INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

Deep Learning-Based Food Classification on The UEC Food-100 Database

[1]A. VASANTHI, [2]D. RAMMOHAN REDDY

[1]PG Scholar, Dept. of MCA, Newton's Institute of Engineering, Guntur, (A.P)

[2]Associate professor, Dept. of CSE, Newton's Institute of Engineering, Guntur, (A.P)

Abstract: *The device has several attractive packages based on automatic food identification, from food waste control to announcements, calorie estimation, and daily food tracking. Despite the importance of this concern, the range of relevant studies is still limited. Additionally, the discrepancy within the literature was currently achieved on standard shooting performance without considering the more general approach of averaging over several tests. This paper reviews the most common deep learning methods used for food categorization, presents a publicly available food database, and publishes benchmark results for food type tests averaged over five trials and beats the best. A shot performance test reaches a state-of-the-art accuracy of 90.02% in the UEC Food-a-Hundred database. The first-level results were obtained by averaging the predictions of the ResNeXt and DenseNet models using the ensemble method. All experiments are run on the UEC Food-100 database, one of the most widely used databases, due to the presence of images of various foods, which must be cropped before processing.*

Keywords: *Food classification, food-100 Database, image classification, machine learning, Artificial neural networks*

I. INTRODUCTION

People all across the world are becoming more health aware as the world becomes more competitive and dynamic. Overweight is becoming a worldwide concern at an alarming rate. India ranks #1 among countries with the largest vegetarian population, with 40% of Asian Indians being vegetarian [1]. It appears that vegans in India are going through a "nutrition transition," with less eating of whole plant meals and more refined carbs, fried foods, and processed foods. This study investigates the link between a vegetarian diet and weight loss. With the rise of nutrition-related diseases, there is an increasing global awareness of the importance of eating a well-balanced and healthy diet. Obesity, diabetes, and cancer can all be prevented by eating a nutritious diet [2]. Although there are a variety of programmes available to detect and classify food, they all require prior knowledge to do so. However, when a new and unfamiliar cuisine is presented, a difficulty occurs. Several studies on the classification of food images have previously been completed. Food image classification is a relatively new sector in the coming applications of deep learning developments. Prior to the development of Deep Learning algorithms, several food categorizations work employed the standard Machine Learning technique for classification. Food-100 data is divided into several subsets. The goal is to use photographs that have been downscaled to allow for speedy testing. HDF5 has been used to reformat the images.

The image classification plays an important role for achieving these systems and it is a challenging task because of the several environmental and technical issues such as light conditions, image quality, noise, orientation, scale, etc. Moreover, the food classification problem becomes complex with different cooking methods in which the same food will have different shapes, e.g., raw and cooked food combined with sauces; that is, the food item is intrinsically deformable with high intra-class variation. Furthermore, the food recognition task becomes even more challenging with multi-food images, i.e., when a dish contains several different comestibles, which may overlap. As opposed to outdoor scene decomposition, the multi-food images do not have any distinct spatial layout, which can help to easier the task; i.e., the type and position of the ingredients of a mixed salad is not predictable. In this case, the common procedure is to utilize a candidate region detection algorithm followed by the classification step. The candidate region detection is a necessary pre-processing step, which is used to extract all the regions that contain only food in an image, i.e. regions of interest, and it allows the subsequent steps of feature extraction, to learn higher level representation, and classification, which lists all variety of comestibles present in an image. Obviously, the food regions extracted from multi-food images have lower quality. Therefore, the databases with multi-food images are more challenging and they are good to work on both detection and recognition tasks

This paper focuses on the food classification issue and uses the UEC Food-100 dataset because it is more arduous, due to the presence of multi-food images. UEC Food-100 database is one of the commonly used datasets to verify the performance of new methods and it also has the advantage of being easily extendable with pictures from the UEC Food-256 dataset. This work completes our previous paper by increasing the number of experiments, providing a detailed analysis of the most common deep learning algorithms used for food classification, giving a comprehensive overview of available databases of food and achieving the state-of-the art performance on the classification experiment of the UEC Food-100 database with 90.02% accuracy.

The proposed research is carried out with the help of the Python programming language and the TensorFlow package. The results were compared to other transfer learning systems when they were completed.

II. LITERATURE SURVEY

The best results on classification in the literature are achieved by the deep learning algorithms, which radically changed the classical approach of machine learning made up of feature extraction and classification. The input to supervised deep learning methods is data with labels; during training the network learns the weights and, at test time, the pipeline of layers predicts the class of the input data. In other words, deep learning models are general-purpose algorithms that are capable of extracting higher level representation of the data, i.e., food items, and classify them.

Lei Zhou [4], gave a basic introduction to deep learning and thorough descriptions of the construction of certain popular deep neural network architectures as well as training methods. hundreds of publications were reviewed that employed deep learning as a data analysis technique to handle difficulties and challenges in the food domain, such as food recognition, calorie calculation, quality detection of fruits, vegetables, meat, and aquatic goods, food supply chain, and food contamination. Each research looked into the individual challenges, datasets, pre-processing methods, networks and frameworks used, performance achieved, and comparisons with other popular solutions. Deep learning's potential for application as an advanced data mining tool in food sensory and consumption studies was also investigated.

Amatul Bushra Akhi et al [5], used to train an image category classifier, employed a pre-trained Convolutional Neural Network (CNN) as a feature extractor. A multiclass linear Support Vector Machine (SVM) classifier trained with extracted CNN features is used to classify fast food photos into ten different kinds. A multiclass linear Support Vector Machine (SVM) classifier trained with extracted CNN features is used to classify fast food photos into 10 different kinds and reached a success rate of 99.5% after working on two distinct benchmark databases, which is higher than the accuracy achieved using bag of features (BoF) and SURF. Chairi Kiourt, explained the three main lines of solutions, namely de-sign from scratch, transfer learning, and platform-based methodologies, are presented, tested, and compared to show the inherent strengths and weaknesses, specifically for the task at hand. Basic background material, a part devoted to important datasets that are critical in light of the empirical methodologies used, and some final notes that highlight future directions complete out the chapter.

Michele De Bonis et al. [6] The paper presents an effective way for constructing a mobile application for food recognition using Convolutional Neural Networks (CNNs). The GoogLeNet, which has the best accuracy and a model size second only to the SqueezeNet with 40Mb of model size and approximately 70% accuracy, and the SqueezeNet, which has a decent degree of accuracy but an extremely limited model size with 3Mb of model size and about 60% accuracy, have been identified. Malina Jiang, built convolutional neural networks from scratch and using pre-trained weights learnt on a bigger picture dataset (transfer learning) to solve the issue of food image classification, reaching an accuracy of 61.4%.

III. PROPOSED SYSTEM

The recent developments in machine learning attest the superior performance of deep learning approaches over the classical methods. The DCNN network allows to analyse various images and videos. Currently, DCNNs are successfully used for image classification, image segmentation, object detection and localization problems. A review on a specific application of deep learning with food is given by Zhou et al.

This paper compares the performances of several deep learning methods using food images of the UEC FOOD100 database; it reaches the state-of-the-art performance of 90.02% best-shot accuracy using the ensemble method with the bagging strategy on DenseNet-161 and ResNeXt-101.

A) Deep Convolutional Neural Network Models.

The Convolutional Neural Network (CNN) provides a feature extraction step with the convolution and the pooling layers and a classification phase with fully connected layers. The convolution layer uses sliding boxes, called filters, to extract hierarchical features out of the images. The number of convolutional layers, pooling and fully connected layers, the coefficients, and the size and the number of the filters are all critical hyper parameters of the deep learning systems. There is also a final layer, which is called the softmax or classification layer. The DCNN became famous in 2012, when Krizhevsky et al. presented the with AlexNet architecture, which won the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) This model is made up of 5 convolutional layers (CL) followed by 3 fully connected layers; the 1st CL uses filters of size 11 × 11, the 2nd CL filters of size 5 × 5, the other layers use filters of size 3 × 3.

The AlexNet tackles the vanishing gradient problem by using the Rectified Linear Unit (ReLU) activation function, instead of the Sigmoid or Tanh functions. Another important change was to reduce the overfitting problem by adding a dropout layer after every fully connected layer. The dropout layer randomly associates to every neuron a probability value and switches off those neurons, which do not reach the pre-fixed threshold. The ConvNet or the VGGNet architecture proposed by the VGG group of Oxford improved the AlexNet model by replacing the big size filters of the 1st and 2nd CL with multiple little filters of size 3 × 3. This trick allows to have the same receptive field, i.e., the size of the area of the input image from with the output depends, while decreasing the total number of parameters, e.g., from (7 × 7) to 3 × (3 × 3). The VGG16 CLs are followed by 3 fully connected layer for a total of 16 layers. The VGG architecture achieved the top-5 accuracy of 92.9% on ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) 2012 dataset and 93.2% on the ILSVRC 2014 dataset.

The complexity of VGG architecture is still too high, because the model is fully connected, i.e., in a layer having input and output channels equal to C, e.g. C = 512, every input channel is connected to every output channel. The GoogLeNet introduced a sparse connected architecture based on the idea that most of the activation in a deep network are not necessary, because redundant, due to the correlations between them. Since kernels for sparse matrix multiplication were not optimized, the GoogLeNet released the Inception module that approximates a sparse CNN with a normal dense construction. Furthermore, the Inception module decreases the number and the size of the convolutional filters and keeps a little width of the network which is reducing the total complexity of the model. Additionally, it uses the convolutional filters with different sizes, i.e., 5, 3 and 1, to consider the details at different resolutions. Other important improvements proposed by this model are: (1) the bottleneck layer implemented by using the filters of size 1 × 1, and (2) the substitution of the fully connected layers with a global averaging pooling layer. The Inception architecture is much faster than the VGG. The Inception V3 achieved the top-5 accuracy of 94.4 % on ILSVRC 2012 dataset.

B) Residual Networks (ResNet)

it was introduced the use of a residual block, which adds the original input to the output feature calculated by processing the input with one or more convolutional layers. That is, when using residual blocks, a layer can feed the following layer and other layers which are several steps ahead in the architecture, i.e., residual blocks are skip connection blocks. The name of this block is related to the implemented mathematical function as it learns the difference (or residual) between the output and the input signals. Finally, like Inception, ResNet uses a global averaging pooling layer before the classification layer. ResNet-152 achieves 96.43% top-5 accuracy in the ILSVRC 2015 dataset.

IV. THE FOOD DATABASE

There are only a small number of food photo databases, with unique features such as number of food classes, total number of photos, type of food, ie. Western (French, Italian, Turkish, ...), Asian (Japanese, Chinese, Thai, ...) or fast food (since it can be considered a type of international food), nice and something like the pictures, that is, against images of single multiple foods and unusual contexts, ie. A single plate may hold more than one food, or a single food item may be separated into multiple plates on a tray. Comparing the different types in the available databases is a difficult task. This research specializes in a publicly available database with more than 60 lessons and 1,000 images. The Pittsburgh Fast Food Photographic Dataset (PFID) [32] is one of the first publicly available databases of food images. Released in 2009, it contains 1,098 images from 61 subjects. The snapshots were taken at a number of popular fast-food locations, where the same item was photographed from specific angles. Additional images were collected under controlled environmental settings in the laboratory. All images in the database contain only one individual dose.

The UEC Food-100 database

The UEC Food-100 database has been introduced by Matsuda et al. in 2012. It stores a total of 12,740 images belonging to 100 classes. Since this work challenges the classification experiment on this dataset.

The UEC Food-100 database stores 12,740 food photos for a total of one hundred lessons. The lessons are mostly Japanese foods, however, there are also foods commonly used in Western cuisine, including red meat, toast, croissants, buns, hamburgers, pizza, sandwiches, spaghetti, sausages, omelets, fried Includes fish, and steamed items. Grilled salmon, pro pork with potatoes, steak, hot dog and chuck. Each image in the database has a floor reality that bounds the containers around the dining area. Most of the snapshots (11,566) contained images of a single food, but there were also images (1,174) with images of more than one food.

Fig.1 Example of pictures from the UEC FOOD-100 dataset

Fig.2 Example pictures from different databases of food

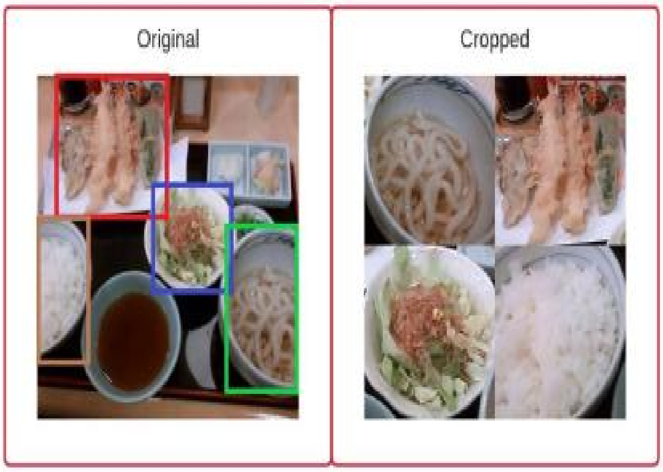
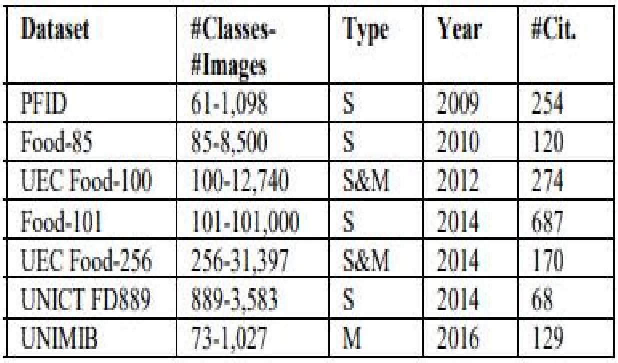


Fig.3 (left) Original multi-food image, (right) Cropped Images

Table.1 Databases of food images. "s" stands for single-food item and "m" for multi-food item image.



Dataset	#Classes-#Images	Type	Year	#Cit.
PFID	61-1,098	S	2009	254
Food-85	85-8,500	S	2010	120
UEC Food-100	100-12,740	S&M	2012	274
Food-101	101-101,000	S	2014	687
UEC Food-256	256-31,397	S&M	2014	170
UNICT FD889	889-3,583	S	2014	68
UNIMIB	73-1,027	M	2016	129

V. CONCLUSION

Despite the big importance of food classification systems, the current number of studies and improvements are still too limited. The main drawback is the lack of big and international databases, which are necessary to train the algorithms. Also, new methods can help to improve the performance on available databases. This work overviews the main algorithms used for food classification, it details the databases of food items currently available and it presents the results of several deep learning algorithms, considering both the best-shot performance as well as the average over five trials. In the best shot performance experiment, this paper reaches the state-of-the-art accuracy of 90.02% on the UEC Food-100 database, improving the previous record by 0.44 percentage points. However, since all methods are very sensitive to the choice of the training and test sets, we believe that comparison based on the average performance over 5-trials is more appropriate. With the best of our knowledge, this is the first work, which reports the accuracy averaged over 5-trials on the UEC Food100 and it can be used as benchmark paper.

REFERENCES

[1] Y. Matsuda, H. Hoashi, and K. Yanai, "Recognition of multiple-food images by detecting candidate regions," in Proc. 2012 IEEE International Conference on Multimedia and Expo, 2012, pp. 25–30.

[2] Y. Kawano and K. Yanai, "Automatic expansion of a food image dataset leveraging existing categories with domain adaptation," in Proc. European Conference on Computer Vision. Springer, 2014, pp. 3–17.

[3] S. Memis¸, B. Arslan, O. Z. Batur, and E. B. Sonmez, "A comparative¨ study of deep learning methods on food classification problem," in Proc. 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), 2020, pp. 1–4.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Proc. of the 25th International Conference on Neural Information Processing Systems Volume 1, ser. NIPS'12. Curran Associates Inc., 2012, p. 1097–1105.

[6] Prasadu Peddi (2021), "Deeper Image Segmentation using Lloyd's Algorithm", ZKGINTERNATIONAL, vol 5, issue 2, pp: 1-7.

[7] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks." In Proc. 2nd International Conference on Learning Representations (ICLR), 2014.

[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[9] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[10] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," Artificial Intelligence Review, vol. 53, no. 8, pp. 5455–5516, 2020.

[11] Prasadu Peddi (2017) "Design of Simulators for Job Group Resource Allocation Scheduling In Grid and Cloud Computing Environments", ISSN: 2319- 8753 volume 6 issue 8 pp: 17805-17811.