

Capstone Proposal

I will build a live camera app for Android to detect house numbers on streets. The project was featured in the Deep Learning course of Udacity.

Domain Background

The domain is number recognition on videos. The app will recognize the numbers on the live image and show it to the user. This project will use [Google's paper](#) that was released in 2014. The paper explains Google's way to recognize multi-digit numbers from static Street View images using Deep Convolutional Neural Networks. This project will do the same using a smartphone with live images. The best part of this project will be combining Machine Learning with Software Engineering which are the field of interests of the author.

Problem Statement

The problem is the fact that house numbers have different formats. The numbers can appear with non-standard baseline, broken outlines, non-standard fonts or bad localization. The goal will be recognizing all of those cases.

Datasets and inputs

Until I design a good architecture, I will use the MNIST database. MNIST database provides handwritten characters and I will use them by concatenating numbers. After I have a good architecture at hand, I will use the SVHN dataset which contains house number images from Google Maps. I won't use SVHN at the beginning because SVHN is a more challenging dataset as its digits are not neatly lined-up and have various skews, fonts and colors. The input will be the live image from the Android smartphone.

Solution Statement

I will use Convolutional Neural Networks to tackle this problem. The input will be the image from Android and the output layer will consist of nodes that represent a digit each. There will also be another node for the length of the house number.

Benchmark Model

We will use the model that is specified in Google's paper. Image as input, hidden layers and an output layer that contains nodes that represent a digit each. The paper also mentions benchmark values for accuracy.

Evaluation Metrics

These benchmark values are coverage, overall accuracy and per character accuracy. The authors of the paper achieved 96.5% coverage, 96% overall accuracy and 97.8% per character accuracy. For coverage, we define a confidence threshold and discard the predictions that are less likelier than the threshold. Coverage is the proportion of non-discarded values to all values.

Project Design

I will first design an architecture using MNIST database. Then improve it with SVHN database. After that, I will put the model into an Android App. The user will point the camera to some house numbers and these numbers will be recognized. The recognized numbers will be shown to the user on the app.