[Team 25] ProjF Proposal : Street View House Number Detection

Anil Kumar Yalla
ayalla@ncsu.edu

Ankur Sharma
asharm48@ncsu.edu

Priya Sharma
psharm23@ncsu.edu

I. Motivation

For humans classifying images is considered a simple task. However, this is a difficult and complex task for machines. We have reliable software like OCR to perform this task for us but they require the images to be properly aligned and be distinguishable from the background. We aim to solve the problem of image classification with real world limitations. The goal of the project is to effectively identify the house numbers found on Google Street View images and translate them into number characters. In this situation, OCR performs inadequately because character detection in natural scenes is challenging due to blurriness, distortion, alignment and font changes in the images, which results in misclassification of numbers. An important use case of this project is being able to identify house numbers on Google Street View and matching them to an address.

II. Data

We are planning to use the Street View House Numbers (SVHN) Dataset. It has over 600k labelled real-world images of house numbers taken from Google Street View. The training dataset consists of 73,257 digits and the test dataset consists of 26,032 digits. Apart from that we have 531,131 additional, somewhat less difficult samples, to use as extra training data. The sequence of numbers in the images are of bounded length. There are 10 classes, 1 for each digit. Digit '1' has label 1, '9' has label 9 and '0' has label 10. This dataset comes in two formats : (1) Original images with character level bounding boxes and (2) MNIST-like 32-by-32 images centered around a single character (many of the images do contain some distractors at the sides). We are planning to use the (2) format of the dataset for our project.

III. Methodology

For the task of classification of street view house numbers we consider basic feed forward neural networks as our baseline model. We implement the other models to analyze the results by varying the architecture of models.

A. Feed Forward Neural Network

We plan on using Feed Forward Neural Network as our baseline model. The information in this network only flows in one direction, forward, from the input nodes to the output nodes, passing through any hidden nodes. In the network, there are no cycles or loops.

B. Convolutional Neural Networks

We plan on improving the baseline model using CNN. It is a type of feed-forward artificial neural network with variations of multi-layer perceptron's designed to use minimal amounts of preprocessing. There are certain advantages offered by convolutional layers when working with image data like fewer parameters, sparsity of connections, parameter sharing and spatial in-variance.

C. Residual Neural Network

Further we plan on improving our CNN model using ResNet. One of the key changes to our CNN model is the addition of the residual block, which adds the original input back to the output feature map obtained by passing the input through one or more convolutional layers. In order to solve the problem of the vanishing/exploding gradient, this architecture introduced the concept called Residual Network. In this network we use a technique called skip connections. The skip connection skips training from a few layers and connects directly to the output. The approach behind this network is instead of layers learn the underlying mapping, we allow network fit the residual mapping.

IV. Evaluation

For Evaluation of this model, we consider 75% as training data and 25% as testing data. And the training data is split into train and validation data. Validation data is of 15% random sample from the training dataset. Our primary evaluation metrics is accuracy for test dataset. We also plan on generating the confusion matrix for the models so as to get a sense of how it performs on each class label and how the misclassifications are distributed.

References

[1] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, Andrew Y. Ng Reading Digits in Natural Images with Unsupervised Feature Learning NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011.

[2] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, Xindong Wu, Object Detection with Deep Learning: A Review, 2019.

[3] Andrew G. Howard, Some Improvements on Deep Convolutional Neural Network Based Image Classification, 2013.

[4] Image Classification using Deep Neural Networks https://medium.com/@tifa2up/imageclassification-using-deep-neural-networks-a-beginner-friendly-approach-using-tensorflow94b0a090ccd4

[5] Dataset-https://www.kaggle.com/stanfordu/street-view-house-numbers