Read Digits from Natural Images using Convolutional Neural Network

Ramesh Kumar
Roberto Cai Wu

November 13, 2017

Motivation

- This is a optical character recognition (OCR) problem
- Digit recognition is used in various applications such as postal mail sorting, bank check processing, form data entry, etc
- Digit recognition is an important component of modern-day map making [1]

Problem Description

- The task is to read digits from natural images
- We use the Street View House Numbers dataset [2], which consists of real-word images taken from house numbers
- We use convolutional neural networks(CNN) for fast processing, accuracy and speed

Challenges

- Wide variability of visual appearance of text: fonts, colors, and orientations [1]
- Different environmental factors: lightning, shadows, and occlusions [1]
- Image acquisition factors: motion, blurring, and resolution [1]

Assumptions

- Images do not contain any characters, other than digits
- Background color will not change and it contain digits of different colors and intensities

Methodology

- Load and Interpret DataSet(done)
- Pre-processing(started)
- Convolutional Neural Network(started)
- Post-processing
- Testing and evaluation

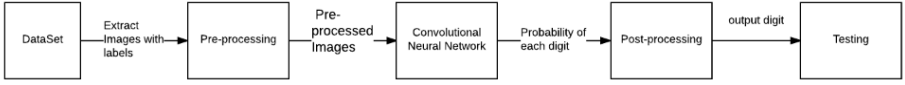


Figure: Block Diagram of System

The Street View House Number Dataset [2]

- 10 classes, 1 for each digit
- Digit 1 has label 1,9 has label 9, and 0 has label 10
- 73257 digits for training, 26032 digits for testing
- Image are from variable-resolution and color

DataSet

Figure: Example images from SVHN dataset [2]

Pre-processing dataset

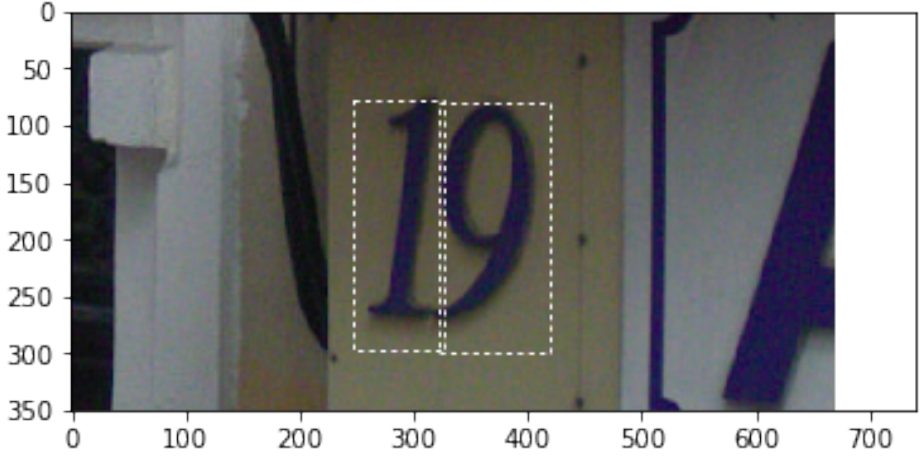


Figure: Reading labeled data and generating bounding box

Pre-processing dataset

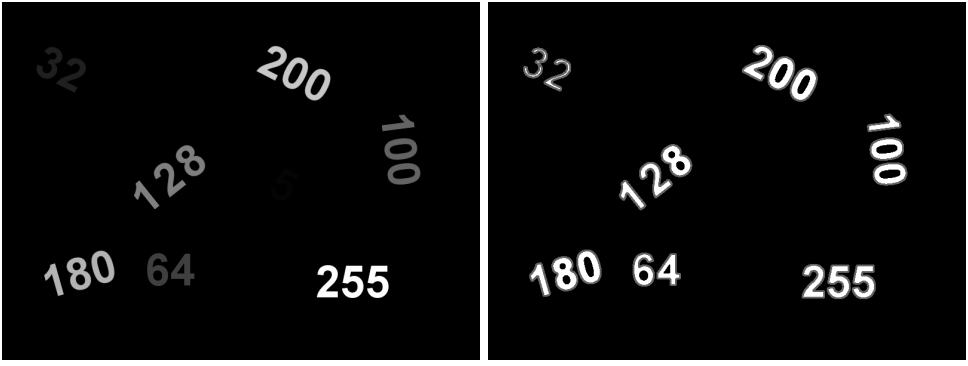
(a) Cropped and converted to gray scale

(b) Cropped and converted to gray scale

Pre-processing from camera

- Resize image to 640x480 pixels
- Convert to gray scale
- Apply Gaussian filter
- Use a binary thresholding
- Find contours
- Draw bounding box around contours

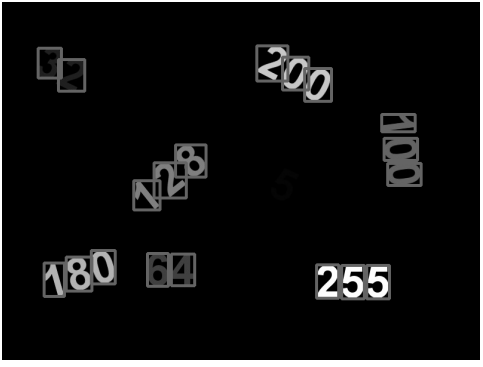
Pre-processing from camera(2)



(a) Original image

(b) Applying threshold and contours

Pre-processing from camera (3)



(a) Bounding box over contours

Convolutional Neural Network(CNN)

- State-of-the-art shows CNN performs better as compare to other approaches[3]
- Extracts features from the images and classify them
- Three type of layers
  - Convolutional: Extract low-level and high-level features
  - Pooling: Reduce amount of parameters and computations
  - Fully Connected: Neurons are fully connected

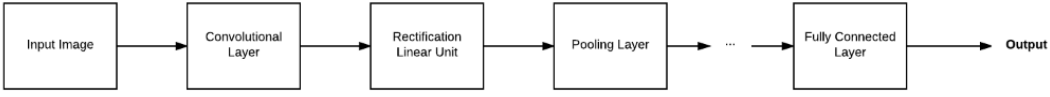


Figure: Basic Architecture of CNN

Testing & Evaluation

- Print numbers in a sheet of paper (different size, font, color, and orientation)
- Test the images of digits from live camera under different conditions (light and perspective)
- Use test set to compute accuracy of model

- Ian J Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, and Vinay Shet. **Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks.** *CoRR*, abs/1312.6:1–13, 2013.
- The Street View House Numbers (SVHN) Dataset. `http://ufldl.stanford.edu/housenumbers.` "[Online; accessed 07-10-2017]".
- Pierre Sermanet, Soumith Chintala, and Yann LeCun. **Convolutional neural networks applied to house numbers digit classification.** In *ICPR*, pages 3288–3291. IEEE Computer Society, 2012.
- Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. **Reading digits in natural images with unsupervised feature learning.**

In *NIPS workshop on deep learning and unsupervised feature learning*, volume 2011, page 5, 2011.