

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 IIT 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 IIT 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.