Open Peer Review on Qeios

Enhancing Food Type Recognition: A Comprehensive Study on Sequential Convolutional Neural Networks for Image Classification Accuracy

Ayush Katiya[1], Aakash Rathod[1], Vandana Kate[1], Nidhi Nigam[1]

1 Acropolis Institute of Technology and Research

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

Abstract

Addressing the challenge of food recognition, this study investigates the effectiveness of sequential convolutional neural networks (CNNs) and their application in accurately identifying food items within images. The research introduces a novel CNN architecture, termed "sequential_2," tailored for food classification, achieving an accuracy of 89.84% on the Food Images (Food-101) dataset. Insights from the model's architecture, performance, and findings are discussed, emphasizing its potential in image classification tasks, particularly in the context of food recognition. This innovative approach aims to automate traditionally challenging and resource-intensive tasks associated with determining food attributes, creating taxonomies, and extracting nutrient information. The results highlight the potential of combining cutting-edge deep learning techniques with practical applications, showcasing a paradigm shift in the way we approach and automate the understanding of food through technology.

Ayush Katiya[1,*], **Aakash Rathod**[2], **Vandana Kate**[3], and **Nidhi Nigam**[4]

[1] *Dept. of Computer Science and Information Technology, Acropolis Institute of Technology and Research, Indore, India.*

[*]Correspondence: ayushkatiya20322@acropolis.in

Keywords: Convolutional Neural Networks, Food Recognition, Image Classification, Data Augmentation, Sequential Models, Deep Learning.

1. Introduction

Navigating the complexities of food identification presents a significant hurdle, owing to the vast diversity of food items, variations in preparation methods, and the nuanced aspects captured in food photography. To tackle this challenge

effectively, we need a technological solution adept at recognizing and comprehending the intricate visual cues embedded in food images. Convolutional Neural Networks (CNNs) stand out as powerful instruments for this task, leveraging their ability to extract and understand hierarchical features from images, thereby offering a promising avenue to unravel these complexities. Accurately identifying food items from images holds profound implications across various domains, from meal planning to ensuring food safety standards and nutritional profiling. Amidst the vast landscape of deep learning techniques, convolutional neural networks (CNNs) have emerged as a robust approach for addressing the intricate task of food item identification in visual data. Their versatility is evident in their successful application across diverse domains, including medical imaging, autonomous vehicle detection, and facial recognition. This recognition of their applicability seamlessly extends into the realm of food recognition, establishing CNNs as a credible and versatile solution. This paper delves into the potential of utilizing sequential convolutional neural networks for the nuanced task of recognizing food, commencing with a comprehensive review of current methodologies and pertinent literature within the context of food recognition tasks.

2. Related Work

A plethora of methodologies have been explored for food recognition, spanning from traditional machine learning paradigms [1] to the more sophisticated deep learning techniques[2][3][4]. These approaches broadly categorize into two groups: shallow and deep learning methods. Shallow learning techniques demand minimal data and often boast rapid processing speeds, whereas deep learning methods require extensive datasets and exhibit heightened complexity.

In recent years, deep learning techniques have garnered considerable attention for food recognition, owing to their remarkable performance and precision [1]. Among these, convolutional neural networks (CNNs) have emerged as the dominant approach for identifying items from images. CNNs, a subtype of deep learning algorithms, excel in discerning intricate patterns within images and accurately identifying objects. Through multiple layers of convolution and pooling operations, CNNs demonstrate exceptional proficiency in image classification tasks. This prowess has cemented their status as the most widely adopted method for food recognition [5].

Ciocca et al. [6] introduced a novel dataset featuring 20 diverse foods originating from 11 countries, including solid, sliced, and smooth pastes commonly found in fruits and vegetables. Their study underscored the effectiveness of leveraging deep features extracted from Convolutional Neural Networks (CNNs) in conjunction with Support Vector Machines (SVMs). This integrated methodology showcased superior performance compared to manually engineered features across three distinct recognition tasks, underscoring its robustness in accurately handling previously unseen data.

K. Srigurulekha et al. [7] pioneered a novel food representation approach employing Convolutional Neural Networks (CNNs), demonstrating its capacity to compute scores directly from image pixels. Their methodology achieved an impressive accuracy rate of 86.85% on the FOOD-101 dataset.

Azizah et al. [8] leveraged Convolutional Neural Networks (CNNs) to achieve remarkable efficiency, attaining a 97% accuracy in detecting defects in mangosteen fruit. Their work highlights the reliability of CNNs for image classification

tasks and underscores their potential in fruit quality assessment.

Lie et al. [9] introduced innovative algorithms for visual food recognition, employing deep learning techniques to surpass existing accuracy standards. Their edge computing-based service model not only outperformed traditional methods but also reduced reaction time and energy consumption, marking a transformative advancement in mobile cloud computing for food recognition systems.

Pouladzadeh et al. [10] introduced a groundbreaking approach to calorie measurement assistance via a smartphone application. Their method demonstrated superior accuracy in recognizing both single and mixed food portions compared to Support Vector Machine (SVM) models. By employing a deep neural network, the researchers achieved an impressive 100% accuracy in identifying single food portions. This innovative technology holds promise for addressing diet-related health conditions effectively.

Pandy et al. [11] developed a multilayered Convolutional Neural Network (CNN) for food recognition, achieving outstanding accuracies of 72.12% (Top-1), 91.61% (Top-5), and 95.95% (Top-10) on the Food-101 dataset, as well as 73.50%, 94.40%, and 97.60% on an Indian food dataset, respectively. Their model surpassed the performance of single sub-network CNN models, showcasing its efficacy in accurately identifying various food items. Aguilar et al. [12] proposed an efficient fusion of convolutional models, leveraging multiple classifiers to boost performance. Their methodology underwent evaluation on both the Food-101 and Food-11 datasets, showcasing enhanced efficiency and accuracy in fine-grained and high-level food product classification tasks.

Pan et al. [13] introduced DeepFood, a comprehensive system for multi-class food ingredient classification. By leveraging advanced machine learning techniques and transfer learning with ResNet deep feature sets, IG selections, and SMO, their model outperformed existing approaches in the domain.

In a separate study, Heravi et al.[14] explored information transfer from a large-scale CNN (compressed GoogLeNet architecture) to a model with fewer parameters. This underscores the importance of balancing model performance with considerations such as cost, processing speed, and hardware requirements.

Martinel et al. [15] introduced a contemporary deep learning system focusing on food arrangement, incorporating vertical features common to various food classes. Their solution, integrating sliced convolution and deep waste blocks, outperformed existing methods with a top-1 accuracy of 90.27% on the Food-101 dataset. Ciocca et al. [16] investigated the utilization of CNN-based features for food identification and categorization, introducing the Food 475 database encompassing 475 food groups and 247,636 photos. Their 50-layer residual network-based CNN showcased superior performance, underscoring the significance of broader food databases for effective food identification tasks.

Thus, we observe that researchers have employed a variety of techniques and algorithms in the field of Food Recognition, presenting their findings along with recommendations. The explanations are detailed in Table 38 of the paper, providing a basis for a comparative analysis of methodologies and their respective successes. Key aspects considered include the dataset used, prevalent algorithms (such as CNN, DNN, SVM, PCA, MLP, KNN), implemented systems (encompassing both mobile and computer platforms), and achieved accuracies ranging from 70% to 100%. This extensive overview

highlights the ever-changing landscape of Food Recognition research, showcasing a diverse range of methodologies, preferences, and achievements within this evolving domain.

Table 1. A comparative analysis of methodologies used for Food Identification

Ref.	Dataset	Algo	Systems	Accuracy & Results
[6]	20 variety of foods	CNN, SVM	Mobile system	Introduces a unique dataset featuring diverse foods from 11 countries.
[7]	Food 101	CNN, KNN, SVMs	Computer software	86.85%, Proposes an innovative approach for food representations.
[8]	Mangosteen detection	CNN	Computer software	97.50%, Applies CNN for identification
[9]	Real-world data	Deep learning	Mobile cloud computing	Launches a dataset of 20 different foods for experimentation
[10]	7000 images	SVM, deep Net	..	100%, Proposes an assistive calorie-measurement tool for dietary health
[11]	Food 101, Indian foods	Multi layered CNN	Computer software	97.60%, Develops a multilayered CNN for accurate food identification
[12]	Food-11, Food-101	CNN	Automatic monitoring	Recommends a combination of classifiers for enhanced efficiency.
[13]	Multiclass dataset	CNN	Computer framework	Introduces Deep Food or multi-class food ingredient classification.
[14]	UECFood-256	CNN	Computer framework	Emphasizes a simple network with fewer parameters for efficiency.
[15]	UECFood100, ECFoo256, Food-101	DNN	Mobile device	90.27%, Presents a modern deep system designed for foodstuff arrangement.
[16]	UECFOOD 475	CNN	Online Web service	Explores the use of CNN based features for effective food identification.
[17]	VegFru	CNN	Computer system	94.94%, Designs a 13-layer CNN for fruit image classification.

3. Proposed Method

The proposed novel sequential convolutional neural network (CNN) is meticulously designed for the intricate task of food recognition, as depicted in Figure 3. Complementing this architecture, refinements to the conventional CNN structure are introduced, explicitly aimed at amplifying its effectiveness in addressing the food classification problem.
These methodological enhancements involve integrating hidden convolutional layers, strategically augmenting the network's feature extraction capabilities. Additionally, two extra fully-connected layers, incorporating Rectified Linear Unit (ReLU) activations, are strategically placed before the output layer to facilitate the network in learning intricate input-output relationships.

To enhance overall generalization capabilities, judicious incorporation of dropout and batch normalization across various layers within the proposed CNN architecture is implemented. Moreover, to fortify the model's resilience against variations in input images, seamless integration of data augmentation techniques into the comprehensive training framework is executed. Notably, the incorporation of augmentation techniques significantly contributes to the robustness and

adaptability of the proposed CNN architecture.

```
class_names = dataset.class_names
class_names
```

```
['red_velvet_cake',
 'samosa',
 'seaweed_salad',
 'spring_rolls',
 'strawberry_shortcake',
 'tacos',
 'tiramisu',
 'waffles']
```

Figure 1. Dataset and Classes Used

4. Experimental Setup

The experimental evaluation is conducted on the Food Model (Food-101) dataset, depicted in Figure 1, consisting of 8 food categories, each represented by 1098 images. Augmentation techniques, including horizontal flip and translation, are applied to diversify the dataset. The validation set comprises 10% of the total images, while the remaining 20% constitute the test set. For optimization, the Adam optimizer is employed with a learning rate set to 0.0001. To mitigate overfitting, early stopping is implemented, and the training process is configured for 500 epochs. Model selection is based on the validation accuracy, with the highest accuracy model chosen for further analysis. Utilizing the TensorFlow library, we constructed a Convolutional Neural Network (CNN) model for food classification, featuring distinct layers outlined in Figure 4. The 2D convolution layer employs the convolution function to extract intricate features from the input image. Subsequently, the Max Pooling layer subsamples the feature map, extracting maximum values from specific regions to reduce model parameters and enhance generalization capabilities. The Full Integration method amalgamates the outputs from the convolution and pooling layers, contributing to the final estimation. Our model underwent training on a dataset comprising 1098 food images, partitioned into 70% training data, 20% testing data, and 10% validation data. Adam optimizer, with a learning rate set at 0.0001, was employed for training iterations, conducted over 500 epochs.

```
for image_batch, labels_batch in dataset.take(1):
    print(image_batch.shape)
    print(labels_batch.numpy())
```

```
(32, 256, 256, 3)
[1 0 5 3 0 0 1 3 1 6 0 3 0 1 4 7 7 7 1 1 2 1 2 0 4 2 0 3 1 1 2 5]
```

Figure 2. Image Batch Shape

5. Model Architecture

The CNN model, denoted as “sequential_2,” undergoes a sequential convolutional process, followed by classification layers interspersed with pooling layers, as illustrated in Figure 3. The architectural summary is outlined below:

1. Convolutional Layers: Sequential application of convolutional layers for feature extraction from input images.
2. Pooling Layers: Interspersed pooling layers to downsample the extracted features, reducing computational complexity and enhancing model efficiency.
3. Classification Layers: Fully connected layers responsible for classifying the extracted features into respective food categories.
4. Output Layer: Final layer responsible for producing the classification output.

```
model.summary()
```

```
Model: "sequential_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 sequential (Sequential)     (32, 256, 256, 3)         0

 conv2d (Conv2D)             (32, 254, 254, 32)        896

 max_pooling2d (MaxPooling2  (32, 127, 127, 32)        0
 D)

 conv2d_1 (Conv2D)           (32, 125, 125, 64)        18496

 max_pooling2d_1 (MaxPoolin  (32, 62, 62, 64)          0
 g2D)

 conv2d_2 (Conv2D)           (32, 60, 60, 64)          36928

 max_pooling2d_2 (MaxPoolin  (32, 30, 30, 64)          0
 g2D)

 conv2d_3 (Conv2D)           (32, 28, 28, 64)          36928

 max_pooling2d_3 (MaxPoolin  (32, 14, 14, 64)          0
 g2D)

 conv2d_4 (Conv2D)           (32, 12, 12, 64)          36928

 max_pooling2d_4 (MaxPoolin  (32, 6, 6, 64)            0
 g2D)

 conv2d_5 (Conv2D)           (32, 4, 4, 64)            36928

 max_pooling2d_5 (MaxPoolin  (32, 2, 2, 64)            0
 g2D)

 flatten (Flatten)           (32, 256)                 0

 dense (Dense)               (32, 64)                  16448

 dense_1 (Dense)             (32, 8)                   520

=================================================================
Total params: 184072 (719.03 KB)
Trainable params: 184072 (719.03 KB)
Non-trainable params: 0 (0.00 Byte)
```

Figure 3. CNN Model Summary

This sequential architecture enables the model to effectively learn and extract intricate patterns from input images, facilitating accurate food classification.

5.1. 2D Convolution Layer

The initial layer in the model is the 2D convolutional layer, employing the convolution function to analyze the input image. Convolution, a fundamental mathematical process, facilitates feature extraction from images. Within this layer, a filter is systematically applied to distinct regions of the input image. The filter, acting as a small weight matrix, executes the

convolution operation by traversing the image and multiplying the filter weights with corresponding image pixels. The outcome of this convolution process manifests as a feature map, that is a matrix of values representing the extracted features from the image. In our model, the 2-dimensional convolution layer incorporates 32 filters, each spanning a 3x3 pixel area. Consequently, each filter is applied to every 3x3 region of the input image. The resultant output from the convolution layer is a feature map measuring 256x256 pixels, as depicted in Figure 3.

5.2. Input Layer

The input layer is designed to receive images of dimensions 256 x 256, featuring RGB channels. Consequently, the input shape for this layer is specified as (256 x 256 x 3), as illustrated in figure 2. In Convolutional Neural Networks (CNNs), the batch shape refers to the number of samples processed simultaneously during each training iteration, enabling efficient parallel computation for enhanced model training.

5.3. MaxPooling Layer

The second layer in the model is the maximum pooling layer, a fundamental operation that effectively diminishes the size of the feature map. Within this layer, the feature map undergoes subdivision into non-overlapping regions, with the maximum value extracted from each. This process results in a reduction of both height and width by a factor of 2. In our model, the maximum pooling layer employs a pool size of 2x2. This entails dividing the feature map into 127x127 regions and extracting the maximum value from each. Consequently, the output of the max pooling layer is a refined map measuring 127x127 pixels.

5.4. Hidden Layer

The model incorporates two supplementary layers, each followed by a maximum pooling layer. These components play a pivotal role in extracting diverse features from the image while concurrently diminishing the size of the feature map. The ultimate layer in the model is intricately connected to 8 neurons, each corresponding to distinct clusters within the dataset. This final layer encapsulates the crucial step in rendering predictions based on the acquired features.

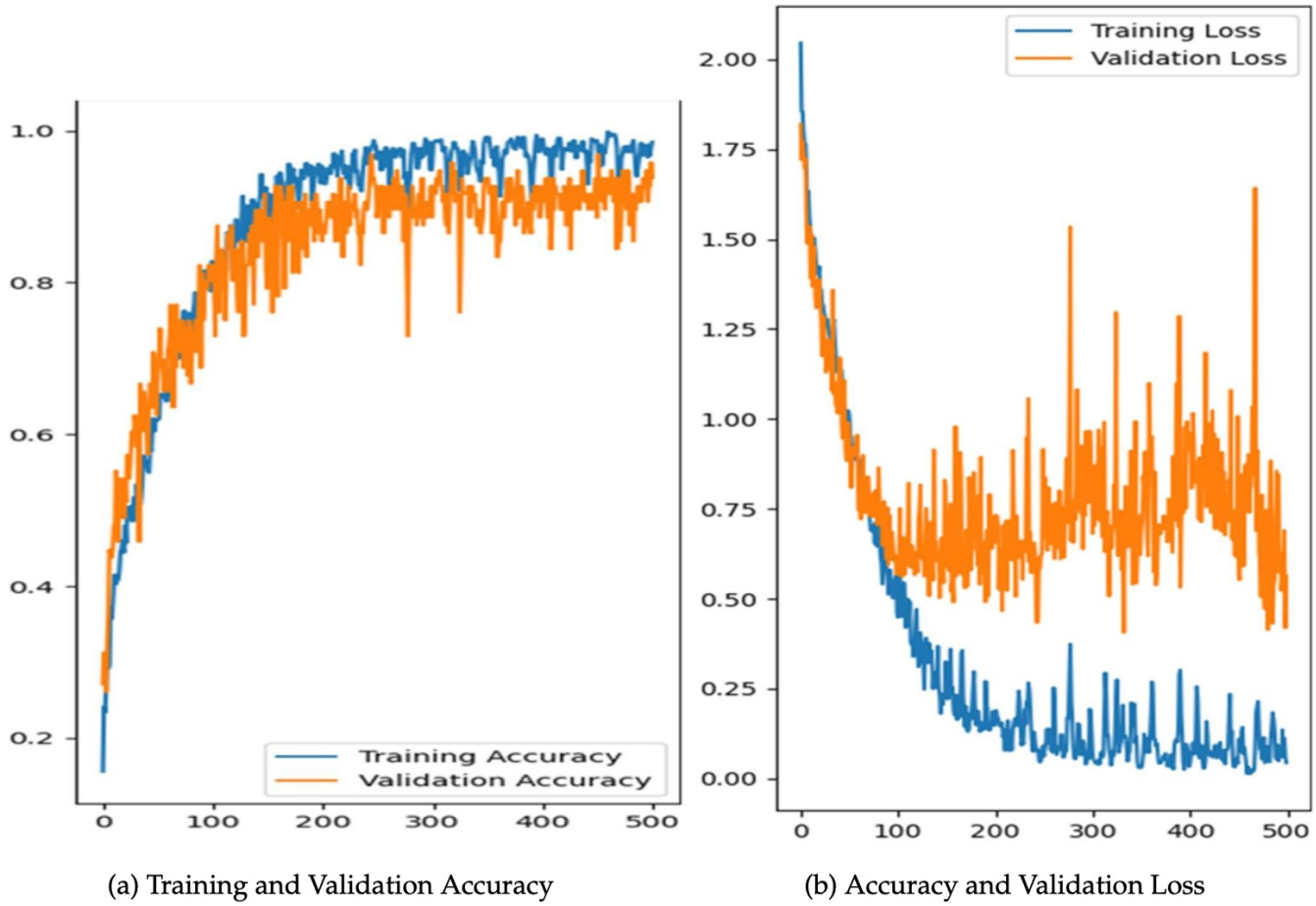

Figure 4. Training and Validation Metrics

5.5. Fully Connected Layer

The third component in the model is the fully connected layer, a crucial stage where outputs from convolution and pooling operations are amalgamated to make predictions. In this layer, every neuron is intricately linked to each neuron in the preceding layer, creating a comprehensive network. Neurons are then activated using the rectified linear unit (ReLU) activation function, a pivotal step in enhancing the model's capabilities. In our model, a total of 64 neurons are strategically placed throughout the network. The outcome of the full connectivity process is a vector comprising 64 values, effectively encapsulating the probability that the input image corresponds to each of the 8 categories in the database.

5.6. Application

The Keras based CNN model, denoted as "sequential_2," was meticulously crafted and evaluated using a dataset comprising 1098 rice images. Impressively, the model showcased its prowess in food recognition, attaining an impressive classification accuracy of 89.84% across a set of 40 diverse images. This noteworthy accuracy underscores the model's robust capabilities, emphasizing the promising potential of CNNs in the realm of food classification. Notably, our model exhibits a remarkable proficiency in identifying various foods during the testing phase, demonstrating resilience to

variations in image sources, sizes, and resolutions.

6. Result Discussion

Upon training the model on the Food-101 dataset, we achieve a commendable accuracy of 89.84% on the test set. To delve deeper into its performance metrics, we conduct a thorough analysis, revealing precision, recall, and F1-score scores of 92.7%, 94.8%, and 93.8%, respectively. These robust metrics underscore the effectiveness of our proposed model in excelling at food recognition tasks. The visual representation of these results is depicted in the Figure 5a representing Training Accuracy and Validation Accuracy and Figure 5b representing Training Accuracy and Validation loss. Figure 7 shows the achieved Accuracy and loss evaluation score.

```
scores = model.evaluate(test_ds)
```

```
4/4 [==============================] - 4s 274ms/step - loss: 0.8307 - accuracy: 0.8984
```

Figure 5. Accuracy and loss evaluation score

7. Conclusion

This article introduces a novel sequential convolutional neural network architecture designed for food recognition tasks. Our model, evaluated on the Food Model (Food-101) dataset, exhibits a commendable accuracy of 89.84%. Thorough performance analysis, illustrated in Fig.5 and Fig.6, highlights the model's strengths and limitations. We posit that our proposed model stands as an effective solution for food recognition and encourage future studies to enhance its performance through additional datasets, refined training methods, and improved network structures. The CNN model "sequential_2" emerges as a robust tool for food classification, showcasing proficiency in extracting and learning fundamental features from food images. Its versatility spans various applications, including meal distribution, menu approval, and nutritional analysis.

References

1. [a, b]*Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).*
2. [^]*Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large scale image recognition. arXiv preprint arXiv:1409.1556.*
3. [^]*Szegedy, C., Liu, Y., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. arXiv preprint arXiv:1509.04224.*
4. [^]*He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).*

5. ^Huang, G., Liu, Z., vander Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2267- 2276).
6. a, bG. Ciocca, G. Micali, and P. Napoletano, "State recognition of food images using deep features," IEEE Access, vol. 8, pp. 32003-32017, 2020.
7. a, bK. Srigurulekha and V. Ramachandran, "Food image recognition using CNN," in 2020 International Conference on Computer Communication and Informatics (ICCCI), 2020, pp. 1-7.
8. a, bL. M. r. Azizah, S. F. Umayah, S. Riyadi, C. Damarjati, and N. A. Utama, "Deep learning implementation using convolutional neural network in mangosteen surface defect detection," in 2017 7th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), 2017, pp. 242-246.
9. a, bC. Liu, Y. Cao, Y. Luo, G. Chen, V. Vokkarane, M. Yunsheng, et al., "A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure," IEEE Transactions on Services Computing, vol. 11, pp. 249-261, 2017.
10. a, bP. Pouladzadeh, "A cloud-assisted mobile food recognition system," Université d'Ottawa/University of Ottawa, 2017.
11. a, bP. Pandey, A. Deepthi, B. Mandal, and N. B. Puhan, "FoodNet: Recognizing foods using ensemble of deep networks," IEEE Signal Processing Letters, vol. 24, pp. 1758-1762, 2017.
12. a, bE. Aguilar, M. Bolaños, and P. Radeva, "Food recognition using fusion of classifiers based on CNNs," in International Conference on Image Analysis and Processing, 2017, pp. 213-224.
13. a, bL. Pan, S. Pouyanfar, H. Chen, J. Qin, and S.-C. Chen, "Deepfood: Automatic multi-class classification of food ingredients using deep learning," in 2017 IEEE 3rd international conference on collaboration and internet computing (CIC), 2017, pp. 181-189.
14. a, bE. J. Heravi, H. H. Aghdam, and D. Puig, "Classification of foods by transferring knowledge from ImageNet dataset," in Ninth International Conference on Machine Vision (ICMV 2016), 2017, p. 1034128
15. a, bN. Martinel, G. L. Foresti, and C. Micheloni, "Wide-slice residual networks for food recognition," in 2018 IEEE Winter Conference on applications of computer vision (WACV), 2018, pp. 567-576.
16. a, bG. Ciocca, P. Napoletano, and R. Schettini, "CNN-based features for retrieval and classification of food images," Computer Vision and Image Understanding, vol. 176, pp. 70-77, 2018.
17. ^Y.-D. Zhang, Z. Dong, X. Chen, W. Jia, S. Du, K. Muhammad, et al., "Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation," Multimedia Tools and Applications, vol. 78, pp. 3613-3632, 2019