BiLinear CNNs Model and Test Time Augmentation for Screening Viral and COVID-19 Pneumonia

Anh-Kiet Duong
University of Science, Ho Chi Minh City, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam
18120046@student.hcmus.edu.vn

Van-Hien Huynh
University of Science, Ho Chi Minh City, Vietnam
Vietnam National University, Ho Chi Minh City, Vietnam
18120361@student.hcmus.edu.vn

***Abstract*—This year, COVID-19, a novel coronavirus has made a worldwide crisis. Several researchers have tried to create machine learning models to diagnose infections via their chest x-ray image. We noticed that some datasets are out of balance because of the bones, characters,... in the image. We present an image preprocessing to rebalance the dataset, test time augmentation to perform random modifications to the test images. And apply bilinear CNNs to the problem of screening Viral and COVID-19 pneumonia because different baseline architecture give different failure results. Our proposed can improve accuracy up to 1.4% compared to state-of-the-art on the dataset given by researchers Muhammad E. H. Chowdhury, Tawsifur Rahman, and their colleagues on the paper "Can AI help in screening Viral and COVID-19 pneumonia?"-IEEE Access. Through this paper, we hope to be able to help in diagnosing infections of COVID-19 and especially in the application of machine learning to this problem.**

I. Introduction

When the COVID-19 virus starts to be recognized, classify chest x-ray images is a way to help doctors screening patients. Scientists apply machine learning to help the classification more accurate and automated.

The problem is the dataset about this novel coronavirus. There are few resources and we face difficulty in finding good data. We get the dataset from the paper "Can AI help in screening viral andCOVID-19 pneumonia?." [1]. But we realize that the dataset of the paper is unbalanced. The set of negatives and the set of positives have multiple different points, not to mention the virus. We do some image preprocessing to make the dataset more neutral.

Also, different models have different advantages and disadvantages. We propose to combine them together, using Bi-Linear CNNs technique. The experiment show that we will get higher result.

II. Related Works

III. Methods

In this section, we describe our approach to screening Viral and COVID-19 pneumonia. First, we demonstrate the image preprocessing for all the images in the dataset. Then we build the BiLinear CNNs model. Finally, apply test time augmentation when testing the model. In *Fig.1*, we show the way our model works.

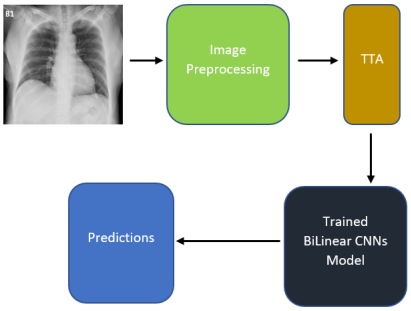

Fig. 1. Method overview

A. Bi-Linear CNNs

After cropping the image results have increased. However, each CNN has a very different result. So we use the bi-Linear model which introduce in [2] for fine-grained visual recognition to get higher results. The bi-linear model can be very heavy, slow and will take a lot of memory to achieve high accuracy. So this method is only suitable for tasks that require very high precision. And screening viral and COVID-19 Pneumonia problem is a suitable problem.

A bi-linear model consists of a quadruple $B = (f_A, f_B, P, C)$. In all these descriptors both f_A and f_B are functions based on CNNs to extract features from images. A feature function is a mapping $f : L \times I \rightarrow \mathbb{R}^{K \times D}$ that takes an image I and a location L and emits a characteristic of size $K \times D$. The outputs are combined at each location using the external product of the matrix, it is, the combination of biLinear characteristics of f_A and f_B at a location l given by: $bilinear(l, I, f_A, f_B) = f_A(l, I)^T f_B(l, I)$ Both fA and fB must have the same characteristic dimension c to be compatible. Here P is a pooling function and C is a classification function.

1) Label Smoothing: This loss is a regularization technique that addresses both problems overfitting, and overconfidence.
$$LS(q_i) = q_i^{'} = (1 - \epsilon)q_i + \frac{\epsilon}{K}$$
2) Focal Loss: This loss is greatly reduces the effect of easy classify areas and small a reduction in areas that

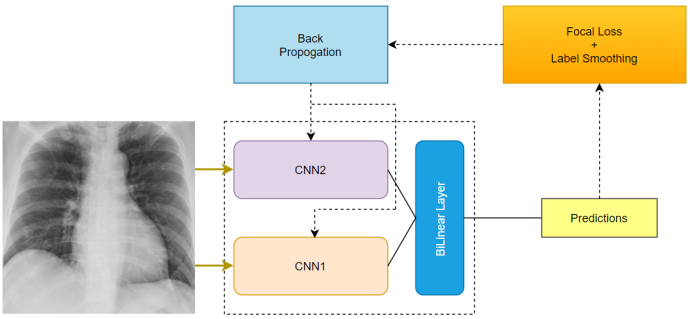

Fig. 2. BiLinear CNNs

are hard to classify. So that the model focuses more on hard case.

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t)$$

We combine the two and use it as loss function. We choose $\alpha = 0.25, \gamma = 2.0, \epsilon = 0.1$ because it works best in experiments.

$$Loss = FL(LS(p))$$

B. Test Time Augmentation (TTA)

The application of data augmentation techniques normally used during training when making predictions. In other words, TTA is simply to apply different transformations to test image like: rotations, flipping, brightness changes,... Then feed these different transformed images to the trained model and average the results to get more confident answer. For example, the image shown below applied transformations together with the original image. All of these images are passed to the same model and the results are averaged.

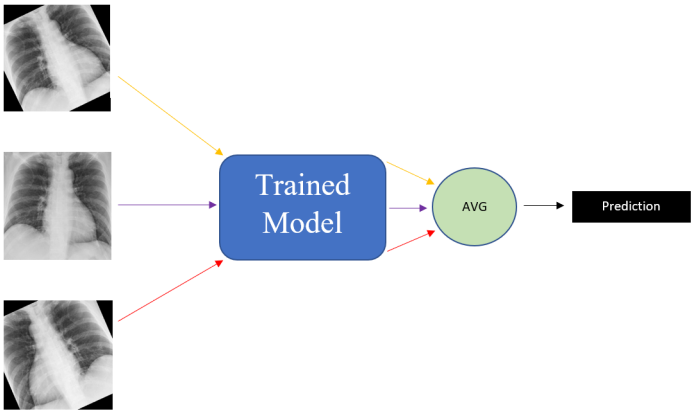

Fig. 3. Test Time Augmentation

IV. Experiments and Results

A. Image Preprocessing

First, take a look at the dataset. We can see that most of the normal images (*Fig.2*) include 'R' character, scapula and lower jaw bone. Most of the Viral Pneumonia images (*Fig.3*) include 'R' character, scapula. While Covid-19 images (*Fig.4*) dose not have these. The imbalance factors will cause the model to bring high results but cannot be applied practically. In order for the model to apply to screening Viral and COVID-19 pneumonia, we must remove those elements from all images of the dataset. This way is specific to this dataset, maybe we don't need to use this for another dataset.

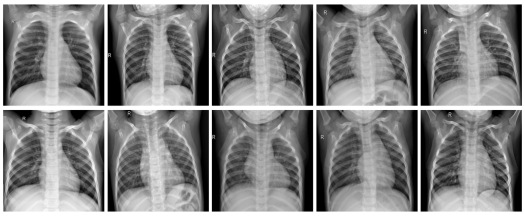

Fig. 4. Normal Images mostly have scapula, jaw bone, blank space...

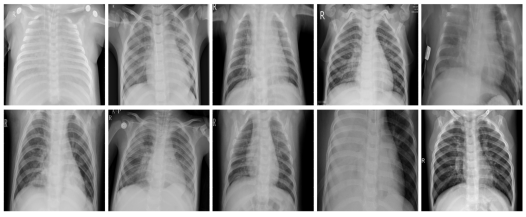

Fig. 5. Viral Pneumonia Images mostly have blank space

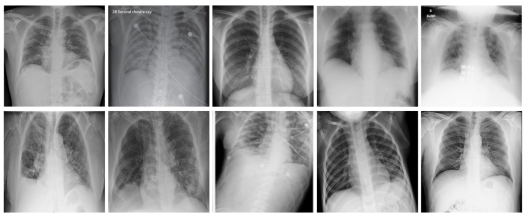

Fig. 6. Majority of COVID-19 images do not have scapula, jaw bone and blank space

So we try to crop from four sides of all the images in the dataset. The results are shown in the table below.

	Best Accuracy
Crop 0% (orginal)	97.94 [1]
Crop 5%	98.03
Crop 10%	98.18
Crop 12.5%	98.21
Crop 15%	98.14
Crop 20%	97.41

TABLE I
THE RESULTS OF IMAGE PREPROCESSING METHOD FROM BEST SINGLE MODEL

The authors crop all images in the dataset (12.5% from 4 sides). This method is to remove unbalanced feature in the image like scapula, lower jaw bone, characters,...

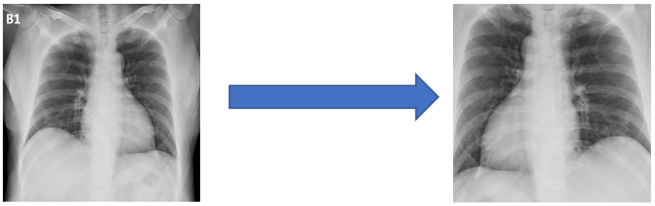

Fig. 7. Image Preprocessing: Crop image to the lung zone

B. Test time Augmentation

We apply some transformations to test images like: rotations (20°), horizontal flipping, changes brightness (5%). The results are shown in the table below.

	Models	Accuracy
Whitout TTA [1]	DenseNet201	97.94
	VGG19	96.00
	InceptionV3	96.20
	MobileNetv2	96.22
Whitout TTA	DenseNet201	98.34
	VGG19	96.83
	InceptionV3	97.28
	MobileNetv2	97.24

TABLE II
THE RESULTS OF TEST TIME AUGMENTATION METHOD

C. Bi-Linear CNNs

The authors use DenseNet201, VGG19, InceptionV3, MobileNetV2 for this method. This is an update of TTA method so all the results applied TTA.The bi-linear model can be very heavy, slow and will take a lot of memory for high accuracy. The results are shown in the table below.

	DenseNet201	VGG19	InceptionV3	MobileNetV2
DenseNet201	-	99.35	98.45	98.18
VGG19	99.35	-	98.72	98.45
InceptionV3	98.45	98.45	-	98.52
MobileNetV2	98.18	98.72	98.52	-

TABLE III
THE RESULTS OF BI-LINEAR CNNS MODEL

REFERENCES

[1] Chowdhury, Muhammad EH, et al. "Can AI help in screening viral and COVID-19 pneumonia?." arXiv preprint arXiv:2003.13145 (2020).

[2] Lin, Tsung-Yu, Aruni RoyChowdhury, and Subhransu Maji. "Bilinear cnn models for fine-grained visual recognition." Proceedings of the IEEE international conference on computer vision. 2015.

[3] He, Tong, et al. "Bag of tricks for image classification with convolutional neural networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

[4] Lin, Tsung-Yi, et al. "Focal loss for dense object detection." Proceedings of the IEEE international conference on computer vision. 2017.

[5] Gao, Yang, et al. "Compact bilinear pooling." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.