Data 2040 Final Project Work Journal

Team: DPU

Members: Enmin Zhou, Huaqi Nie, Siyuan Li, Yangyin Ke, Haoda Song

**Overall Goal**

In the following few weeks, we will build an Optical Character Recognition (OCR) system to recognize words from images of natural scenes.

## **Relative Work Links**

### GitHub: <https://github.com/siyuanli1202/DATA2040_Final_Project>

### Blogs:

* + Initial Blog: <https://huaqi-nie.medium.com/optical-recognition-system-in-natural-scenes-7cd3c26b03e7>
  + Midway Blog:
  + Final Blog:

### Google Colabs:

* + Notebooks could be found in Team Shared Drive since currently we cannot use brown.edu address to access google colab

## **Week 1 (Apr. 1 - Apr. 4)**

### **Achievements:**

* + Clarified the overall goal of the project, building an OCR system to recognize words from the images of natural scenes. The process could be separated into two sections, text detection and text recognition in word level. (All members)
  + Researched the natural scenes image datasets for potential uses. Decided to use the COCO-2014 Text Recognition dataset as the inputs for text detection (Haoda Song, Yangyin Ke) and Google IIIT 5k word recognition dataset as the inputs for text recognition (Enmin Zhou).
  + Researched CRNN models for text recognition (Siyuan Li, Huaqi Nie) and implement EAST model from scratch (Haoda Song)
  + Explored the COCO-2014 Text Recognition dataset (82784 images) and Google IIIT 5k word recognition dataset. (All members)
  + Extracted 13880 images with text from COCO-2014 for text detection use. (Yangyin Ke)
  + Performed Fast-RCNN as the baseline text detection model. (Yangyin Ke, Enmin Zhou)
    - Faster-RCNN: <https://medium.com/jumio/object-detection-tutorial-with-torchvision-82b8f269f6ff>
  + Reduced the noise for the images after resizing and performed EAST-model as the baseline text recognition model. (Siyuan Li, Huaqi Nie)
    - EAST model:
      * <https://arxiv.org/abs/1704.03155v2>
      * <https://jaafarbenabderrazak-info.medium.com/opencv-east-model-and-tesseract-for-detection-and-recognition-of-text-in-natural-scene-1fa48335c4d1>

### **Further Steps:**

* + Explore other models' architectures in order to improve the performance. Other models could include models in YOLO family, SSD, RetinaNet:
    - <https://lilianweng.github.io/lil-log/2018/12/27/object-detection-part-4.html#yolov3>
  + Tune the hyperparameters and re-design the architectures of the models in use.

## **Week 2 (Apr. 5 - Apr. 11)**

### **Achievements:**

* + Implemented SSD (with VGG16, MobileNetv2 and EfficientNetB07) to detect bounding boxes of text in Google Street View dataset. (Yangyin Ke, Enmin Zhou, Haoda Song) <http://vision.ucsd.edu/~kai/svt/>
  + Improved performance of SSD model with MobileNetv2 so the model could capture tiny and vague text in the SVT pictures. (Haoda Song, Yangyin Ke, Enmin Zhou)
  + Researched and discussed the differences between models in YOLO family, RetinaNet and SSD for small text detection, especially different loss functions, to improve the future models. (Enmin Zhou, Yangyin Ke, Haoda Song) <https://jonathan-hui.medium.com/real-time-object-detection-with-yolo-yolov2-28b1b93e2088>
  + Implemented the tesseract model on the image data of Google IIIT 5k word recognition to evaluate the model performance. (Huaqi Nie, Siyuan Li) <https://nanonets.com/blog/ocr-with-tesseract/>
  + Implemented the CRNN (convolutional recurrent neural network) on the image data of Google IIIT 5k word recognition to perform text detection tasks. (Huaqi Nie, Siyuan Li)
  + Built the CRNN model architecture based on the paper *An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition* <https://arxiv.org/abs/1507.05717> (Huaqi Nie, Siyuan Li)
  + Experimented on image preprocessing to increase prediction accuracy. (Huaqi Nie, Siyuan Li)
  + Researched on the Attention-based OCR model for image text recognition.

### **Further Steps:**

* + Improve the models for text detection so that the model could better capture the tiny text in natural senses photos.
  + Improve the models for text recognition so that the model would better detect the letters and numbers in a special or artistic style.
  + Pipe the text detection model and text recognition model in Tensorflow.
  + Implement the final model in Google MLkit so that anyone could recognize the text in the natural-view photos through the phone cameras.

<https://github.com/MaybeShewill-CV/CRNN_Tensorflow>

<https://arxiv.org/pdf/1601.01100.pdf>

<https://github.com/FLming/CRNN.tf2>

<https://www.kaggle.com/samfc10/handwriting-recognition-using-crnn-in-keras>

Dan :

<https://github.com/ultralytics/yolov5>

<https://github.com/zzh8829/yolov3-tf2>

<https://arxiv.org/pdf/2102.12627.pdf>

<https://www.reddit.com/r/MachineLearning/comments/m3rbk4/project_geoffrey_hintons_glom_for_mnist_using/>