A Review of Deep Learning Method for Image Recognition and Nutritional Assessment of Dishes

Chen Xieyu[1,a,*], **Tang Na**[1,b]

[1]School of Business, Geely University of China, Chengdu, 641423, China
[a]124elaine@gmail.com, [b]tangna@guc.edu.cn
[*]Corresponding author

Abstract: *Image recognition and nutritional assessment of dishes is a research field that has received wide attention in recent years. With the improvement of people's living standards and the demand for a healthy diet, the development of image-based dish recognition and nutritional assessment methods is therefore of great practical importance. This paper reviews the research on dish image recognition and nutritional assessment in terms of recognition types, learning methods, and nutritional assessment models and their application areas. First, this paper introduces different types of dish image recognition, including manual feature-based dish image recognition and deep feature-based image recognition. Second, this paper discusses the commonly used learning methods, including transfer learning and small sample learning. Then, this paper discusses the relationship between dish image recognition and nutritional assessment and introduces some commonly used assessment methods, such as multimodal and pre-trained RNN(Recurrent Neural Network). Finally, this paper summarizes the application areas of dish image recognition and nutritional assessment methods, including smart catering, health management, and food safety. This paper will help researchers gain a deeper understanding of the latest advances in dish image recognition and nutritional assessment, and provide references for research and applications in related fields.*

Keywords: *Image Recognition, Nutritional Assessment, Deep Learning*

1. Introduction

With the continuous development of deep learning methods as well as people's increasing health awareness and the growing importance of balanced nutritional intake, researchers are keen on recognizing, classifying, and detecting food, and have continuously proposed better and lighter recognition algorithms. There are rich and diverse types of dishes, and the existing food image recognition mainly focuses on specific classifications under a major category of dishes, such as dish recognition, ingredient recognition, fruit and vegetable recognition, and packaged dish recognition[1]. The recognition results can be used to advance various applications, for example, the calorie content of the diet and nutritional composition of the diet can be evaluated through dish and ingredient recognition. For dish recognition, the current image recognition technology applications are mainly focused on Western and Japanese cuisine[2], due to the large differences in the color of the dishes and the complexity of the cooking methods, so there are not many applications of image recognition technology to identify dishes, and even fewer nutritional assessment studies specifically for the dishes. Therefore, this paper will focus on the current research progress of dish image recognition and nutritional assessment.

2. Research progress

This section focuses on the current research status of dish image recognition and nutritional assessment in terms of recognition types, learning methods, and nutritional assessment models and their application areas, as shown in Figure 1.

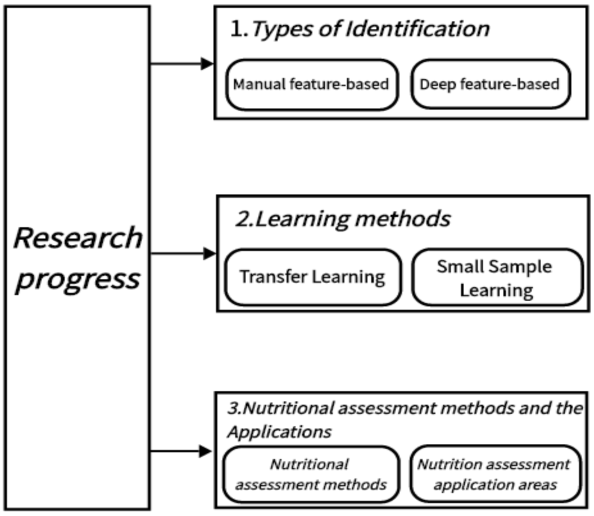
Research progress
1.Types of Identification
Manual feature-based
Deep feature-based
2.Learning methods
Transfer Learning
Small Sample Learning
3.Nutritional assessment methods and the Applications
Nutritional assessment methods
Nutrition assessment application areas

Figure 1: The overall roadmap of Research Progress

2.1. Types of Identification

Feature extraction is the most critical part of the image classification task, according to the type of features can be categorized into the following two types of dish image recognition: one is based on manual features of dish image recognition; the other is based on the deep features of the dish product image recognition.

2.1.1. Manual feature-based

In dish image recognition, handcrafted features are manually designed and extracted features are used to describe the content of the image. These features can reflect information about the shape, color, texture, and boundary of the dish[3]. Since dish images are rich in feature information, the study of manual feature extraction was focused on early dish image classification tasks. Traditional dish image recognition consists of two steps: feature extraction of dish images and training of classification models[4]. In the early work, some simple image processing methods are usually used to extract manual features and perform dish image recognition, Zhu and other scholars[5] proposed a moving segmentation dish image recognition method. The method first separates the food from the background, then extracts features such as color, texture, and shape from different regions of the image, and finally classifies the extracted features using machine learning algorithms to classify the image into different food categories, which effectively solves the limitations of traditional food classification methods. Du Meijun and other scholars[3] proposed an image recognition method for dishes based on texture features and multiclassification support vector machines, this study confirmed that texture features can provide unique features about dishes, while multiclassification support vector machines can achieve effective classification and recognition. Kong. F and other scholars[6] proposed an algorithm DietCam based on computer vision technology, the researchers carried out this algorithm system on a self-constructed dataset. The researchers tested this algorithm system on a self-constructed data set, and the experimental results showed that the algorithm showed high accuracy and robustness in regular shape food recognition.

Due to the complementarity between different handcrafted features, various types of integration algorithms are also widely used for the fusion between different features. For example, Sifan Deng et al. and Wang et al.[7-8] Both use Local Binary Patterns (LBP) and Color Moments to extract the feature vectors of each image and then use three different integrated learning algorithms to classify the images: AdaBoost, Random Forest, and Support Vector Machine Embedded AdaBoost (SVM-AdaBoost). Forest). All these algorithms are based on different ideas and principles to improve the classification performance by combining multiple weak classifiers, and the integrated classifiers provide better performance compared to individual classifiers.

2.1.2. Deep feature-based

Convolutional Neural Network (CNN) is the most commonly used algorithm for learning image features layer by layer, which simulates the mechanism of the human brain to process information hierarchically and can automatically learn visual features directly from the original pixels. Figure 2 illustrates the structure of CNN. Due to the powerful expressive ability of CNN, it was soon applied to the field of dish image recognition as well. Most researchers use the pre-training and fine-tuning model, i.e., the features are extracted directly from the pre-trained network, and then fine-tuning is performed on the existing deep network on the image dataset. Adopting this model can reduce the time required for training and the need for large amounts of labeled data, which can lead to better feature representations and reduce the risk of overfitting, thus improving the performance of the model on specific tasks[9]. Scholars such as Zhao Ming[10] adapted networks such as VGGNet and ResNet to specific image recognition tasks by fine-tuning them to improve their recognition performance. Scholars such as Li Hua et al.[11], on the other hand, exploited the generalization ability of the Inception-v3 model pre-trained on a large-scale image dataset and achieved good performance on the task of classification of dish images. Similarly, Smith et al.[12] applied the pre-trained MobileNet model to dish image recognition. By fine-tuning the MobileNet model, the model showed high accuracy and robustness in recognizing various types of dish images. While Pouladzadeh et al.[13] studied a mobile dish recognition system. In the training phase, candidate regions are first generated using the region suggestion algorithm, and features of all regions are extracted by CNN, followed by region mining using maximum coverage using the proposed submodular optimization method to select positive regions for each dish category.

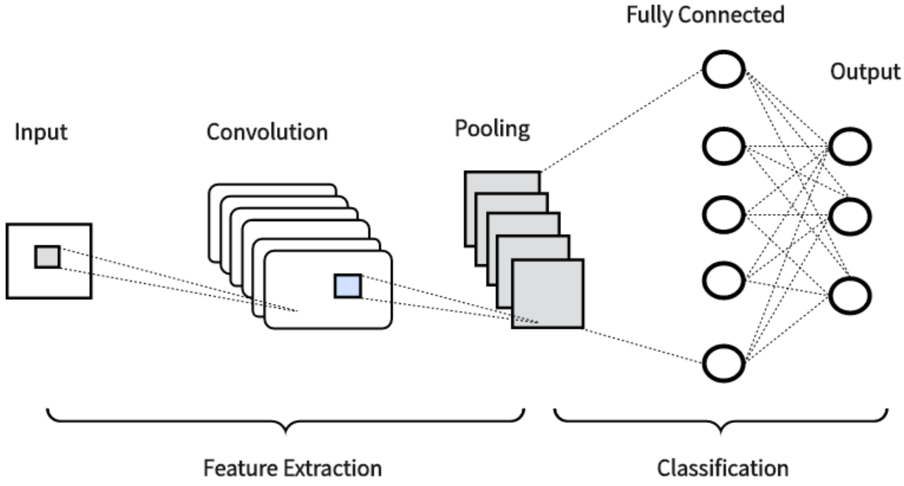

Figure 2: The structure of CNN

2.2. Learning methods for dish image recognition

Dish image recognition is important in many real-world applications in areas such as restaurant management, health monitoring and social media. However, the dish image recognition task is still challenging due to the visual diversity of dishes and the lack of samples. To overcome these challenges, researchers have proposed various approaches based on transfer learning, small sample learning, etc. to improve the performance of dish image recognition by utilizing the knowledge of pre-trained models.

2.2.1. Transfer Learning

Li et al.[14] pre-trained the model on a large task and applied it to a dish image recognition task. The authors used a common strategy in migration learning, i.e., the parameters of the pre-trained model were used as initial weights and fine-tuned. Through migration learning, the accuracy of dish image recognition can be significantly improved, enabling the model to better learn and capture the features of the dish.

Deng and other scholars[15] proposed a transfer learning method based on Generative Adversarial Networks (GAN) for solving the small sample problem in dish image recognition. They expanded the dataset by generating additional samples of dish images using GAN and combined it with a convolutional neural network (CNN)-based classifier for training. The experimental results show that the method achieves satisfactory dish image recognition results with small samples. By introducing the generative model, the researchers effectively solved the problem of insufficient data and improved the generalization ability of the model. In addition to this, Wu and other scholars[16] proposed an adaptive transfer learning-based approach for improving the performance of dish image recognition. They realized adaptive feature extraction for the dish recognition task by learning the association between source and target domains. Specifically, the study proposes a new loss function to measure the difference between the source and target domains and adaptively adjusts the network structure according to the difference. The method achieves better results than traditional transfer learning methods in the dish image recognition task, illustrating the importance of adaptive transfer learning for dish image recognition.

2.2.2. Small Sample Learning

Su et al.[17] proposed a small sample learning method based on dense connectivity for dish image recognition. The researchers achieved more effective feature extraction and classification by introducing a dense connection block and a global average pooling layer in the convolutional neural network. Experiments proved that this method achieved better performance for dish image recognition in small-sample learning scenarios compared with traditional neural network structures. Zhuang et al.[18] introduced a small-sample learning method based on multi-scale spatial pyramid deep networks for dish image recognition. The researchers proposed a novel network structure that fuses features from multiple scales together to improve the representation of dish images. Experimental results show that the method can effectively recognize and classify dish images in small sample learning scenarios. Wang et al.[19] proposed a small sample learning method based on contrast self-supervised learning for dish image recognition. Using self-supervised learning, the researchers enable the model to train itself on unlabeled data, thus improving the model's ability to represent dish images. Specifically, they designed a contrast loss function that maximizes the distance between pairs of positive and negative samples, thus enabling the model to better distinguish between different dish categories.

2.3. Nutritional assessment methods and the Applications

2.3.1. Nutritional assessment methods

Most scholars have performed joint learning or feature fusion to improve the accuracy and robustness of nutritional assessment by using a transfer learning model and combining it with a model for multimodal data to fuse images and text. For example, Johnson et al.[14] used pre-trained RNN models for feature extraction and modeling while introducing an attention mechanism to capture important image and text features. The method achieved better results on the dish calorie estimation task. In addition, Liu et al.[15] proposed an adaptive migration learning method that combines image, text, and user-labeled data and captures the correlation between different modal data by learning a shared feature space. Next, an adaptively tuned classifier is used for prediction, which effectively improves the accuracy and generalization of the nutritional value assessment of dishes. Yang et al.[16] designed a Bayesian network-based framework that includes modeling and prediction of different modal features. The model can capture the underlying relationships of the dishes and thus perform nutritional value assessment to optimize the accuracy of the prediction results. Zhang et al.[17] used a multi-task learning model to simultaneously process image and text data by sharing convolutional and fully connected layers. The model can extract shared features from different modal data and perform prediction in dish classification and nutritional value assessment tasks.

2.3.2. Nutrition assessment application areas

With the development of vision technology and deep learning techniques, scholars have proposed the concept of passive dietary monitoring[18-19], i.e., the use of sensors, such as video cameras, to record one's diet and perform food recognition and subsequent tasks. Therefore, among the application areas of nutritional assessment, dish image recognition is the first and crucial step.

Liao and other scholars[18] developed a passive diet monitoring technique based on a convolutional neural network and recurrent neural network. By combining data from multiple sensors it can achieve accurate monitoring of the user's dietary behavior, process and analyze multimodal data, provide personalized dietary advice, and promote health management. Zhang Gang and other scholars[19] researched and designed a passive diet monitoring system for smart aging. The system uses sensors to collect data on the user's eating behavior, including the use of utensils, eating speed, and the amount of food eaten. By designing and implementing the monitoring system and using data mining techniques to analyze users' eating behavior, personalized dietary advice and health management support can be provided to the elderly population. Huang and other scholars proposed a passive diet monitoring method based on sensing technology. The researchers used sensors to monitor the user's eating behavior, such as the number of chewing times and eating speed and extracted the features of the user's eating behavior. It can be seen that using the passive diet detection method, the user's eating behavior can be monitored in real-time and accurately to further understand the eating habits and help the user improve their eating habits and manage their health.

In summary, it is found that most of domestic research scholars prefer extracting manual features and combining machine learning models for dish image recognition, and fewer scholars combine deep learning models. On the contrary, foreign researchers and scholars mainly adopt the pre-training and fine-tuning model for dish image recognition and segmentation. In addition, most of the domestic scholars' research on nutritional assessment focuses on application areas, and most of them use passive diet detection systems to monitor users' dietary behavior. In contrast, foreign scholars prefer to use migration learning models and combine the models of multimodal data to fuse images and texts for joint learning or feature fusion to improve the accuracy and robustness of nutritional assessment.

3. Conclusion

Nowadays, with the rapid development of the food industry, the phenomena of degradation and nutritional loss of prefabricated dishes have gradually become the core of public opinion, and there is a lack of theoretical mechanisms related to this research. Image recognition classification of dishes and assessment of their nutritional value will help dishes to achieve the automated classification of dishes, but also for the standardization of the cooking process, the choice of ingredients to provide theoretical support. At present, there is a lack of theoretical research on image recognition of dishes, and domestic scholars tend to use manual feature extraction combined with the machine learning model method, this method has some limitations to a certain extent, the extraction of manual features needs to rely on manual experience and domain knowledge, and requires a lot of time and energy. In addition, the machine learning model may not be flexible and accurate enough for complex dish feature representation. Foreign scholars prefer to use pre-training and fine-tuning deep learning models for dish image recognition and segmentation, which can better capture the details and features of dish images by utilizing large-scale data and the powerful representation capability of deep neural networks. However, the drawbacks are that the pre-training to fine-tuning process requires more computational resources and time, and fine-tuning for a specific task may require a large amount of labeled data. In this paper, we summarize the research progress of food image recognition from three dimensions, namely, recognition type and learning method and nutritional assessment method and application, and this study explores the application of deep learning in the image recognition of dish dishes, which will help to improve the accuracy and efficiency of the image recognition and nutritional assessment of dish dishes and reduce the reliance on handmade features by adopting the deep learning model, which is more important for the promotion of the image recognition technology in the field of dish dishes. This is of more important significance to promote the development of image recognition technology in the field of dishes.

References

[1] Lukas Bossard, Matthieu Guillaumin, Luc Van Gool. food-101 - mining discriminative components with random forests//Proceedings of the European Conference on Computer Vision. Zurich, Switzerland, 2014: 446-461

[2] Liao Enhong, Li Huifang, Wang Hua, Pang Xiongwen. Image Recognition of Dishes Based on Convolutional Neural Networks. Journal of South China Normal University (Natural Science Edition), 2019, 51(4): 113-119

[3] Du Meijun, Ye Chunyang, Chen Dongxiao, et al. Image recognition of dishes based on texture features and multi-classification support vector machine[J]. Computer Engineering and Applications, 2017, 53(22): 130-134.

[4] Zhou, Y., Xu, X., Yang, X., & Liao, W. (2019). Deep learning-based automatic food recognition: a review. journal of food engineering, 270, 10-20.

[5] Zhu F, Bosch M, Schap T R, et al. Segmentation assisted food classification for dietary assessment// Proceedings of the International Society for Segmentation assisted food classification for dietary assessment// Proceedings of the International Society for Optical Engineering. San Francisco, USA, 2011: 78730B

[6] Kong F, Tan J. Dietcam: Regular shape food recognition with a camera phone//Proceedings of the International Conference on Body Sensor Networks. Dallas Dallas, USA, 2011: 127-132

[7] Deng Sifan, Xie Yingli. Research on image classification algorithms based on manual features and integrated learning [J]. Computer Applications and Software, 2020, 9(6): 180-185.

[8] Wang, L., Peng, J., Huang, Q., & Jiang, S. (2018). An effective feature integration framework for

image classification. IEEE Transactions on Multimedia, 20(3), 637-647.

[9] Kawano, Y., Yanai, K., & Sumiya, K. (2014). Food image recognition using deep convolutional network with pre-training and fine-tuning. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (pp. 429-433).

[10] Zhao M, Wang G. Research on image recognition model based on pre-training. Electronics and Software Engineering, 2020, 9(5): 102-106.

[11] Li H, Wang C. Research on image recognition of dishes based on fine-tuning. Journal of University of Electronic Science and Technology, 2019, 46(3): 520-526.

[12] Smith, A., et al. Fine-Tuning Pretrained Convolutional Neural Networks for Food Classification. IEEE Transactions on Image Processing, 2017, 26(2): 1059-1073.

[13] Pouladzadeh P, Shirmohammadi S. Mobile multi-food recognition using deep learning. ACM Transactions on Multimedia Computing, Communications, and Applications. 2017, 13(3s): 1-21

[14] Li, H., Zhang, Z., Huang, Z., & Xian, Z. (2018). "Transfer learning for food recognition: an empirical study." Journal of Food Engineering, 218, 14-21.

[15] Deng, C., Li, Y., Yu, E., & Wu, F. (2017). "Food recognition through transfer learning with generative model." IEEE International Conference on Multimedia and Expo, 802-807.

[16] Wu, B., Shang, J., & Zhang, H. (2019). "Adaptive transfer learning for food recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 450-459.

[17] Su, X., Xu, M., & Hu, W. (2018). Densely connected small sample learning for food image recognition. Journal of Visual Communication and Image Representation, 57, 231-239.

[18] Zhuang, F., Tsai, S. S., Kuo, Y. H., & Liao, H. Y. (2017). "Food recognition based on a multi-scale spatial pyramid deep network." Neurocomputing, 249, 70-78.

[19] Wang, Z., Zeng, Y., Hu, J., & Huang, G. (2020). "Few-shot food recognition via contrastive self-supervised learning." Pattern Recognition, 98, 107082.