

# Computer Vision to reduce food wastage

Sharath Giri and Anitesh Reddy  
April 14, 2023

## Abstract

*In the last decade, computer vision has taken massive strides towards helping humans to better understand the universe around us. This has aided the process in many fields and impacting individuals and society at various levels. Our focus is to utilize the advancements in science of computer vision to address a key issue of food wastage around the globe. We believe that there is a lot of food wastage in many industries due to poor feedback mechanism such as, food served in flights, buffets in restaurants, even at homes etc. Our project will explore the possibility of combining various established models to achieve this task for instance, GLIP, RESNET50 etc. Our intention is to create a model which can be used by airlines, restaurants and individuals to save food.*

## 1. Introduction

Airlines spend around USD 20 Billion on food and beverages out of which approximately USD 4 Billion worth of food is wastage [2]. Absence of a structured feedback mechanism and government regulations are the major factors for this wastage. Roughly one-third of the edible parts of food produced for human consumption, gets lost or wasted globally, which is about 1.3 billion ton per year. In North-America and Europe around 10-15% of this wastage by consumers.[?] Our final goal is to create a model which

can be run on an edge device and capture videos of the completed food trays and feeds it into the model which in turn gives feedback on food wasted. For instance, we would plant this device on top of service carts which will capture the video of trays being collected after food consumption. Similarly, these devices can be placed on top of collection trays at restaurants (Example of a food tray before and after consumption in a n aircraft: Figure1). Over a period of time, if we consistently get feedback that a dish is not completely consumed, we can conclude either the portion size can be reduced or the taste is to be question.

Our model should be able to capture figure 1b and process it and give feedback that main dish is completely eaten



Figure 1. Shows figure 1a and 1b, our model will read figure 1b and identify the different containers to predict how much of that dish is wasted

and salad is half eaten. In some cases, we may face a situation where the container is not clearly visible and the model should be able to identify those and not give any feedback.

## 2. Related Work

There is quite some research done to quantify the amount of food that is being wasted along the supply chain from production to consumption. There are behaviour studies done as well to understand what is behaviour aspect is causing the wastage of food ref3. Airlines have also started to adopt AI to bring down food wastage. Very recently (11th

April 2023) Airbus announced usage that they are working on food scanner to help airlines reduce food wastage [3]. Similarly, Lufthansa has announced (on 23rd Mar 2023) that they will be using similar approach as our model to reduce food wastage but will only be able to roll it out in 2024 [4]. Limitics is a Singapore based start-up which is at forefront of the concept to use AI to reduce food wastage in many industries such as Airlines, Hospitality etc. They have partnered with Eithad Airways to come up with similar approach as ours and have been consuming data from aircrafts for over a year now. [1] Similarly, there are efforts

made to reduce food wastage at restaurants using similar concepts. Overall industries are definitely moving in the direction of adopting AI to reduce food wastage. We haven't come across papers publishing usage of computer vision to reduce food wastage.



Figure 2. Airbus Prototype of food scanner which will capture the images and process them to generate insights on food wastage

### 3. Model

#### 3.1. Data Collection

**DALL-E for Image Generation:** The first step in our proposed system is to use DALL-E, a generative deep learning model that produces high-quality images based on a given text prompt. In our case, we use DALL-E to generate images of meals with varying levels of food consumption. These images will then be processed by the GLIP model for cropping. Currently we are targeting on predicting the consumption of Rice only. Hence, we have generated an image dataset of around 900 images and manually classified them into 4 categories namely as eaten, over-served, tasted, untouched.

Apart from DALL-E we are collecting images manually which are corrupt. Our definition of corrupt image is an image where the food is not visible correctly. example: Figure 3.

#### 3.2. Model Architecture

The proposed system consists of a combination of GLIP (Generative Latent Image Perception) followed by a model for identifying corrupted images and a fine-tuned ResNet50 architecture integrated with transformers for classifying the degree of food consumption. The system demonstrates promising performance in classifying unseen data, suggesting potential applications in the food industry for efficient food waste management and meal planning. High level model architecture is provided in figure 4.

**GLIP Model for Image Cropping:** We use the GLIP (Generative Latent Image Perception) model to automatically crop the generated meal images, focusing on the regions containing food items. This step ensures that our system processes only relevant portions of the images, which will improve the accuracy and efficiency of the subsequent classification model. We plan to test our model on unseen data as well. Thus, we plan on using few images manually



Figure 3. Image of a corrupt dish

obtained from the internet and try to crop out the relevant portions where the food is present.

**Image Corruption Detection:** Before classifying the degree of food consumption, it is crucial to ensure that the images used are of high quality and free from any corruption. Images are classified as corrupt if there are foils/tissue present in the tray or if the lid is closed. For this purpose, we employ a separate deep learning model specifically trained to identify corrupted images. This model acts as a pre-processing step, filtering out any unusable images before they are inputted into the main classification model. To obtain the training dataset for this model, we are planning on creating the dataset at home and capture the images through our mobile.

**Fine-tuned ResNet50 Model with Transformers for Consumption Detection:** To classify the degree of food consumption in meals, we use a fine-tuned version of the ResNet50 architecture integrated with transformers. This hybrid model combines the strengths of both CNNs and transformers, resulting in a powerful and accurate system for classifying food consumption levels. We train the model on a dataset of labelled meal images, covering a wide range of consumption levels. The fine-tuned ResNet50 model with transformers is then able to classify the degree of food consumption in unseen meal images. At the moment we have classified single dish uncorrupted images using ResNet50 and gained an accuracy of 77.8% on 15 images. We have observed that the model is performing well even on unseen data such as Noodles or different types of plates.

