2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Mediterranean Food Image Recognition Using Deep Convolutional Networks

Fotios S. Konstantakopoulos, Eleni I. Georga, *Member, IEEE* and Dimitrios I. Fotiadis, *Fellow, IEEE*

Abstract— **We present a new dataset of food images that can be used to evaluate food recognition systems and dietary assessment systems. The Mediterranean Greek food - MedGRFood dataset consists of food images from the Mediterranean cuisine, and mainly from the Greek cuisine. The dataset contains 42,880 food images belonging to 132 food classes which have been collected from the web. Based on the EfficientNet family of convolutional neural networks, specifically the EfficientNetB2, we propose a new deep learning schema that achieves 83.4% top-1 accuracy and 97.8% top-5 accuracy in the MedGRFood dataset for food recognition. This schema includes the use of the fine tuning, transfer learning and data augmentation technique.**

I. Introduction

The modern way of life and its ever more faster rhythms make it difficult for most people to adopt a daily healthy diet. It is a fact that people nowadays consume more and more foods high in calories and high in fats. This results in a steady increase in unhealthy eating related diseases, such as obesity, diabetes, and cardiovascular disease. Worldwide obesity has reached epidemic proportions, with 2.8 million people dying each year from being overweight or obese. Although associated with high-income countries, obesity is now prevalent in low- and middle-income countries. The number of obese people has almost tripled since 1975. Moreover, 38 million children under the age of five were overweight or obese in 2019 [1]. Diabetes has become a common disease of the modern way of life and is considered a metabolic disorder that causes high levels of blood sugar. There are two main types of diabetes: Type 1 which is characterized by a lack of insulin and Type 2, which characterized by the body's insulin resistance [2]. Today about 463 million adults live with diabetes, while this number is expected to raise to 700 million in 2045. To date, diabetes is estimated to have caused 4.2 million deaths globally [3]. Cardiovascular diseases (CVDs) are a group of disorders of the heart and blood vessels, and are the number one cause of death worldwide. The number of deaths from CVDs has increased from 2 million in 2000 to almost 9 million in 2019. CVDs account for 16% of all deaths from all causes [4]. A common factor that can affect the prevention and treatment of the above diseases is the management of the daily diet.

Today, the progress in the field of artificial intelligence (AI) and computer vision enables individuals to monitor their diet on a daily basis, through the usage of appropriate applications. Recent studies have shown that AI applications are more popular among users, compared to traditional methods of recording nutritional composition. Important components of these applications are the food image dataset, as well as the food image recognition system. A dataset of food images consisting of many and quality images, is essential for the use of deep learning techniques in food image recognition. Food recognition systems are responsible for identifying food through an image, that can be captured via a smart device, such as a smartphone.

There are many food image datasets that are used either for food image recognition systems or for their evaluation. These datasets are distinguished by the total number of images, the number of food classes, the type of cuisine as well as the source of obtaining the images. For example, ChineseFoodNet [5] represents the Chinese cuisine, while FFoCat [6] refers to Mediterranean cuisine. The PFID dataset [7], on the other hand, consists of 61 classes of food with a total number of 1,098 images, captured in fast-food restaurants, while NutriNet [8] contains 225,953 food images belonging to 520 classes downloaded from the web. Large food image databases, such as Food-101 [9], UEC-Food100 [10], VIREO Food-172 [11] and UEC-Food256 [12] are used to evaluate deep learning models.

Food recognition is an important step in dietary assessment systems and is responsible for the correct identification of the image of food entering the system. The existing food recognition techniques can be divided into: (i) traditional machine learning techniques, and (ii) deep learning techniques. In addition, top-1 and top-5 accuracy metrics are used to evaluate food image recognition methods. Top 1 accuracy is the accuracy where true class matches with the most probable classes predicted by the model, while top 5 accuracy is the accuracy where true class matches with any one of the 5 most probable classes predicted by the model. In traditional machine learning approaches, a feature extractor, such as scale invariant feature transform (SIFT) [13] or speed-up robust features (SURF) [14], is selected and then the

This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness. Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T1EDK-03990).

F. S. Konstantakopoulos, E. I. Georga, and D. I. Fotiadis are with the Unit of Medical Technology and Intelligent Information Systems, Materials Science and Engineering Department, University of Ioannina, Ioannina, GR 45110 Greece (e-mail: fotkonstan@uoi.gr, egeorga@uoi.gr, corresponding author phone: +302651009006; fax: +302651008889; e-mail: fotiadis@uoi.gr).

D. I. Fotiadis is with the Institute of Molecular Biology and Biotechnology, Biomedical Research Department, FORTH, University of Ioannina, Ioannina, GR 45110 Greece.

Fig. 1 Food images of the proposed MedGRFood dataset. From the left to right: cabbage rolls, dolmades, galaktoboureko, fava, Greek salad, imam bayildi, lamb fricassee, soutzoukakia, moussaka, pastitsio, pork souvlaki, rabbit stew with onions, pita gyro, ravani, shrimp saganaki and tzatziki.

extracted features are into to a classifier for training the prediction model using machine learning algorithms, such as support vector machine (SVM) [15] or random forests (RF) [9], achieving top-1 accuracy 82.2% and 50.8% in its own and Food-101 dataset, respectively.

In recent years, Deep Neural Networks (DNNs) and especially Convolutional Neural Networks (CNNs), have become the state-of-the-art method of image recognition. Compared to other image recognition methods, CNNs actually use little preprocessing. CNNs are used to extract features from images, through which the layers of the network learn, while the network is trained in a set of images. This feature makes DNNs ideal for computer vision tasks. There are many CNNs used to recognize food images that are either already built or built from scratch. For example, IG-GMAN [16], WISeR [17] and DenseFood [18] are networks that have been built from scratch and can achieve top-1 accuracy 90.4%, 83.2% and 81.2% in datasets Food-101, UEC-Food256 and VIREO-172, respectively. In addition, in many cases pre-trained CNNs are used, such as Inception V3 [19] or Vgg-16 [20] achieving top-1 accuracy 81.5% and 71.7% in UEC-Food100 and in its own dataset, respectively.

In this study we present a new dataset of food images of Mediterranean cuisine, appropriate for food recognition systems. For its evaluation, first we use the fine-tuning deep learning model EfficientNet, applying the techniques of transfer learning and data augmentation. Then, we compare it with the same deep learning food recognition model, without the application of fine-tuning technique.

II. MedGRFood Image Dataset

The Mediterranean diet incorporates the traditional healthy habits of people from countries bordering the Mediterranean Sea. It varies by country and region, but the principal aspects of this diet include high consumption of olive oil, vegetables, legumes, fish and low consumption of meat products [21]. The Mediterranean diet is linked with good health and low risk for many diseases, such as obesity, diabetes and cardiovascular diseases [22]. Therefore, we created the MedGRFood dataset which consists of Mediterranean cuisine food images, which are divided into 11 food groups, based on the Greek Food Composition Dataset by the Hellenic Health Foundation [23]. The food groups are the following: (i) Milk, dairy product or milk substitute, (ii) Egg or egg products, (iii) Meat or meat products, (iv) Seafood or related products, (v) Fat or oil, (vi) Grain or grain products, (vii) Nut, seed or kernel products, (viii) Vegetable or vegetable products, (ix) Fruit or fruit products, (x) Sugar or sugar products, and (xi) Miscellaneous food products. The dataset consists of 42,880 food images which belong to 132 food classes. Most of the images have been collected from the web, while the rest have been taken under specific conditions, completing the required number of images per food class for a balanced dataset. The dataset contains some popular Mediterranean dishes, such as those shown in Fig. 1. The MedGRFood dataset is added to the existing datasets of food images, being a new food image dataset, with different food classes of current datasets, with quite high-resolution images, making it ideal for evaluating food images algorithms, but also for its use in dietary assessment systems. This is the first dataset of food images of Greek cuisine and one of the most representative food image datasets of the Mediterranean cuisine.

III. Food Recognition System

A. EfficientNet model

To recognize food images we use a model from the EfficientNets family [24] as a base model. In EfficientNet a new scaling method is proposed (called compound scaling) to increase the model's size in order to achieve maximum accuracy. The compound scaling method can be used to existing CNN architectures, such as the ResNet [25]. The previous CNN models follow the conventional approach of scaling the dimensions arbitrarily by adding more layers. In EfficientNet simultaneous and uniform scaling of dimensions is applied by a fixed amount, achieving much better performance. More specific, EfficientNets using a weighted scale of three-connected hyperparameters of the of the input model resolution, depth and width of the network in a principled way:

$$\begin{aligned} depth: d &= \alpha^{\varphi} \\ width: w &= \beta^{\varphi} \\ resolution: r &= \gamma^{\varphi} \end{aligned} \quad , \tag{1}$$

Where φ, the compound coefficient, is a user defined global scaling factor that controls how many resources are available, while α, β, γ determine how these resources are allocated in depth, width and network resolution, respectively. When φ is set to 1, the base configuration is acquired the first version of EfficientNets models family, the EfficientNet-B0. Next, this configuration is used in a grid search to find the α, β and γ coefficients, that optimize the following equation,

under the constraint: $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$,

$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1 \ , \quad (2)$$

Once the coefficients α, β and γ are estimated, then the compound coefficient φ can be increased to get deeper but more accurate models. This is how the EfficientNet-B1 to EfficientNet-B7 models are constructed, with the integer at the end of the name indicating the value of compound coefficient. In this study we use the EfficientNet-B2 as base model for the recognition of food images.

B. Transfer learning and fine-tuning

Transfer learning is the improvement of learning in a new task, through the transfer of knowledge from a related task that has already been learned. The general intuition behind transfer learning to image recognition problems, is that by training a model in a large dataset it can be afterwards used effectively as a recognition model in a smaller dataset. We apply transfer learning in the base model, using the weights of pre-trained EfficientNet-B2 model in the ImageNet LSVRC-2012 dataset [26]. Knowing that features in CNNs are more generic in early layers (i.e., edge detection) and more specific to the original dataset at later layers, we take advantage of the generality and large number of ImageNet images, by transferring knowledge to our own task.

Then we apply the fine-tuning technique: we unfreeze the last two blocks of layers in the base model, and we train it in MedGRFood dataset. Along with these layers, we also train the additional fully connected activation and drop-out layers by adapting the based model to our classification problem. In total we add three blocks of these layers. These additional layers further improve the performance of the model and prevent overfitting

C. Data augmentation

Data augmentation are techniques used to increase the amount and the diversity of training data (images), by adding slightly modified but realistic copies of already existing data. Data augmentation acta as a regularizer and helps to reduce overfitting when training a CNN model. Geometric transformations flipping, cropping, zooming and rotation techniques are applied to augment the number of training food images.

D. Implementation

We used the Anaconda environment with the python programming language to implement the CNN model. We also used the cuda toolkit, and the cudnn and tensorflow libraries, for model training through the Nvidia GeForce RTX 3080 graphic processing unit.

IV. Results

For the evaluation of food system recognition, we have constructed and trained two CNN models. In the proposed model the fine-tuning technique is used and in the second model the fine-tuning is not used. Table I presents the performance of the two models. We observe that in the proposed model we achieve 83.4 and 97.8% top-1 and top-5 accuracy, respectively. In the second model we achieve 78.9 and 95.9% top-1 and top-5 accuracy, respectively. In addition, there is a difference in the value of loss, in the training time of each step (millisecond/step) and in the number of total parameters which are created. The results clearly show that the fine-tuning technique improves the performance of a CNN food recognition model. The total number of parameters created between the two models is high, which requires more training time for the first model. Nevertheless, the improvement of the model accuracy and the reduction of the loss value, make the CNN model with fine-tuning technique a better choice for food image classification problems. Top-1 accuracy of fine-tuning model is very good, considering that in MedGRFood dataset there are several foods which may have a similar appearance (i.e., pastitsio, moussaka, papoutsakia, pasta pie, cheese pie, spinach pie and more), making the correct classification of foods difficult. The fact that several foods look alike is shown by the top-5 accuracy, which has an excellent value, and is one of the best top-5 values according to the literature. The top-1 accuracy, top-5 accuracy and loss curves are presented in Fig. 2. The model has been trained for 250 epochs using the stochastic gradient descent (SGD) optimizer. Moreover, we choose a scaled learning rate, with an initial value of 0.0001 and for activation function we use the Swish [27].

TABLE I. MODEL PERFORMANCE WITH AND WITHOUT FINE TUNING

Model	Top-1 accuracy (%)	To-5 accuracy (%)	Loss	Training time (ms/step)	Number of parameters (x10^6)
Fine tuning	83.4	97.8	0.65	400	300.988
Without fine tuning	78.9	95.9	0.72	370	16.116

V. Discussion and Conclusion

In food image datasets, the use of deep learning techniques for food recognition, required the development of datasets with a large number of images for each food class. However, the existing databases are limited to the number of food classes, depending on the dietary habits of the dataset constructor. Furthermore, the quality of the images affects either positively or negatively the performance of recognition systems, so it is equally important that the food image database contains high-resolution images. The MedGRFood image dataset is a first step towards the creation of a complete dataset of food images according to the Greek Food Composition Dataset by the Hellenic Health Foundation, representing Mediterranean cuisine. Expanding the dataset, with the addition of images from missing foods classes, is the next step in completing it.

By using the new deep learning model EfficientNetB2 for food image recognition, and by applying the fine tuning, transfer learning and data augmentation techniques, we achieved 83.4 and 97.8% top-1 and top-5 accuracy respectively, in the MedGRFood food image dataset. EfficientNet is a model that performs better, compared to previous CNN image classification models. In addition, the computational cost, as well as the time required to train the model based on EfficientNet is significantly lower. In this study we prove the significance of pre-trained CNN models and of the fine-tuning technique, as both the top-1 and top-5

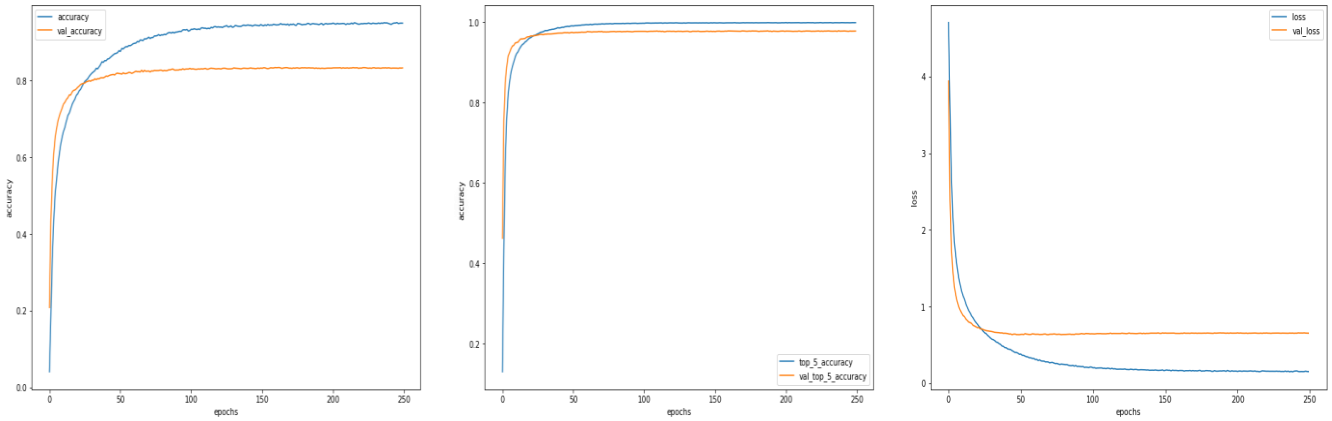

Fig. 2 Top-1 accuracy, top-5 accuracy and loss curves for EfficientNetB2 fine-tuning model.

accuracy were improved, compared to equally deep learning model. It is noteworthy to mention that the required training time does not differ significantly between the two models.

To sum up, we presented a new food image dataset and we proposed a deep learning schema for food image classification. We combined a pre-trained CNN model and applied fine-tuning, transfer learning and data augmentation techniques to improve classification results in a food recognition model. Compared to the classification model without fine-tuning technique, we achieved a 4.5% improvement in top-1 accuracy, 1.9% improvement in top-5 accuracy and significant reduction in the loss index.

REFERENCES

[1] World Health Organization. (1 April 2020,). *Obesity and overweight*. Available: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

[2] International Diabetes Federation. (20 March 2020,). *Type 1 diabetes*. Available: https://www.idf.org/aboutdiabetes/type-1-diabetes.html

[3] International Diabetes Federation. (12 February 2020,). *Diabetes facts & figures*. Available: https://www.idf.org/aboutdiabetes/what-is-diabetes/facts-figures.html

[4] World Health Organization. (17 May 2017,). *Cardiovascular dieases (CVDs)*. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[5] X. Chen, Y. Zhu, H. Zhou, L. Diao, and D. Wang, "Chinesefoodnet: A large-scale image dataset for chinese food recognition," *arXiv preprint arXiv:1705.02743,* 2017.

[6] I. Donadello and M. Dragoni, "Ontology-Driven Food Category Classification in Images," in *International Conference on Image Analysis and Processing*, 2019, pp. 607-617: Springer.

[7] M. Chen, K. Dhingra, W. Wu, L. Yang, R. Sukthankar, and J. Yang, "PFID: Pittsburgh fast-food image dataset," in *2009 16th IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 289-292: IEEE.

[8] S. Mezgec and B. Koroušić Seljak, "NutriNet: a deep learning food and drink image recognition system for dietary assessment," *Nutrients,* vol. 9, no. 7, p. 657, 2017.

[9] L. Bossard, M. Guillaumin, and L. Van Gool, "Food-101–mining discriminative components with random forests," in *European conference on computer vision*, 2014, pp. 446-461: Springer.

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

[11] J. Chen and C.-W. Ngo, "Deep-based ingredient recognition for cooking recipe retrieval," in *Proceedings of the 24th ACM international conference on Multimedia*, 2016, pp. 32-41.

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

[13] M.-Y. Chen *et al.*, "Automatic chinese food identification and quantity estimation," in *SIGGRAPH Asia 2012 Technical Briefs*, 2012, pp. 1-4.

[14] Y. Kawano and K. Yanai, "Real-time mobile food recognition system," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 1-7.

[15] S. Christodoulidis, M. Anthimopoulos, and S. Mougiakakou, "Food recognition for dietary assessment using deep convolutional neural networks," in *International Conference on Image Analysis and Processing*, 2015, pp. 458-465: Springer.

[16] W. Min, L. Liu, Z. Luo, and S. Jiang, "Ingredient-Guided Cascaded Multi-Attention Network for Food Recognition," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 1331-1339.

[17] N. 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: IEEE.

[18] A.-S. Metwalli, W. Shen, and C. Q. Wu, "Food Image Recognition Based on Densely Connected Convolutional Neural Networks," in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2020, pp. 027-032: IEEE.

[19] H. Hassannejad, G. Matrella, P. Ciampolini, I. De Munari, M. Mordonini, and S. Cagnoni, "Food image recognition using very deep convolutional networks," in *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, 2016, pp. 41-49.

[20] L. Jiang, B. Qiu, X. Liu, C. Huang, and K. Lin, "DeepFood: Food Image Analysis and Dietary Assessment via Deep Model," *IEEE Access,* vol. 8, pp. 47477-47489, 2020.

[21] A. Trichopoulou *et al.*, "Definitions and potential health benefits of the Mediterranean diet: views from experts around the world," *BMC Medicine,* vol. 12, no. 1, p. 112, 2014/07/24 2014.

[22] V. Tosti, B. Bertozzi, and L. Fontana, "Health Benefits of the Mediterranean Diet: Metabolic and Molecular Mechanisms," *The Journals of Gerontology: Series A,* vol. 73, no. 3, pp. 318-326, 2018.

[23] EuroFIR AISBL e-book Collection, The Greek Food Composition Dataset by the Hellenic Health Foundation, 1ed ed., 2011. [Online]. Available.

[24] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, 2019, pp. 6105-6114: PMLR.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[26] A. Krizhevsky, I. Sutskever, and G. E. J. A. i. n. i. p. s. Hinton, "Imagenet classification with deep convolutional neural networks," vol. 25, pp. 1097-1105, 2012.

[27] P. Ramachandran, B. Zoph, and Q. V. J. a. p. a. Le, "Searching for activation functions," *arXiv preprint arXiv:1710.05941,* 2017.