Advances in Computer and Communications, 2023, 4(2), 109-113
https://www.hillpublisher.com/journals/acc/
ISSN Online: 2767-2875

Application of Computer Vision in Food Nutrition Evaluation

Ruijie Li*, **Xuexin Li**

Ocean University of China, Qingdao, Shandong, China.

How to cite this paper: Ruijie Li, Xuexin Li. (2023) Application of Computer Vision in Food Nutrition Evaluation. *Advances in Computer and Communication*, 4(2), 109-113.
DOI: 10.26855/acc.2023.04.009

Received: April 15, 2023
Accepted: May 12, 2023
Published: June 9, 2023

***Corresponding author:** Ruijie Li, Ocean University of China, Qingdao, Shandong, China.

Abstract

In recent years, with the rapid development of artificial intelligence and some significant results proposed, people pay more and more attention to it. Food is related to every aspect of everyone's life. With the gradual improvement of people's living standards, problems such as how to nutritively match their own diet or design weight-loss recipes have attracted more and more attention. In such an environment, the combination of the two reflects a certain significance. This paper analyzes the main advantages and potential defects of artificial intelligence in processing food nutrition assessment, and thinks about its progress space and technical pain points, discusses the technical advantages of artificial intelligence, and analyzes its application in data, computing power, algorithms and other aspects. As well as the possible advantages and industry prospects brought by the integration of the two, we hope to bring a new theoretical basis for the further in-depth discussion of the integration development of the two in academia and industry.

Keywords

Artificial intelligence, convolutional neural networks, food, nutrition evaluation

1. The connection between artificial intelligence and food nutrition assessment

Food is something we need to eat every day, which is rich in nutrients for us is extremely important. On the one hand, all kinds of nutrition provide the most basic material for human activities. Protein, fat, sugar and other components rich in food can provide energy for human body or constitute the basic structure of human body. On the other hand, balanced nutrition can improve the body's health condition, enhance the body's resistance to disease, improve immunity, and make people healthier and healthier.

Compared with the primitive period, people's living conditions are much better than before, people are more and more aware of the importance of food nutrition for their own. However, the number of dietitians who are familiar with various foods is not enough to meet the needs of most people for nutrition collocation. Computer vision, as a hot field in recent years, has the ability to carry out targeted work on specific tasks, and will play a key role in the assessment of food nutrition and further recipe collocation.

2. How does computer vision work in food

2.1 Convolutional neural networks

The research on convolutional neural networks began in the 1980s and 1990s, and time delay network and LeNet-5 [1] are the earliest convolutional neural networks. After the 21st century, with the introduction of the theory of deep learning and the improvement of numerical computing equipment, convolutional neural networks have developed rapidly and been applied to computer vision, natural language processing and other fields.

Most convolutional neural networks are composed of many modules stacked together, among which the most important module is the convolutional layer. The purpose of using the convolutional layer is to identify patterns in input data by learning local features. In the convolutional layer, we use a series of learnable filters that are convolved with the input data to produce what is called a feature graph. In essence, feature map represents the relation between the original input data in spatial location information and specific features. The convolutional layer is designed to allow the network to automatically learn important features in the input data without manually adjusting filter parameters. In addition, in order to ensure the uniform size of the image, so as to facilitate the preprocessing, there are usually many important modules such as the pooling layer and the normalization layer.

For the processing of color images, convolutional neural network usually obtains the RGB information of images, extracts its feature matrix, and then obtains the final result based on its features. Due to the different selection of convolution kernel, these features may contain various features such as spatial features and overall regional features of the image. Although convolutional neural network is a black box model, we do not know every detail of its specific implementation, but it can often achieve amazing results in image processing. Therefore, there have been a lot of worsk using convolutional neural network to process food images in the past.

2.2 Work in recent years

Recent work has focused on two directions. One of the main directions is to build a neural network or agent with good performance to solve the specific application problem. For example, in 2019, Kaylen J. Pfisterer et al. [2] published an article on CVPR. They propose a novel approach to model the association between color and vitamin A content using transmission imaging of a pureed food dilution series in a computer vision-powered nutrient sensing system, via a fine-tuned deep autoencoder network. The network was trained to predict the relative concentration of sweet potato purees and achieved excellent performance, achieving 80% accuracy on primary (6 months) and intermediate (8 months) commercially prepared pureed sweet potato samples. And an inverse cooking system built by Amaia Salvador et al., 2019, that recreates cooking recipes given images of food. Our system predicts ingredients assets through a novel architecture, models their dependencies without imposing any order, and then generates cooking instructions by simultaneously focusing on the image and its inferred ingredients.

Another important level of work is to develop an evaluation system that is close to human standards, which includes various parameters such as evaluation function and loss function. This is also a quite important work, because it can guide the training of the model and play a key role in some self-supervised or semi-supervised models. On the other hand, this set of standards can also be transferred to other fields and achieve excellent results in the food production industry. As Haotian Jianget al. [3] published in IEEE in 2018, a distributed food big data query method based on Spark ecosystem is proposed to realize intelligent and scalable food health query. Based on a large amount of information related to food, such as its ingredients, environmental conditions of food production, health status of consumers, it applies semantics to a large amount of information from different data sources, which greatly improves the accuracy of food quality scoring.

In addition, the establishment of a dataset is also of great significance. Large-scale datasets can improve the performance of the model, and are also the basis for the training of deep neural network models with millions or billions of parameters. Data sets like Recipe1M+ [4] require a lot of manpower and resources to build, which is also a pretty hard job. In addition, more accurate annotation of the dataset (even at the pixel level) can be extremely helpful for tasks such as image semantic segmentation, which can be done with some carefully processed datasets, but also requires considerable effort.

3. Advantages of using computer vision to process food images

3.1 Efficiency, automation and low cost

Computer vision has significant efficiency and automation advantages in processing food images to extract nutrients. In traditional nutritional analysis, experts need to spend a lot of time to manually identify and analyze food ingredients, such as using instruments to detect food nutrient content, or scanning it to obtain spectral data for inference. Not only does it cost a lot to train experts, but it also requires fairly expensive equipment. However, using computer vision techniques, we can process a large number of food pictures in a short period of time, automatically identify food ingredients, and extract relevant nutritional information. This efficient and automated processing method greatly improves the speed and efficiency of data processing, provides the possibility of large-scale food nutrition analysis, and makes nutrition calculation more convenient, real-time and personalized.

3.2 Rich visual information

Computer vision processing food images can provide rich visual information, which is crucial for a deep understanding of the nutritional composition and quality of food. For example, for a simple fruit such as an apple, we can infer its rigidities by its color; By size and shape, we can estimate its breed; By the texture, we can tell whether it is fresh or damaged. All this information may affect the nutritional value and taste of the apple. For more complex foods, such as a pizza, computer vision can help us identify the various toppings on the pizza, including tomatoes, cheese, ham, mushrooms, etc., each of which has a specific nutritional content, and their proportion and distribution may affect the overall nutritional value of the pizza. In addition, by analyzing the visual information of food, we can also infer the preparation process and cooking method of food, such as baking, frying or steaming, which also have an impact on the nutritional composition of food. In short, the visual information obtained by computer vision processing food images can not only help us understand the nutritional content of food, but also help us understand the freshness of food, the preparation method and the quality of food ingredients, providing a comprehensive food quality assessment.

3.3 Extensibility and generality

Another advantage of computer vision techniques in processing food images is their great scalability and versatility. Once a well-trained model is built, it can be applied to a variety of different types of food images, be it fruits, vegetables, or complex cooking foods, and computer vision models can do this given enough data to train on. This means that we can use the same computer vision model to process a variety of foods, making our work in food nutrition analysis more extensive and general, greatly increasing the possibility and application range of food nutrition analysis. At the same time, it also means that each family can customize a "special model" to meet its needs.

4. Current computer vision processing difficulties

4.1 Complexity of component identification

A major difficulty faced by computer vision systems when processing food images to extract nutrients is to accurately identify the ingredients of the food. This is because, in most cases, the composition of a food product and its proportions directly determine its nutritional value. But some foods may use dyes to change their color, which can make it difficult for computer vision systems to determine their true ingredients. Similarly, if the food is mixed with other ingredients (e.g., complex cooking ingredients), it is more difficult to identify each ingredient and its proportion. This requires a high degree of accuracy and subtle judgment in computer vision systems, which is still a challenge at the current state of the art.

4.2 Physical and chemical changes during cooking

During cooking, both physical and chemical properties of food may change, which may have a considerable impact on nutrient composition recognition by computer vision systems. For example, when cooking food, the Maillard Reaction causes the surface of the food to take on a special brown color. This can mislead a computer vision system into thinking that the nutritional composition of the food has changed significantly. In addition, water evaporation during cooking may also alter the appearance of food, thus complicating the identification of its nutritional composition.

4.3 The diversity of visual representations of food

The same food may visually exhibit huge differences due to different factors such as preparation method, shape, size, photo Angle, etc. This diversity is a great challenge for computer vision systems, as the system must understand and adapt to these variations to accurately identify food products and their nutrients. This requires a high degree of complexity and flexibility in computer vision systems to handle a wide range of possible visual representations.

5. Exploration of computer vision in food

In the past two years, with the development of the neural network field, a large number of more excellent models have been proposed, such as resnet, transformer and other modules. These modules have been proven to have good performance in the field of visual detection. Maybe we can try their effect in the field of food nutrition analysis to observe whether they can also achieve good performance.

On the other hand, the field of visual segmentation is also important to the field of food image processing, as it can provide pixel-wise analysis of the image, which makes the detection more perfect. In recent years, image segmentation models such as mask RCNN [5] and blendmask [6] have been continuously developed. Both segmentation accuracy and image semantic understanding has been greatly improved compared with the past, and amazing results have been achieved in practice. It is conceivable that in the future, techniques combined with excellent image segmentation models will achieve quite good results on food image detection.

In addition, it is more worth mentioning that the big model, the full name of the big model is very large scale pre-training model. As an integration in the field of deep learning, the big model has only gone through five years of development, but it has swept the major AI lists and test data sets represented by NLP with powerful algorithm effects. ChatGPT4 was released this year and surprised everyone with how well it did. And as early as the end of 2022, AI painting based on the big model diffusion model has also won the champion of human painting competition. The big model under AIGC is like a super big warehouse, which is able to store and process a large amount of data and information. It can not only do the tasks that normal models can do, but also handle more complex and large data sets. These large models often consist of billions or even tens of billions of parameters and require a lot of computing resources and storage space to run. Just as the human brain has extremely large neurons and connections between neurons, it can complete more complex and advanced thinking and decision-making. If a large model is used for food image nutrition detection, we can learn richer visual features of food, such as color, shape, texture, and the relationship between these features and food composition and nutrition information from a large number of food image data. This deep understanding is helpful to more accurately identify food and its nutrients, and then improve the accuracy of food nutrition testing. At the same time, the multi-modal processing ability of the pre-trained model also allows us to combine image and text data for more comprehensive food nutrition detection.

On the other hand, the application of ultra-large scale pre-trained models could also have a profound impact on the food industry. This can not only help individual users with automated food nutrition monitoring and management, but also drive innovation in the food industry. For example, restaurants can use this technology to automatically calculate the nutritional information of their dishes to provide healthier dining options, and food manufacturers can use this technology to monitor and optimize the nutritional quality of their food. However, while the future looks very bright, we must also recognize the challenges when applying these models, including issues such as data privacy and model interpretability. The in-depth study and proper handling of these problems will help to realize the healthy development of large-scale pre-trained models in the field of food image nutrition detection.

6. Conclusion

Data, computing power and algorithms are the three main elements of artificial intelligence development. It is impossible to develop artificial intelligence without huge data and powerful computing power. In the future, relevant enterprises may need to participate in it and use more capital or manpower to drive its development. At the same time, the excellent effect brought by artificial intelligence will also feed back these enterprises who are willing to invest a lot of costs in it, and finally bring them considerable competitive advantages.

References

[1] Le Cun Y, Léon Bottou, Bengio Y, et al. (1998). Gradient-Based Learning Applied to Document Recognition [D]. 1998.

[2] Kaylen J. Pfisterer, Robert Amelard, Robert Amelard, et al. (2019). Towards computer vision powered color-nutrient assessment of pureed food [D]. 2019.

[3] Jiang H, Starkman J, Liu M, et al. (2018). Food Nutrition Visualization on Google Glass: Design Tradeoff and Field Evaluation [D]. 2018.

[4] Salvador A, Hynes N, Aritro Biswas et al. (2021). Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images [D]. 2021.

[5] Girshick K H G G P D R. (2018). Mask R-CNN [D]. 2018.

[6] Chen H, Sun K, Tian Z, et al. (2020). BlendMask: Top-Down Meets Bottom-Up for Instance Segmentation [D]. 2020.

[7] Bajcsy, P., & Yang, J. (2018). Computer vision for food quality evaluation. Journal of Food Engineering, 2018, 220, 1-15. doi: 10.1016/j.jfoodeng.2017.09.012.

[8] Njume, C., & Zhang, X. (2019). Food portion estimation using computer vision: A review of recent progress. Journal of Food Measurement and Characterization, 2019, 13(2), 703-712. doi: 10.1007/s11694-018-00494-7.

[9] Mezgebo, T. T., de Sousa, R. M., & Voltolini, T. V. (2020). Advances in computer vision technology for food safety and quality evaluation: A review. Trends in Food Science & Technology, 2020, 104, 29-41. doi: 10.1016/j.tifs.2020.09.004.

[10] Li, S., Zhang, L., Yang, G., & Wang, Y. (2021). Deep learning-based food nutritional value analysis: A review. IEEE Access, 2021, 9, 13941-13952. doi: 10.1109/ACCESS.2021.3051075.

[11] da Silva, M. G., dos Santos-Filho, S. D., & de Oliveira, L. S. (2022). Computer vision-based food recognition for waste minimization and sustainability. Trends in Food Science & Technology, 2022, 118, 452-471. doi: 10.1016/j.tifs.202.