

Project Summary

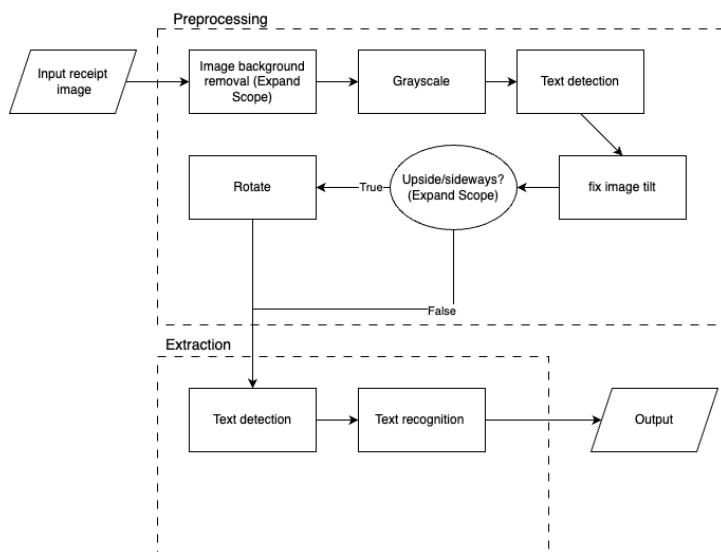
Our project is a machine learning model that can turn receipt into text. This is useful for users who may want to keep track of their physical receipts as a digital text file. Furthermore, down the line we can calculate total, ect with the digital text information.

What Has Been Accomplished So Far:

- Created unified environment amongst us
- Created github repo and issues
 - Github Link: <https://github.com/andyrzwang/cmpt310-rcpt-scan>
- Created unified mind map/data flow diagram
- Updated project scope and timeline to be more realistic
- Gathered data
 - Pretrained data with verified data
 - (supervised training)
 - Data to train model
- Blocker faced :
 - Downgrade python for tensorflow compatibility (solved)
- Reviewed github procedures with members.
- Created diagram
- Decided on libraries to use (reflected in github repo readme)
 - opencv-python
 - Pytesseract
 - Ocr
 - pandas
 - easyocr
 - Also ocr but better for noisy images such as receipts

Any Project Changes (brief):

We have revised our project plan(shown below) and we have also adjusted our scope since the professor has said our scope is quite big. We have made some items optional in case we find that our project is completed too quickly/easily.



Revised Project plan/Timeline

Milestone 1 – July 2

Goal: Basic environment setup, find good datasets of receipts, ensure we can focus on pure coding for the future

- Collect sample datasets of receipt images from Kaggle and public datasets.
- create proper environments to ensure smooth future coding

Expected Deliverable: Project setup, and structure of tasks along with technical tools needed to ensure straightforward coding process.

Milestone 2 – July 30

Goal: Model that extracts text from multiple printed receipts with simple ui

- Can extract text/numbers from a messy/ugly printed receipt
- Make a simple text based UI
- Add support for batch processing multiple receipts.
- Optionally: Output raw extracted fields into a CSV or DataFrame, make a better ui, find total/ect

Expected Deliverable: A more robust system that accurately converts a printed receipt into computer text/numbers.