Pneumonia Chest X-Ray Classification Using Transfer Learning

DSU DAKOTA STATE

Martinson Ofori, Rajesh Godasu, Evan Miles

DL Class Poster

JULY 27, 2019

Abstract

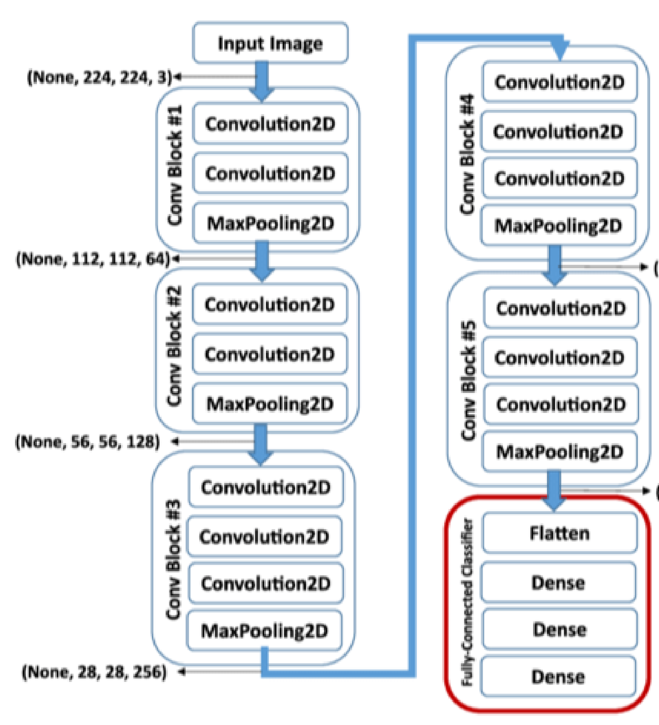
Computer Aided Diagnoses (CAD) is important to modern medical ailment detection. A key part of CAD is Medical Image Classification (MIC). The aim of MIC is to classify various medical images into categories that aid doctors in disease diagnosis. Recent advancements in Deep Learning (DL) has allowed researchers to apply DL techniques such as Transfer Learning (TL) to MIC. Overfitting is still a major problem while training such models. We attempt to address this problem with a novel architecture that makes use of pooling and dropouts to control overfitting using the publicly available pneumonia dataset.

Background

Simonyan & Zisserman (2014) introduced **VGG-16,** the core of our proposed model. The VGG-16 uses **16 weight layers**, three fully connected layers, and a final **softmax** layer for image classification. The hidden layers utilize **ReLU** activation. Using **smaller (3x3) convolution filters** than previous models, VGG-16 boosted the accuracy of **ConvNets** over its predecessor, **AlexNet**.

For the purposes of this research poster, we use the **pneumonia image dataset** which consists of **5,863 X-ray images**, split in to two categories, Pneumonia and Normal.

The dataset was provided by the Guangzhou Women and Children's Medical Center and is of pediatric patients between one and five years old. Three expert physicians graded each image in to the two categories.



Literature Review

Solutions for **medical image classification (MIC)** problems can be achieved efficiently and accurately using **Deep Learning (DL)** techniques (Sahiner et al., 2019). Robust growth in DL models and powerful GPUs in recent history has aided these advancements.

Transfer Learning (TL) is the current trend in achieving MIC solutions (Xu et al., 2018; Ausawalaithong et al., 2018; Gessert et al., 2018). TL allows researchers to benefit from **pre-existing models** by borrowing their **architecture,** and often, **pre-trained weights**.

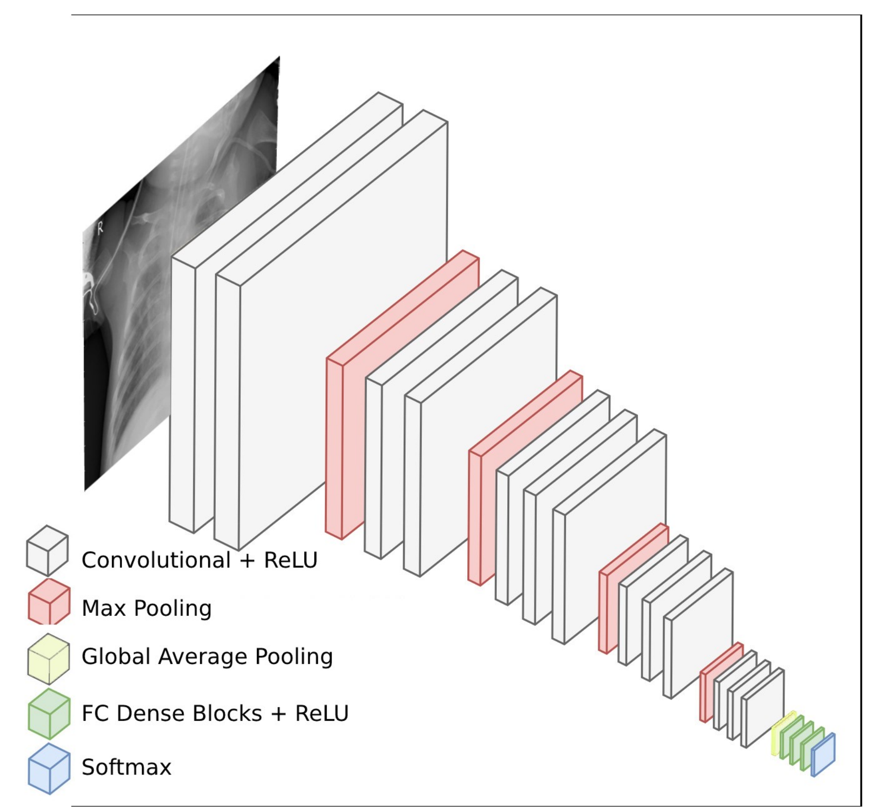
Such classic architectures include **Resnet** (Zhang et al., 2016), **Densenet** (Huang et al, 2017), **Googlenet** (Szegedy et al., 2015) and **VGG-16** (Simonyan & Zisserman, 2014), and have been widely implemented through TL on various MIC problems (Guan et al., 2018; Tan et al., 2018; Majtner et al. 2018).

In recent history, researchers have successfully used VGG-16 for various MIC problems such as **detection of Alzheimer's disease** (Hon & Kan, 2017), classification of **thyroid nodules in ultra sound images** (Liu et al., 2017)**,** and three class **brain tumor classification** problem (Wong et al., 2018).

Chest X-ray 14 dataset has been used in various DL, and TL, research for **pneumonia detection** (Wang et al, 2017). Previous works on X-ray datasets include **ChexNet** with pneumonia identification as a **binary classification** task (Rajpurkar et al. 2017). Zongyuan et. al (2018) used same to perform **multi-class classification** and addressed the problem of **multi-labels** and **data imbalances**.

Methodology

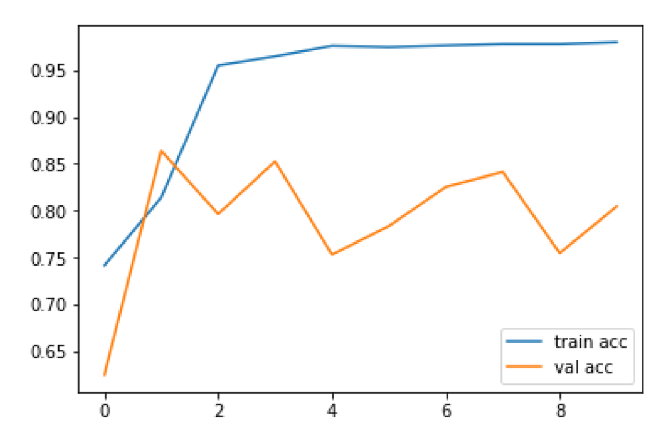
A novel architecture embedding three dense blocks on top of VGG-16 model and freezing the top 15 layers is used. To reduce overfitting Global Average Pooling (GAP) layer is introduced before the dense blocks. GAP layers perform dimensionality reduction by averaging tensor dimensions to 1×1×d. We also make use of dropouts of 50%, 40% and 30% after each FC block.



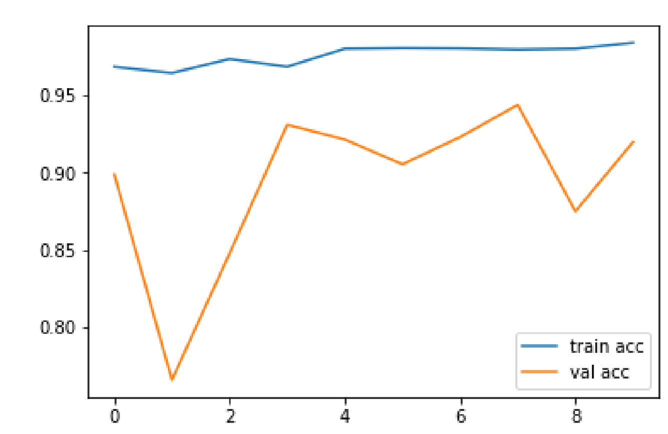
LR step decay, early stopping and model checkpoint saving is used to ensure the best weights are retained during training.

Results

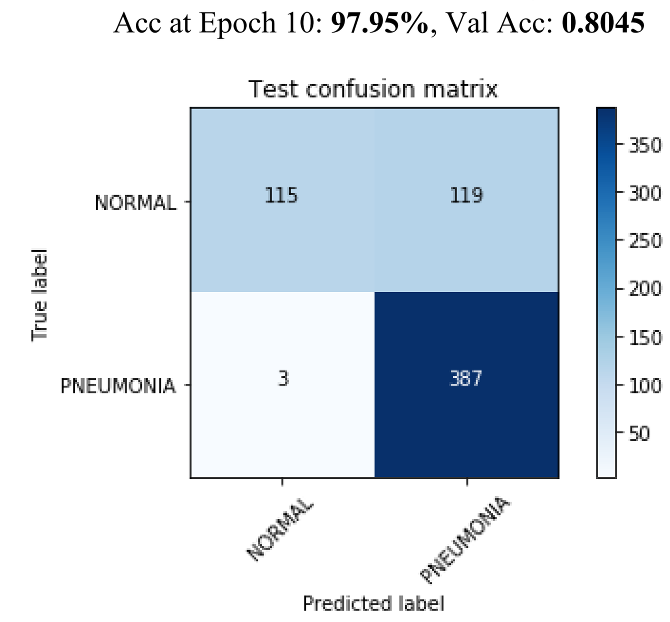
We trained our model for 10 epochs on Google Colab by making use of the free GPU: 1xTesla K80 , having 2496 CUDA cores, compute 3.7, and 12GB GDDR5 VRAM.



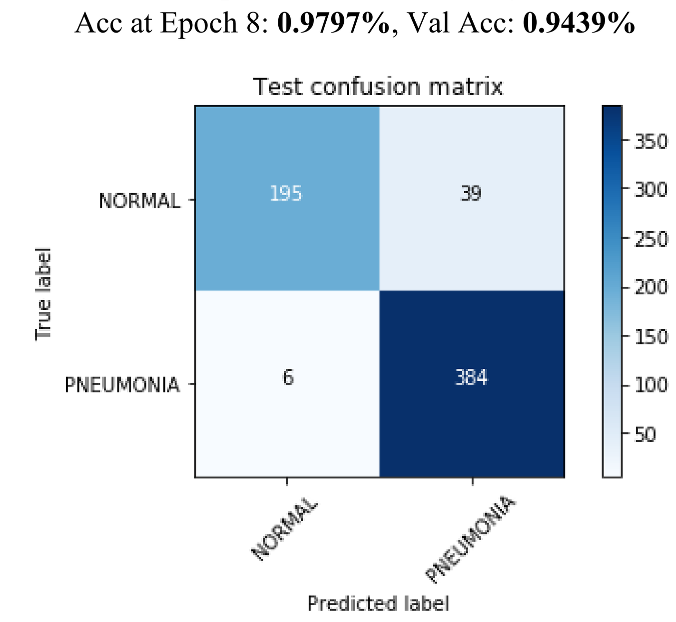
Pretrained VGG 16 Accuracy
Acc at Epoch 10: **97.95%**, Val Acc: **0.8045**



Proposed Model Accuracy
Acc at Epoch 8: **0.9797%**, Val Acc: **0.9439%**



Pretrained VGG 16 Confusion Matrix
Recall: **99%**, Precision: **76%**



Proposed Model Confusion Matrix
Recall: **98%**, Precision: **91%**

From the result, although VGG-16 has 1% better recall, the proposed model has much better precision of 91% compared to 76% of the VGG-16 model.

Conclusions

VGG-16 famously achieved 92.7% on ImageNet dataset but quickly overfits on medical X-ray images.

Taking inspiration from VGG-19, a variation of VGG-16 with additional 3 FC layers, we make use of global average pooling and dropouts after each of the 3 FC layers.

We control overfitting leading to higher performance on chest X-ray image classification for pneumonia dataset.

Future work: Retrain the proposed network for longer epochs. Apply the model to the larger Chest X-ray 14 dataset. Compare performance of other well known architectures to the proposed model.