Machine Learning Engineer Nanodegree Program

Capstone Proposal

STEM NGUYEN

Blindness Detection (Kaggle Competition)

Domain Background

Imagine being able to detect blindness before it happened.

Millions of people suffer from diabetic retinopathy, the leading cause of blindness among working aged adults. Aravind Eye Hospital in India hopes to detect and prevent this disease among people living in rural areas where medical screening is difficult to conduct. Successful entries in this competition will improve the hospital's ability to identify potential patients. Further, the solutions will be spread to other Ophthalmologists through the 4th Asia Pacific Tele-Ophthalmology Society (APTOS) Symposium.

Currently, Aravind technicians travel to these rural areas to capture images and then rely on highly trained doctors to review the images and provide diagnosis. Their goal is to scale their efforts through technology; to gain the ability to automatically screen images for disease and provide information on how severe the condition may be.

Problem Statement

This is a deep learning problem. Inputs are the images and the goal is to predict severity of diabetic retinopathy on a scale of 0 to 4:

0 - No DR

1 - Mild

2 - Moderate

3 - Severe

4 - Proliferative DR

Datasets and Inputs

The datasets are provided on Kaggle competition website. They are free to download.

Dataset files

train.csv - the training labels

test.csv - the test set (you must predict the diagnosis value for these variables)

sample_submission.csv - a sample submission file in the correct format

train.zip - the training set images

test.zip - the public test set images