In the name of God,

This document is the English translation of my Master's thesis, authored by me, Shayan Rokhva, graduate student of the Department of Information Technology Engineering, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran.

This thesis was successfully defended on September 21, 2025 (30 Shahrivar 1404), receiving a perfect score of 20 out of 20, and I am now an officially certified graduate.

Please note that this version is NOT an official translation, but rather an accurate personal translation prepared by me, a fluent English speaker.

Its purpose is to familiarize you with the thesis, its topic, and the technical and research skills I developed throughout this work.

Should you have any questions regarding this thesis, please feel free to contact me.

Shayan Rokhva
Department of Information Technology Engineering
Faculty of Industrial and Systems Engineering
Tarbiat Modares University, Tehran, Iran

🖂 shayanrokhva1999@gmail.com (suggested)
🖂 shayan1999rokh@yahoo.com

WhatsApp: +98 939 397 2774

Note:

I have made every effort to ensure accuracy; however, since this translation (from Persian to English) was completed within approximately one day, minor errors may still exist. If you notice any, please feel free to contact me.

Tarbiat Modares University

Department of Information Technology Engineering

Faculty of Industrial & Systems Engineering

M.Sc. Thesis

Food Waste Recognition & Estimation Using Deep Learning and Image Processing

Author

Shayan Rokhva

Supervisor

Dr. Babak Teimourpour

September 2025

Endless gratitude to the Almighty for granting humankind the power of thought, enabling us to pursue the path of progress and enlightenment. I am deeply thankful that divine grace allowed me to take steps on this path with my limited scientific knowledge.

I wish to express my sincere appreciation and heartfelt gratitude to my esteemed supervisor, Dr. Babak Teimourpour, for his invaluable guidance, patience, and insightful advice that greatly contributed to the quality and direction of this research. Under his supervision, I not only gained knowledge but also learned the virtues of perseverance, hope, patience, and dedication.

I also extend my gratitude to all my professors, each of whom has generously and sincerely shared their knowledge and experience throughout my academic journey.

Abstract

Food waste is one of the critical challenges of the 21st century, with far-reaching economic, social, and environmental consequences, making its reduction an undeniable necessity. This study develops an intelligent framework based on deep learning and image processing for the automatic estimation of food waste in real-world environments such as university dining halls and restaurants. For this purpose, plate images were collected in two stages — *before and after consumption* — and prepared through semantic labeling and the creation of pixel-wise masks. To mitigate overfitting and enhance accuracy, the data underwent preprocessing and intelligent augmentation. Subsequently, conventional deep learning architectures as well as customized variants tailored to the research requirements were implemented and evaluated using standard image segmentation metrics alongside problem-specific criteria.

The results on unseen test data revealed that the best-performing models achieved a minimum Dice score of 0.85 across all food categories or higher, while the proposed distributional pixel accuracy metric recorded values of at least 0.9 for all categories. From a computational efficiency perspective, all models performed adequately, achieving inference speeds of at least 20 images per second. However, the optimized, lightweight U-Net reached approximately 83 images per second, enabling real-time deployment in practical applications. Moreover, the customized versions—with modified architectural components and loss functions—demonstrated competitive, and in some cases superior, performance to baseline models while reducing processing time. These enhancements enabled comparative analysis of consumption behavior across different food combinations and identification of each component's contribution.

Overall, the proposed framework is innovative and efficient, leveraging cutting-edge technologies such as deep learning and computer vision. It can serve as an effective tool for intelligent food waste monitoring and as a decision-support system toward enhancing the sustainability of the food ecosystem.

Keywords:

Food waste, Semantic segmentation, Deep learning, Image processing, Computer vision, Food sustainability

Table of Contents

Chapter 1

Introduction and General Framework of the Research

1.1.Introduction

Food, as one of humanity's most fundamental needs, plays a crucial role in food security, public health, and environmental sustainability. However, global statistics reveal that a significant portion of the food produced each year is wasted at various stages of the supply chain—particularly during consumption. These losses are estimated to range between 25% and 40% across different sources. Food waste not only leads to the loss of valuable resources such as water, energy, and labor from an economic perspective but also causes extensive environmental consequences, including increased greenhouse gas emissions and pressure on waste management systems. Moreover, the occurrence of food waste while a large segment of the global population struggles with food insecurity adds ethical and social dimensions to the issue, further underscoring its importance (Ahmadzadeh et al., 2023; Slorach et al., 2019).

In recent decades, emerging technologies—particularly artificial intelligence (AI) and its subfields—have gained a prominent role in addressing complex challenges. Among these technologies, deep learning has emerged as one of the most advanced approaches, creating remarkable transformations across diverse domains. With its ability to automatically extract complex features from large-scale data, deep learning has achieved notable success in fields such as natural language processing, speech recognition, medical image analysis, and computer vision systems. Leveraging such capabilities can enhance efficiency, optimize processes, and reduce costs across various industries (Talaei Khoei et al., 2023).

In the context of food waste management, the use of deep learning and image processing represents an innovative and effective approach. These technologies can automatically analyze images of food at different consumption stages to estimate leftovers and identify patterns associated with waste. Such capabilities provide valuable insights for decision-makers and managers in the food industry, restaurants, and catering services, serving as practical tools for waste reduction, supply planning improvement, and sustainability enhancement (Lubura et al., 2022).

Accordingly, the objective of this study is to develop and evaluate a deep learning- and image-processing-based framework for food waste estimation. Achieving such a framework can not only contribute to resource conservation but also lay the groundwork for developing intelligent strategies and policies that promote sustainable food systems and improve quality of life.

1.2.Problem Statement

As previously discussed, food waste has become one of the most critical challenges facing modern societies, with multifaceted economic, social, and environmental consequences. According to international reports, nearly one-third of the food produced globally each year ends up as waste. This massive loss not only squanders valuable resources such as water, land, energy, and human labor but also contributes significantly to greenhouse gas emissions, thereby exacerbating climate change (Cederberg and Sonesson, 2011; Lubura et al., 2022). From a social perspective, while a large portion of the global population continues to suffer from food insecurity, the increase in food waste reflects an unjust inequality in resource distribution. Therefore, identifying scientific and technological solutions to mitigate this problem is an undeniable necessity (Attiq et al., 2021).

Among the major contributors to food waste are university dining halls, public canteens, and restaurants. In such facilities, the large volume of prepared and served meals often leads to substantial food waste due to various reasons, such as mismatched portion sizes relative to individual needs, changes in taste preferences, or unsatisfactory food quality. Consequently, obtaining an accurate estimation of food waste in these environments can assist managers and policymakers in making intelligent decisions for waste reduction, optimizing raw material procurement, and even developing broader policies (Faezirad et al., 2021a; Turker, 2025). Despite the significance of this issue, conventional methods for estimating food waste—such as weighing meals before and after consumption—suffer from several limitations. These methods require precise instruments and human labor, are time-consuming and costly, and are unsuitable for large-scale or automated applications (Lubura et al., 2022; Rokhva and Teimourpour, 2025a).

In recent years, advances in modern technologies—especially in the field of artificial intelligence—have opened new opportunities for addressing complex problems. Among these, deep learning, as one of the most powerful subfields of machine learning, has achieved remarkable success in areas such as speech recognition, natural language processing, medical image analysis, and computer vision. The main advantage of deep learning lies in its ability to automatically extract complex patterns and features from large-scale data, making it particularly suitable for processing food images and estimating food waste (Talaei Khoei et al., 2023).

Specifically, the integration of deep learning with image processing enables the development of systems capable of analyzing food images before and after consumption to estimate leftovers and, consequently, the amount of waste. This approach operates rapidly and autonomously, without the need for human intervention, offering higher accuracy and efficiency than traditional methods. Furthermore, the data obtained from this process can be utilized to analyze consumer behavior, identify waste patterns, and support the design of targeted waste reduction policies (Lubura et al., 2022; Mazloumian et al., 2020).

Accordingly, the central problem of this study lies in exploring how modern technologies—particularly deep learning and image processing—can be utilized for automatic and efficient estimation of food waste in real-world environments such as university dining halls. This central question leads to several subproblems. First, the research is particularly significant in large-scale food service settings (e.g., university canteens and large restaurants), where the level of waste is typically higher (Faezirad et al., 2021b, 2021a). Second, estimating waste per food category can provide managerial insights for improving menu design and portion control, thereby reducing food waste, environmental damage, and associated human and social costs (Lubura et al., 2022; Wang et al., 2017). Third, from a deep learning perspective, the study can offer valuable insights, such as identifying which architectures are most effective for food image analysis, how to address challenges like visual diversity of dishes, varying lighting conditions, or non-food items on plates, and how to present the results in actionable forms for decision-makers to promote waste reduction and sustainability (Lubura et al., 2022; Rokhva and Teimourpour, 2025a).

In summary, the main research questions of this study can be formulated as follows:

- ✓ How can classification, detection, and segmentation techniques—as subfields of artificial intelligence and image processing—be utilized to identify and estimate food waste, and how can the models' performance be evaluated in terms of accuracy and speed?

- ✓ What is the estimated amount of food waste for each food category, and what errors, considerations, and limitations exist in the proposed model that can serve as directions for future research and improvement?

1.3.Research Objectives

In general, the objectives of this research can be summarized as follows:

- ❖ To estimate the approximate amount of waste for each food category using deep learning–based image processing.
- ❖ To evaluate the accuracy of the estimations using both conventional and innovative (customized) evaluation metrics.
- ❖ To customize, optimize, and improve the loss functions, optimizers, and model parameters, as well as to define new evaluation criteria where applicable.
- ❖ To address the research questions by analyzing and comparing the obtained results, discussing key findings and challenges, and proposing suitable directions for future research.

1.4.Research Steps

The steps required to conduct this research are outlined below:

(1) Literature Review:

In the first stage, an in-depth literature review is conducted to explore various dimensions of the research topic. This includes examining the global and national status of food waste, its economic and environmental consequences, and a review of deep learning architectures—particularly their applications in the food industry and related sectors such as agriculture. More importantly, the review encompasses nearly all studies published in the past three years related to the intersection of artificial intelligence technologies and the food industry. Naturally, this stage also covers key concepts of image classification, object detection, and image segmentation as subfields of computer vision and artificial intelligence.
To ensure reliability and scientific rigor, the related articles are collected from reputable and peer-reviewed databases such as *ScienceDirect*, *Springer*, *Wiley*, *MDPI*, and *Nature*, rather than general search engines like Google Scholar. In total, more than 225 papers were initially reviewed, of which approximately 80 highly relevant studies were examined in detail and utilized throughout the research process.

(2) Data Collection:

In this study, data are collected firsthand through field observations, classifying them as primary data. Images of meals are captured in two conditions—before and after consumption—in the university dining hall to enable food waste estimation. To enhance accuracy and generalizability, data are recorded under various lighting conditions, camera angles, and plate types. Particular attention is given to standardization (e.g., maintaining a nearly constant distance from the plate, ensuring proper resolution, and using a controlled background) as well as ethical considerations (e.g., removing any images containing faces or personal information of students, if present). The outcome of this step is a well-prepared dataset forming the foundation of the research.

(3) Data Labeling:

In this step, the collected images are meticulously labeled to make them suitable for semantic image segmentation. Data labeling is performed using the platform www.Roboflow.com, which allows the creation of precise pixel-level masks for different food components (e.g., rice, stew, potatoes, etc.). This phase is time-intensive because the accuracy of the masks directly influences the model’s performance. After labeling, the dataset is divided into training and testing subsets to prevent any data leakage, ensuring that the model remains unbiased and that test data are not exposed to training patterns. It is important to note that data augmentation (e.g., rotation, brightness adjustment, cropping, or filtering) is applied only after this division and exclusively to the training set. This fair separation guarantees an objective evaluation of the deep learning model.

(4) Model Implementation and Data Analysis:

At this stage, appropriate deep learning architectures—both standard versions found in the literature and modified or optimized versions—are selected and implemented. This process involves designing the entire pipeline, selecting suitable components (e.g., optimizer, loss function), tuning hyperparameters such as the learning rate and its scheduling policy, and applying various enhancements to the loss function or architecture to achieve higher accuracy without significantly compromising inference speed. The ultimate goal is to develop a model that is both accurate and efficient, enabling real-time applicability in real-world environments.

(5) Validation and Iterative Improvement:

In this step, the trained models are evaluated using both standard quantitative metrics from the literature and custom-defined metrics tailored to the problem at hand (if applicable). Based on this evaluation, areas for improvement are identified, leading to an iterative feedback process aimed at achieving optimal performance. This cycle continues until a model with a balanced trade-off between accuracy and processing speed is achieved.

(6) Documentation of Results:

Finally, all results are systematically documented. This includes writing the thesis chapters, preparing analytical tables and figures, and comparing different models and their respective versions. It is noteworthy that the entire research process generally follows the CRISP-DM cycle (Cross-Industry Standard Process for Data Mining), illustrated in **Figure 1-1**. which represents a widely adopted framework in data mining–related processes.

1.5.Research Hypothesis

This study does not include formal hypotheses; however, it is based on several assumptions outlined below. In this research, to estimate food consumption and waste using images, the post-consumption state is compared with the *average* pre-consumption meal rather than paired before-and-after plates. This simplifying assumption aligns with real-world conditions, as accurately tracking individual plates in large-scale dining facilities such as university cafeterias is nearly impossible, or, even with advanced image processing, would require substantial computational resources. Therefore, this simplification is both reasonable and practical.

Moreover, meals served to students in the university cafeteria are nearly identical in quantity and composition for everyone. Consequently, the pre-consumption state (with minor tolerance) can be considered uniform across individuals, meaning that comparing each post-consumption plate with the average pre-consumption image is almost equivalent to comparing it with that individual's own pre-consumption plate. This rationale further supports and justifies the assumption, which will be discussed in more detail in the methodology section.

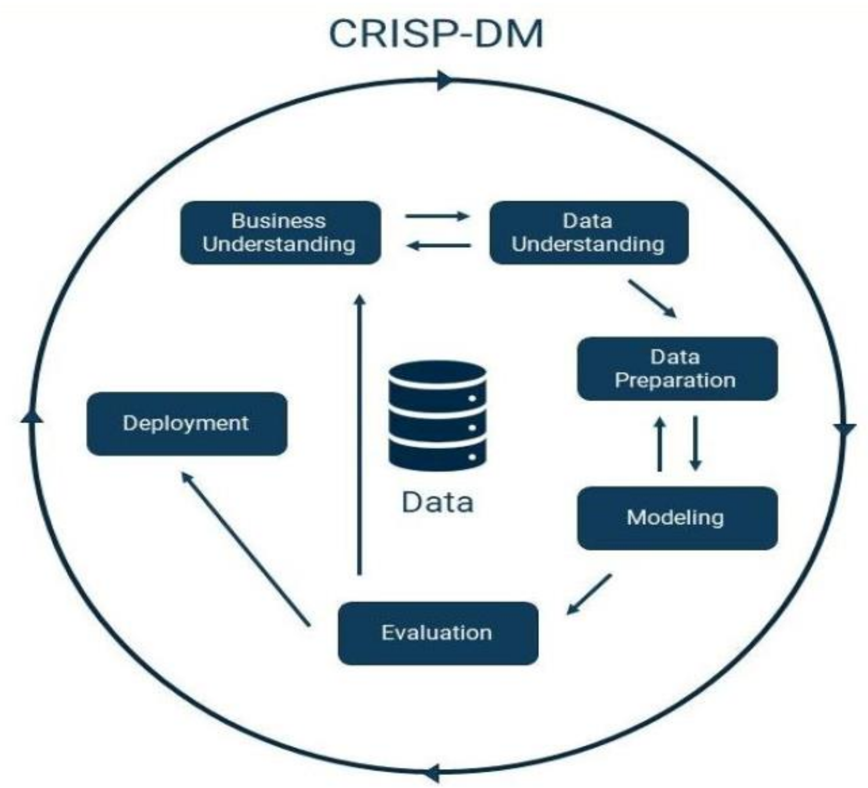

Figure 1.1. The CRISP-DM Process (Schröer et al., 2021)

1.6.Research Applications

The findings of this study can be utilized at various managerial, operational, and policy-making levels. Developing an intelligent system for the automatic estimation of food waste can assist food service managers in accurately assessing food consumption and leftovers, thereby enabling optimal planning for meal preparation and distribution. Furthermore, the results of this research can help reduce costs associated with food waste, promote more efficient resource utilization, and ultimately enhance environmental sustainability. From a social perspective, providing precise data on consumption and waste patterns can contribute to revising nutritional policies and designing educational interventions aimed at reducing food waste. From a scientific standpoint, this research offers a foundation for developing and evaluating advanced models in deep learning and image processing, serving as a basis for future studies in related fields.

1.7.Research Contributions

The innovations of this study can be presented across several distinct yet complementary levels.

First, the primary contribution lies in employing image processing based on deep learning for a real-world application—estimating food waste. Instead of traditional methods such as weighing or those relying on manual labor, this approach leverages computer vision to analyze the average patterns of pre- and post-consumption images of meals, thereby estimating the overall and component-wise food waste. This method introduces meaningful simplification and automation, representing a tangible and practical application of deep learning and image processing in real-life contexts.

Second, beyond field data collection, preprocessing, and data refinement, the study emphasizes the design and development of an operational framework grounded in deep learning and image processing—constituting a central innovation of this research. Within this framework, both established models from the literature and customized lightweight versions—with fewer parameters and structures tailored to the nature of the problem and dataset—have been designed. These customized architectures aim to increase training and inference speed, reduce computational load, mitigate overfitting, and enhance potential performance. Such architectural personalization improves efficiency and practicality for real-world applications, while systematic comparisons with conventional models provide transparent insights toward achieving the study's objectives.

Third, beyond model architecture, key components of the learning process within the framework have also been adapted—ranging from the loss function and optimizer to hyperparameter tuning—using a goal-oriented strategy for sustained performance improvement. Alongside standard image segmentation metrics, a customized metric aligned with the research objective has been designed to more precisely and fairly assess the proportion of food components and waste. This element also constitutes a technical innovation within the proposed framework.

Finally, through comprehensive analysis of results, identification of limitations, and discussion of findings, the research offers a scientific–practical framework that integrates theoretical and applied perspectives. This hybrid framework holds strong potential for intelligent food waste management and sustainability enhancement. The deep learning and image processing–based framework thus forms the core innovation of this study, providing a scalable foundation for the implementation of real-time monitoring and decision-support systems.

1.8.Chapter Summary

This chapter first introduced the preliminary concepts of the research topic and examined the problem from multiple perspectives to provide greater clarity. It then discussed the objectives of the study and presented the research questions guiding the investigation. A brief overview of the research methodology and its main stages was also provided. Finally, the chapter outlined the practical applications and key innovations of the study.

Chapter 2

Literature Review

2.1. Introduction

This chapter provides a detailed review of previous studies related to food waste, deep learning, and artificial intelligence, as well as their integration within the food industry. It also highlights the most relevant research works in this domain. Finally, a literature review table is presented to identify the research gap that this thesis aims to address, thereby emphasizing the necessity and importance of conducting this research.

2.2. Food Waste & Its Importance

According to the literature, food waste is undoubtedly one of the most complex and multidimensional challenges of the 21st century, with far-reaching environmental, economic, and social implications.

From an environmental perspective, enormous resources such as water, energy, soil, and human labor—invested throughout food production and distribution—are lost through wastage. The decomposition of food waste in landfills produces greenhouse gases such as methane, which accounts for a significant portion—between 8% and 10%—of total global greenhouse gas emissions (Sarangi et al., 2024). Some researchers have even equated the environmental impact of food waste with that of coal-based power plants. Therefore, reducing food waste is not only an environmental necessity but also a strategic measure to alleviate pressure on natural resources (Kohli et al., 2024).

From an economic perspective, food waste represents the destruction of value added throughout the entire supply chain—from production to consumption. This loss harms both producers and consumers and imposes significant costs on waste collection and management systems. The global economic value of food waste is estimated in the trillions of dollars annually, and when hidden environmental costs are included, the figure rises to several thousand billion dollars. Reducing waste in the food service sector can enhance productivity, generate positive economic outcomes, and increase employment. Moreover, studies in the European Union have shown that reducing food waste has a notably positive effect on economic growth and job creation (Lekavičius et al., 2023; Sarangi et al., 2024).

From a social perspective, food waste has profound ethical and cultural implications. Although the world produces enough food for everyone, billions of people still suffer from hunger and food insecurity. The massive loss of edible food, therefore, reflects systemic inequality and poses a significant ethical and psychological burden, particularly on vulnerable communities (Seberini, 2020).

Alarmingly, most food waste occurs at the consumption stage, especially in large-scale catering facilities such as restaurants, university dining halls, hotels, hospitals, and other institutional food providers. Due to high volumes, variable conditions, and monitoring difficulties, waste generation happens rapidly and on a large scale. Research emphasizes that focusing on monitoring and reducing waste at this stage of the food supply chain can significantly mitigate economic and environmental pressures while enhancing the sustainability of the overall food system (Faezirad et al., 2021a, 2021b).

2.3. Deep Learning & Neural Networks

Deep learning is a subfield of machine learning and artificial intelligence that is built upon multilayer artificial neural networks. Its core idea is to mimic the brain's information processing mechanisms through layered structures composed of artificial neurons. By increasing the number of layers and network depth, these models can extract more abstract and hierarchical representations of data. The first generation of such networks, as illustrated in **Figure 2-1**, emerged as multilayer perceptrons (MLPs), consisting of input, hidden, and output layers. Each neuron in an MLP computes a weighted linear combination of its inputs and passes the result through a nonlinear activation function to enhance the model's capacity for approximating complex functions (Sarraf et al., 2021).

The learning process in these networks is based on error backpropagation and stochastic gradient descent optimization. In this approach, the network's output is first compared with the target value to compute the error, which is then propagated backward through the layers using the chain rule to obtain weight gradients. By iteratively updating these weights, the network gradually learns the hidden patterns within the data. Although MLPs performed well on simple problems, their limitations became evident with high-dimensional data such as images, text, or temporal sequences, highlighting the need for more specialized architectures (Sarraf et al., 2021).

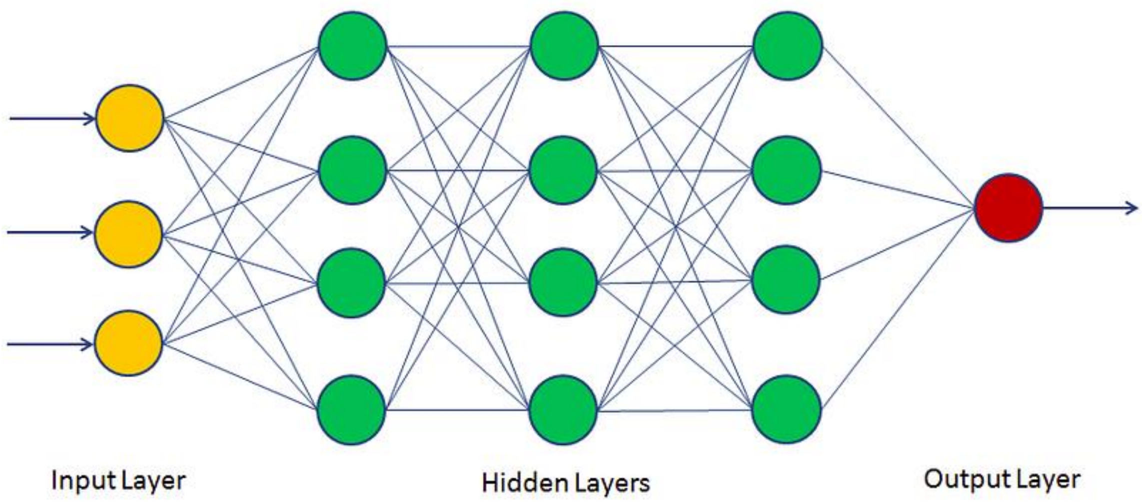

Figure 2-1. Structure of a Multilayer Perceptron (MLP) Network

Following these limitations, specialized neural network architectures were developed for different data domains. Convolutional Neural Networks (CNNs) were introduced to process spatially structured data such as images. By employing *convolutional* and *pooling* layers, CNNs effectively extract local and hierarchical features. In contrast, Recurrent Neural Networks (RNNs) were designed for sequential data, enabling the modeling of temporal and contextual dependencies. Although RNNs initially suffered from issues such as vanishing gradients, the development of architectures like LSTM and GRU helped to alleviate these problems. Moreover, Generative Adversarial Networks (GANs)—comprising a *generator* and a *discriminator*—revolutionized the generation of synthetic and realistic data. Subsequently, the emergence of Transformers and the attention mechanism introduced a novel paradigm, particularly in Natural Language Processing (NLP) and, more recently, in computer vision, allowing the learning of long-range dependencies and efficient parallel computation (Celard et al., 2023).

In this study, the main focus is on Convolutional Neural Networks (CNNs) (**Figure 2-2**), as the image-based data used requires a model capable of automatically identifying and analyzing spatial and local features. The CNN architecture consists of three main types of layers:

1. **Convolutional layers**, which apply small filters over the image to extract features ranging from low-level edges to high-level abstractions;
2. **Pooling layers**, which reduce feature dimensionality and improve robustness against minor variations; and
3. **Fully connected layers**, which integrate the extracted features to perform final tasks such as classification (Chauhan et al., 2018).

The learning process in CNNs, similar to MLPs, is based on backpropagation; however, due to shared filter weights, the number of learnable parameters is significantly reduced. This characteristic makes CNNs both accurate and computationally efficient. Furthermore, the hierarchical feature extraction capability of CNNs allows the model to detect edges and textures in early layers and progressively understand more complex objects and spatial relationships in deeper layers (Bajlan et al., 2021).

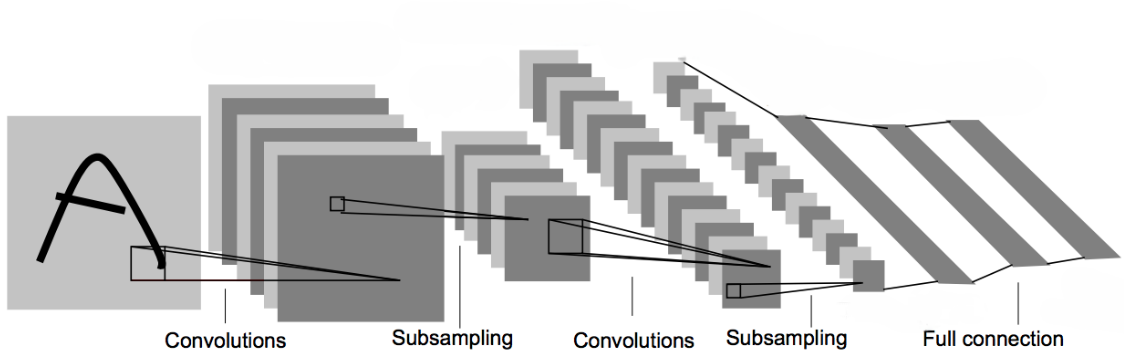

Figure 2-2. Structure of a Convolutional Neural Network (CNN)

It is worth noting that the applications of Convolutional Neural Networks (CNNs) extend well beyond image classification, encompassing tasks such as object detection and image segmentation—both semantic and instance-based. These capabilities make CNNs an ideal choice for studies focused on identifying and quantifying components within images. In the context of food waste analysis, such properties enable the accurate detection of food items and the measurement of the remaining or consumed portions, directly aligning with the objectives of this research (Chauhan et al., 2018).

Overall, the evolution of neural networks—from early MLPs to advanced architectures such as CNNs, RNNs, GANs, and Transformers—illustrates a progressive adaptation to the diverse nature of data and computational demands. Among these architectures, CNNs remain the backbone of computer vision, serving as the foundation for many successful image processing studies. These characteristics underpin their selection as the core framework of the present research for monitoring and estimating food waste based on image data (Chauhan et al., 2018; Sarraf et al., 2021).

2.4.Image Processing & Computer Vision

It can be asserted that computer vision and image processing are two closely related fields sharing the common goal of extracting and interpreting meaningful information from visual data. Image processing primarily focuses on low-level operations such as quality enhancement, filtering, noise removal, edge detection, and geometric transformations. In contrast, computer vision operates at a higher level of abstraction, aiming to understand content and relationships—essentially answering the questions *"what is where and doing what?"* Prior to the advent of deep learning, these tasks largely relied on handcrafted features and classical classifiers. The emergence of deep neural networks transformed this paradigm: through hierarchical representation learning, models now automatically extract suitable features directly from raw data, achieving accuracies that surpass traditional approaches in many tasks (Guo et al., 2022; Sarraf et al., 2021).

The range of problems in this domain is vast, yet three main task families are recognized as its foundational pillars: image classification, object detection, and image segmentation. In classification, the entire image is assigned a single label. Object detection goes a step further by identifying not only *what* is present but also *where* each object is located through bounding boxes. Segmentation, the most fine-grained approach, assigns a category label to each individual pixel, offering a detailed understanding of spatial structure. **Figure 2-3** illustrates a comparative overview of these three major tasks (Alzubaidi et al., 2021; Assad et al., 2024).

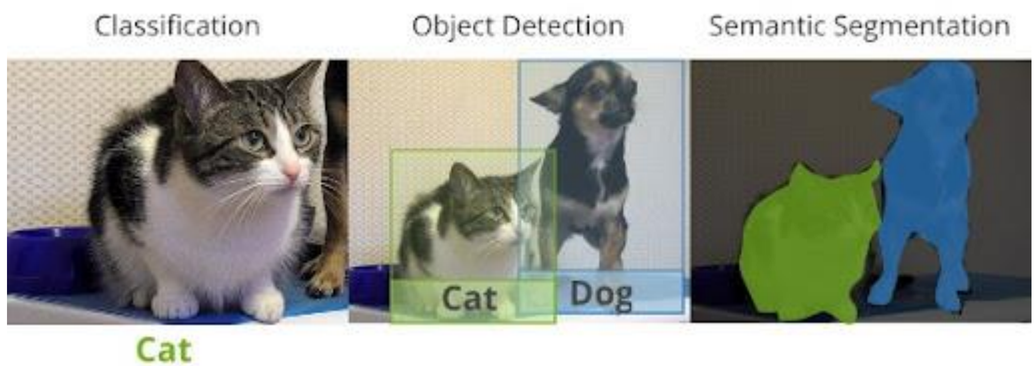

Figure 2-3. Comparison of Image Classification, Object Detection, and Semantic Segmenatiomn

In addition to the aforementioned tasks, two major approaches exist within image segmentation: semantic segmentation and instance segmentation. In *semantic segmentation*, all pixels belonging to the same class are labeled identically, without distinguishing between individual instances. In contrast, *instance segmentation* not only assigns class labels but also differentiates each distinct instance within a class. This distinction is illustrated in **Figure 2-4.**

For greater clarity, the present study focuses on semantic segmentation, as it sufficiently serves the purpose of estimating food waste based on the surface area of components remaining on the plate. In this context, it is only necessary to differentiate food categories based on their semantic type—for example, distinguishing *potatoes* from *chicken* or *rice*—without the need to separate individual rice grains or potato sticks. Thus, the chosen methodological framework is fully aligned with the research objective and provides a reliable foundation for inference from pre- and post-consumption images (Chauhan et al., 2018; Chen et al., 2017).

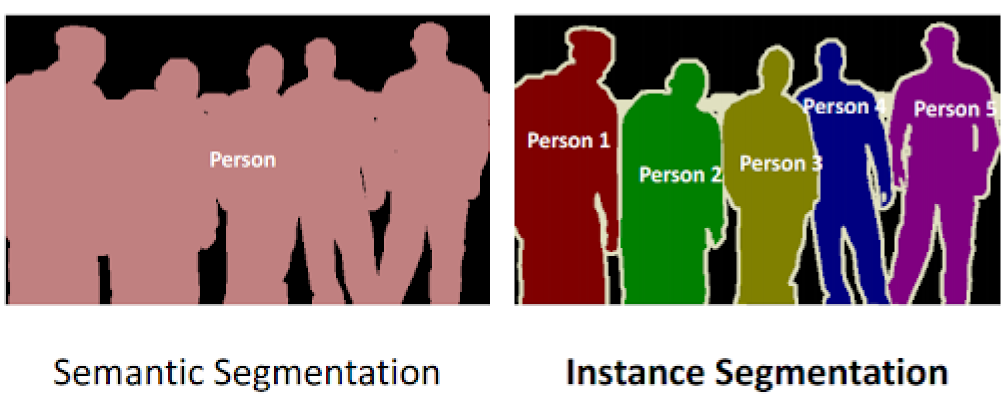

Figure 2-4. Difference between Semantic Segmentation and Instance Segmentation

The evolutionary path of image classification clearly demonstrates how models have progressed from relatively shallow networks to today's highly efficient architectures. The first major breakthroughs arose with deep convolutional architectures, followed by more structured and deeper designs that introduced residual blocks and optimization of the accuracy-to-parameter ratio. The transfer of knowledge from large-scale datasets to smaller and domain-specific ones through transfer learning became the standard, effectively mitigating data scarcity challenges. Simultaneously, techniques such as batch normalization, regularization, and learning rate scheduling enhanced training stability and convergence speed. As a result, modern models can learn hierarchical representations that range from low-level features (edges and textures) to high-level concepts (shapes and patterns) (Alzubaidi et al., 2021).

In object detection, development has transitioned from two-stage region proposal-based methods to high-speed one-stage frameworks, reflecting the continual trade-off between *accuracy and speed*. Early two-stage approaches generated candidate regions and refined them in a second stage, achieving high accuracy but at the cost of computational complexity and numerous parameters. Over time, one-stage detectors such as YOLO integrated region proposal and classification into a single pass, dramatically improving speed. By incorporating strategies such as scale alignment, hard positive mining, and balanced loss weighting, they achieved competitive accuracy. Today, object detection performs reliably in dense scenes and under variations in illumination, occlusion, and scale; however, it still lacks pixel-level precision and thus remains insufficient for fine-grained surface estimation (Alzubaidi et al., 2021).

Image segmentation, extending beyond classification and detection, provides the most detailed spatial representation. Semantic segmentation was the first major milestone, mapping every pixel to a class label—typically using fully convolutional architectures that restored spatial resolution through upsampling paths. Later advancements introduced higher-accuracy variants featuring multi-scale feature aggregation, dilated convolutions, and attention modules, which improved boundary reconstruction and fine-detail recognition. Subsequently, instance segmentation combined concepts from detection and segmentation to generate distinct masks for individual objects, addressing the challenge of separating multiple instances of the same class, as illustrated in **Figure 2-4.** This level of precision is particularly advantageous when the goal involves measuring the area of each component and comparing pre- and post-process states—valuable in scenarios demanding enhanced accuracy (Alzubaidi et al., 2021; Assad et al., 2024).

Beyond architecture, practical considerations play a decisive role. Factors such as pixel-level labeling quality, class balance, data augmentation strategy design, and proper dataset partitioning (training/validation/testing) are critical to ensuring model generalizability and avoiding data leakage. Additionally, choosing a loss function aligned with the task objective (e.g., class-weighted or boundary-focused losses) and applying effective optimization policies (such as learning rate scheduling and early stopping) are vital to achieving a balanced trade-off between high spatial accuracy and reasonable computational cost. In real-time applications, computational efficiency, model size, and hardware feasibility are equally crucial performance determinants (Agarla et al., 2023; Rokhva et al., 2024).

2.5.Related Works

This section reviews existing literature in the food industry that employs deep learning and computer vision techniques, aiming to identify the research gap relevant to this study.

Farinella et al. (2020) conducted a study titled *"Food Waste Detection through Object Recognition in the Pierce Dining Hall"* near Stevens Institute of Technology, USA. They utilized cameras and supervised deep learning methods—specifically Convolutional Neural Networks (CNNs)—to classify food waste from images. Alongside image data, variables such as the total cooked food mass, discarded mass, and the amount served from the buffet line were also used for waste analysis. Importantly, the images were employed only for *waste identification*, while *quantitative estimation* was based on mass measurements. During implementation, the research team used a Raspberry Pi and a low-cost portable sensor to automatically collect images and transmit data to a remote database. These data were subsequently used to train the neural network for recognizing food and waste categories. Although the system was successfully designed and the model effectively trained, the COVID-19 pandemic limited large-scale experimental validation. The authors highlighted that such systems can play a supportive role in food waste management and serve as a foundation for future intelligent monitoring systems (Farinella et al., 2020).

Mazloumian et al. (2020) conducted a pioneering study in Switzerland titled *"Deep Learning for Food and Waste Classification"*, which applied deep learning and computer vision techniques to food waste management. Emphasizing that nearly one-third of global food production is wasted—with significant environmental, economic, and social consequences—the researchers introduced deep learning and image processing as scalable and efficient tools for monitoring and managing food waste. Their dataset consisted of approximately half a million images of discarded food categorized into 20 food classes. The study aimed to perform both classification and segmentation of food waste, improving model accuracy for future applications, particularly in decision-support systems for waste management (Mazloumian et al., 2020).

For the classification task, the authors employed transfer learning using the pretrained VGG16 model, which accelerated convergence and improved accuracy—a crucial advantage given the large dataset size. Although VGG16 has more parameters and is less efficient compared to modern architectures, it was considered satisfactory for the 2020 context. For segmentation, the U-Net architecture was adopted, known for its precise pixel-level predictions. Due to computational constraints, only 1,000 images were used for segmentation, split into 70% training, 20% validation, and 10% testing subsets. The results were highly promising, achieving 83.5% accuracy in classifying waste categories and 98.5% pixel accuracy in distinguishing waste from non-waste regions. A key finding was that classification enables the identification of waste types, while segmentation determines their exact locations, thereby enhancing the system's practical utility. **Figure 2-5** illustrates an example of image segmentation from this study (Mazloumian et al., 2020).

Faezirad et al. (2021a) conducted a study titled *"Preventing Food Waste in a Subsidized University Cafeteria Using an Artificial Neural Network Model under Uncertainty."* The aim was to reduce waste in the dining systems of a major public university in Tehran. The importance of this topic arises from the large volume and variety of meals prepared daily in such settings, which leads to considerable food waste. The study focused on the role of meal reservation and distribution systems and utilized seven years of reservation and student behavior data from the university. The dataset, in tabular format, included information such as student attendance or absence, the number of meals reserved, received, or unclaimed, time data, and similar attributes. To predict consumption and manage waste, a hybrid model was developed comprising an Artificial Neural Network (ANN) component for deterministic demand estimation and a statistical component for uncertainty analysis. The uncertainty model considered not only the costs of consumed and wasted materials but also the potential costs of meal shortages. Results showed that the proposed model could reduce food waste by up to 79%, revealing significant potential for saving food resources, conserving natural assets, and mitigating environmental impacts (Faezirad et al., 2021a).

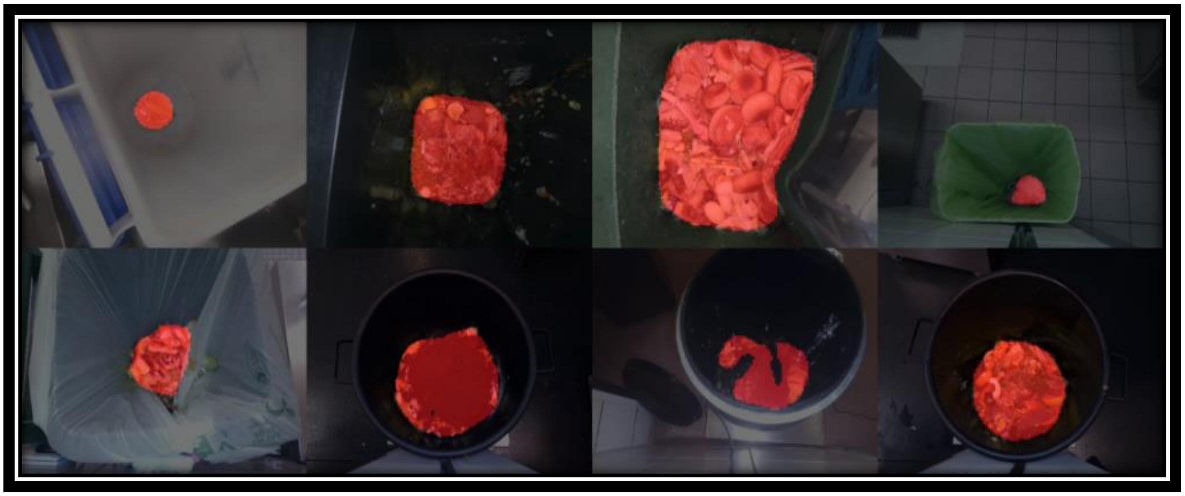

Figure 2-5. Illustration of Image Segmentation in the Study by Mazloumian et al. (2020)

In another study, Faezirad et al. (2021b) presented *"Food Waste Reduction through Demand Forecasting in University Meal Reservation Systems Using Artificial Neural Networks with Weighted Error Functions."* While conceptually similar to their previous work, this study focused specifically on accurate food demand prediction without incorporating uncertainty modeling. The researchers designed an ANN model using a diverse set of input features such as daily reservations, number of present students, day of the week, meal price, alternative meal options, accommodation status, education level, and actual meals received. The model effectively predicted deterministic demand, enabling optimization of ingredient procurement. Results indicated that the developed network achieved high prediction accuracy, leading to an 81% reduction in food waste. Moreover, the model provided feature-importance analysis, identifying key variables influencing demand patterns. In the examined university, over 23,000 kilograms of food were prepared annually, and the observed waste reduction had a remarkable effect on efficiency and budget savings. The findings highlighted the potential of neural networks as decision-support tools for optimizing raw material management, minimizing costs, and preventing food resource wastage (Faezirad et al., 2021b).

Yadav and Chand (2021) focused on food classification from images using a lightweight model, MobileNetV2, chosen for its low parameter count and fast training speed, making it suitable for mobile and real-time applications. This enables deployment even on devices lacking powerful graphical hardware. Their approach involved feature extraction through MobileNetV2 followed by classification using a Support Vector Machine (SVM). Features were extracted from three different layers of the network, and the model was evaluated using standard metrics such as accuracy, precision, recall, F1-score, and confusion matrix. The reported classification accuracies were 84%, 87.26%, and 83.6% across different layers. The F1-score reached 82.43%, reflecting a balanced trade-off between precision and recall. Other metrics also exceeded 80%, confirming the model's robustness. The authors emphasized that accuracy alone cannot sufficiently evaluate model quality, especially for imbalanced data, thus underscoring the importance of the F1-score. The findings suggested that this model could not only classify food but also assist in nutrient tracking and food waste identification, paving the way for decision-support or recommendation systems on mobile platforms (Yadav & Chand, 2021).

Geetha et al. (2022) designed a waste management system using an ensemble of neural networks, which, although not directly focused on food waste, aligns with the broader objective of waste detection, recognition, and classification through computer vision and deep learning. The main goal was to create an image-based system capable of estimating the type, size, hazard level, and volume of waste, and then transmitting this information to dispatch appropriate robotic cleaning units. The methodology integrated several computer vision techniques, including image preprocessing, object detection with bounding boxes for localization, and pixel-wise segmentation for precise boundary mapping. The system was implemented as a mobile application that could detect and classify waste and, using the phone's GPS, summon the nearest cleaning robot based on the waste category—an approach proposed within the smart city framework. The dataset consisted of 2,527 waste images across six categories: plastic, glass, metal, paper, cardboard, and organic waste. The organic class, ranging from fruit peels to infectious hospital waste, posed the greatest challenge. Images were captured from multiple angles to enhance diversity and model generalizability, and data were split 70–15–15 into training, validation, and testing sets (Geetha et al., 2022).

Several models were employed in this study, including a customized version of VGG16, YOLO-V5, and Mask R-CNN, which were used respectively for classification, object detection, and image segmentation. The sample images presented in the paper demonstrated that the R-CNN and YOLO models were capable of detecting waste objects in both scattered and densely packed environments. Examples of these results were illustrated in **Figures 2-6 & 2-7**. The findings indicated an accuracy exceeding 90% for waste detection and classification—surpassing many similar studies. Moreover, the use of multiple evaluation metrics and comprehensive comparisons with the latest research represented one of the study's key strengths. Finally, the trained models were integrated into a mobile application, performing seamlessly across tasks such as waste detection and classification, quantity estimation, and geotagging to notify the nearest cleaning unit. This demonstrated that the proposed approach was highly effective not only from a scientific standpoint but also in practical terms for smart waste management (Geetha et al., 2022).

Figure 2-6. Illustration of waste detection in a scattered environment using the **RCNN** model (Geetha et al., 2022).

Figure 2-7. View of waste detection within a pile of waste using the YOLO algorithm (Geetha et al., 2022).

Said et al. (2023) presented an article titled *"An Intelligent Approach for Sustainable Management and Valorization of Food Waste,"* which explored the potential of three technologies—Internet of Things (IoT), Machine Learning (ML), and Deep Learning (DL)—to reduce food waste and design a smart management system. At the beginning of the study, the researchers emphasized the necessity of reducing food loss to enhance the sustainability of the food supply chain. The study also highlighted the importance of food waste valorization, both at the macro level and by specific waste types, even considering the value of waste that could be prevented through modern technologies. The research was primarily qualitative and focused less on numerical or quantitative aspects. Specifically, the authors demonstrated that through IoT, data related to food loss could be collected rapidly, in real time, and in a relatively structured manner, although the quality of such data might require preprocessing and refinement. Subsequently, ML algorithms could analyze these data to provide valuable insights for researchers and policymakers, including a better understanding of the types and amounts of food loss, identification of areas with the highest waste levels, and estimation of the economic value of these losses. Such insights can be highly effective in formulating policies and practical actions aimed at reducing food waste and improving efficiency (Said et al., 2023).

Ahmadzadeh et al. (2023) also published a review article titled *"A Comprehensive Review of Food Waste Reduction Based on the Internet of Things and Big Data Technologies,"* which investigated the role of IoT, ML, and Big Data in reducing food waste throughout the entire food supply chain. The study examined food waste from production on farms and livestock facilities, harvesting, transportation, and industrial processing, to household consumption and even post-meal stages, showing that modern technologies can play a crucial role at every stage. The article emphasized that integrating IoT, Big Data, and real-time analytics can significantly reduce waste while unlocking substantial economic, social, and human opportunities. The authors proposed a conceptual four-layer system architecture, schematically shown in **Figure 2-8,** which includes: sensors for collecting environmental and waste-related data; a network layer for fast and organized data transmission; a service layer for data mining and predictive analysis using ML methods; and an application layer for presenting the results through visualization and key management indicators.

This proposed framework highlighted the importance of choosing appropriate sensors, ensuring network quality and speed, improving the efficiency of ML models, and, ultimately, enhancing the interpretability and understandability of results for decision-makers. The study also demonstrated that ML algorithms could be applied at various stages of this cycle—from data preprocessing to final analysis and visualization. Overall, the article by Ahmadzadeh et al. presented a qualitative roadmap for developing decision-support systems in the domain of food waste reduction and showed that the combination of IoT and Big Data can provide an effective infrastructure for sustainable food supply chain management (Ahmadzadeh et al., 2023).

Another study published in 2023 by Moumane et al. titled *"Food Recognition and Nutrient Estimation Using MobileNetV2 Architecture and Transfer Learning"* highlighted that food classification can serve as a fundamental step toward identifying and estimating food waste using intelligent systems, thus underscoring its significance. In this research, the MobileNetV2 model—a type of Convolutional Neural Network (CNN)—was employed to classify 190 food categories, encompassing a wide range of Western dishes, local cuisines, and beverages. The rationale for using MobileNetV2 was consistent with previous studies, emphasizing its fewer parameters, higher trainability, faster training process, and overall high accuracy, all of which make it particularly suitable for real-time analysis within intelligent decision-support systems. Finally, a mobile application was developed based on this model to classify food items with relatively high accuracy. The study emphasized the importance of food classification as a prerequisite for systems operating with deep learning frameworks, which can subsequently be extended to food waste detection and estimation tasks (Moumane et al., 2023).

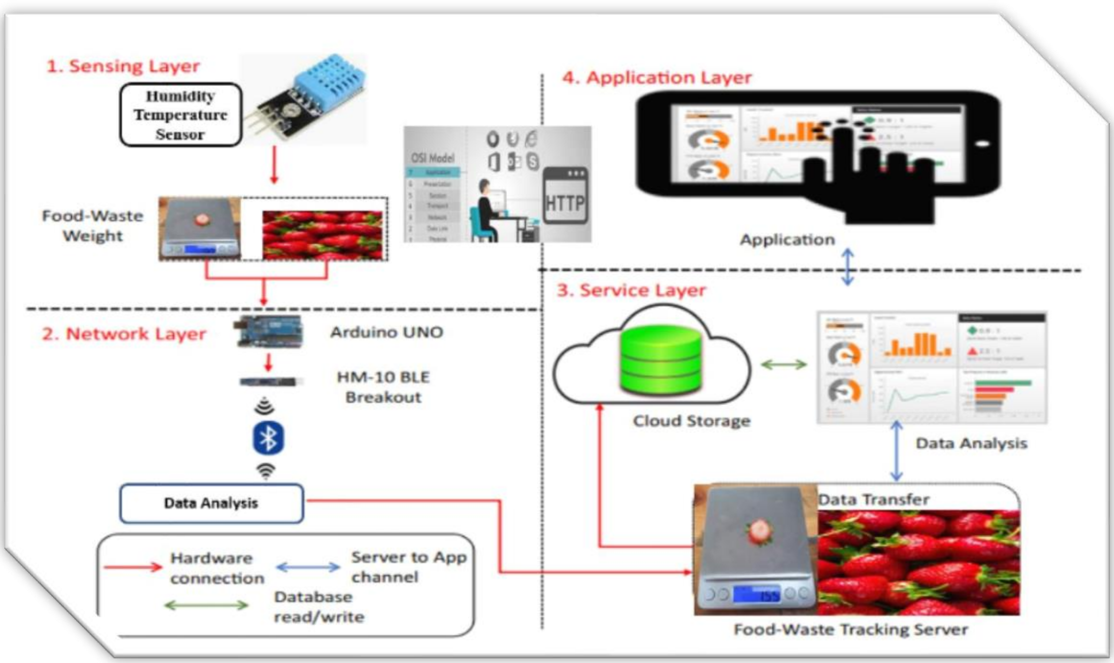

Figure 2-8. Illustration of the different layers used in the architecture of the proposed system (Ahmadzadeh et al., 2023).

In 2023, de Almeida Oroski and da Silva published a review study titled *"Understanding Food Waste Platforms: A Mini Review,"* which emphasized the role of modern technologies in supporting intelligent systems for food waste reduction, particularly within the food supply chain. The study demonstrated that digital technologies can empower management platforms to perform functions such as selling near-expiry food products, monitoring and measuring waste, and even providing education and awareness within the food industry. A key aspect of these systems lies in their decision-support capabilities, which can be applied at various stages of the supply chain. However, the authors clarified that technology alone is insufficient to effectively reduce waste. They highlighted the importance of complementary factors such as sustainable business models, circular economy principles, and continuous monitoring, all of which must work alongside technology to achieve meaningful waste reduction. Consequently, the study concluded that while smart and AI-based technologies are essential for food waste management, true progress can only be achieved by integrating them with reliable business frameworks and sustainable managerial approaches (de Almeida Oroski and da Silva, 2023).

Rokhva et al. (2024) published an article titled *"Computer Vision in the Food Industry: Accurate, Real-Time, and Automated Food Recognition Using a Pretrained MobileNetV2 – A Lightweight and Fast Framework for Automatic Food Identification."* The study aimed to develop a highly accurate and cost-effective system for real-world applications capable of maintaining performance even under limited hardware conditions. The proposed framework employed MobileNetV2 with transfer learning and fine-tuning of all parameters. The model was trained on the Food11 dataset, consisting of 16,643 images across 11 food categories, with image resolutions ranging from 32 to 256 pixels. A visual overview of these images is presented in **Figure 2-9.**

After preprocessing and normalization, data augmentation techniques—such as random rotation, horizontal flipping, and color variation—were applied to improve generalization. The model training utilized the Cross-Entropy loss function, SGD optimizer with momentum and Nesterov acceleration, and a dynamic learning rate schedule. To prevent overfitting, L2 regularization was applied. Each experiment was repeated five times per resolution to ensure result stability. The results indicated that 256×256 resolution achieved the best performance, reaching 92.97% accuracy and a processing speed of 291 images per second. Some categories, such as *Soup* and *Noodles/Pasta*, achieved near 99% accuracy, while others like *Bread* and *Dairy Products* showed slightly lower accuracy (84–87%). At lower resolutions, the model's speed increased, but accuracy dropped sharply (down to 60%). A confusion matrix analysis further revealed that most classification errors were due to visual similarities among certain food categories. To address this issue, the researchers suggested incorporating attention modules in future work (Rokhva et al., 2024).

Rokhva and Teimourpour (2025b) conducted a study titled *"Accurate and Real-Time Food Classification through the Synergy of EfficientNetB7, CBAM Attention Module, Transfer Learning, and Data Augmentation,"* aiming to enhance classification accuracy while maintaining high processing speed in food recognition tasks. The main objective of this research was to address key challenges such as high inter-class similarity, intra-class diversity, and dataset imbalance.

In this study, EfficientNetB7 was employed as the base model, leveraging transfer learning to accelerate convergence and reduce the dependence on large datasets. To improve feature extraction quality and emphasize the most informative regions of each image, the Convolutional Block Attention Module (CBAM) was integrated into the architecture, strengthening both spatial and channel-wise attention. To prevent overfitting and increase data diversity, various data augmentation techniques—including geometric and color transformations—were applied. Furthermore, careful hyperparameter tuning, such as optimizing learning rate schedules and applying regularization methods, contributed significantly to model stability.

The experiments were conducted on the Food11 dataset, and the proposed model achieved an impressive accuracy of 96.6%, the highest among all single-model approaches. Beyond accuracy, the model also demonstrated a processing speed exceeding 60 images per second, making it suitable for real-time and practical applications. Overall, the integration of the four main strategies—EfficientNetB7, CBAM, transfer learning, and data augmentation—resulted in a powerful and efficient framework for food image classification, with potential applicability across various domains (Rokhva and Teimourpour, 2025b).

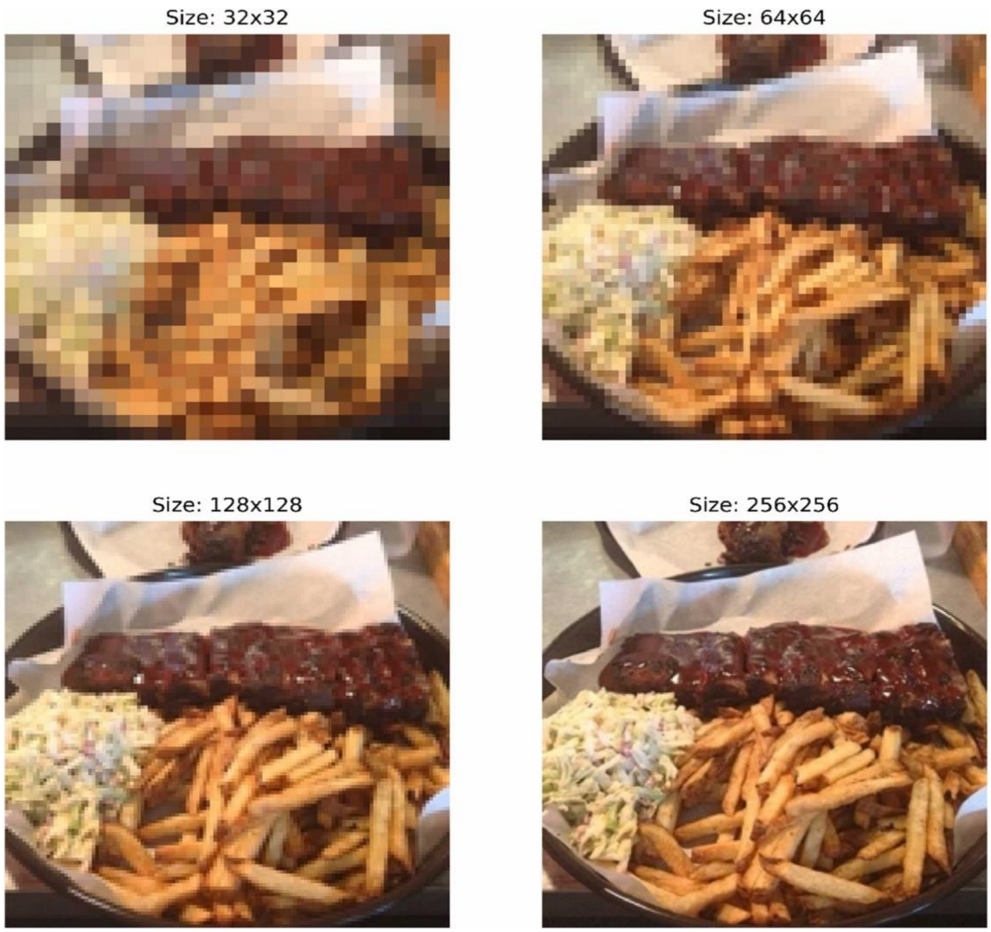

Figure 2-9. Sample images of different resolutions used in the study (Rokhva et al., 2024).

One of the most significant related studies was conducted by Lubura et al. (2022), titled *"Food Recognition and Food Waste Estimation Using Convolutional Neural Networks."* The paper first discussed the importance and necessity of automatic food recognition by computers using image processing and computer vision techniques. After an exploratory discussion, the authors demonstrated that computer vision techniques can play a crucial role in reducing food waste, with their outcomes serving as the foundation for intelligent automated systems. The study also reviewed the enormous scale of global food waste, estimated at one-quarter to one-third of all food produced worldwide, and examined its negative consequences, including economic losses, social challenges, environmental degradation, and threats to global food security. The authors concluded by emphasizing the importance of automating these processes within a smart system framework.

Regarding data, the study analyzed daily meal information of individuals—particularly students and young adults aged 20–30 years—in Serbia, covering the period from January 1 to April 31, 2022 (a six-month duration). To construct a dataset for food classification purposes, food item images were collected from the internet, forming an image database comprising 157 food categories. Each category contained 50–200 images, bringing the total to 23,552 images. The large size and diversity of this dataset reflected the generalizability and robustness of the results for the intended application. **Figure 2-10** provides a limited visual overview of the images used in this study, corresponding to the image classification section (Lubura et al., 2022).

Figure 2-10. A limited view of the images and data used in the study (Lubura et al., 2022).

Regarding the methodology, a Convolutional Neural Network (CNN) was initially developed and employed for food image classification, serving as the foundation for subsequent stages of the research. **Figure 2-11** provides an overview of the CNN architecture that was built and utilized from scratch (Lubura et al., 2022). According to the study, while the large number and diversity of food types enhanced the generalizability of the model, they also made the evaluation process highly challenging. Nevertheless, both the study and its literature review confirmed that deep learning, along with computer vision and image processing techniques, offers a highly effective approach for such tasks.

Since visual evaluation and confusion matrix construction for 150 food categories were impractical, all items were grouped into 12 broader composite categories. For instance, bananas, apples, tangerines, pineapples, and other fruits were all merged into a single “fruits” class. This consolidation allowed for an interpretable 12×12 confusion matrix to be generated.

The proposed CNN model was trained for 20 epochs, ultimately demonstrating strong classification and prediction performance for food images within their respective categories. The evaluation metrics showed that after optimized training, the network achieved an accuracy of 98.8% and a loss value of 0.102. The resulting confusion matrix confirmed that nearly all categories were predicted very accurately, with consistently high precision across all food classes (Lubura et al., 2022).

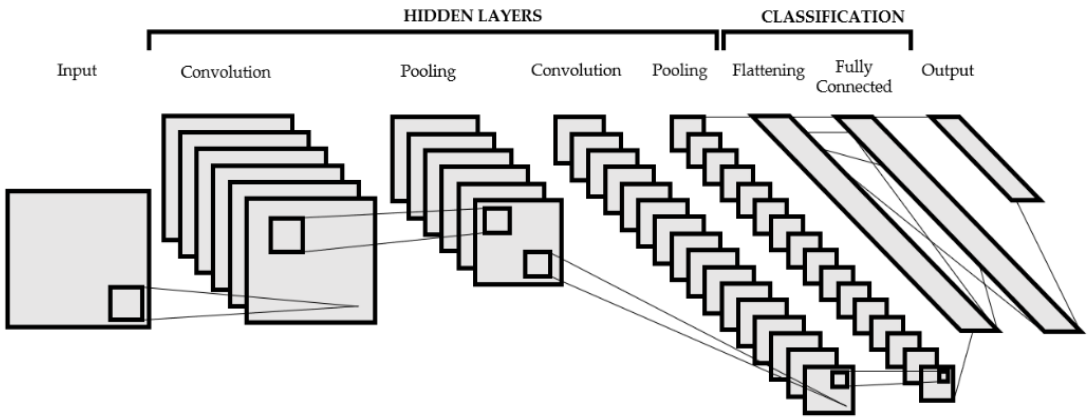

Figure 2-11. Overview of the Convolutional Neural Network employed for food image classification (Lubura et al., 2022).

Subsequently, by masking the background of the images taken before and after meal consumption and utilizing a set of optimized hyperparameters (presented in tabular form in the study's appendix), the amount of food waste was estimated by comparing the surface area of the food portions before and after eating. It is noteworthy that the analysis was based on surface estimation rather than volume, which represents one of the limitations of this study and a potential direction for future work.

According to the paper, in this stage—namely, food waste estimation through surface analysis—only a portion of the available images, precisely 1,354 images, was examined and analyzed. Based on these images, the estimated amount of waste, representing the proportion of uneaten food, was found to be 21.3%. **Figure 2-12** provides a schematic illustration of the background masking process and the surface-based estimation of food waste (Lubura et al., 2022).

Furthermore, one of the noteworthy and commendable aspects of this study was its transparent acknowledgment of limitations and challenges encountered by the image processing model—particularly in the image segmentation phase. For instance, the model produced errors in cases where spoons or forks were placed in the middle of the plate, where food containers had unusual shapes, or where plate colors and patterns closely resembled food textures, even in otherwise standard dishware. Another major limitation was the analysis of surface area rather than volume for estimating food waste (Lubura et al., 2022).

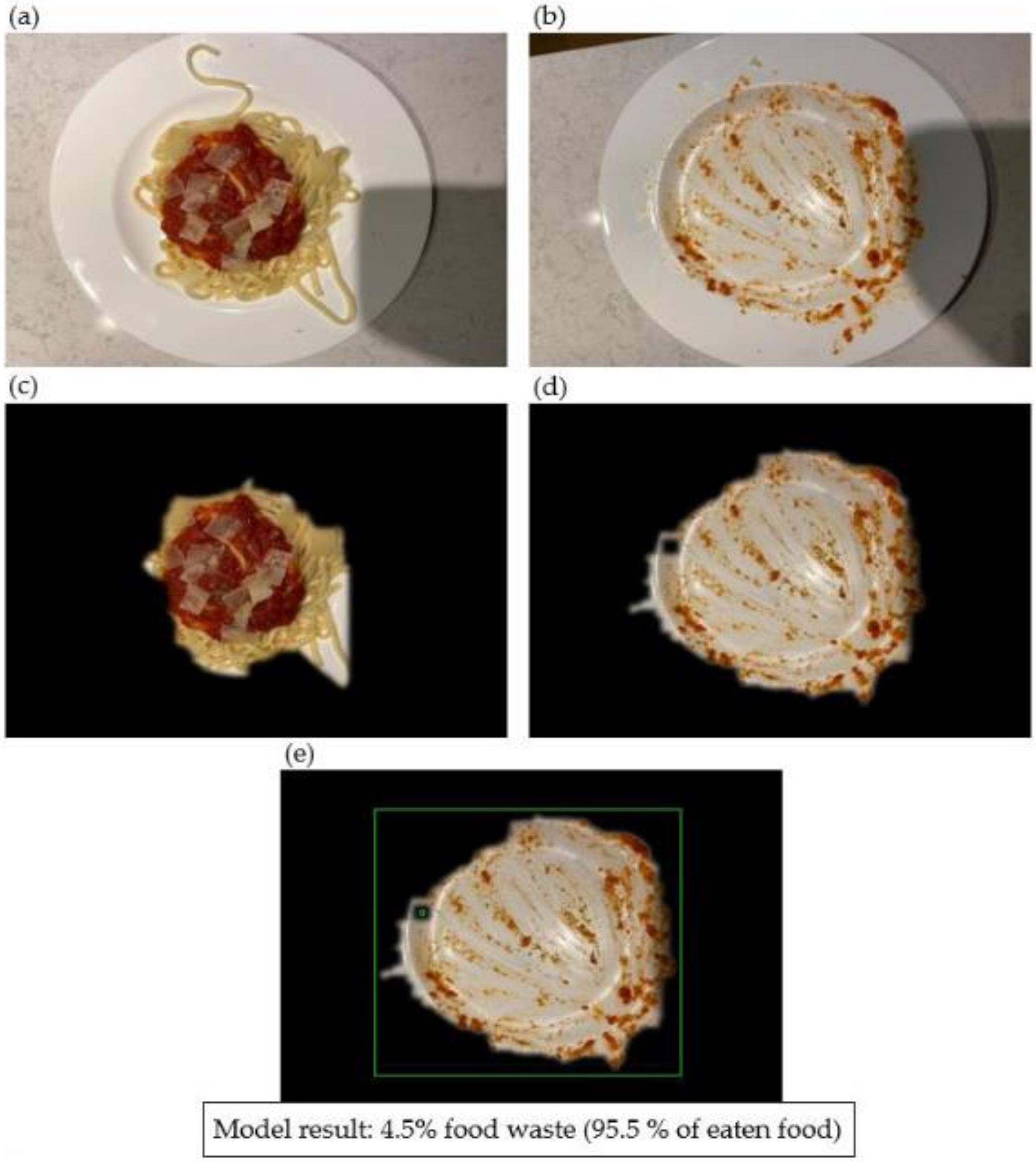

Figure 2-12. Schematic illustration of the background masking process and surface-based estimation of food waste (Lubura et al., 2022).

At this point, it is important to differentiate the present study from that of Lubura et al. (2022). In the mentioned research, food classification was performed using a customized CNN (similar to VGG16) and evaluated with various performance metrics, which were clearly and rigorously presented. However, in the waste estimation component—despite obtaining a numerical result of 21.3% food waste—several significant shortcomings were evident. Notably, no learning-based model (such as one utilizing deep learning) or image segmentation method was applied to crop and compare the before- and after-consumption images. Instead, the study merely listed a set of parameters based on which classical (traditional) image processing techniques were used to separate foreground and background, followed by a comparative analysis between pre- and post-consumption plates to estimate food waste. The complete list of these parameters is presented in **Table 2-1** (Lubura et al., 2022).

Table 2-1. Parameters used for foreground and background separation in food waste estimation (Lubura et al., 2022).

Variable	Description	Assigned Value
Blur	It affects the smoothness of the boundary between the foreground and the background.	21
Low canny	The minimum intensity threshold required for edge detection.	10
High canny	The maximum intensity threshold required for edge detection.	200
Dilation iteration	The number of dilation iterations applied during the masking process.	10
Erode iteration	The number of erosion iterations applied during the masking process.	10
Mask color	Background Mask	(0, 0, 0)

It can be stated that although this approach may fulfill the intended objective of the study, it cannot be functionally compared with intelligent deep learning models. Another major drawback is that, based on the procedures described in the study, no evaluation metric was provided to assess the accuracy of the waste estimation, which is understandable given the non-learning nature of the method. In other words, only the optimality of the parameters listed in **Table 2-1** could be compared with their counterparts; however, due to the absence of a learning model, this represents a significant limitation clearly reflected in the study. Unlike the classification section, the waste estimation phase lacked any metric to assess segmentation accuracy or foreground–background separation quality (Lubura et al., 2022).

Ferreira et al. (2025) recently conducted a study aimed at developing an innovative system for detecting and monitoring food waste on school cafeteria trays. The authors sought to bridge the gap between practical applications of deep learning models and their real-world deployment for food waste recognition using computer vision and YOLO algorithms. For this purpose, a custom dataset comprising 175 real images of plates at various meal stages was collected and expanded to 860 images through data augmentation techniques such as rotation and flipping. Annotation was performed using the Roboflow tool, allowing differentiation among food items, waste components, and non-edible objects (e.g., cutlery or napkins).

Multiple YOLO versions—v5, v8, and v11—were then compared. The results showed that YOLOv11, owing to its advanced architecture and pixel-level segmentation capabilities, achieved the best performance, particularly in detecting small food remnants and adapting to variable cafeteria conditions.

Experiments conducted on both the custom dataset and three additional reference datasets demonstrated that YOLOv11 achieved precision = 0.62, recall = 0.32, and mAP@50 = 0.343, indicating acceptable performance in identifying both food waste and non-edible objects. The main contribution of this study lies not in proposing a new model but in implementing one of the most advanced YOLO versions in a real-world educational setting. An example of food waste detection using this approach is illustrated in **Figure 2-13** (Ferreira et al., 2025).

Figure 2-13. Examples of food waste detection by (Ferreira et al., 2025).

2.6.Review Table

Table 2-2. Key studies were reviewed in this research, followed by a summary highlighting the identified research gap.

Reference	Dataset	Methodology	Findings
(Farinella et al., 2020)	Food-101 dataset from Kaggle and images from Google and Bing	CNN + MobileNetV2 + Raspberry Pi	Accurate identification of 10 food categories without loss of accuracy
(Mazloumian et al., 2020)	Half a million RGB images in 20 categories	U-Net + VGG16 + an additional layer for performance improvement	95.8% pixel accuracy in segmentation and 83.5% classification accuracy
(Faezirad et al., 2021a)	Tabular data, national university dataset, years 2013–2019	Artificial Neural Network for consumption estimation under certainty and statistical modeling for uncertainty conditions	Prediction accuracy between 73–75%; potential 79.6% reduction in food waste
(Faezirad et al., 2021b)	Tabular data, national university dataset, over 56,000 meal records (2012–2018)	Artificial Neural Network with customized weighted loss function	University cafeteria waste estimated at 10%; potential reduction of 82% with RMSE ≈ 0.02
(Yadav and Chand, 2021)	5,000 images across 10 food categories	MobileNetV2 for feature vector extraction and SVM for classification	Highest accuracy of 87.27% from intermediate layers; all accuracies above 80%
(Geetha et al., 2022)	2,527 waste images	Mask R-CNN + YOLOv5 + EfficientNet + VGG16 + SFM (ensemble)	Accuracies: detection 92%, classification 93.65%, organic vs. non-organic classification 95.6%
(Lubura et al., 2022)	23,552 waste images in 157 categories; 1,354 plate images for cropping	CNN built from scratch for image classification; classical image processing for segmentation and foreground/background separation	98.8% image classification accuracy; 21.3% waste estimation; no segmentation or waste evaluation metrics provided
(Moumane et al., 2023)	190 RGB food categories	MobileNetV2 with transfer learning	High accuracy and real-time classification speed

(Rokhva et al., 2024)	Food-11 dataset from Kaggle	MobileNetV2 + transfer learning + data augmentation + multi-size image analysis + speed evaluation	93% accuracy for 256×256 size; 60% for 32×32; real-time classification
(Ferreira et al., 2025)	175 custom images, 3 reference datasets	YOLOv5–V8–V11 with V11 performing best; data augmentation, object detection	mAP = 0.343, Precision = 0.62, Recall = 0.322
(Rokhva and Teimourpour, 2025b)	Food-11 dataset from Kaggle	EfficientNetB7 + CBAM, 256×256 image size, transfer learning	96.6% accuracy; outperforming single-model studies and close to 96.8% achieved by ensemble networks

The literature review revealed that most studies on food waste have primarily focused on food classification or component detection from images, while quantitative and precise estimation of food waste has received much less attention. Many of these works have merely identified the presence or absence of food items on plates rather than performing pixel-level analytical estimation. Although such approaches can provide valuable qualitative insights, they fail to deliver accurate quantitative measurements of leftover food. Furthermore, a substantial portion of previous research has been conducted under controlled environments or with limited and synthetic datasets, rarely addressing real-world scenarios such as university cafeterias or restaurants.

These limitations indicate a significant research gap in developing a practical, accurate, and real-time framework for automatic food waste estimation. Accordingly, the present study seeks to bridge this gap by leveraging deep learning and image processing techniques. The proposed framework advances this field not only by introducing a real-world dataset of pre- and post-consumption plate images but also by designing customized architectures and dedicated evaluation metrics to enable precise and instantaneous waste estimation. The innovation of this research lies in the integration of deep models' semantic segmentation capabilities with the operational requirements of real environments, providing outcomes that can contribute to optimizing food service processes and supporting large-scale policy decisions aimed at reducing waste and enhancing the sustainability of the food system.

2.7.Chapter summary

This chapter reviewed the existing body of research related to food waste detection and estimation using computer vision and deep learning techniques. The literature revealed that most prior studies have concentrated on food classification or object detection, while relatively few have addressed the quantitative estimation of food waste. Furthermore, many of these studies were conducted in controlled environments or relied on limited datasets, reducing their applicability to real-world contexts such as university cafeterias.

The reviewed works demonstrated the progressive use of CNN-based architectures such as VGG16, MobileNetV2, EfficientNetB7, and hybrid models incorporating attention mechanisms (e.g., CBAM) and transfer learning, leading to remarkable improvements in accuracy and processing efficiency. However, the lack of precise pixel-level segmentation and real-time estimation frameworks remains a major gap in this domain.

Based on these findings, the current research aims to fill this gap by developing a customized, deep learning–based framework capable of accurately and efficiently estimating food waste from real-world plate images. By combining semantic segmentation, data augmentation, and specialized evaluation metrics, the proposed system moves beyond simple recognition toward quantitative waste analysis. Ultimately, this approach contributes to sustainable food management, resource optimization, and policy-level decision-making for reducing food waste.

Chapter 3
Research Methodology

3.1. Introduction

In this chapter, we detail the research workflow, including the procedures for collecting the various datasets required throughout the study, the image segmentation process and mask creation, data preprocessing steps, the method for estimating food waste, and the deep learning models employed—along with their components, architectures, and other methodology-related aspects. The methodology is intentionally described at length to facilitate straightforward reproducibility by other researchers. The following chapter presents the experimental results and provides a focused discussion to thoroughly interpret and explain the findings.

3.2. Research Methodology

An overview of the workflow adopted in this study, based on the CRISP-DM methodology, is illustrated in **Figure 3-1.** Following a structured literature review, the data collection process begins. The next phase involves data preparation, which includes generating precise segmentation masks and performing data preprocessing operations such as data augmentation, normalization, and image resizing.

Subsequently, the food waste estimation method is introduced, which relies on surface analysis to approximate the amount of waste and consumption. This method is first implemented on the raw dataset, after which deep learning–based intelligent models are defined, described, and applied. During this stage, structural modifications are made to the models to improve their performance according to data volume, complexity, and the intended application, resulting in customized and enhanced architectures.

Once properly trained, these models can automate the estimation process, providing real-time waste prediction from input images. Evaluation metrics are then defined to measure model performance. The proposed models are validated across multiple hyperparameter settings, and the best-performing configurations are selected. The final optimized architectures are tested on unseen data to prevent bias or data leakage, and results are assessed using both standard evaluation metrics and custom performance criteria defined for this study. Each methodological component is explained in detail in the following sections.

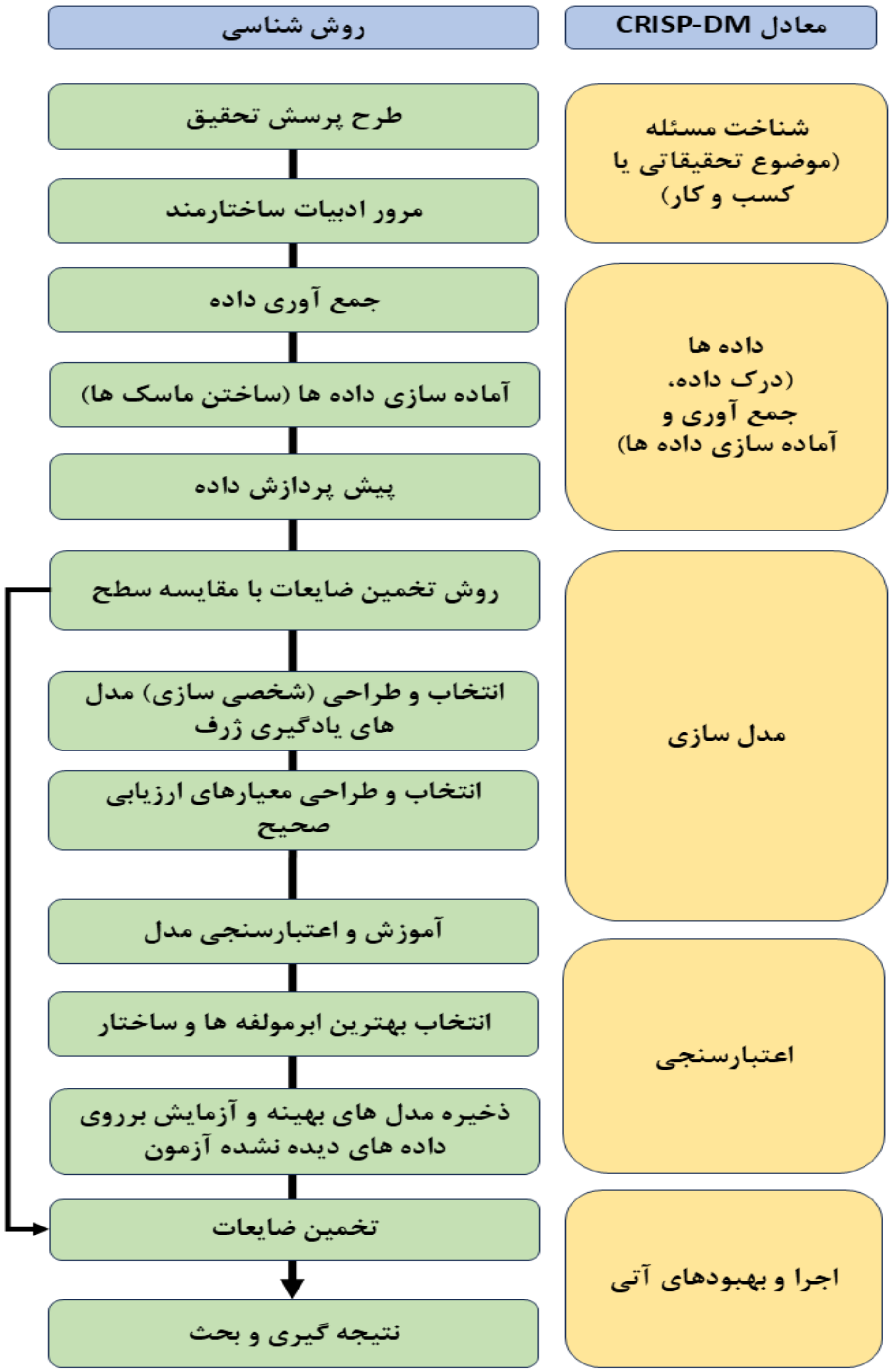

Figure 3-1. Overview of the workflow adopted in this study and its alignment with the CRISP-DM methodology.

3.3. Data

3.3.1. Data collection

The data collection process in this study was conducted at the men's dining hall of Tarbiat Modares University. The primary objective of this phase was to obtain real-world images of student meals under natural operational conditions, without any artificial intervention, ensuring that the dataset accurately reflected real-life scenarios. Given the relatively high diversity of meals—approximately 15 different dishes served over two weeks—only five of the most frequently consumed meals were selected as representative samples.

These categories included Adas Polo (Lentil Rice), Chelo Goosht (Rice with Meat), Fesenjan with Rice, Gheymeh Bademjan with Rice, and Protein-and-Fries combinations. The main reason for choosing these five dishes was their high frequency and popularity among students. For example, among stews, Fesenjan and Gheymeh Bademjan were chosen as common representatives, while Adas Polo and Chelo Goosht represented rice-based dishes. The protein-based group included dishes such as chicken schnitzel, cordon bleu, and chicken nuggets, all typically served with French fries.

Additionally, the presence of diverse structural textures—ranging from dry and semi-liquid to protein-rich dishes—was considered an advantage, as it enabled a more comprehensive evaluation of model performance and made it possible to determine which food types were more challenging for recognition and waste estimation.

During data collection, images were captured before and after meal consumption for each food category. All images were taken using a Samsung A54 smartphone, ensuring high-resolution quality and image stability. To minimize camera distance bias, photographs were taken from relatively consistent angles and distances. However, due to the dynamic and high-volume environment of the cafeteria, some variations were inevitable, which in turn added realism and diversity to the dataset.

Moreover, while maintaining authenticity, efforts were made to capture images under slightly varying lighting conditions and angles. These variations are later addressed during the data preparation, augmentation, and preprocessing stages, which aim to enhance the robustness and generalization ability of the deep learning models.

3.3.2. Data Preparation

After the initial image collection from the selected food categories, the next phase involved data preparation and the creation of semantic masks for each image. This process was essential for preparing the dataset to train deep learning–based segmentation models and represented one of the most fundamental steps of the research. All collected images were uploaded to the Roboflow platform, where the manual annotation process was carried out with high precision. For each food category, classes were defined according to the main components of the dish. For instance, in *Chelo Goosht (Rice with Meat)*, three classes—meat, rice, and background—were defined, while in *Adas Polo (Lentil Rice)*, only one main class (Adas Polo) and a background class were used.

Each image was annotated pixel by pixel, meaning that every pixel was assigned to one of the defined categories so that the segmentation models could accurately learn the boundaries between different food components and the background. The annotation was performed entirely manually, making it a highly time-consuming process, as even the smallest food residues after consumption had to be carefully labeled.

For example, in the pre-consumption images, almost all defined categories were present and clearly visible, whereas in the post-consumption images, one or more components might have been entirely consumed or only partially visible (e.g., meat fully eaten, or just small scattered fragments remaining). This issue was particularly noticeable in semi-liquid dishes such as *Fesenjan* or *Gheymeh Bademjan*, where after consumption, only smears or small scattered portions of stew remained on the plate—making boundary definition both complex and somewhat subjective.

A key methodological decision at this stage was the use of semantic segmentation instead of instance segmentation. The reason for this choice was the nature of the problem. For example, in *Gheymeh Bademjan*, all parts of the stew are conceptually identical to the model regardless of their position on the plate, and there is no need to distinguish them as separate instances. Similarly, in the *Protein and Fries* category, all fries share the same identity and are best treated as a single unified class, rather than as individual objects. Thus, semantic segmentation was the most appropriate and efficient approach, reducing model complexity while aligning with the intrinsic characteristics of the data. To improve clarity and interpretability, distinct colors were used for different classes in the masks. For instance, in the *Protein and Fries* dataset, three separate colors were applied to represent protein, fries, and background, making the class boundaries clearly visible (though this color distinction does not affect model training).

Ultimately, hundreds of images were precisely labeled with their corresponding semantic masks, forming the core dataset for training and testing the deep learning models. It can be stated that this was the most time-intensive and challenging stage of the entire project. The manual creation of detailed masks required weeks of continuous effort, but the investment was worthwhile since the quality of the semantic masks directly impacts the final model performance—and even small errors in this stage could have weakened the overall outcomes of the research.

3.3.3. Data Trasnformation & Augementation

Given that the dataset collected for this research was relatively smaller in size compared to standard benchmark datasets, data preprocessing—particularly data augmentation—was regarded as a critical necessity. This process not only enhances the performance of deep learning models (by improving accuracy) but also strengthens their generalization ability under real-world and variable conditions (e.g., different lighting intensities, camera angles, cropped or rotated images, and similar variations). The limited number of images and insufficient diversity in lighting and angle conditions could easily lead to overfitting; therefore, designing and implementing a comprehensive set of data transformations was essential to reduce such risks.

In the first step, all collected images were resized to a fixed resolution of 256×256 pixels. This dimension was chosen as a balance between visual detail preservation and computational efficiency. Increasing the resolution further would significantly raise computational costs, whereas excessive reduction would cause the loss of fine details such as rice grains or small potato pieces. Moreover, image sizes in the range of 128–324 pixels have been used in several prior studies (Bohlol et al., 2025; Lubura et al., 2022; Rokhva et al., 2024), and among them, 256 serves as both a median value and a power of two—making it more compatible with pooling and convolutional layers in deep learning architectures. In addition to resizing, pixel values were normalized to the range [0–1] to ensure stable model training and minimize fluctuations caused by differences in brightness intensity.

In the next stage, to increase data diversity and prevent models from overfitting to specific patterns, a set of random and probabilistic transformations was applied to the images. These transformations included 90-degree rotations, small random rotations (±15°), horizontal and vertical flips, brightness and saturation adjustments, exposure variations, color spectrum shifts, and partial blurring. Each transformation was designed to simulate natural variations encountered in real cafeteria conditions. For instance, brightness and exposure changes represented variations in lighting, while partial blurring simulated realistic factors such as hand tremors or imperfect camera focus.

Although this research was conducted manually and with a limited dataset, data augmentation and diversification were crucial steps, as they simulate the real-world conditions that an automated camera system might face in practical deployment—such as varying lighting, color tones, and angular perspectives. Training models on such diversified data significantly improves their robustness and performance on unseen images.

These augmentations and transformations were applied only to the training dataset, and in a probabilistic and randomized manner. This means that each image could undergo no transformation, a single transformation, or a combination of several transformations applied sequentially. This approach produced a dataset with high visual diversity without introducing unrealistic distortions. Consequently, the probability of all transformations or no transformation being applied to the same image was very low, while the wide range of possible combinations effectively tripled the size of the training data.

Finally, the dataset was split into two separate subsets: a training set and a testing set. The test set ratio for each food category ranged between 20% and 30%, ensuring both sufficient diversity in test data and an adequate number of samples for model training. The split was carefully performed to prevent data leakage—ensuring that no similar or duplicate images appeared in both training and testing sets, thereby avoiding any bias in model evaluation. The remaining augmented data were assigned to the training set so that the models could learn from a broad range of lighting, angular, and visual conditions.

Overall, the procedures described in this section played a crucial role in enhancing model stability, preventing overfitting, and improving generalization. Moreover, these steps ensured that the training conditions closely mirrored real-world environments, enabling the models to maintain robust and reliable performance when confronted with new or unseen data.

3.4. Method for Estimating Consumption and Food Waste

The main objective of this stage is to estimate the amount of food consumed and the rate of remaining (wasted) food by analyzing the areas covered by different food components and the background in each image. By calculating the proportion of each component from the segmentation masks, the food waste ratio—defined as the complement of the consumption rate—can be determined.

Since these masks were manually created by the authors, they are assumed to be highly accurate, thus serving as a reliable reference for training AI models and evaluating segmentation performance. This also eliminates the need for repeated manual labeling.

To estimate food waste, let us assume that a plate contains n categories, indexed from 0 to n–1, and that the input images have dimensions of 256×256 pixels. The proportion of category i, where $i \in \{0, \ldots, n-1\}$, is calculated using **Equation (1-3).** Using this formula, the proportion of each category is computed for all images of a given food type (both before and after consumption). These values can then be averaged across all plates of the same food category to obtain overall quantitative estimates.

$$\text{Class } i \text{ proportion} = \left(\frac{\text{Total pixels of class } i}{256 \times 256}\right) \times 100 \tag{1-3}$$

In the pre-consumption state, the portions of food served to students generally fall within a consistent range, although some variation may exist in their spatial distribution on the plate. Therefore, it is expected that the distribution of each category remains within a reasonable range. Nevertheless, whether or not this assumption strictly holds, the weighted average proportion of each category in the pre-consumption state can be used as a reference metric (as applied in this study).

Accordingly, the consumption rate for a specific category within a given meal is obtained by comparing its surface area after consumption with the average reference value before consumption. Based on this comparison, the consumption rate of each category (excluding the background) is calculated using **Equation (2-3).**

$$\text{Eating Rate}_i = \left(\frac{P_i^{pre} - P_i^{post}}{P_i^{pre}}\right) \times 100 \tag{2-3}$$

After calculating the consumption rate for a specific category within an individual plate, the average remaining rate for that category is obtained by averaging these values across all corresponding post-consumption plates of the same food type, as expressed in **Equation (3-3).** In this equation, N denotes the number of images/masks used in the calculation.

$$\overline{\text{Eating Rate}}_{\text{i}} = \left(\frac{1}{N}\sum_{\text{j}=1}^{N} \text{Eating Rate}_{\text{i}}^{(\text{j})}\right) \tag{3-3}$$

Since the remaining rate is the direct complement of the consumption rate, the average remaining rate is calculated using **Equation (4-3).**

$$\overline{Remaining\ Rate}_i = 100 - \overline{Eating\ Rate}_i \tag{4-3}$$

3.5. Deep Learning Models

This section introduces the deep learning models employed in this research for food image segmentation. Specifically, four image segmentation architectures based on the widely recognized U-Net and U-Net++ frameworks are presented, each designed to generate precise segmentation masks. These models enable accurate prediction while minimizing dependence on manual labeling, thereby allowing automated estimation of food waste.

As a baseline, the original U-Net architecture (Ronneberger et al., 2015) was adopted due to its well-established simplicity, interpretability, and effectiveness. Next, to address the computational requirements of real-time food segmentation, which demands rapid processing, and considering the relatively small dataset size in this study compared to large-scale datasets, a lightweight version of U-Net was developed. Models with fewer parameters are advantageous in such scenarios, as smaller datasets increase the risk of overfitting, and lighter architectures can help mitigate that risk.

This modified version retains the structural framework of the standard U-Net but reduces the number of convolutional filters in each stage. Specifically, while the standard U-Net employs 64, 128, 256, 512, and 1024 filters at successive depths, the lightweight version uses 32, 64, 128, 256, and 512 filters, respectively. These architectural modifications—illustrated in **Figure 3-2**—significantly reduce the total number of parameters and increase processing speed.

However, the reduced representational capacity of the lightweight model may (though not necessarily) limit its performance on more complex segmentation tasks. Conversely, in simpler tasks and in the absence of large-scale datasets, this reduction can actually improve generalization on test data by decreasing model complexity and minimizing overfitting. This trade-off will be further examined and discussed in the Results section during model evaluation.

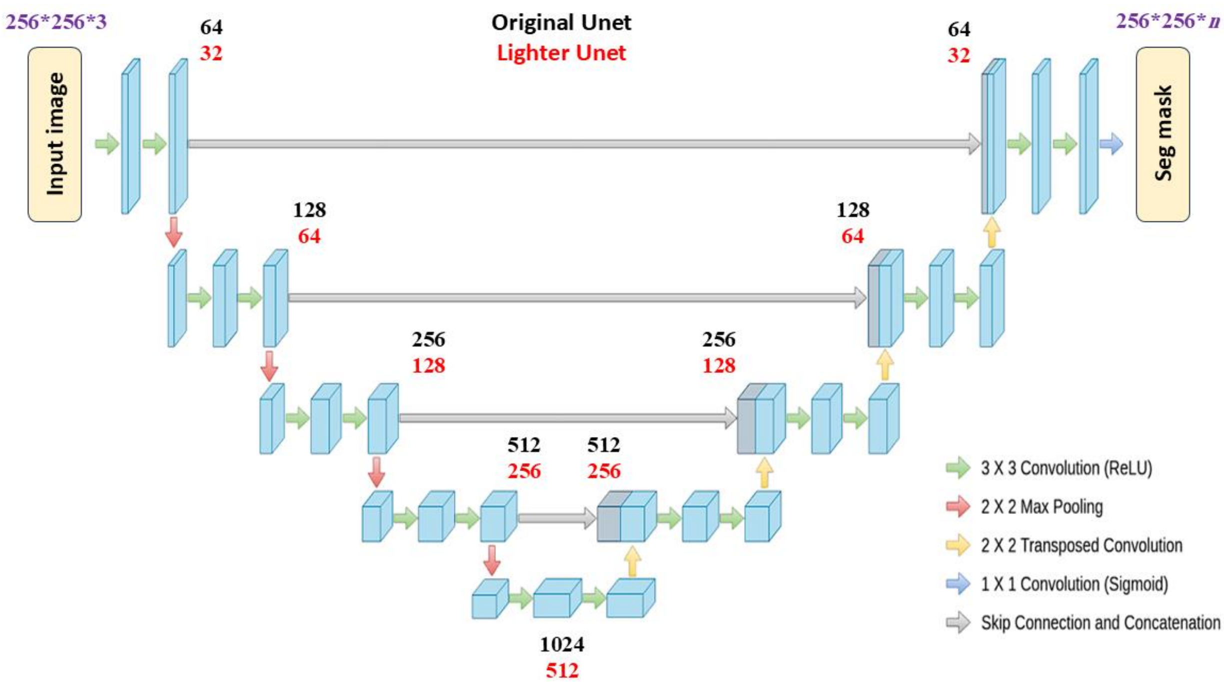

Figure 3-2. Comparison between the original U-Net and the developed lightweight version.

In addition to the standard U-Net, U-Net++—also known as NestNet—was included in this study due to its enhanced capability to preserve fine-grained segmentation details through nested skip connections. Although these nested structures can introduce greater complexity in hyperparameter tuning, they may (though not necessarily) offer improved performance in more challenging segmentation tasks.

Accordingly, both the original version and a lightweight, reduced-depth version of U-Net++ are proposed in this research, as illustrated in **Figure 3-3**. Furthermore, to compare the architectural complexity of the models, **Table 3-1** presents the number of parameters for each proposed model along with the tensor shape at the bottleneck layer.

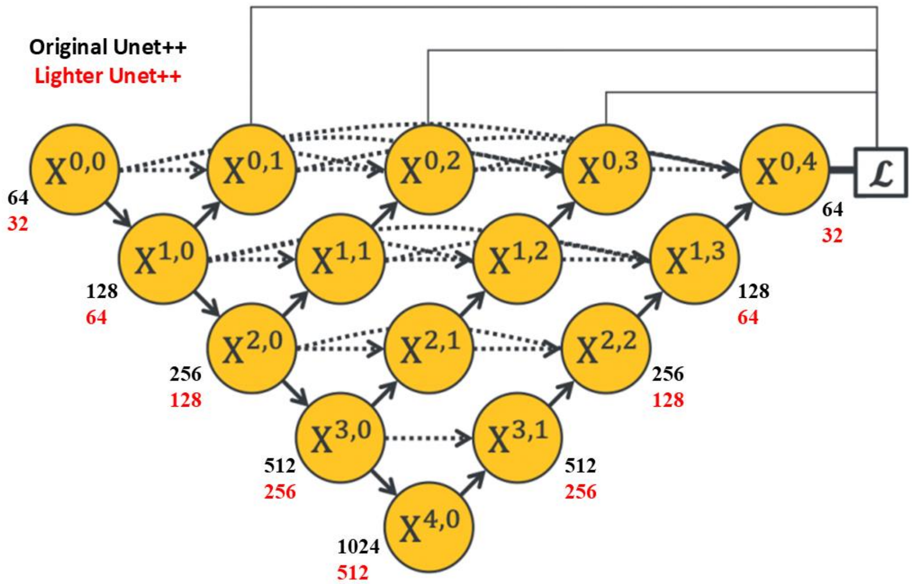

Figure 3-3. Comparison between the original U-Net++ and the developed lightweight version.

Table 3-1. Comparison of the models, their parameter counts, and the final tensor shapes at their bottleneck layers.

Model	Tensor shape at bottleneck	Param counts
U-Net	16*16*1024	31.04 m
Light Unet	16*16*512	7.77 m
Unet++	16*16*1024	36.15 m
Light Unet++	16*16*512	9.04 m

3.6. Components and Hyperparameter Configuration

3.6.1. Customized Loss Function

To address the class imbalance challenge in semantic segmentation tasks, we employ a modified dynamic weighting strategy within the cross-entropy loss function. While the standard cross-entropy assumes equal contribution from all classes, this assumption fails when pixel distributions are highly imbalanced—for example, when the background or a dominant class like rice occupies most of the image area in food segmentation tasks.

In such cases, minority classes may be underrepresented during training, leading to poor learning performance and low segmentation accuracy. This problem becomes particularly evident in post-consumption images, where only small portions of food remain and the proportion of certain classes relative to the background is extremely small. Therefore, a loss function capable of handling this imbalance is essential.

Our proposed solution is based on a dynamic weighting approach, where the weight of each class is computed inversely proportional to its frequency within each batch. If fcf_cfc represents the number of pixels belonging to class ccc and CCC denotes the total number of classes, the raw inverse-frequency weight for class ccc is defined according to **Equation (5-3)**

$$w_c = \frac{1}{f_c + \varepsilon} \tag{5-3}$$

For numerical stability, a small constant ε\varepsilonε is added. Overall, this approach enhances the learning of minority classes by assigning them greater importance during training. However, it also introduces a potential drawback: when certain classes contain only a very small number of pixels, their corresponding weights can become excessively large, leading to over-penalization or gradient explosion. To prevent this issue, a capping mechanism is introduced to limit the range of class weights, ensuring stable training. This adjustment is expressed in **Equation (6-3).**

$$w_c^{(capped)} = min\left(w_c, min_j\left(w_j\right) \cdot \lambda\right) \tag{6-3}$$

Here, λ∈R+\lambda \in \mathbb{R}^+λ∈R+ is a hyperparameter that defines the maximum ratio between the largest and smallest class weights (in our experiments, λ=10.0\lambda = 10.0λ=10.0 was used). This ensures that extremely rare classes do not exert a disproportionate influence on the total loss value, while still maintaining a meaningful prioritization during loss computation.

After applying this upper-bound constraint, the weights are normalized to establish a balanced overall contribution among all classes, as expressed in **Equation (7-3).**

$$\hat{w}_c = \left(w_c^{(capped)} / \left(\Sigma_{(j=1)}^{C} w_j^{(capped)}\right)\right) \cdot C \tag{7-3}$$

Finally, the loss function is computed as the pixel-wise weighted cross-entropy, as formulated in **Equation (8-3).**

$$\mathcal{L} = -(1/N) * \Sigma_{(i=1)}^{N} \hat{w}_{(y_i)} \cdot log\left(p_{(y_i)}\right) \tag{8-3}$$

In this formulation, N denotes the total number of pixels, yi∈{0,1,…, C−1}y_i \in \{0, 1, \ldots, C-1\}yi ∈{0,1,…, C−1} represents the true label of pixel i, and p(yi)p_{(y_i)}p(yi) is the predicted probability from the softmax activation function corresponding to the correct class.

This capped inverse-frequency formulation achieves a balance between sensitivity to minority classes and overall loss stability. It prevents rare classes from dominating the training process while ensuring they are not overlooked. Such a property is particularly valuable in real-world applications like food segmentation, where small yet semantically significant regions must be recognized without disrupting the training dynamics of the model.

3.6.2. Suitable Optimizer

The initial optimizer selected for this study was Adam, due to its ability to adaptively adjust the learning rate for each parameter and its momentum-like behavior achieved through exponential moving averages of gradients. These properties allow Adam to achieve faster convergence and make it less sensitive to hyperparameter tuning compared to traditional optimizers such as SGD.

However, since one of our objectives was to apply L2 regularization to mitigate overfitting, the standard Adam optimizer exhibited certain limitations in this regard. In Adam, the weight update rule is defined as shown in **Equation (9-3).**

$$\theta_{(t+1)} = \theta_t - \eta \cdot \left(m_t / \left(\sqrt{(v_t)} + \varepsilon\right)\right) \tag{9-3}$$

When L2 regularization is applied in this formulation, it becomes coupled with the gradient, thereby being influenced by Adam's adaptive learning rates. This interaction can make the regularization effect inconsistent and less effective across parameters, particularly in deep architectures where parameter variances differ substantially.

To overcome this limitation, we employed the AdamW optimizer, which decouples the weight decay term from the gradient update and applies it directly to the weights. This approach ensures more stable and uniform regularization across all parameters. The update rule for AdamW is presented in **Equation (10-3).**

$$\theta_{(t+1)} = \theta_t \cdot (1 - \eta \cdot \lambda) - \eta \cdot \left(m_t / \left(\sqrt{(v_t)} + \varepsilon\right)\right) \tag{10-3}$$

This decoupling ensures that the weight decay functions as intended, remaining independent of the adaptive gradient scaling. As a result, training becomes more stable and yields better generalization. Therefore, in this study, AdamW was selected as the final optimizer, combining the advantages of fast convergence with effective regularization.

3.6.3. Definition and Configuration of Components and Hyperparameters

Accurate definition and configuration of components and hyperparameters are essential for achieving stable convergence, efficient training, and results that are close to the optimal solution. Therefore, each parameter must be carefully tuned and explicitly reported.

Hardware Configuration: Although hardware specifications are not hyperparameters per se, reporting them is crucial because they have a significant impact on training efficiency and inference speed. Training deep learning models of this scale is practically impossible using only a CPU. Hence, all experiments in this study were conducted using an NVIDIA T4 GPU on the Google Colab platform. This GPU substantially accelerated the training process and ensured timely convergence.

Image Size: The input image resolution was set to 256×256 pixels, a commonly used size for classification, detection, and semantic segmentation tasks. Increasing the resolution can improve model performance by preserving finer spatial details (provided other hyperparameters are properly adjusted), but it also increases computational cost and inference time. Conversely, reducing the resolution enhances computational efficiency but can severely degrade performance due to the loss of visual and contextual details. Thus, a 256-pixel resolution was chosen as a balanced compromise between performance and efficiency.

Batch Size: During all phases—training, validation, and testing—the batch size was set to 4. Unlike classification tasks where batch sizes of 16, 32, or higher are common, segmentation tasks involve dense pixel-wise prediction, which requires much more memory per sample. This relatively small batch size was selected to optimize GPU memory usage while maintaining gradient stability and effective convergence.

Learning Rate: The learning rate is one of the most critical hyperparameters in deep learning, as it directly controls the optimization dynamics. If set too high, the model diverges and fails to converge; if too low, training becomes excessively slow or stalls. Therefore, careful empirical calibration is essential. Initial experiments were conducted with learning rates of 0.1, 0.01, 0.001, and 0.0001. Using the AdamW optimizer, the learning rate of 0.001 produced the best initial convergence behavior (fastest, with only minor oscillations). Larger values (0.1 and 0.01) caused instability and strong loss fluctuations, while 0.0001 yielded stable but slower convergence, which was suboptimal during early epochs.

Dynamic Learning Rate: To promote adaptive learning and improve training effectiveness, a progressive learning rate scheduler was implemented. Although AdamW inherently adapts learning rates for each parameter (reducing dependence on explicit scheduling compared to SGD), incorporating a scheduler still offers important benefits—particularly faster learning in early stages and finer adjustments later in training. This behavior is illustrated in Figure **3-4.**

This multi-step decay strategy provides a robust mechanism for stable and efficient convergence. Even when the initial rate (e.g., 0.001) is slightly aggressive, the subsequent decay steps act as a compensatory mechanism, ensuring long-term stability. Additionally, final learning rates such as 0.00001, though small, do not hinder convergence; instead, they facilitate fine-tuning in the final stages. Overall, this dynamic scheduling approach ensures both rapid initial optimization and effective convergence toward optimal minima.

Regularization: L2 regularization was implemented through the weight decay parameter in the PyTorch environment. This mechanism is vital for controlling overfitting by penalizing large weights and stabilizing the training process. Considering the dataset size, chosen input resolution, task type (semantic segmentation), and model complexity, a value of 0.0001 was empirically selected. This configuration effectively reduced performance oscillations and prevented overfitting.

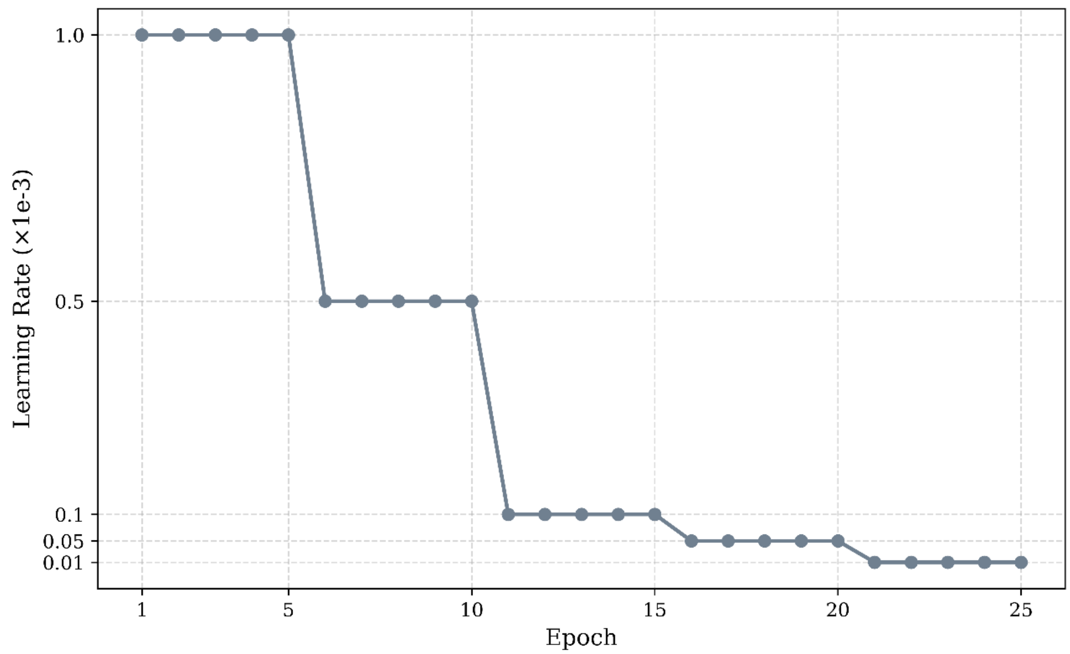

Figure 3-4. Learning rate decay during the training process, contributing to model training stability.

3.7. Metrics

3.7.1. Common Metrics for Image Segmentation

To evaluate the performance of image segmentation, we employ three well-established and widely used metrics. These metrics are chosen for their ability to assess both the overall accuracy and the class-wise overlap, making them highly suitable for segmentation tasks.

Pixel Accuracy (PA):

As shown in **Equation (11-3),** this metric measures the ratio of correctly classified pixels to the total number of pixels in the image. It provides a straightforward indication of segmentation accuracy. However, an important consideration is that this metric can be dominated by large or background classes, potentially overstating model performance in imbalanced datasets. Therefore, it must be complemented with additional metrics—particularly in fine-grained or class-sensitive segmentation tasks.

$$PixelAccuracy = \left(\Sigma_{(c=0)}^{(C-1)} TP_c\right) / \left(\Sigma_{(c=0)}^{(C-1)} (TP_c + FP_c + FN_c)\right) \tag{11-3}$$

Jaccard Index (Intersection over Union, IoU):

This metric quantifies the overlap between the predicted segmentation and the ground-truth mask for each class. It measures how well the model's predicted regions align with the actual labeled regions. The formula for IoU is presented in **Equation (12-3).**

$$IoU_c = TP_c/(TP_c + FP_c + FN_c) \tag{12-3}$$

Dice Score (Dice Coefficient):

Similar to IoU, this metric focuses on the overlap between the predicted and ground-truth regions. It emphasizes both precision and recall by measuring the harmonic mean of these two quantities. The formula for the Dice Score is given in **Equation (13).**

$$Dice_c = (2 \cdot TP_c)/(2 \cdot TP_c + FP_c + FN_c) \quad (13\text{-}3)$$

Together, these metrics ensure a comprehensive evaluation of segmentation performance: while Pixel Accuracy provides an overall measure, it can be biased toward dominant classes, whereas IoU and Dice Score serve as stricter and more discriminative metrics, offering deeper insights into class-wise performance, particularly when dealing with underrepresented or minority classes.

3.7.2. Tailored Metric for This Study

In addition to employing standard evaluation metrics, we introduce a novel metric called Distributed Pixel Accuracy (DPA), which serves as a more tolerant variant of traditional pixel accuracy, specifically designed to meet the requirements of this study. In the context of food waste estimation, the goal is not necessarily to determine the exact spatial position of every individual pixel, but rather to estimate the relative proportions (distribution) of different classes within each segmentation mask. To achieve this, DPA, defined in Equation (14), evaluates the similarity of pixel ratios for each class between the predicted mask and the ground-truth mask, focusing more on distributional agreement than on precise spatial alignment.

$$DPA_c = 1 - \left|\left(P_c^{pred}/T^{pred}\right) - \left(P_c^{gt}/T^{gt}\right)\right| \quad (14\text{-}3)$$

In which :

$*P_c^{pred}$ و P_c^{gt} denote the number of predicted and ground-truth pixels for class ccc, respectively.

$*T^{pred}$ و T^{gt} represent the total number of pixels in the predicted mask and the ground-truth mask, respectively.

This metric penalizes large deviations in the ratio of each class, thereby providing a distribution-based measure that aligns with the estimative goals of this study. Naturally, it is even more lenient than traditional Pixel Accuracy (which itself is more tolerant than IoU and Dice); however, given the nature of this research, it serves as an appropriate and meaningful metric.

To illustrate, **Figure 3-5** presents several examples, including an extreme case in which conventional metrics (Pixel Accuracy, IoU, and Dice) yield values close to zero due to spatial misalignment, whereas the DPA achieves a value close to 1, as it correctly captures the relative distribution ratios of the classes. This rare yet insightful example highlights the unconventional but valuable nature of the DPA metric.

As mentioned, it should be noted that DPA is not suitable for applications requiring high spatial precision, such as medical imaging, autonomous driving, or other safety-critical systems. However, in the context of food waste estimation, where the proportional coverage of food items is more important than their exact location, DPA proves highly useful. This metric enables relative surface-level comparisons among classes and provides a practical perspective for evaluating food waste estimation tasks.

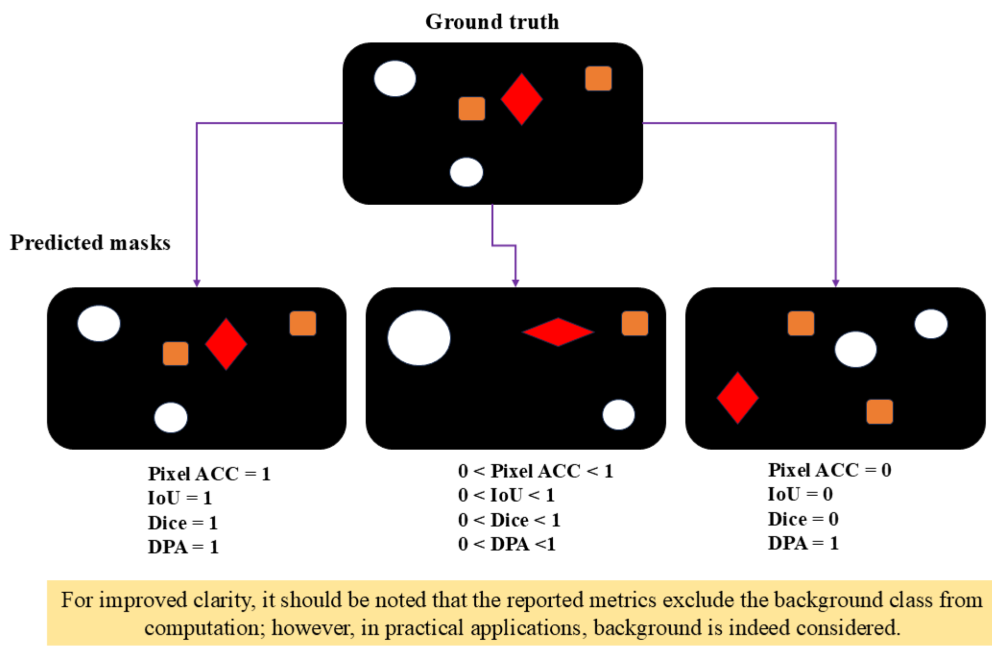

Figure 3-5. Comparison of the proposed DPA metric with conventional evaluation metrics.

3.7.3. Selection of Metrics Based on Suitability (Weighted vs. Unweighted)

Selecting appropriate evaluation metrics—whether standard or customized—is essential for accurately measuring model performance.

Macro (Unweighted) Average:

In this approach, all classes are treated equally. Each metric is computed independently for every class, and their simple mean is then calculated. This method is more sensitive to minority classes, making it particularly useful when rare categories are of special importance.

Weighted Average:

Here, metrics are computed for each class and then weighted according to the number of true samples in each category. This produces a more representative overall score, as dominant classes have a proportionally greater influence on the final evaluation.

The choice between these two approaches should reflect the specific needs of the application domain. For instance, in medical image analysis, where detecting small and critical regions is vital, macro-averaged metrics are preferable because they give greater importance to minority classes. In contrast, in food waste estimation, weighted metrics are more appropriate. Minor segmentation errors—such as mislabeling a few grains of rice or slight boundary mismatches—have negligible effects on the overall estimation accuracy. Thus, relying solely on macro averages could unfairly penalize the model, as small errors in rare classes would be treated equally to large errors in dominant ones.

In summary, although this study employs a custom dynamic weighted loss function to enhance segmentation performance for smaller classes (by emphasizing inverse-frequency weighting to capture even the smallest components), the final model selection and optimal epoch determination are based primarily on weighted metrics. This approach better reflects the real-world objectives and acceptable error thresholds of the task.

3.8. Chapter Summary

In this chapter, the research methodology was presented within a comprehensive and structured framework. This framework serves as the backbone of the study, encompassing all key stages—from data collection and preprocessing to the design, selection, and configuration of model components—and systematically illustrating their interconnections.

Furthermore, both standard evaluation metrics and the custom metrics developed for this research were integrated into the framework to enable a multidimensional performance assessment. The emphasis on developing this unified framework ensured that the methodology transcended a mere set of technical steps, evolving instead into a coherent roadmap for data management, model optimization, and result validation.

Thus, the proposed framework enhances the transparency, reproducibility, and generalizability of the research process while providing a foundation for comparison with other related approaches. In the next chapter, building upon this developed framework, the empirical findings and both quantitative and qualitative results of the study will be analyzed to evaluate the practical effectiveness and innovative contribution of the framework in food waste estimation.

Chapter 4

Research Findings

4.1. Introduction

In the previous chapter, the research methodology—including the approach for estimating food waste, the deep learning-based segmentation models, model components such as the custom loss function, optimizer selection, hyperparameter tuning, and the definition of newly introduced components—was discussed in detail. In this chapter, the results and findings of the study are presented and discussed, followed by an examination of the limitations of the research and potential directions for future work.

4.2. Insights From the Data

This section presents the findings derived from the collected data. Initially, **Figure 4-1** illustrates sample images captured before and after consumption, representing the outcomes of the data collection process.

It should be noted that the images were not precisely paired (before and after) from the same specific plate; rather, they were selected from the overall set of similar samples. More specifically, the images taken before consumption were captured from the food serving line, where all dishes had equal portions and identical compositions. The images after consumption were collected from the dish collection line, where food waste was gathered. Undoubtedly, this approach not only poses no issue but also realistically simulates real-world conditions.

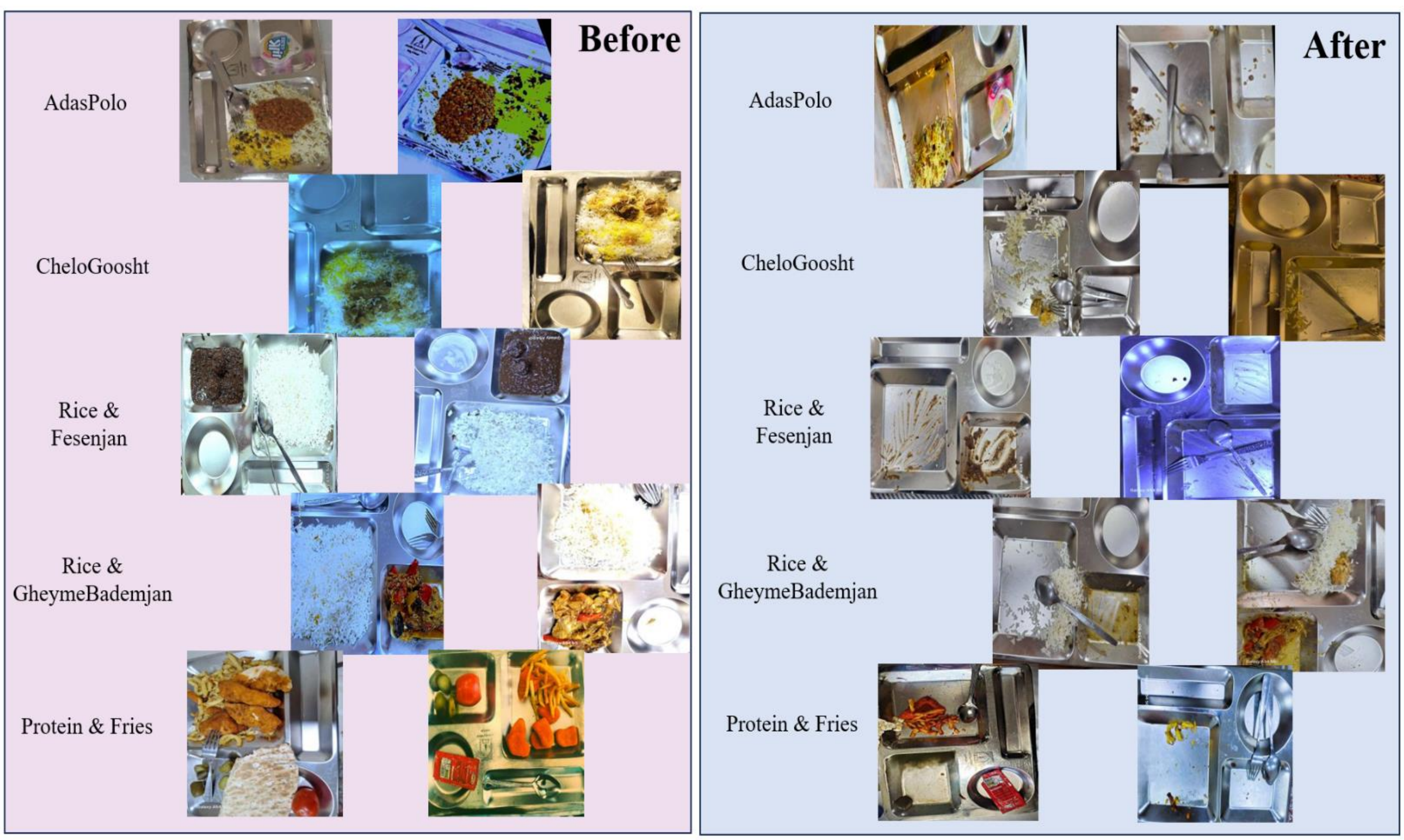

Figure 4-1. Sample images from the dataset: left for before consumption and right for after consumption.

In other words, in an actual setting—such as a cafeteria serving over 1,000 students—it is practically impossible to track individual plates unless massive hardware and highly advanced image processing systems are employed. Therefore, comparing the plates after consumption with the *average* of the plates before consumption is a logical method.

Moreover, since the quantity, proportion, and composition of food items on the plates are fixed before consumption, comparing post-consumption images with the average pre-consumption ones is both reasonable and rational. As mentioned earlier, the next step involves generating masks based on specific classes, the details of which are precisely presented in **Table 4-1**. **Figure 4-2** illustrates the results of mask generation, depicted in various colors.

In the next stage, the findings obtained from data augmentation and diversification, the construction process of which was explained in detail in the methodology section, are presented. **Table 2-3** and **Figure 4-3** illustrate the order, degree, and intensity of data diversification, as well as the transformation of a specific image under the influence of data augmentation and expansion.

Table 4-1. Number and types of annotations for each food category

Type of Food	Number of Categories	Categories (+ 0: Background)
Adas Polo	2	1: Adas Polo
Chelo Goosht	3	1: Meat 2: Rice
Fesenjan	3	1: Fesenjan Stew 2: Rice
Gheymeh Bademjan	3	1: Gheymeh Bademjan Stew 2: Rice
Protein and Fries	3	1: Fries 2: Protein

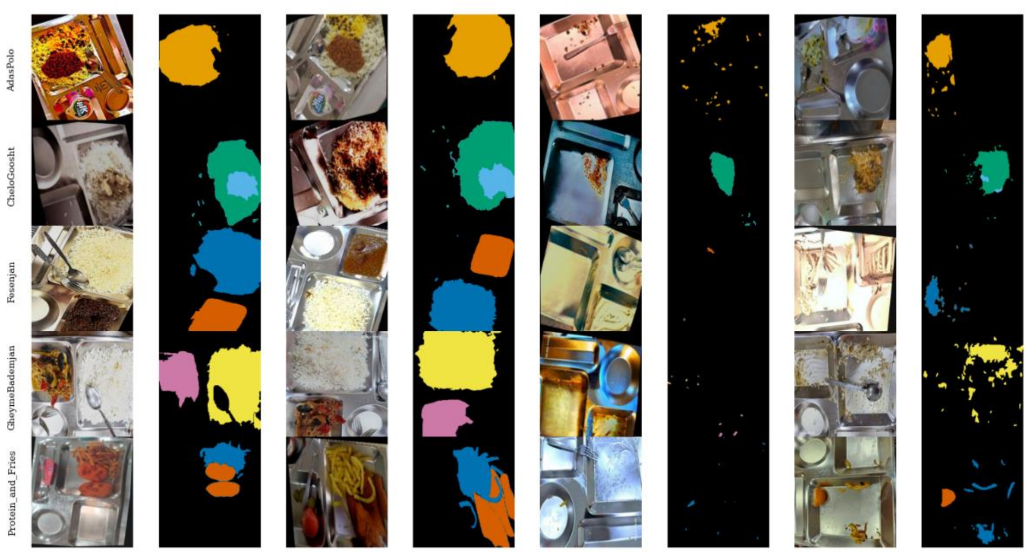

Figure 4-2. Sample pre- and post-consumption images along with semantic masks in five food categories

Table 4-2. Data augmentation pipeline and the size of training and test sets

Food Type	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	Stage 7	Stage 8	Stage 9	Test Size	Training Size (Before Augmentation)
Adas Polo (Lentil Rice)	Horizontal and vertical flipping	90° clockwise, counterclockwise, and full inversion rotation	Rotation between ±15°	Shear transformation between ±15° horizontally and vertically	Saturation between −15 and +15	Brightness between −15 and +15	Exposure (−3 to +3)	Blur up to 1 px	–	63	264
Chelo Goosht (Rice with Meat)	Horizontal and vertical flipping	90° clockwise, counterclockwise, and full inversion rotation	Rotation between ±15°	Shear transformation between ±15° horizontally and vertically	Color spectrum variation between ±15	Saturation between −20 and +20	Brightness between −15 and +15	Exposure between −5 and +5	–	64	207
Fesenjan (Pomegranate Walnut Stew with Rice)	Horizontal and vertical flipping	90° clockwise, counterclockwise, and full inversion rotation	Rotation between ±15°	Shear transformation between ±15° horizontally and vertically	Color spectrum variation between ±15	Saturation between −15 and +15	Brightness between −15 and +15	Exposure between −10 and +10	Blur up to 1.2 px	61	148
Gheyme Bademjan (Split Pea & Eggplant Stew with Rice)	Horizontal and vertical flipping	90° clockwise, counterclockwise, and full inversion rotation	Rotation between ±15°	Shear transformation between ±15° horizontally and vertically	Color spectrum variation between ±15	Saturation between −15 and +15	Brightness between −15 and +15	Blur up to 1.2 px	–	63	167
Protein & Fries	Horizontal and vertical flipping	90° clockwise, counterclockwise, and full inversion rotation	Rotation between ±15°	Shear transformation between ±15° horizontally and vertically	Color spectrum variation between ±15	Saturation between −15 and +15	Brightness between −15 and +15	Exposure between −5 and +5	Blur up to 1.2 px	103	273

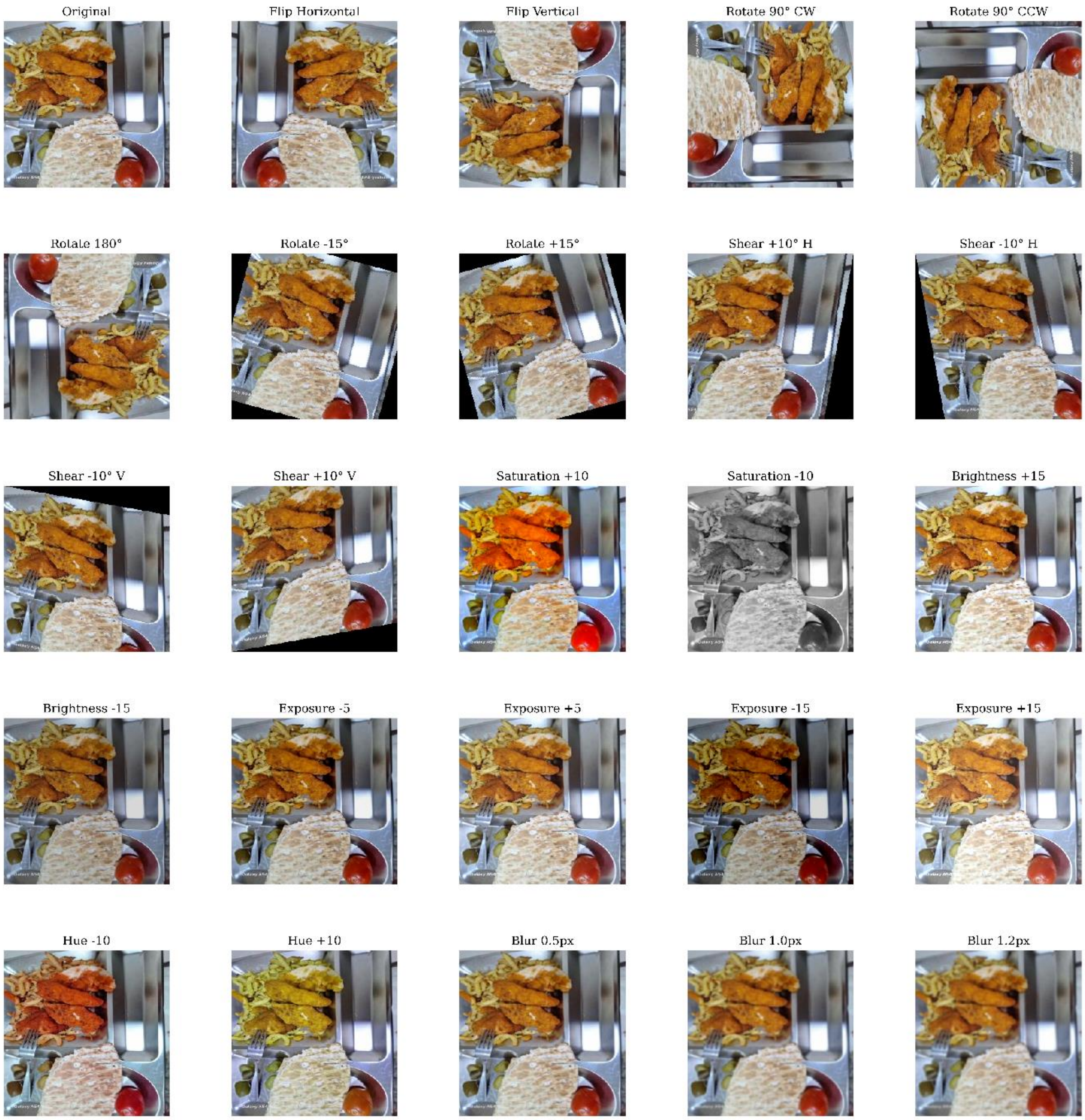

Figure 4-3. View of an image under diverse conditions, showing the variations applied during data augmentation

4.3. Food Waste Estimation

This section presents the results related to food waste estimation. To achieve a comprehensive analysis, the entire dataset — including both training and test subsets — was examined as an integrated whole. However, as logically expected, the estimations were independently performed for the pre- and post-consumption images to clearly highlight the actual differences between these two stages. This analysis was conducted solely based on the segmented masks and utilized all available data, with the primary objective of providing an overall picture of food waste patterns across different categories. Naturally, in the deep learning phase, model training was conducted exclusively on the training set, and model performance was evaluated only on the test data to prevent any bias or data leakage.

The estimation of food waste was performed according to Equations (1)–(4), by calculating the occupied area of each category (food component) within the masks. Within this framework, **Figures 4-4 to 4-8** present the results as histogram charts, illustrating the surface distribution of food components in the pre- and post-consumption masks, along with their weighted mean values (for each). These figures provide a clear representation of the consumption and remaining portions for each food category.

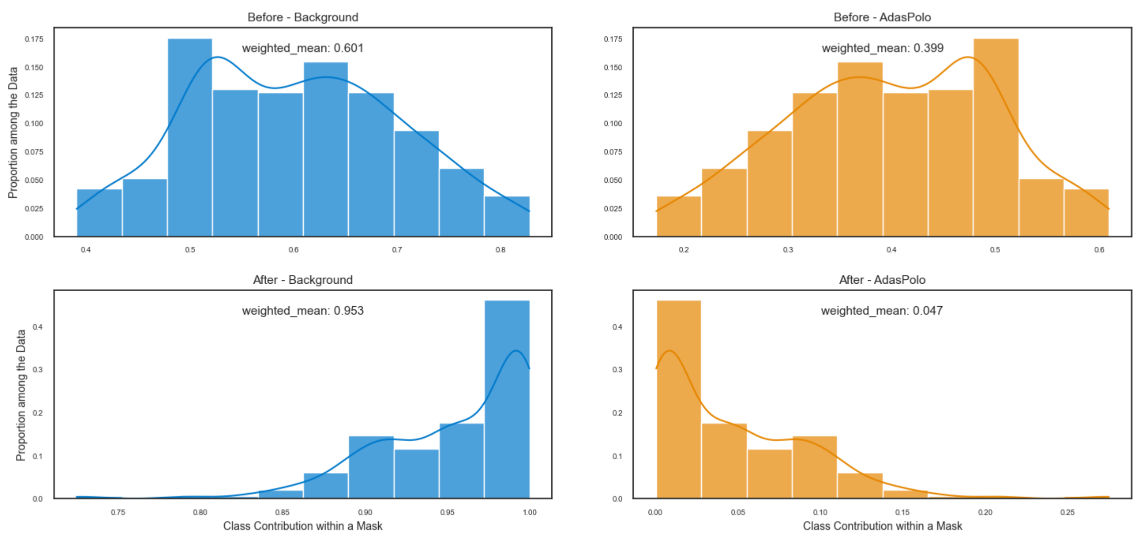

Figure 4-4. Histogram of surface area proportions of categories for Adas Polo before and after consumption

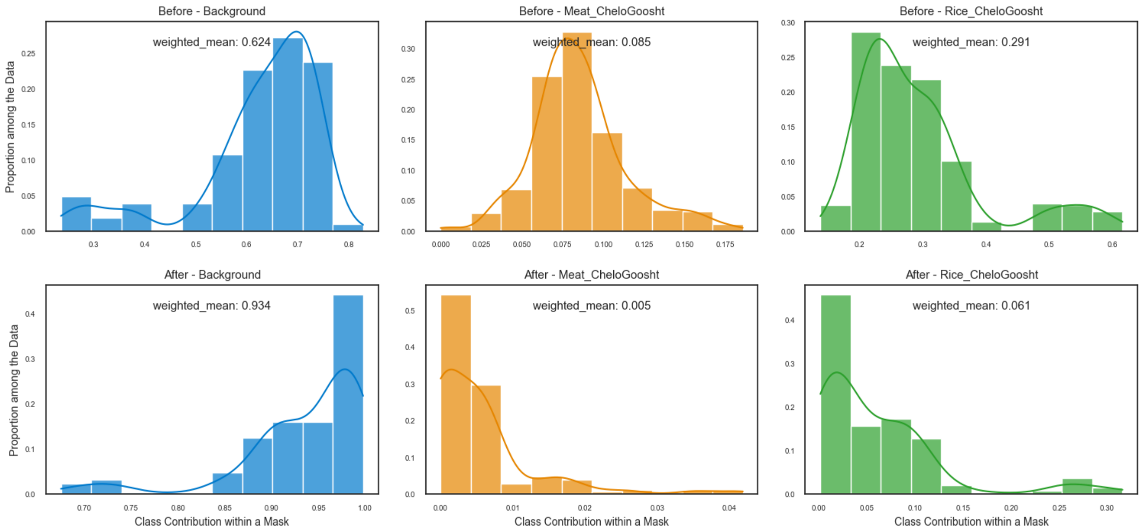

Figure 4-5. Histogram of surface area proportions of categories for Chelo Goosht before and after consumption.

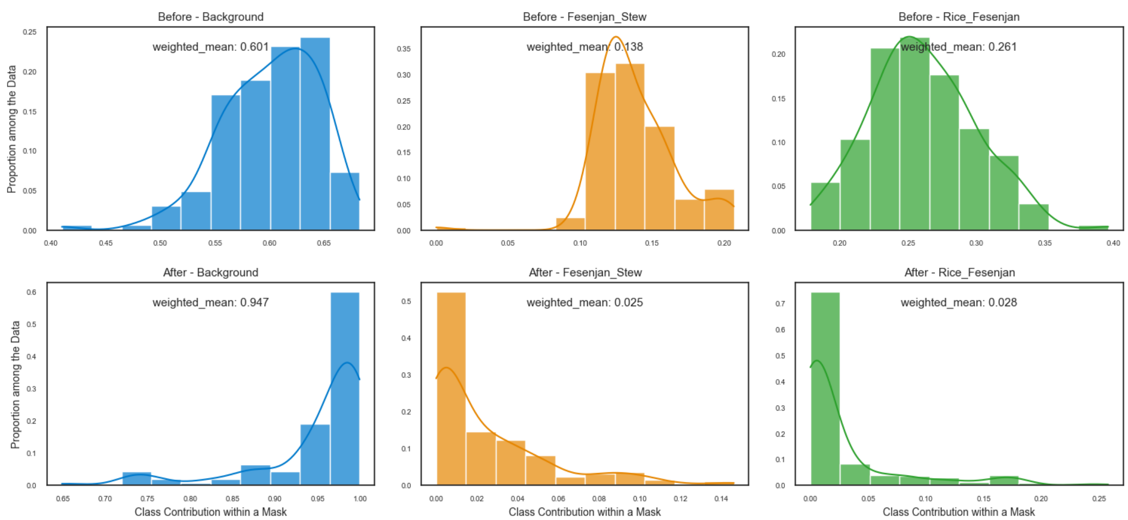

Figure 4-6. Histogram of surface area proportions of categories for Fesenjan before and after consumption.

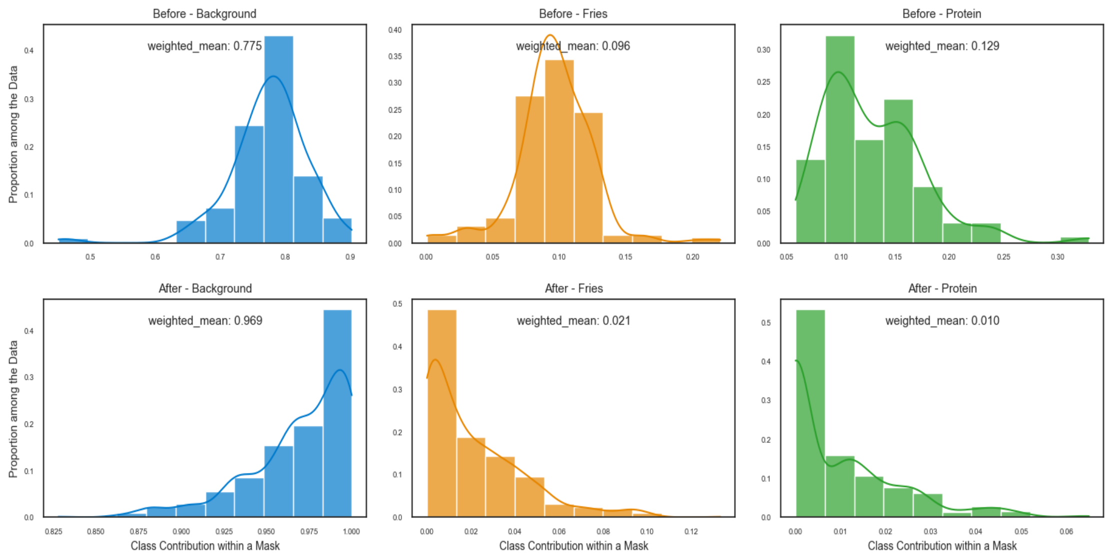

Figure 4-7. Histogram of surface area proportions of categories for Gheymeh Bademjan before and after consumption.

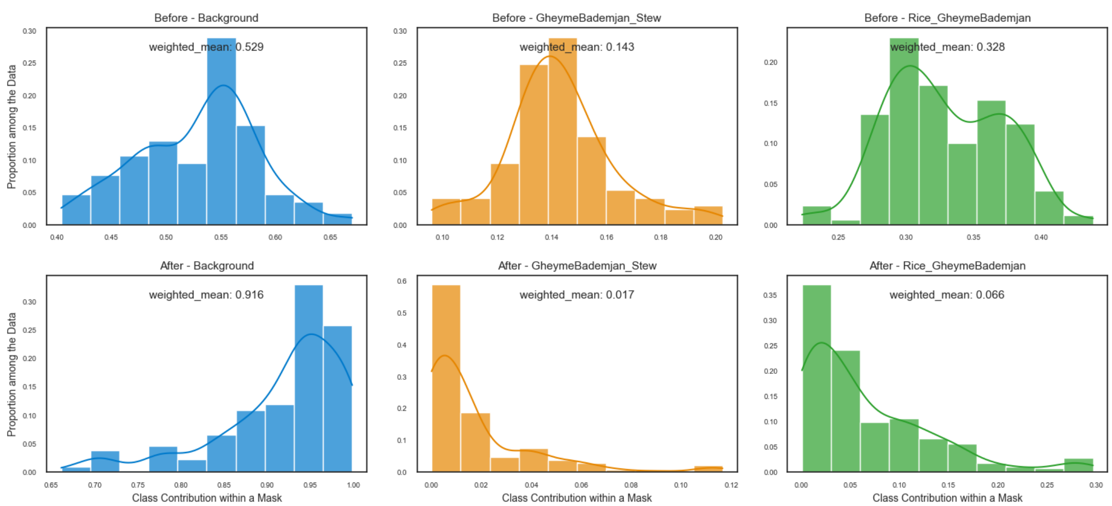

Figure 4-8. Histogram of surface area proportions of categories for Protein and Fries before and after consumption.

Furthermore, to establish a valid basis for comparison between the two stages, category ratios were calculated using weighted means to provide stronger and more interpretable quantitative metrics. Using the mean values obtained from the histograms (**Figures 4-4 to 4-8**) and Equations (1) to (4), the average consumption rates and the estimated food waste values were computed. The complete details of these results are presented in **Table 4-3**, which shows the average rates of food consumption and leftovers for each case across different food categories (based on surface area ratios).

Table 4-3. Calculation of Eating Rate and Remaining Rate

Food Type	Category	Average Weight Before Consumption	Average Weight After Consumption	Eating Rate (%)	Remaining Rate (%)
Adas Polo	Lentil Rice	0.399	0.047	88.2	11.8
Meat (Chelo Goosht)	Chelo Goosht	0.085	0.005	94.1	5.9
Rice (Chelo Goosht)	Chelo Goosht	0.291	0.061	79.0	21.0
Fesenjan Stew	Fesenjan	0.138	0.025	81.9	18.1
Rice with Fesenjan	Fesenjan	0.261	0.028	89.3	10.7
Gheymeh Bademjan Stew	Gheymeh Bademjan	0.143	0.017	88.1	11.9
Rice (Gheymeh Bademjan)	Gheymeh Bademjan	0.328	0.066	79.9	20.1
French Fries (with Protein)	Protein & Fries	0.096	0.021	78.1	21.9
Protein	Protein & Fries	0.129	0.010	92.2	6.8

4.4. Models' Training & Validation

As previously explained, the test set was separated from the very beginning to remain completely unseen and to prevent any data leakage or bias. The training set was then divided into two parts: 85% for training and 15% for validation. This validation subset was used to identify the best-performing model — the one achieving the highest weighted IoU score on the validation data (which, in most cases, also corresponds to the highest weighted Dice score). This model was saved and ultimately used for evaluation on the unseen test set.

To provide a visual understanding of model performance over time, **Figure 4-9** illustrates the evolution of the weighted Dice score during training for all four models across different food categories. It is noteworthy that, while some models exhibit fluctuations in the initial epochs, the early training phase proceeds rapidly, establishing a strong foundation for employing smaller learning rates in later stages to achieve smoother convergence. This figure also highlights the models' success in converging to their optimal state based on the chosen evaluation metrics.

Moreover, **Figure 4-9** shows that the *Adas Polo* and *Protein and Fries* categories not only achieved the highest validation Dice scores but also demonstrated the most stable and consistent convergence patterns. It further reveals that each pair of models — U-Net and its smaller variant, as well as U-Net++ and its smaller counterpart — exhibited similar performance and convergence behaviors compared to models outside their respective pairs. This observation suggests that, at least in this application, the model's structure and internal connectivity may have a greater impact than the sheer number of convolutional layers responsible for extracting higher-level features.

Additionally, in nearly all cases, either U-Net or its smaller version performed as well as or better than U-Net++ and its smaller variant. Notably, the smaller U-Net++ produced results that either surpassed or matched those of the full U-Net++. This indicates that, given the dataset size and the moderate complexity of the current task, the problem is not overly challenging. In simpler terms, from a research perspective, this finding implies that simpler models with fewer parameters and a lower risk of overfitting can perform as well as — or even better than — more complex architectures in relatively straightforward tasks. However, it should be emphasized that this observation is specific to this task, and exploring various architectures remains a valid and recommended practice depending on dataset size and task complexity.

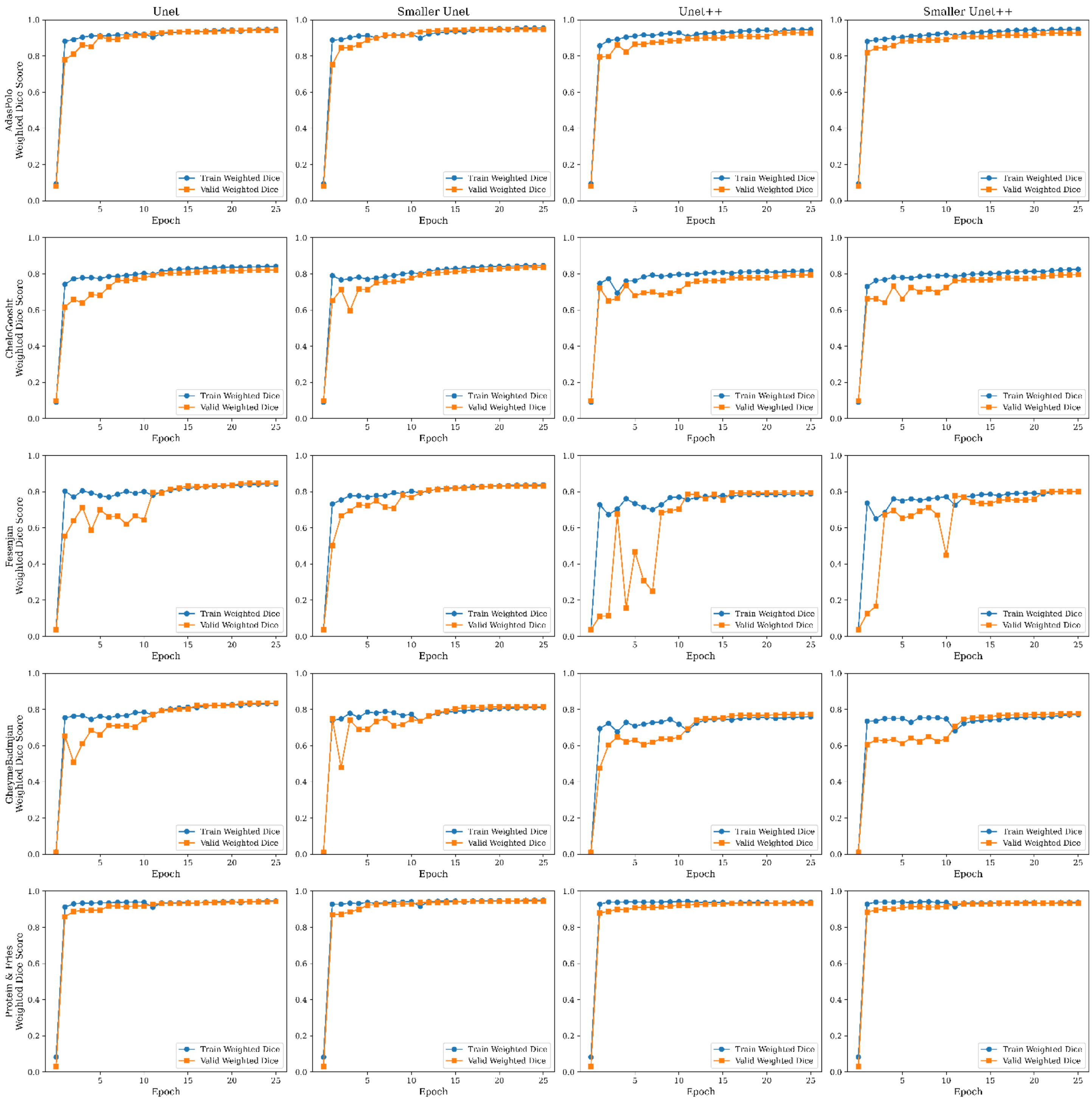

Figure 4-9. Training and evaluation of the deep learning models developed for the five food categories.

4.5. Test Results

To identify and apply the optimal model during training, the model achieving the highest weighted IoU on the validation set was saved and then evaluated on the unseen test data. The results are summarized in **Figure 4-10.**

This figure presents several valuable observations. First, all performance metrics—especially DPA, a customized metric designed for this study—highlight the effectiveness of the proposed approach. The highest DPA value among all models for all food types reached approximately 90%, indicating the strong potential of these predictive models to estimate food waste with minimal reliance on manual mask labeling. Moreover, the models also achieved high performance in conventional deep learning metrics such as IoU and Dice, which are considered strict evaluation criteria, generally exceeding 85% in the Dice Score metric. Second, the consistent and strong performance of the U-Net and Smaller U-Net models indicates that model selection should align with task complexity. This finding emphasizes that heavier models with more parameters are not necessarily superior. In scenarios like the present study, lightweight architectures with fewer parameters and faster inference times can deliver comparable or even better performance than more complex counterparts.

Third, when focusing on the most common and widely accepted segmentation metrics such as Dice and IoU, the similarity in performance between each model pair—U-Net and its smaller version, and U-Net++ and its smaller version—becomes evident. This observation further supports the discussion in Section 4.3, confirming that lighter model variants can maintain comparable accuracy.

Finally, the *Adas Polo* and *Protein and Fries* food categories achieved the highest performance across all metrics, while stew-based dishes such as *Gheymeh Bademjan* and *Fesenjan* exhibited relatively lower results. This difference is likely due to the inherent visual complexity and fluid nature of stews, which affect both the accuracy of mask labeling and the model's prediction performance. This insight warrants further attention and underscores an important aspect of data diversity in segmentation tasks.

.

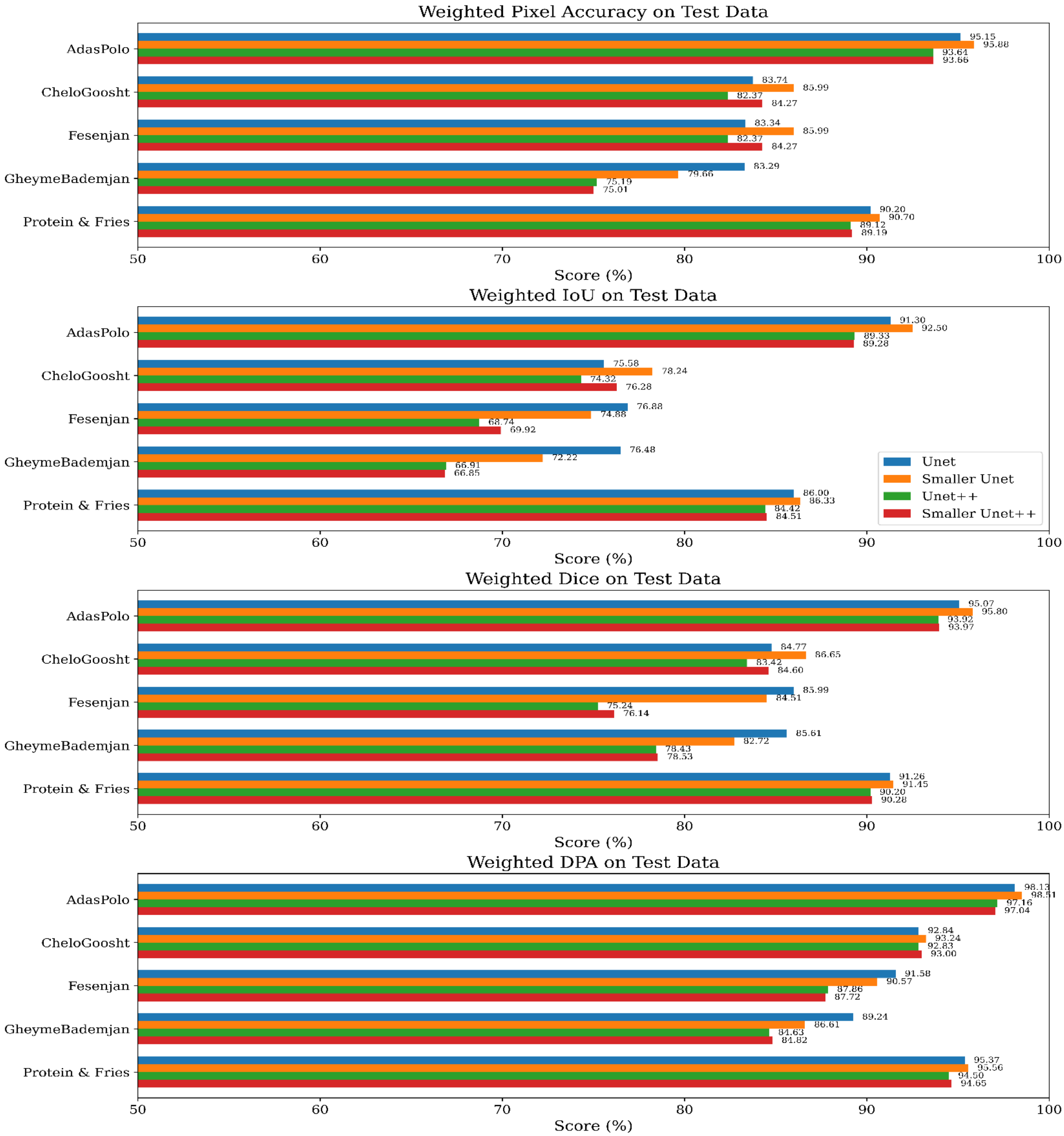

Figure 4-10. Model's Results on Unseen Test Data

4.6. Models' Computational Complexity & Speed

To better understand how model complexity and parameter count affect computational efficiency, a detailed comparison is presented in **Figure 4-11**. As previously mentioned, these measurements were conducted using a T4 GPU within the Google Colab environment. It should also be noted that the reported speeds may vary slightly due to factors such as internet connectivity and system load. Therefore, the values shown in the figure represent a range between the minimum and maximum observed speeds, defining the performance spectrum recorded in this study. As expected, training speed is generally lower since backpropagation introduces additional computational overhead. Moreover, model size shows an inverse relationship with speed, models with a higher number of parameters typically operate more slowly.

Importantly, inference speed holds greater significance in real-world applications, as models are primarily deployed for prediction. Fortunately, all evaluated models in this study achieved inference speeds exceeding 20 images per second, indicating their practical potential for real deployment. Notably, the Smaller U-Net, with approximately 7.8 million parameters, not only demonstrated strong performance but also achieved an inference speed of over 80 images per second. This result further reinforces the notion that, depending on task requirements and data characteristics, lightweight architectures can offer an effective trade-off between speed and performance.

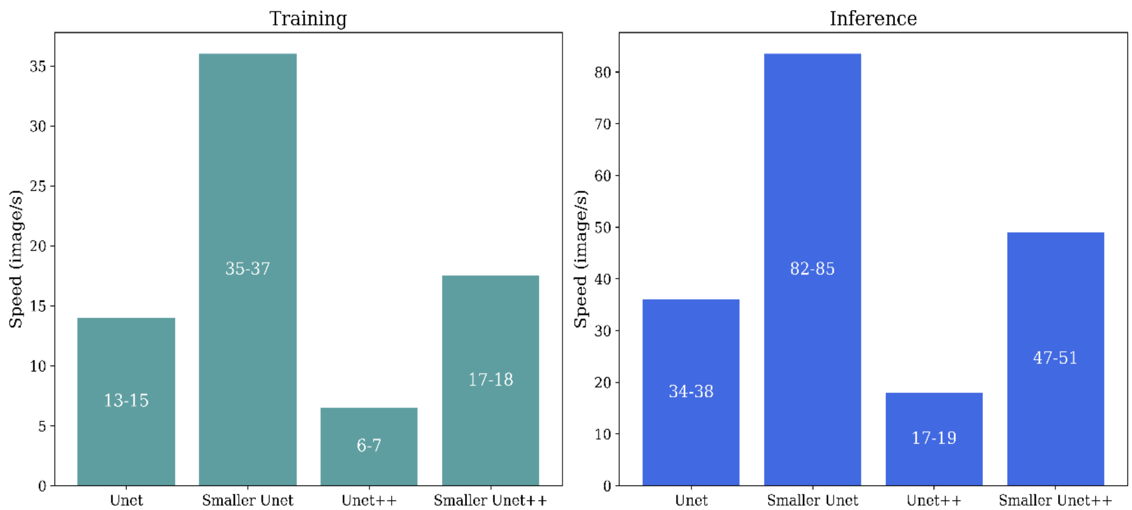

Figure 4-11. Model speeds during the training and inference processes.

4.7. Discussion, Limitations, & Future Works

4.7.1. Qualitative Analysis

For evaluation, a comprehensive set of common metrics, including Pixel Accuracy, IoU, and Dice, was employed. In addition, the innovative DPA metric was introduced, which, instead of focusing solely on spatial pixel alignment, considers the ratio and distribution of pixels. This approach is arguably more suitable for the nature of this study, as its ability to measure the consistency of pixel distribution between the model's predictions and the ground truth provides more reliable results.

Regarding the results, the high DPA values—particularly above 0.9 in models such as U-Net and its lighter version- indicate strong agreement with manually labeled masks and, consequently, high accuracy in estimating consumption and waste rates. This consistency reflects the internal validity of the entire experimental process and confirms the reliability of the quantitative findings presented.

Furthermore, the strong performance of lightweight models such as the Smaller U-Net, combined with fewer parameters, lower FLOPs, and significantly faster inference speeds, highlights the practical applicability of the proposed approach in real-time environments. This feature is particularly valuable in scenarios where efficiency and speed are of critical importance. Nevertheless, model selection should always align with the task complexity and dataset size, as deeper architectures such as U-Net++ may still be more appropriate for more complex problems or larger datasets.

4.7.2. Limitaitons & Considerations

Although this study has made progress in estimating and quantifying food waste using computer vision and semantic segmentation, it is not without limitations and challenges. The most important ones are discussed below.

(a) <u>Two-Dimensional vs. Three-Dimensional Analysis</u>

One inherent limitation of estimating food waste through 2D image segmentation is the lack of depth information. This limitation prevents the model from adequately understanding volume variations, density differences, and overlapping food layers. For example, in meals containing granular components such as rice, visually similar regions in pre- and post-consumption 2D images may represent very different quantities because, before consumption, the image may show a larger volume simply due to layering. Such differences can lead to moderate estimation errors in consumption or leftover quantities, particularly in applications where volumetric precision is critical.

In contrast, 3D methods that utilize depth data (e.g., point clouds or voxel representations) can more accurately capture spatial and volumetric features, resulting in higher estimation accuracy. However, practical implementation of these methods faces significant computational, logistical, and resource-related constraints. Processing 3D data requires powerful hardware, much longer execution times, and 10–100 times higher memory consumption than 2D methods. Furthermore, generating 3D labels (such as voxel masks or depth-aligned annotations) is highly costly and time-consuming, making it impractical or undesirable for many research and operational scenarios.

Therefore, the use of 2D segmentation in this study represents a practical and cost-effective choice, striking a balance between feasibility and accuracy. Nevertheless, future research may explore hybrid approaches such as depth simulation or depth estimation from 2D images to bridge the gap between 2D simplicity and volumetric precision, thereby improving the reliability of food waste estimation.

(b) <u>Risk of Overestimation</u>

It should be noted that calculating consumption and leftover rates based on manually labeled masks may slightly bias the results toward underestimating consumption and consequently overestimating leftovers. This bias partly stems from the same lack of volumetric information. For instance, in pre-consumption images, a single central pixel might represent several grains of rice, whereas the same position in a post-consumption image may correspond to only one grain or a small residue.

Additionally, the manual labeling strategy in this study aimed to create highly precise data by annotating even the smallest regions. Although this approach improves training quality, it can introduce minor deviations in consumption estimation, leading to slightly lower consumption and higher leftover values than the actual quantities. These findings indicate that future research should pursue corrective mechanisms, such as reduction factors, that better align pixel-based estimations with real volumetric changes to enhance result realism.

(c) Challenges in Segmenting Stews and Semi-Liquid Foods

A major challenge in food segmentation, particularly in post-consumption images, involves dishes with stew-like or semi-liquid characteristics. Due to their viscous yet fluid nature, these foods tend to spread on the plate during consumption, often leaving irregular stains or scattered residues. This behavior complicates the precise detection of food boundaries.

Unlike solid foods, which typically retain distinguishable boundaries even when fragmented, segmenting stew residues is far more complex. In such cases, even human annotators struggle to determine whether dispersed stains or small particles should be considered part of the edible content. Furthermore, the visual characteristics of these foods vary in intensity — darker, denser regions at the center gradually fade to lighter tones at the edges. This gradual color change increases labeling difficulty and introduces subjectivity and ambiguity into the masks.

Such ambiguity in human-labeled data directly affects deep learning model performance. Models relying on these annotations face difficulties when dealing with unclear or i

rregular boundaries, as reflected in the lower evaluation scores for these categories in this study. Since the root cause lies in the lack of a precise definition of "actual remaining food" after consumption, there is a need for methods capable of handling visual uncertainty. One promising approach is Type-2 Fuzzy Logic, which can model higher-level uncertainty (Castillo et al., 2017), especially in visually ambiguous regions. However, further experimental validation is still required in the context of food segmentation. Moreover, although attention mechanisms cannot eliminate the inherent ambiguity of manual labeling, they can improve segmentation performance by helping the network focus more effectively on salient regions of the image.

(d) Edible vs. Non-Edible Waste and Presence of Utensils

In this study, nearly all served meals were entirely edible; hence, there was no need to distinguish between edible and non-edible components in post-consumption segmentation. However, this issue is worth mentioning, as in practical scenarios, non-edible elements such as bones from meat or chicken may remain. In such cases, a separate class should be defined for these components, and they must be excluded from consumption rate calculations to preserve analytical accuracy.

Another real-world challenge is the unintended presence of utensils such as spoons and forks in the images. These objects can interfere with the segmentation process because their color or texture may resemble food components. Therefore, the segmentation framework should include a detection layer capable of identifying and ignoring such non-edible elements. Specifically, for bone-in dishes such as chicken or fish, distinguishing edible from non-edible parts becomes particularly important; otherwise, consumption rates may be underestimated. This issue adds complexity compared to fully edible foods like rice or boneless stews. Thus, designing a precise classification strategy to separate "fully edible" from "bone-in" foods and to manage the presence of utensils in images is essential for extending this framework to real-world applications.

Furthermore, in real environments such as university cafeterias, public restaurants, or hospitals, food variety is very high, and combinations of edible and non-edible items on a single plate are common. For example, bones, chicken or fish skin, fruit pits, or even decorative inedible items may cause segmentation errors. Ignoring such elements could lead to inaccurate waste volume estimates and reduced reliability of research findings.

Therefore, future versions of the proposed framework should incorporate mechanisms to detect and remove these components to ensure accurate performance under complex real-world conditions. Additionally, the presence of utensils like spoons and forks poses challenges not only in segmentation but also during data preprocessing. These objects often occupy a considerable portion of the image and may introduce bias into the model. To mitigate this issue, hybrid approaches could be employed — for instance, using object detection models or multi-stage frameworks that first identify and remove non-edible elements (such as utensils or bones) before performing segmentation exclusively on food components. Such solutions could elevate the framework to a more advanced level and enable its deployment in a wider range of operational environments.

4.7.3. Practical Applications

Overall, the methodology employed in this study can have significant implications for educational institutions, environmental policymakers, and food service managers. Despite the aforementioned limitations, the proposed computer vision and deep learning–based framework enables automated and relatively accurate monitoring of food consumption and waste in a low-cost, rapid, and non-invasive manner. This capability is particularly valuable in high-traffic environments such as university dining halls, hospitals, and accommodation centers, where manual monitoring is often difficult, expensive, and prone to human error.

From a policy-making perspective, a system built upon this approach has the potential to provide reliable managerial insights even under constrained conditions. With access to a sufficiently large and high-quality set of segmentation masks, the model can predict consumption patterns across different food components. The findings revealed that when protein is served alongside rice (as in *Chelo Goosht*) or fried potatoes, the rate of protein consumption significantly increases. Such results can assist dining managers in optimizing portion sizes of each component based on actual consumption patterns, thereby reducing waste. In university cafeterias, this not only decreases food waste but also enhances student satisfaction and improves cost efficiency.

Moreover, the proposed approach demonstrated that even with lightweight architectures and mid-range hardware, satisfactory performance can be achieved. Therefore, deploying this framework in real-world environments does not require expensive or specialized equipment, making it feasible for resource-limited settings such as schools, dormitories, and military centers. This characteristic is particularly important for university dining facilities, which typically face budget constraints and large numbers of patrons—conditions where any efficiency gain directly translates into higher operational productivity. Ultimately, the presented framework can serve as a foundation for long-term behavioral analysis and consumption trend monitoring, paving the way for future research in nutrition, data-driven food service optimization, and industrial engineering. Future studies should focus on expanding the dataset and incorporating more diverse sources to enhance the generalizability and robustness of results.

In summary, although this study faced certain limitations—such as a relatively small dataset, lack of a fixed imaging system, and reliance on two-dimensional surface analysis—it still provides a practical, scalable, and cost-effective framework for intelligent food waste monitoring. It should be emphasized that this research has an academic and exploratory nature, aiming to lay a scientific foundation for future studies rather than direct commercial deployment. Nevertheless, its findings can facilitate real-world adaptations and applied developments, particularly within university dining environments.

4.8. Chapter Summary

In this section of the chapter, the research findings were comprehensively presented and analyzed, with an effort to clearly and coherently explain the results using quantitative relationships, charts, and tables. Alongside the presentation of results, complementary discussions were provided to compare the findings with previous studies and to examine the strengths and limitations of the current work. The main focus of this chapter was the evaluation of the developed deep learning and image processing–based framework, which successfully achieved a balanced performance between accuracy, speed, and computational cost by employing lightweight architectures, data augmentation techniques, and customized evaluation metrics.

This framework demonstrated its feasibility for deployment even in resource-constrained environments such as university dining halls, serving as a practical alternative to manual and costly waste monitoring methods. From an application standpoint, the findings highlight that this framework can assist cafeteria, hospital, and food service managers in everyday decision-making—for instance, adjusting portion sizes based on real consumption patterns or identifying high-waste meals to improve menus and reduce waste. Furthermore, the framework holds the potential to evolve into an integrated intelligent system for behavioral monitoring and providing long-term managerial insights.

Chapter 5

Summary & Conclusion

5.1. Introduction

In this chapter, the conclusions and a concise explanation of the research findings are presented. Additionally, the questions introduced in Chapter 1 are addressed. The chapter further discusses the key innovations of the study and provides suggestions for future research directions.

5.2. Reviewing Previous Chapters

Chapter One introduced the importance of the food waste issue and its economic, social, and environmental consequences. It then presented the main research problem, objectives, implementation steps, hypotheses, potential applications, and innovations. The primary goal was defined as developing a deep learning and image processing–based framework for accurately estimating food waste.

Chapter Two reviewed the theoretical foundations and related studies. It began by explaining the importance and impacts of food waste, followed by discussions on deep learning concepts, neural networks, and image processing. Previous research was then analyzed, and its limitations were identified — particularly the lack of effective use of learning-based models for food waste estimation. Finally, the research gap that this study aimed to address was clearly defined.

Chapter Three described the research procedure, including data collection (images before and after consumption in the university cafeteria), creation of semantic masks and labeling, preprocessing and data augmentation, design and implementation of deep learning models, hyperparameter tuning, development of a customized loss function, and definition of evaluation metrics. The method for estimating consumption and waste based on pixel-level image analysis was also explained in detail.

Chapter Four presented the results of the developed models. It first reported data-related findings and overall food waste estimations across the dataset, followed by model training, validation, and quantitative evaluation based on the defined metrics. Computational performance and model speed were also assessed. Finally, the discussion section addressed the findings, limitations, and real-world applications, demonstrating that the proposed framework can serve as a practical and scalable tool for intelligent food waste monitoring.

5.3. Research's Contributions & Achievements

This research was designed around two main questions: first, how can artificial intelligence and image processing techniques be used to identify and estimate food waste? And second, to what extent can the waste of each food category be estimated, and how do limitations, errors, and existing considerations affect model performance? The experimental results demonstrated that the proposed framework provides both a practical and scientific answer to these questions. More precisely, the use of pre- and post-consumption images combined with deep learning architectures enabled a relatively accurate estimation of leftover food. The best-performing models achieved at least a Dice score of 0.85 and a distributed pixel accuracy above 0.9 across all food categories on unseen data. This is significant, as it shows that the proposed approach delivers not only conceptual but also quantitative reliability for real-world applications.

One of the key achievements of this study was the design and implementation of a complete pipeline—from field data collection to the final estimation of food waste. This pipeline included essential stages such as standardized image acquisition of plates before and after consumption, precise semantic labeling of food components, data preprocessing and augmentation, and the deployment of various deep learning architectures. The main innovation lies in combining semantic segmentation techniques with problem-specific evaluation metrics, allowing the model to be assessed not only at the level of food component recognition but also in terms of actual consumption and waste estimation. For example, the analysis revealed that the rate of protein consumption significantly increased when served with rice or potatoes—an insight that can serve as a practical foundation for portion control and meal planning policies.

Another major contribution of this study was the simultaneous consideration of both accuracy and computational efficiency. Unlike many previous studies that focused solely on accuracy, this research also evaluated processing speed and the feasibility of real-time implementation. The results showed that although more complex models achieved slightly higher accuracy, the lighter versions performed comparably well. For instance, the smallest variant of the developed U-Net processed approximately 83 images per second, while the minimum speed among other models was 20 images per second. These figures indicate that the proposed framework can provide both the accuracy and processing speed necessary for real-world applications in high-density environments such as university cafeterias or hospitals.

Ultimately, the key innovation of this research lies in developing a framework capable of automatically and in real time estimating food waste while providing valuable data for decision-makers. The proposed system not only contributes to cost reduction and optimization of food preparation and distribution processes but also enables more precise behavioral analysis of consumption. For instance, data analysis revealed that combining protein with rice significantly increases protein consumption—a finding that can directly inform menu design and portion adjustment strategies. Therefore, this study not only addressed its core research questions but also delivered clear quantitative results (Dice ≥ 0.85, Accuracy ≥ 0.9, and inference speed up to 83 images per second), generating both scientific and practical value and opening new avenues for future research and large-scale implementations.

5.4. Future Works

Despite the valuable contributions of this research, there remains substantial room for future work. For example, although two-dimensional (2D) analysis has been the most widely used approach in recent studies due to its simplicity and fast processing, three-dimensional (3D) analysis continues to be one of the most accurate tools for estimating the actual volume of food. The use of 3D data such as point clouds or voxel-based segmentation, could provide a much more realistic representation of remaining food in dishes with irregular depth or heterogeneous distribution. However, challenges such as the need for advanced imaging equipment, high computational costs, significant financial requirements, and the difficulty of generating 3D masks make this approach feasible only under suitable infrastructure and strong financial support. If such conditions are met, integrating 3D analysis with the current approaches could meaningfully enhance the accuracy of food waste estimation models.

Given the practical limitations of 3D analysis, improving 2D methods remains a more logical and cost-effective option for future studies. In this research, food surface comparisons before and after consumption were used; however, this approach inherently introduces errors. For instance, a single pixel in the initial image may represent several grains of rice, whereas the same pixel after consumption may correspond to just a small particle. To reduce such bias, hybrid approaches can be applied — for example, combining 2D segmentation with object detection methods or incorporating weighted estimations into image-based results. Such integration can maintain the advantages of speed and low computational cost while achieving higher precision and compensating for the limitations of surface-level analysis.

Another major challenge in this study was the precise segmentation of stew-like and semi-liquid foods, which, due to ambiguous boundaries and uneven color distribution, are difficult even for human annotators to label accurately. This uncertainty introduces noise into training data and reduces deep learning model performance. Type-2 fuzzy logic could be a promising solution in this context, as it models uncertainty within membership functions, allowing for the representation of ambiguous and complex visual conditions. Integrating this logic into the preprocessing or post-processing stages can assign a continuous degree of membership to each pixel, leading to more accurate boundary delineation of food components. Therefore, incorporating Type-2 fuzzy logic into the segmentation pipeline could be an effective step toward better handling uncertainty and improving prediction accuracy in future studies.

Another important direction for future research is developing methods for detecting and removing non-edible elements in food images. In real-world settings, many meals contain bone-in components such as chicken or fish, where non-edible parts like bones, skin, or seeds naturally remain. If these components are not excluded from computation, the estimated consumption will be lower than the actual value, leading to inaccurate results. Moreover, fully edible foods like rice or boneless stews are inherently easier to segment, whereas bone-in dishes require separate categorization and more refined processing. Another challenge is the unintended presence of serving utensils such as spoons and forks in the images, which can severely disrupt segmentation. These objects often share color or texture similarities with food components and, if misclassified, increase the error rate. Future studies can address this issue by incorporating preliminary detection layers or object detection algorithms to identify and remove such non-edible elements before performing the main segmentation. This enhancement would allow the proposed framework to operate with higher accuracy and robustness in real-world settings, such as university cafeterias or hospitals, and improve its generalizability.

Finally, one of the most important future directions involves the scalability and expandability of the proposed system. Due to current limitations, this study focused only on five common food categories; however, real-world environments such as universities, hospitals, and military facilities feature a much wider variety of dishes. Therefore, future research should expand the dataset scope, conduct more extensive labeling, and develop automated tools for data collection. Moreover, using large general-purpose models alone may not be sufficient; instead, these models should be pre-trained on diverse datasets and then fine-tuned with environment-specific data to maintain generalization while enhancing real-world performance. Such an approach ensures that intelligent systems can retain high accuracy and efficiency even when deployed at scale.

References

Agarla, M., Napoletano, P. and Schettini, R. (2023), “Quasi Real-Time Apple Defect Segmentation Using Deep Learning”, *Sensors*, Vol. 23 No. 18, p. 7893, doi: 10.3390/s23187893.

Ahmadzadeh, S., Ajmal, T., Ramanathan, R. and Duan, Y. (2023), “A Comprehensive Review on Food Waste Reduction Based on IoT and Big Data Technologies”, *Sustainability*, Multidisciplinary Digital Publishing Institute, Vol. 15 No. 4, p. 3482, doi: 10.3390/su15043482.

de Almeida Oroski, F. and da Silva, J.M. (2023), “Understanding food waste-reducing platforms: A mini-review”, *Waste Management & Research*, SAGE Publications Ltd STM, Vol. 41 No. 4, pp. 816–827, doi: 10.1177/0734242X221135248.

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., *et al.* (2021), “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions”, *Journal of Big Data*, Vol. 8 No. 1, p. 53, doi: 10.1186/s40537-021-00444-8.

Assad, T., Jabeen, A., Roy, S., Bhat, N., Maqbool, N., Yadav, A. and Aijaz, T. (2024), “Using an image processing technique, correlating the lycopene and moisture content in dried tomatoes”, *Food and Humanity*, Elsevier BV, Vol. 2, p. 100186, doi: 10.1016/j.foohum.2023.11.013.

Attiq, S., Danish Habib, M., Kaur, P., Junaid Shahid Hasni, M. and Dhir, A. (2021), “Drivers of food waste reduction behaviour in the household context”, *Food Quality and Preference*, Vol. 94, p. 104300, doi: 10.1016/j.foodqual.2021.104300.

Bohlol, P., Hosseinpour, S. and Firouz, M.S. (2025), “Improved food recognition using a refined ResNet50 architecture with improved fully connected layers”, *Current Research in Food Science*, Elsevier, Vol. 10, p. 101005.

Castillo, O., Sanchez, M.A., Gonzalez, C.I. and Martinez, G.E. (2017), “Review of Recent Type-2 Fuzzy Image Processing Applications”, *Information*, Multidisciplinary Digital Publishing Institute, Vol. 8 No. 3, p. 97, doi: 10.3390/info8030097.

Cederberg, C. and Sonesson, U. (2011), *Global Food Losses and Food Waste: Extent, Causes and Prevention; Study Conducted for the International Congress Save Food! At Interpack 2011, [16 - 17 May], Düsseldorf, Germany*, edited by Gustavsson, J., presented at the International Congress Save Food!, Food and Agriculture Organization of the United Nations, Rome.

Celard, P., Iglesias, E.L., Sorribes-Fdez, J.M., Romero, R., Vieira, A.S. and Borrajo, L. (2023), “A survey on deep learning applied to medical images: from simple artificial neural networks to generative models”, *Neural Computing and Applications*, Vol. 35 No. 3, pp. 2291–2323, doi: 10.1007/s00521-022-07953-4.

Chauhan, R., Ghanshala, K.K. and Joshi, R.C. (2018), “Convolutional neural network (CNN) for image detection and recognition”, *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, IEEE, pp. 278–282.

Chen, L.-C., Papandreou, G., Schroff, F. and Adam, H. (2017), “Rethinking Atrous Convolution for Semantic Image Segmentation”, arXiv, 5 December, doi: 10.48550/arXiv.1706.05587.

Faezirad, M., Pooya, A., Naji-Azimi, Z. and Amir Haeri, M. (2021a), “Preventing food waste in subsidy-based university dining systems: An artificial neural network-aided model under uncertainty”, *Waste Management & Research: The Journal for a Sustainable Circular Economy*, Vol. 39 No. 8, pp. 1027–1038, doi: 10.1177/0734242X211017974.

Faezirad, M., Pooya, A., Naji-Azimi, Z. and Amir Haeri, M. (2021b), “Demand Prediction in University Booking Systems to Reduce Food Waste Using Neural Networks Including Weighted Error Function”, *Industrial Management Journal*, University of Tehran, Vol. 13 No. 2, pp. 193–170, doi: 10.22059/imj.2021.318760.1007821.

Farinella, L., Fernandes, E., Michener, N., Polimeni, M. and Vesonder, G. (2020), “Identification of Food Waste through Object Recognition”, *2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, presented at the 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), pp. 0496–0499, doi: 10.1109/UEMCON51285.2020.9298165.

Ferreira, J., Cerqueira, P. and Ribeiro, J. (2025), “Food Waste Detection in Canteen Plates Using YOLOv11”, *Applied Sciences*, Vol. 15 No. 13, p. 7137, doi: 10.3390/app15137137.

Geetha, S., Saha, J., Dasgupta, I., Bera, R., Lawal, I.A. and Kadry, S. (2022), "Design of Waste Management System Using Ensemble Neural Networks", *Designs*, Multidisciplinary Digital Publishing Institute, Vol. 6 No. 2, p. 27, doi: 10.3390/designs6020027.

Guo, M.-H., Xu, T.-X., Liu, J.-J., Liu, Z.-N., Jiang, P.-T., Mu, T.-J., Zhang, S.-H., *et al.* (2022), "Attention mechanisms in computer vision: A survey", *Computational Visual Media*, Vol. 8 No. 3, pp. 331–368, doi: 10.1007/s41095-022-0271-y.

Kohli, K., Prajapati, R., Shah, R., Das, M. and Sharma, B.K. (2024), "Food waste: environmental impact and possible solutions", *Sustainable Food Technology*, Vol. 2 No. 1, pp. 70–80, doi: 10.1039/D3FB00141E.

Lekavičius, V., Bobinaitė, V., Kliaugaitė, D. and Rimkūnaitė, K. (2023), "Socioeconomic Impacts of Food Waste Reduction in the European Union", *Sustainability*, Multidisciplinary Digital Publishing Institute, Vol. 15 No. 13, p. 10151, doi: 10.3390/su151310151.

Lubura, J., Pezo, L., Sandu, M.A., Voronova, V., Donsì, F., Šic Žlabur, J., Ribić, B., *et al.* (2022), "Food Recognition and Food Waste Estimation Using Convolutional Neural Network", *Electronics*, MDPI, Vol. 11 No. 22, p. 3746.

Mazloumian, A., Rosenthal, M. and Gelke, H. (2020), "Deep Learning for Classifying Food Waste", arXiv, 6 February, doi: 10.48550/arXiv.2002.03786.

Moumane, K., El Asri, I., Cheniguer, T. and Elbiki, S. (2023), "Food Recognition and Nutrition Estimation using MobileNetV2 CNN architecture and Transfer Learning", *2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA)*, presented at the 2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA), pp. 1–7, doi: 10.1109/SITA60746.2023.10373725.

Rokhva, S. and Teimourpour, B. (2025a), "Artificial Intelligence in the Food Industry: Food Waste Estimation based on Computer Vision, a Brief Case Study in a University Dining Hall", arXiv, 19 July, doi: 10.48550/arXiv.2507.14662.

Rokhva, S. and Teimourpour, B. (2025b), "Accurate & real-time food classification through the synergistic integration of EfficientNetB7, CBAM, transfer learning, and data augmentation", *Food and Humanity*, Vol. 4, p. 100492, doi: 10.1016/j.foohum.2024.100492.

Rokhva, S., Teimourpour, B. and Soltani, A.H. (2024), "Computer vision in the food industry: Accurate, real-time, and automatic food recognition with pretrained MobileNetV2", *Food and Humanity*, Vol. 3, p. 100378, doi: 10.1016/j.foohum.2024.100378.

Ronneberger, O., Fischer, P. and Brox, T. (2015), "U-Net: Convolutional Networks for Biomedical Image Segmentation", arXiv, 18 May, doi: 10.48550/arXiv.1505.04597.

Said, Z., Sharma, P., Thi Bich Nhuong, Q., Bora, B.J., Lichtfouse, E., Khalid, H.M., Luque, R., *et al.* (2023), "Intelligent approaches for sustainable management and valorisation of food waste", *Bioresource Technology*, Vol. 377, p. 128952, doi: 10.1016/j.biortech.2023.128952.

Sarangi, P.K., Pal, P., Singh, A.K., Sahoo, U.K. and Prus, P. (2024), "Food Waste to Food Security: Transition from Bioresources to Sustainability", *Resources*, Multidisciplinary Digital Publishing Institute, Vol. 13 No. 12, p. 164, doi: 10.3390/resources13120164.

Sarraf, A., Azhdari, M. and Sarraf, S. (2021), "A Comprehensive Review of Deep Learning Architectures for Computer Vision Applications", *American Scientific Research Journal for Engineering, Technology, and Sciences*, Vol. 77 No. 1, pp. 1–29.

Seberini, A. (2020), "Economic, social and environmental world impacts of food waste on society and Zero waste as a global approach to their elimination", edited by Kliestik, T.*SHS Web of Conferences*, Vol. 74, p. 03010, doi: 10.1051/shsconf/20207403010.

Slorach, P.C., Jeswani, H.K., Cuéllar-Franca, R. and Azapagic, A. (2019), "Environmental and economic implications of recovering resources from food waste in a circular economy", *Science of The Total Environment*, Vol. 693, p. 133516, doi: 10.1016/j.scitotenv.2019.07.322.

Talaei Khoei, T., Ould Slimane, H. and Kaabouch, N. (2023), "Deep learning: systematic review, models, challenges, and research directions", *Neural Computing and Applications*, Vol. 35 No. 31, pp. 23103–23124, doi: 10.1007/s00521-023-08957-4.

Turker, G.F. (2025), “Reducing Food Waste in Campus Dining: A Data-Driven Approach to Demand Prediction and Sustainability”, *Sustainability*, Vol. 17 No. 2, p. 379, doi: 10.3390/su17020379.

Wang, L., Liu, G., Liu, X., Liu, Y., Gao, J., Zhou, B., Gao, S., *et al.* (2017), “The weight of unfinished plate: A survey based characterization of restaurant food waste in Chinese cities”, *Waste Management*, Elsevier BV, Vol. 66, pp. 3–12, doi: 10.1016/j.wasman.2017.04.007.

Yadav, S. and Chand, S. (2021), “Food image recognition based on Mobile NetV2 using support vector machine”, *AIJR Proceedings*, AIJR Publisher.

باجلان، فاطمه؛ ویسی، هادی؛ و خانساری، محمد. «مروری بر شبکه عصبی پیچشی و معماری‌های مختلف آن». *علوم پایه دانشگاه تهران*، دوره ۶، شماره ۱، بهار ۱۴۰۰، صص. ۲۰–۲۹.