Reflection

Introduction

To help small restaurants visualize customized food, we propose a deep learning-based system that enables automatic image editing. Customers can select different ingredients to include in their dishes and receive realistic images to help them visualize their decisions. In our design, the categories of dishes include burgers, sandwiches, tacos, and sides accompanying main dishes.

We decided to use Mask R-CNN to segment and classify dish components precisely. Then, we applied diffusion-based inpainting to change the ingredients in the image at high resolution, realistically. This solution enhances customized food visualization from the customer's side and reduces photography costs for restaurants. Our method would offer an efficient and scalable way to food visualization and ultimately empower small-scale restaurant businesses.

Challenges

Since we decided to change the database, the size of the database increased dramatically, and the complexity of the images also increased at the same time. Although we believe the new database will allow for a broader range of applications, we still need more time for training and fine-tuning the model. Beyond that, since we apply two different models in our system, streamlining the work processes of these two models will also be a challenge for us.

For the diffusion-based inpainting model’s evaluation, we also discuss the objectivity of the evaluation. Our original plan was solely based on human evaluation, which is inevitably more subjective and unstable. To help address this issue, we are working on a scoreboard method to enable a more objective evaluation of our output.

Insights

Right now, we are still working on fine-tuning the process of our models. The model could output general food segmentation and classification, but it doesn’t meet our expectations. We expect the model to be strong enough to support the expected function

Plan

For now, we are planning to allocate more time to fine-tune our models. Different from last time, we have decided to use a new dataset called FoodSeg103. Compared to the original dataset (UNIMIB2016), the new dataset offers greater variability in terms of food categories, food positioning, and picture styles. The size of the dataset also increases significantly(from 1,027 to 7,118 images), which enables us to better fine-tune the model. We will focus on fine-tuning the model and integrating the two models in the future. We also think about using the ChatGPT API to determine whether the food is healthy or not. So far, we have also proposed replacing unhealthy food with healthy alternatives as part of a health monitoring system, but we will assess the feasibility of this feature after completing the model fine-tuning.