Classification and Detection with Convolutional Neural Networks

Computer Vision (FALL 2019) Final Project

Aiping Zheng
azheng39@gatech.edu

Introduction

- Recognizing multi-character text in natural image still could be very challenging.

- Deep Convolutional Neural Network (CNN) were developed to solve this problem and achieved huge success.

Data and preprocessing

- 32x32 RGB images in Format 2 of the Street View House Numbers (SVHN) Dataset were used for training.
- digit 0 was originally labeled 10, after preprocessing, it was changed to 0.
- 10000 images from CIFAR10 was used as negative control (no digit in the image) and labeled 10.
- Resize
- normalize

Models architecture and training

- Model 1: VGG16 with pretrained weight
- Model 2: VGG16 without pretrained weight

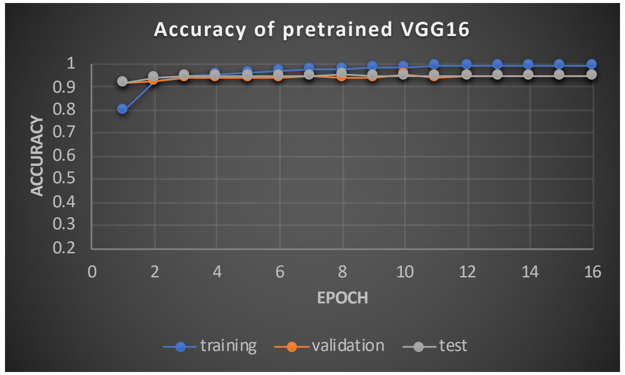
- Final output layer of both model 1 and model 2 was also changed to 11 classes from 1000 classes.

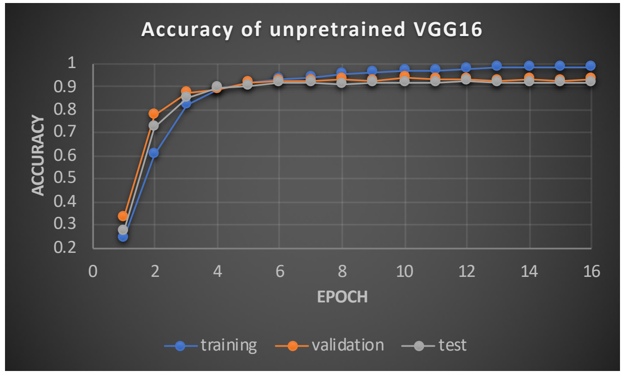
- Model 3: the same architecture as the one in Goodfellow’s paper. The output layer were also changed to suit 11 classes, and dropout rate was set to 0.2.

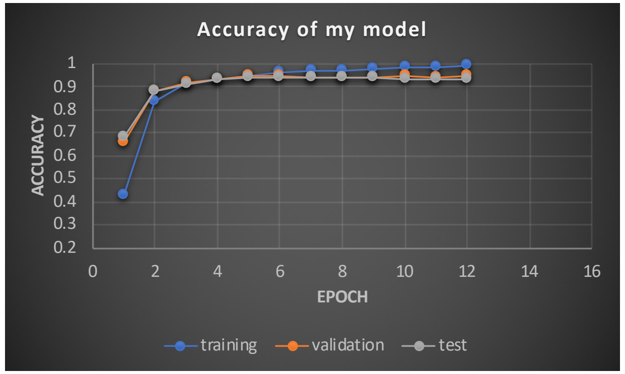
Overall pipeline

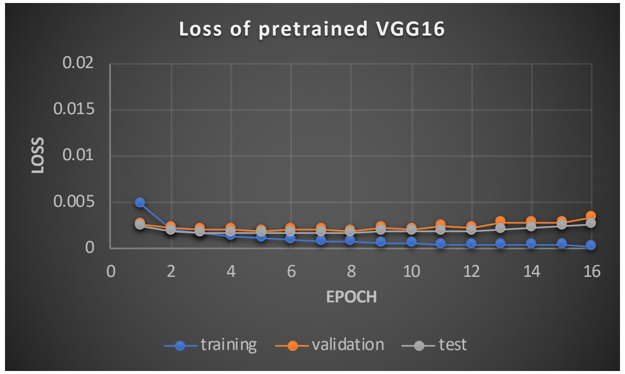
- maximally stable extremal regions (MSER) of openCV
- trained CNN **model 3** was used to detect digits.
- All bounding boxes without digits inside were removed.
- Some other thresholds were also applied for bounding box selection, including width and height ratio of bounding boxes.
- Non-maximum Suppression (NMS)
- Clustering
- The digits were combined into one number.

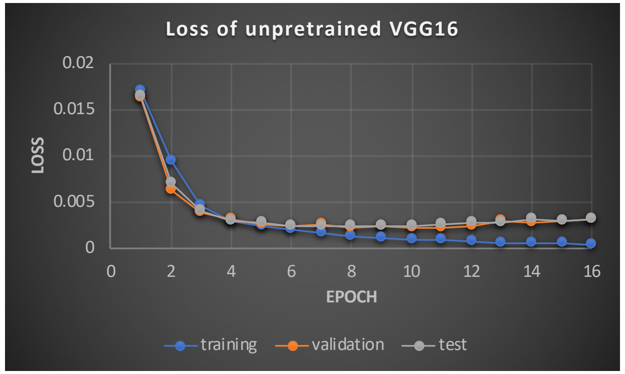
Model performance

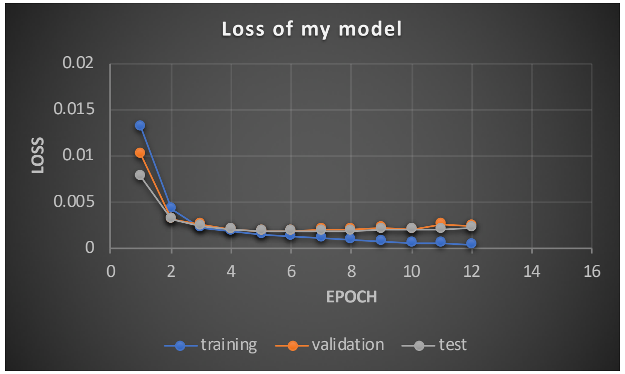

Accuracy of pretrained VGG16


Accuracy of unpretrained VGG16

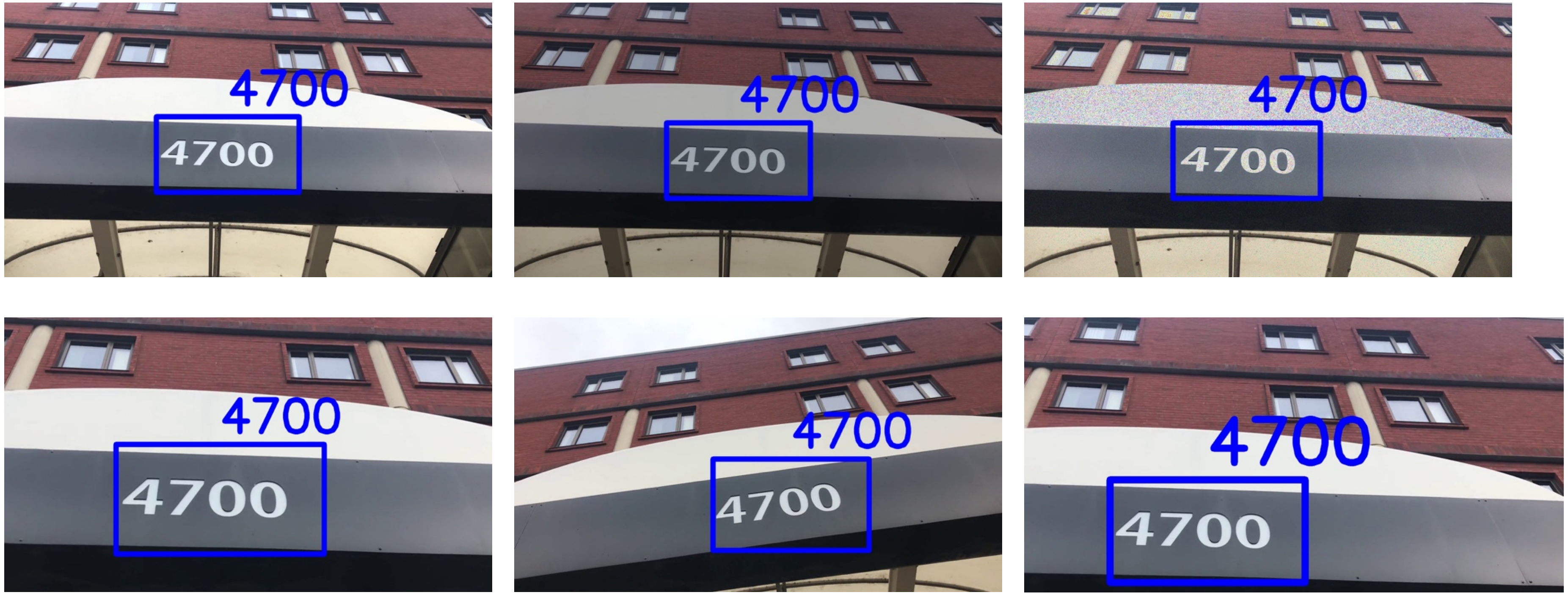

Accuracy of my model


Loss of pretrained VGG16

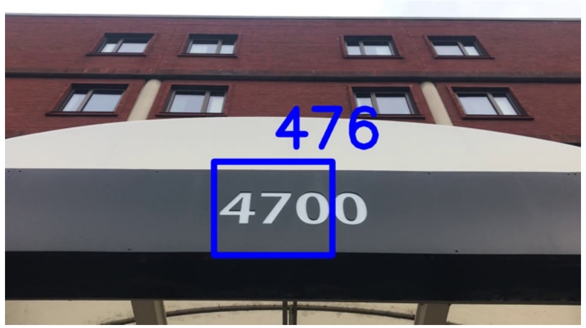

Loss of unpretrained VGG16

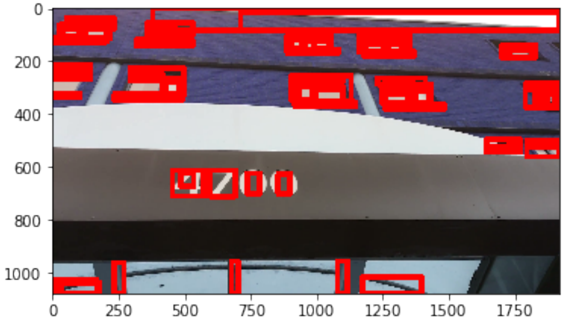

Loss of my model

Detection on real images



Negative result





Discussion

- **Comparison to state-of-art work**

- The testing accuracy of our model is 93.45%

- I. J. Goodfellow et al 96%, Y. Netzer et al. 91%

- human performance is 98%.

- **Future improvements**

- parameters for MSER need to be further tuned

- Negative control datasets better than CIFAR10

- Extra training dataset