Project Document: Pneumonia Classification on Radiology Using TensorFlow

Members: Grace Sigalla, Gowtham Vaddanam, Sreerekha Rajendran

INFO 607 Spring 2022

Table of Contents

1. Goals
2. Context
3. Scope
4. Project Plan
5. Software and Hardware
6. Experimentation Plan
7. Deliverables
8. References
9. **Goals**

The key goal of our project is to fully understand and master TensorFlow, which is a platform that enables easy building and execution of machine learning applications. Our project will be focused on Pneumonia classification on X-rays, and we will utilize TensorFlow to create our machine learning models that will detect patients with Pneumonia using the dataset we attained from Kaggle. Finally, we will combine our knowledge of TensorFlow with our understanding of MySQL to compare performance between the two.

1. **Context**

TensorFlow is a Google Open-Source Machine Learning Framework that may be used to create dataflows for a variety of applications. The graph's nodes represent mathematical processes, while the graph's edges represent the multi-dimensional data arrays that are exchanged between them. Tensors are just multidimensional arrays, which are a 2-dimensional table's extension to data with a higher dimension. TensorFlow has several properties that make it suitable for Deep Learning, and its core open-source library aids in the development and training of machine learning models. It enables easy model building, robust ML, and powerful experimentation. Ragged Tensors, TensorFlow Probability, Tensor2Tensor, and BERT are just a few of the sophisticated add-on libraries and models available in TensorFlow.

It is a large-scale machine learning system that works in a variety of contexts. Dataflow graphs are used by TensorFlow to describe computation, shared state, and operations that change that state. It transfers the nodes of a dataflow graph over several computers in a cluster, as well as various processing devices within a single system, such as multicore CPUs, general-purpose GPUs, and custom-designed ASICs known as Tensor Processing Units (TPUs). TensorFlow's architecture affords application developers more flexibility. TensorFlow allows developers to experiment with new optimizations and training techniques. TensorFlow covers a wide range of applications, with a focus on deep neural network training and inference.

In this project, we are using TensorFlow for image classification. Image classification is something a human can do in seconds, but it has been a difficult one for machines until recent advances in Artificial Intelligence and Deep Learning. For example, self-driving cars can recognize things and take necessary actions in real time, thanks in large part to TensorFlow Image Classification.

Healthcare and biomedical research can benefit from these advances the most. Healthcare is a data-rich industry, and there is a vast quantity of untapped data that can be used to drive innovation. We can use Image Classification to predict if a person has Pneumonia or not. All we must do is examine the procedures and instruments used to categorize X-Ray pictures into one of two categories: normal or pneumonia. We can do this using TensorFlow. On the TensorFlow/Keras platform, we'll utilize a Convolutional Neural Network to do this. The project will be done using Kaggle’s dataset that contains chest x-ray images.

1. **Scope**

IN-SCOPE:

To classify the radiological images as ‘pneumonia’ and ‘others’, this project involves with Convolutional Neural Networks (CNN) on TensorFlow platform and keras – which is a high-level API of TensorFlow; for decreasing the number of user actions required for common use cases. The dataset consists of 150\*150-pixel(s) units per image, which is compressed in a zip file to make scaling better. Building up the customized model based on the state-of-the-art architecture called VGG-16 reference.



All types of convolutional layers are used including max pooling and fully connected. Drop-out and learning rate reduction methods are used. Augmentation to be done to avoid high variance in the learning curve i.e., overfitting and any imbalances in the training data. And further packages in python like seaborn for visualizing the predictors, how well the model is performing.

OUT-SCOPE:

For our project, the dataset used is considerably huge but static in generalizing the task/problem statement. We are not exploring the complexities involved in TensorFlow for hyperparameter tuning like regularization and optimization tasks like RMS Prop. No pre-trained model is used via transfer learning.

**IV.** **Project Plan**

The first step of our project is to install and learn TensorFlow. There are a wide range of resources available on both TensorFlow’s website and YouTube.

Once we have learned the basics about TensorFlow from the two learning avenues we mentioned, we will begin to build our machine learning models using the publicly available Pneumonia dataset referenced above in Section 2 of context.

1. **Software and Hardware**

The software involved in this project is Python libraries, which can be run on either Apple iOS or Windows operating system using an Internet connection. TensorFlow/keras packages will be Jupyter notebook command system and baseline in computing the big data. No specific hardware is required.

**VI.** **Experimentation Plan**

Datasets: Upon completion of the preparation phase, and after gaining familiarity with TensorFlow, we will begin the process of building machine learning models to predict which patients based on the data have pneumonia. Using the dataset obtained from Kaggle <https://www.kaggle.com/code/amyjang/tensorflow-pneumonia-classification-on-x-rays/data> which contains x-ray images that we will be deducting if they have pneumonia or not, by building an image classification model in order to predict the results.

Techniques: The main technique we will employ in this project is TensorFlow, which will enable us to train a machine learning model for predicting the results. We will be writing our code using python programming language.

Implementation: Once we load the data, we will begin training our model using TensorFlow right away as the dataset we are using is already cleaned and processed.

**VII.** **Deliverables**

The final project will contain deliverables which will illustrate and explain in detail what we understood from the entire process of working with TensorFlow. A properly optimized model is the result of trial and error, as well as making the right predictions about hyperparameters and epochs to run. One of the most basic challenges in the overall effectiveness of the model is setting up the training data. Through our project we will try to make sure the above-mentioned challenges are overcome, and we present good results.

We will provide the references to the dataset used in the project, the source code, and a clear and precise documentation of the whole process and working of the models. Once the project is complete, we will demonstrate the implementation and the results through a presentation.

As outlined in the Project Description handout, we will submit a zip file that contains the following:

* A PowerPoint file used to present to the class.
* A word file that explains the whole project including installation procedure, code, and output, etc.
* The data set used in the project.

**VIII. References**

*Tensorflow*. TensorFlow. (n.d.). Retrieved April 30, 2022, from <http://www.tensorflow.org/>

TensorFlow. 2022. Introduction to TensorFlow. [online] Available at: <https://www.tensorflow.org/learn> [Accessed 1 May 2022]

En.wikipedia.org. 2022. TensorFlow - Wikipedia. [online] Available at: <https://en.wikipedia.org/wiki/TensorFlow> [Accessed 1 May 2022].

Sharma, V. and Tensorflow, I., 2022. *Image Classification Using Tensorflow | Pluralsight*. [online] Pluralsight.com. Available at: <https://www.pluralsight.com/guides/image-classification-using-tensorflow> [Accessed 1 May 2022].

Tensorflow is a really powerful tool. I would suggest that you don’t focus entirely on getting goo accuracy, but rather also focus on exploring different ways you can use Tensorflow and implement different models and do some comparisons. You’re welcome to adjust scope and goal if needed. I look forward to seeing your final product.

Grade: 10/10