CS230

Covid-Net-Lite: SARS-CoV-2 Detection with Data Augmentation and a Lightweight Convolutional Neural Network

Miguel Ayala
Department of Computer Science
Stanford University
mayala3@stanford.edu

Abstract

In this study we show that with a relatively simple data augmentation technique and a lightweight convolutional neural network we are able to detect SARS-CoV-2 from chest radiographs at levels exceeding or matching state of the art models. Covid-Net-Lite, the model we create, only requires 2 convolutional layers, a representation size of around 0.15MB per image and only 729 trainable parameters. Despite this, when trained on an augmented and distorted dataset of 9654 radiographs we are able to overcome the problems created by the lack of SARS-CoV-2 images and outperform the likes of VGGNet and ResNet50 with an accuracy of 98.69%.

1 Motivation

1.1 A Global Pandemic

As of June 8 2020, SARS-CoV-2, a coronavirus, has spread worldwide infecting 6,931,000 people and causing 400,857 deaths [17]. One of the main issues preventing the pandemic's suppression has been the lack of adequate testing.

1.2 Limitations on Testing

For a variety of reasons, the number of testing kits produced and the number of kits that can be analyzed at any given time falls far short of the numbers [1] seen in countries that have successfully stemmed the spread of the disease [14]. If production cannot be scaled up to necessary levels, alternative tests to the current molecular-based approaches should be explored. Aside from serological-based tests [11] [20], there has been promise in detection of the virus via chest radiography. Early findings from China suggest that patients with SARS-CoV-2 seem to exhibit abnormal structural differences in their lungs including small pleural effusions and pulmonary consolidation [22]. It may therefore, be worthwhile to see if analysis of chest scans may help diagnose SARS-CoV-2 patients. Additionally, chest scans are a promising avenue to explore because though the equipment required is widespread [13], the expertise needed to provide accurate readings is much more scarce [8].

2 Previous Approaches

2.1 CNNs + Pulmonary Health

There has been plenty of work in using CNNs to classify various pulmonary ailments such as pneumonia, asthma and tuberculosis [2] [9] [3] [15] with reasonable success. Aside from the detection of these diseases other researchers have used class activation maps to localize a host of numerous pathologies, to give a more granular view of patient lung health [12]. In some cases, the performance of these systems is better than that

[1] in terms of a test per capita basis

CS230: Deep Learning, Spring 2020, Stanford University, CA.

of experienced radiologists in terms of accuracy and specificity.

2.2 CNNs + SARS-CoV-2

Given the recent success of CNNs in analyzing pulmonary radiographs, it is only natural that researchers have attempted to use CNNs to help detect SARS-CoV-2 from chest radiographs. Wang et al [16] curated a large dataset of radiographs from multiple sources and built a deep neural network - COVID-Net - of stacked convolutional layers. The modules were arranged to promote long range connectivity. In the study they compared the results of their network to those of VGG-19 and ResNet-50 when run on the same dataset. Narin et al [10] used pre-trained models followed by an average pooling layer, ReLU activation and a softmax layer to predict whether or not a patient had SARS-CoV-2. They experimented on 50 positive images and 50 negative images using pretrained models from InceptionV3, ResNet50 and InceptionResNetV2. Abbas et al [1] attempted a similar approach in their DeTraC model where the process was broken down into several phases consisting of a pretrained CNN and transfer learning with ResNet18. While these papers all produce accuracy scores well above 85%, there are improvements that can be made to their datasets. Though the Narin paper does consciously attempt to balance their data, only having 100 examples in training may severely limit their model's real world performance. The Wang paper found many more examples of SARS-CoV-2 radiographs, however, only 358 of their 13,975 examples are SARS-CoV-2 examples. It is possible that their model may perform even better with a balanced dataset. The Abbas study's data approach combined balance and volume by using data augmentation. Nevertheless, they could probably have benefited from having more than 1,700 examples. Additionally these models may be better suited for real life use if they were designed more economically. All of the models reviewed are incredibly large. The smallest of the models surveyed - COVID-Net - had 11.5 million parameters. If the idea is to make testing as accessible as possible, an ideal model would be much smaller than this.

3 Goals

What differentiates this study from previous ones is that, in addition to test accuracy, it is focused on generating a sizeable balanced dataset and minimizing the resources needed to train or use the model. Both of these goals are necessary because we believe that the former will help make the latter feasible. These goals will help produce a model that may be helpful in providing accesible testing on a large-scale.

4 Data

4.1 Datasets

This study uses images from 2 main sources. The first is a dataset curated by Praveen Govi [6]. The dataset is composed of images collected from various countries comprising 1576 healthy individuals, 58 patients who have tested positive for SARS-CoV-2 and 4276 patients with respiratory issues other than SARS-CoV-2. The second source is the Covid-Chestxray-Dataset which was curated by researchers from the University of Montreal [4]. In total it has 343 frontal chest X-Ray images including 238 SARS-CoV-2 images. While this leaves us with a good number of images overall, it is still incredibly imbalanced and varied in terms of dimensionality. Image files were incredibly large with some exceeding dimensions of 1300 x 1400 x 3.

4.2 Preprocessing

The goal of preprocessing the images was to minimize the resource overhead on each run without significant losses in model performance. To that end, we decided to run a simple logistic regression model on datasets with varying image dimensions and normalization strategies. The variants we tested were 1-channel grayscale, 3-channel colour, 32 x 32 image size, 64 x 64 image size, 128 x 128 image size, 256 x 256 image size, original image size, normalized and unnormalized. After a grid search on these, we surmised that we could reduce each image to normalized 64 x 64 grayscale (one channel) representations without any significant loss in performance. In fact, it actually helped to reduce overfitting on the training set. Below is a representation of our images after preprocessing:

4.3 Resampling

While the preprocessing will help our model perform efficiently, it does not help us solve the imbalance problem in the dataset. To fix this, we will use data resampling with a few modifications. Let ρ be the number of positive examples in our training set. Let $K = \frac{\rho}{1-\rho}$. K is equivalent to the number of positive examples in our dataset for every negative examples. Because we have less positive examples than negative examples, we balance our dataset by sampling each positive example $\frac{1}{K}$ times. Mathematically resampling is

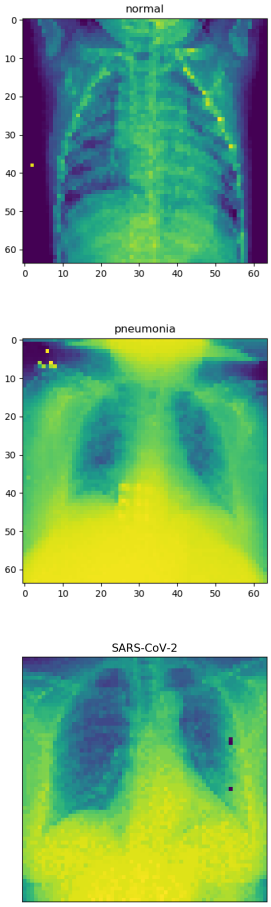

Figure 1: A comparison of our preprocessed images. We have a healthy radiograph (Top-Left), a pneumonia positive radiograph (Top-Right) and a SARS-CoV-2 positive radiograph (Bottom).

equivalent to a reweighting of the positive examples. Each positive example will have a weight $\frac{1}{K}$ times greater than any negative example. In other words, each positive example has a greater influence on the model's performance. While this method is effective in reducing the influence of negative examples, it can cause the model to overfit on these positive examples which is what we see here. With such a high resampling factor, approximately 22, it is plausible that we end up over fitting on the small set of positive examples we have in the dataset.

4.4 Distortion

As a result, we distort the resampled images so that they appear as if from the same distribution yet are different enough so that the model can generalize better.We stochastically apply 2 distortion techniques: flipping and Gaussian noise injection. Flipping involves a horizontal reflection of the image where the axis is at x = 32. Gaussian noise involves generating a 64 x 64 matrix of pixels sampled from a normal distribution and adding this matrix to our existing image. We evaluate a uniformly random Bernoulli variable to see whether or not we flip the image. Similarly, we evaluate another uniformly random variable to see whether or not we apply our Gaussian noise to the image. Additionally, the parameters of the normal distribution used to generate the noise is sampled from a random distribution. The addition of these probabilistic decisions ensures that the chance of any 2 images being the same is negligible.

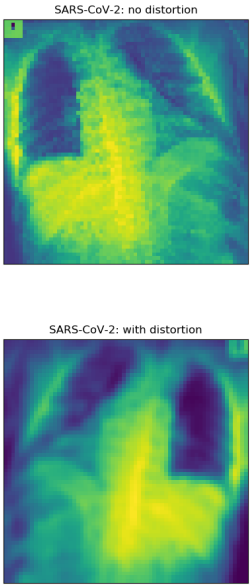

Figure 2: A comparison of a SARS-CoV-2 image before and after applying our distortion

4.5 Augmentation Results

The process of resampling and distorting our SARS-CoV-2 positive examples has the effect of making our dataset more separable. To visualize this, we use t-distributed Stochastic Neighbor Embedding - a process for reducing the dimensions of datapoints non-linearly. It will allow us to visualize our datasets as if they were in 2 dimensions:

In our t-SNE visualization of the unagumented data, we see that there are very few SARS-CoV-2 datapoints and that there is heavy overlap between these points and the non-SARS-CoV-2 datapoints. However, after augmentation we see

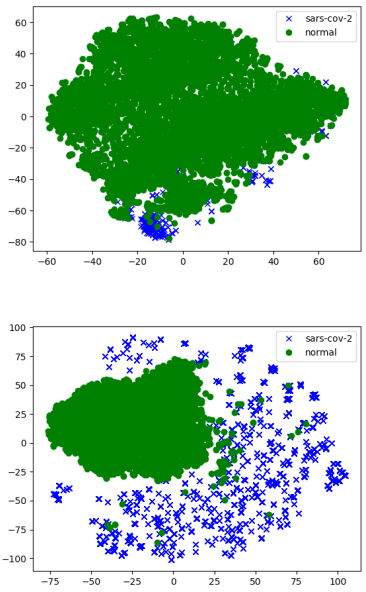

Figure 3: t-SNE visualizations of our trainset before (left) and after (right) augmentation. A green circle represents a radiograph with no SARS-CoV-2, while a blue cross represents a radiograph with SARS-CoV-2.

that there is much clearer separation between the blue points and the green points. This gives us hope that SARS-CoV-2 examples and non-SARS-CoV-2 examples will be separable with high-dimensional models. The only concern is that this process of augmentation may have shifted the distribution of our positive examples so much so that our model fails to generalize to the original distribution of SARS-CoV-2 radiographs. The outcome of this should be apparent in our model's test results. Otherwise, this augmentation appears to have been a success. Overall, after resampling and preprocessing we have a dataset of 11,028 images. The train set has 9654 images including 4433 SARS-CoV-2 positive examples and 5221 SARS-CoV-2 negative examples (healthy and non-healthy). Our validation and test set do not contain any unaugmented images and are still imbalanced. Each has 687, of which 63 are SARS-CoV-2 positive and 624 of which are SARS-CoV-2 negative.

5 Model

5.1 Problem Formulation

Our task is a simple binary classification task. As such, we model our loss as binary cross entropy loss: $-\frac{1}{N}\sum_i^N y^{(i)}\log(h(x^{(i)})) + (1 - y^{(i)})\log(1-h(x^{(i)}))$ where $h(x^{(i)})$ is the prediction of our model.

5.2 Covid-Net-Lite Architecture

To satisfy our requirement of low resource overhead, we design a CNN with only 2 convolutional layers.

Our weights are initialized with Xavier initialization. The first convolutional layer has 8 filters with a kernel size of 4 and 'SAME' padding. After this is a max pool layer with a pool size of 8. The second convolutional layer has 16 filters that each have a kernel size of 2 and 'SAME' padding. The following max pool layer has a pool size of 4. Each convolutional layer is followed by a ReLU activation and a final fully connected layer with sigmoid activation produces our prediction. In total we only have 729 trainable parameters. The model utilizes ADAM optimization.

5.3 Hyperparameters

The only hyperparameter that we tuned on the validation set was the number of epochs to train the model. We tested 10, 50, 100, 150 and 200 epochs. We found that 100 epochs had the best balance between bias and variance.

5.4 Implementation

In keeping with our requirements for a lightweight accesible model, we wanted to ensure that the our network was trainable and executable on standard hardware. Our model was created on top of Keras and Tensorflow and trained on a CPU with 16GB of memory. With this setup, 100 epochs of training took around 20 mins.

6 Results

Below are the results for Covid-Net-Lite trained on the original non-augmented dataset and on the augmented dataset over 100 epochs.

	Non-Augmented	Augmented
Train Accuracy	95.72%	99.96%
Test Accuracy	90.83%	98.69%

7 Analysis

7.1 Non-Augmented vs Augmented

While it is impressive that our augmented dataset achieved higher train and test accuracies, it is even more impressive considering the performance on SARS-CoV-2 positive examples. The

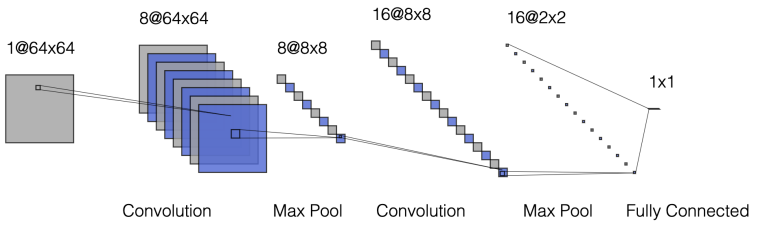

Figure 4: Architecture Diagram

non-augmented scores are high, but the model did not correctly predict any of the positive examples. In fact, with so few positive examples in the non-augmented dataset, the model learned to just predict non-SARS-CoV-2 every single time. Clearly, the data augmentation was able to reweight the positive examples without changing the distribution or overfitting, resulting in not only higher performance but also higher balanced accuracy.

7.2 Covid-Net-Lite vs Other Models

The accuracy scores achieved by Covid-Net-Lite exceeded those of the models produced by Wang et al, Abbas et al and Narin et al.

Model	Test Accuracy
COVID-Net	93.30%
DeTraC	92.50%
VGGNet	83.00%
ResNet50	90.60%
Narin InceptionV3	97.00%
Narin InceptionResNetV2	87.00%
Covid-Net-Lite	98.69%

The only metric where it falls short is that of recall. Covid-Net-Lite only achieves a recall of $\frac{57}{63} = 0.904$, which lags behind COVID-Net, the Narin models and DeTraC.

In terms of resource overhead, Covid-Net-Lite outperforms the others considerably. Per image, the model requires a representation size of 0.15MB. For the sake of comparison, a single image propagated through VGGNet requires 93MB. In terms of parameters, Covid-Net-Lite only requires 729 parameters, while COVID-Net requires 11.75 million, VGG-19, 20.37 million and ResNet50, 24.97 million. Overall, I think Covid-Net-Lite achieves the goals of high accuracy with low resource overhead.

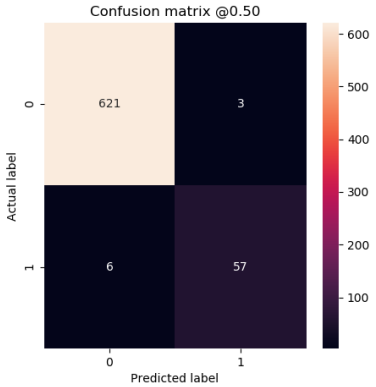

Figure 5: Confusion Matrix of the Covid-Net-Lite predictions

8 Next Steps

While this was a surprising success in terms of computation requirements and accuracy, the low recall is concerning with a disease like Covid-19. A high number of false positives will accelerate the rate of the disease exponentially. Future iterations of Covid-Net-Lite should prioritize the reduction of the false positive rate. One way we can do this is by modifying the loss function so that misclassifed positive examples incur a much greater penalty than normal misclassifications. Additionally, the low recall may be a result of distribution shift caused by the data augmentation. The process of data augmentation used here was not scientifically derived. In the future, a data augmentation process should be engineered so as to minimize the Kullback-Leibler divergence of the unaugmented dataset and the augmented dataset.

9 Github Link

$https://github.com/mayala3/covid\_net$

References

[1] Asmaa Abbas, Mohammed M Abdelsamea, and Mohamed Medhat Gaber. "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network". In: *arXiv preprint arXiv:2003.13815* (2020).

[2] Rahib H Abiyev and Mohammad Khaleel Sallam Ma'aitah. "Deep convolutional neural networks for chest diseases detection". In: *Journal of healthcare engineering* 2018 (2018).

[3] Marios Anthimopoulos et al. "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network". In: *IEEE transactions on medical imaging* 35.5 (2016), pp. 1207–1216.

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

[5] Joseph Paul Cohen, Paul Morrison, and Lan Dao. *COVID-19 Image Data Collection*. 2020. arXiv: 2003.11597 [eess.IV].

[6] Praveen Govi. "CoronaHack -Chest X-Ray-Dataset". In: *arXiv 2003.11597* (2020). URL: https://www.kaggle.com/praveengovi/coronahack-chest-xraydataset/.

[7] Melina Hosseiny et al. "Radiology perspective of coronavirus disease 2019 (COVID-19): lessons from severe acute respiratory syndrome and Middle East respiratory syndrome". In: *American Journal of Roentgenology* (2020), pp. 1–5.

[8] Andrew Kesselman et al. "2015 RAD-AID conference on international radiology for developing countries: the evolving global radiology landscape". In: *Journal of the American College of Radiology* 13.9 (2016), pp. 1139–1144.

[9] Qing Li et al. "Medical image classification with convolutional neural network". In: *2014 13th international conference on control automation robotics & vision (ICARCV)*. IEEE. 2014, pp. 844–848.

[10] Ali Narin, Ceren Kaya, and Ziynet Pamuk. "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks". In: *arXiv preprint arXiv:2003.10849* (2020).

[11] World Health Organization et al. *Laboratory testing for coronavirus disease 2019 (COVID-19) in suspected human cases: interim guidance, 2 March 2020*. Tech. rep. World Health Organization, 2020.

[12] Pranav Rajpurkar et al. "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning". In: *arXiv preprint arXiv:1711.05225* (2017).

[13] Suhail Raoof et al. "Interpretation of plain chest roentgenogram". In: *Chest* 141.2 (2012), pp. 545–558.

[14] Max Roser, Hannah Ritchie, and Esteban Ortiz-Ospina. "Coronavirus Disease (COVID-19)–Statistics and Research". In: *Our World in data* (2020).

[15] Wei Shen et al. "Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification". In: *Pattern Recognition* 61 (2017), pp. 663–673.

[16] Linda Wang and Alexander Wong. "COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images". In: *arXiv preprint arXiv:2003.09871* (2020).

[17] *WHO Coronavirus Disease(COVID-19) Dashboard*. 2020. URL: https://covid19.who.int/?gclid=Cj0KCQjww_f2BRC-ARIsAP3zarEDu9iTOeUJ_EFDaGIAHBciNtUtET8yU6lWOJKQ1qi8GVINOzz2hpwaAuWDEALw_wcB.

[18] Ho Yuen Frank Wong et al. "Frequency and distribution of chest radiographic findings in COVID-19 positive patients". In: *Radiology* (2020), p. 201160.

[19] Xiaowei Xu et al. "Deep learning system to screen coronavirus disease 2019 pneumonia". In: *arXiv preprint arXiv:2002.09334* (2020).

[20] Li Yan et al. "Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan". In: *medRxiv* (2020).

[21] Jianpeng Zhang et al. "Covid-19 screening on chest x-ray images using deep learning based anomaly detection". In: *arXiv preprint arXiv:2003.12338* (2020).

[22] Zi Yue Zu et al. "Coronavirus disease 2019 (COVID-19): a perspective from China". In: *Radiology* (2020), p. 200490.