Spring 4-2020

Explainable Deep Learning for Medical Image Analysis

Brennan Rhoadarmer
University of Nebraska-Lincoln, brhoadarmer@huskers.unl.edu

Explainable Deep Learning for Medical Image Analysis

Brennan Rhoadarmer

Deep Learning

Deep learning is a branch of machine learning in which neural networks are based on learning data representation. These neural networks are sequential layers of nodes, with each node taking input from previous layers and preforming a few basic operations on that data and then outputting to the next layer of nodes. These neural networks, or models, then have large amounts of data put through the layers, checking the final output against the known output, and updating the model's operations to try and improve the model's output. A visualization of a model can be seen in figure 1.

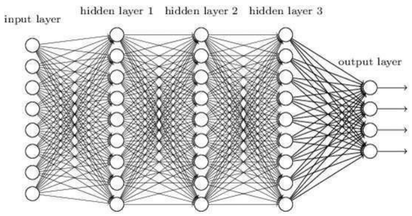

Figure 1: A visualization of a small neural network. Nielsen, M.(2018). Retreived from http://neuralnetworksanddeeplearning.com/chap6.html

The Problem

One of the problems with using deep learning in a medical field, where a diagnosis can be the difference between life and death, a correct treatment or a wrong one, is that deep learning models are often black boxes, as demonstrated in figure 2. That is, we do not know why they produce the answers that they do. This is a problem if a doctor wants to know why a model is suggesting a certain diagnosis. So we need to find a way to make a deep learning model that is both accurate for medical images and can explain why it is giving the result that it is.

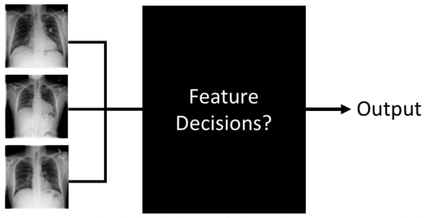

Figure 2: The black box of deep learning models

GradCAM

A recent jump forward in the idea of making transparent the black box that surrounds deep learning models is called GradCAM[3], which stands for Gradient-weighted Class Activation Mapping. GradCAM uses trace backs through the layers of the model to go back through the layers and determine which pieces of the input influenced the decision for that classification the most.

After GradCAM has determined which part of the input most influences the output, it can then create a "heat map" so that users can visually inspect how the model is determining the classification. See figure 4 for examples of these "heat maps".

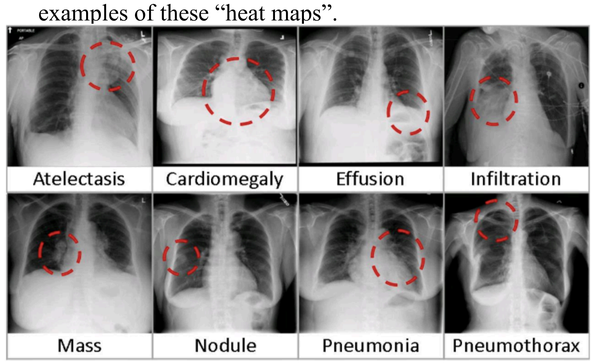

Figure 3: Examples of medical images that we wanted the model to be able to categorize, without the red circles, which identify what doctors would refer to for diagnosis (Taken from the NIH Chest X-ray Dataset)[1]

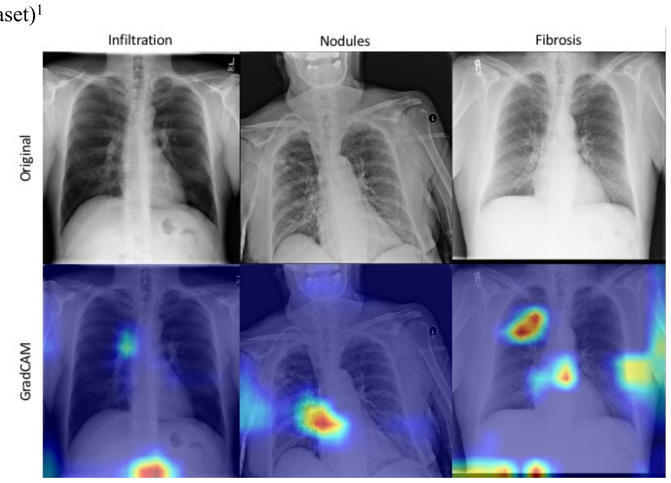

Figure 4: Examples of images produced from GradCAM along side their originals. Originals from the NIH image set. GradCAM generated by me

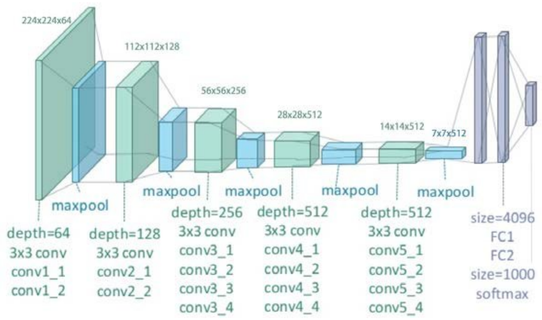

Figure 5: A visual representation of the VGG-19 architecture. Retrieved from paper: Breast Cancer Screening Using Convolutional Neural Network and Follow-up Digital Mammography[7]

Model Training

Before being able to implement GradCAM on a model, we needed a model that had been trained on our data and was of the right shape, which for our implementation was VGG-19, seen in figure 5, a well known and powerful deep learning model architecture.

We trained a model using procedures from a similar experiment that was also attempting to classify medical images by a Stanford research team[4]. Using those methods and simply changing the base model, we trained our own neural network, which we then ran GradCAM over.

Current Results

The accuracy of our current model is not very good right now, with an average AUC (area under the curve) of 0.2 for our multi-class data. We want something much closer to 1, so the model is not very good at classifying our data. However, we were still able to run GradCAM over several inputs for our model and found why it was classifying those inputs the way it was. A few of those can be seen in figure 4, but a majority of the GradCAM results show that the model does not focus on items that it should and instead pays more attention to the patient's positioning in the image rather than anything that might lead to a diagnosis.

Therefore, moving forward, we are going to try and use a pretrained model[5] that is top rated on CheXpert data, which is a large data set of chest X-rays, very similar to the NIH data we used in our GradCAM implementation. This model uses Densenet[6] as a basis, so we will have to modify our implementation of GradCAM to work with Densenet instead of VGG19.

References

Medical image data from:

1.NIH Chest X-Ray Dataset from https://www.kaggle.com/nih-chest-xrays/data/home

2.CheXpert Dataset: https://stanfordmlgroup.github.io/competitions/chexpert/

3.GradCAM paper:https://arxiv.org/pdf/1610.02391.pdf

4.Stanford paper: https://arxiv.org/pdf/1610.02391.pdf

5.Pre-trained Densenet: https://github.com/jfhealthcare/Chexpert

6. Densenet: https://arxiv.org/pdf/1608.06993v3.pdf

7. Zheng, Yufeng & Yang, Clifford & Merkulov, Aleksey. (2018). Breast cancer screening using convolutional neural network and follow-up digital mammography. 4. 10.1117/12.2304564.

UCARE

University of Nebraska Lincoln