Snap2Text Project Proposal

Fiona Feng | Yichen Li

Project Summary

The present study aims to develop a robust and efficient system for text extraction from images that can accurately recognize text amidst varying backgrounds, lighting conditions, and font styles. We propose the implementation of an Optical Character Recognition (OCR) system based on Convolutional Neural Networks (CNN).

The primary focus of our research will be to enhance the performance of the OCR system and optimize its accuracy through rigorous experimentation and algorithmic refinement.

Moreover, we plan to extend the capabilities of our OCR system to enable real-time text recognition from video subtitles or text appearing in a video. This extension will cause the development of novel algorithms for detecting and extracting text from video frames and the seamless integration of the OCR system to facilitate real-time recognition of the extracted text.

Goals and Objectives

- Research various optical character recognition (OCR) approach in the following prospectives:
  - Preprocessing–prepare the image for better text recognition
  - Character Extraction–extract characters based on position and sentences
  - Identification- align the extract character with the correspondent character
- Develop an optical OCR system using convolutional neural networks (CNN) for text recognition from images
- Investigate the feasibility of implementing real-time OCR for video subtitles and display translation text.
  - Identify the screen areas for text recognition
  - Display translation text in the correspondent location

Related Works

From [1], it proposes using convolutional neural networks (CNNs) for text detection directly from raw color pixels. This approach involves joint feature extraction and classification in one step, with feature extraction being learned via examples using the back-propagation algorithm. The network is trained to reject badly localized text and emphasize good localization of text lines by the network itself, instead of relying on tedious geometric constraints and local image processing. The article presents the proposed network topology, training details, and experimental results.
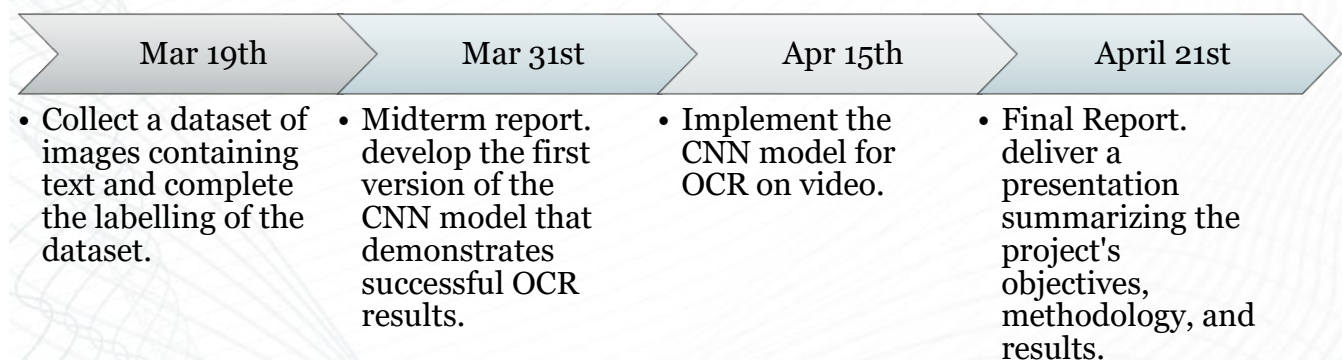
[2] describes a method for text recognition in natural scenes using CNN and transfer learning. The method involves two parts: character recognition and scene text processing. In the character recognition part, the algorithm is trained on alphabet character images using data loading and augmentation techniques to improve the model's accuracy. The trained model is then evaluated on validation and test data to find the model with the best accuracy. In the scene text processing part, the algorithm processes the scene text images using image processing techniques to detect the bounding box of each character. These bounding boxes are then cropped and saved as new images, which are used to evaluate the trained model's accuracy. The method uses transfer learning to improve the performance of the model on the target dataset with limited training samples.

Proposed Approach

This research project will entail the following steps:

1. Acquisition of a dataset of images featuring text from a variety of sources.
2. Preprocessing of the images and character labeling to produce a training dataset.
3. Development of a CNN model architecture incorporating appropriate parameters, using deep learning frameworks such as TensorFlow, Keras, or PyTorch.
4. Training of the CNN model using the labeled training dataset.
5. Evaluation of the trained CNN model by testing it against a separate dataset that was not used for training.
6. Refinement of the CNN model through fine-tuning via parameter adjustments or retraining with additional data to enhance its accuracy.
7. Assessment of the CNN model's performance on videos.

Timeline

Mar 19th	Mar 31st	Apr 15th	April 21st
• Collect a dataset of images containing text and complete the labelling of the dataset.	• Midterm report. develop the first version of the CNN model that demonstrates successful OCR results.	• Implement the CNN model for OCR on video.	• Final Report. deliver a presentation summarizing the project's objectives, methodology, and results.

Duties

Fiona: Conduct research on various CNN and deep learning models for OCR

Ethan: Gathering a dataset and writing code in Python

References

[1] Delakis, M., & Garcia, C. (2008, January). text Detection with Convolutional Neural Networks. In *VISAPP (2)* (pp. 290-294).

[2] Fajar Rizky, A., Yudistira, N., & Santoso, E. (2023). Text recognition on images using pre-trained CNN. *arXiv e-prints*, arXiv-2302.