Automated Food Recognition and Volume Estimation for Dietary Assessment and Nutritional Monitoring

Komal Nale[1]
komalnale1@gmail.com
Research scholar[1], Sinhgad College of Engineering, Vadgaon Bk, Pune, Maharashtra

Supriya Rajankar[2]
Research Guide[2], Sinhgad College of Engineering, Vadgaon Bk, Pune, Maharashtra

Vijay Baru[3]
Research Co- guide[3], Sinhgad College of Engineering, Vadgaon Bk, Pune

Madan Mali[4]
Head Of Department[4], Sinhgad College of Engineering, Vadgaon Bk, Pune

ABSTRACT

This research explores the integration of computer vision and artificial intelligence (AI) for food detection and volume estimation to enhance nutritional analysis. As chronic diseases like obesity, diabetes, and cardiovascular conditions rise, accurate monitoring of food intake becomes critical. Traditional methods such as self-reported food diaries have limitations in precision and reliability. By using deep learning models, particularly Convolutional Neural Networks (CNNs) and object detection algorithms like YOLOv11, food items can be detected and segmented from images, followed by volume estimation to determine portion sizes. This approach allows for automated, real-time analysis of food consumption, reducing biases from self-reported data and offering more accurate assessments of nutritional content. The research highlights the potential of combining food recognition, volume estimation, and nutritional databases to calculate calorie and macronutrient information. By addressing challenges such as food shape variation and dataset diversity, these AI-powered systems can be further refined to improve accuracy and generalization across different food types and cultural cuisines. This methodology has the potential to revolutionize dietary tracking, offering real-time, personalized nutrition management, benefiting individuals, healthcare professionals, and large- scale studies. The findings suggest that with continued advancements, food detection and volume estimation systems could play a crucial role in public health strategies for better dietary management.

Keywords: *(Food Detection, Volume Estimation, Nutritional Analysis, Convolutional Neural Networks, YOLOv, Dietary Tracking)*

1. INTRODUCTION

Food detection and volume estimation are essential aspects of nutritional analysis, particularly in an era where there is an increasing focus on health and wellness. As the prevalence of chronic diseases such as obesity, diabetes, and cardiovascular conditions rise, it is crucial to develop accurate and reliable methods for monitoring and analyzing food intake. Traditional techniques, like self-reported food diaries and questionnaires, often suffer from biases, inaccuracies, and subjectivity, limiting their effectiveness in providing precise nutritional data. To address these limitations, the integration of computer vision and artificial intelligence offers a promising solution for food analysis. Specifically, image-based food segmentation and volume estimation have garnered significant attention in recent years. This approach utilizes deep learning algorithms, such as Convolutional Neural Networks (CNNs), to detect and segment food items in images, followed by volume estimation to calculate the portion sizes of food. These advancements enable automated, real-time analysis of food consumption, reducing the reliance on self-reported data and improving the accuracy of nutritional evaluations. By estimating the volume of food items and linking this data to nutritional databases, the system can provide precise information on the calorie and macronutrient content of a meal. This research explores the potential of combining CNNs with object detection models like YOLOv11 to enhance food detection and volume estimation, providing an efficient tool for personalized dietary tracking, healthcare management, and large-scale dietary studies. As this technology evolves, it could revolutionize the way we approach nutrition, offering practical solutions for healthier eating habits and improved public health outcomes.

2. RELATED WORK

The need for accurate dietary assessment has never been greater, as nutrition plays a pivotal role in the prevention and management of various health conditions, such as obesity, diabetes, and cardiovascular diseases. Traditional methods of dietary assessment, such as food diaries and self-reported questionnaires, have inherent limitations including inaccuracies, biases, and time-consuming procedures. Several recent studies have explored the potential of using technology, particularly computer vision and deep learning models, to address these challenges. For example, Fotios S. Konstantakopoulos et al. (2024) [10] highlighted the role of computer vision in real-time nutrient estimation through food images captured by smartphones. They systematically reviewed food image segmentation, classification, and volume estimation methods, discussing the strengths and weaknesses of various techniques. The integration of these methods enables the automatic analysis of food items, which could revolutionize dietary assessment and provide more accurate insights for managing chronic diseases. Similarly, Wei Wang et al. (2022) [5] focused on Vision-Based Dietary Assessment (VBDA), presenting multi-stage architectures involving food image analysis, volume estimation, and nutrient derivation. They highlighted the progress made toward end-to-end dietary assessment models, which simplify the workflow and improve accuracy, paving the way for real-time dietary tracking.

Several researchers have incorporated deep learning techniques, particularly Convolutional Neural Networks (CNNs), to enhance food detection and volume estimation accuracy. Kalliopi V. Dalakleidi et al. (2022) [15] conducted a thorough review of image-based food recognition systems (IBFRS) that combine mobile camera data with deep learning algorithms. They found that CNNs outperformed other methods, especially when dealing with large food image datasets. These systems segment food images, classify food types, and estimate volumes and nutrients, thus improving dietary monitoring. Lameck Mbangula Amugongo et al. (2022) [17] also acknowledged the importance of CNNs in food recognition, noting that mobile-based solutions for food recognition and volume estimation can significantly help individuals manage their diets and health. Their study emphasizes the potential of mobile apps, leveraging CNNs, to make dietary assessment more accessible and user-friendly, especially in healthcare settings. They also pointed out the need for applications to provide explanations for their predictions to improve user trust, which remains a challenge for many mobile-based solutions. Furthermore, Peihua Ma et al. (2022) [16] demonstrated the practical application of deep learning models in dietary assessment by developing a food image database (ChinaMartFood-109) and applying deep convolutional neural networks for food nutrient estimation. Their study, which utilized the InceptionV3 model, showed significant improvements in food classification accuracy, achieving top-1 and top-5 accuracy rates of 78% and 94%, respectively. Such progress in deep learning techniques enhances the accuracy and scalability of food detection systems.

In addition to advancements in food detection and volume estimation, researchers have highlighted the importance of food volume estimation in the accuracy of nutritional analysis. Druwil Jain et al. (2024) [11] introduced the "Food Diet Recaller App - FDRA," which utilizes AI and computer vision to estimate food quantity and nutritional content. This app uses the InceptionV3 model for image classification, achieving high accuracy in food recognition. The app calculates food volume based on linear-shaped items, allowing for accurate estimation of nutritional content. This study emphasizes the growing role of AI-powered apps in dietary management, offering users real-time insights into their food intake. Afnan Ahmed Crystal et al. (2024) [12] explored the role of automated technologies in reducing the burden of diet monitoring, particularly for chronic disease management such as diabetes. They systematically reviewed food image recognition systems (FIRS) and noted the potential of these systems in improving diet tracking, with many tools offering suggestions for better meal planning. This review further emphasizes how mobile vision- based technologies can simplify and automate food intake tracking, eliminating the errors and biases inherent in manual reporting methods. These automated systems are especially valuable in managing diseases that require precise diet control, such as diabetes, where carbohydrate tracking is essential. Finally, Sushant Kaushal et al. (2024) [13] underscored the growing importance of vision-based technologies in improving the accuracy and efficiency of nutritional evaluation. They highlighted the remarkable performance of deep learning models in food classification and nutrient estimation, advocating for the integration of computer vision and AI to revolutionize the future of dietary assessment.

The literature reveals a consistent trend toward developing robust, accurate, and user-friendly food detection and volume estimation systems based on AI and computer vision. Researchers have made significant strides in developing methodologies that can automatically identify and segment food items, estimate their volume, and calculate their nutritional content. However, challenges remain, such as the need for more comprehensive food image datasets, the ability to handle various food shapes and sizes, and the integration of real-time analysis in diverse environments. As highlighted by Jamalia Sultana et al. (2023) [14], the variability in food presentation and occlusions presents obstacles

for image-based food analysis, requiring the development of more robust algorithms that can handle different conditions. The large-scale deployment of these technologies in mobile applications could lead to better personalized nutrition management and assist healthcare professionals in providing more precise dietary advice. Additionally, the development of large food image datasets, such as those discussed by Peihua Ma et al. (2022) [16], and the incorporation of advanced data augmentation techniques will be critical in overcoming the current limitations of food recognition models. Future research should focus on enhancing the scalability and generalizability of food recognition systems, ensuring their accuracy across different cultural food groups, and improving user trust through transparent explanations of predictions. As mobile technology and deep learning continue to evolve, the integration of food detection and volume estimation tools will likely become an indispensable part of everyday health management, offering a more accurate and efficient way to monitor and optimize nutrition.

3. METHODOLOGY

The process for food detection and nutritional estimation shown in figure 1 typically involves following key steps

Fig.1 Food Detection and Nutritional Estimation Process Flow

Input Image The process starts with capturing an image of the food item. This image can be taken using a smartphone, camera, or any imaging device. The quality and clarity of the image are important for accurate recognition and analysis. The captured image serves as the input for the subsequent stages.

Food Recognition: This step involves identifying the type of food present in the image using a Convolutional Neural Network (CNN) model. The image is preprocessed (resized, normalized, etc.) and passed through a trained CNN model designed to classify different food types. The model identifies the food item by analyzing its visual features, enabling the system to recognize whether the item is, for instance, an apple, rice, or pizza. Food recognition is essential, as the type of food will influence the volume, weight, and calorie estimations.

Volume Estimation: Once the food type is identified, the next step is to estimate its volume. This can be achieved through various techniques, such as stereo-based methods, model-based approaches, or depth-camera measurements. For example:

Depth Camera-Based Approach: Depth sensors like Time of Flight (TOF) cameras can capture depth information, allowing for direct volume estimation.

Image-Based Techniques: Methods like perspective transformation or stereo images can approximate the volume based on shape and size.

Calories Count: After estimating the volume and weight of the food, the system can calculate the calorie content. It does this by referencing a nutritional database that provides calorie information per gram for various food items. By multiplying the weight with the calorie value per gram, the total calorie content of the food item is determined. This step completes the nutritional analysis by providing a calorie count, which can then be displayed or logged for the user.

4. RESULTS

Fig.2 Image and Mask Comparison of a Pizza

The figure 2 shows an image of a pizza on the left, with toppings such as pepperoni and olives clearly visible. On the right, the corresponding mask highlights different areas of the pizza, using color-coded regions to represent various components of the pizza, such as the pepperoni slices (marked in pink) and other elements (possibly the crust or olives) in different shades. This mask serves as a simplified version of the original image, potentially used for segmentation or analysis purpose.

```
Downloading data from https://storage.googleapis.com/tensorflow/keras-applications/resnet/resnet50_weights_tf_dim_ordering_tf_kernels.h5
102967424/102967424 ━━━━━━━━━━━━━━━━━━━━ 1s 0us/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 2s 2s/step
Downloading data from https://storage.googleapis.com/download.tensorflow.org/data/imagenet_class_index.json
35363/35363 ━━━━━━━━━━━━━━━━━━━━ 0s 0us/step
Top 3 Predictions:
diamondback: 33.56%
pizza: 25.43%
hognose_snake: 13.86%
```

Fig.3 Top Predictions for an Image Classification Model

The figure 3 displays the output from an image classification model, showing the top three predictions for a given image. The model predicts the highest likelihood for the class "diamondback" at 33.56%, followed by "pizza" at 25.43%, and "hognose_snake" at 13.86%. This suggests that the image may resemble a "diamondback" more closely than the other categories, despite "pizza" being another plausible label based on the model's prediction.

Fig.4 Model Predictions and Mask Image for Pizza

The figure 4 shows a matrix of numerical values, likely representing model predictions or feature activations, followed by a pizza image labeled "New Mask Image for Inspection" at the bottom. The matrix of numbers may correspond to pixel-level confidence values or a segmented mask that the model has generated, while the pizza image serves as the visual reference being analyzed. These outputs could be part of a machine learning model inspecting or classifying parts of the pizza based on the input image.

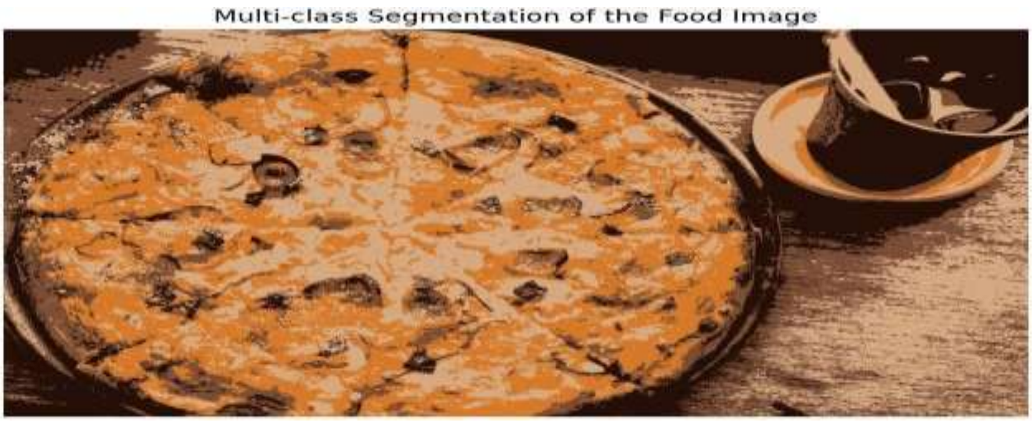

Fig.5 Multi-class Segmentation of the Food Image

The figure 5 shows a segmented image of a pizza with different regions highlighted in a distinctive color, representing the application of multi-class segmentation to the food image. This segmentation method breaks the image into distinct classes or components, making it easier to identify and analyze individual elements like toppings, crust, or other features within the pizza.

Fig.6 Two Different Views of Pizza

The figure 6 image displays two different shots of pizza: one of a whole pizza on a wooden tray with some seasonings and a drink beside it, and the other of a single slice of pepperoni pizza with a beer and seasoning bottles in the background. These images showcase different presentations of pizza, offering a variety of views from a full pizza to a close-up of a slice ready to be enjoyed.

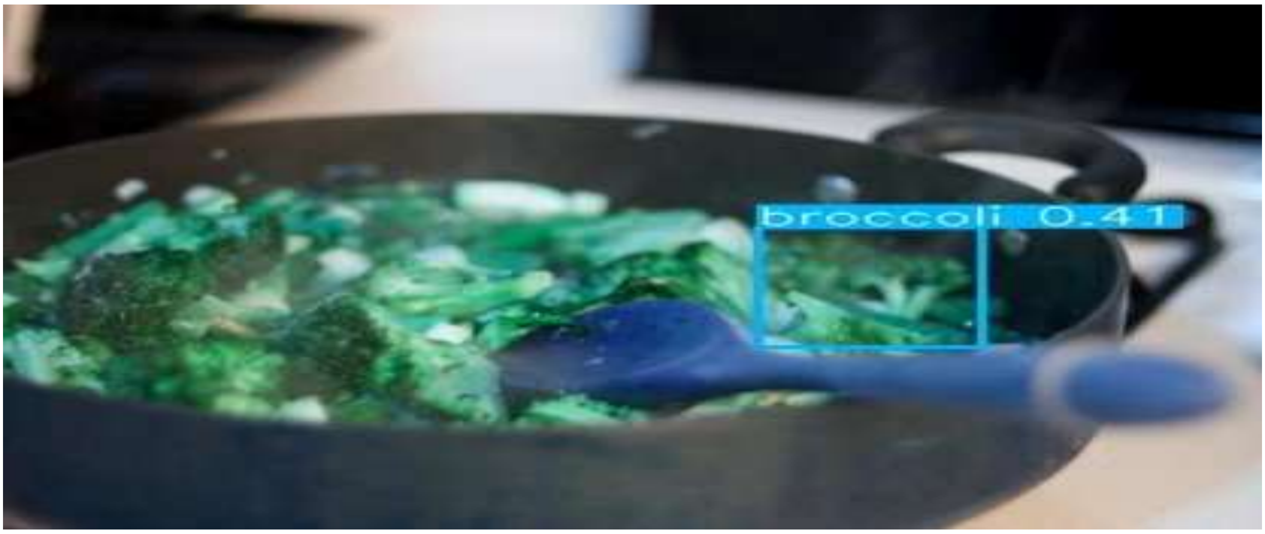

Fig.7 Cooking Broccoli in a Pan

The figure 7 image depicts a close-up of broccoli being sautéed in a pan with steam rising, and a blue spatula stirring the mixture. The highlighted text on the image indicates that the software has recognized broccoli in the scene with a confidence score of 0.41. This suggests that the system is identifying the broccoli as an object in the cooking process, though the recognition confidence is relatively low.

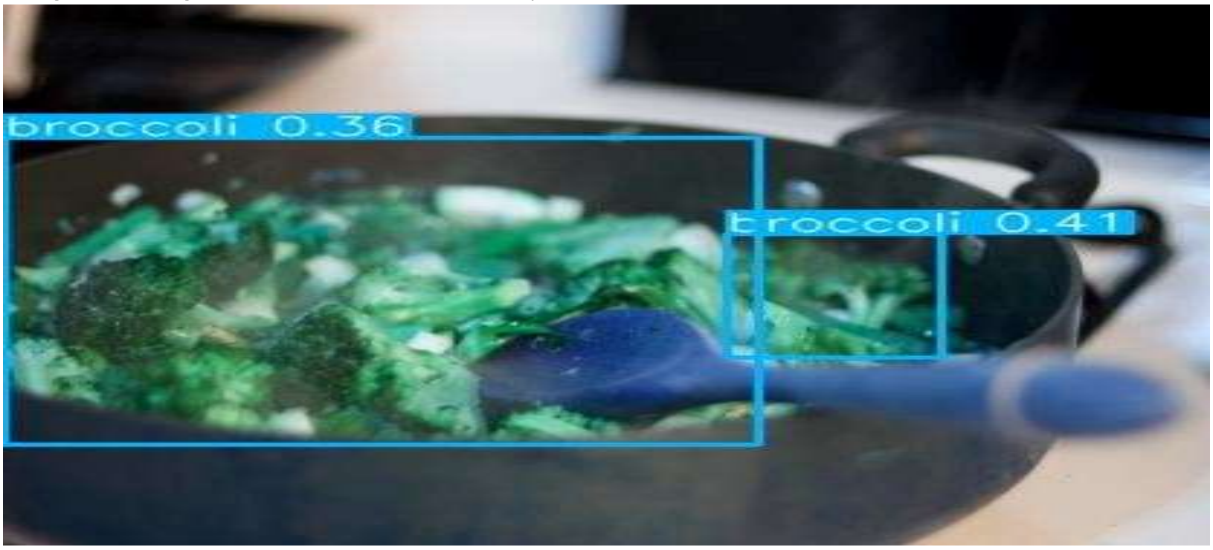

Fig.8 Sautéing Broccoli in a Pan with Multiple Object Recognition

The figure 8 shows a pan with broccoli being cooked, with two separate areas highlighted by the object detection system, indicating the recognition of broccoli with confidence scores of 0.36 and 0.41. This suggests that the system is identifying and labeling the broccoli in the pan, though with relatively low confidence, as seen in the differing confidence levels for each recognition box.

```
plate_diameter_cm = 25

plate_diameter_pixels = 200

scale_factor = plate_diameter_cm / plate_diameter_pixels
print(f"Scale factor: {scale_factor} cm per pixel")

food_area_cm2 = food_area_pixels * (scale_factor ** 2)
print(f"Estimated Area (in cm²): {food_area_cm2}")
```

```
Scale factor: 0.125 cm per pixel
Estimated Area (in cm²): 13141.25
```

```
food_height_cm = 2

food_volume_cm3 = food_area_cm2 * food_height_cm
print(f"Estimated Volume (in cm³): {food_volume_cm3}")
```

```
Estimated Volume (in cm³): 26282.5
```

Fig.9 Estimation of Food Area and Volume from Image Data

The figure 9 image displays Python code used to estimate the area and volume of food based on pixel measurements. The code first calculates the scale factor from the known diameter of a plate, then uses it to estimate the area of the food in square centimeters and its volume in cubic centimeters, with an assumed food height.

```
        nutrition = food_nutrition[food_type.lower()]
        print(f"Nutritional information for {food_type}:")
        print(f"Calories: {nutrition['calories']} kcal")
        print(f"Protein: {nutrition['protein']} g")
        print(f"Fat: {nutrition['fat']} g")
        print(f"Carbs: {nutrition['carbs']} g")
    else:
        print(f"Nutritional information for {food_type} not found.")
```

```
Nutritional information for pizza:
Calories: 285 kcal
Protein: 12 g
Fat: 10 g
Carbs: 36 g
```

Fig.10 Displaying Nutritional Information for Food Type

The figure 10 image shows Python code that retrieves and displays nutritional information based on the food type entered. In this case, the code successfully retrieves and prints the nutritional details for pizza, including calories, protein, fat, and carbs, by referencing a dictionary of food data.

5. CONCLUSION

Food detection and volume estimation for nutritional analysis hold significant promise in enhancing dietary monitoring and personalized health management. By leveraging advanced computer vision techniques and deep learning models, such as Convolutional Neural Networks (CNNs) and object detection algorithms like YOLOv11, it is possible to accurately detect food items and estimate their portion sizes in real-time. This approach eliminates the reliance on self-reported data, which is often prone to inaccuracies and biases, thus offering a more objective and efficient method of assessing dietary intake. The integration of these technologies into mobile applications and wearable devices can empower users to track their food consumption more effectively, enabling better health decisions and disease management, particularly for chronic conditions like diabetes, obesity, and cardiovascular diseases.

Moreover, the methodology combining food recognition, volume estimation, and nutritional content calculation provides a comprehensive tool for assessing food intake. The ability to segment food items and estimate their volume with high accuracy directly impacts the precision of nutritional analysis, such as calorie counting and macronutrient breakdown. Studies reviewed in this paper demonstrate the potential of CNNs to outperform traditional methods, providing significant improvements in food classification and volume estimation accuracy. Furthermore, the integration of depth-sensing technologies and stereo-based methods for volume estimation enhances the overall accuracy of the system, offering more reliable insights into food intake. This technology, in combination with vast food image datasets, contributes to improving the scalability and generalizability of food recognition models across diverse cuisines and food presentations.

In conclusion, the continuous development of food detection and volume estimation systems powered by AI and computer vision is poised to revolutionize the field of nutritional analysis. As these systems evolve, their accuracy and applicability will expand, providing critical tools for individuals, healthcare professionals, and researchers. Future research should focus on overcoming challenges such as variations in food presentation, dataset diversity, and model transparency. With these advancements, automated, real-time food analysis could become an indispensable part of public health strategies, aiding in personalized nutrition management and enabling more accurate dietary assessments across various populations.

REFERENCES

[1] L.-I. Coman, M. Ianculescu, E.-A. Paraschiv, A. Alexandru, and I.-A. Bădărău, "Smart Solutions for Diet-Related Disease Management: Connected Care, Remote Health Monitoring Systems, and Integrated Insights for Advanced Evaluation," *Appl. Sci.*, vol. 14, no. 6, p. 2351, Mar. 2024, doi: 10.3390/app14062351.

[2] F. P. W. Lo, Y. Sun, J. Qiu, and B. Lo, "Image-Based Food Classification and Volume Estimation for Dietary Assessment: A Review," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 7, pp. 1926–1939, Jul. 2020, doi: 10.1109/JBHI.2020.2987943.

[3] T. Brosnan and D.-W. Sun, "Inspection and grading of agricultural and food products by computer vision systems—a review," *Comput. Electron. Agric.*, vol. 36, no. 2–3, pp. 193–213, Nov. 2002, doi: 10.1016/S0168-1699(02)00101-1.

[4] Z. Deng, T. Wang, Y. Zheng, W. Zhang, and Y.-H. Yun, "Deep learning in food authenticity: Recent advances and future trends," *Trends Food Sci. Technol.*, vol. 144, p. 104344, Feb. 2024, doi: 10.1016/j.tifs.2024.104344.

[5] Wei Wang, Weiqing Min , Tianhao Li , Xiaoxiao Dong , Haisheng Li , Shuqiang Jiang b , "A review on vision-based analysis for automatic dietary assessment" Science Direct, March 2022, pp.1-20.

[6] P. Klasnja and W. Pratt, "Healthcare in the pocket: Mapping the space of mobile-phone health interventions," *J. Biomed. Inform.*, vol. 45, no. 1, pp. 184–198, Feb. 2012, doi: 10.1016/j.jbi.2011.08.017.

[7] F. P. W. Lo, Y. Sun, J. Qiu, and B. Lo, "Image-Based Food Classification and Volume Estimation for Dietary Assessment: A Review," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 7, pp. 1926–1939, Jul. 2020, doi: 10.1109/JBHI.2020.2987943.

[8] C. Liu *et al.*, "A New Deep Learning-Based Food Recognition System for Dietary Assessment on An Edge Computing Service Infrastructure," *IEEE Trans. Serv. Comput.*, vol. 11, no. 2, pp. 249–261, Mar. 2018, doi: 10.1109/TSC.2017.2662008.

[9] Z. Zhuang *et al.*, "A Survey of Point Cloud Completion," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 17, pp. 5691–5711, 2024, doi: 10.1109/JSTARS.2024.3362476.

[10] F. S. Konstantakopoulos, E. I. Georga, and D. I. Fotiadis, "A Review of Image-Based Food Recognition and Volume Estimation Artificial Intelligence Systems," *IEEE Rev. Biomed. Eng.*, vol. 17, pp. 136–152, 2024, doi: 10.1109/RBME.2023.3283149.

[11] D. Jain, A. Uphade, S. Ugale, S. Kumbhar, and V. C. Wangikar, "Food Image Recognition and Volume

Estimation: A Comprehensive Study for Dietary Assessment," in *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, Pune, India: IEEE, Apr. 2024, pp. 1–8. doi: 10.1109/I2CT61223.2024.10543812.

[12] A. A. Crystal, M. Valero, V. Nino, and K. H. Ingram, "Empowering Diabetics: Advancements in Smartphone-Based Food Classification, Volume Measurement, and Nutritional Estimation," *Sensors*, vol. 24, no. 13, p. 4089, Jun. 2024, doi: 10.3390/s24134089.

[13] S. Kaushal, D. K. Tammineni, P. Rana, M. Sharma, K. Sridhar, and H.-H. Chen, "Computer vision and deep learning-based approaches for detection of food nutrients/nutrition: New insights and advances," *Trends Food Sci. Technol.*, vol. 146, p. 104408, Apr. 2024, doi: 10.1016/j.tifs.2024.104408.

[14] J. Sultana, B. Md. Ahmed, M. M. Masud, A. K. O. Huq, M. E. Ali, and M. Naznin, "A Study on Food Value Estimation From Images: Taxonomies, Datasets, and Techniques," *IEEE Access*, vol. 11, pp. 45910–45935, 2023, doi: 10.1109/ACCESS.2023.3274475.

[15] K. V. Dalakleidi, M. Papadelli, I. Kapolos, and K. Papadimitriou, "Applying Image-Based Food-Recognition Systems on Dietary Assessment: A Systematic Review," *Adv. Nutr.*, vol. 13, no. 6, pp. 2590–2619, Nov. 2022, doi: 10.1093/advances/nmac078.

[16] Peihua Ma , Chun Pong Lau , Ning Yu , An Li , Jiping Sheng. "Application of deep learning for image-based Chinese market food nutrients estimation" Science Direct, Food Chemistry, Volume 373, Part B, 30 March 2022, 130994

[17] Lameck Mbangula Amugongo, Alexander Kriebitz, Auxane Boch and Christoph Lütge, "Mobile Computer Vision-Based Applications for Food Recognition and Volume and Calorific Estimation: A Systematic Review", 11, 59. https://doi.org/10.3390/healthcare 11010059 , December 2022.