Computer Vision Final Project (Fall 2018)

Classification and Detection with Convolutional Neural Networks

Rupal Gupta
rgupta338@gatech.edu

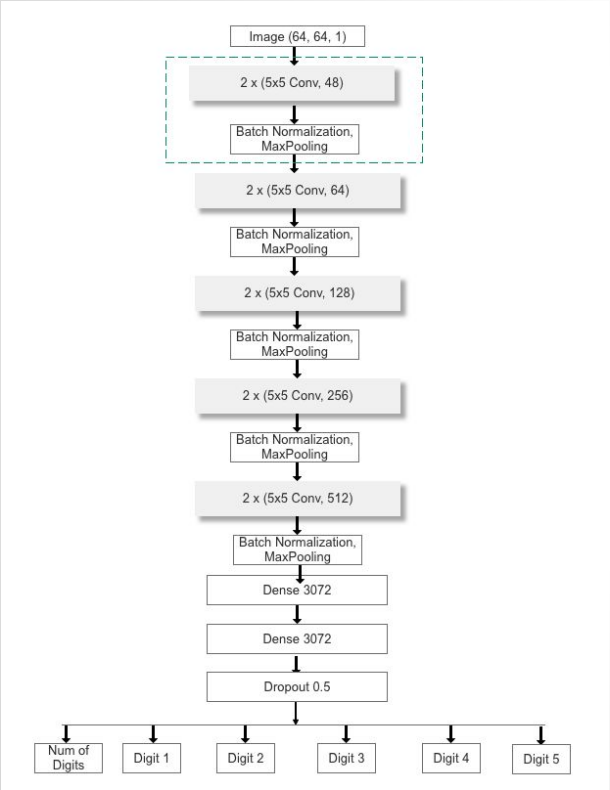
Data Analysis and Preparation

- Multi-Digit Detection (entire sequence): SVHN Full format images
- Image patches / bounding box regions resized to 64x64
- Negative training examples formed by cropping patches in regions outside the bounding box
- Assumption: Sequence length of digits 5
- Normalization using the mean and standard deviation of training dataset
- Custom CNN model: used grayscale normalized images
- VGG 16 models: used RGB normalized images
- Validation data created by shuffling training set and randomly selecting 20% of data

Transfer Learning using VGG16 Model

- Performed Transfer Learning using VGG 16 models in two ways:
  - Training VGG 16 model from scratch (randomly initialized weights)
  - Using VGG 16 model with pre-trained (ImageNet) weights
- Loaded VGG 16 model without the top layer (consisting of FC layer)
- Added two dense layers with 3072 units
- Six output layers with Softmax Activation function
- The first output generates the total number of digits in a sequence and the rest of 5 outputs predict the sequence of digits in the image.

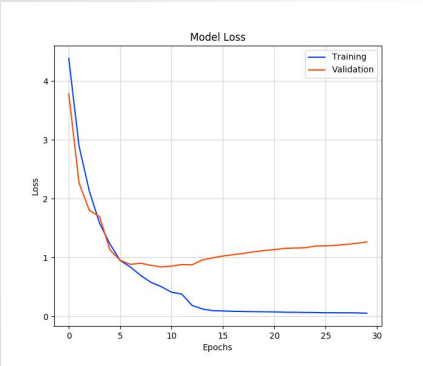
Custom CNN Architecture



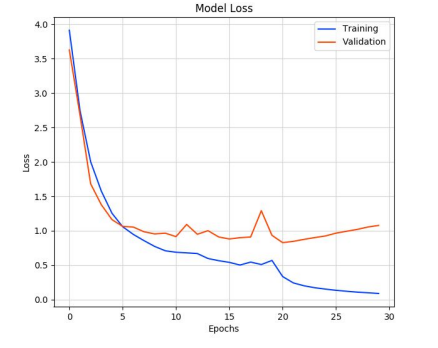
- 10 Conv Layers, 5 Batch Normalization Layers and 5 Max Pooling Layers.
- Followed by 2 Dense layers with ReLu
- 6 Dense Layers with Softmax
- Optimizer: Adam
- Batch Size: 64
- Learning Rate: 10^{-3} (ReduceLROnPlateau)
- Loss Function: Sparse Categorical Cross Entropy
- Epochs = 30; Early Stopping
- Image Augmentation

Model Performance Comparison

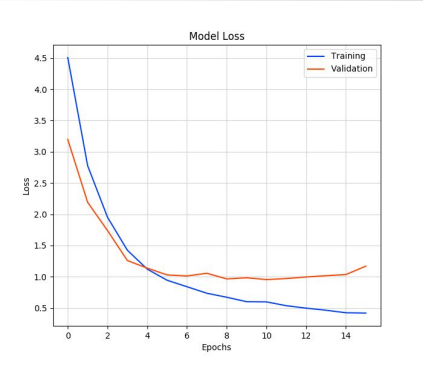
Model	Overall Train Sequence Percentage Accuracy	Overall Test Sequence Percentage Accuracy
VGG 16 Scratch	85.89%	78.42%
VGG 16 pre-trained weights	88.52%	83.72%
Custom CNN	92.16%	86.72%



Custom CNN



VGG 16 pre-trained weights



VGG 16 scratch

Steps for Real Image Processing

- Read in the input image
- Create 64x64 patches using sliding window over Image Pyramids
- Normalize the image patches
- Feed to the digit detector CNN model
- Windows where confidence > 0.88 are saved (others discarded)
- Use Non Maxima Suppression to get the correct most voted region
- Only these windows are fed to multi-digit classifier CNN
- The final result is the one with confidence > 0.98

My Results on Real Images

Images that worked well:

Images that did not work well:

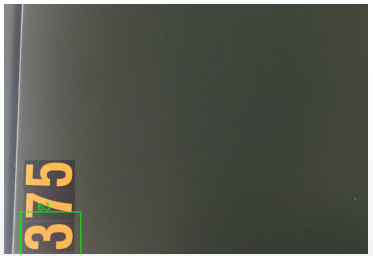


Image and Video Result Links

Image Results for all 5 images can be accessed at: Office 365 Drive Images Link or Google Drive Images Link

Video Results can be viewed at: Office 365 Drive Link or Google Drive Link

All Plots: All Error, Loss and Accuracy Plots or Plots Google Drive Link