Pneumonia Detection with SEResNets CS7180: Final Project

Erica Shephard
Computer Science
Northeastern University
Boston, USA
shepherd.e@northeastern.edu

Nicholas Wen
Robotics, Electrical and Computer Engineering
Northeastern University
Boston, US
wen.n@northeastern.edu

Sumegha Singhania
Robotics, Electrical and Computer Engineering
Northeastern University
Boston, USA
singhania.s@northeastern.edu

***Abstract*—Pneumonia is a vicious disease that kills almost 50,000 people in the United States alone. A chest X-ray (CXR) is the most common way to diagnose pneumonia, however identifying the disease can be very difficult due to the high visual similarity between pathologies. While work has been done to leverage convectional neural network models for pneumonia detection, some initial findings are uncorroborated. In this paper, we attempt to examine some of these claims and see if their results are reasonable within a time and processing power constraint.**

I. Introduction

Over one million people are hospitalized with pneumonia every year. Currently, the best way to diagnose pneumonia is via a chest x-ray, however this task often requires special training as a radiologist. In recent years, researchers have been utilizing neural networks to identify various illnesses, including pneumonia, in lieu of a medical expert. In this work, we attempt to create one such of these papers, Attention-Guided Convolutional Neural Network for Detecting Pneumonia on Chest X-Rays, and attempt to verify their findings and expand their choice of network and parameters to see if anything might be improved.

II. Related Works

A. Chest X-Rays and Pneumonia Detection

Chest X-rays(CXR) are the most common way to detect pneumonia. Usually identification is done via regions of higher opacity within the CXR scan [1]. This method however is not consistently reliable, as factors such as patient position, degree of inspiration (positioning of ribs and diaphragm), and prior medical conditions/procedures can all affect the diagnosis and analysis of pneumonia. In addition, symptoms of pneumonia often overlap with other maladies which can easily lead to clinicians struggling to properly identify pneumonia within a CXR and giving misdiagnoses that can result in further complications or loss of life [2] [3].

B. Deep Learning and Pneumonia

In recent years, the development of deep learning has been moving into medical spaces. Convolutional Neural Networks(CNN) specifically has been used on CXRs to help with the identification of various diseases. In Rajpurkar, et al. [4] they develop an 121 layer end-to-end network that outputs the probabilities of different pathologies. This paper correctly identified pneumonia 78.02% accurately. While this number is impressive, this percentage is not within a reasonable range for use in medical technologies.

More recently in Sedai et al [5], we see neural networks used to identify small pathologies with bounding boxes in CXR images . In this experiment, all training examples are known to have a pathology in them, but the specific illness and its location are unknown to the network. In order to identify smaller masses or signs of illness utilizing an attention map to focus the network. Notably, while this paper begins to look at smaller areas, the percentage of positive pneumonia detection is only 40% while the false positive rate is 80%, which is worse than Rajpurkar, et al. [4] and is still not suitable for more widespread usage.

1) Squeeze and Excitation: In 2018,Hu et al. [6] introduced Squeeze and Excitation blocks(SEblocks) which demonstrated huge improvement in attention gating for neural networks. The main concept is to use the global characteristics of channels to determine important features and weights. The SEblock works on each channel of the network. First, a global pooling layer is used to reduce each channel down to a single value (squeezing) boiling down to a 1x1xC vector. After this, the vector is fed through 2 dense layers, the first with a ReLU activation layer and the second with a sigmoid activation layer as a smooth gating function. Finally these values are used as the weights for the original channels via multiplication. The SEblocks act as an attention gating mechanism without adding much overhead to a network.

The usage of SEblocks for Pneumonia detection becomes a main point in Li et al [7], where an improved squeeze and excitation network (SENet) is used to suppress useless features. This paper uses the SE blocks similarly to Sedai et al's use of attention maps. These blocks help prioritize the locations of potential pneumonia locations in our image. These SEblocks are used in between initial convolution layers, and the result after each step is sent as a residual connection fused with a 1x1 convolution to be utilized in future up-sampling layers with different hierarchical layers. These upsampling layers end with the feature maps integrated to a single channel and a sigmoid activation. Finally each output pixel represents probability of pneumonia detection, which allows a bounding box to be found after a thresholding function is used.

This paper, while ostensibly the next step in pneumonia, has some confounding elements, including statistical calculations being inconsistent with equations as well as several elements of the experiment left unusually vague. Due to this we intend to focus on this paper, recreating its key elements to either confirm its finding or explore shortcomings with the initial findings.

III. Data

We used training and testing data from the Radiological Society of North America (RSNA) pneumonia challenge [8]. The training data consists of 26684 chest x-ray images and the testing data consists of 3000 images. These images are 1024x1024 in size and dicom format. In addition, a text file with patientIds, positive or negative pneumonia evidence and bounding boxes for detected pneumonia is given. After looking at the data carefully, we realized that there were no ground truth values provided for the testing dataset. Hence we made our own datasets: a smaller dataset consisting of 3724 training images and 300 testing images and a bigger dataset consisting of 29469 training images and 2631 testing images.

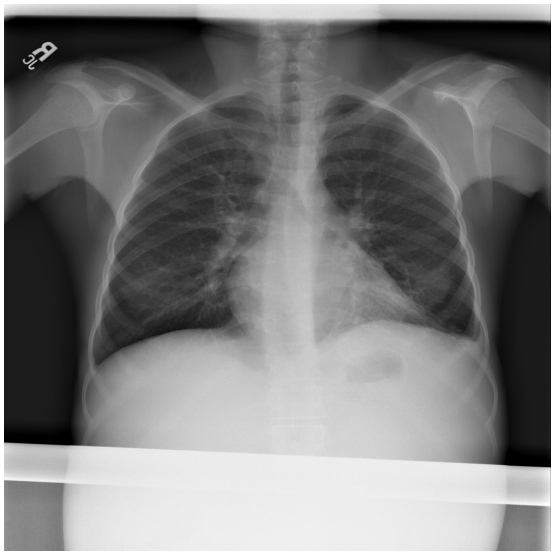

Fig. 1. An example of an image from the RSNA pneumonia challenge dataset [8]

IV. Methods

A. Data Pre-Processing

There is a huge variation in CXR images, depending on the individual and other conditions, which might make it difficult for the network to distinguish between pneumonia and other unexpected noise. To counter this, as suggested by the paper, we erased the pneumonia pixels in the CXR images, by replacing them with the average value of the whole image, to give the network a better understanding of non-pneumonia characteristics of a CXR. In addition to this, we scale down images to 128x128 to fit well with our computational limitations and apply data augmentation in the form of rotation, translation, flipping and scaling to expand our dataset.

The smaller dataset only contains scaled down images and the larger dataset contains images with erased pneumonia pixels and augmented images. This is done to measure the impact of the pre-processing mentioned in the paper.

B. SEResNets

Inspired by Li et al., we implemented SEResNet-34 using cross-entropy loss and the Stochastic Gradient Descent (SGD) algorithm for the optimizer with a learning rate of 0.001, weight decay of 0.0001, and momentum of 0.9. For experimentation purposes, we also implemented SEResNet-18, SEResNet-50, and SEResNet-101 to test the effects of using a deeper network. The squeeze-and-excitation modules were modeled according to Figure2 [6].

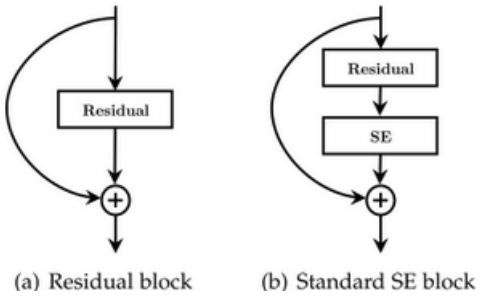

Fig. 2. SE Resnet Module

These modules were then added to Pytorch implementations of the residual basic block and bottleneck architecture in Standard SE format as shown in the Figure 3.

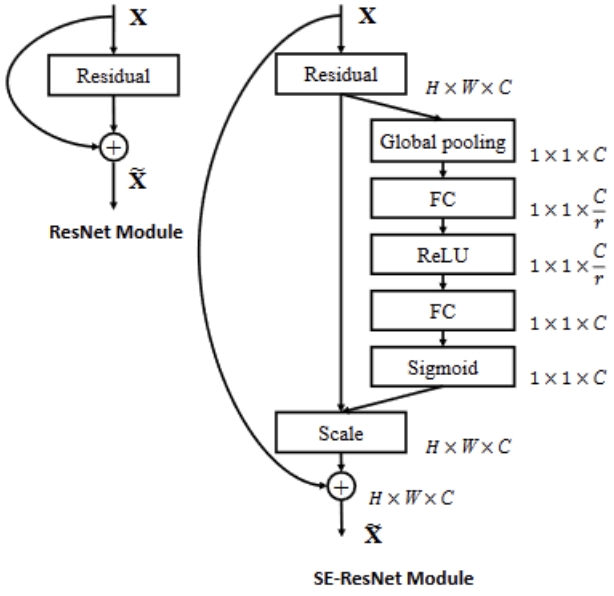

Fig. 3. Taken from [9]

C. Data

SEResNet-18 and SEResNet-34 used the modified Pytorch basic blocks while SEResNet-50 and SEResNet-101 used the modified Pytorch bottleneck architectures. These were then used in the Pytorch implementation of the respective ResNet. All four networks were trained with cross-entropy loss and a SGD optimizer. These models were also trained at different learning rates of 0.001, 0.005, and 0.0001, weight decays of 0.0001, 0.00001, batch sizes of 1 and 64, and finally a momentum of 0.9. Each network was trained for 10 or 15 epochs on the smaller and larger pre-processed datasets of the RSNA data.

D. Prediction and Accuracy

The networks were trained to predict an output of 0 or 1: 0 being no pneumonia present in the XR image, and 1 being pneumonia present in the XR image. We used an accuracy function of $\frac{TP+TN}{TP+FP+TN+FN}$, with TP being true positives, TN as true negatives, FP as false positives, and FN as false negatives.

V. Experiments and Results

We tested 4 network models with our test images: Resnet18, Resnet34,Resnet50, and Resnet 101. For each of these, the number following describes the number of residual blocks there are in each network. We found that the optimal learning rate we found varied from network to network. All networks were trained with 128x128 pixel images. We trained all the networks on the RSNA dataset, however we created a larger dataset by artificially removing pneumonia clusters from some examples. Unfortunately since some of the images have more than 1 instance, our modification did not work correctly. Here we show the results with the larger dataset with flawed images, followed by several networks trained on a smaller (unmodified) dataset.

When testing with a larger dataset, Resnet18 was trained with a learning rate of 0.0001, and all other models were trained with a learning rate of 0.0005. Due to time constraints, we trained the two smaller networks (Resnet18 and Resnet34) for 15 epochs and the two larger networks (Resnet50 and Resnet101) for 10 epochs each4.

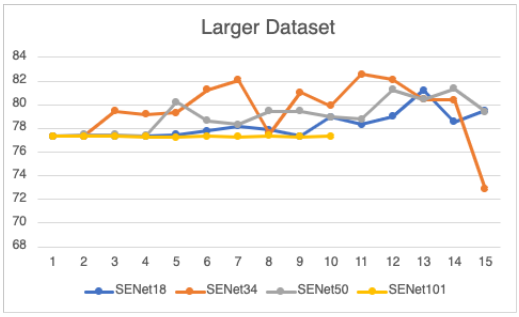

Fig. 4. The accuracy graph taken when using the larger(flawed) dataset

On the smaller datasets, we trained several different networks for each model size, varying the learning rate and batch size. Each graph can be seen in figures 5.1-5.4 with their accuracy over time as well as the network parameters.

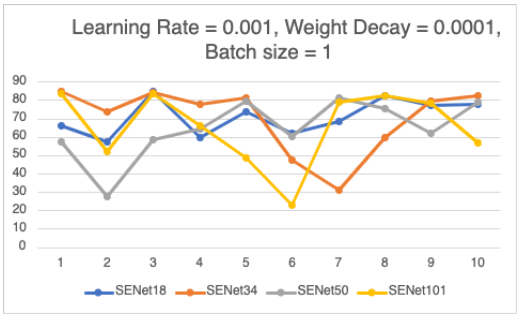

Fig. 5.1. Learning Rate:0.0001, Batch Size:1, Weight Decay 0.0001

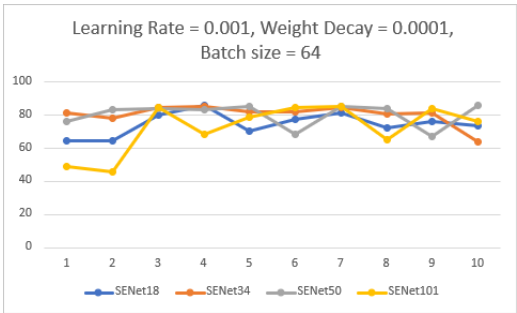

Fig. 5.2. Learning Rate:0.0001, Batch Size:64, Weight Decay 0.0001

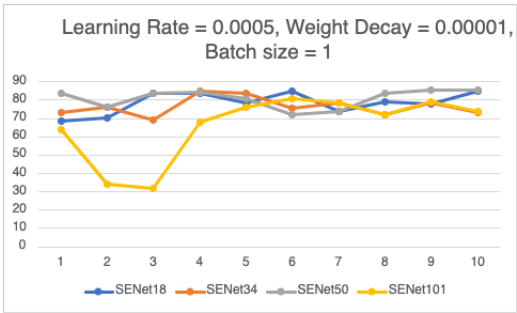

Fig. 5.3. Learning Rate:0.0005, Batch Size:1, Weight Decay 0.00001

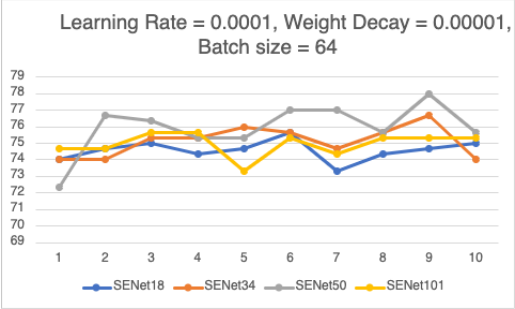

Fig. 5.4. Learning Rate:0.0001, Batch Size:64, Weight Decay 0.0001

It should be noted that the first epoch shows a high accuracy, but it is actually a false indication of accuracy since the network predicts all negatives in the first run and there are more negative CXR images than positives. For both the datasets, we see that the shallower networks perform better than the deeper ones. This makes sense because deeper networks generally take longer to train. Our hypothesis is that given more training time, the deeper networks could perform much better, especially with images of higher dimensions. Additionally, if measuring the impact of having additional erased pneumonia scans as normal chest x-rays, it does not seem to have a huge impact on accuracy. But since, there are some mistakes in our erased pneumonia images, we cannot say that this hypothesis is accurate.

VI. Discussion and Summary

While we did have some significant insights from this experiment, some aspects of this project left more room for exploration into this topic. We hope that researchers in the future attempt to recreate the original paper with full 1024x1024 images and expand our predictions to include bounding boxes. In addition, we hope to one day have the processing power to see the effect on the networks with more epochs of training. Ideally someone will also recreate the modified dataset with correct removal of Pneumonia examples.

Overall, these results confirm several known observations about neural networks, but despite the performance increase we saw from the literature we examined, this technology is still not reliable enough for use in widespread medical spaces. We observed that a higher learning rate caused larger leaps and achievements in learning, but had the potential to become unstable. On the other hand, a smaller learning rate showed slow improvement that tended to be stable. A larger batch rate also seemed to help stabilize the accuracy.

References

[1] Franquet, Tomás. ”Imaging of community-acquired pneumonia.” Journal of thoracic imaging 33.5 (2018): 282-294.

[2] Kelly, Barry. ”The Chest Radiograph.” Ulster Medical Journal” 81.3(2012):143-8.

[3] Wang, Xiaosong, et al. ”Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases.” Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on. IEEE, 2017

[4] Rajpurkar, Pranav, et al. ”Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning.” arXiv preprint arXiv:1711.05225 (2017).

[5] Sedai, Suman et al. “Deep multiscale convolutional feature learning for weakly supervised localization of chest pathologies in X-ray images.” MLMI@MICCAI (2018).

[6] Hu, Jie, Li Shen, and Gang Sun. ”Squeeze-and-excitation networks.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[7] B. Li, G. Kang, K. Cheng and N. Zhang, ”Attention-Guided Convolutional Neural Network for Detecting Pneumonia on Chest X-Rays,” 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 4851-4854, doi: 10.1109/EMBC.2019.8857277.

[8] “RSNA Pneumonia Detection Challenge,” Kaggle. [Online]. Available: https://www.kaggle.com/competitions/rsna-pneumonia-detection-challenge/data. [Accessed: 20-Nov-2022].

[9] amaarora, “Squeeze and excitation networks explained with pytorch implementation,” Committed towards better future, 24-Jul-2020. [Online]. Available: https://amaarora.github.io/2020/07/24/SeNet.html. [Accessed: 15-Dec-2022].