# McDonald's food target recognition and calorie display based on YOLOv5 algorithm

**Yuhao Zhang**

Sichuan University, Xihangang Street, Shuangliu District, Chengdu, Sichuan, 610225, China

591256546@qq.com

**Abstract.** This research paper delves into the development and assessment of a novel food recognition and evaluation system tailored for McDonald's menu items, leveraging the capabilities of the YOLOv5 algorithm. The study demonstrates that the system can successfully identify McDonald's food items from images and seamlessly query calorie and nutritional information from a backend database. The data is then presented to the user, aiding in more informed dietary choices and promoting public health awareness. The system has particular utility for McDonald's customers, facilitating real-time decisions that align with individual health goals and nutritional requirements. Our experimental findings show a high degree of accuracy and efficiency, although the system's scope is currently limited to five key menu items. Future directions for this work include expanding the range of recognizable food categories and implementing user feedback mechanisms to refine recognition accuracy. Moreover, the paper discusses potential optimizations for reducing system response time and further enhancing the practical utility of the technology. This research serves as a significant step towards utilizing computer vision technologies for public health interventions, aiming to combat the rise of obesity and related diseases.

**Keywords:** McDonald's, fast food, target recognition, YOLO v5, caloric intake.

## 1. Introduction

The fast-food sector serves as an intriguing artifact of contemporary society, epitomizing the acceleration of modern lifestyles while commanding widespread consumer affection. McDonald's, the preeminent global entity in this industry, initially established in the 1950s by the McDonald brothers and Ray Kroc, has revolutionized the gastronomic experience by implementing innovative service paradigms and employing packaging strategies that facilitate portability [1]. McDonald's, the leading multinational entity in this domain, not only revolutionized the gastronomic paradigm by implementing groundbreaking service strategies and portable packaging technologies but also permeates modern cultural expressions through its location, interior furnishings, food offerings, and packaging [2]. Originally established in the 1950s in the United States by the McDonald brothers and Ray Kroc, the operational model of McDonald's utilizes an efficient, assembly-line-based culinary process that has captivated a global audience and fundamentally altered food consumption patterns [3-4].

Since the early 1970s, the role of fast food in American dietary habits has undergone a significant transformation, marked by an exponential surge in consumption frequency [5]. According to a survey

disseminated by the National Restaurant Association, approximately 30% of consumers regard fast-food establishments as an integral facet of their lifestyle [6]. However, the convenience and palatability associated with fast food come at the expense of elevated caloric content, thereby exacerbating the prevalence of obesity. Data from the Centers for Disease Control and Prevention (CDC) indicate that as of 2009, over one-third of American adults were classified as obese, with a Body Mass Index (BMI) exceeding 30.0 kg/m^2 [7]. Several empirical investigations have substantiated a significant correlation between elevated levels of fast-food consumption and BMI, underscoring the urgent need for comprehensive public health initiatives [8-10].

This present study endeavors to integrate computer vision and nutritional informatics by utilizing the YOLOv5 algorithm for McDonald's food item recognition. We aim to develop a software application that allows users to capture photographs of McDonald's food items, which are then processed by our algorithm to identify the specific food type. Upon identification, the corresponding caloric information is retrieved from a back-end database and displayed to the user. By visualizing caloric metrics, this initiative seeks to promote informed consumer choices, thereby mitigating obesity risk factors.

The reminder of this work is organized as follows. Section II introduces related work. Section III presents the experimental process of McDonald's food recognition based on YOLOv5 algorithm. Section IV shows the program design and conclusion is drawn in Section V.

## 2. Related Work

### 2.1. YOLOv5 algorithm

The You Only Live Once (YOLO) algorithm was first proposed by Joseph Redmon, Santosh Divvala, and others in 2016. Prior to this, the Region based Convolutional Neural Network (R-CNN) series of algorithms achieved high detection accuracy in the field of object detection. However, due to its two-level network structure, its detection speed cannot meet the real-time requirements. In contrast, based on a single level object detection network, the YOLO algorithm has a very fast detection speed, can process 45 frames per second, and is easier to run in real-time, causing wide attention [12]. And, the YOLO series have V1 - V5 five basic versions so far [13]. Its main idea of YOLO algorithm is to uniformly conduct dense sampling at different positions of the image. Different scales and aspect ratios can be used for sampling, and then Convolutional Neural Network (CNN) can be used to extract features and directly perform classification and regression. In practical applications, YOLO algorithm has been widely used in various computer vision applications, such as autonomous driving, monitoring, and robotics.

Combined with the advantages and disadvantages of each version of the YOLO network, this paper selects the newest YOLOv5 network as the basic network for the research of McDonald's food detection algorithm. Its main feature is the adoption of a new lightweight backbone network Cross Stage Partial (CSP) architecture, while also using Bag of Freebies (BOF) and Bag of Specials (BoS) technologies to reduce computational complexity and improve generalization ability. In addition, the YOLOv5 algorithm applies Monte Carlo data augmentation technology to increase sample diversity, thereby improving the model's generalization ability and robustness [13]. Overall, YOLOv5 algorithm is an efficient and accurate object detection algorithm that can be applied to object detection tasks in various scenarios.

### 2.2. Food identification and health management

In recent years, with the increasing public attention to diet related health issues such as obesity, computer-aided food identification technology and health management have received more attention [14]. Also, the widespread popularity of mobile devices and wireless communication networks, such as 5G telecommunications, has better met the needs of individuals to use electronic health-care systems [14-18], which enable them to record their daily life routine, such as diet, sleep quality and exercises. With mobile phones in their hands, people's health information can be analyzed and evaluated through programs, thus achieving data visualization of basic health management.

Since good dietary habits play an important role in human health management, it has been a popular topic for many scholars on food identification and dietary health analysis in recent years. Aizawa et al. propose FoodLog, a web-based system that allows people to record their dietary intake by taking and uploading photos of the food they eat [17]. Concretely, the system attempts to locate and analyze the nutritional components of a diet from photos, and calculates dietary balance, dividing food into five groups: grains, vegetables, meat, beans, milk, and fruits. Despite the broad prospects, one of the main obstacles in applying food recognition and evaluation systems to practice is how to design and develop effective algorithms and systems to obtain food information (such as food type) from images, ensuring accuracy while reducing system response time [18-21]. Keigo Kitamura et al. design a improved system of food detection and health estimation, which additionly allows users to modify errors that may occur during image analysis for a more accurate results [18]. They propose two strategies to improve performance, one is to pre classify before estimating food balance, and the other is to personalize the food image estimator. Researchers find that when using personalized estimators, the accuracy of each user is almost the same, while the accuracy improvement for each food category is completely different, especially for meat and beans. The experimental result shows that the overall accuracy with both techniques is 44%, but the method of using users' own food images for online training has a higher accuracy improvement compared to image pre classification. Chang Liu et al. propose a classification method based on deep learning, and integrated other image analysis algorithms into a real-time computing system based on edge computing to improve recognition accuracy and reduce response time [21]. The result indicates that this real-time system effectively achieves their goals. Given that deep learning algorithms are often time-consuming, researchers still need to further optimize the algorithms and consider designing new deep learning algorithms that can be executed on mobile devices.

## 3. Research methods

This paper aims to research on the food recognition and evaluation systems based on YOLOv5 algorithm, using McDonald's food images as recognition samples and a corresponding calorie datasets which are entered into program's backend database after processing. Users only need to take photos and upload the McDonald's food they consume on our developed program. Finally, the program utilizes front-end and back-end interaction technology to achieve McDonald's food recognition and calorie display, helping users grasp the current calorie intake of food, thereby controlling their diet reasonably, maintaining healthy BMI levels, and reducing the likelihood of obesity. The computer used in the experiment is based on an x64 operating system, Windows 10. Device name is LAPTOP-Q2D9ECLP. And the processor of the computer is Intel (R) Core (TM) i5-10210U CPU @ 1.60GHz 2.11 GHz. The on-board RAM is 8.0GB.

The program architecture diagram is shown in Figure 1. There are primarily four logical layers: Data layer, Program back-end, Algorithm layer and Program front-end. Data layer is the most basic logical layer. It provide the program with the required McDonald's food names and their corresponding information in the form of a dataset. Program front-end layer can transmit images to the Algorithm layer, which need to processes the images and interacts with the back-end through the recognized food names. In the meanwhile, Program back-end layer is used to store the processed dataset, and provides information such as heat and name to the front-end layer of the program through the processing of Algorithm layer. Additionally, all the layers interact with each other through application programming interfaces (API). We will provide a detailed introduction to the four logical layers in the following text.
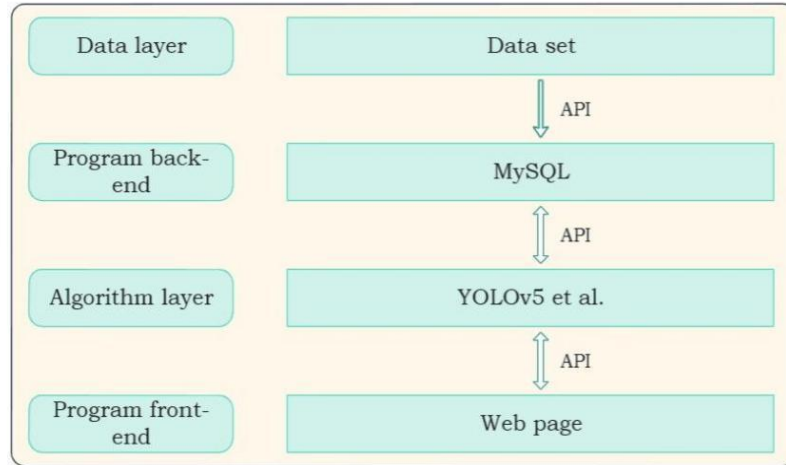
**Figure 1.** McDonald's food identification and calorie analysis program architecture diagram.

### 3.1. Data layer

The Data layer is an integral component that encompasses all the information culled from the McDonald's Nutrition Dataset. This dataset, curated by Priyanshu Sethi and available on the Kaggle platform, serves as a comprehensive source for our food image recognition experiments and the resultant application. The dataset is expansive, including a vast array of McDonald's menu items, each detailed with specific categories and nutritional information. Structurally, the data table is comprised of 14 columns, each elucidating distinct attributes of the menu items. These attributes range from the serial number and name of the menu item to its serving size, either quantified in grams or milliliters for liquid products. Further, the dataset offers granular nutritional details, specifying the caloric content in kilocalories, protein content in grams, and various types of fats in grams, including saturated and trans fats. Additionally, it provides information on cholesterol levels, carbohydrate content, added sugars, and sodium levels in milligrams. The final column categorizes each item under a specific menu category, enabling more nuanced analyses and applications.

### 3.2. Program back-end layer

The Program Back-End layer serves as the nerve center for data management operations, encompassing the addition, deletion, modification, and querying of data. For this research experiment, we leverage Navicat Premium (version 11.0.8) as our database management tool, primarily to manage the MySQL database where the McDonald's food dataset is stored. Notably, Navicat Premium enables a singular program to establish simultaneous connections with up to seven diverse databases, including but not limited to MySQL, MariaDB, SQL Server, SQLite, Oracle, MongoDB, and PostgreSQL. It also facilitates expeditious data transfer between these database systems, supporting specified SQL formats and plaintext files with designated encodings.

MySQL stands out as a premier choice for our Relational Database Management System (RDBMS) due to its prevalent adoption for large-scale data processing tasks. It offers myriad advantages such as high availability, excellent scalability, ease of management, and superlative query performance, making it particularly amenable for web applications.

Prior to utilizing the experimental dataset, it's imperative to undertake data preprocessing tasks. These commonly consist of data cleaning, data integration, data transformation, and data reduction. For the scope of this experiment, we primarily focus on data cleaning activities. The cleaning phase targets missing values, outliers, and duplicate entries within the dataset. Through a combination of techniques like discarding, filling, replacing, and deduplication, we aim to rectify anomalies, correct inaccuracies, and populate gaps within the dataset.

To accomplish this, Python programming language is employed, facilitated through the PyCharm 2021.1.3 x64 Integrated Development Environment (IDE). Upon successful preprocessing, the sanitized dataset is subsequently imported into the MySQL database for downstream utilization in the experiment.

### 3.3. Algorithm layer

The Algorithm Layer serves as the computational epicenter of the system, implementing sophisticated methodologies for problem-solving and experimental goal attainment. In this research, we predominantly incorporate the YOLOv5 algorithm for the task of food type recognition within McDonald's image samples. As delineated in Section II, the YOLO algorithm revolutionizes object detection by framing it as a regression problem rather than a conventional classification task.

In this approach, a singular Convolutional Neural Network (CNN) scans the entire image in one pass, partitioning it into a grid-like structure. For each grid cell, the network forecasts both the class probability and the bounding box attributes. These attributes are encapsulated by four descriptors: the centroid coordinates, dimensions (height and width), and a confidence score mapped to the respective class of the object within the bounding box. Furthermore, the algorithm estimates an object-existence probability for each bounding box.

During the training phase, the objective is to associate each object within the image with a single, optimal bounding box. This is accomplished by maximizing the overlap between predicted boxes and ground-truth boxes, commonly referred to as Intersection over Union (IoU). Upon completion of this phase, the algorithm employs a technique known as "Non-Maximum Suppression" to filter out redundant bounding boxes based on a predetermined confidence threshold, thereby ensuring that the final output is a set of high-probability, non-overlapping boxes.

Schematic diagrams elaborating on these identification steps are visually represented in Figure 2, providing a comprehensive understanding of the algorithm's functional intricacies.
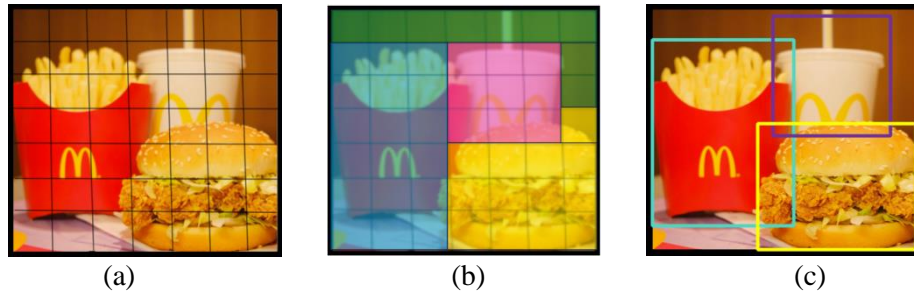


|       (a)       |       (b)       |       (c)       |

**Figure 2.** Images of (a) S*S grid on input (b) Class probability map (c) Final detection.

The experimental pipeline for the McDonald's food recognition system is bifurcated into two pivotal segments: model training and real-time recognition. For the model training phase, the experiment focuses on five quintessential McDonald's offerings: the Big Mac Burger, McSpicy Chicken Filet Burger, Coca Cola, French fries, and Chicken McNuggets. A voluminous corpus of images is collected for each food item, aggregating to a dataset of 300 images, partitioned into 60 images for each item category. Subsequently, these images are annotated using LabelImg, a graphical user interface-based tool that facilitates manual annotation of object classes within the images, pertinent for YOLO's object detection mechanism. Each successful annotation is saved as a text file, encapsulating the coordinates, dimensions, and class indices for the respective bounding boxes. The dataset, now adequately pre-processed, is ingested into the Spyder IDE (Python 3.9) for the training process.

Regarding the model hyperparameters, each training iteration involves two images, culminating in a total of 200 epochs. The YOLOv5 model variant employed is YOLOv5l (Large), characterized by its augmented model architecture and input resolution, which consequently confers enhanced detection accuracy and feature robustness.

Transitioning to the recognition segment, upon completion of the training, the algorithm generates a weight file, aptly named "best.pt," which serves as the computational foundation for object detection in

user-supplied test images. To evaluate the efficacy of the trained model, a quartet of test images are processed through the algorithm. The subsequent results, delineated in Figure 3, confirm the successful recognition of the Big Mac Burger, McSpicy Chicken Filet Burger, Coca Cola, French fries, and Chicken McNuggets, albeit with slight variations in the label nomenclature such as 'McDonald's Big MAC,' 'Burger,' 'Coca-Cola,' 'Chips,' and 'Chicken McNuggets.' Thus, the experiment substantiates the utility and efficacy of the YOLOv5l algorithm in the context of fast-food item recognition, with particular emphasis on McDonald's menu items.



| (a) | (b) | (c) | (d) |

**Figure 3.** The final recognition images of (a) McDonald's Big Mac Burger (b) Mixed food types (c) Chicken McNuggets (d) McSpicy Chicken Filet Burger.

*3.4. Program front-end layer*

The Program Front-End Layer serves as the user interface and interaction gateway, bridging the computational functionalities encapsulated in the back-end with the user experience on the front-end. This layer is underpinned by a sophisticated blend of database architecture, software frameworks, and extensible solutions to deliver an intuitive, responsive, and robust interface.

Specifically, the front-end is tasked with capturing user-input images of McDonald's food items, initiating calls to the back-end for data retrieval and processing, and subsequently displaying the calorie content and nutritional information in a legible format. The architectural blueprint for this front-end layer, as depicted in Figure 4, elucidates the design intricacies and the interaction flow between various front-end components and the back-end services.

This layer makes judicious use of contemporary web development frameworks to construct a modular, scalable, and maintainable user interface. Additionally, it employs asynchronous communication methods and Application Programming Interfaces (APIs) to achieve seamless data exchange with the back-end layers, thereby ensuring real-time updates and interactivity. Consequently, the Program Front-End Layer plays a pivotal role in enhancing user engagement, facilitating ease of use, and ensuring the functional cohesiveness of the McDonald's food recognition and calorie assessment system.
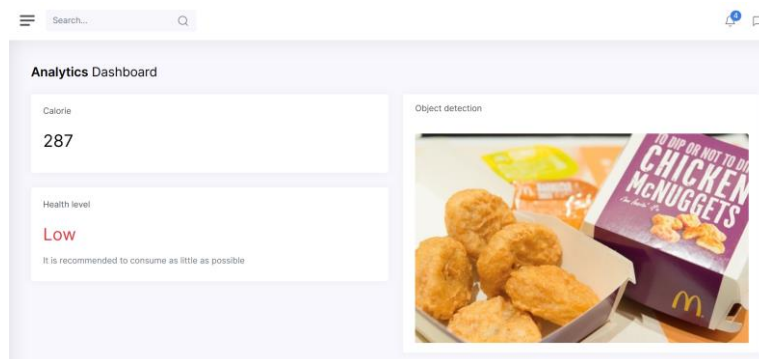


**Figure 4.** The front-end page framework of McDonald's food recognition and calorie assessment system.

The page mainly consists of four content boxes: search box, object detection box, heat output box, and food health assessment box. After logging into the system, users can directly query the ingredient related information of a certain type of food through the search box, or click on the object detection box to upload the captured image for query. Then, the front-end will interact with the back-end based on user actions. If the consumer chooses to input picture for recognition, the system will then transmit the photos to the algorithm layer for object recognition, and the resulting names will be used as query criteria to further transmit to the backend database for retrieval of their relevant component information. The final heat information will be displayed on the front-end layer, which is shown in heat output box. In addition, the system will further evaluate the food safety index through backend data such as calories and sugar content, and the results will be displayed in the food health assessment box in the form of a rating, which can be low, medium, high, or extremely high.

## 4. Discussion

In the realm of automated McDonald's food detection, the performance metrics generated from the Spyder terminal offer valuable insights into the model's computational capabilities. With 267 layers and a whopping 46,189,053 parameters, the model demonstrates significant computational power, as reflected by its 107.9 GFLOPs (Giga Floating-point Operations Per Second). GFLOPs serve as an efficacious measure of hardware performance, denoting the computational bandwidth available for executing floating-point operations per second. The computational capability, gauged through floating-point operations, essentially encapsulates the raw power of the model in terms of arithmetic calculations. Parameters, on the other hand, indicate the model's complexity and expressive ability. Though higher parameter numbers imply stronger feature representation, they also necessitate increased computational resources and data for effective training and inference.

From a temporal perspective, the experimental output showed latencies of 1.1ms for preprocessing, 977.5ms for inference, and 1.9ms for Non-Maximum Suppression (NMS) per image at a resolution of 640x640. The NMS algorithm plays a pivotal role in expunging superfluous bounding boxes, thereby refining the final detection results. Its utilization ensures that only the most accurate and relevant bounding boxes are retained, thus augmenting the precision of the detection process.

Given these performance indicators, the system demonstrates a compelling combination of speed and accuracy, thereby serving as a robust tool for dietary management for McDonald's patrons. It makes nutritional data both transparent and easily accessible, especially catering to individuals who are particularly conscious about their dietary habits, such as fitness enthusiasts or those aiming for weight loss.

However, there exist several avenues for further optimization. The current model's scope is restricted to recognizing only five categories of McDonald's food items, which undeniably limits its applicability given the extensive menu offered by the fast-food chain. Future endeavors should focus on diversifying the types of food items recognized by the system. Additionally, the reliability of the YOLOv5 algorithm, while impressive, is not infallible. Implementing a 'User Feedback' mechanism could serve as an instrumental tool for iterative model refinement, enabling administrators to finetune the model based on real-world usage data. Lastly, system latency remains an issue that needs to be addressed. Techniques such as lazy loading, reducing HTTP request frequencies, and CSS and image optimization could contribute significantly to enhancing the overall user experience by reducing page load time.

## 5. Conclusion

This research paper elucidates the architecture and efficacy of a McDonald's food recognition and evaluation system predicated on the YOLOv5 algorithm. The experimental findings corroborate the algorithm's prowess in identifying diverse McDonald's food items in user-provided images. The system successfully integrates these identifications with a backend database, thereby allowing for real-time query and presentation of calorific and nutritional information to the end-users.

By providing instantaneous and accurate caloric information, this system serves as a seminal tool for dietary management. It enables consumers to make informed decisions regarding their meal choices at

McDonald's, thereby fostering a culture of mindful eating. The ramifications of this are twofold: firstly, it encourages individuals to be vigilant about their caloric intake, thereby aiding in the maintenance of a healthy Body Mass Index (BMI); secondly, by creating a more informed consumer base, it can indirectly contribute to combating the rising epidemic of obesity and associated comorbidities such as diabetes, hypertension, and cardiovascular diseases.

Though the system has demonstrated considerable promise, it remains a work in progress with multiple avenues for enhancement, including expanding the catalog of recognized food items and fine-tuning recognition accuracy based on user feedback. Regardless, the current iteration of the system stands as a viable solution for promoting healthy eating habits, especially within the context of fast-food consumption.

In summation, this research contributes not only to the burgeoning field of computer vision-based object recognition but also offers a pragmatic solution aimed at public health improvement. As the system matures and gains more traction, it has the potential to become an indispensable tool for effective dietary management and, by extension, a bulwark against the proliferation of obesity and related health ailments.

## References

[1] Dong Shuliang. The King of Fast Food - McDonald's [J]. Science Popular (Primary School Edition), 2023 (05): 28-29.

[2] Li Guishu McDonald's [J] Crazy English: Middle School Edition, 2003 (14): 36-36

[3] Deng Zhiping, Huang Zhuomining The Ceremony of Integrating into the City - The Meaning of McDonald's Consumption for Young Migrant Workers [J] Youth Exploration, 2006 (5): 65-67

[4] Vignali C. McDonald's:"think global, act local"–the marketing mix[J]. British food journal, 2001, 103(2): 97-111.

[5] Dave J M, An L C, Jeffery R W, et al. Relationship of attitudes toward fast food and frequency of fast-food intake in adults[J]. Obesity, 2009, 17(6): 1164-1170.

[6] Association NR. Restaurant Industry Pocket Factbook. National Restaurant Association: Washington, DC, 2008.

[7] CDC. Overweight and ob esityhttp://www.cdc.gov/nccdphp/dnpa/obesity/index. Htm (2009) Accessed 21 June 2010

[8] Ginny G P D, Thankam S S P D, Hinojosa P. The Fast Food and Obesity Link: Consumption Patterns and Severity of Obesity[J].Obesity Surgery, 2012, 22(5):p.810-818.

[9] Jeffery, R.W., Baxter, J., McGuire, M. et al. Are fast food restaurants an environmental risk factor for obesity?. Int J Behav Nutr Phys Act 3, 2 (2006).

[10] Flegal KM , Carroll MD , Ogden CL , et al . Prevalence and trends inobesity among US adults , 1999-2008 . JAMA . 2010 : 303 : 235.

[11] M. S T A,T. R M. Modeling dynamics of fast food and obesity for evaluating the peer pressure effect and workout impact[J]. Advances in Difference Equations,2021,2021(1).

[12] Ting L, Baijun Z, Yongsheng Z, et al. Ship detection algorithm based on improved YOLO V5[C]//2021 6th International Conference on Automation, Control and Robotics Engineering (CACRE). IEEE, 2021: 483-487.

[13] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified Real-Time Object Detection", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788, 2016.

[14] Chen, Mei Yun , et al. "Automatic Chinese food identification and quantity estimation." Siggraph Asia Technical Briefs ACM, 2012.

[15] R. Steele, "An overview of the state of the art of automated capture of dietary intake information", Critical Rev. Food Sci. Nutrition, vol. 55, pp. 1929-1938, 2013.

[16] Y. Matsuda and K. Yanai, "Multiple-food recognition considering co-occurrence employing manifold ranking", Proc. 21st Int. Conf. Pattern Recognit., pp. 2017-2020, 2012.

[17] Kitamura, et al. "Food log by analyzing food images." Acm Multimedia (2008):999-1000.

[18] K. Kitamura, C. de Silva, T. Yamasaki and K. Aizawa, "Image processing based approach to food balance analysis for personal food logging," 2010 IEEE International Conference on Multimedia and Expo, Singapore, 2010, pp. 625-630, doi: 10.1109/ICME.2010.5583021.

[19] Y. Matsuda and K. Yanai, "Multiple-food recognition considering co-occurrence employing manifold ranking", Proc. 21st Int. Conf. Pattern Recognit., pp. 2017-2020, 2012.

[20] C. K. Martin, T. Nicklas, B. Gunturk, J. B. Correa, H. R. Allen and C. Champagne, "Measuring food intake with digital photography", J. Hum. Nutr. Diet., vol. 27, no. Suppl 1, pp. 72-81, Jan. 2014.

[21] C. Liu et al., "A New Deep Learning-Based Food Recognition System for Dietary Assessment on An Edge Computing Service Infrastructure," in IEEE Transactions on Services Computing, vol. 11, no. 2, pp. 249-261, 1 March-April 2018, doi: 10.1109/TSC.2017.2662008.