Shop Sign Detection and Text Recognition in Video

This document summarizes the approach, challenges, solutions, and potential improvements for the provided code that detects shop signs, recognizes text from those signs, and compiles the data into an Excel spreadsheet.

Approach:

The code leverages a pre-trained deep learning model called YOLOv5 for object detection. This model identifies shop signs and logos within individual frames of a video. Once a shop sign is detected, the code extracts the relevant portion of the frame (region of interest or ROI) for further processing. To enhance text recognition accuracy, pre-processing steps are applied to the ROI, including converting it to grayscale and performing basic noise filtering. Finally, the Tesseract OCR engine is used to recognize text within the pre-processed ROI. The extracted shop names and corresponding timestamps are compiled into a Pandas DataFrame and saved as an Excel spreadsheet.

Challenges:

Several factors can influence the accuracy of shop sign detection and text recognition, including:

- Video quality: Factors like blurriness and lighting conditions in the video can affect accuracy.
- Shop sign appearance: Variations in shop sign design, such as fonts and styles, can pose challenges for detection and recognition.
- Noise: Video frames might contain noise that can hinder the ability to recognize text.

Beyond these general video quality issues, there are more specific challenges to consider:

- Inaccurate object detection: The YOLOv5 model might not perfectly distinguish shop signs from other objects, leading to irrelevant text extraction.
- Poor image quality: Blurry or low-contrast video footage can make it difficult for Tesseract to accurately recognize text.
- Text complexity: Unusual fonts, cluttered backgrounds, or overlapping text on signs can be problematic for OCR engines.

Solutions:

The code addresses some of these challenges through the following solutions:

- Confidence threshold: The code implements a confidence threshold to filter out detections with low confidence scores. This helps to reduce the number of false positives (incorrectly identified shop signs).
- Pre-processing: Grayscale conversion and noise filtering are applied as pre-processing steps to improve the quality of the input for the OCR engine.

- Data cleaning (optional): The code includes a commented-out section for potential data cleaning. This could be used to remove nonsensical entries after text recognition.

Potential Areas for Improvement:

There are several ways to potentially improve the accuracy and robustness of the system:

- Fine-tuning YOLOv5: By fine-tuning the YOLOv5 model with a dataset of shop signs specific to the video domain, detection accuracy can be improved.
- Alternative OCR engines: Exploring OCR engines trained for specific fonts or languages relevant to the shop signs might enhance recognition results.
- Advanced pre-processing: Techniques like adaptive thresholding or CLAHE (Contrast Limited Adaptive Histogram Equalization) could be explored for further noise reduction and contrast enhancement during pre-processing.

Future Work:

The extracted text could be further classified to categorize shop types based on the recognized names. Additionally, optimizations could be made to enable real-time processing of video streams.