# Image Classification and Unsupervised Image Object Removal in the Street View House Numbers Dataset

# Group 7 Final Project Report

**Bill Grieser -- Darshan Kasat -- Shivam Thassu**

# Introduction

In this project, we will use three Neural Network Frameworks that we encountered in class to classify the Street View House Numbers (SVHN) dataset. (Netzer, et al., 2011) The three frameworks are:

* TensorFlow
* Caffe
* PyTorch

In addition, we will use the by-products of the PyTorch classification model to attempt unsupervised object removal of digits in the SVHN test data set.

We chose this dataset because it gave us the opportunity to utilize Convolutional Layers and other topics that we covered in class. The SVHN is considerably more complicated than the MNIST dataset, as it contains confounding objects in the images, the digits appear in different angles in the images, and the dataset is in color.

For the Image Object Removal, the goal of this task it to use by-products of the PyTorch classification network to draw inferences about the foreground and background region of a SVHN digit image, and use the inferences to train a Generalized Regression Neural Network to regenerate the image in a photorealistic manner but without the foreground object in it – in this case, the digit.

For the image classification task, we will measure success by the overall accuracy rate in classifying the test data. We have set a target of 90% accuracy. For the Object Removal task, the goal is to explore the concept and determine areas of further study.

# Description of the Data

The SVHN data was collected by Netzer, et al. from the Street View images in Google. In their work, they took a two-step approach: first, identify the digits in an image, and then, classify the digit as 0 through 9. In the data they publish, we are relying on their first step: identifying the bounding boxes in images that contain a digit. We are addressing the second task, which is to recognize the specific digit. (Note: This data was also used in a Kaggle challenge but we have obtained the data from the website of the original project and we have done our own pre-processing).

The data provided by Netzer, et al. are a set of parent images containing a street view image that includes digits. Metadata for each image identifies one or more digits within the parent image.

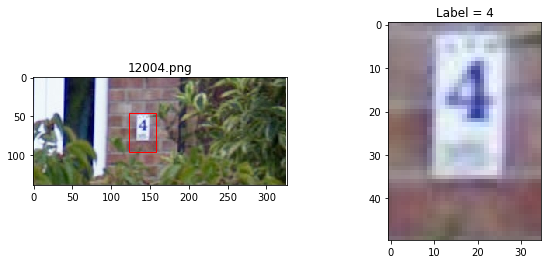


Figure Example of Parent Image from the dataset and the individual digit extracted by our preprocessing

(The bounding box is not in the original but is generated based on the metadata). To see several examples of parent images with their digits extracted based on the metadata, in the code directory, run:

python3 read\_pickles.py

This will display several examples of the data.

The data is provided in two sets: a training set with 33,402 parent images and a test set with 13,068 parent images. These contain 73,257 individual digits in the training set and 26,032 digits in the test set.

# Image Classification Task

For the image classification task, we used three frameworks to train models

## Caffe

## TensorFlow

## Pytorch

# Digit Removal from Images

# Summary and Conclusions

# References

Alilou, V., & Yaghmaee, F. (2015). Application of GRNN neural network in non-texture image inpainting and restoration. *Pattern Recognition Letters*, 24-31.

Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., & Ng, A. Y. (2011). Reading Digits in Natural Images with Unsupervised Feature Learning. *NIPS Workshop on Deep Learning and Unsupervised Feature Learning.* Retrieved November 18, 2018, from http://ufldl.stanford.edu/housenumbers