Estimating the potential of Traditional Geometric Data Augmentation in improving the classification accuracy of Covid-19 detection algorithm

Shahin Saleh **Zohib Sekandari** **Muhammad Hafeez Saeed**

Abstract

This paper estimates the potential of Traditional Geometrical Data Augmentation (TGA) techniques to enhance the classification accuracy of COVID-net model over unbalanced chest x-ray (CXR) images dataset. In the light of achieved results, it has been proved that TGA techniques help improve the model's classification performance for COVID-19 virus. In the end, the report also presents potential improvement suggestions for the COVID-net model regarding the way dataset could be augmented better.

1 Introduction

With COVID-19 virus spreading in a rapid phase and classified as a pandemic, many sectors have shifted their focus to try and aid the healthcare. One of the prominent contributions is the COVID-Net tool developed by researchers Linda Wang and Alexander Wonga from the University of Waterloo and Canadian startup Darwin AI. COVID-net is a deep convolutional neural network design tailored for the detection of COVID-19 cases from chest radiography images[Wang, Lin, and Wong 2020]. The data set that they have shared with the public contains three different types normal, pneumonia and COVID-19 images, but the data set is unbalanced and has few COVID-19 radiography images. to increase the COVID-19 radiography examples, our project implements various traditional geometrical data augmentation (TGA) techniques.

This project deals with radiography images, therefore, in order to maintain the useful information content only a few data augmentation techniques were reasonable to implement in order to avoid erratically transformed images. The results of this project have been presented in this report by using confusion matrices to observe the difference between the classification results when the COVID-net model is trained with and without data augmentation. A clear concern related to overfitting has been found when training without TGA, which seems an outcome of unbalanced COVID-19 example images in comaprison to large number of Normal and Pneumonia images. Through this project, we aim to overcome this issue by increasing the portion of COVID-19 images through TGA, so that our model gets better trained on every patient category. Through the obtained results from this project, we aim to contribute to the large open source project of COVID-net and provide some useful suggestions for the model improvement.

2 Related work

A related research where the project team had unbalanced data of chest radiography images, they tried various augmentation techniques to enhance their dataset size and thus improve classification accuracy. The augmentation technique that they used was Pix2Pix Generative Adversarial Network (GAN) [Xing et al. 2019]. To confirm the quality of the generated images, two radiologists were hired, which were shown 4 different image sets: one real image set and other three were GAN generated, and the GAN generated images were found satisfactory.

Another research conducted by Ganesan et al. 2019, confirms that for CXR images and abnormality detection classifier, following conclusions hold valid: (1) GAN Augmentation (GA) is not always superior to Traditional Geometrical Augmentation (TGA) for improving the classifier's performance; (2) in comparison to No Augmentation, however, both TGA and GA leads to a significant performance improvement; and, (3) increasing the quantity of images in TGA and GA strategies also improves the classifier's performance.

3 Data

Three different datasets were used for the prupose of this project i.e. *covid-chestxray-dataset, Figure1-COVID-chestxray-dataset and rsna-pneumonia-detection-challenge*. These datasets were combined together, resulting into three classes: normal, pneumonia and COVID-19. Thereby, resulting into a large homogeneous dataset called COVIDx with following breakdown:-

Table 1: Total Samples Breakdown

Normal	Pneumonia	COVID-19
8851	6043	254

A breif description about the individual datasets is given below:

3.1 covid-chestxray-dataset

This dataset is a publicly available collection of CT and X-ray images from patients that are suspected or have been confirmed with COVID-19 virus[Joseph, Paul, and Lan 2020]. The dataset contains 354 sample images and the metadata contains 28 different variables. We have used 287 out of these images as the rest are unlabelled. Finally, a total of 260 images have been taken with the following breakdown. Normal : 0, Pneumonia: 33 and COVID-19: 227.

3.2 Figure1-COVID-chestxray-dataset

This is a small dataset that was initially used by Darwin AI, having a total of 27 different images all labeled as COVID-19 [Wang, Lin, and Wong 2020]. The metadata contains 12 different variables. This dataset is available free online and growing everyday as more COVID-19 positive CXR images become available.

3.3 rsna-pneumonia-detection-challenge

This dataset is a public dataset and it is available at Kaggle by Radiological Society of North America. It was used for a competition to create a model that can detect pneumonia from chest x-ray images. The dataset does not consist of any COVID–19 examples but has many Normal and Pneumonia examples which could help better train our model. Overall, this dataset provides us with sample images with the following breakdown. Normal : 8851 and Pneumonia: 6012.

4 Method

4.1 COVID-net

The main aim of this project was to examine the effects of data augmentation on the classifier's performance. Therefore, the used COVID-net model in our project was taken from online initiative started by Wang et. al (2020). It has been pre-trained on ImageNet (http://www.image-net.org/) and afterwards trained on COVIDx dataset using Adam Optimizer using a learning rate policy where the learning rate decreases when learning stagnates for a period of time (i.e., 'patience') [Kingma and Ba 2019]. The implementation of COVID-net model has been done on Keras deep learning library with tensorflow as back end. The COVID-net architecure is based upon Generative Synthesis (SynSenth) technique, which ensures a final COVID-net work architecture:

1. COVID-19 sensitivity $\geq$ 80 %
2. COVID-19 Positive Predictive Value (PPV) $\geq$ 80 %

4.2 Data Augmentation

The considered data augmentation approaches were traditional geometrical augmentation (TGA) and Generative Adversarial Network Augmentation (GAN-A). Due to limited time scope and lack of computational resources it was challenging to train a GAN that would then synthesize new CXR images. Also it has been observed in case of CXR images, there is no substantial enhancement in model's accuracy which is trained by GAN-A as compared to TGA. [Ganesan et al. 2019]. Therefore, we examined various TGA methods on the three datasets mentioned in section 3. To apply data augmentation we used Keras and tensorflow as front end and back end respectively. More specifically the keras class ImageDataGenerator() was used to implement the data augmentation.

The various data augmentation techniques that were implemented using Keras ImageDataGenerator() class are mentioned below:

4.2.1 Rotation

Considering a controlled hospital environment, and aiming to keep the augmented CXR images close to reality, we put a rotation cap at 15 °and floor at 10 °

Rotation Range: 10-15 degrees

4.2.2 Translation

Translation to an image means moving all pixels of the image in one direction, such as horizontally or vertically, while keeping the image dimensions the same. Considering a controlled hospital environment, not much translations/movements are expected during CXR. Therefore, we introduce small translation parameters to train the model to look around a bit while classifying any CXR image

Width Shift Range: 0.1-0.2
Height Shift Range: 0.1-0.2

4.2.3 Variation in Brightness

The brightness of the image can be augmented by either randomly darkening images, brightening images, or both. The intent is to allow a model to generalize across images taken on different lighting levels. Therefore, we keep the brightness parameter such that it augments the original pictures with dark and bright images

Brightness Range: 0.8-1.1

4.2.4 Zoom Range

Zoom augmentation works by randomly zooming the picture in or out as provided in the range. It has ability to add new pixels around the image or interpolate the image pixel values if needed. To keep the augmented pictures close to reality we keep a safe zoom range:

Zoom Range: 0.8-1.2

4.2.5 Flipping

Flipping simply means reversing the rows or columns of the pixels in case of a horizontal or vertical flip. Practically, vertical flips are not possible for the CXR patient images. However, we have tested both vertical and horizontal flip augmentation to check their effects on model's accuracy.

Horizontal flipping: Enabled
Vertical flipping: Enabled

4.2.6 Fill Mode

For all the cases we have set fill mode argument equal to "constant", this parameter helps to fill in the missing pixels caused due to rotation or translation. This way the augmented images make more sense and model is better trained.

Fill Mode: Constant

4.3 Evaluation Metrics

For the purpose of evaluating the classification performance of the trained COVID-net model with and without data augmentation, we used confusion matrices. A 3x3 confusion matrix was developed at every epoch. Using confusion matrix, recall and precision were calculated. Evaluation metrics were implemented on python using train_test_split() function from sklearn library. For a classifier to be trained with an unbalanced dataset, a precision - recall trade-off needs to be maintained [Flach and Kull 2015], Therefore, for every epoch, both recall (or Sensitivity) and precision (or PPV Value) were plotted and the maximum epoch number was decided where both had optimum values at the same time. Moreover, the loss curve was also plotted for checking classifier's overall accuracy with and without data augmentation.

5 Experiments and Analysis

For the purpose of training and validation of the classifier we have divided the entire data set (COVIDx) into training and test set. The breakdown of training and test set has been shown in table 1 and 2 respectively. However, it is important to be considered, for the case of TGA based training, we have augmented the COVID-19 images so that we have at least 30% COVID-19 images available for training.

Table 2: Train Set

Normal	Pneumonia	COVID-19
8751	5943	223

Table 3: Test Set

Normal	Pneumonia	COVID-19
100	100	31

5.1 Simple Training without Data Augmentation

We trained the model without TGA for 10 Epochs. The following Positive Predictive Value (PPV) and Sensitivity scores were obtained during training:

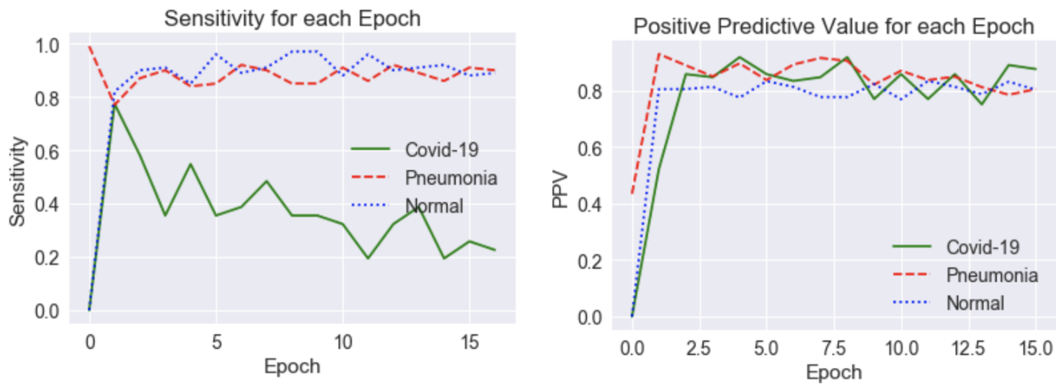

Figure 1: The Sensitivity and PPV plots for training the model without Data Augmentation.

In our case, having a decent figure of sensitivity (recall) and PPV (precision) was crucial. We have defined sensitivity as the number of True Positive predictions of COVID-19 out of total available COVID-19 examples, while PPV tells the number of people that when predicted with COVID-19, actually had the disease. In Figure 1, we can see that our model flattens out at sensitivity of around 90% for the normal and pneumonia cases. However, for the Covid-19 cases after epoch two (2), the sensitivity starts dropping from 80% to around 23% near the Epoch ten (10). The PPV score for normal, pneumonia and Covid-19 are fluctuating between 80% to 97%. Therefore, in case of no TGA, the best model training would be achieved at epoch four (4) since, we get least Sensitivity - PPV trade-off.

To estimate the model's classifying ability on training data, we calculated the loss function in each Epoch. The graph of loss curve is given in Figure 2:

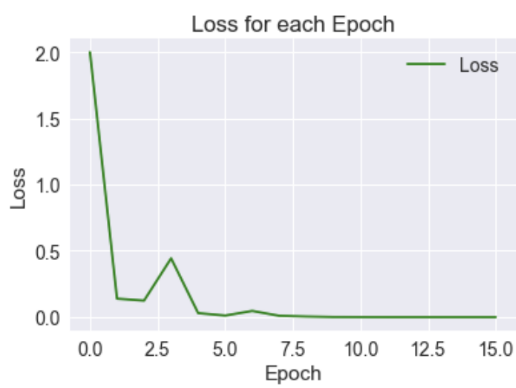

Figure 2: The Loss plot for training the model without Data Augmentation.

From the above figure, it is clear that the model is truly learning on the training data and thus have a clear decline of loss function.

To accurately number the classification performance of the model on the training and test data, the confusion matrices were obtained for Epochs 5 and 10. These are presented in Figure 3:

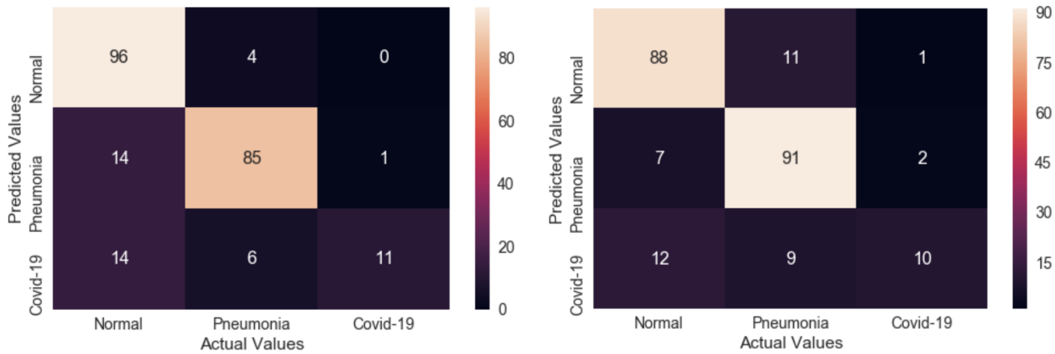

Figure 3: The Confusion Matrix for Epoch 5 (left) and 10 (right)

In the above figure, it is clear that as the Epoch number increase the model shows signs of overfitiing and thus unable to correctly predict the True COVID-19 examples. For example: at Epoch five (5), a total of 14 normal patients and 6 pneumonia patients have been classified incorrectly as COVID-19. The same trend can be observed for later Epochs, which shows that the model overfits and lacks generalizability on test data.

5.2 Training with Data Augmentation

Initially, we used each of the basic TGA techniques for training our model i.e. Rotation, Translation, Variation in Brightness, Zooming, Flipping (Vertical and Horizontal) and Fill Mode. We got gradually better results when we applied every individual technique one by one, except for the Vertical Flipping, which caused a decrease in model's performance. Therefore, when we trained our model we omitted the Vertical Flip and used the rest of the TGA Techniques. We trained the model with data augmentation for 10 Epochs. The following Positive Predictive Value (PPV) and Sensitivity scores were obtained during training:

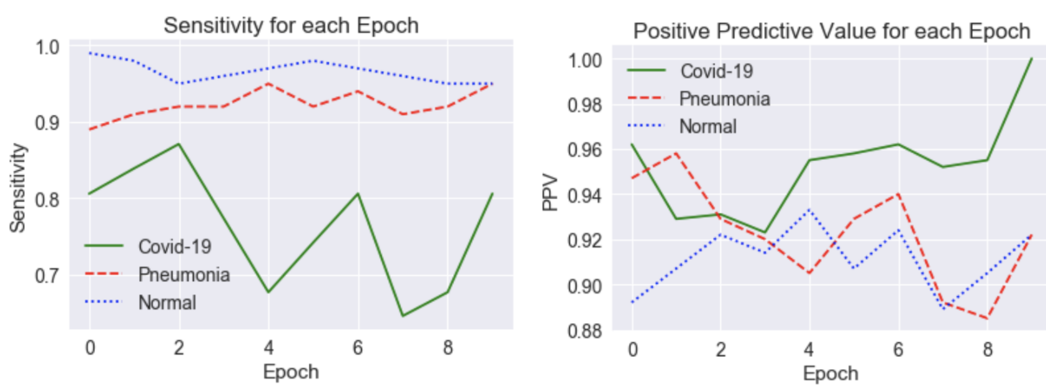

Figure 4: The Sensitivity and PPV plots for training the model with Data Augmentation.

In Figure 4 above, although the pattern is similar to that of Figure 1 (training without TGA), but we can see much less fluctuations in the sensitivity plot and an overall increase in both the sensitivity and PPV values. This indicates that the model is now able to classify images much accurately and does not visibly miss out any true COVID-19 case. Moreover, in case of TGA applied, the best model training would be achieved at epoch six (6) since, we get least Sensitivity - PPV trade-off.

To estimate the model's classifying ability on training data, we calculated the loss function in each Epoch. The graph of loss function is given in Figure 5:

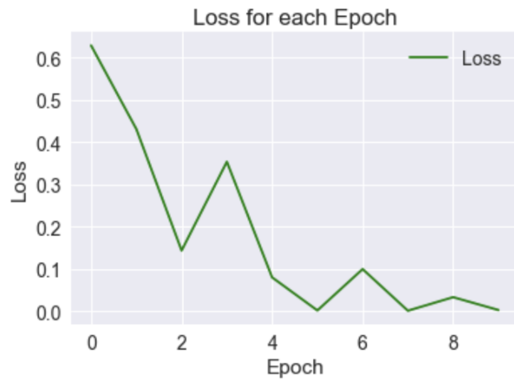

Figure 5: The Loss plot for training the model without Data Augmentation.

From the above figure, it is clear that the model is again, truly learning on the training data and thus have a declining loss function.

To accurately number the classification performance of the model with TGA applied on the training and test data, the confusion matrices were obtained for Epochs 5 and 10. These are presented in Figure 6:

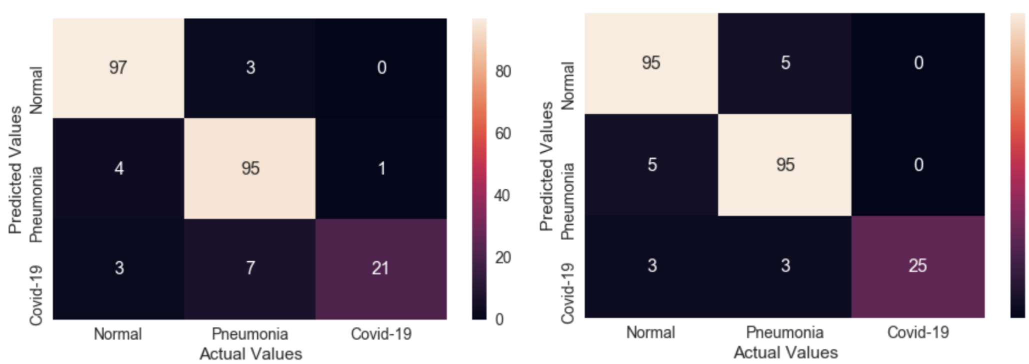

Figure 6: The Confusion Matrix for Epoch 5 (left) and 10 (right)

In the confusion matrices, we can see clear improvement in number of correctly classified cases for all three categories (normal, pneumonia and Covid-19). Most noticeable improvement is that observed in COVID-19 cases, where the number of false predicted COVID-19 cases have decreased to 10 at epoch five (5) as compared to NO TGA results (20 at epoch five (5)). However, the overfitting trends would again appear if the number of training epochs keep on increasing, therefore in order to select the optimal model with optimal train and test accuracy we seek for the least Sensitivity - PPV trade off, which was found to at epoch six (6) using Figure 4.

6 Conclusion

The results obtained through the experiment section show that the Traditional Geometrical data Augmentation (TGA) techniques generally help increase the classifier's test accuracy on the COVID-19 cases. Whereas, without TGA there is a visible decrease in classifier's performance. However, one thing worth considering is that not all the TGA techniques render improved results, for instance, vertical flipping data augmentation when applied resulted into poor precision and recall results. Moreover, after analyzing the loss curve, sensitivity (or recall) and PPV (or precision) plots, we conclude that maximum number of training epochs should be limited to six (6) with data augmentation applied. Whereas, the maximum number of training epochs should be limited to four (4) in case of no data augmentation. Considering the success we achieved with TGA, we still think there is scope for more advanced data augmentation techniques like: GAN generated data augmentation, that could be tested and verified against COVID-net model and potentially enhance the screening accuracy of COVID-19 virus. Overall, this project helps understand the potential of TGA techniques in improving classifier's performance in case of highly unbalanced training data like COVIDx dataset.

References

Flach Peter, A and Meelis Kull (2015). "Precision-Recall-Gain Curves: PR Analysis Done Right". In: *NIPS*. DOI: `https://papers.nips.cc/paper/5867-precision-recall-gain-curves-pr-analysis-done-right.pdf`.

Ganesan, P et al. (2019). "Assessment of Data Augmentation Strategies Toward Performance Improvement of Abnormality Classification in Chest Radiographs". In: *Lister Hill National Center for Biomedical Communications*. DOI: `https://lhncbc.nlm.nih.gov/publication/pub9938`.

Joseph Paul, Cohen, Morrison Paul, and Dao Lan (2020). "COVID-19 image data collection". In: *arXiv 2003.11597*. URL: `https://github.com/ieee8023/covid-chestxray-dataset`.

Kingma Diederik, P and L Ba J (2019). "ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION". In: *ICLR*. DOI: `https://arxiv.org/pdf/1412.6980.pdf`.

Wang, L, Q Lin Z, and A Wong (2020). *COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest Radiography Images*. arXiv: `2003.09871 [cs.CV]`.

Xing, Yunyan et al. (2019). "Adversarial Pulmonary Pathology Translation for Pairwise Chest X-ray Data Augmentation". In: *arXiv preprint arXiv:1910.04961*.