COSE474-2024F: Final Project Proposal
"Nutrition Information Extraction from Nutrition Facts Using OCR"

Yoo Jiyeon

1. Introduction

One of the most essential aspects of people's lives is health. By consuming a balanced intake of nutrients suited to our needs, we can better manage our health. Processed foods, which we commonly encounter, include a nutrition facts that outlines their nutritional ingredients according to specific standards and methods. However, due to various factors such as complex labeling and small font sizes, individuals face limitations in fully understanding the nutritional information and making informed food choices.

In this project, we utilize OCR technology to extract key nutritional information—such as calories, protein, fat, and sugars—from nutrition labels and present it to users in a simplified text format.

2. Problem definition & chanllenges

The project focuses on automating the process of extracting key nutritional information from food labels using Optical Character Recognition (OCR) and presenting it to users in an easily understandable format.

It is essential to clearly distinguish between letters, numbers, and the units of each nutrient on the nutrition label. Typically, nutrition labels use various units such as g, mg and etc. Accurately recognizing the values and identifying which nutrient each unit belongs to is crucial.

Next, nutrition labels vary in format depending on the country. In this project, we will train the system based on nutrition labels that follow FDA regulations. To improve the overall accuracy, the system must be capable of extracting the desired nutrient information regardless of the label format.)

3. Related Works

Huang, Zheng, et al. "Icdar2019 competition on scanned receipt ocr and information extraction." 2019 International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2019.

Baek, Youngmin, et al. "Character region awareness for text detection." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

OCR consists of two stages: text recognition and text detection. For the text detection model, we plan to use CRAFT, and for the text recognition model, TPS-ResNet will be employed. .

4. Datasets

I plan to create a dataset consisting of nutrition label images with features such as calories, carbohydrates, fat, protein, and sugars. Additionally, we will improve the performance of the text recognition model by using a dataset composed of general text images and text.

5. State-of-the-art methods and baselines

Baseline Methods: Traditional OCR systems, such as Tesseract, often serve as baseline methods. While Tesseract performs well for clean and simple text documents, its accuracy decreases significantly with complex layouts, small font sizes, and low-quality images—issues that are common in nutrition labels. This highlights the advantage of using more advanced SOTA models like CRAFT and TPS-ResNet for our project, as these models are better suited to handle the variability in nutrition label designs.

In the field of Optical Character Recognition (OCR), the combination of CRAFT for text detection and TPS-ResNet for text recognition represents one of the most effective SOTA approaches for structured text detection.

6. Schedule & Roles (if you have a teammate)

- collect nutrition label images for training and testing

- implement OCR using CRAFT for text detection and TPS-ResNet for text recognition

Build the model to extract key nutrition information from the OCR output.