I. J. Education and Management Engineering, **2024, 2, 53-60**
Published Online on April 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijeme.2024.02.05

Classification of Food Objects Using Deep Convolutional Neural Network Using Transfer Learning

Dipta Gomes*
University of Ulster, Belfast, Northern Ireland, UK
Email: diptagomes@gmail.com
ORCID ID: https://orcid.org/0000-0003-3019-6051
*Corresponding author

Received: 17 June, 2023; Revised: 04 August, 2023; Accepted: 17 September, 2023; Published: 08 April, 2024

Abstract: With the advancements of Deep Learning technologies, its application has broadened into the fields of food classification from image recognition using Convolutional Neural Network, since food ingredient classification is a very important aspect for eating habit recognition and also reducing food waste. This research is an addition to the previous research with a clear illustration for deep learning approaches and how to maximize the classification accuracy to get a more profound framework for food ingredient classification. A fine-tuned model based on the Xception Convolutional Neural Network model trained with transfer learning has been proposed with a promising accuracy of 95.20% which indicates a greater scope of accurately classifying food objects with Xception deep learning model. Higher rate of accuracy opens the door of further research of identifying various new types of food objects through a robust approach. The main contribution in the research is better fine-tuning features of food classification. The dataset used in this research is the Food-101 Dataset containing 101 classes of food object images in the dataset.

Index Terms: Food classification, Deep Learning, Convolutional Neural Network, Data Augmentation

1. Introduction

Classification of food had been a major challenge in the research field of Computer Vision and Object detection and automation of the ingredient detection had been the most awaited aspect of the field of Deep Learning. Applications such as Food Detection System, Calorie Measuring System and Diet Monitoring System had been important tools for everyday life for all levels of users around the world. Recent research on Image processing and detection of object from images had reached a breakthrough in different fields of research [1,2,3,4,5,6] like surveillance systems, medical imaging, remote sensing, emotion detection and object detection. Various research has shown machine learning techniques proved successful in detecting food objects from images successfully [1,7,8]. In most recent years, it was shown deep learning tools such as Convolutional Neural Networks (CNN) proved a very successful tools for high-efficiency detection model [9].

Object detection and classification from images is a challenging job due to the presence of low image quality, distortions, disturbances, intensity of light and small image dataset size. Comparisons among various Underwater Object detection from images has been represented by Dipta et al. [10], where various deep learning methods and conventional methods are compared, and a detailed evaluation of methods had been illustrated. Beside all these obstacles, food normally gets deformed and the shape and appearance of food changes with time as classification of food had always been a hard task for researchers and food cooked in different recipes give the images high variations and forms. In every food object detection system, it is required to be fast as well as accurate and needs to take consideration of the color, texture and quality of food in an image. A Real Time Sign Language Detection model had been put forward by P. P. Urmee et al. [11] where the Xception Model had been used on top of a proposed Bangla Sign Language dataset called BdSLInfinite. A convolutional neural network model (CNN) model had been proposed on top of Xception Model by P. P. Urmee et al. [11] to detect real time Bangla Sign Language. Here in the research work a dataset called BdSLInfinite dataset had been proposed for Bangla Sign Language for people with hearing difficulties. The model achieves an accuracy of 98.93% and a average response time of 48.53 minutes.

The dataset to be used for training is the Food101 dataset containing 101 food categories with over 100,000 images. Due to this high volume of images and classes of food, this dataset proved to be an excellent choice for this research.

The training images for the dataset requires some pre-processing thus providing extra scope of improvement in the research process.

Through the research, a highly efficient transfer learning model has been proposed which has been trained using Food-101 dataset containing a huge variety of food images of various classes of food objects. Image classification has been provided using the Convolutional Neural Network deep learning model due to its high efficiency in learning with complex identification features. The parameters of the model are then fine-tuned from the pre-trained model which helps the model to be more robust in nature. The research work is mainly focused on increasing accuracy rate of existing food object detection models and introduces a more robust approach to address the problem of extracting food detection features and further improve the research scope of food object detection.

2. Literature Review

Previous comparison of deep learning and conventional method for food classification has been put forward by Sefer Memis et. al [9]. The UEC FOOD-101 dataset was used for evaluation in this research among the models ResNet-18 [12], Inception-V3[13], ResNet-50[14], DenseNet-121[15], Wide ResNet-50 [16] and ResNet-50 [17]. It was found the performance of Resnet-50 without Mixed Precision (MP) has shown the best result with a classification accuracy of 87.7%. All the above models had been the building blocks of concurrent research approaches.

DeepFood multi-class food classification framework had been put forward by Lili Pan et al. [18] where on top of a pre-trained Convolutional Neural Network, trained with the large ImageNet Dataset, a set of transfer learning algorithms had been used to find deep learning features. The dataset used is the Mealcome MLC-41 Dataset. The classification was carried out first through training by Sequential Minimal Optimization (SMO). Here, the ResNet using SMO provided a top1 accuracy of 87.781%.

Z. Zong et al. in their paper [19] has introduced a food image classification method using local texture patterns through using Scale Invariant Feature Transformation (SIFT) with Local Binary Pattern (LBP) feature. The proposed method is then evaluated using the Pittsburgh Fast-Food Image using the Support Vector Machine classifier. The proposed method shoed an accuracy of 67%.

Research had been carried out on top of UEC-FOOD100 dataset by K. Yanai et al. [20] where a fine-tuned Deep Convolutional Neural Network (DCNN) CaffeNet has been used which was pre-trained using the large ImageNet dataset with a accuracy of 78.77%. The model was pre-trained with 2000 categories from the ImageNet dataset.

Table 1. Deep Learning models and results with different datasets

Model	Features	Dataset	Accuracy
Ensemble Network Architecture on top of AlexNet, GroupNet, ResNet [21]	Features from other CNN models uses Ensemble Network Architecture	Food101 Dataset	95.95%
Random Forest [26]	Random forest features	Food 101 Dataset	50.76%
DeepFood based on GoogleNet [27]	Top 5% Accuracy	Food 101 Dataset	93.70%
Fine-tuned Deep CNN CaffeeNet [20]	ColorFV features RootHoG features DCNN features DCNN-Food features Top 5% Accuracy	UEC-Food100	94.85%
		UEC-Food256	88.97%
ResNet-18 [9]	ResNet-18 11.7M features	UEC Food 100	84.4%
Inception V3 [9]	Inception V3 26.7M features		84.90%
Resnet-50 [9]	Resnet-50 25.8M features		86.5%
DenseNet-121 [9]	DenseNet-121 7.1 M features Top 1% accuracy		87.1%
AlexNet [18]	F1 Measure	MLC-41 Dataset	80.41%
CaffeeNet [18]			80.75%
ResNet-50 [18]			87.78%
SIFT Detector [19]	Colour Histogram SIFT Features Accuracy	Pittsburgh Fast-Food Image Dataset	82%
Submodular Optimization Method [22]	Accuracy	FoodDD Dataset	94.11%
ResNet-50 [23]	Top 5% Accuracy	Food 475 Dataset	95.5%

Pandey et al. [21] proposed a multi-layered CNN model for automated food recognition model on top of AlexNet with Food101 dataset and a Indian dataset obtaining a Top1% of 72.12% Top-5 % of 91.61 %, and an accuracy rate of 95.95%. Here the author used a CNN model that uses features from other CNN models and then preprocess images that are of importance. A Ensemble Network Architecture has been built on top of fine-tuned AlexNet, GoogleNet and ResNet which is concatenated and later activated using SoftMax function to fund the scores of the features during training of the data.

Pouladzadeh et al. in [22] developed a mobile application platform to detect food constituents using Convolutional Neural Network (CNN) for training. At first the region of interest has been deduced using regional proposal algorithm, then using CNN to find the features of the regions. The positive are of interest is then obtained using their researched submodular optimization method. The dataset here used is the FoodDD Dataset with a average recall rate of 90.98%, precision rate of 93.05% and an accuracy of 94.11%.

Ciocca et al. in [23] has proposed a food dataset Food 475 dataset that has been developed from Food 524 dataset where each of the image is described as a feature in CNN model. A food detection model using a 50 layered architectural network CNN RestNet-50 model has been proposed in this research work. The research deduces an Top 5% accuracy of 95.5%.

Salim et al. [24] had provided a review of several researches on food detection on deep learning using Convolutional Neural Networks (CNN) had been successful in detecting one-to-three-dimensional data for multiple formats of images.

V. H. Reddy et al. [25] had put forward a calorie measuring system from uploaded food images, showing the constituent food calorie. A dataset had been developed with 20 classes with 50 images in each class. A specific Convolutional Neural Network with 6 layers had been proposed with a detection accuracy of 78.7% and a training accuracy of 93.29%.

S. Memiş et al. [9] carried out a analytical comparison on various deep learning models on UEC-100 food dataset. The models they compared are ResNet-18, Inception-V3, Resnet-50, Densenet-121, Wide Resnet-50 and ResNext-50 with image size set as 320x320 and 299x299. In order to cope up a finite dataset like UEC-100, they took the transfer learning approach with a promising result of 87.7% accuracy with ResNet-50 model.

Bossard et al. in [26] introduced the Food-101 dataset to automate food image classification to detect object of interest using random forest showing a conventional method of food object detection with a accuracy rate of 50.76%.

Chang Liu et al. [27] proposed a method to calculate calorie intake in a dietary plan through a deep learning model based on Convolutional Neural Network known as Deep Food. The research had been carried out on top of Food 101 dataset and UEC-256 Dataset. For the UEC-256 Dataset, for 72000 iterations, the model gave the highest Top-5% Accuracy of 81.5%. For Food-101 dataset, for 250,00 iterations the highest accuracy was 93.7%.

3. Research Methodology

In this research we focus on a transfer learning based Convolutional Neural Network Xception Model F. Chollet et al. [28] which is trained on large Food 101 dataset and to detect food objects from images.

A. Dataset and platform of implementation

In the light of the research statement, the dataset selected is the Food101 Dataset containing a sum of 101 classes of food of various variations and ranges. In the paper by Bossard et al. [26], the Food 101 Dataset was first used to classify food objects from images using a conventional data mining method of Random Forest. The dataset consists of 750 training and 250 testing images per class, summing up to a total of 101, 000 number of images. The images selected were the top 101 categories of food and pre-processed in the form of correction of high intensity colors and wrong labelling in the training set [28]. A section of the dataset is illustrated below using the Keras Framework by Tensorflow on Google Colab in Figure 1.

Fig. 1. Sample of First Nine classes in Food-101 Dataset

Due to the constraint of high requirement of resources to train the Xception CNN Model with the large Food-101 Dataset, the model was trained on the Google Colab platform and the runtime was powered by a NVIDIA V100 GPU Hardware Accelerator for training the model.

B. Proposed Methodology

The proposed methodology comprises of the selection of the Food 101 Dataset and pre-processing with setting the image size to (299,299,3). The dataset was downloaded from the http://data.vision.ee.ethz.ch/cvl/food-101.tar.gz using the Keras get_file() function.The dataset was split into test and train set with a ratio of 75,7500 and 25,250 respectively from the food-101 meta file containing the labels for test and training set respectively.

Data Augmentation is a process to extend a dataset into a larger volume, which had been carried out on top of the large Food 101 dataset. Firstly, the image had been rescaled with 1. / 255. The rotation range had been set to 0.2 with width shift rage of 0.2 and height shift range of 0.2. The shear range of 0.2 has been used and zoom range of 0.2 had been selected. The images have also been horizontal flipped. The Augmentation in the dataset makes the dataset more robust in nature and bulking the dataset for more accurate detection of images. A section of the augmented data has been illustrated in Figure 2.

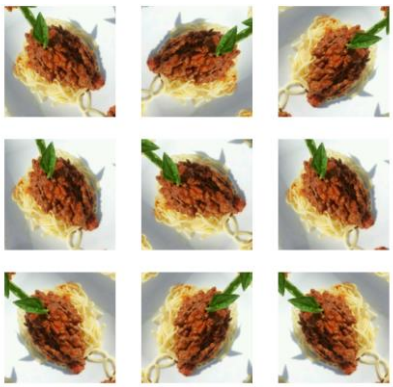

Fig. 2. Augmented form of the dataset

The pre-trained model selected is the famous Deep Learning Xception Model [11] by F. Chollet due to its high Top 5% Accuracy rate for Food object classification. The Xception Model is based on a depth wise separable convolutional network which is also known as Extreme Inception model. The Xception Model in [11] trained with the large ImageNet dataset showed a high Top 5% accuracy rate of 94.5 % compared against VGG-16 of 90.10 %, ResNet-152 of 93.3 % and Inception V3 of 94.10%. The Top 1% accuracy of Xception Model also proved a very promising result of 79 % which is higher than all the other models. In this research, the proposed model is transfer learned using this specific model to overcome the difficulty in reaching the 90% top 5% accuracy benchmark for the Food-101 Dataset.

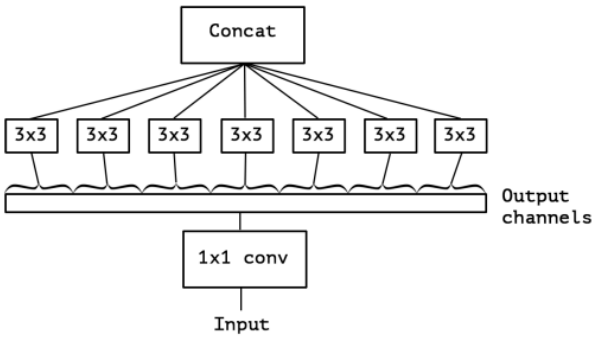

Fig. 3. Xception Model in [11]

The dataset is first pre-processed using Data Augmentation and setting the image size to (299,299,3). The Xception model is first initiated using the Keras built-in function setting the activation function as ‘relu’. The Xception model is also known as Xtreme Inception Model. At first the fully connected Xception model is initiated. A Global average pooling layer is initiated with input from the output from Xception model feature selection. A Dense layer of 128 units is then added with activation function set to ‘relu’, since the consumption of resources for the project was crucial a higher number of units had not been selected. The dropout layer is then added to the model with a dropout rate of 40%. All the layers of the pre-trained model is then frozen to stop learning from all the layers.

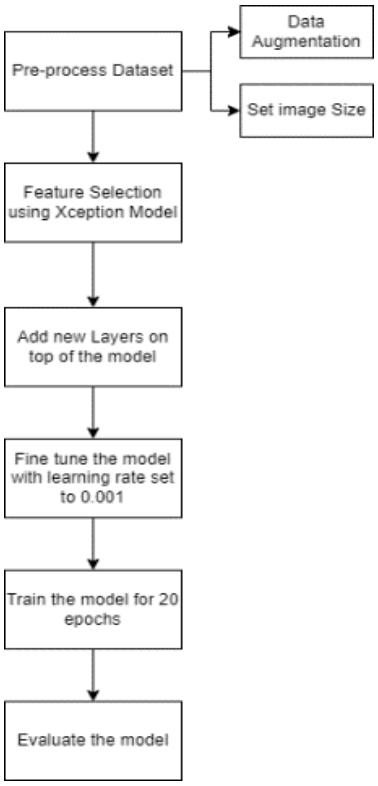

Fig. 4. Proposed Model

The regulirization phase is then initiated that comprises with the activation function 'softmax' that compromises overfitting of the model where neurons in the network are dropped from trainng the model resulting other neurons to step up and learn a pattern in the image. For fine tuning the model, all the layers after 'add_3' layer is the selected from the model to be trained and all layers before the add_3 later are frozen from being trained and the weights after the add_3 layer are used for training. The model is then compiled with a fine tuned learning rate of 0.001 with corresponding measure of loss and accuracy for 20 epochs. The model hence compiled with loss calculated using the 'categorical_crossentropy' and the accuracy using the Top 1% and Top 5% accuracy. The model is then fitted to find the optimum accuracy of the model.

The proposed methodology provides a simpler and robust approach to more accurate food object detection and use of transfer learning on top of a Xception model gives a globally accepted model. The dataset Food-101 had also been proved to be a very accurate dataset to work on food object detection models.

4. Result and Discussion

The proposed model based on the Xception Model had been trained and the validation Top 5% accuracy and the corresponding loss are also calculated successfully. The accuracy hence obtained is thus plot on a graph in Fig 5. comprising only the validation accuracy, which is the accuracy obtained by testing the prediction accuracy of the model. It was observed that the accuracy of model reached a Top 5% accuracy of 95.20% at epoch 18.

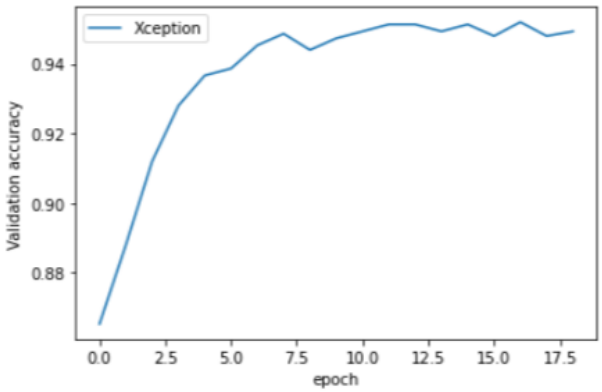

Fig. 5. Validation Accuracy of the model

The loss graph is plotted in Fig 6 showing a gradual decrease of loss average with the number of epoch. Here initially the loss ranged from 150% and later decreased to 11.65 %. The promising result with the model proves a good model to train the Food 101 dataset.

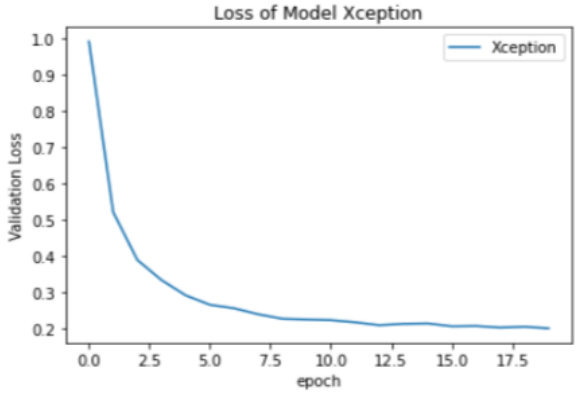

Fig. 6. Validation Loss of the model

To compare among peer models with promising results, the model is then compared with two of the state of art models trained on Food 101 dataset. A miniature Food-101 dataset with 10 images per class is taken into consideration while training the models. In Fig 7 the Xception model shows better convergence with respect to other models such as Inception V3 and ResNet50 since the accuracy rate measured is higher compared to that of the other models. For a minimal dataset, it was validated that the model based on Xception gives a higher Top 5% accuracy rate compared to that of ResNet50 for the Food 101 Dataset.

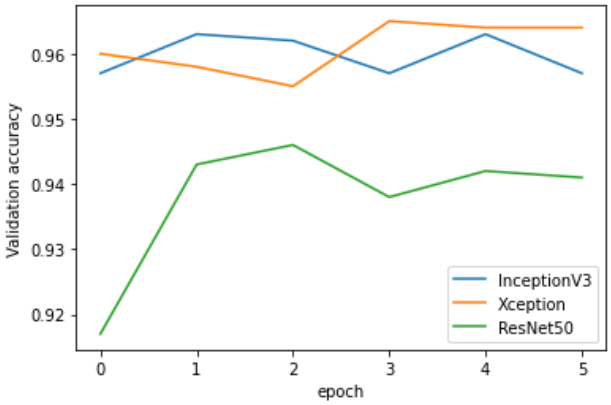

Fig. 7. Comparison of Accuracy rate between models with Food-101 Dataset

5. Conclusion

The model that has been proposed in this research based on the fine-tuned Xception model through transfer learning that has achieved a high accuracy rate of 95.20% which is a great achievement in comparison to other models that had previously been used using the same dataset. The research had the sole purpose of detecting food objects successfully from images and through the evaluation it was shown the model has a very low loss rate with a learning rate of 0.001%. The research still has shortcomings that need to be addressed in future work like a more varied dataset and a better model of food object detection. An accurate calorie measuring method needs to be proposed in future version of this research. Even though the field of food science and health sector have huge demand in tackling with correct diet managing method, this research is a beginning of such research as well that will be highly beneficial to the society. The Data Augmentation of the images proved a very important aspect of the research in the pre-processing phase and fine tuning the model has been the main novelty of the work that had been done in this research.

References

[1] S. Pouyanfar and S.-C. Chen, “Semantic concept detection using weighted discretization multiple correspondence analysis for disaster information management,” in the 17th IEEE International Conference on Information Reuse and Integration, 2016, pp. 556-564.

[2] M.-L. Shyu, C. Haruechaiyasak, S.-C. Chen, and N. Zhao, “Collaborative filtering by mining association rules from user access sequences,” in IEEE International Workshop on Challenges in Web Information Retrieval and Integration, 2005, pp. 128-135.
[3] X. Chen, C. Zhang, S.-C. Chen, and S. Rubin, “A human-centered multiple instance learning framework for semantic video retrieval,” IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 39, no. 2, pp. 228-233, 2009.
[4] Q. Zhu, L. Lin, M.-L. Shyu, and S.-C. Chen, “Effective supervised discretization for classification based on correlation maximization,” in IEEE International Conference on Information Reuse and Integration, 2011, pp. 390-395.
[5] C. Chen, Q. Zhu, L. Lin, and M.-L. Shyu, “Web media semantic concept retrieval via tag removal and model fusion,” ACM Transactions on Intelligent Systems and Technology, vol. 4, no. 4, pp. 1-22,2013.
[6] T. Meng, and M.-L. Shyu, “Leveraging concept association network for multimedia rare concept mining and retrieval,” in IEEE International Conference on Multimedia and Expo, 2012, pp. 860-865.
[7] K. Yanai, and Y. Kawano, “Food image recognition using deep convolutional network with pre-training and fine-tuning,” in IEEE International Conference on Multimedia & Expo Workshops, 2015, pp.1-6.
[8] T. Joutou, and K. Yanai, “A food image recognition system with Multiple Kernel Learning,” in 16th IEEE International Conference on Image Processing, 2009, pp. 285-288.
[9] S. Memiş, B. Arslan, O. Z. Batur and E. B. Sönmez, "A Comparative Study of Deep Learning Methods on Food Classification Problem," 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), 2020, pp. 1-4, doi: 10.1109/ASYU50717.2020.9259904.
[10] Dipta Gomes, A. F. M. Saifuddin Saif, and Dip Nandi. 2020. Robust Underwater Object Detection with Autonomous Underwater Vehicle: A Comprehensive Study. In Proceedings of the International Conference on Computing Advancements (ICCA 2020). Association for Computing Machinery, New York, NY, USA, Article 17, 1–10. https://doi.org/10.1145/3377049.3377052
[11] P. P. Urmee, M. A. A. Mashud, J. Akter, A. S. M. M. Jameel and S. Islam, "Real-time Bangla Sign Language Detection using Xception Model with Augmented Dataset," 2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), 2019, pp. 1-5, doi: 10.1109/WIECON-ECE48653.2019.9019934
[12] He, K., Zhang, X., Ren, S., and Sun, J., “Deep Residual Learning for Image Recognition”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, 2016.
[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z.B. Wojna,“Rethinking the Inception Architecture for Computer Vision”, in proceedings IEEE CVPR, Las Vegas, 2016.
[14] He, K., Zhang, X., Ren, S., and Sun, J., “Deep Residual Learning for Image Recognition”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, 2016
[15] G. Huang, Z. Liu, L.V.D Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks”, in proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261-2269, Honolulu, 2017, doi: 10.1109/CVPR.2017.243.
[16] S. Zagoruyko and N. Komodakis. “Wide Residual Networks”, in proc. British Machine Vision Conference, Sept. 2016.
[17] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, “Aggregated Residual Transformations for Deep Neural Networks”. In proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5987-5995, Honolulu, 2016.
[18] L. Pan, S. Pouyanfar, H. Chen, J. Qin and S. -C. Chen, "DeepFood: Automatic Multi-Class Classification of Food Ingredients Using Deep Learning," 2017 IEEE 3rd International Conference on Collaboration and Internet Computing (CIC), 2017, pp. 181-189, doi: 10.1109/CIC.2017.00033.
[19] Z. Zong, D. T. Nguyen, P. Ogunbona and W. Li, "On the Combination of Local Texture and Global Structure for Food Classification," IEEE International Symposium on Multimedia, Taichung, pp. 204-211, 2010, doi: 10.1109/ISM.2010.37.
[20] K. Yanai and Y. Kawano, "Food image recognition using deep convolutional network with pre-training and fine-tuning," IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 1-6, Turin, 2015
[21] P. 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.
[22] Pouladzadeh, Parisa & Shirmohammadi, Shervin. (2017). Mobile Multi-Food Recognition Using Deep Learning. ACM Transactions on Multimedia Computing, Communications, and Applications. 13. 1-21. 10.1145/3063592.
[23] G. 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.
[24] Nareen O. M. Salim et al 2021 J. Phys.: Conf. Ser. 1963 012014
[25] V. H. Reddy, S. Kumari, V. Muralidharan, K. Gigoo, and B. S. Thakare, "Food Recognition and Calorie Measurement using Image Processing and Convolutional Neural Network," in 2019 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), 2019, pp. 109-115.
[26] Bossard, L., Guillaumin, M., Van Gool, L. (2014). Food-101 – Mining Discriminative Components with Random Forests. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8694. Springer, Cham. https://doi.org/10.1007/978-3-319-10599-4_29
[27] 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: Proceedings of the 14th International Conference on Inclusive Smart Cities and Digital Health - Vol. 9677, pp. 37–48.
[28] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1800-1807, doi: 10.1109/CVPR.2017.195.

Author's Profile

Dipta Gomes completed his Masters in Computer Science from University of Ulster, Belfast, Northern Ireland in 2023. Previously he completed his undergraduate BSc in 2017 and Masters in Intelligent Systems from American International University-Bangladesh (AIUB) in 2019. Most of his ongoing and current contributions are in the fields of Machine Learning, Deep Learning, Computer Vision and Algorithms. Previously He worked as a Lecturer of the department of Computer Science in American International University-Bangladesh and currently he is pursuing his PhD in Computer Science. His research interest includes Artificial Intelligence, Computer Vision, Deep Learning, Image Processing, Pattern Recognition and Machine Learning.

How to cite this paper: Dipta Gomes, "Classification of Food Objects Using Deep Convolutional Neural Network Using Transfer Learning", International Journal of Education and Management Engineering (IJEME), Vol.14, No.2, pp. 53-60, 2024. DOI:10.5815/ijeme.2024.02.05