A Review on Food Classification using Convolutional Neural Networks

Sruthy R

Guest. Lecturer, Dept. of Electronics and Communication, NSS Polytechnic College, Pandalam, Kerala, India

---***---

Abstract - *In recent years, researchers have developed algorithms leveraging machine learning techniques that can automatically detect food and nutritional information on images. Food recognition techniques will also assist with the assessment of calories, the detection of poor quality food, and the development of systems that monitor a diet to fight obesity. A classic approach would not recognize complex features of food images. Due to the fact that Convolutional Neural Networks (CNN) can analyze massive data sets and are capable of automatically identifying complex features, they have been used to classify food. A comprehensive review of food classification from food images using convolutional neural networks is provided in this paper.*

Key Words*: *Convolutional Neural Network, Deep learning Food Classification*, *Feature extraction, Classification

1. INTRODUCTION

In the last few decades, researchers have focused on using computer vision and machine learning to identify food and nutritional information automatically from captured images. It is critical to accurately estimate the calorie content of food when determining dietary intake. Individuals tend to consume too many calories and do not exercise enough. Amid their hectic schedules and stress, most people nowadays lose sight of what they eat. Therefore a reliable classification of foods is essential to keep track of our food intake.

The task of automatically recognizing foods from images is particularly complicated. Image-based food recognition is more complex since many food items resemble one another in appearance, size, and color, and different regions have a different food preparation and cooking styles. Deep learning improved the accuracy of classification tasks over traditional image analysis methods. The low inter-class variance and high intra-class variance of food images make Machine Learning methods unable to discern some complex features, whereas CNNs can easily do so[1]. Contrary to conventional machine learning approaches, deep learning algorithms do not require a feature extraction algorithm [2]. The main purpose of this paper is to give an overview of current techniques in food image classification using Convolutional Neural Networks. Section 3 addresses the different food classification models used in the study. Section 4 presents a comparison of the performance of the different CNN models. Section 5 presents a summary of the study.

2. LITERATURE REVIEW

In [3], a 2d Convolutional neural network with the max pooling function has been used for food classification. The food-101 dataset has been chosen as the database in this method. By analyzing the amount of food visible in the image, this method identifies the food type and calculates the calorie value. The pre-trained CNN models like AlexNet, VGG16 (Visual Geometry Group), ResNet-50 (Residual Network), and Inception V3 have been used for food classification in [4]. Transfer learning works reasonably well with these models since the earlier pre-trained layers have already learned many of the features required for food image recognition.

In [5] Liu. et al. proposed an image recognition system influenced by LeNet-5, AlexNet, and GoogleNet. This system uses a CNN with 3 convolutional layers, 2 subsampling layers, and one fully connected layer. Publicly available databases, UEC-100/UEC-256 and Food 101, were used in this system. An Inception V3 based customized food classification have proposed in [6]. This CNN-based model was trained over more than a thousand food images and obtained an accuracy of 98.7%.

A novel CNN model and Inception V3 based food classification system is implemented in [7]. Food 101 is the dataset used for this study. Before training the convolutional neural network with the images, a pre-processing step have applied to lower the computational complexity and improve the accuracy. Stochastic gradient descent, Nesterov's accelerated gradient, and Adaptive Moment Estimation were used to train three separate deep CNNs for food classification from scratch [8]. Then the networks are compared with AlexNet and CaffeNet. They get fine-tuned using the same learning techniques. The Food11 and Food101 datasets were used to create training, validation, and testing datasets.

The Inception V-3 and V-4 models were fine-tuned to identify the foods in [9], and an attribute estimation model was developed to quantify the food attributes. Data augmentation, multi-cropping, and other approaches were used to improve the result. The presented classification and attribute extraction model has an accuracy of 85 percent. SqueezeNet and VGG-16 are used for food identification in [10]. SqueezeNet achieves an accuracy of 77.20 percent with few parameters. Then, the VGG-16 network improves the performance of the automatic classification of food images. This VGG-16 has achieved a high accuracy of 85.07 percent due to the increased network depth. Table 1 shows the comparison table of different food classification systems.

The pre-trained 'Alexnet' and 'Caffenet' models were fine-tuned, and a new complementary structure trained with food images from the Food-11, FooDD, Food100, and web archives dataset was used in [8] to classify foods.

3. METHODOLOGY

The flowchart of the food classification system using the Convolutional Neural Network is shown in Fig. 1. The food classification system consists of two phases. The first phase is the training phase, in which feature extraction and classification of food images are performed to develop a prediction model. In the testing phase, this prediction model is used to classify the food test images.

Here, CNN performs automatic feature extraction and classification of food images. CNN's have an input layer, an output layer, and hidden layers, all of which aid in image classification and processing. The convolutional layer is the foundation of CNN. By sliding a kernel over the input images, this layer performs convolution and generates a feature map. The dimension of the feature map is minimized by the pooling layer. Max pooling is the most common type of pooling. Pooling is achieved by sliding a window over the input and taking the highest value in the window. The output of the last pooling layer is converted to a one-dimensional vector by the FC layer [11].

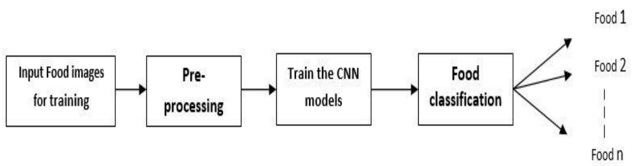

Fig-1:- Flow diagram

The various CNN models involved in this survey are

3.1 AlexNet

AlexNet [12] is an eight-layer convolutional neural network architecture. The dimension of the input image is 227×227×3. It consists of 5 convolutional layers and 3 fully connected layers (FC). The first two convolutional layers are coupled with overlapping max-pooling layers to extract many features. Softmax function is utilized for activation in the output layer. A dropout layer reduces over fitting during training. The 'dropped out' neurons do not participate in the 'forward pass' or back propagation. The layers with the dropped out neurons are located in the first two FC layers.

3.2 VGG-16

VGG-16 [13][14] is a CNN model with 16 convolutional layers. It remains one of the most widely used image recognition models today. The VGG network accepts images with a size of 224×224×3 pixels. VGG16 is a very large network with a total of 138 million parameters. The main disadvantage of the VGG16 network is that, as a large network, it requires more time for the training process.

3.3 ResNet-50

ResNet50 [15] [16] is a CNN model with 48 convolutional layers, 1 MaxPool layer, and 1 Average Pool layer. Skip connections are used by residual neural networks to jump past some layers. Skip connections can function in two ways. First, they solve the vanishing gradient problem by creating an alternative path for the gradient to follow. They also allow the model to learn an identity function. In this way, the higher layers of the model work equally well as the lower layers. The default input size of ResNet-50 is 224×224×3.

3.4 Inception V3

Inception V3 [15] [17] is a CNN model developed by Google. Inception models have parallel layers and no deep layers. Multiple Inception modules make up the Inception model. The core module of the Inception V1 model consists of four parallel layers. Inception V3 is the extended version of Inception V1. It uses auxiliary classifiers and has 42 layers.

3.5 LeNet-5

The network is named Lenet-5 [18] [19] because it contains five layers of learnable parameters. The network consists of a single channel because the input is a 32 x 32×3 grayscale image. It uses a combination of average pooling and three pairs of convolutional layers. It has two fully connected layers. Finally, a Softmax classifier classifies the images into their specific categories.

3.6 SqueezeNet

SquuezeNet [15] [20] is a smaller, deep neural network with 18 layers. It has Fire modules that "squeeze" parameters through 1x1 convolutions. The architecture of SqueezeNet consists of "squeeze" and "expand" layers. These are sent through a combination of 1×1 and 3×3 convolutional filters in an "Expand" layer. SqueezeNet starts with a single convolutional layer, followed by 8 Fire modules, and finally a final convolutional layer. There are no FC layer in this neural network architecture. Relu activation function is employed in SqueezeNet.

3.7 CaffeNet

CaffeNet [21] is an AlexNet variation. In contrast to AlexNet, CaffeNet does pooling before normalizing. The input image has a size of 224 ×224× 3 pixels.

3.8 GoogleNet

This CNN model is based on the Inception architecture. GoogleNet [15][22] uses a stack of nine inception blocks and global average pooling, and can run on individual devices even with limited processing capabilities. This network is influenced by LeNet architecture. Three filter sizes are used in the convolution: (1 * 1), (3 * 3), and (5 × 5). In combination with the convolutions, a max-pooling operation is performed, and then the resulting data is data is sent to the next inception module.

4. COMPARISON

Table-1:- Comparison table

No	Title	CNN used	Dataset	Accuracy
1	Food Classification from Images Using Convolutional Neural Networks [3]	Customized 2D CNN architecture	FOOD-101	86.97%
2	Food Image Classification with Convolutional Neural Networks [4]	AlexNet VGG16 ResNet-50 Inception-V3	FOOD-101	61.9% 43.5% 71.4% 85.2%
3	DeepFood: Deep Learning-Based Food Image Recognition for Computer-Aided Dietary Assessment [5]	Customized CNN	UEC-256 FOOD-101	87.2% 93.7%
4	Transfer Learning: Inception-V3 Based Custom Classification Approach For Food Images[6]	Customized CNN based on Inception-V3	Real-time food images from various resources.	96.27%
5	Food Image Classification with Convolutional Neural Network [7]	Customized CNN Pre-trained Inception-V3	FOOD-11	74.70% 92.86%
6	Comparison of convolutional neural network models for food image classification [8]	AlexNet CaffeNet Structure 1 Structure 2 Structure 3	Extended version of FOOD-11	86.92% 83.7% 73.87% 71.7% 69.92%
7	Machine Learning Based Approach on Food Recognition and Nutrition Estimation [9]	Fine tuned Inception-V3 Inception- V4	Created dataset	95% 94.7%
8	Automated Food image Classification using Deep Learning approach[10]	SqueezeNet VGG-16	FOOD-101	92.83% 94.02%

5. CONCLUSION

Food plays an essential role in human life, providing various nutrients, and therefore the consumption of food is crucial for our health. Food classification is therefore a crucial aspect in maintaining a healthy lifestyle. In the world of health and medicine, food image classification is an emerging research area. A survey of automatic food classification methods based on Convolutional Neural Networks has been presented. The majority of the work uses the Food-101 dataset to train the models. Among the different approaches, InceptionV3-based systems provide higher accuracy in food image classification.

REFERENCES

[1] S. Yadav, Alpana and S. Chand, "Automated Food image Classification using Deep Learning approach," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, pp. 542-545, doi: 10.1109/ICACCS51430.2021.9441889.

[2] Abdulnaser Fakhrou, Jayakanth Kunhoth, Somaya Al Maadeed "Smartphone-based food recognition system using multiple deep CNN models", Multimedia Tools and Applications volume 80, pages33011–33032 (2021)

[3] D. J. Attokaren, I. G. Fernandes, A. Sriram, Y. V. S. Murthy and S. G. Koolagudi, "Food classification from images using convolutional neural networks," TENCON 2017 - 2017 IEEE Region 10 Conference, 2017, pp. 2801-2806, doi: 10.1109/TENCON.2017.8228338.

[4] Jiang, Malina. "Food Image Classification with Convolutional Neural Networks." (2019).

[5] Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., Ma, Y. (2016). DeepFood: Deep Learning-Based Food Image Recognition for Computer-Aided Dietary Assessment. In: Chang, C., Chiari, L., Cao, Y., Jin, H., Mokhtari, M., Aloulou, H. (eds) Inclusive Smart Cities and Digital Health. ICOST 2016. Lecture Notes in Computer Science (), vol 9677. Springer, Cham. https://doi.org/10.1007/978-3-319-39601-9_4

[6] Vishwanath C Burkapalli, Priyadarshini C Patil, "Transfer Learning: Inception-V3 Based Custom Classification Approach For Food Images", ICTACT Journal on Image and Video Processing (Volume: 11 , Issue: 1),August 2020, doi: 10.21917/ijivp.2020.0321

[7] M. T. Islam, B. M. N. Karim Siddique, S. Rahman and T. Jabid, "Food Image Classification with Convolutional Neural Network," 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), 2018, pp. 257-262, doi: 10.1109/ICIIBMS.2018.8550005.

[8] Gözde Özsert Yiğit & B. Melis Özyildirim (2018) Comparison of convolutional neural network models for food image classification, Journal of Information and Telecommunication,2:3, 347-357, DOI: 10.1080/24751839.2018.1446236

[9] Shen, Zhidong & Shehzad, Adnan & Chen, Si & Sun, Hui & Liu, Jin. (2020). Machine Learning Based Approach on Food Recognition and Nutrition Estimation. Procedia Computer Science. 174. 448-453. 10.1016/j.procs.2020.06.113.

[10] S. Yadav, Alpana and S. Chand, "Automated Food image Classification using Deep Learning approach," 2021 7th International Co nference on Advanced Computing and Communication Systems (ICACCS), 2021, pp. 542-545, doi: 10.1109/ICACCS51430.2021.9441889.

[11] A. Arya and M. Manuel, "Intelligence Quotient Classification from Human MRI Brain Images Using Convolutional Neural Network," 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), 2020, pp. 75-80, doi: 10.1109/CICN49253.2020.9242552.

[12] Shipra Saxena, "Introduction to The Architecture of Alexnet", 2021

[13] https://neurohive.io/en/popular-networks/vgg16/

[14] https://medium.com/@mygreatlearning/what-is-vgg16-introduction-to-vgg16-f2d63849f615

[15] https://blog.paperspace.com/popular-deep-learning-architectures-resnet-inceptionv3-squeezenet/

[16] https://iq.opengenus.org/resnet50-architecture/

[17] https://iq.opengenus.org/inception-v3-model-architecture/

[18] https://www.analyticsvidhya.com/blog/2021/03/the-architecture-of-lenet-5/

[19] https://www.datasciencecentral.com/lenet-5-a-classic-cnn-architecture/

[20] https://towardsdatascience.com/review-squeezenet-image-classification-e7414825581a

[21] https://medium.com/coinmonks/paper-review-of-alexnet-caffenet-winner-in-ilsvrc-2012-image-classification-b93598314160

[22] https://towardsdatascience.com/deep-learning-googlenet-explained-de8861c82765