Grenze International Journal of Engineering and Technology, January Issue

A Holistic Approach to Know Food Image Recognition Technology using Inception v3 an Image Classification Model

Dr. R. J. Shelke[1], Mrs. M. N. Shrigandhi[2] and Pooja P. Goyal[3]
[1-3]Walchand Institute of Technology, Solapur, India
Email: rjshelke@witsolapur.org, mapasnur@witsolapur.org, goyal.pooja201991@gmail.com

***Abstract*— This research helps to understand the efficient way to know the food image recognition using a robust method Inception V3, a CNN model for recognizing the food image. Food has been a very crucial thing in our day-to-day life, and to measure what we eat is really important. Inception V3 Convolutional Neural Network model is proposed for the efficient recognition of the food images. In this model the Inception V3 model is combined with the concatenated features to enhance the performance. It enhances the feature extraction and gives it to the image classification. The Inception V3 model is built in Tensor Flow environment using Python for deploying the CNN models.**

***Index Terms*— Convolutional Neural Network, Machine Learning, Inception V3, Food Image, Image Classification, Deep Learning.**

I. INTRODUCTION

Food Image Recognition is one of the crucial fields in today's era where lot of research work is happening in the deep learning algorithms. Social media platform is contributing for the change in the food habits. Image processing, Neural Networks these are some of the areas where food image recognition techniques are implemented. Related to this food recognition many applications are developed in market for food logging, tracking of diet, diet evaluation, marketing of food products, know what you eat. The proposed system uses Inception V3 a CNN model to extract the features from the food image and predict the food image [1].

It is very tedious for a human eye to do task like calorie estimation, knowing the ingredient of a recipe, labelling of food, all these tasks become easy with the help of machine learning algorithms. The current generation has become very conscious about the food they eat, tracking the calorie, diabetics control food, hence to solve these issues automatic food recognition methods from machine learning and computer vision are coming up in market. Labelling the food, knowing your food is very important as people are unable to track the calorie deficit in this stressed and busy life, where comes the significance of food classification and recognition. Smart applications are built using machine learning and deep learning methods for an efficient food labelling.

With the advancement in memories, computational efficiency, processing technologies it has made quite easier and effective to train the machine learning models within less time. In this system Inception V3 a CNN model is used to build a food image recognition model. In this model, while designing Inception V3 model several layers for pooling, convolution is proceeded by linked layer for categorization in CNN.

Grenze ID: 01.GIJET.12.1.807

These layers take the flattened output from these convolutional and pooling layers and proceed to classify the input image, assigning the output to its specific category class.
The model is trained and tested using a large dataset to increase the accuracy of the model. It is implemented using Python in Tensor Flow environment which facilitates in the implementation of machine learning CNN Inception V3 model. Inception V3 is pre-trained on ImageNet dataset, which makes it proficient in classifying a wide range of images [2]. It has become very easier for machine learning models to train and test the various algorithms in deep learning like Convolutional Neural Networks and Inception V3 model with computer vision techniques.

II. Related Works

Food is a very important topic for research work. There are many aspects to finding food recognition based on depth, quantity, volume estimation, location-based GPS, taste estimation, multi food recognition and real-time recognition.
Weiqing et al. [22] witnesses remarkable advancements in visual recognition for larger datasets in food domain. It gives importance to the computer vision and food-oriented vision and multimodal tasks. Food2K the largest food recognition dataset with 2000 categories and over 1 million images. It is proposed a deep progressive region enhancement network, it consists of two features, local feature learning and region feature enhancement. It is further more explored with nutritional understanding of food, backbone to improve the performance of food related tasks.
Landu et al. [12] generates system of proposals by Region Proposal Network (RPN) derived from Faster R-CNN model. UEC-FOOD100 and UEC-FOOD256 extensively used as datasets. The system analyzes nutritional ingredients and dietary assessment report by calculating the number of calories, fat, protein, carbohydrate.
Mohammed et al. [16] proposes the concerns of many dieticians and healthcare conventions about consuming the right amount of food. Local Binary Patterns (LBP), Gabor Filters, Scale Invariant Feature Transform (SIFT) are applied to extract features from food image, this improves the overall classification accuracy. The deep learning approaches the feature extraction automatically through a series of connected layers with fully connection layer for final classification.
Gianluigi Ciocca et al. [8] proposes Convolutional Neural Network (CNN) architectures on food categories, food state and both. Support Vector Machine (SVM) as end-to-end classification deep features outperform hand-crafted features. Computer Vision techniques help in monitoring the daily dietary habits, daily recorded manuals. Hand Crafted (HC) and Learned Features (LF) are features extracted from manual predefined algorithm and Convolutional Neural Network (CNN).
Mengyang et al. [14] proposes a sequential learning method for Neural Network (NN based) model producing the ingredients following a cooking logic for the recipes. A Double flow feature fusion module (DFFF) is proposed to obtain features from two image based with original image features are feed into Bidirectional long short time (Bi-LSTM) memory to produce ingredients.

III. Methodology

The implementation of the proposed system which uses a model of Inception v3 a CNN model is discussed in the methodology.

A. Dataset Collection

This system takes dataset from Kaggle. Many datasets are available for machine learning applications on the web platform Kaggle. In this system the Kaggle dataset Indian Food Images Top 20 is taken. The total images in dataset are 5838 images for 20 classes. Out of which 3503 images are for training and 2335 images are for testing.

B. Network Architecture

The Inception V3 model is pre-trained on a massive dataset like ImageNet, it is 48-layer deep CNN structure [21]. In this model it signifies that it saves a lot of training time and enhaces higher accuracy. The model is trained on Indian Food Images Top 20 kaggle dataset, this is categorized images from India's traditional and state famous food items. Here the model fine-tune the weights of the new layers and also sometimes pre-trained layers for the classification of food items more accurately. Data Augmentation technique is applied on the existing image dataset like rotation, flip or zoom to increase the size and diversity of the dataset. This helps to generalize better and also prevent from the risk of overfitting in the dataset. This model acts as a powerful

feature extractor using CNN and Inception V3 model for the food images, hence identifying patterns like shapes, textures and colors [3]. Thereafter these features are fed into the new layers for classification. The final layer in the model takes the features and makes a prediction of the food image category. The softmax function works like the output function similar to a max-layer. The system here uses Convolutional Neural Network CNN with Inception V3 for the food images leveraging the power of deep learning for classifying and identifying food images within the image dataset. Fig. 1 shows the System Architecture of Convolutional Neural Network and Inception V3 for the feature extraction and classification of the food image. Average Pooling is 2-Dimentional function with a pool size of (8,8), it decreases the data variance and computational complexity.

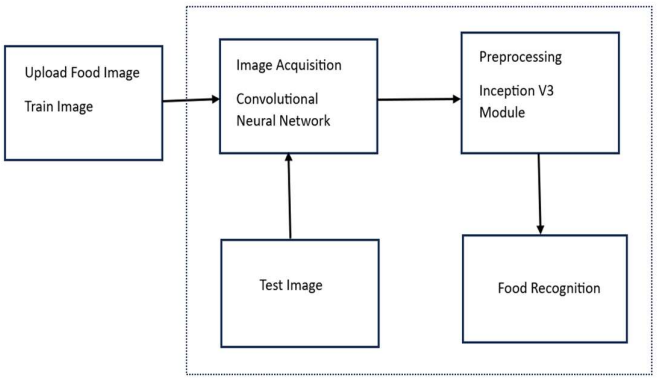

Figure. 1. System Architecture for Food Image Recognition

This approach helps for efficient training on the food image dataset, for achieving high accuracy in identifying and categorizing different food classes [4]. Here it involves convolutional layers to extract features from food images. During the classification process methods like edges, textures and shapes of the food image is considered.

Also the system makes Inception v3 with the convoltuional layers more sophisticated within its Inception module to extract the rich features from the image which is widely used in deep learning. It uses the multi-level feature extraction by incorporating filters of various sizes in the same layer.

C. CNN Architecture

Convolutional Neural Network is a special type of neural network which is effective of processing a dataset of images in a grid-like structure. The pixel form of data is fed into the network for feature extraction and classification of data.

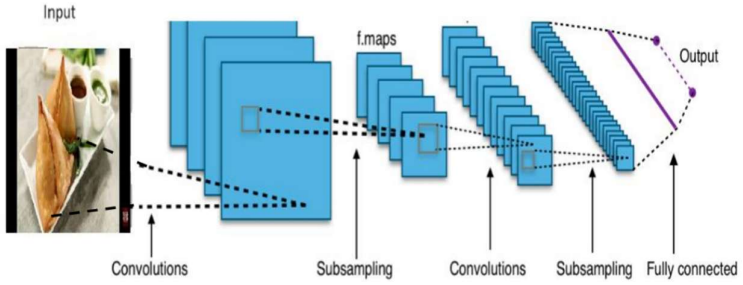

Figure. 2. CNN Architecture

Fig. 2 shows the CNN Architecture model which represents the different layers of a Convolutional Neural Network. The layers comprise of Convolutional Layers, Pooling Layers and Fully Connected Layers. The Activation function ReLU (Rectified Linear Unit) introduces non-linearity into the network, which helps it to learn more complex problems. Kernels are filters in Convolutional Layers; it extracts the features from the input data. These Kernels slide across the input image in the convolution operation so that it can extract feature maps. Pooling Layer is a down sampling operation; it reduces the spatial dimensions of the input feature maps while retaining the most important information [20]. Fully Connected Layer is the dense layer it takes the flattened input and learning complex relationships between features to make predictions [5]. The CNN here does feature

extraction and classification of food image by extracting the features and gives it to fully connected layer to classify the food image.

D. Inception V3 Architecture

The Inception module uses parallel convolutional layers of different sizes 1 X 1, 3 X 3, 5 X 5 to capture features at different scales. The output of each filter is concatenated and passes to the next layer. This approach captures the local and global features of the input image. Factorized convolutions are used to decrease computational cost and number of parameters. Here Inception V3 replaces larger convolutions with smaller ones. This model reduces the grid size and also maintaining feature maps, this prevents information loss and improves performance. Auxiliary classifier is used during the training for regularization to improve model performance. This model utilizes RMSProp Optimizer an adaptive learning rate optimization algorithm. In deep learning models it increases the performance and speed and also a variant of gradient descent algorithm. Batch Normalization is used to enhance the training stability and normalizes layer input to improve the speed [6]. This model includes Average Pooling, Convolution, Max-Pool, Concatenation, Dropout, Fully-connected, SoftMax layers.

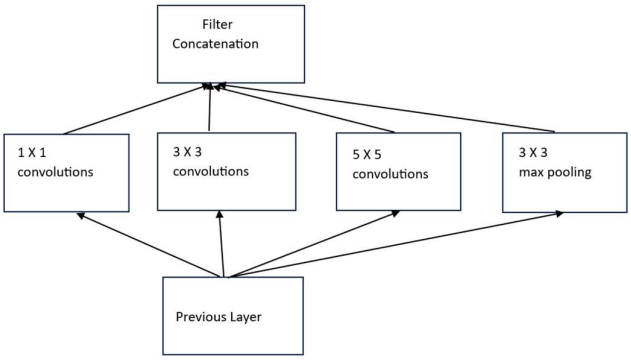

Figure. 3. Inception V3 Architecture Module

In CNN the layers are stacked sequentially, here Inception module uses parallel branches with different filter sizes 1 x 1, 3 x 3, 5 x 5 convolutions and pooling operations on the feature map. In Inception V3 Module now the output of these parallel branches is now concatenated as shown in the Fig. 3, simultaneously the network captures the best features at multiple scales and resolutions. This utilizes computational efficiency and extracts rich features efficiently from the image.

E. Training Phase

Using the Kaggle dataset for Indian Top 20 Food Images we will train the Inception V3 Convolutional Neural Network Model. In this system the food image recognition process of training it uses the Inception V3 a pre-trained model to recognize the food image [7]. The pre-trained weights from ImageNet a large dataset with 1000 object categories are used as a starting point and the model is then fine-tuned on a food image dataset. In the pre-processing process data collection is done, by gathering the food images from Kaggle dataset with the different classes. TensorFlow helps to build a machine learning model, implemented in Microsoft Visual Studio Code using Python code to develop the Inception V3 Convolutional Neural Network interface.

Training for the dataset with Inception V3 model is used with Convolutional Neural Network (CNN) to classify the food images. It uses the Kaggle dataset Indian Food Images Top 20 to recognize and fed to the model for further classification and recognition using the Inception V3 algorithm with the deep learning features of Convolutional Neural Networks. The total of different 20 food classes items are available for training the model.

F. Train Inception V3 CNN Model

In this system, the Inception V3 uses Convolutional Neural Network as a building block to recognize the food image from the dataset. The Inception V3 convolutional layer extracts the feature maps from the input image, it uses Kernels to convolve the features from an image [9]. Parameter Sharing is applied across the same Kernel for the entire input; this allows it to learn the patterns in the image regardless on its location inside the entire image.

ReLU activation function is applied for non-linearity into the network, it allows it to learn complex relationships between input feature maps and the final output.

$$f(x) = \max(0, x) \text{ --------- ReLU activation function}$$

Here x is the input food image in the Convolutional Neural Network for an activation function.
For the optimization in this system, Stochastic Gradient Descent (SGD) is used in machine learning to minimize a loss function in the training of Convolutional Neural Network Inception V3. SGD updates the model parameters such as weights to reduce the error measured by loss function.

G. Testing Phase

Training for the dataset with Inception V3 model is used with Convolutional Neural Network (CNN) to classify the food images. It uses the Kaggle dataset Indian Food Images Top 20 to recognize and fed to the model for further classification and recognition using the Inception V3 algorithm with the deep learning features of Convolutional Neural Networks [10]. The total of different 20 food classes items are available for training the model.

IV. Experimental Results

The system is developed in Python using TensorFlow for developing the food image recognition system. The system built is developed in Windows 10 using Visual Studio Code. The dataset used for the food recognition system used is Kaggle Indian Top 20 food images, shown in Fig. 4 as the food image dataset.

Figure. 4. Food Image Dataset

Figure. 5. Input Image Chapati

In the Fig. 5 Input image Chapati is shown to recognize the food image using deep learning algorithm. The image size taken is 299 x 299 for resolution from the dataset for training and testing purpose while also reducing the losses [11]. Now the input image is fed to the system which uses the Convolutional Neural Network Inception V3 algorithm to recognize the food image, which gives the predicted output. The classes used for 20 different top 20 Indian Food Images. The batch size is set to 32 in the food image recognition in the model, so that it processes 32 images at a time. As the training progresses, the model adjusts its parameters and learns to make more accurate predictions, leading to a reduction in loss. It uses Anaconda Prompt for displaying the command line in Python for the predicting the food image label [19]. In this system, 20 epochs are run for food image recognition to balance the training time with model performance for the convolutional neural network model. As it is a larger dataset, it required fewer epochs to converge, as there is more data to learn from the huge dataset. The image size takes more time for the epochs to process effectively as it contains higher resolution. Global average pooling in CNNs averages each feature map to single value reducing spatial dimensions of the output. Average pooling is a down sampling operation in deep learning it reduces the spatial dimension of a feature map. For the input image we consider 299 X 299 size image. It takes the average of the feature maps by average pooling function on the image dataset.
Here the top layers of the Inception V3 model are fine-tuned, they are replaced or retrained with fully connected layers designed only for a specific food image dataset. Data augmentation is performed on the training data to prevent overfitting. It increases the size and diversity of the dataset by creating modified copies of the existing data. This makes the model more robust and increases its performance. It generalizes to new and unseen data also reduces overfitting. The model architecture for classification and recognition of the images includes convolutional and pooling layers. The last layer is completely connected which represents the output classification of the network [13]. With their ability to recognize individual training set patterns that are prone to overfitting in deep neural networks, here comes regularization. It generalizes the data which it does not recognize by keeping the learning model simple. In this model, the dropout function is used for regularization, it

decreases the over-fitting in the model. A fully connected layer, known as dense layer in deep neural network where every neuron is connected to every neuron in the previous layer.

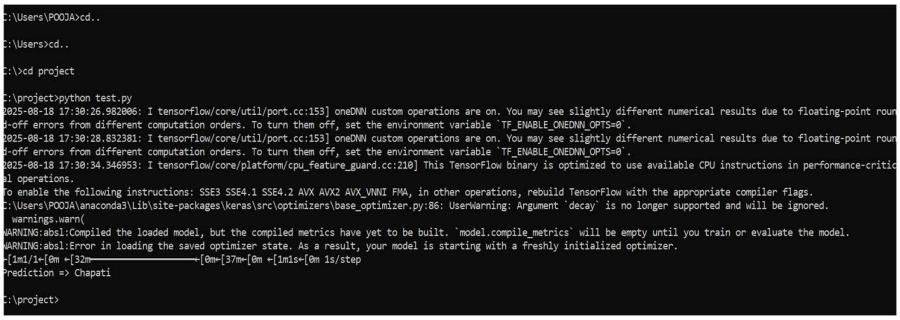

Figure. 6. Predicted Food Image as Chapati

In Fig. 6 Predicted Food image as Chapati is recognized by the Inception V3 Convolutional Neural Network algorithm.

Figure. 7. Input Image Fried Rice

In the Fig. 7 Input image Fried Rice is shown to recognize the food image using deep learning algorithm.

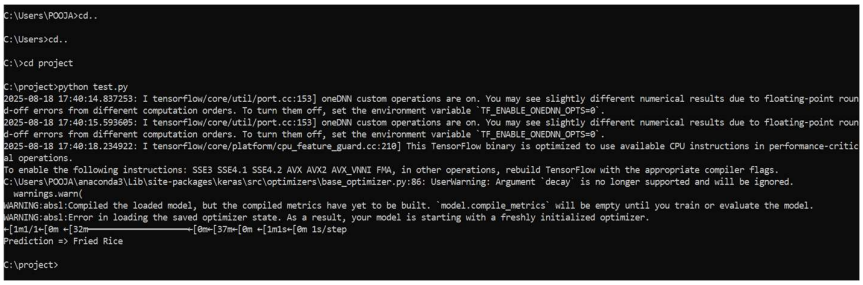

Figure. 8. Predicted Food Image as Fried Rice

In Fig. 8 Predicted Food image as Fried Rice is recognized by the Inception V3 Convolutional Neural Network algorithm.

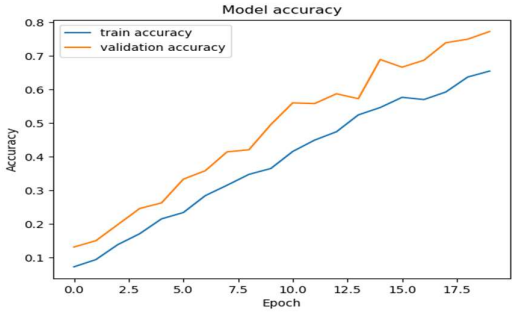

Figure. 9. Accuracy of the Proposed Model

Fig. 9. shows the Accuracy of the Proposed Model using Inception V3 and Convolutional Neural Network (CNN). In this system the accuracy of the model with Inception V3 using Convolutional Neural Network (CNN) gives 85% with a superior performance in predicting the food image in the proposed model [15].
The model shows high prediction performance of the food image recognition using Inception V3 on the ImageNet dataset, with the number of training epochs as 20 for more computational efficiency, also the Kaggle Dataset used for Top 20 Indian food images.
In this method when compared to other neural network algorithm like the traditional ones, it is seen that the Inception V3 with Convolutional Neural Network gives better results with higher accuracy and speed for Food Image Recognition using the Kaggle Image dataset in TensorFlow. The model loss in categorical cross-entropy loss is to quantify the predicted class probabilities and its true labels for food images. The Inception V3 model is more efficient and robust as it requires less training data than building a CNN from scratch.

V. CONCLUSION

Convolutional Neural Network works better with Inception V3 for the food image recognition model. It gives higher accuracy and efficiency when the large dataset is used to train the model. The model gives 85% accuracy for the food image recognition at the training phase we run epochs to get maximum accuracy and remove the model loss which is trained over maximum Kaggle Indian Food Image dataset at the top of ImageNet weights, which increases the model capability. Inception V3 with CNN makes the image recognition of food images easier and faster by extracting the features from the dataset images [17]. The proposed model works excellently with the dataset using the Inception V3 with CNN to sustain the model accuracy for a large number of classes for image classification and identification of the food image. The model gives more robust approach for the Inception V3 using Convolutional Neural Network architecture requiring less training data and leveraging high accuracy.

FUTURE SCOPE

Food Image recognition using Convolutional Neural Network and Inception V3 algorithm future work in this model involves developing personalized nutrition plans based on individual health needs and preferences. Advancements in real-time mobile applications for accurate calorie and nutritional tracking, integration with social media for the food supply chain in personalized recommendations [18]. Categorize users based on food preferences from their shared images for more efficient and personalized marketing.

REFERENCES

[1] Asish Bera, Z. W. (2021). Attend and Guide (AG-Net): A Keypoints-Driven Attention-Based Deep Network for Image Recognition. IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 30, 2021, 14.
[2] Bin Zhu, C. W. (2022). Learning From Web Recipe Image Pairs for Food Recognition. IEEE Transactions on Multimedia, 24, 11.
[3] Chaitanya A, J. S. (2023). Food Image Classification and Data Extraction Using Convolutional Neural Network and Web Crawlers. Procedia Computer Science, 10.
[4] Chang Liu, Y. C. (2018). A New Deep Learning-Based Food Recognition System for Dietary Assessment on An Edge Computing Service Infrastructure. IEEE TRANSACTIONS ON SERVICES COMPUTING, VOL.11, NO.2, MARCH/APRIL 2018, 13.
[5] Chengxu Liu, Y. L. (2021). Food and Ingredient Joint Learning for Fine-Grained Recognition. IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 31, NO. 6, JUNE 2021, 14.
[6] Fotios S. Konstantakopoulos, E. I. (2024). A Review of Image-Based Food Recognition and Volume Estimation Artificial Intelligence Systems. IEEE REVIEWS IN BIOMEDICAL ENGINEERING, VOL. 17, 2024, 17.
[7] Frank Po Wen Lo, Y. S. (2020). Image Based Food Classification and Volume Estimation for Dietary Assessment. IEEE Journal of Biomedical and Health Informatics, 24, 14.
[8] Gianluigi Ciocca, P. N. (2017). Food Recognition: A New Dataset, Experiments, and Results. IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 21, NO. 3, MAY 2017, 11.
[9] Haozan Liang, G. W. (2021). MVANet: Multi-Task Guided Multi-View Attention Network for Chinese Food Recognition. IEEE Transactions on Multimedia, 11.
[10] Jingjing Chen, B. Z.-W.-S.-G. (2021). A Study of Multi-Task and Region-Wise Deep Learning for Food Ingredient Recognition. IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 30, 2021, 13.
[11] Jixiang Gao, J. C.-G. (2023). Dynamic MIxup for Multi-Label Long Tailed Food Ingredient Recognition. IEEE Transactions on Multimedia , 25, 10.

[12] LANDU JIANG, B. Q. (2020). DeepFood: Food Image Analysis and Dietary Assessment via Deep Model. SPECIAL SECTION ON DATA-ENABLED INTELLIGENCE FOR DIGITAL HEALTH IEEE Access, 13.
[13] Luciano Oliveira, V. C. (2014). A mobile, lightweight, poll-based food identification system. Pattern Recognition 47, 12.
[14] Mengyang Zhang, G. T. (2023). Sequential Learning for Ingredient Recognition From Images. IEEE Transactions on Circuits and Systems for Video Technology, 14.
[15] MINKI CHUN, H. H. (2022). Development of Korean Food Image Classification Model Using Public Food Image Dataset and Deep Learning Methods. IEEE Access, 10.
[16] Mohammed Ahmed Subhi, S. H. (2019). Vision Based Approaches for Automatic Food Recognition and Dietary Assessment. IEEE Access, 12.
[17] Paritosh Pandey, A. D. (Dec 2017). FoodNet: Recognizing Foods Using Ensemble of Deep Networks. IEEE Signal Processing Letters, 24.
[18] Shota Horiguchi, S. A. (2018). Personalized Classifier for Food Image Recognition. IEEE Transactions on Multimedia, 13.
[19] Shuo Feng, Y. W. (2023). A fine-grained recognition technique for identifying Chinese food images. Heliyon 9(31 October 2023), 12.
[20] Stanley Giovany, A. P. (2017). Machine Learning and SIFT Approach for Indonesian Food Image Recognition. Computer Science and Computational Intelligence(14 October 2017), 9.
[21] Wandong Zhang, J. W. (2020). Wi-HSNN: A subnetwork based encoding structure for dimension reduction and food classification via harnessing multi-CNN model high-level features. Neurocomputing 414(14 July 2020), 10.
[22] Weiqing Min, Z. W. (2023). Large Scale Visual Food Recognition. IEEE Transactions on Pattern Analysis, 45, 18.