Capstone Project Proposal -- (Chinese Character OCR )

* ***Bosheng Wu, Mentor: Rahul***
* **Project Introduction:** 
  + Optical character recognition (OCR) is a technique used for detecting text from images, which can be applied in plenty of area. For Healthcare, it can help doctor detect text from patient’s health record document. In translator, it can be used for detecting text from document or natural image. Also, for traffic system, OCR is used for detecting car license plate number. Traditional OCR can be separated based on two classes: document text image and scenes text. There are a lot of state of art algorithm which has a really good performance for doc-text OCR. For scenes text, however, it still is a challenge task to improve. Since noise, fuzzy and other uncertain condition in image, detection model used in doc-text can not easily solve scenes image problem.
  + **Goal**: When people travel to a new country, it will be convenient for them has an App to help them translate the word such as street sign, cafe menu, etc. My plan is trying to make an App which can translate Chinese character. Which can help foreign people when they travel in China.
  + **Demo**
    - High Scope: Develop an end-to-end model for detecting & recognizing Chinese Character from scenes images. Also, it can translate character sequence into english. And deploy this model into a IOS App which can translate Chinese character from image captured by my iphone camera.
    - Low Scope: Develop an model for detecting & recognizing Chinese Character from scenes images. Deploy this model into website which can recognize character from picture you uploaded.
  + **Dataset**:
    - Chinese Character:
      * CTW- Scenes Image Data with Chinese Character: <https://ctwdataset.github.io/downloads.html>
      * CASIA - Chinese handwritten data: <http://www.nlpr.ia.ac.cn/databases/handwriting/Home.html>
    - English Character:
      * ICDAR competition
      * MSRA-TD500 (small dataset): <http://www.iapr-tc11.org/mediawiki/index.php/MSRA_Text_Detection_500_Database_(MSRA-TD500)>
    - Numerical Image:
      * SVHN (Street Number): <http://ufldl.stanford.edu/housenumbers/>
  + **State of art solution review**
    - Whole OCR pipeline can be separated into two parts: Text detection & recognition
    - Text Detection: To understand the character content, we need to localize characters in the scenes image. Which means using a bounding box to mark characters region.
      * **Classical Image Processing**
        + Stroke Width Transform(SWT): Apply algorithm “SWT” to find word location. <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/1509.pdf>
        + Maximally Stable External Regions (MSER): Before using SWT candidate detection, using MSER to improve character candidate detection. Also they apply SVM based on HOG features & CNN features to classify object & background. <https://www.researchgate.net/publication/322780311_Text_Detection_Based_on_MSER_and_CNN_Features>
        + Other Classical Image processing might work in doc-text detection but usually has bad performance in the scenes text.
        + Effort and Resource:

OpenCV has built-in function of these two algorithms.

Need use dataset for character/background classification. (But not very expensive training).

* + - * **Machine Learning**
        + AdaBoost Classifier: This model used a group of features for weak classifiers and merge them based on boosting to make a good performance. <http://www.cs.jhu.edu/~ayuille/pubs/ucla/A188_xchen_CVPR2004.pdf>
        + Effort and Resource;

Have source code in github.

Need more time for training each weak classifier and merge together.

* + - * **Deep Learning**
        + Faster-RCNN: two stages framework
        + SSD: one stages framework
        + Convolutional Character Network: This is the most recent scene-text detection model in 2019, which is the Top1 in Leaderboard. <http://www.cs.jhu.edu/~ayuille/pubs/ucla/A188_xchen_CVPR2004.pdf>
        + Effort and Resource:

Have source code framework in github

Need more time for training, also need to consider for per-trained model and transfer learning.

* + - Text Character Recognition: After we find the ROI of Text, next step is recognizing character. For character recognition, there are two types: characters segmentation + single char recognition, characters sequence recognition. Machine Learning has best performance in both areas.
      * **Machine learning:**
        + Character segmentation + Neural Network: Firstly, we segment character sequence in bounding box into single characters (“abc” -> ‘a’,’b’,’c’) and use NN to classify label of each char. (or We can change NN to CNN for better performance)

Network is easy to implement, just need to find the right dataset to train.

Also need some time to figure out how to segment single character.

* + - * **Deep Learning**:
        + CRNN+CTC(loss): create a end2end model with CNN + RNN (or LSTM) + CTC loss to recognize char sequence. <https://arxiv.org/pdf/1507.05717.pdf>

Have source code in github

Need more time in training, need to find correct dataset. Also need to consider using transfer learning.