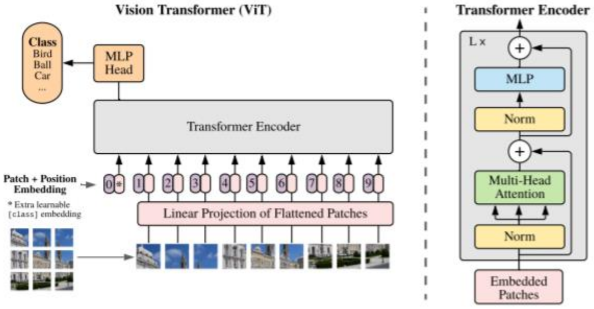

Figure 2: Architecture diagram of VIT

In object section, I try to apply a YOLOv5 model.

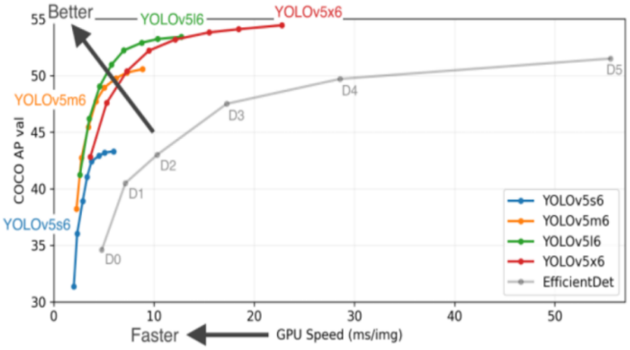

Figure 3: YOLOv5

D. Dataset

Negative for COVID-19, its class is 'Negative for Pneumonia' and three major classes for positive. The three types are'Typical','Indeterminate','Atypical'.

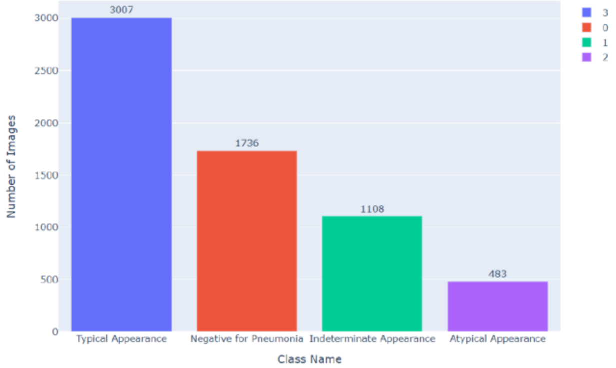

Figure 4: class distribution

E. Data Preprocessing

In order to train the model, we have done a lot of data processing so that we can get make our dataset in required format and make the model run on TPU more efficiently. We have done the following processing on the plant's dataset.

- Change DIM format to JPG.
- Removed duplicates from the dataset.
- Resized the images to 512x512.

F. Data Augmentation

In order to achieve high accuracy. Data augmentation played a major role in it [6]. As we can see from the class distribution diagram the dataset was not balanced so, there were some chances that model would converge to particular class but, we need a more generalized model [5]. Applying data augmentation with image data generator class in keras was also an option but it was really time consuming while training the model and can affect the TPU performance so, we have used Albumentation library to apply the data augmentation like vertical flip, horizontal flip, random rotate, shift scale rotates, median blur etc. following diagram, shows some data augmentation results.

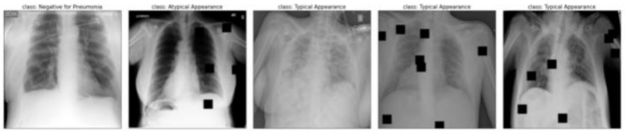
class: Negative for Pneumonia
class: Atypical Appearance
class: Typical Appearance
class: Typical Appearance
class: Typical Appearance

Figure 5: Data Augmentation

G. EfficientNet b7 Architecture

The basic architecture of the model that shows the scaling is given in figure 1. we used a new method for a scaling up of the model, which makes use of a simple but effective CNN factor analysis in a more structured way. In contrast to the traditional methods used in any network, efficientnet set up the dimensions, such as width, height, and resolution, the model scale each dimension is evenly spread, with a fixed set of scaling factors. On the basis of the new scaling method, and the recent advances in AutoML, our method belongs to a family of models is referred to as EfficientNets that 10 times smaller and faster.

Compound Scaling:

To understand the effect of the network on the scale, they have investigated the effect of scale model sizes. At the scale of the individual dimensions of the model's performance, we observed that the consideration of all the dimensions of the network width, depth, and clarity with the best available resources, to enhance the overall system performance. The first step in a full zoom and a method for performing a grid-search on the relationship between a variety of scalable dimensions, the underlying network is under constant resource constraints. This will determine the appropriate zoom level for each of the dimensions mentioned above.

The range of the efficiency of the model is highly dependent on the underlying network. Therefore, in order to further improve efficiency, they developed a new core network, the search for a neural architecture with the AutoML MNAS framework, which allows for optimal accuracy and efficiency. This architecture allows for the use of a mobile Inverted bottleneck convolution (MBConv), which is comparable with MobileNetV2 and MnasNet, but it is slightly larger due to the increase in the project on the FLOP.

IV. Experiments and Evaluation

H. Image Classification

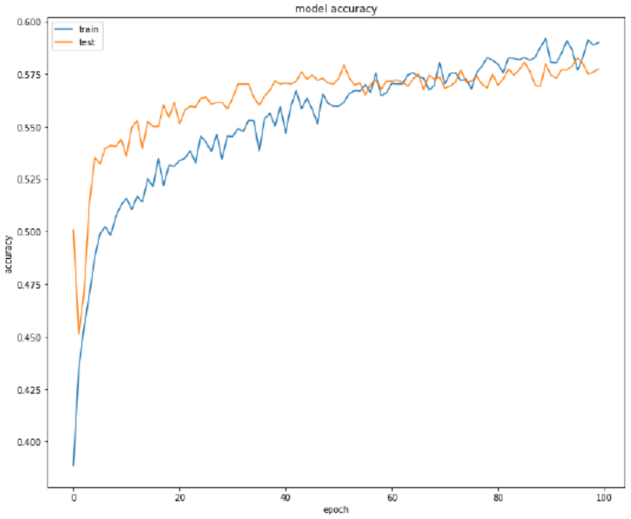

Figure 6: VIT performance

The VIT model showed nearly 60% performance on train and test datas.

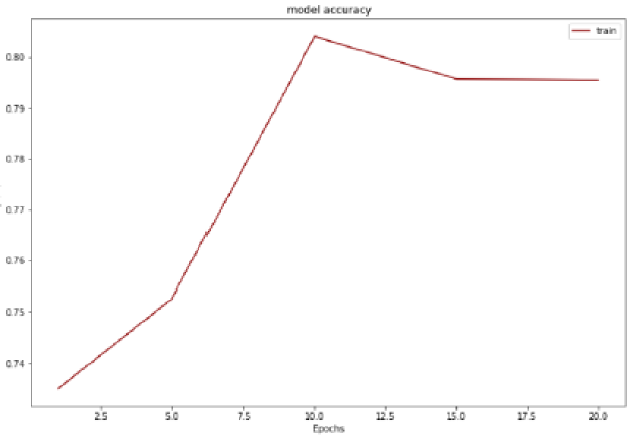

Figure 7: EfficientnetB7 performance

The EfficientnetB7 model showed nearly 80% performance on train data.

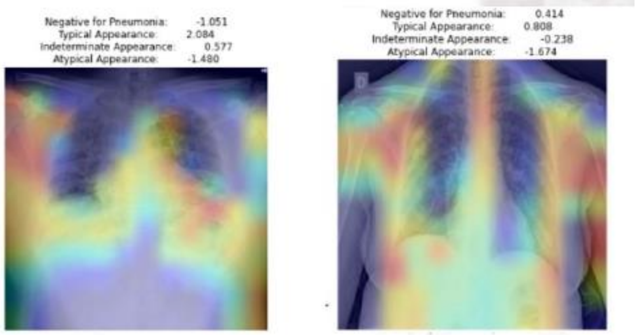
Negative for Pneumonia: -1.051
Typical Appearance: 2.084
Indeterminate Appearance: 0.577
Atypical Appearance: -1.480
Negative for Pneumonia: 0.414
Typical Appearance: 0.808
Indeterminate Appearance: -0.238
Atypical Appearance: -1.674

Figure 8: Heat-Map in radiographs

The heat map shows which part of each class is paying attention to.

I. Object Detection

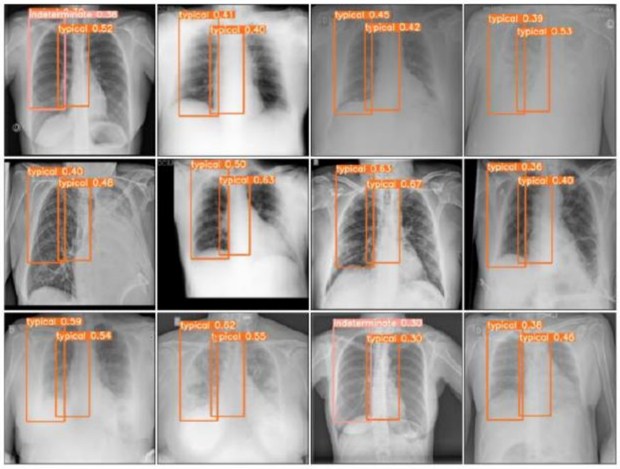

Figure 9: YOLOv5 model result on radiographs

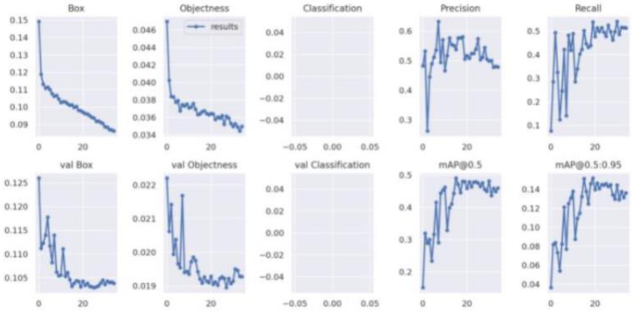

Figure 10: Performance changes with epochs. (Precision,Recall,mAP)

V. Conclusion

The classification was performed by applying efficientnetb7, and localization applied yolov5. The final performance was 59.8% based on the mAP.

VI. References

[1] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun. "Faster R-CNN:Towards Real-Time Object Detection with Region Proposal Networks." *arXiv preprint arXiv:1506.01497.(2015)*

[2] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. "You Only Look Once: Unified, Real-Time Object Detection." *arXiv preprint arXiv:1506.02640.(2015)*

[3] Mingxing Tan, Quoc V. Le. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks " *International Conference on Machine Learning, 2019*, ICML 2019.

[4] Alexey, et al. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale” " *International Conference on Machine Learning, 2019*, ICML 2021.

[5] Van Dyk, David A., and Xiao-Li Meng. "The art of data augmentation." *Journal of Computational and Graphical Statistics* 10.1 (2001): 1-50.

[6] Perez, Luis, and Jason Wang. "The effectiveness of data augmentation in image classification using deep learning." *arXiv preprint arXiv:1712.04621* (2017).

[7] Marques, Gonçalo, Deevyankar Agarwal, and Isabel de la Torre Díez. "Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network." *Applied Soft Computing* 96 (2020): 106691.

[8] Gessert, Nils, et al. "Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data." *MethodsX* 7 (2020): 100864.

[9] Wang, Jun, et al. "Boosted efficientnet: detection of lymph node metastases in breast cancer using convolutional neural networks." *Cancers* 13.4 (2021): 661.

[10] Duong, Linh T., et al. "Automated fruit recognition using EfficientNet and MixNet." *Computers and Electronics in Agriculture* 171 (2020): 105326.

[11] Munien, Chanaleä, and Serestina Viriri. "Classification of Hematoxylin and Eosin-Stained Breast Cancer Histology Microscopy Images Using Transfer Learning with EfficientNets." *Computational Intelligence and Neuroscience* 2021 (2021).