Review of the deep learning for food image processing

Chenrui Niu[1], Xiayang Ying[2*], Gan Pei[1], Menghan Hu[1*], Guangtao Zhai[3]

(1. *Shanghai Key Laboratory of Multidimensional Information Processing, School of Communication and Electronic Engineering, East China Normal University, Shanghai 200241, China*;
2. *Department of General Surgery, Pancreatic Disease Center, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai 200025, China*;
3. *Istitute of Image Communication and Network Engineering, Shanghai Jiao Tong University Shanghai 200240, China*)

Abstract: As deep learning techniques are increasingly applied with greater depth and sophistication in the food industry, the realm of food image processing has progressively emerged as a central focus of research interest. This work provides an overview of key practices in food image processing techniques, detailing common processing tasks including classification, recognition, detection, segmentation, and image retrieval, as well as outlining metrics for evaluating task performance and thoroughly examining existing food image datasets, along with specialized food-related datasets. In terms of methodology, this work offers insight into the evolution of food image processing, tracing its development from traditional methods extracting low and intermediate-level features to advanced deep learning techniques for high-level feature extraction, along with some synergistic fusion of these approaches. It is believed that these methods will play a significant role in practical application scenarios such as self-checkout systems, dietary health management, intelligent food service, disease etiology tracing, chronic disease management, and food safety monitoring. However, due to the complex content and various types of distortions in food images, further improvements in related methods are needed to meet the requirements of practical applications in the future. It is believed that this study can help researchers to further understand the research in the field of food imaging and provide some contribution to the advancement of research in this field.

Keywords: deep learning, food image processing, feature extraction, dietary health
DOI: 10.25165/j.ijabe.20241705.8975

Citation: Niu C R, Ying X Y, Pei G, Hu M H, Zhai G T. Review of the deep learning for food image processing. Int J Agric & Biol Eng, 2024; 17(5): 15–30.

1 Introduction

As society continues to develop and the standard of living for people improves, there is a growing concern about food safety issues and dietary nutrition. This concern, which has become one of the major health challenges facing society, stems from the real-life data of food safety issues, rising obesity, and chronic diseases globally. In 2010, there were an estimated 600 million foodborne illnesses and 420 000 deaths worldwide in just that year, while in the United States foodborne illnesses result in 325 000 hospitalizations and 5000 deaths annually[1]. Globally, it is estimated that there are more than 1 billion episodes of diarrhea associated with food poisoning each year, resulting in the deaths of about 3 million children[2]. At the same time, modern diets generally favor foods high in sugar, fat, and salt, and long-term intake of excess calories and unbalanced nutrients can lead to obesity, high blood pressure, high blood sugar, and other health problems. Since 1975, the prevalence of overweight and obesity has almost tripled[3], and according to the World Obesity Federation, it is predicted that by 2030, 1 billion people will be suffering from obesity globally. The impact of obesity is not only limited to adults, and is expected to increase significantly in the coming years. The COVID-19 pandemic has limited people's overall mobility[4], and this, coupled with the sedentary lifestyle of many, will further exacerbate the obesity crisis[5]. If effective measures are not taken, the prevalence of obesity may be accelerated. Meanwhile, non-communicable diseases (NCDs), such as cardiovascular diseases, diabetes, and certain cancers, have become one of the major contributing factors to global mortality[6]. Overall, these health problems are often closely related to poor dietary habits, so food work for dietary health will be of increasing public interest (Figure 1)[7-9].

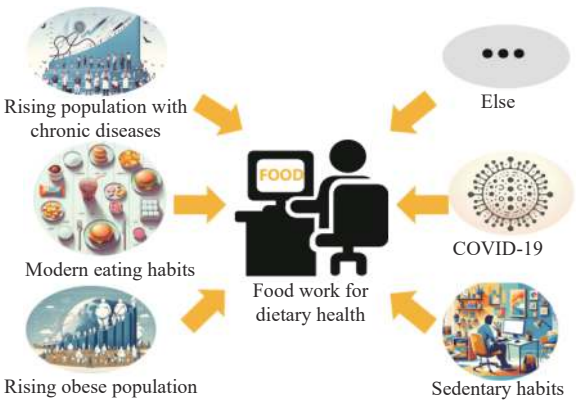


Figure 1 Factors contributing to the development of food image processing for healthy diets

Received date: 2024-04-01 **Accepted date:** 2024-08-25
Biographies: Chenrui Niu, MS, research interest: intelligent information processing of medical images, Email: 51265904046@stu.ecnu.edu.cn; **Gan Pei**, MS, research interest: intelligent information processing of medical images, Email: 51265904018@stu.ecnu.edu.cn; **Guangtao Zhai**, Professor, research interest: multimedia signal processing, Email: zhaiguangtao@sjtu.edu.cn.
***Corresponding author: Xiayang Ying**, MD, research interest: pancreatic cancer diagnosis and treatment. Shanghai Jiaotong University School of Medicine, Shanghai 200025, China. Email: yingxiayang@hotmail.com; **Menghan Hu**, Associate Professor, research interest: intelligent information processing for rehabilitation medicine. School of Communication and Electronic Engineering, East China Normal University, Shanghai 200241, China. Email: mhhu@ce.ecnu.edu.cn.

Automated food image analysis contributes to intelligent dietary management and disease causality analysis. Therefore, in recent years, researchers have employed deep learning methods to analyze food images. Deep learning has emerged as a cutting-edge technique for analyzing large data, demonstrating wide and effective applications across various tasks, including image processing, classification, recognition, object detection, and image understanding[10]. The advent of deep learning has ushered in innovative methodologies and tools for food inspection, nutritional evaluation, and health management. Utilizing deep learning-based image recognition algorithms, it is possible to swiftly and accurately assess the freshness of food items and identify any spoilage to ensure food safety[11]. Moreover, deep learning plays a pivotal role in analyzing food composition and nutritional content, offering personalized nutritional advice to individuals or healthcare professionals. This contributes significantly to the enhancement of dietary structures and promotes dietary health[12].

Initially, the detection of food safety was carried out manually[13-15], a method inherently prone to delays, subjectivity, and inevitable human errors, often resulting in low efficiency in recording and analyzing critical data. Even with comprehensive records, the lack of professional nutritional knowledge among the general populace hindered independent analysis and subsequent dietary planning. This situation led to prolonged dietary imbalances, underscoring the vital importance of advancements in food-related work. The focus on dietary health today involves maintaining the body's well-being through balanced intake of nutrients, including carbohydrates, fats, proteins, vitamins, minerals, and water. It emphasizes the diversity, balance, and moderation of diet to ensure the body receives requisite nutrients while avoiding excessive intake of unhealthy components like added sugars, high salt, and fats. Crafting a sound dietary health plan should be based on factors such as an individual's age, gender, level of activity, health condition, and personal preferences. Personalized dietary strategies are instrumental in preventing various chronic diseases, such as heart disease, obesity[3], diabetes, and some forms of cancer[9], thereby significantly enhancing an individual's health level and quality of life.

In today's digital age, when smartphones, computers, and smartwatches are ubiquitous, tracking and recording dietary intake will become exceptionally simpler. This technological progress has spurred the development of various dietary mobile applications, such as DietLens[16], Dietcam[17], and Im2Calories[18], which combine food recognition technology with artificial intelligence models to automate food and ingredient detection and logging. In the early work on food image processing, researchers primarily relied on manual features to recognize and classify food images[14,15], showing some effectiveness in handling simple food scenes. Relying solely on manually designed features has proved insufficient to capture the diversity of foods, complex backgrounds, and the nuanced semantic information in changing food images. This limitation hindered the accuracy and robustness of food image processing efforts. The rise of deep learning has brought about a revolutionary change in food image processing. Utilizing deep learning architectures, e.g., Convolutional Neural Networks (CNNs), it is possible to automatically learn deep features from extensive food image datasets without manual intervention in feature design and selection. This significant leap forward not only improved the performance of food image analysis but also paved the way for intelligent health management systems to offer realtime monitoring and personalized dietary advice. By integrating deep learning algorithms, these systems can now assist in refining individual dietary habits and nutrition, making professional guidance more accessible than ever before. The structure of this survey is shown in Figure 2, and the distribution of the number of references used in recent years is shown in Figure 3. The main objectives of this survey are as follows:

1) Differentiate and introduce common food image processing efforts and outline the commonly used evaluation metrics for these tasks. Additionally, compile and collect commonly used datasets related to food images and provide open-source addresses;

2) Highlight the unique challenges and difficulties currently faced in food image processing, from traditional methods to deep learning approaches, summarize and showcase existing solutions;

3) Discuss the specific applications of deep learning in food image processing and look into future work in the field.

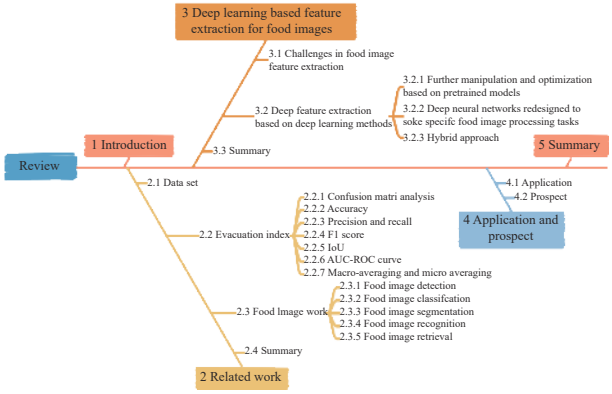


Figure 2 Overall framework of the survey.

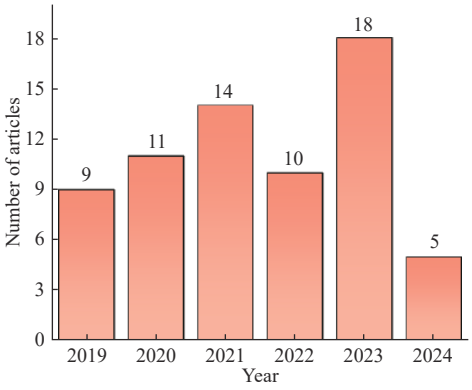


Figure 3 Count of research articles included in the study

2 Related work

The broad scope of food categorization covers a wide range of food products, from the basic categories of meat, vegetables, and fruits to the more detailed subcategories of beef, lamb, and fish, which can even be further subdivided into more specific categories such as yak meat and yellow beef. This hierarchical categorization system reflects not only the diversity of foods, but also the gradual increase in the importance and awareness of different ingredients. The categorization of food is not limited to single ingredients but also depends on the dish preparation methods and the combination of ingredients, resulting in dishes that exhibit multiple levels of granularity. For instance, the same meat ingredient, with different cooking methods and matching ingredients, may be categorized into different dishes or food categories, such as steamed fish or braised fish. Therefore, food-related work needs to take into account not only the basic categorization of ingredients, but also factors such as

the method of preparation of the dish, the ingredients used in combination, and the variability of the food in different geographical and cultural contexts. This nuanced classification system puts forward higher requirements for research and application in food-related fields and offers more possibilities for better understanding and utilization of food resources.

2.1 Dataset

CNNs have been around since the late 1980s to the early 1990s, but their really rapid development occurred in the early 21st century. With the development of CNNs, large-scale datasets have also seen significant progress and development in general-purpose visual recognition, but have largely lagged behind in the food domain[19,20]. However, the size of food-centered datasets has been steadily rising. These include PFID, a Western food dataset with 101 categories totaling 4545 images in 2009[21]; Food85, which consists of 85 categories totaling 8500 food items in 2010[22]; and the Japanese food, which was made publicly available by a team of researchers from the University of Electricity and Power of Tokyo (UEC) in 2012 and 2014. Other such datasets include UECFood-100[23] and UECFood-256[24], with a total of tens of thousands of images; the Western food dataset ETH Food101[25], constructed by Bossard et al. using 101 000 images from 101 food categories; ChineseFoodNet, proposed by Chen et al.[26] and including a total of 192 000 images of 208 categories of Chinese food; FoodX-251[27], and ISIA Food-500[28], both of which were introduced in 2019 and include hundreds of thousands of food image samples; and the large-scale dataset ImageNet[29]. In 2023, Min et al., with the help of Meituan, completed the largest food dataset currently available, Food2k[20], which contains a total of one million images of Chinese and Western food products, with a total of 2000 food categories. The Food2k dataset is at least an order of magnitude higher than the existing datasets in terms of the number of images and the size and number of categories. These datasets are more or less different in terms of data volume, classification methods, classification levels, and label formats. Detailed statistics are summarized in Table 1. In addition, Figure 4 is a comparative scatter plot of the number of samples from these datasets, and Figure 5 shows the image samples and label samples from several datasets. There are also other food-related recipe datasets, such as Epicurious Recipes Dataset[30], Yummly66K[16], and Recipe1M[31], which contain ingredients and recipes used in the preparation of dishes, and can be used to assist food classification tasks by using the internal linkages or semantics of dishes. For instance, lettuce in “vegetable salad” is more likely than fish, and the meat in “stir-fried yellow beef” is more likely to be beef. This can be used to assist in food categorization, or in multimodal food image and text joint learning, food image datasets based on hyperspectral imaging HSIFoodIngr-64[32], and other special food datasets.

2.2 Evaluation index

The selection of appropriate evaluation metrics is essential to assess model performance. This survey summarizes the commonly used evaluation metrics in food-related work studies, including accuracy, precision, recall, F1 score, IoU and confusion matrices, AUC-ROC curves, and microaveraging. These evaluation metrics play a great role in various food imaging tasks. Definitions, commonly used scenarios and formulas for some of these metrics are given in Table 2 and subsequent paragraphs. Specifically, these evaluations need to take into account the unique challenges associated with food images, such as different presentation styles, occlusions, and complex textures. The discussion will therefore focus on the relevance of these metrics to food image processing.

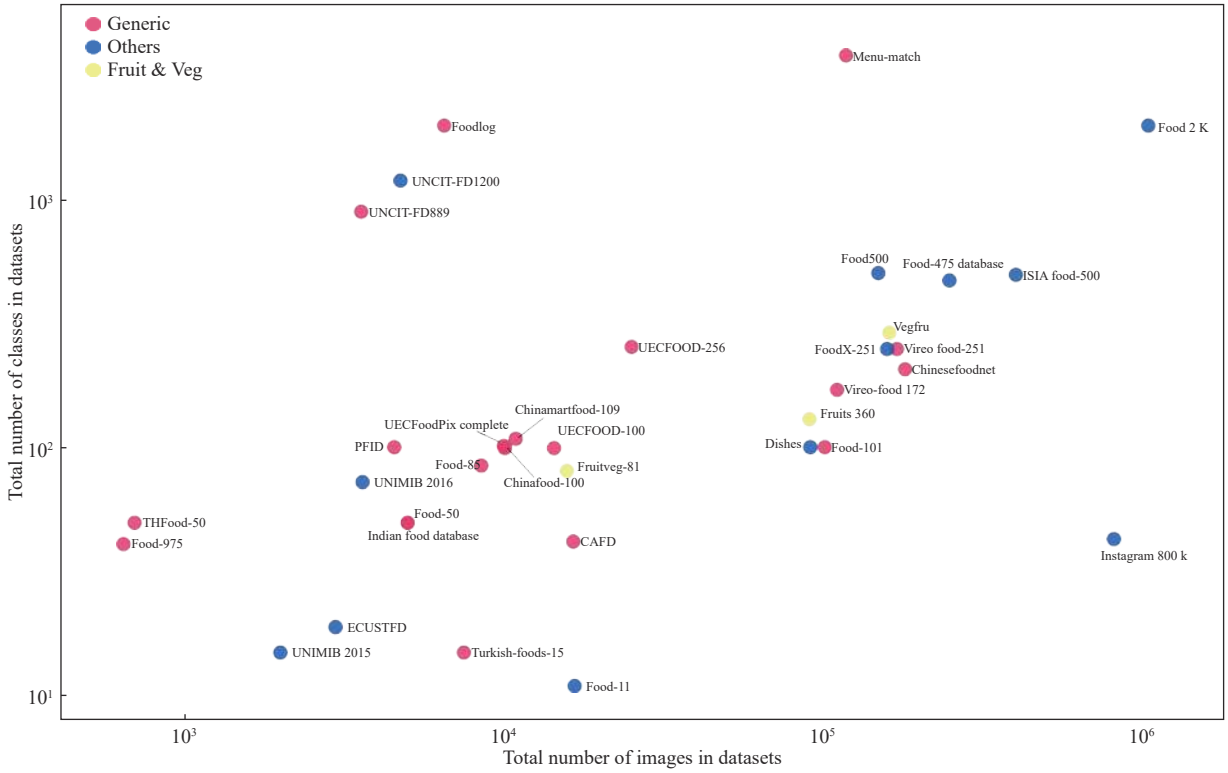
Table 1 Food image datasets

Year	Food image dataset	Category	No.of image	No.of class	Public access
2009	PFID[21]	American Foods	4545	101	×
2010	Food-50[33]	Japanese Foods	5000	50	×
2010	Food-85[22]	Japanese Foods	8500	85	×
2011	Foodlog[34]	Japanese Foods	6512	2000	×
2012	UECFOOD-100[23]	Japanese Foods	14 361	100	√
2014	UECFOOD-256[24]	Japanese Foods	25 088	256	√
2014	Food-101[25]	American Foods	101 000	101	√
2014	Rice dataset[35]	Generic (Rice)	×	1	√
2014	UNCIT-FD889[36]	Italian Foods	3583	899	√
2014	UNIMIB 2015[37]	Generic	2000	15	×
2015	FOOD201-Segmented[18]	American Foods	12 625	×	×
2015	UPMC Food-101[38]	Generic	90 840	101	√
2015	TADA(19 foods)[39]	American Foods	×	19	×
2015	Dishes[40]	Chinese Foods	117 504	3832	×
2015	Menu-Match[41]	Chinese Foods	646	41	√
2016	Food-975[42]	Chinese Foods	37 785	975	×
2016	Vireo-Food 172[43]	Chinese Foods	110 241	172	√
2016	UNIMIB 2016[44]	Generic	3616	73	√
2016	Food500[45]	Generic	148 408	508	×
2016	Food-11[46]	Generic	16 643	11	√
2016	UNCIT-FD1200[47]	Generic	4754	1200	√
2016	Food-475 Database[48]	Generic	247 636	475	√
2016	Instagram 800k[49]	Generic	808 964	43	√
2017	ChineseFoodNet[25]	Chinese Foods	179 920	208	√
2017	ECUSTFD[50]	Generic	2978	19	√
2017	Turkish-Foods-15[51]	Turkish Dishes	7500	15	×
2017	Indian Food Database[52]	Indian Foods	5000	50	√
2017	THFood-50[53]	Thai Food	700	50	√
2017	VegFru[54]	Generic (Fruit and VEG)	160 731	292	√
2017	FruitVeg-81[55]	Generic (Fruit and VEG)	15 737	81	√
2018	Fruits 360[56]	Fruits	90 380	131	√
2019	FoodX-251[26]	Generic	158 000	251	√
2020	ISIA Food-500[57]	Generic	399 726	500	√
2021	Vireo Food-251[58]	Chinese Foods	169 673	251	√
2021	UECFoodPix Complete[59]	Japanese Foods	10 000	102	√
2021	ChinaFood-100[60]	Chinese Foods	10 074	100	√
2021	Food2K[19]	Generic	1 036 564	2000	√
2022	ChinaMartFood-109[61]	Chinese Foods	10 900	109	×
2022	CNFOOD-241[62]	Chinese Foods	191 881	241	√
2023	CAFD[63]	Central Asian Foods	16 499	42	√

Note: Public access data source: https://www.kaggle.com/datasets/rkuo2000/uec-food100; https://www.kaggle.com/datasets/rkuo2000/uecfood256; https://www.kaggle.com/datasets/kmader/food41; https://www.kaggle.com/datasets/muratkoklu-dataset/rice-image-dataset; https://iplab.dmi.unict.it/UNICT-FD889/; https://www.kaggle.com/datasets/gianmarco96/upmcfood101; https://neelj.com/projects/menu-match/data/; https://fvl.fudan.edu.cn/dataset/vireofood172/list.htm; https://www.v7labs.com/open-datasets/unimib-food-database; https://www.kaggle.com/datasets/trolukovich/food11-image-dataset; https://iplab.dmi.unict.it/UNICT-FD1200/; https://www.v7labs.com/open-datasets/food-475-database; https://www.instagram.com/explore/tags/800k/; https://sites.google.com/view/chinesefoodnet/; https://paperswithcode.com/dataset/ecustfd; https://www.ifct2017.com/; https://paperswithcode.com/dataset/thfood-50; https://paperswithcode.com/dataset/vegfru; https://www.payititi.com/opendatasets/show-1266.html; https://www.kaggle.com/datasets/moltean/fruits; https://paperswithcode.com/dataset/foodx-251; https://paperswithcode.com/dataset/isia-food-500; https://fvl.fudan.edu.cn/dataset/vireofood251/list.htm; https://mm.cs.uec.ac.jp/uecfoodpix/; https://cinnamonsociety.com/recipes/blog/100-chinese-foods-to-try-before-you-die; https://paperswithcode.com/dataset/food2k; https://www.kaggle.com/datasets/zachaluza/cnfood-241; https://paperswithcode.com/dataset/cafd.

2.2.1 Confusion matrix

The confusion matrix provides a granular view of the model’s performance by displaying the distribution of true positives, false positives, true negatives, and false negatives across different food



Note: Red indicates generic food image datasets, yellow indicates vegetable and fruit image datasets, and blue indicates a single country or region food image datasets.

Figure 4 Distribution of sample size in current food datasets

Table 2 Index commonly used in food image processing

Evaluation index	Definition	Use (general)	Main formula or expression
Confusion matrix	A table that describes the performance of the classification model, including the number of true, false positive, true negative, and false negative.	Classification, Detection	Figure 6a
Accuracy	Ratio of correctly predicted outcomes to total outcomes.	Classification	$\frac{(TP+FN)}{(TP+FP+FN+TN)}$
Precision	Proportion of predicted positive outcomes that are actually positive.	Classification, Recognition, Detection, Image Retrieval	$\frac{TP}{(TP+FP)}$
Recall	Proportion of actual positive outcomes that are correctly predicted to be positive.	Classification, Recognition, Detection, Image Retrieval	$\frac{TP}{(TP+FN)}$
F1 score	Harmonized mean of precision and recall rates.	Classification, Recognition, Detection, Segmentation	$\frac{2\times(Precision\times Recall)}{Precision+Recall}$
IoU	Ratio of intersection and concatenation of predicted and actual objects.	Detection, Segmentation	$\frac{Regional\ Convergence}{Regional\ Union}$
AUC-ROC Curve	Plot of true and false positive rates with better performance for larger areas.	Recognition, Image Retrieval	$TPR=\frac{TP}{TP+FN}$, $FPR=\frac{FP}{FP+TN}$ Figure 6b
Macro-averaging	In multi-class classification, Macro-averaging refers to calculating metrics (e.g., Precision, Recall) individually for each class and then averaging these metrics across all classes. Each class is treated equally, with the same weight in the final average.	Image Retrieval	$Macro-Aver=\frac{1}{N}\sum_{i=1}^{N}Metric\ for\ Class i$
Micro-averaging	Total TP, FP, and FN are counted across all categorization decisions and then performance metrics are calculated.	Image Retrieval	$Micro-Aver=\frac{Overall\ TP, FP, TN, or\ FN}{Overall\ number\ of\ instances}$

categories. In food image classification, this is crucial for understanding the model's ability to distinguish between visually similar foods, such as different types of pasta or varying cooking stages of the same dish, as well as between foods with similar textures, like soups and sauces.

2.2.2 Accuracy

Accuracy measures the overall proportion of correctly classified food images, but it can be misleading in imbalanced datasets. A high accuracy might mask poor performance on rare or visually similar categories, highlighting the need for additional metrics.

2.2.3 Precision and recall

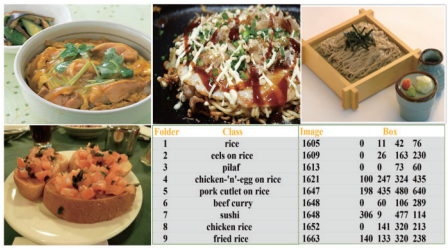
Precision and recall are particularly important in food image tasks where the distinction between categories can be subtle. Precision indicates the proportion of correctly identified positive instances (e.g., correctly identifying a salad), which is critical when the consequences of false positives are significant, such as in dietary assessments. Recall measures the model's ability to identify all relevant instances of a food category, which is essential for comprehensive food recognition, ensuring that all instances of a particular food type are detected, regardless of variations in presentation or occlusion.

2.2.4 F1 score

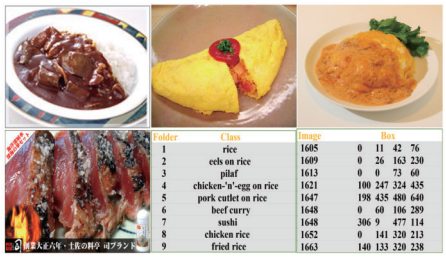
The F1 score, which balances precision and recall, is valuable in food image processing, particularly for handling imbalanced

a. Food-101

b. Food-2k



c. UECFood100



d. UECFood-256

Figure 5 Sample images and labels

datasets where some food categories are underrepresented. It ensures that the model's performance is robust across both common and rare food types, providing a more comprehensive evaluation.

2.2.5 IoU

IoU is a critical metric in food image segmentation tasks, where the goal is to accurately delineate food items from the background or other objects on a plate. IoU measures the overlap between the predicted segmentation mask and the ground truth, and a higher IoU indicates better performance. This is particularly important in applications like portion size estimation or ingredient identification, where precise segmentation is necessary.

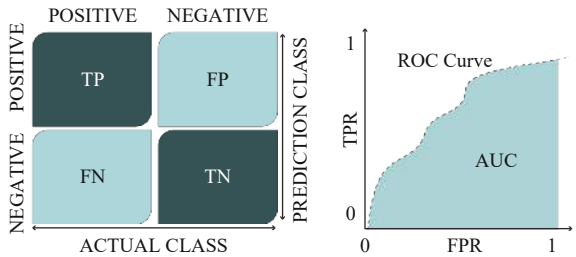
2.2.6 AUC-ROC curve

The AUC-ROC curve, which plots the true positive rate against the false positive rate, is useful in food image classification to evaluate model performance across different thresholds. This is relevant in scenarios like allergen detection, where the model's ability to minimize false negatives is critical, and a high area under the curve indicates a model that can reliably distinguish between allergenic and non-allergenic foods.

2.2.7 Macro-averaging and micro-averaging

In multi-class food categorization tasks, where the dataset includes a wide variety of food categories, macro-averaging and micro-averaging are used to evaluate overall model performance. Macro-averaging treats all categories equally, providing insight into the model's ability to perform well across all types of food, regardless of frequency. Micro-averaging, on the other hand, gives more weight to categories with more samples, making it useful when some food types are more prevalent in the dataset.

Beyond commonly used metrics such as accuracy, recall, and IoU, several other evaluation metrics are also critical in food image processing. For example, the R^2 score is valuable in regression tasks like food quantity estimation, measuring the correlation between the model's predictions and actual values. The Dice Coefficient is essential for segmentation tasks, assessing the overlap between predicted and actual segmentation masks. Metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are often used in tasks like calorie estimation to quantify prediction errors. Sensitivity and Specificity are crucial in tasks of the correct identification of positive and negative examples, such as detecting allergens in food. The Fowlkes-Mallows Index (FMI) and Adjusted Rand Index (ARI) can be applied in clustering tasks to evaluate the quality of groupings within food image data. These diverse metrics ensure a more comprehensive assessment of model performance, addressing the various challenges encountered in food image analysis.



a. Confusion matrix b. ROC and AUC

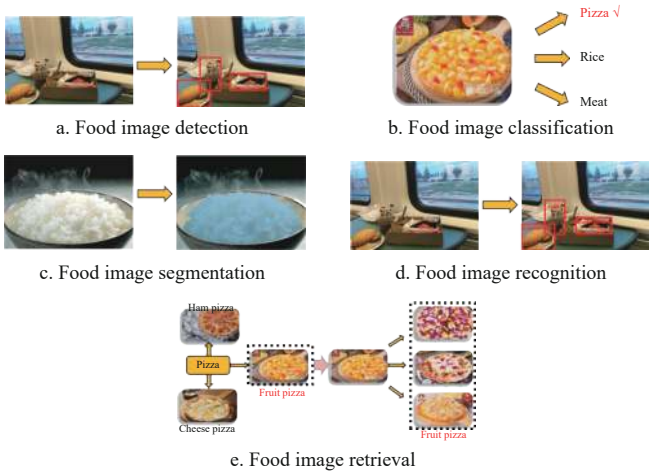
Figure 6 Supporting charts for Table 2

2.3 Task types in food image processing

2.3.1 Food image detection

Food detection involves identifying the presence of food items within images or videos, typically requiring the precise localization of these items. This task is critical for applying deep learning to food safety and quality monitoring, with objectives that include automatically identifying foreign objects in food, detecting food quality, and recognizing signs of spoilage or contamination. Achieving these goals relies on advanced image analysis techniques and complex algorithmic models to ensure safe and healthy food consumption (Figure 7a). However, food image detection presents several unique challenges. The visual appearance of food items can vary greatly due to factors like cooking methods, lighting conditions, and the presence of overlapping or occluding objects. These variations make it difficult to consistently detect and identify food items across different contexts. Additionally, distinguishing between visually similar food items and differentiating food from non-food objects, particularly foreign contaminants, can be challenging. For instance, Rong et al.[64] introduced a method for the automatic detection of foreign objects in walnuts using deep learning. The key evaluation metrics included the percentage of correctly segmented object regions and the percentage of foreign objects correctly classified. Their approach successfully learns features directly from the training data, eliminating the need for manual feature extraction. This method addresses the significant challenge of distinguishing between walnuts and foreign objects in real images, highlighting the complexities involved in food image

detection. In summary, while food image detection is a powerful tool for enhancing food safety and quality, it requires overcoming significant challenges related to the variability of food appearance and the need for precise differentiation between food items and potential contaminants.



a. Food image detection b. Food image classification

c. Food image segmentation d. Food image recognition

e. Food image retrieval

Figure 7 Simple schematic of task types in food image processing.

2.3.2 Food image classification

Food image classification is a vital area of deep learning, focusing on the automatic recognition and categorization of food items in images. The primary objective is to accurately classify food images into predefined categories such as fruits, vegetables, meats, and bread, while also distinguishing specific food products within these broader categories. However, food image classification presents unique challenges, including high variability in food appearance due to differences in cooking styles, presentation, lighting conditions, and camera angles; these factors lead to significant intra-class variability, making it difficult for models to consistently identify and classify food items. Additionally, certain foods may share similar textures, colors, or shapes, resulting in inter-class similarities that further complicate classification. To address these challenges, researchers have explored various deep learning models and techniques. Memiç et al.[65] conducted extensive experiments using the UEC Food-100 dataset, evaluating models such as ResNet-18, Inception-V3, ResNet-50, DenseNet-121, Wide ResNet-50, and ResNext-50, and found that ResNext-50 achieved the highest accuracy at 87.7%. Ghosh et al.[66] identified ResNet as an effective model for classifying food intake from sensor signals in their study on advanced time series deep learning algorithms. Min et al.[19] demonstrated a significant improvement over other state-of-the-art methods on the Food2k dataset with their proposed PRENet model; Yang et al.[67] developed four models - 2D-CNN, 3D-CNN, R-2D-CNN, and R-3D-CNN - to improve hyperspectral image classification of food products, with their R-3D-CNN model excelling in accuracy and convergence speed. These findings underscore the potential of deep learning models in overcoming the unique challenges of food image classification, paving the way for more accurate and efficient categorization in this complex domain.

2.3.3 Food image segmentation

The food image segmentation task is pivotal in accurately distinguishing between different food items and their backgrounds, enabling the precise identification and analysis of each food item's location, shape, and size. This task becomes especially critical in complex food scenes, where multiple foods are mixed together. High-precision segmentation not only facilitates the analysis of food type, quantity, and placement but also provides essential data for nutritional analysis, calorie estimation, and catering services, as depicted in Figure 7c. For example, Siemon et al.[68] proposed a hierarchical clustering sequential transfer learning method that enhanced food segmentation accuracy by 6% and demonstrated greater robustness in monitoring Danish school children's food services. Furthermore, Meta's development of the Segment Anything Model (SAM)[69] for general image segmentation, and its subsequent iteration, SAM2[70], for both image and video segmentation, marked significant advancements in the field. Building on the foundation of SAM, Lan et al.[48] introduced an innovative zero-shot framework called FoodSAM. Rather than simply fine-tuning the SAM model, FoodSAM integrates the coarse semantic masks generated by SAM with category-agnostic masks, resulting in improved semantic segmentation quality. Additionally, FoodSAM extends its zero-shot capabilities to include instance segmentation, panoptic segmentation, and promptable segmentation, supporting a wide range of prompt variants. Experiments conducted on the FoodSeg103 and UECFoodPix Complete datasets have demonstrated that FoodSAM outperforms existing state-of-the-art methods across various segmentation tasks, making it a notable advancement in food image segmentation research.

2.3.4 Food image recognition

Food recognition is a complex task that involves identifying various types of food from images and categorizing them. The process goes beyond detecting the presence of food, and includes accurately assigning food to specific categories, determining its composition, and even estimating its calorie content. Achieving high performance in food recognition usually requires the integration of advanced image processing and pattern recognition techniques, complemented by large datasets containing images of various foods. Recent advances in this field have led to the development of more efficient methods, such as the Progressive Self-Distillation (PSD) method proposed by Zhu et al.[71] PSD is designed to incrementally enhance the network's ability to extract more detailed and discriminative features from food images. Unlike traditional methods that may focus on recognizing multiple regions in an image, PSD employs a self-distillation process whereby a teacher network and a student network are trained simultaneously. These networks share a common embedding network and the student network receives modified images from the teacher network. Certain regions of these images are occluded, forcing the student network to improve its ability to find and focus on the most informative parts of the image. As training proceeds, the teacher network itself improves, becoming progressively better at recognizing key details and thus improving recognition accuracy; Tan et al.[72] conducted a comparative study for food recognition using the school lunch dataset, the UECFOOD-100 dataset, and the UECFOOD-256 dataset. The results of the study highlight the robust performance of the different models under different conditions. For example, in the school lunch dataset, Faster R-CNN combined with Mobilenet-V3 achieved an impressive mean accuracy (mAP) of 0.931, showing its high accuracy in food recognition. On the other hand, YOLOv5 achieved a mAP of 0.774 and 0.701 on the UECFOOD-100 and UECFOOD-256 datasets, respectively. These results highlight the effectiveness of models such as YOLOv5 when dealing with smaller input image sizes and larger batches, which further solidifies their utility in practical applications.

2.3.5 Food image retrieval

The food image retrieval task aims to retrieve food images that are similar or related to the query image from a large image database. This task is valuable for understanding and analyzing food images, and can help users quickly find relevant information about a specific food product, such as the name of the food product, ingredients, nutritional value, and cooking method. Food image retrieval relies on powerful feature extraction and image matching techniques that analyze the visual content to find items that are similar to the target image (Figure 7e). Saritha et al.[73] demonstrated the application of deep learning in content-based image retrieval (CBIR) by using deep belief networks (DBN) for feature extraction and classification to improve the retrieval performance. Dubey[74] provided a comprehensive survey of the development of deep learning in the last decade of CBIR by categorizing and summarizing the existing state-of-the-art methods, and demonstrates the advances in deep learning techniques through performance analysis; Min et al.[19] conducted a cross-modal recipe-food image task to validate the generalization ability of Food2k dataset, which was evaluated using median retrieval rank (MedR) and recall percentage of top K (Recall K).

2.4 Summary

This section provides an overview of the significant aspects of deep learning in the food domain, focusing on the development of datasets, evaluation metrics, and methods for tasks such as food classification, recognition, detection, segmentation, and image retrieval. It reviews the evolution of food datasets from early small-scale collections to recent extensive and diverse ones, discussing the variety and uniqueness of datasets, including recipe and hyperspectral imaging datasets crucial for food categorization and identification. It also summarizes common food image datasets in terms of number and categories, along with open-source access channels. Furthermore, the section introduces common evaluation metrics in food image processing, essential for accurately assessing and comparing the performance of different models. It explores several key deep learning tasks in the food domain, introducing the objectives of these tasks while showcasing recent related work. These tasks demonstrate the wide applicability of deep learning in food image processing, highlighting current challenges and future research directions, hopefully to provide references and insights for further research.

3 Deep learning-based feature extraction for food images

In the field of dietary health, image feature extraction is crucial for tasks such as food recognition, volume estimation, and nutritional derivation[75,76]. Before the popularity of deep learning methods, researchers applied hand-crafted features[14,15], including features ranging from simple food colors, image textures, shapes, edges, and spatial relationships to LBP (Local Binary Patterns)[77], SIFT (Scale-Invariant Feature Transform)[78], and Gabor Filters[79]. As listed in Table 3, this study demonstrates traditional manual feature-based food image processing efforts. In recent times, the swift evolution of deep learning technology in the realm of image recognition and processing has established it as the go-to approach for extracting features from food images. This progress, particularly in handling food image characteristics, has spurred significant advancements in research related to food identification, categorization, and nutritional assessment. This section summarizes deep learning-based feature extraction methods for food images, introduces advanced feature extraction methods, and aims to provide a more comprehensive framework for understanding how these methods can play a key role in food image analysis and what the future holds for them. Through in-depth analysis of recent excellent research results and application cases, the study reveals how deep learning is revolutionizing the field of food image processing, and expects to provide help for future research and applications.

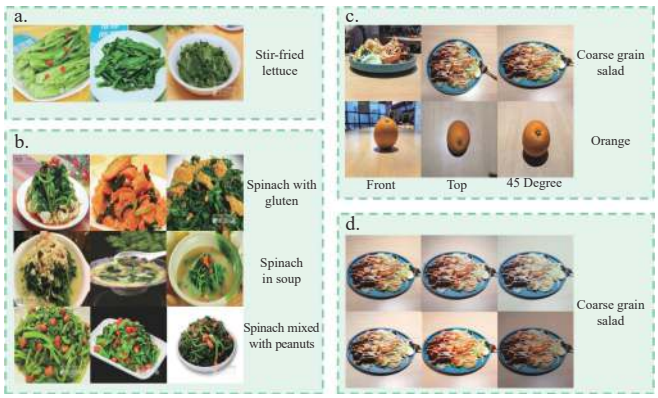
Table 3 Food image processing based on hand-crafted features

Author	Year	Feature	Method	Work	Database
Yang et al.[80]	2010	Spatial feature	STF	Classification	PFID
Matsuda et al.[22]	2012	Spatial Pyramid Representation, HoG, Gabor Texture Feature, and SIFT	SVM	Recognition, Classification	Matsuda
Chen et al.[81]	2012	SIFT, LBP, Gabor, and Color	SVM	Recognition, Classification	Chen
Anthimopoulos et al.[82]	2014	Color, Size, Texture, and Shape	SVM, Random Forest	Recognition, Segmentation	Marios
Oliveira et al.[83]	2014	Color (HSV, LAB, RGB) and Texture (SVC, DoG)	SVM	Recognition, Classification	Oliveira
Kawano et al.[84]	2014	Fisher Vector and RootHoG	One-vs-rest Linear, Classifier	Classification	UECFood-256
Tammachat et al.[85]	2014	Color and Texture	SVM	Recognition, Classification	Tammachat
He et al.[86]	2014	Color (HSV), Texture (EFD, GFD), and Local Area Features (SIFT, MDSIFT)	KNN, Tree	Classification	He
Abdulrahman et al.[87]	2015	Color and SURF Feature	Color and SURF Feature	Image Retrieval, Recognition	Abdulrahman
Ahsani et al.[88]	2019	Color and Texture	GLCM, Lab Color	Image Retrieval, Recognition	Ahsani

3.1 Challenges in food image feature extraction

Deep learning has made significant progress in food image processing, but it still faces a series of unique challenges across various tasks in practical applications. First, the inherent diversity of food is a major challenge, characterized by large intra-class variations and small inter-class variations[89]. Different foods exhibit significant differences in shape, color, size, and texture, especially in cooked foods processed by different cooking techniques, which makes accurate identification and classification more challenging. Moreover, the high diversity of food images demands that deep learning models possess the ability to capture and understand various food categories and their variations. In addition to these challenges, food images are often captured under varying conditions, such as different equipment, lighting, and complex backgrounds. These factors not only increase the complexity of image processing but can also introduce noise that affects the accuracy of feature extraction. Furthermore, issues such as occlusion, overlap between food items, morphological changes during cooking, and interference from non-food objects (such as utensils, plates, or hands) further complicate processing. The

diversity in food presentation and portion sizes also presents a significant challenge. Different plating techniques and food states (such as partially eaten, mixed, or combined with other ingredients) can make it difficult for models to correctly identify the food items. Additionally, the seasonal and regional variations in food types and availability require models to have a broad adaptability. Figure 8 illustrates some of these common challenges encountered in food image processing tasks.



Note: Sample food images: a. lettuce and b. spinach belong to the same vegetable category, with high visual similarity reflecting small intra-class differences; b. shows the same food under different cooking techniques; c. shows the same food seen from different angles; and d. shows the same food seen with different lighting.

Figure 8 Samples of food images that differ due to means of capture and environment

3.2 Deep feature extraction based on deep learning methods

Since AlexNet[90] achieved excellent results in the ImageNet competition in 2012, CNNs have seen a surge of novel and innovative methods and frameworks in recent years, including Multi-task Learning, Domain Adaptation, Weakly Supervised and Unsupervised Learning, Contrastive Learning, Graph Convolutional Networks, Vision Transformers, and Diffusion Models. These methods, especially in food image processing, are often used in conjunction with or improved upon CNNs, or even multiple methods are used in combination as a way to achieve the best results. Recently, in the field of food image processing, the development of deep learning methods has focused on two directions (Figure 9). The first is the further manipulation and optimization based on pre-trained models, which aims to utilize existing deep neural networks trained on large datasets. This is done

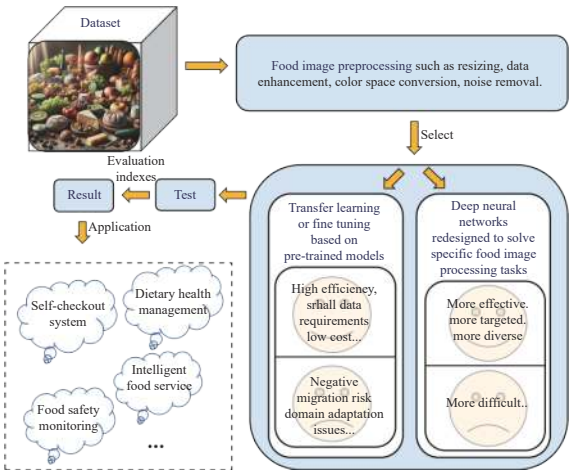


Figure 9 Development directions of deep learning methods in food image processing

to enhance the extraction of deep features from food images through fine-tuning or specific adaptive improvements. The second is the redesign of deep neural networks. These customized models focus more on specific architecture and algorithm designs tailored to the unique properties and task requirements of food images.

3.2.1 Further manipulation and optimization based on pre-trained models

The simplest approach is to use a transfer learning method that applies pre-trained CNN network models (e.g., AlexNet, VGG[91], etc.) on large-scale datasets, such as the ImageNet dataset, as a feature extractor for food images. Since the ImageNet dataset is a generic object dataset, even though the entire ImageNet contains more than 1000 categories of food-related images, it contains spices, kitchenware, and other categories that are not “food”, and therefore the results often have room for improvement. On top of that, there are ways to fine-tune the network so that the CNN can capture more discriminative features[92]. Kagaya et al.[93] applied deep learning networks to dish image recognition by fine-tuning the AlexNet network to extract image features. Wu et al.[94] investigated the use of pre-trained models, with a focus on Contrastive Language-Image Pretraining (CLIP), to address the challenge of food recognition with a limited number of samples. CLIP, developed by OpenAI, is a versatile model that learns to associate images with textual descriptions by jointly training on large-scale image-text pairs. This extensive training endows CLIP with a broad understanding of visual concepts. By fine-tuning CLIP specifically for food recognition, Wu et al. effectively combined its rich prior knowledge with the new insights gleaned from a small number of training samples. This synergy allowed the model to achieve strong recognition performance, even when faced with the challenge of limited data availability. In 2023, Xiong et al.[95] obtained highly accurate food category prediction based on the tuning and fine-tuning of ResNet network. In recent years, several other studies have delved into the performance implications of migration learning from large-scale non-food datasets to small-scale food image datasets. Cai et al.[96] explored how deep CNNs can learn by migration from knowledge gained from large datasets such as ImageNet to small-scale food image datasets. They found that the models that performed better on ImageNet had better migration performance on small-scale target datasets. In particular, the fine-tuning approach performed better compared to the feature extraction approach, and data augmentation also contributed to the migration learning process. Al-Rubaye et al.[97] investigated the application of deep transfer learning methods in food image classification, extending the Food-101 dataset through data augmentation techniques, and achieving a best result of 96.13% using the EfficientNetB1 classifier. This outcome not only validates the effectiveness of data augmentation but also demonstrates the importance of transfer learning in enhancing model performance. The success of EfficientNetB1 is rooted in its scientific basis of compound scaling, a strategy that simultaneously optimizes the depth, width, and resolution of the network, enabling the model to retain high efficiency while possessing stronger feature extraction capabilities, particularly suitable for handling the diversity and complex textures in food images. On the other hand, Dao et al.[98] employed a transfer learning strategy using PubMedCLIP as a pre-trained model, evaluating the performance across multiple datasets under limited training data conditions. Although this study primarily focused on medical image classification, it also provides valuable insights into the small-sample problem in food image recognition. PubMedCLIP leverages pre-training on large-scale text-image pairs,

utilizing a cross-modal embedding space to extract semantic information from images. Its network architecture features robust semantic understanding and fine-grained feature extraction capabilities, which effectively enhance classification performance under small-sample conditions. Together, these studies highlight the significant effectiveness of transfer learning and data augmentation strategies in improving food image recognition and classification tasks. Table 4 lists recent relevant studies.

Table 4 Food image processing with advanced features acquired by pre-trained models

Author and year	Pre-model	Work	Evaluation index	Database	Summary
Kagaya et al. 2014[93]	CNN	Detection, Recognition	Accuracy	FoodLog	Optimizing CNN hyperparameters and compared to traditional methods, CNN significantly outperforms SVM in food image recognition, with color features playing a key role.
Yanai et al. 2015[99]	DCNN	Classification	Top-1 Accuracy	ImageNet, UECFOOD-100, and UECFOOD-256	Pre-training DCNN on ImageNet and fine-tuning on UEC-FOOD datasets significantly improved food image recognition accuracy and Twitter photo mining efficiency.
Azizpour et al. 2015[92]	ConvNet	Recognition	mAP, Accuracy, Recall	ImageNet	A study on ConvNet representation transferability in visual tasks identifies factors and optimizations, leading to up to 50% error reduction across 16 tasks.
Zhu et al. 2018[100]	AlexNet	Classification	Accuracy	ImageNet	AlexNet achieves a 92.1% accuracy in classifying five vegetables using data enhancement, outperforming BP networks and SVMs, highlighting the importance of dataset size.
Pehlic et al. 2019[101]	AlexNet, GooleNet, Vgg16	Recognition	Accuracy	Pehlic	Modifying the Fully Connected Layer for fine-tuning, AlexNet, GoogleNet, and VGG16 achieved 100% accuracy on a small, author-collected dataset.
Sun et al. 2019[102]	CNN	Classification, Detection	mAP, IoU	Food-5K, UECFood100, UECFood256	This deep learning approach transfers knowledge from a food/non-food classifier to food detection, significantly enhancing object detection accuracy across CNN networks, achieving notable mAP improvements.
Wu et al. 2022[94]	CLIP	Sample Less Identification	Accuracy	VIREO Food-172, UECFood-256	Fine-tuning CLIP with an adapter improves food recognition, with top accuracies of 72.18% on VIREO Food-172 and 68.64% on UECFood-256 datasets at 30 shots.
Cai et al. 2023[96]	DCNN	Classification	Top-1 Accuracy	Food-5K, Food-11, Food-101- Sub, FoodX-251- Sub	Transfer learning enables effective training of large DCNN models with small food datasets, showing that fine-tuning outperforms direct feature extraction but requires more computational resources and training time; data augmentation significantly enhances performance on target tasks.
Gadhiya et al. 2023[103]	GoogleNet, VGG16, ResNet50, MobileNet	Classification	Precision, Recall, F1 Score, FLOPs	Gujarat	ResNet-50 excels in Gujarat image classification with 97.44% training and 88.89% testing accuracy, showing strong results in key metrics.
Xiong et al. 2023[95]	ResNet	Recognition	Accuracy, Precision, Recall, F1 score	Food-101	Training ResNet models on preprocessed data showed over 90% accuracy in food recognition, with significant improvement and specificity across categories, highlighting effectiveness for automated identification.

3.2.2 Deep neural networks redesigned to solve specific food image processing tasks

Deep neural networks redesigned specifically for solving specific tasks show a diverse trend. These methods or frameworks are not only designed and optimized for the characteristics of food image processing, but also incorporate contextual information, external knowledge, and even try to fuse handmade features with deep features to obtain more accurate and diverse processing results. In the following, this work will introduce the core ideas and implementation strategies of these research directions in detail Table 5.

1) Deep learning method for obtaining deep features based on local or specific attributes of food.

In food image processing, deep learning techniques excel at automatically extracting deep features that are intricately tied to food-specific attributes such as structure, color, and composition. These attributes are crucial for accurately capturing the complex diversity inherent in food images. For instance, when analyzing food items with distinct vertical structures, like burgers and sandwiches[80] (Figure 10), deep learning models are able to effectively capture this structural information through specialized analysis modules or network branches, leading to enhanced model performance.

Wiatowski et al.[104] introduced a theoretical framework that provides insight into how deep convolutional neural networks (CNNs) can efficiently extract features from hierarchically structured images. Their theory, which is grounded in multiscale analysis and compressed sensing, demonstrates how CNNs use convolutional operations and nonlinear activation functions to iteratively extract and aggregate information from an image. By doing so, deep CNNs can capture local features at lower levels and progressively integrate these features at higher levels, culminating in a comprehensive understanding of complex image patterns, especially those with hierarchical structures. This theory offers a robust mathematical foundation for understanding the effectiveness of deep CNNs in processing complex food images. Building on this foundation, Metwalli et al.[105] developed the DenseFood model, a densely connected convolutional neural network specifically designed for food image recognition tasks. The model is structured to extract spatial features of food images, incorporating an initial layer, dense block layers, transition layers, and fully connected layers. By utilizing batch normalization, the ELU activation function, and 3×3 convolutional layers, the DenseFood model enhances feature transfer and extraction. Additionally, during training, the model employs a combination of softmax loss and central loss to minimize intra-class variation and maximize inter-class differentiation. This approach has been proven effective, as demonstrated by the model's 81.23% accuracy on the VIREO-172 dataset, outperforming other models like DenseNet121 and ResNet50. Similarly, Panitchakorn et al.[106] leveraged transfer learning and pre-trained models, including VGG16 and InceptionV3, to construct a deep CNN model tailored for the binary classification of artificial snowflake beef images. These examples underscore how deep learning models are being crafted with a keen focus on the unique attributes of food, thereby achieving higher accuracy and performance in food image recognition tasks.

Table 5 Food image processing combining contextual or external knowledge for advanced feature acquisition

Author and year	Work	Method	Result
Xu et al. 2015[40]	Recognition	Geolocalized models for improved dish recognition in restaurants, leveraging geolocation and external restaurant information, introducing two effective and scalable strategies.	The study curated a restaurant-focused food dataset, showing geolocation use improves recognition by 30%, with geolocalized models adding 3-8% gains and fivefold faster training.
Herranz et al. 2017[110]	Recognition	A probabilistic framework integrating visual, location, and external restaurant knowledge for food recognition in restaurants, improving performance across tasks.	Combining visual cues, geographic context, and external restaurant information via a probabilistic model significantly enhances food recognition accuracy in restaurant settings.
Chen et al. 2018[107]	Proposing a generalized distillation framework	SDNet introduces adversarial distillation for efficient recommendations, integrating external knowledge via teacher-student models, enhancing prediction while maintaining test efficiency.	SDNet significantly boosts prediction accuracy and efficiency across datasets, outperforming baselines with marked reductions in prediction time and notable accuracy improvements.
Tian et al. 2020[111]	Classification and recognition capabilities for 2D images	The method introduces recurrent neural networks into CNNs, constructing ShortCut3-ResNet for parallel deep feature learning, with dual optimization improving accuracy.	The proposed CNN algorithm, featuring recurrent neural networks and ShortCut3-ResNet, enhances image recognition accuracy and efficiency on the CIFAR-10 dataset.
Jiang et al. 2020[112]	Recognition	The Multi-Scale Multi-View Feature Aggregation (MSMVFA) scheme for food recognition aggregates high-level semantic, mid-level attribute, and deep visual features, enhancing recognition accuracy through multi-scale CNN activations fusion.	The MSMVFA approach achieved state-of-the-art food recognition performance on three large-scale benchmark datasets, demonstrating superior top-1 accuracy and effectively capturing food image semantics through its multi-scale, multi-view feature aggregation methodology.
Song et al. 2020[113]	Recognition	A Hybrid Attention-Based Prototypical Network enhances few-shot restaurant food image recognition using ResNet-50 and attention mechanisms for instance and feature focus.	Experiments on a large restaurant database confirm the proposed model' s superiority in accuracy and hit@K over state-of-the-art methods, highlighting the effectiveness of hybrid attention in few-shot food image recognition.
Metwalli et al. 2020[105]	Recognition	The DenseFood model, based on DenseNet architecture, combines softmax and center loss for training, aiming to minimize within-class variation and maximize between-class variation, outperforming other models.	DenseFood achieved 81.23% accuracy on VIREO-172 dataset, surpassing others and showing significant performance gains through fine-tuning, proving its effectiveness in food image recognition.
Lohala et al. 2021[114]	Classification and recognition capabilities for 2D images	This method employs a deep CNN with a modified loss function for fast-food image classification, enhancing accuracy and reducing processing time.	The proposed deep learning approach significantly improves fast-food image classification accuracy by 5% on average and reduces processing time by 40-50 seconds.
Wang et al. 2021[115]	Maturity and composition prediction, different taxonomic themes and plant disease detection	Applying deep learning techniques for hyperspectral image analysis.	Deep learning techniques can more effectively utilize spatial and spectral information in hyperspectral image analysis.
Min et al. 2023[19]	Food2K, PRENet	PRENet, for large-scale food recognition, integrates progressive local feature learning with region feature enhancement, employing self-attention to enrich feature representations.	PRENet boosts Food2K accuracy by 2.24% and 1.75% compared to ResNet50 and ResNet101. Pretrained models generalize well across food-related tasks.
Zhu et al. 2023[71]	Recognition	PSD enhances food recognition by progressively boosting detail discovery through self-distillation, using shared embedding networks and learning from modified images to find informative regions.	PSD demonstrated effectiveness across three datasets, notably improving Top-1 and Top-5 accuracies on ETHZ Food-101 and ISIA Food-500 with the Swin Transformer architecture.
He et al. 2023[116]	Classification	This study proposes an improved CNN architecture for food image classification, focusing on optimizing performance through various optimizers, loss functions, and activation functions.	MobileNet V2 model, optimized with data augmentation techniques, delivers superior accuracy in food image classification, showcasing the efficacy of CNNs in this domain.
Zhou et al. 2024[117]	Zero Sample Food Detection	ZSFDet leverages Multi-Source Graph Fusion and Region Feature Diffusion Model to synthesize enhanced features for effective detection of unseen food categories during training.	ZSFDet utilizes food domain knowledge and graph-based fusion for distinctive feature generation, achieving a 6.1% ZSD mAP and significant GZSD performance improvements on the FOWA dataset.

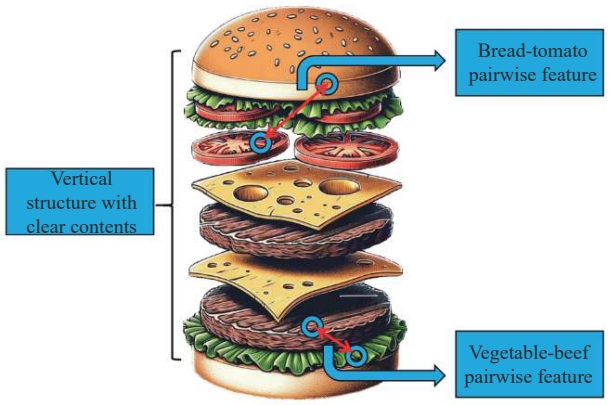
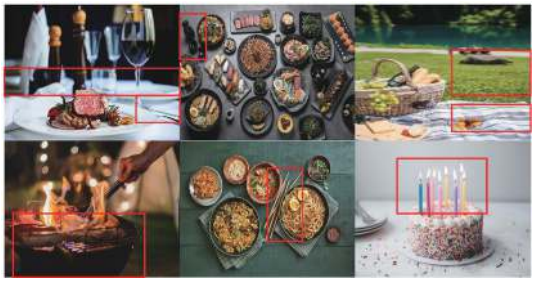


Figure 10 Exploiting spatial relationships between ingredients using pairwise feature statistics improves food recognition accuracy[80]

2) Contextual and external knowledge fusion is considered for network design.

Food image processing is particularly challenging due to its highly complex scene information and dependence on external knowledge. Thus food images do not only involve the visual features of the food itself, but also contain a large amount of contextual information, such as tableware, background environment. Meanwhile, the types and attributes of the food are often closely related to a wide range of external knowledge, such as food ingredient databases, nutritional information, and food recipes. Therefore, effectively incorporating this contextual information and external knowledge into deep learning models is crucial for improving the accuracy and depth of understanding of food image recognition. Contextual information plays an indispensable role in improving the recognition ability of the model. In food images, in addition to the food itself, incidental contextual elements, such as other foods, drinks, and tableware on the table, can provide important clues about the nature of the subject food. For instance, an image featuring knives, forks, and red wine is more likely to be associated with Western cuisine rather than Chinese cuisine, with the presence of chopsticks indicating the opposite. Specific scenes such as picnic blankets or kitchen countertops can also provide critical contextual clues for identifying food categories (Figure 11). This information can assist in model predictions. Incorporating

external knowledge into the model can greatly extend the knowledge boundary and depth of understanding of the model [107-109]. By accessing external food composition databases and nutritional information, the model can not only learn the visual characteristics of different food products but also their composition and nutritional value, which is important for performing complex food recognition tasks, such as distinguishing between food products that are similar in appearance but have different compositions, e.g., different kinds of nuts or grains. In addition, this external knowledge can help the model provide more comprehensive information about the food, not only by identifying the food category but also by analyzing and evaluating its nutritional composition.



Note: Six pairs of diagrams frame contextual information outside of food in the following order: fork, white tablecloth; Japanese-specific sake pot; picnic cloth, grass; barbecue grill; chopsticks; candles.

Figure 11 Contextual information that can aid in model prediction

To achieve this goal, techniques such as multimodal learning, attention mechanisms, and knowledge graphs are widely used in web design. Multimodal learning allows the model to simultaneously process data from different sources, such as image features and textual information from external databases, so that the model can have a more comprehensive understanding of the food and its related attributes. The attention mechanism helps the model to focus on the most critical parts of the image as well as the most important external knowledge points, thus improving the accuracy and efficiency of the recognition. On the other hand, the knowledge graph provides a rich external knowledge framework for the model, which introduces a priori knowledge to help the model's judgment and work, such as introducing geographic location information and other information to obtain the food preferences of the residents of the place, which can help the model's recognition and judgment. For example, Xu et al.[40] proposed an innovative framework for implementing location-based dish recognition in a restaurant environment. The core idea of the framework is to simplify the classification problem by using geo-location and external information about the restaurant. Herranz et al.[110] also proposed an approach using contextual information such as restaurant menus and geo-location, which significantly improves the accuracy of food recognition. The system they developed cleverly fuses visual features, geolocation, and external knowledge to construct a probabilistic model to associate dishes, restaurants, and locations. Experimental results on multiple datasets show that this multi-evidence integration approach not only enhances the accuracy of dish recognition, but also improves the efficacy of restaurant recognition and location refinement. Specifically, the system first reduces the complexity of the problem by filtering out unlikely dish categories that are geographically distant from the test image. The problem was then reconstructed by the researchers to connect dishes, restaurants, and locations using a probabilistic model, effectively exploiting geographic context and menu information to optimize recognition performance. Song et al.[113] proposed an innovative prototype network with a hybrid attentional mechanism specifically for the task of recognizing a small number of shots of unfamiliar restaurant food images. The network was evaluated on different deep learning architectures such as ResNet-50, AlexNet, VGG-16, and GoogLeNet, and the model performance was tested on benchmark datasets such as Food101, FoodCam256, and VIREO Food-172. The results also show that the hybrid attention mechanism, which combines both instance-level and feature-level attention, significantly improves the recognition accuracy of the model. Specifically, the instance-level attention mechanism is able to filter out more informative instances from the support set and reduce the influence of noisy instances during the training process, while the feature-level attention mechanism is able to alleviate the feature sparsity problem by highlighting important dimensions in the feature space and tailoring specific distance functions for different relationships. This also makes the model more effective and robust in focusing on important instances and features, and ultimately achieves excellent performance in the field of small number of lenses learning by calculating the distance to each category prototype for classification. Zhou et al.[117] proposed a new framework called Zero-Shot Food Detection (ZSFDet), which utilizes Knowledge Enhanced Feature Synthesizer (KEFS) to solve the fine-grained problem in ZSFDet. The model improves the mAP of their method by 9.6% in the GZSD setting compared to the latest baseline TCB on the PASCAL VOC dataset, in the MS COCO "48/17" and "65/15" divisions by 12.6% and 5.0%, respectively. In the subsequent ablation experiments, it is demonstrated that integrating word vectors and attribute vectors can improve the ZSD performance. The contributions of different graphs and fusion strategies are also verified, in which the application of knowledge graphs achieves a significant performance enhancement on the FOWA dataset. Zhu et al.[71] proposed a method named Progressive Self-Distillation (PSD) for the food recognition task, and improved 2.78% and 2.57% in top-1 accuracy and 0.9% and 1.24% in top-5 accuracy relative to the baseline approach on the ETHZ Food101, Vireo Food-172, and ISIA Food-500 datasets using two architectures, Swin-B and DenseNet161, respectively. Fan et al.[62] introduced a novel multi-model fusion method based on stacking for the first time in food recognition, further enhancing the model's robustness and accuracy. The top-1 accuracy achieved on the CNFOOD-241 dataset reached 82.88%. The Progressive Region Enhancement Network (PRENet)[19] introduced primarily focuses on progressive local feature learning and enhancing region features. PRENet learns complementary multi-scale finer local features through a progressive training strategy and uses a self-attention mechanism to incorporate richer multi-scale context into local features to enhance its representation. In the experiments, PRENet performs well on the Food2K dataset and achieves good top-1 classification accuracy, e.g., 89.91% when using ResNet50 as the backbone network. These latter approaches, which introduce an attention mechanism, allow the network to focus on the parts of the image that are critical for classification and extract more discriminative features from them, which is a fine-grained recognition effort. Chen et al.[118] introduced a novel model for fine-grained food classification, Res-VMamba, which utilizes a method enhanced by deep residual learning on a state space model. Achieving a Top-1 accuracy of 79.54% on the CNFOOD-241 dataset without using pretrained weights, it surpasses existing advanced methods, establishing a new benchmark for cutting-edge performance in the field of food recognition.

3.2.3 Hybrid approach

Furthermore, in addition to acquiring deep features individually, there are works focusing on fusing manual features with deep features extracted by deep learning. This integrated approach combines the advantages of both features, which not only makes full use of the powerful feature learning capability of deep learning, but also retains the sensitivity of manual features in dealing with specific image details. In this way, a richer and more subtle feature representation is provided for fine-grained classification of food images[119]. Studies have explored diverse hybrid approaches by integrating the automatic feature extraction capabilities of deep learning models with the advantages of other traditional machine learning techniques, such as combining the powerful classification capabilities of support vector machines (SVMs) or applying decision tree algorithms to enhance the model's explanatory power, such as new methods for nonlinear data classification based on decision trees and deep neural networks (DNNs) as proposed by Arifuzzaman et al.[120] These innovative approaches show the versatility and flexibility of deep learning, and through continuous innovation and improvement, it is believed that researchers will be able to develop more efficient and accurate deep neural network models for specific food image processing tasks, thus advancing the field of food science and technology.

3.3 Summary

This section discusses feature extraction from food images using deep learning, emphasizing its critical role. Early research focused on manual features like color and texture, while modern methods automate feature extraction. Despite progress, challenges such as food diversity and varying conditions remain. The report also details fine-tuning with pre-trained models and building networks for specific tasks to enhance deep feature extraction from food images. Additionally, it explores integrating manual and deep learning features and using multimodal learning, attention mechanisms, and knowledge graphs to improve accuracy and understanding in food image recognition. These advancements are transforming research and applications in food image processing.

4 Applications and prospects

4.1 Applications

The application of food image processing has emerged as a significant research area, leveraging deep learning to offer unprecedented convenience in daily life and health management. Here are some specific application examples:

1) Self-checkout systems

In many modern supermarkets and convenience stores, self-checkout systems have been implemented using food image recognition technology[121]. Customers simply place their products in the scanning area, and the system automatically recognizes the items and calculates the total amount without the need for barcodes. This technology not only enhances checkout efficiency but also reduces operational costs and improves customer satisfaction by speeding up the checkout process.

2) Dietary health management

Deep learning can analyze food images to identify food types and their nutrients, providing users with dietary recommendations. This is particularly useful for individuals managing their diets, such as diabetics and athletes. Applications using this technology are becoming popular in health and wellness apps, where they offer personalized dietary advice and help users track their food intake more accurately.

3) Intelligent food service

In the food service industry, food image processing technology helps restaurants automate dish identification and manage inventory. It also enables personalized food recommendations based on image analysis, enhancing customer experience and satisfaction. For example, companies like Meituan use image recognition for tailored food suggestions, boosting customer engagement and driving sales.

4) Food safety monitoring

Deep learning aids in food safety by detecting foreign objects or signs of spoilage in food. This technology is being integrated into production lines to ensure quality control and compliance with safety standards, thus protecting consumers and enhancing brand reliability.

5) Other applications

Deep learning innovations include methods like FIRE for generating food titles and recipes from images[122], and smartphone-based food dye quantification[123]. These advancements are opening new commercial opportunities in food technology and analysis, contributing to more efficient and cost-effective solutions in the industry.

4.2 Prospects

With the continuous development of deep learning, the field of food image processing has made remarkable progress, bringing many surprises. However, looking ahead, this field will still face a series of unique and complex challenges. Firstly, the diversity of food types, the variability of cooking methods, and the diversity of backgrounds make image recognition tasks exceptionally complex. Additionally, compared to other types of images, food images have finer-grained characteristics, meaning that models must capture subtle differences, such as slight variations between similar types of food, to achieve accurate classification and recognition. To address this issue, Min et al.[19] proposed the PRENet network, a deep progressive regional enhancement network for food image recognition. This network locates hot spots containing ingredients in complex environments through progressive local feature learning modules and further enhances feature representation through regional feature enhancement modules. Similarly, the PSD method proposed by Zhu et al.[71] also progressively enhances the network's ability to capture food recognition details. These methods provide valuable research insights by transitioning from coarse to fine granularity, which future researchers can refer to and learn from.

At the same time, the demand for lightweight models is continuously increasing, especially in scenarios with limited computing resources such as mobile devices or embedded systems. Developing efficient deep learning models that occupy minimal resources has become particularly important. Currently, several outstanding networks have demonstrated excellent performance in this area, such as MobileNet[124], which uses depthwise separable convolutions, and EfficientNet[125], which adopts a compound scaling method. These challenges require innovation in model structure design and algorithm optimization to reduce computational complexity while ensuring accuracy.

In the field of food image processing, the application of multimodal methods has made significant progress in recent years. These methods combine image data with other data sources (e.g., text, speech, or sensor data) to improve the accuracy of analyzing food information[126,127]. For example, the fusion of image and text descriptions can be utilized to identify food types and characteristics more comprehensively, while the combination of speech input can enhance the contextual understanding of food products. Meanwhile, the fusion of sensor data (e.g., weight and temperature sensors) with

image data provides more comprehensive food analyses and improves the accuracy of nutritional estimation. However, multimodal approaches, while providing more accurate estimates, rely on complex data fusion. High-quality multimodal datasets are difficult to obtain, and difficulties in synchronizing data sources may also affect estimation accuracy. In addition, in the specific direction of food nutrition estimation, although 3D reconstruction techniques[128,129] have been available to significantly improve the accuracy of volumetric estimation, their high cost and technology also limit their application in consumer devices. Future research is also needed to simplify these methods and enhance their feasibility in practical applications.

5 Summary

With the continuous improvement in quality of life, people pay more attention to healthy lifestyle, which is especially obvious in the era of intelligent informationization. The application of deep learning in the field of food image processing demonstrates its strong potential and vigor. This survey first details the key aspects of deep learning applications in the food domain, including the establishment and development of datasets, the formulation of evaluation metrics, and their applications in tasks such as food classification, recognition, detection, segmentation, and image retrieval. In particular, this study mentions commonly used food datasets and their special application scenarios to better understand the research trends and challenges in this area. Next, this survey explores the importance of deep learning for feature extraction in food image processing, focusing on further manipulation and optimization based on pre-trained models and deep neural networks designed for specific food image processing tasks. These methods significantly enhance the recognition of deep features in food images through fine-tuning or specific adaptive improvements, providing strong technical support for accurate classification and recognition of food images. Furthermore, we summarize current application examples in the field of food image processing and provide a prospective discussion on future development directions. Combining innovative means such as large models and AR technology, food image processing technology is expected to have a wide range of applications in providing personalized dietary advice and enhancing the food selection experience. It is expected that with the continuous progress of deep learning, research and applications in the food field can realize more in-depth development, so as to better promote public health and facilitate the popularization and practice of healthy lifestyles.

[References]

[1] Havelaar A H, Kirk M D, Torgerson P R, Gibb H J, Hald T, Lake R J, et al. World health organization global estimates and regional comparisons of the burden of foodborne disease in 2010. PLoS Medicine, 2015; 12(12): e1001923.

[2] Fung F, Wang H S, Menon S. Food safety in the 21st century. Biomedical Journal, 2018; 41(2): 88–95.

[3] Manasa H S, Mahadevaswamy R K. Obesity and its effects on health. Saudi J Nurs Health Care, 2023; 6(1): 16–17.

[4] Honce R, Schultz-Cherry S. A tale of two pandemics: Obesity and Covid-19. Journal of Travel Medicine, 2020; 27(5): taaa097.

[5] Daniels N F, Burrin C, Chan T, Fusco F. A systematic review of the impact of the first year of covid-19 on obesity risk factors: A pandemic fueling a pandemic? Current Developments in Nutrition, 2022; 6(4): nzac011.

[6] Das M. WHO urges immediate action to tackle noncommunicable diseases. The Lancet Oncology, 2022; 23(11): 1361.

[7] Mayen A-L, Aglago E K, Knaze V, Cordova R, Schalkwijk C G, Wagner K H, et al. Dietary intake of advanced glycation endproducts and risk of hepatobiliary cancers: A multinational cohort study. International Journal of Cancer, 2021; 149(4): 854–864.

[8] Ji X-W, Wang J, Shen Q-M, Li Z-Y, Jiang Y-F, Liu D-K, et al. Dietary fat intake and liver cancer incidence: A population-based cohort study in Chinese men. International Journal of Cancer, 2021; 148(12): 2982–2996.

[9] Steck S E, Murphy E A. Dietary patterns and cancer risk. Nature Reviews Cancer, 2020; 20(2): 125–138.

[10] Huang C C. Special issue on deep learning-based neural information processing for big data analytics. Neural Computing and Applications, 2020; 32(6): 1513–1515.

[11] Lin Y D, Ma J, Cheng J-H, Sun D-W. Visible detection of chilled beef freshness using a paper-based colourimetric sensor array combining with deep learning algorithms. Food Chemistry, 2024; 441: 138344.

[12] Panesar A. Artificial intelligence and machine learning in precision health. In: Precision Health and Artificial Intelligence. Panesar A (Ed.). Berkeley: Apress. 2023; pp.67–85. doi: 10.1007/978-1-4842-9162-7_4.

[13] Graves M, Smith A, Batchelor B. Approaches to foreign body detection in foods. Trends in Food Science and Technology, 1998; 9: 21–27.

[14] Kumar G, Bhatia P K. A detailed review of feature extraction in image processing systems. In: IEEE Fourth International Conference on Advanced Computing & Communication Technologies, Rohtak, Haryana, India: IEEE, 2014; pp.5–12.

[15] Davies E R. Image processing for the food industry. World Scientific Press. 2000; 312p. doi: 10.1142/4182.

[16] Ming Z-Y, Chen J J, Cao Y, Forde C, Ngo C-W, Chua T S. Food photo recognition for dietary tracking: System and experiment. In: 24th International Conference on Multi Media Modelling (MMM), Bangkok, Thailand: Springer, 2018; pp.129–141.

[17] Kong F Y, Tan J D. Dietcam: Automatic dietary assessment with mobile camera phones. Pervasive and Mobile Computing, 2012; 8(1): 147–163.

[18] Meyers A, Johnston N, Rathod V, Korattikara A, Gorban A, Silberman N, et al. Im2calories: Towards an automated mobile vision food diary. In: 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile: IEEE, 2015; pp.1233–1241.

[19] Jacobs Jr C G. Challenges to the quality of data quality measures. Food Chemistry, 2009; 113(3): 754–758.

[20] Min W Q, Wang Z L, Liu Y X, Luo M J, Kang L P, Wei X M, et al. Large scale visual food recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023; 45(8): 9932–9949.

[21] Chen M, Dhingra K, Wu W, Yang L, Sukthankar R, Yang J. Pfid: Pittsburgh fast-food image dataset. In: 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt: IEEE, 2009; pp.289–292.

[22] Hoashi H, Joutou T, Yanai K. Image recognition of 85 food categories by feature fusion. In: 2010 IEEE International Symposium on Multimedia, Taichung, Taiwan: IEEE, 2010; pp.296–301.

[23] Matsuda Y, Yanai K. Multiple-food recognition considering co-occurrence employing manifold ranking. In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), Tsukuba, Japan: IEEE, 2012; pp.2017–2020.

[24] Kawano Y, Yanai K. Automatic expansion of a food image dataset leveraging existing categories with domain adaptation. In: Computer Vision-ECCV 2014 Workshops, Zurich, Switzerland: Springer, 2015; pp.3–17.

[25] Bossard L, Guillaumin M, Van Gool L. Food101 - mining discriminative components with random forests. In: Computer Vision–ECCV 2014. Zurich, Switzerland: Springer, 2014; pp.446–461.

[26] Chen X, Zhu Y, Zhou H, Diao L, Wang D. Chinese foodnet: A large-scale image dataset for Chinese food recognition. arXiv preprint arXiv: 1705.02743, 2017; In press. doi: 10.48550/arXiv.1705.02743.

[27] Kaur P, Sikka K, Wang W, Belongie S, Divakaran A. Foodx-251: A dataset for fine-grained food classification. arXiv preprint arXiv: 1907.06167, 2019; doi: 10.48550/arXiv.1907.06167.

[28] Sahoo D, Hao W, Ke S, Wu X W, Le H. Foodai: Food image recognition via deep learning for smart food logging. In: ACM SIGKDD Conference on Knowledge Discovery and Data Mining, New York, NY, USA: Association for Computing Machinery, 2019; pp.2260–2268.

[29] Deng J, Dong W, Socher R, Li L-J, Li K, Li F F. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA: IEEE, 2009; pp.248–255.

[30] Liu C, Guo C, Dakota D, Rajagopalan S, Li W, Kübler S, et al. My curiosity was satisfied, but not in a good way": Predicting user ratings for online recipes. In: Proceedings of the Second Workshop on Natural Language Processing for Social Media (SocialNLP), Dublin, Ireland: Association for Computational Linguistics and Dublin City University, 2014; pp.12–21.

[31] Marín J, Biswas A, Ofli F, Hynes N, Salvador A, Aytar Y, et al. Recipe1m+: A dataset for learning cross-modal embeddings for cooking recipes and food images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021; 43(1): 187–203.

[32] Xia X J, Liu W, Wang L A, Sun J. Hsifoodingr-64: A dataset for hyperspectral food-related studies and a benchmark method on food ingredient retrieval. IEEE Access, 2023; 11: 13152–13162.

[33] Joutou T, Yanai K. A food image recognition system with multiple kernel learning. In: 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt: IEEE, 2009; pp.285–288.

[34] Miyazaki T, de Silva G C, Aizawa K. Imagebased calorie content estimation for dietary assessment. In: 2011 IEEE International Symposium on Multimedia, Dana Point, CA, USA: IEEE, 2011; pp.363–368.

[35] Stutz T, Dinic R, Domhardt M, Ginzinger S. Can mobile augmented reality systems assist in portion estimation? A user study. In: 2014 IEEE international symposium on mixed and augmented reality-media, art, social science, humanities and design (ISMAR-MASH'D), Munich, Germany: IEEE, 2014; pp.51–57.

[36] Farinella G M, Allegra D. Stanco F. A benchmark dataset to study the representation of food images. In: Computer Vision - ECCV 2014 Workshops, Springer, Cham, 2015; pp.584–599. doi: 10.1007/978-3-319-16199-0_41.

[37] Ciocca G, Napoletano P, Schettini R. Food recognition and leftover estimation for daily diet monitoring. In: New Trends in Image Analysis and Processing - ICIAP 2015 Workshops: ICIAP 2015, Genoa, Italy, Springer, 2015; pp.334–341.

[38] Wang X, Kumar D, Thome N, Cord M, Precioso F. Recipe recognition with large multimodal food dataset. In: 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin: IEEE, 2015; pp.1–6.

[39] Fang S, Liu C, Zhu F, Delp E J, Boushey C J. Single-view food portion estimation based on geometric models. In: 2015 IEEE International Symposium on Multimedia (ISM), Miami, FL, USA: IEEE, 2015; pp.385–390.

[40] Xu R J, Herranz L, Jiang S Q, Wang S, Song X H, Jain R. Geolocalized modeling for dish recognition. IEEE Transactions on Multimedia, 2015; 17(8): 1187–1199.

[41] Beijbom O, Joshi N, Morris D, Saponas S, Khullar S. Menu-match: Restaurant-specific food logging from images. In: 2015 IEEE Winter Conference on Applications of Computer Vision, Waikoloa: IEEE, 2015; pp.844–851.

[42] Zhou F, Lin Y Q. Fine-grained image classification by exploring bipartite-graph labels. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas: IEEE, 2016; pp.1124–1133.

[43] Chen J, Ngo C-W. Deep-based ingredient recognition for cooking recipe retrieval. In: Proceedings of the 24th ACM international conference on Multimedia, New York: Association for Computing Machinery, 2016; pp.32–41.

[44] Ciocca G, Napoletano P, Schettini R. Food recognition: A new dataset, experiments, and results. IEEE Journal of Biomedical and Health Informatics, 2016; 21(3): 588–598.

[45] Merler M, Wu H, Uceda-Sosa R, Nguyen Q-B, Smith J R. Snap, eat, repeat: A food recognition engine for dietary logging. In: Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management, New York, NY, USA: Association for Computing Machinery, 2016; pp.31–40.

[46] Singla A, Yuan L, Ebrahimi T. Food/non-food image classification and food categorization using pretrained googlenet model. In: Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management, New York, NY, USA: Association for Computing Machinery, 2016; pp.3–11.

[47] Farinella G M, Allegra D, Moltisanti M, Stanco F, Battiato S. Retrieval and classification of food images. Computers in Biology and Medicine, 2016; 77: 23–39.

[48] Ciocca G, Napoletano P, Schettini R. CNN-based features for retrieval and classification of food images. Computer Vision and Image Understanding, 2018, 176–177: 70–77.

[49] Rich J, Haddadi H, Hospedales T M. Towards bottom-up analysis of social food. In: Proceedings of the 6th International Conference on Digital Health Conference, New York, NY, USA: Association for Computing Machinery, 2016; pp.111–120.

[50] Liang Y C, Li J H. Computer vision-based food calorie estimation: Dataset, method, and experiment. arXiv preprint arXiv: 1705.07632, 2017; In press. doi: 10.48550/arXiv.1705.07632.

[51] Güngör C, Baltacı F, Erdem A, Erdem E. Turkish cuisine: A benchmark dataset with turkish meals for food recognition. In: 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, Turkey: IEEE, 2017; pp.1–4.

[52] Pandey P, Deepthi A, Mandal B, Puhan N B. Foodnet: Recognizing foods using ensemble of deep networks. IEEE Signal Processing Letters, 2017; 24(12): 1758–1762.

[53] Termritthikun C, Muneesawang P, Kanprachar S. "Nu-Innet: Thai food image recognition using convolutional neural networks on smartphone. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 2017; 9(2-6): 63–67.

[54] Hou S, Feng Y, Wang Z. Vegfru: A domainspecific dataset for fine-grained visual categorization. In: 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy: IEEE, 2017; pp.541–549.

[55] Waltner G, Schwarz M, Ladstatter S, Weber A, Luley P, Lindschinger M, et al. Personalized dietary self-management using mobile visionbased assistance. In: New Trends in Image Analysis and Processing–ICIAP 2017: ICIAP International Workshops, Catania, Italy: Springer, 2017; pp.385–393.

[56] Muresan H, Oltean M. Fruit recognition from images using deep learning. Acta Universitatis Sapientiae, Informatica, 2018; 10(1): 26–42.

[57] Min W Q, Liu L H, Wang Z L, Luo Z D, Wei X M, Wei X L, et al. Isia food-500: A dataset for large-scale food recognition via stacked global-local attention network. In: Proceedings of the 28th ACM International Conference on Multimedia. ACM New York, NY, USA: Association for Computing Machinery, 2020; pp.393–401.

[58] Chen J J, Zhu B, Ngo C-W, Chua T-S, Jiang Y-G. A study of multi-task and region-wise deep learning for food ingredient recognition. IEEE Transactions on Image Processing, 2020; 30: 1514–1526.

[59] Okamoto K, Yanai K. Uec-foodpix complete: A large-scale food image segmentation dataset. In: Pattern Recognition. ICPR International Workshops and Challenges, Springer, Cham, 2021, pp.647–659. doi: 10.1007/978-3-030-68821-9_51.

[60] Ma P, Lau C P, Yu N, Li A, Liu P, Wang Q, et al. Image-based nutrient estimation for chinese dishes using deep learning. Food Research International, 2021; 147: 110437.

[61] Ma P H, Lau C P, Yu N, Li A, Sheng J P. Application of deep learning for image-based chinese market food nutrients estimation. Food Chemistry, 2022; 373: 130994.

[62] Fan B K, Li W Q, Dong L, Li J Z, Nie Z D. Automatic chinese food recognition based on a stacking fusion model. In: 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia: IEEE, 2023; pp.1–4.

[63] Karabay A, Bolatov A, Varol H A, Chan M-Y. A central asian food dataset for personalized dietary interventions. Nutrients, 2023; 15(7): 1728.

[64] Rong D, Xie L J, Ying Y B. Computer vision detection of foreign objects in walnuts using deep learning. Computers and Electronics in Agriculture, 2019; 162: 1001–1010.

[65] Memis S, Arslan B, Batur O Z, Sonmez E B. A comparative study of deep learning methods on food classification problem. In: 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), Istanbul, Turkey: IEEE, 2020; pp.1–4.

[66] Ghosh T, Sazonov E. A comparative study of deep learning algorithms for detecting food intake. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, Scotland, United Kingdom: IEEE, 2022; pp.2993–2996.

[67] Yang X F, Ye Y M, Li X T, Lau R Y K, Zhang X F, Huang X H. Hyperspectral image classification with deep learning models. IEEE Transactions on Geoscience and Remote Sensing, 2018; 56(9): 5408–5423.

[68] Siemon M S N, Shihavuddin A S M, Ravn-Haren G. Sequential transfer learning based on hierarchical clustering for improved performance in

deep learning based food segmentation. Scientific Reports, 2021; 11(1): 813.
[69] Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. Segment anything. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023; pp.4015–4026.
[70] Ravi N, Gabeur V, Hu Y T, Hu R H, Ryali C, Ma T Y, et al. Sam 2: Segment anything in images and videos. arXiv: 2408.00714, 2024. doi: 10.48550/arXiv.2408.00714.
[71] Zhu Y H, Liu L H, Tian J. Learn more for food recognition via progressive self-distillation. Proceedings of the AAAI Conference on Artificial Intelligence, 2023; 37: 3879–3887.
[72] Tan S W, Lee C P, Lim K M, Lim J Y. Food detection and recognition with deep learning: A comparative study. In: 2023 11th International Conference on Information and Communication Technology (ICoICT), Melaka, Malaysia: IEEE, 2023; pp.283–288.
[73] Saritha R R, Paul V, Kumar P G. Content based image retrieval using deep learning process. Cluster Computing, 2019; 22: 4187–4200.
[74] Dubey S R. A decade survey of content based image retrieval using deep learning. IEEE Transactions on Circuits and Systems for Video Technology, 2021; 32(5): 2687–2704.
[75] Konstantakopoulos F S, Georga E I, Fotiadis D I. A review of image-based food recognition and volume estimation artificial intelligence systems. IEEE Reviews in Biomedical Engineering, 2023; 17: 136–152.
[76] Wei P C, Wang B. Food image classification and image retrieval based on visual features and machine learning. Multimedia Systems, 2022; 28: 2053–2064.
[77] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002; 24(7): 971–987.
[78] Lowe D G. Distinctive image features from scaleinvariant keypoints. International Journal of Computer Vision, 2004; 60: 91–110.
[79] Jain A K, Farrokhnia F. Unsupervised texture segmentation using gabor filters. Pattern recognition, 1991; 24(12): 1167–1186.
[80] Yang S L, Chen M, Pomerleau D, Sukthankar R. Food recognition using statistics of pairwise local features. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco: IEEE, 2010; pp.2249–2256.
[81] Chen M-Y, Yang Y-H, Ho C-J, Wang S-H, Liu S-M, Chang Y, et al. Automatic Chinese food identification and quantity estimation. SIGGRAPH Asia 2012 Technical Briefs, 2012; 1–4. doi: 10.1145/2407746.240777.
[82] Anthimopoulos M M, Gianola L, Scarnato L, Diem P, Mougiakakou S G. A food recognition system for diabetic patients based on an optimized bag-of-features model. IEEE Journal of Biomedical and Health Informatics, 2014; 18(4): 1261–1271.
[83] Oliveira L, Costa V, Neves G, Oliveira T, Jorge E, Lizarraga M. A mobile, lightweight, poll-based food identification system. Pattern Recognition, 2014; 47(5): 1941–1952.
[84] Kawano Y, Yanai K. Foodcam-256: A largescale real-time mobile food recognitionsystem employing high-dimensional features and compression of classifier weights. In: Proceedings of the 22nd ACM international conference on Multimedia, New York, NY, USA: Association for Computing Machinery, 2014; pp.761–762.
[85] Tammachat N, Pantuwong N. Calories analysis of food intake using image recognition. In: 2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, Indonesia: IEEE, 2014; pp.1–4.
[86] He Y, Xu C, Khanna N, Boushey C J, Delp E J. Analysis of food images: Features and classification. In: 2014 IEEE international conference on image processing (ICIP), Paris, France: IEEE, 2014; pp.2744–2748.
[87] Abdulrahman A, Ozeki T. Food image recognition by using bag of surf: Color features. In: Proceedings of the 3rd International Conference on Human-Agent Interaction, New York: Association for Computing Machinery, 2015; pp.207–208. doi: 10.1145/2814940.2814976.
[88] Ahsani A F, Sari Y A, Adikara P P. Food image retrieval with gray level co-occurrence matrix texture feature and cie l* a* b* color moments feature. In: 2019 International Conference on Sustainable Information Engineering and Technology (SIET), Lombok, Indonesia: IEEE, 2019; pp.130–134.
[89] Mezgec S, Koroušić Seljak B. Nutrinet: A deep learning food and drink image recognition system for dietary assessment. Nutrients, 2017; 9(7): 657.
[90] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2017; 60(6): 84–90.
[91] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2015. doi: 10.48550/arXiv.1409.1556.
[92] Azizpour H, Sharif Razavian A, Sullivan J, Maki A, Carlsson S. From generic to specific deep representations for visual recognition. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, USA: IEEE, 2015; pp.36–45.
[93] Kagaya H, Aizawa K, Ogawa M. Food detection and recognition using convolutional neural network. In: Proceedings of the 22nd ACM international conference on Multimedia, New York, NY, USA: Association for Computing Machinery, 2014; pp.1085–1088.
[94] Wu Y Q, Song X, Chen J J. Few-shot food recognition with pre-trained model. In: Proceedings of the 1st International Workshop on Multimedia for Cooking, Eating, and related APPlications, New York, NY, USA: Association for Computing Machinery, 2022; pp.45–48.
[95] Xiong Y M. Food image recognition based on ResNet. Applied and Computational Engineering, 2023; 8: 20230284.
[96] Cai L Z, Tang L M, Lim S. Transfer learning with deep models for small food datasets. In: International Conference on Automation Control, Algorithm, and Intelligent Bionics (ACAIB 2023), Xiamen, China, 2023. doi: 10.1117/12.2686482.
[97] Al-Rubaye D A, Ayvaz S. Deep transfer learning and data augmentation for food image classification. In: 2022 Iraqi International Conference on Communication and Information Technologies, Basrah, Iraq, 2022; pp.125–130.
[98] Dao H N, Quang T N, Paik I. Transfer learning for medical image classification on multiple datasets using PubMedCLIP. In: 2022 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia), Yeosu, Korea: IEEE, 2022; pp.1–4.
[99] Yanai K, Kawano Y. Food image recognition using deep convolutional network with pre-training and fine-tuning. In: 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, Italy: IEEE, 2015; pp.1–6.
[100] Zhu L, Li Z B, Li C, Wu J, Yue J. High performance vegetable classification from images based on AlexNet deep learning model. Int J Agric & Biol Eng, 2018; 11(4): 217–223.
[101] Pehlic A, Abd Almisreb A, Kunovac M, Skopljak E, Begovic M. Deep transfer learning for food recognition. Southeast Europe Journal of Soft Computing, 2019; 8(2): 182.
[102] Sun J, Radecka K, Zilic Z. Exploring better food detection via transfer learning. In: 2019 16th International Conference on Machine Vision Applications (MVA), Tokyo, Japan: IEEE, 2019; pp.1–6.
[103] Gadhiya J, Khatik A, Kodinariya S, Ramoliya D. Classification of regional food using pre-trained transfer learning models. In: 2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India: IEEE, 2023, pp.1237–1241.
[104] Wiatowski T, Bölcskei H. A mathematical theory of deep convolutional neural networks for feature extraction. IEEE Transactions on Information Theory, 2018; 64: 1845–1866.
[105] Metwalli A-S, Shen W, Wu C Q. Food image recognition based on densely connected convolutional neural networks. In: 2020 international conference on artificial intelligence in information and communication (ICAIIC), Fukuoka, Japan: IEEE, 2020; pp.27–32.
[106] Panitchakorn G, Limpiyakorn Y. Convolutional neural networks for artificial marbling beef classification. In: 2021 10th International Conference on Internet Computing for Science and Engineering, New York, NY, USA: Association for Computing Machinery, 2021; pp.101–104.
[107] Chen X, Zhang Y, Xu H, Qin Z, Zha H. Adversarial distillation for efficient recommendation with external knowledge. Transactions on Information Systems (TOIS), 2018; 37(1): 1–28.
[108] Marra G, Giannini F, Diligenti M, Gori M. Integrating learning and reasoning with deep logic models. In: Brefeld U, Fromont E, Hotho A, Knobbe A, Maathuis M, Robardet C. (eds), Machine Learning and Knowledge Discovery in Databases. Springer, Cham. 2019, pp.517–532. doi: 10.1007/978-3-030-46147-8_31.
[109] Xie X Z, Niu J W, Liu X F, Chen Z S, Tang S J, Yu S. A survey on incorporating domain knowledge into deep learning for medical image

analysis. Medical Image Analysis, 2021; 69: 101985.
[110] Herranz L, Jiang S Q, Xu R H. Modeling restaurant context for food recognition. IEEE Transactions on Multimedia, 2016; 19(2): 430–440.
[111] Tian Y H. Artificial intelligence image recognition method based on convolutional neural network algorithm. IEEE Access, 2020; 8: 125731–125744.
[112] Jiang L D, Qiu B J, Liu X, Huang C X, Lin K H. Deepfood: Food image analysis and dietary assessment via deep model. IEEE Access, 2020; 8: 47477–47489.
[113] Song G, Tao Z, Huang X, Cao G, Liu W, Yang L. Hybrid attention-based prototypical network for unfamiliar restaurant food image few-shot recognition. IEEE Access, 2020; 8: 14893–14900.
[114] Lohala S, Alsadoon A, Prasad P, Ali R S, Altaay A J. A novel deep learning neural network for fast-food image classification and prediction using modified loss function. Multimedia Tools and Applications, 2021; 80(17): 25453–25476.
[115] Wang C Y, Liu B H, Liu L P, Zhu Y J, Hou J L, Liu P, et al. A review of deep learning used in the hyperspectral image analysis for agriculture. Artificial Intelligence Review, 2021; 54(7): 5205–5253.
[116] He Z, Zhang Z, Feng G, Yan Z, Yi L, Yi Z, et al, Dishes recognition system based on deep learning. Academic Journal of Computing & Information Science, 2022; 5(2): 48–53.
[117] Zhou P F, Min W Q, Song J J, Zhang Y, Jiang S Q. Synthesizing knowledge-enhanced features for real-world zero-shot food detection. IEEE Transactions on Image Processing, 2024; 33: 1285–1298.
[118] Chen C-S, Chen G-Y, Zhou D, Jiang D, Chen D-S, Res-vmamba: Fine-grained food category visual classification using selective state space models with deep residual learning. arXiv: 2402.15761, 2024. doi: 10.48550/arXiv.2402.15761.
[119] Song X F, Zou Y, Shi Z, Yang Y F. Image matching and localization based on fusion of handcrafted and deep features. IEEE Sensors Journal, 2023; 23(19): 22967–22983.
[120] Arifuzzaman M, Hasan M R, Toma T J, Hassan S B, Paul A K. An advanced decision tree-based deep neural network in nonlinear data classification. Technologies, 2023; 11(1): 24.
[121] Ciapas B, Treigys P. Automated barcodeless product classifier for food retail self-checkout images. The Visual Computer, 2024; 40: 6245–6259.
[122] Antela K U, Sáez-Hernández R, Morales-Rubio A, Cervera M L, Luque M J. Smartphone-based procedure to determine content of single synthetic dyes in food using the arata-possetto extraction method. Talanta, 2024; 270: 125537.
[123] Saadati M. Smartphone-based digital image analysis for determination of some food dyes in commercial products. Food Analytical Methods, 2021; 14: 2367–2374.
[124] Howard A G, Zhu M L, Chen B, Kalenichenko D, Wang W J, Weyand T et al. Mo-bileNets: Efficient convolutional neural networks for mobile vision applications. arXiv: 1704.04861, 2017. doi: 10.48550/arXiv.1704.04861.
[125] Tan M X, Le Q C. EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv: 1905.11946v5, 2019, doi: 10.48550/arXiv.1905.11946.
[126] Min W Q, Jiang S Q, Sang J T, Wang H Y, Liu X D, Herranz L. Being a supercook: Joint food attributes and multimodal content modeling for recipe retrieval and exploration. IEEE transactions on multimedia, 2016; 19(5): 1100–1113.
[127] Zhao H, Yap K-H, Kot A C. Fusion learning using semantics and graph convolutional network for visual food recognition. In: 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA: IEEE, 2021; pp.1710–1719.
[128] Dehais J, Anthimopoulos M, Shevchik S, Mougiakakou S, et al. Two-view 3D reconstruction for food volume estimation. IEEE transactions on multimedia, 2016; 19(5): 1090–1099.
[129] Naritomi S, Yanai K. Real scale hungry networks: Real scale 3D reconstruction of a dish and a plate using implicit function and a single RGB-D image. In: Proceedings of the 7th International Workshop on Multimedia Assisted Dietary Management, New York, NY, USA: Association for Computing Machinery, 2022; pp.3–10.