Read Digits in Natural Scene Images using Convolutional Neural Networks

Ramesh Kumar(Matr. Nr. 9029726)
Roberto Cai Wu(Matri. Nr. 9029544)

October 26, 2017

1 Motivation

How this project related to Computed Vision?

- Since computer vision is about interpreting the images from the camera such as image processing, classifying images based on features, etcetera.
- In this project, we also use computer vision methods for pre-processing and post-processing such as image size reduction, image conversion, and many more, other than convolutional neural network.
- Digit recognition is a computer vision problem used in applications such as postal mail sorting, bank check processing, form data entry, etcetera.

Why Convolutional Neural Network

- Since fast processing, accuracy and speed for these applications is important therefore, convolutional neural network is useful in this case.
- Furthermore, according to state-of-the-art, it shows better performance as compare to other approaches[1]

2 Approach

The first part of this task consists of gathering data to train the network. We will use an already available dataset, Street View House Number (SVNH)[2]. Images in this dataset come in various resolutions, colors, perspective, etcetera. In addition to this, we will create our own dataset to train and test on it. Images from dataset will be preprocessed before feeding to the network. Later on, post-processing will be needed in cases where an image contains multiple digits to get a final result by combining result of individual digit.

- Few possible approaches to solve this task are:
  - Multiple hand-crafted features
  - Template matching(might work on this also if time permits)
  - Convolutional Neural Network(we use this)

- Convolutional Neural Networks are used to extract features from the images and classify them. These networks are consists of different layers such as convolutional layers, pooling layers, and fully connected layers. Convolutional layer is first layer of the network that takes input as image and fully connected layer is final layer/output layer of the network that provides possible outcome based on input image.
- **Challenges**
  - Different Lighting conditions
  - Design an appropriate network
  - Different perspective view
  - Blurred digits
  - Can be more when we start implementation
- For testing, we will test the images from live camera under various possible conditions such as hand-written digits, computerized-written digits, different lighting conditions, different font sizes, etcetera.
- Possible outcomes will be within 10 digit classes; one for each digit from label "0" to "9" and probability score of each outcome.
- Furthermore outcome will also include locations of all recognized digits as well as information about what digits belong together and form a complete number

3 Contribution

- Pre-processing using available dataset(Roberto Cai)
- Create the dataset and integrate with publicly available dataset (Ramesh Kumar)
- Create the Network(pair)
- Post-processing (Roberto Cai)
- Probability Score (Ramesh Kumar)

- Individual number combination in final result (Roberto Cai)
- Evaluation of approach (Ramesh Kumar)
- Documentation(pair)

References

[1] Pierre Sermanet, Soumith Chintala and Yann LeCun, "Convolutional Neural Networks Applied to House Numbers Digit Classification", The Courant Institute of Mathematical Sciences - New York University

[2] Yuval Netzer , Tao Wang , Adam Coates , Alessandro Bissacco1 , Bo Wu , Andrew Y. Ng, "Reading Digits in Natural Images with Unsupervised Feature Learning", In NIPS Workshop on Deep Learning and Unsupervised Feature Learning, 2011.