SD1 - Yunhua Zhao American Dream

•••

Martin Lahoumh
Brandon Tjandra
Miguel Luna
Jiazhou Zhang

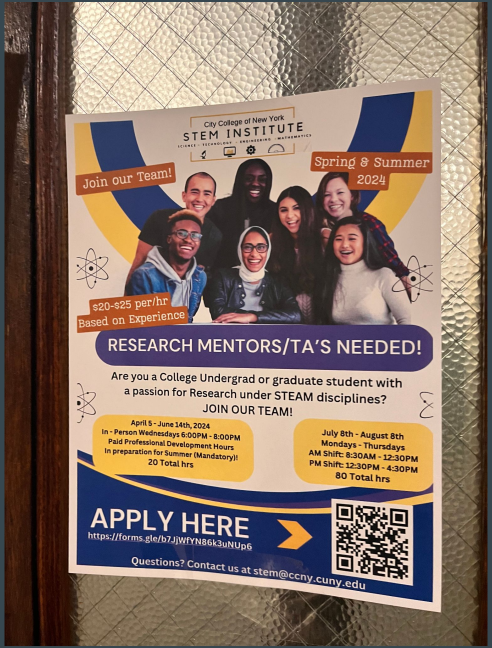
Proposal

The goal of this project is to create a web application that:

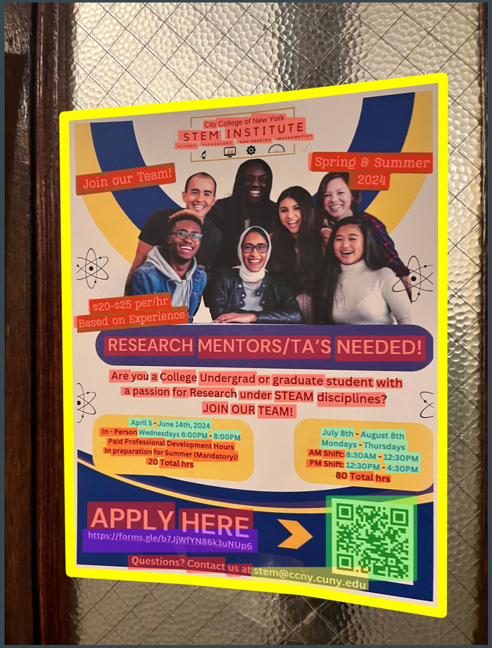
1. Takes an image (supplied by the user) that contains text in some language
2. Extracts the text
3. Translates it to a user's preferred language
4. Places the translated text on the appropriate parts of the image

The machine learning aspect of this project is detecting where in an image text is located and extracting the text.

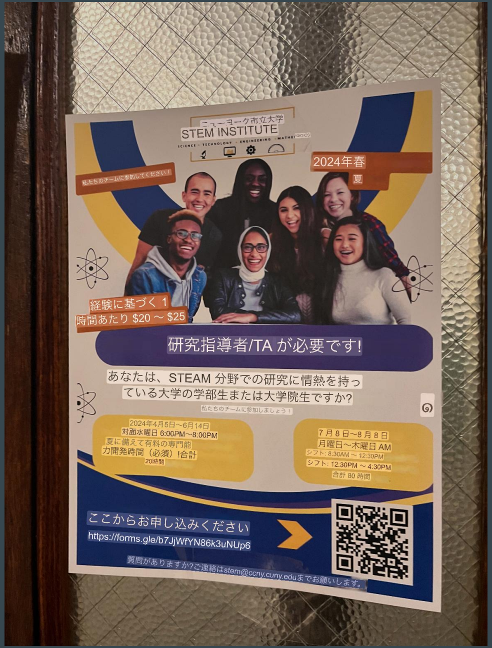
Usage example


City College of New York
STEM INSTITUTE
Join our Team!
Spring & Summer 2024
$20-$25 per/hr Based on Experience
RESEARCH MENTORS/TA'S NEEDED!
Are you a College Undergrad or graduate student with a passion for Research under STEAM disciplines?
JOIN OUR TEAM!
April 5 - June 14th, 2024
In - Person Wednesdays 6:00PM - 8:00PM
Paid Professional Development Hours
In preparation for Summer (Mandatory)!
20 Total hrs
July 8th - August 8th
Mondays - Thursdays
AM Shift: 8:30AM - 12:30PM
PM Shift: 12:30PM - 4:30PM
80 Total hrs
APPLY HERE
https://forms.gle/b7JjWfYN86k3uNUp6
Questions? Contact us at stem@ccny.cuny.edu

1. User supplies image to the application


City College of New York
STEM INSTITUTE
Join our Team!
Spring & Summer 2024
$20-$25 per/hr Based on Experience
RESEARCH MENTORS/TA'S NEEDED!
APPLY HERE
https://forms.gle/b7JjWfYN86k3uNUp6
Questions? Contact us at stem@ccny.cuny.edu

2. Apply AI text detection and recognition


ニューヨーク市立大学
STEM INSTITUTE
2024年春 夏
経験に基づく 1 時間あたり $20 ～ $25
研究指導者/TA が必要です!
あなたは、STEAM 分野での研究に情熱を持っている大学の学部生または大学院生ですか?
7月8日～8月8日
月曜日～木曜日 AM
ここからお申し込みください
https://forms.gle/b7JjWfYN86k3uNUp6

3. Translate text and return translation over image

Research & Current Knowledge

This project will require the use of a R-CNN and OCR.

- R-CNN (Region Based Convolutional Neural Network):
  - A CNN used for image segmentation
  - Image segmentation creates bounding boxes for each identified object
  - This will be needed for **text detection**

- OCR (optical character reader):
  - The process of converting text in an image to machine-encoded text
  - This will be needed for **text recognition**

Our methods of choice

For our R-CNN, we will use two:

- With **Mask R-CNN**, we will identify the text sign within the whole image. It is more useful for identifying sign objects rather than the sign's text.

  There is a lot of material online to reference due to its popularity amongst image detection projects.

- With **OpenCV**, we will identify the individual text characters from the sign. It has methods that can distinguish text that are to close together or poorly written, so it makes the approach more broad.

For our OCR, we will use **Tesseract-OCR**, more specifically **pytesseract**, for its compatibility and ease of use with OpenCV.

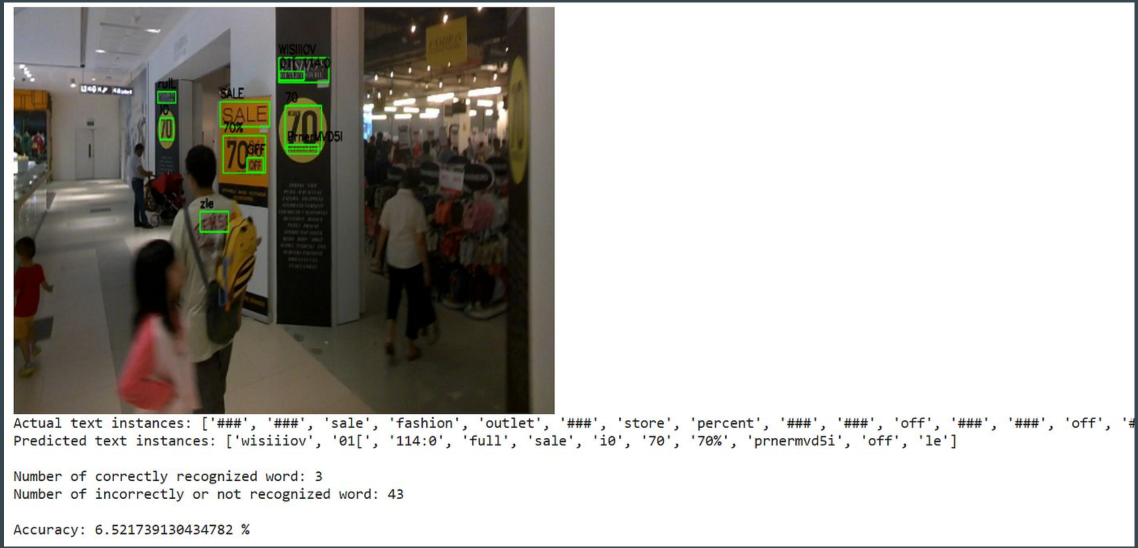
Once text is extracted, we will use **Google's Cloud Translation API** to translate the text to the desired language the user wants. We will overlap the translated text on top of where the previous text was on the image.

Potential methods (cont.)

The following are used in existing Python OCR models we have found during research:

- **MSER** (Maximally Stable Extremal Regions)
- **Tesseract-OCR**
- **EasyOCR**

(1): EasyOCR model example



```
Actual text instances: ['###', '###', 'sale', 'fashion', 'outlet', '###', 'store', 'percent', '###', '###', 'off', '###', '###', 'off', '#
Predicted text instances: ['wisiiiov', '01[', '114:0', 'full', 'sale', 'i0', '70', '70%', 'prnermvd5i', 'off', 'le']

Number of correctly recognized word: 3
Number of incorrectly or not recognized word: 43

Accuracy: 6.521739130434782 %
```

Datasets

Name	Description	Number of entries	Date last updated
COCO	Object detection, segmentation, and captioning dataset	330K images, >200K labeled	2017
COCO-Text	Large scale dataset for text detection and recognition in natural images	63,686 images, 145,859 text instances	2018
TextOCR	Text-recognition on arbitrary shaped scene-text present on natural images	28,134 images and 903,069 annotated words	2021
ICDAR 2019	Used for a oriented scene multilingual text detection and spotting challenge	20,000 images containing ten different languages	2019
Curve Text (CUTE80)	The first public curved text dataset, images contain text with complex transformations and obstructions	80 images, labelled sequentially via XML sheet	2014
Total-Text	Word-level based English curve text dataset, built on top of CUTE80	1,555 images, character masks, and text stored sequentially via .gif files	2022
SCUT-CTW1500	Text-line based dataset with both English and Chinese instances	1,500 images, labelled sequentially via .txt file	2020

Development timeline

For the alpha version:

→ Our goal is to have a working and satisfactory text recognition model.

For the beta version:

→ Our goal is to have working translation overlays from text recognition.

We may develop the GUI (the website for the project) either simultaneously or after the completion of these two main objectives.

Beyond this point, we may consider looking into implementing more features.

Feasibility & next steps

With where we are at currently, we believe that we will not have much difficulty completing the application within the allotted time in the next semester.

Should we finish early, the following are features we may add to expand the application **(from least to most difficult)**:

1. Categorize based on type of signage, including traffic and street signs
2. Identify and provide interaction with supplemental information in an image, such as QR codes, phone numbers, etc.
3. Add support for recognizing source languages besides English
4. Create a mobile application for this project
5. Add augmented reality (AR) support for the application to identify and translate signs with camera use **(desirable, but unlikely!)**

Thank you!