VOL.8 NO.2 September 2019 - ISSN 2233 – 1859

UOIuBIH
ORSinBIH
Operations Research Society in
Bosnia and Herzegovina

Southeast Europe Journal of Soft Computing
Available online: http://scjournal.ius.edu.ba

IUS Soft Computing
Research Group

Deep Transfer Learning for Food Recognition

Abdulah Pehlić*[1], Ali Abd Almisreb[1], Melisa Kunovac[1], Elmedin Skopljak[1] and Merjem Begovic[1]
[1]Faculty of Engineering and Natural Sciences,
International University of Sarajevo,
Hrasnickacesta 15, Ilidža 71210 Sarajevo,
Bosnia and Herzegovina
*abdulah.pehlic@gmail.com

Article Info

Article history:

Article received on 13 August 2019
Received in revised form 25 August 2019

Keywords:

Deep Transfer Learning; Food Recognition; GoogleNet; AlexNet; Vgg16.

Abstract

Food Recognition is an essential topic in the area of computer vision. One of its target applications is to avoid achieving a cashier at the dining place. In this paper, we investigate the application of Deep Transfer Learning for food recognition. We fine-tune three well-known deep learning models namely; AlexNet, GoogleNet, and Vgg16. The fine-tuning procedure depends on removing the last three layers of each model and adds another five new layers. The training and validation of each model conducted through food a dataset collected from our university's canteen. The dataset contains 39 food types, 20 images for each type. The fine-tuned models show similar training and validation performance and achieved 100% accuracy over the small-scale dataset.

1. INTRODUCTION

The recent advent of transfer deep learning has achieved successes in many areas such as classification and recognition [1]–[4]. One of the most promising visual object recognition applications is food recognition, since it helps to estimate food calories and analyze eating habits of people to maintain their health [5]. Those applications started to open new challenges to the computer vision and object recognition algorithms. Most of existing methods of food recognition directly extract visual features of the whole image using popular deep networks for food recognition without considering its own characteristics [6]. Food recognition is gaining more attention in the multimedia community due to its various applications, e.g., multimodal food-log and personalized healthcare [7]. It is common that one dish can be served in several ways. Therefore, in this paper, we will explain the utilization of deep transfer learning concept for small-scale food dataset captured directly from the tray. The rest of the paper is organized as follows: Section 2 covers the Literature Review, Section 3 contains details of the materials used in this paper and methodologies. Section 4 elaborates on the achieved results and finally the conclusion.

2. LITERATURE REVIEW

Recently the use of artificial intelligence is increasing rapidly. Almost, in every branch of our lives we are facing with the beginnings of its use. That is, also, the case with food recognition applications. More and more researches are done in this field, and many of them are there to help improving health of human body. Recognizing of food type and its features, as the content of meal brought automatic or semi-automatic dietary estimations to help people control their eating habits and help them in their diets and daily food income [8]. Those applications were improved using CNN methods, the powerful class of models in various problems, applied to the database taken from 23 restaurants, to predict the calories and nutrition of their meals from one single image [5, 8]. More and more researchers are using CNN such as ResNet, GoogleNet, MobileNet and VGG-Net. Some of the researches mention that the GoogleNet has the highest validation accuracy value, with the lowest number of epochs[10]. This lead to the use of DCNN (deep convolutional neural network), which is very suitable for large-scale image data, since it takes only 0.03 seconds to classify one food photo with GPU. The experiments done with this on ETH Food-101, UEC FOOD 100, and UEC FOOD 256 datasets show that it has achieved the accuracy of 88.28%, 81.45%, and 76.17% as top-1 accuracy and 96.88%, 97.27%, and 92.58% as top-5 accuracy [11]. The further working was concentrated on a buffet-style restaurant. The results showed that, using real data can achieve 0.79 in F-measure and 9.4% error in energy, much better than the previous approach [12]. The Supervised Extreme Learning Committee (SELC) takes as many features as possible but shows just the features which are proposed for the classification of the food. Each ELM presented a particular type feature [13]. The classification rate of 55.8 % is reached in the approach of recognizing multiple images by detecting candidate regions and classifying them with various features [12]. Newly proposed system, the visual attention analysis, has shown that the network is able to self-identify the relevant portions of the image that should be considered for classification [6]. Going further, the Ingredient-Guided Cascaded Multi-Attention Network (IG-CMAN) brings the state-of-the-art recognition performance with new dataset WikiFood-200 [7]. The improvement of meal-recognition systems were done by Tensorflow based machine-learning process was performed, where an Expert.js-based semantic network was constructed. Recognizing the Korean, Chinese and Italian food brought the result of 55.3 % of the food recognition accuracy rate [14]. When it was not just enough to recognize the meal on the plate, the system for retrieving recipes from the picture was set. The joint relationship between food and ingredient labels through multi-task learning is exploited by deep architecture. It is done by learning contextual relationships of ingredients from a large textual corpus of recipes [15]. To see the quality of the food one of the most important parameters is its expiration date. Mobile application was developed to be used in recognizing of printed expiration dates [16]. Furthermore, the ear-warn device was proposed for identification of the temporal similarity between different types of food. It is aimed to record the fluctuations on the glucose level during the satiation and satiety periods of the user, in order to keep track on daily intake of calories and their deficit, to provide a better dietary experience for users [17].

3. MATERIALS AND METHODOLOGY

3.1 Dataset

The dataset is consisted of images of 39 different sorts of foods. Each food type contains 20 images. Therefore, the total number of images in the dataset is 780. smartphone used to collect the dataset. Table 1 contains a full list of the food names included in this dataset. The dataset includes pictures of soups, main dishes, appetizers, and salads. Figure 1 shows samples of the dataset images. All the images collected from foods provided at the ccanteen of the International University of Sarajevo.

Table 1: Food labels and number of images for each type

No.	Label	Count	No.	Label	Count
1	Arap Tava	20	21	Hurmašica	20
2	Baked drumsticks	20	22	Lece	20
3	Baked potatoes	20	23	Macaroni	20
4	Baklava	20	24	Macaroni and Cheese	20
5	Baklava i Pistacija	20	25	Mashed potatoes	20
6	Barbecue steak	20	26	Meat and eggs	20
7	Burek	20	27	Mercimek soup	20
8	Cacik	20	28	Mixed salad	20
9	Chicken cream soup	20	29	Musaka	20
10	Chicken croquette	20	30	Pileci duvec	20
11	Chicken Curry	20	31	Pizza	20
12	Chicken drumstick and potatoes	20	32	Rice	20
13	Chicken wings	20	33	Season salad	20
14	Chickpeas	20	34	Šehrija	20
15	Çoban Kavurma	20	35	Sirnica	20
16	Çoban salad	20	36	Stuffed pepper	20
17	Dana Rosto	20	37	Turkish Manti	20
18	Duvec	20	38	Zeljanica	20
19	Et Haslama	20	39	Grašak	20
20	Ezogelin soup	20			

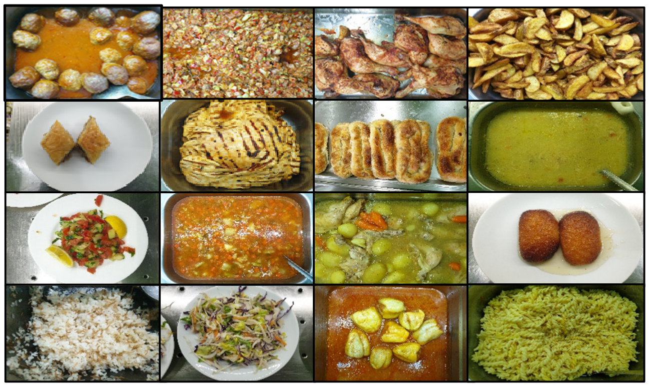

Figure 1: Samples of food images

3.2 Methods

The success of [18] to deal with largescale dataset bacame a revolution in the fild of machine learning. in contrast with the shallow neural network, deep neurla nwtwork compromises of large number of layers. Therefore, it ganerate a large number of parameters which needs ahardware with high capability such as GPUs working in parallel. Hence, Transfer learning concepts emerged to assests in train deep neural network models with small scale datasets. In this paper, we fine tuned 3 of the ealiest deep learning models namely; AlexNet [18], GoogleNet [19] and Vgg16 [20]. The fine-tuning depends on removing the last 3 layers of each model and replace it with 5 new layers. Fine-tuning process is tabulated in Table 2.

Table 2: AlexNet, GoogleNet and Vgg16 fine tuning

AlexNet	**GoogleNet**	**Vgg16**	**Fine-tune**
'fc8': Fully Connected	'loss3-classifier': Fully Connected	'fc8': Fully Connected	'fc8': Fully connected
'prob': Softmax	'prob': Softmax	'prob' : Softmax	'relu8': Relu
'output': Classification Output	'output': Classification Output	'output': Classification Output	'fc9': Fully connected
			'prob': Softmax
			'output': Classification output

Fine-tuning make the models hybridized of trained and newly added layers. Therefore, in order to have balance between all layers in terms of the learning speed, we assign higher learning rate for the new layers to boost their learning and freeze the old layers weight.

4. RESULTS

The deep transfer neural networks that were used in this research were trained in MATLAB, and after the training process accomplished, the results of the training are seen. The results show how accuracy and loss have been behaving from the first to the last iteration. Accuracy represents how well the network has learned the image, and loss function is actually the opposite of accuracy, the percentage of the loss in training. We can see from the presented figures that loss is basically an opposite graph of the accuracy.

4.1 Training Evaluation

A. AlexNet

Firstly, we notice from Figure 2 (a) that AlexNet training accuracy risen up very quickly. This shows how handy transfer learning is in practice. The main reason it rises so fast is that all of the layers not containing new data for the learning process were already trained. Almost after 200 iterations approximately, the accuracy is fully developed. This means that the accuracy of the learning process is done, and the network has learned 100% of the data we already gave it for processing. More deatils about the performance of AlexNet training during the earlier iterations is shown in Figure 2(b).

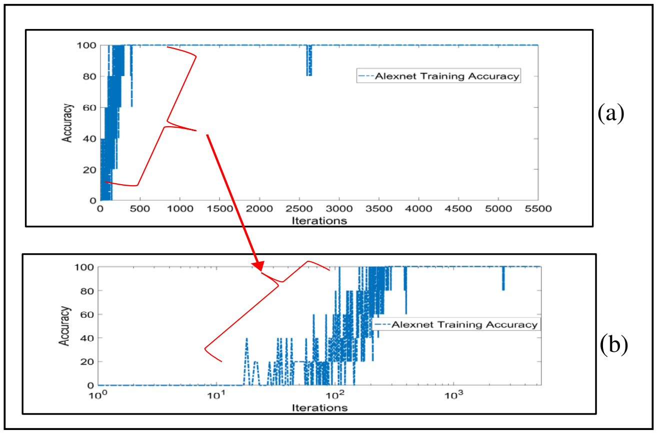

Figure 2: (a)AlexNet training accuracy, (b) more details of AlexNetearlier training iterations

B. GoogleNet

Secondly, with the GoogleNet training, we can see in Figure 3(a) that the training accuracy made a quick rise, but now we can see that it has reached 100% even faster than AlexNet. There are no visible oscillations throughout the training except during the fist few iterations as illustrated in Figure 3(b). We also notice that the loss started decreasing earlier, not on the top of the scale, since the neural network itself already has some training, which is also a benefit from transfer learning.

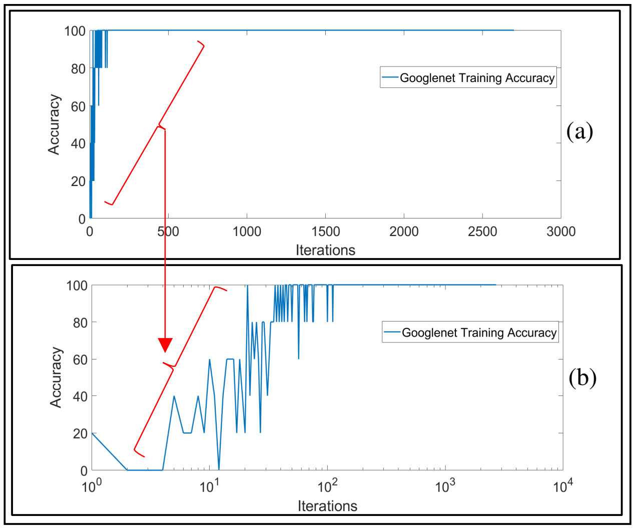

Figure 3: (a) GoogleNettraining accuracy, (b) more details of GoogleNetearlier training iterations

C. Vgg16

Figure 4(a) represents Vgg16 training accuracy at different iterations. We can see a little different performance, with more oscillations than the previous ones. This happens when too much similarity is present in the data. Since the data here had pictures of food that are really hard to differ, the accuracy was lowered at some times.

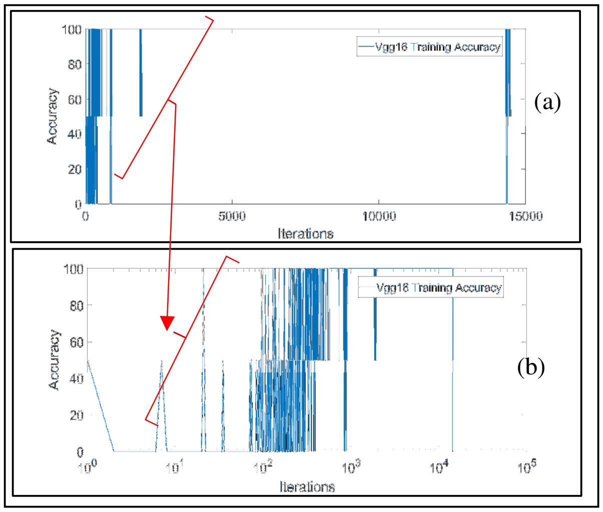

Figure 4: (a) Vgg16 training accuracy, (b) more details of Vgg16 earlier training iterations

4.2 Loss Evaluation

Loss function is an important part in Deep Neural Networks evaluation, which is used to measure the inconsistency between predicted value and actual label. For AlexNet model, we can observe from the Figure 5(a) that as accuracy increases, loss decreases. This is also a representation of the benefits from transfer learning.

For GoogleNet model, as shown in Figure 5(b), the decrease of loss indicate that the network learned rapidly. After it reaches 0 level, it is consistent at the bottom, with no visible oscillations. We can see that the training process of GoogleNet went somewhat better than training with AlexNet.

Lastly, Figure 5 (c) represents loss in training with VGG16. We can see that exactly at the points where accuracy increases, loss decreases. There is slightly more loss than in the previous graphs. Even though we can see some oscillations, the training still went well, and the final accuracy of the learning process is 100%.

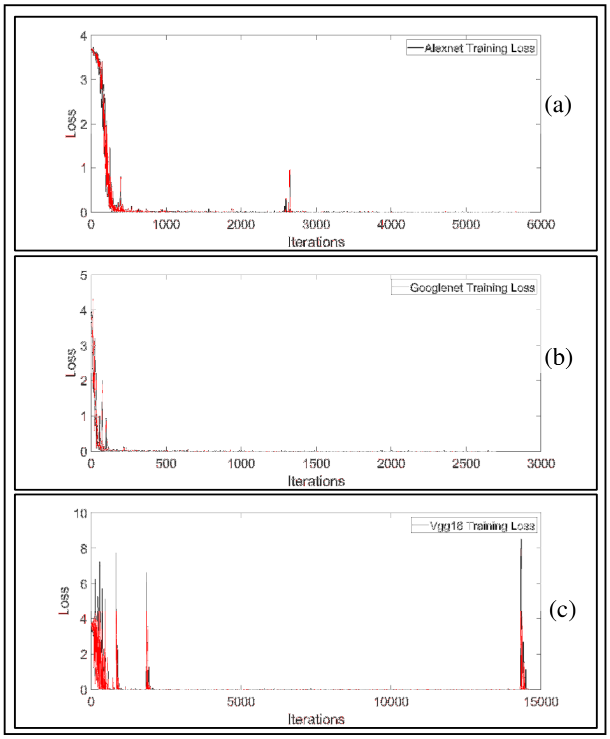

Figure 5: (a) AlexNet loss, (b) GoogLeNet loss, (c) Vgg16 loss

4.3 Validation

it is an important to show the validation of each model. As presented in Figure 6(a), the models shows different behaviours at the beginning of the traning, while the rest of the performace is similar for all models. Figure 6(b) represents the models performance in terms of validation loss. It shows additional evidence that all the models has different performance at the begining of the traning.

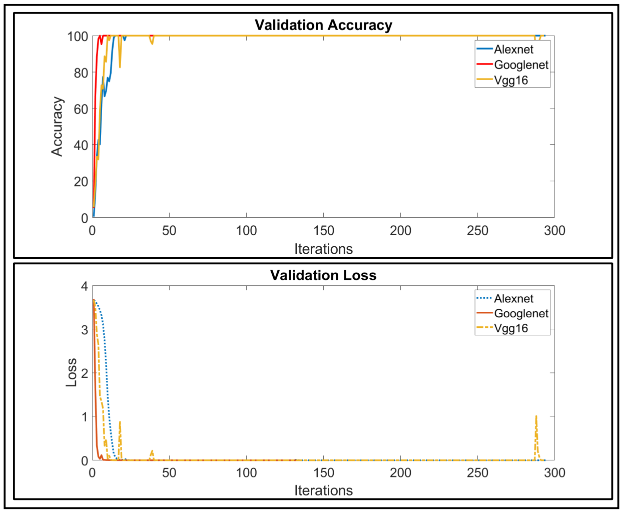

Figure 6: (a) Validation accuracy, (b) Validation loss

5. CONCLUTION

In this paper, we covered the utilization of three fine-tuned deep transfer learning models. AlexNet, GoogleNet and Vgg16 performed similarly in terms of training, validation and loss. The experiments shows the GoogleNet acheived higher scores faster than AlexNet and Vgg16. However, the fine-tuned models recorded 100% as training and validation accuracy. In the future, this work can be combained with currancy recognition to automat the casheir position and develop a fully self-check-in counter.

REFERENCES

[1] A. A. Almisreb and N. Jamil, "Automated ear segmentation in various illumination conditions," in *Proceedings - 2012 IEEE 8th International Colloquium on Signal Processing and Its Applications, CSPA 2012*, 2012.

[2] A. A. Almisreb, N. Jamil, and N. M. Din, "Utilizing AlexNet Deep Transfer Learning for Ear Recognition," in *Proceedings - 2018 4th International Conference on Information Retrieval and Knowledge Management: Diving into Data Sciences, CAMP 2018*, 2018.

[3] N. Jamil, A. A. Almisreb, S. M. Z. S. Z. Ariffin, N. Md Din, and R. Hamzah, "Can convolution neural network (CNN) triumph in ear recognition of uniform illumination invariant?," *Indones. J. Electr. Eng. Comput. Sci.*, 2018.

[4] A. A. Almisreb and M. A. Saleh, "Transfer Learning Utilization for Banknote Recognition: a Comparative Study Based on Bosnian Currency," *Southeast Eur. J. Soft Comput.*, 2019.

[5] Y. Kawano and K. Yanai, "Food image recognition with deep convolutional features pre-trained with food-related categories," *IEEE Int. Conf. Multimed. Expo Work.*, pp. 1–6, 2015.

[6] N. Martinel, G. L. Foresti, and C. Micheloni, "Wide-slice residual networks for food recognition," *Proc. - 2018 IEEE Winter Conf. Appl. Comput. Vision, WACV 2018*, vol. 2018-Janua, pp. 567–576, 2018.

[7] W. Min, L. Liu, Z. Luo, and S. Jiang, "Ingredient-guided cascaded multi-attention network for food recognition," *MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia*, 2019. .

[8] M. A. Subhi, S. H. Ali, and M. A. Mohammed, "Vision-Based Approaches for Automatic Food Recognition and Dietary Assessment: A Survey," *IEEE Access*, vol. 7, pp. 35370–35381, 2019.

[9] A. Myers *et al.*, "Im2Calories: Towards an automated mobile vision food diary," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, no. December, pp. 1233–1241, 2015.

[10] M. Taskiran and N. Kahraman, "Comparison of CNN Tolerances to Intra Class Variety in Food Recognition," *2019 IEEE Int. Symp. Innov. Intell. Syst. Appl.*, pp. 1–5, 2019.

[11] K. Yanai and Y. Kawano, "FOOD IMAGE RECOGNITION USING DEEP CONVOLUTIONAL NETWORK WITH PRE-TRAINING AND FINE-TUNING Keiji Yanai Yoshiyuki Kawano Department of Informatics , The University of Electro-Communications , Tokyo , Japan," 2014.

[12] Y. Matsuda, H. Hoashi, and K. Yanai, "Recognition of multiple-food images by detecting candidate regions," *Proc. - IEEE Int. Conf. Multimed. Expo*, pp. 25–30, 2012.

[13] N. Martinel, C. Piciarelli, and C. Micheloni, "A supervised extreme learning committee for food recognition," *Comput. Vis. Image Underst.*, vol. 148, pp. 67–86, 2016.

[14] H. J. Suh and K. H. Lee, "Real-time calorie extraction and cuisine classification through food-image recognition," *Int. J. Grid Distrib. Comput.*, vol. 11, no. 6, pp. 69–78, 2018.

[15] J. Chen and C. W. Ngo, "Deep-based ingredient recognition for cooking recipe retrieval," *MM 2016 - Proc. 2016 ACM Multimed. Conf.*, pp. 32–41, 2016.

[16] K. Hosozawa *et al.*, "Recognition of Expiration Dates Written on Food Packages with Open Source OCR," *Int. J. Comput. Theory Eng.*, vol. 10, no. 5, pp. 170–174, 2018.

[17] B. G. Rosa, S. Anastasova-Ivanova, B. Lo, and G. Z. Yang, "Towards a Fully Automatic Food Intake Recognition System Using Acoustic, Image Capturing and Glucose Measurements," *2019 IEEE 16th Int. Conf. Wearable Implant. Body Sens. Networks*, pp. 1–4, 2019.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2012.

[19] C. Szegedy *et al.*, "Going deeper with convolutions," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 1–9, 2015.

[20] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *Int. Conf. Learn. Represent.*, 2015.