Create and deploy Convolutional Neural networks using Keras for Image data.

Table of contents

Acquiring data

The following list is different data sources that I accumulated over the period of the coronavirus spreading. Currently, this is the best I can get from the internet because the data being collected in many countries and hospitals is classified information. Even if it was not classified information, you still need consent from each patient.

- GitHub: COVID-19 Chest X-Ray and CT images. This data set is continually updated as more cases appear.
- Kaggle 1: Chest X-Ray images with both positive and negative cases of pneumonia.
- Kaggle 2: Chest X-Ray images with positive and negative cases of pneumonia.

Problem with mixed data sources

All of the data available on the internet has not been subjected to the same preprocessing, and all of the images are, as explained below, clearly different in the amount of black bars. Most of the positive COVID-19 data has the entire X-ray take up most of the screen, with little to no black bars on the sides (except for a few). However, the data set with negative COVID-19 cases has mostly black bars on the side of each image.

19CST302&Neural Networks and Deep Learning   **S.VASUKI**

This becomes a problem because the model might later learn that it just needs to look at the black bars on the side to know whether a new sample is a positive or negative case of COVID-19.

After manual inspection of the data set, it becomes apparent that almost all of the negative cases have these black bars, while approximately 10-20% of the positive cases have black bars. Wouldn't it be quite a coincidence if the accuracy of predicting positive or negative later turns out to be around the 80-90% mark?

In the previous image, the left side shows an image with a negative COVID-19 case (black side and top bars). The image on the right shows the same image, but is cropped toward the center and scaled up to match sizes.

19CST302&Neural Networks and Deep Learning   **S.VASUKI**