Flower Classification with Deep CNN and Machine Learning Algorithms

Büşra Rümeysa Mete
Computer Engineering Department
Istanbul Kultur University
Istanbul, Turkey
b.mete@iku.edu.tr

Tolga Ensari
Computer Engineering Department
Istanbul University - Cerrahpasa
Istanbul, Turkey
ensari@istanbul.edu.tr

***Abstract*—Development of the recognition of rare plant species will be advantageous in the fields such as the pharmaceutical industry, botany, agricultural, and trade activities. It was also very challenging that there is diversity of flower species and it is very hard to classify them when they can be very similar to each other indeed. Therefore, this subject has already become crucial. In this context, this paper presents a classification system for flower images by using Deep CNN and Data Augmentation. Recently, Deep CNN techniques have become the latest technology for such problems. However, the fact is that getting better performance for the flower classification is stuck due to the lack of labeled data. In the study, there are three primary contributions: First, we proposed a classification model to cultivate the performance of classifying of flower images by using Deep CNN for extracting the features and various machine learning algorithms for classifying purposes. Second, we demonstrated the use of image augmentation for achieving better performance results. Last, we compared the performances of the machine-learning classifiers such as SVM, Random Forest, KNN, and Multi-Layer Perceptron(MLP). In the study, we evaluated our classification system using two datasets: Oxford-17 Flowers, and Oxford-102 Flowers. We divided each dataset into the training and test sets by 0.8 and 0.2, respectively. As a result, we obtained the best accuracy for Oxford 102-Flowers Dataset as 98.5% using SVM Classifier. For Oxford 17-Flowers Dataset, we found the best accuracy as 99.8% with MLP Classifier. These results are better than others' that classify the same datasets in the literature.**

***Index Terms*—machine learning; cnn; feature extraction; data augmentation; flower classification**

I. Introduction

Flowers are the most important producers of the earth that can grow in a wide variety of climates in terms of their habitats. They also keep on playing a very important role in the food chain by feeding almost all insect species in the world. In addition to this they play an important role in the food chain, and many drugs can be produced by using their healing properties. For such reasons, having a good knowledge of flowers and knowing their species is very important in terms of recognizing a new or rare plant species. Otherwise, many plants may be damaged because they are considered harmful to one's farmland or may be sold at very cheap prices. And all this occurs due to the inadequate recognition of the plant species. However, it is a real phenomenon that many of the plants grown in nature can be cultivated. In addition, increasing the recognition capacity of numerous endemic plant species, such as elecampane, verbascum thapsus whose life is limited to a specific area and grown only under special climatic conditions, will support the development of the pharmaceutical industry. Thanks to the studies in this field, little-known plants will be able to see the true value they deserve.

II. Related Work

In order to classify flowers, researchers were paying more attention to segmenting images and selecting them in some ways that can be called as artificial. This traditional method is more primitive today when development of the computer science and the technology -coming with it- have been getting improved unceasingly, because it is not running in a fully automatic process since it is required human interposition. Moreover, this method can't achieve high accuracy enough.

In [1] a visual vocabulary was introduced by the authors from Oxford University. This vocabulary includes many different features such as colour, shape and texture that discriminate between flowers from different classes. They also presented a study [2] combining four different features of flowers in order to classify them by using a multiple kernel framework with Support Vector Machine (SVM) classifier. So, they reached 88.33% accuracy on the quite challenging datasets of Oxford 102 Flowers which is also one of the dataset used in our study.

In the study [3], fine-grained classification was used by the authors on the same datasets used in our study that are Oxford 17-Flowers and Oxford-102 Flowers. In this study [3] they obtained 93.14% and 79.1 accuracy for the datasets of Oxford-17 Flowers and Oxford-102 Flowers, respectively.

In 2017, the study [4] used Google's pre-trained model named GoogLeNet with inception-v3 module for the purpose of classifying flowers. The authors selected two datasets same as the studies have mentioned above and also in our study: Oxford 17-Flowers and Oxford-102 Flowers. Although both datasets are not quite large, they have reached quite high accuracy of 95% and 94%, respectively.

In order to make a good comparison, we used the multi-class datasets prepared by Oxford Geometry Group, which are also used in other studies and also the most challenging datasets in this field. The original side of the study is that we use GoogleNet's inception-v3 module to extract features. Moreover, due to the lack of labeled data, we also used

978-1-7281-3789-6/19/$31.00 ©IEEE, 2019.

a data augmentation technique before the feature extraction phase. After dividing our data into training and testing, we classified the features we extracted using inception-v3 with various machine learning algorithms.

III. Dataset

In this study, we use 17-Flowers and 102-Flowers datasets belonging to Oxford Visual Geometry Group. Both of these datasets are composed of various flowers that are common in England. The 17-Flowers dataset contains 80 images from each flower group while 102-Flowers dataset contains at least 40 and up to 200 images for each category. Images have large scale, exposure and light variations. In addition to this, there are many classes that show great differences in the dataset as well as many classes with similarities to other classes. These well-known datasets are used for fine-grained recognition explained in Section I. Sample images of 20 species of flowers from Oxford-102 Flowers dataset can be seen in Fig. 1.

Fig. 1: Samples of 20 species of flowers in Oxford 102-Flower Dataset

IV. Proposed Classification System

The proposed classification system in the study includes four fundamental components: the modules of data, and data augmentation, and feature extraction, and classifier. It is shown in Fig. 2

A. Data Augmentation

We used one of the data augmentation techniques for the purpose of obtaining better performance on natural image classification to handle the disadvantages of the lack of labeled data. This method is very common as well as easy to diminish overfitting problem of scarceness of tagged data in flower classification. In the study, we used a particular data augmentation

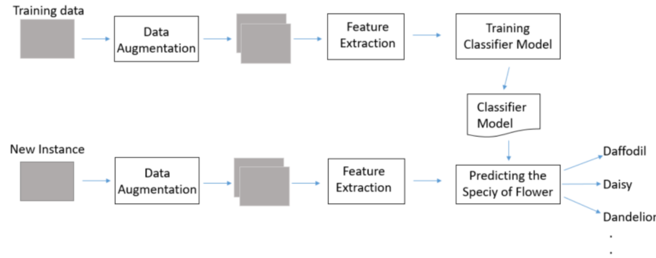


Fig. 2: Proposed classification system

Fig. 3: Samples of vertically flipped images

technique. We have doubled the number of images in both datasets by vertically flipping all the images.

B. Feature Extraction with Deep CNN

There are several CNN models has been introduced and utilized in the literature. For instance, in 2013, Min Lin and her colleagues proposed a novel deep network architecture [5]. The name of this new deep network architecture is "Network in Network" which is the name of the article as well. With this article, they have been the pioneers of the idea that linear filters in convolutional neural networks can increase the learning capacity with the usage of non-linear activation functions. They obtained the model by adding a multilayer sensor structure (perceptron) into the convolutional neural network which is in linear structure. This model was successful in data sets such as MNIST, CIFAR10, and CIFAR100. Furthermore, it is observed that the performance is increased for all datasets after data augmentation process and dropout process were applied. However, in 2015 with the article [6] named "Going Deeper with Convolutions", Google introduced the idea of going deeper with the architecture named GoogLeNet (inception-v1). With the article [6] published in 2015, a deeper and wider model was designed. They obtained the concept of width within the model by the module called as inception.

There are 4 modules for GoogLeNet model in the literature that are called as inception-v1 [6], inception-v2 [7], inception-v3 [7], and inception-v4 [8]. Inception-v3 includes two parts that are feature extraction part and classification part. Its feature extraction part includes convolutional neural network. On the other hand its classification part includes fully connected and softmax layers. All the mentioned layers in inception-v3 has shown in Fig. 4[1].

In this study, we used the previous module of GoogLeNet which is inception-v3 as a feature extractor by removing its classification part. In Fig. 4, feature extraction part and removed classification part of the module are seen. We mentioned the images in datasets in Section III and we fed them into the model that we used in the study after removing its classification layer to achieve the purpose of producing a group of tagged feature vectors.

Inception-v3 is a well-known architecture. The network's input must be an image with sizes of 299x299 pixels.

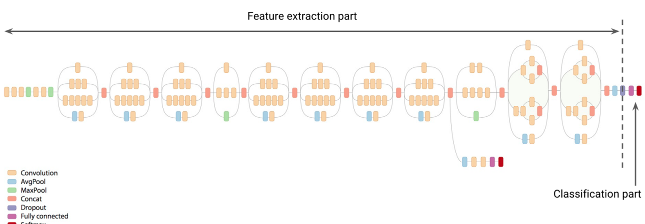


Fig. 4: Feature extraction and classification parts of Inception-v3.

C. Dimension Reduction

We experienced that it is very difficult to understand or discover the relationships between features for a multidimensional data set during the experiments. So we applied the dimension reduction process and showed its outputs. Note that it is just an informative step.

While Principle Component Analysis (PCA) is a linear feature extraction technique, t-SNE is a non-linear technique for dimensionality reduction and it is more suitable for the visualization of high-dimensional datasets [9].

In this study, after CNN features obtained and saved t-distributed stochastic neighbor embedding (t-SNE) technique was applied for an only informative reason. Fig. 5 demonstrates the transforming high dimensional features into two dimensional feature plane for the datasets Oxford-17 Flowers and Oxford 102-Flowers, respectively.

It is also seen the same color points are frequently clustered jointly in Fig. 5. In this way, it was foreboded that the features to train our classifier with superior accuracy.

[1]The figure has been taken from https://codelabs.developers.google.com in 06.09.2019

D. Support Vector Machine

Support Vector Machine (SVM) is one of the most effective and simplest supervised learning algorithm used in classification. However, it is often used in classification problems.

In Support Vector Machines, there is no prior knowledge or assumption about the distribution of input data. Input data can be separated linearly or non-linearly. In addition, there is no over-fitting problem in SVM. In artificial neural networks, over-fitting may occur unless cross validation is applied. Moreover, various kernel functions can be used to make inseparable problems detachable and to map data in better viewing space. Kernel-based algorithms are quite flexible. Because, the algorithm is independent from the hyper parameters such as learning rate and fixing parameters. The other reason for this is that it is sufficient to change the kernel function when the problem area changes [10].

The most used kernel functions such as polynomial kernel (1), linear kernel (2) and Gaussian/RBF (3) kernel were given in the specified equations.

$$K(x,y) = (x \times y + 1)^d \tag{1}$$

$$K(x,y) = x \times y \tag{2}$$

$$K(x,y) = exp(-\frac{||x-y||^2}{2\sigma^2}) \tag{3}$$

In this study, we obtained the highest success using SVM classifier. Having experienced different types of kernels such as polynomial, linear, Gaussian/RBF, we observed that the highest accuracy of 98,5% in 114,99 seconds on 102 Flowers Dataset. We obtained these results by using the LinearSVC library of sklearn with default values of its parameters.

E. Random Forest

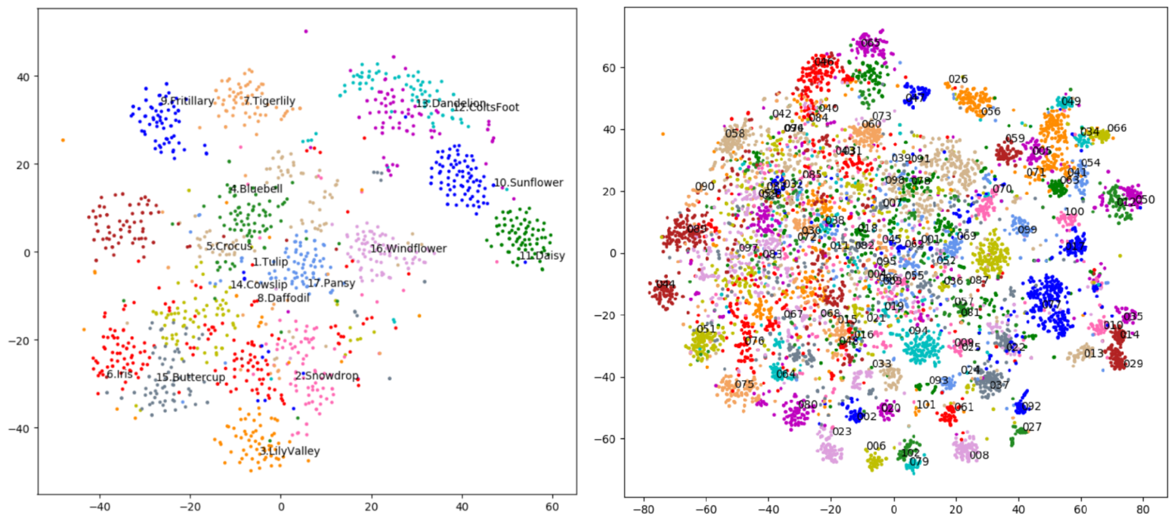
Random Forest is one of the popular models of machine learning, because it can be applied to both regression and classification problems. Furthermore, it gives good results without hyper parameter estimation. In Random Forest, it is aimed to increase the classification value by using more than one decision tree during the classification process. One of the biggest problems of decision trees is over-learning that can also called as memorizing data or more technically overfitting. To solve this problem, the random forest model selects and trains tens or even hundreds of different sets of randomly from both the data set and the feature set. With this method, hundreds of decision trees are created and each decision tree is individually estimated.

It is obtained the accuracy of 86.4% in 140.586323 seconds with Oxford 102 Flowers Dataset.

F. KNN

In KNN, there are many different distance formulas in finding distance such as Euclidean Distance, Manhattan Distance and Minkowski Distance. Their equations have given in the following Formulas 4, 5, 6, respectively.

$$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \tag{4}$$



a) Oxford 17 Flowers Dataset

b) Oxford 102 Flowers Dataset

Fig. 5: Transforming of the datasets into two dimensional feature plane without applying data augmentation

$$\sum_{i=1}^{k} |x_i - y_i| \tag{5}$$

$$(\sum_{i=1}^{k} (|x_i - y_i|)^q)^{1/q} \tag{6}$$

In the study, the best accuracy was obtained with manhattan distance as 96.2% in 41.943569 seconds. On the other hand, it was obtained the accuracy of 96.0% with minkowski distance in 39.699353 seconds. These findings were obtained with Oxford 102 Flowers Dataset.

G. Multi Layer Perceptron

Deep Learning Networks are feedforward neural networks with multiple layers. It is also known as Multilayer Perceptrons or MLPs. General feedforward deep learning networks do not include feedback. In a multilayer perceptron, the neurons are arranged into an input layer, an output layer and one or more hidden layers. We used ReLU as activation function in the study with other defaults of MLPClassifier library of sklearn.

We obtained the second highest accuracy of 96,5% in 101,58 seconds at Oxford-102 Flowers Dataset.

V. Comparative Results

For the purpose of focusing on the improvements gained with data augmentation, the study proposes the classification system which can be seen in Fig. 2. In this study, we used two datasets containing flower images with 17 and 102 classes. We conducted the all experiments on a computer with a NVIDIA GEFORCE GTX 960M display card.

From the experiments, we observed the most efficient finding for Oxford 17-Flowers Dataset with Multi-Layer Perceptron Classifier (MLPC) as 99,8% in 21,89 seconds. The second highest accuracy was gained with SVM Classifier as 99.6% in 12,01 second. The following highest accuracy was obtained from KNN classifier by using manhattan distance as 99.1% in 1.53 seconds. Finally, the lowest accuracy was obtained from the Random Forest classifier as 95,4% in 5,03 seconds. The findings were demonstrated in Fig.6

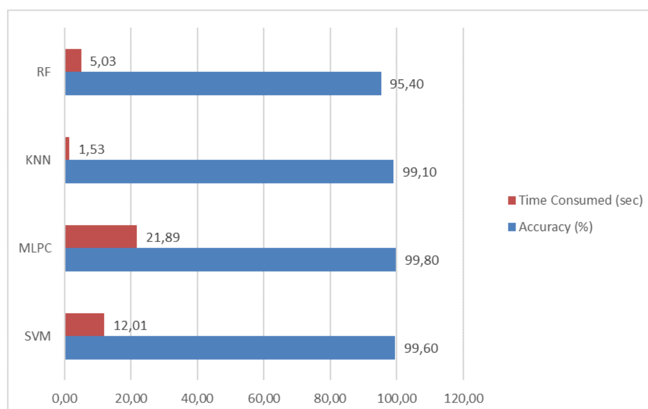


Fig. 6: Comparative results for Oxford-17 Flowers Dataset

Similiarly, we obtained the most efficient result for Oxford-102 Flowers Dataset with SVM classifier as 98.5% in 114,99 seconds. The second highest accuracy was gained with Multi-Layer Perceptron Classifier as 96,5% in 101,58 seconds. The following highest accuracy was obtained from KNN classifier by using manhattan distance as 96,2% in 39,58 seconds. Finally, the lowest accuracy was obtained from the Random Forest classifier as 87,6% in 150,06 seconds. The findings were demonstrated in Fig.7

In figure 8, we demonstrated the comparative results for Oxford-17 Flowers and Oxford-102 Flowers datasets. As seen in the figure the highest accuracy rates belong to the classifiers named SVM and Multilayer Perceptron.

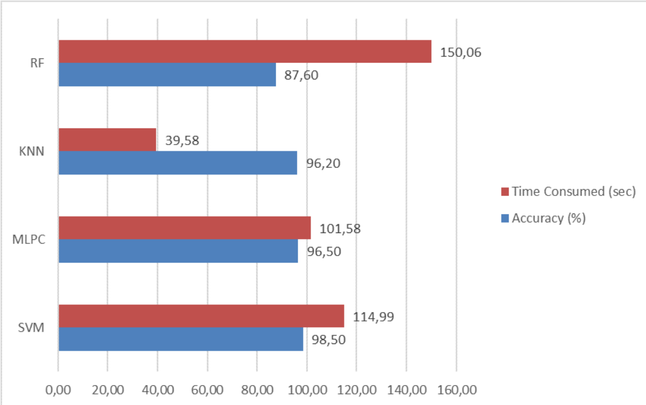


Fig. 7: Comparative results for Oxford-102 Flowers Dataset

Acknowledgement

This study derives from the author's master thesis. It is supported by Technology and Project Support Unit of Istanbul Kultur University.

VI. Conclusion

In visual recognition area, fine-grained recognition is known as a kind of a challenge that has been getting more popular especially in recent years. In this study, that kind of a challenge -which is required developing algorithms in order to selecting classes with quite small subtle visual distinctions- has been carried out. For this, two benchmark datasets including different types of flowers mostly common in England were used: Oxford 17-Flowers [1] and Oxford-102 Flowers [2].

In this study, with the proposed classification system and two flower datasets containing many classes (17 and 102), we obtained the highest result in the literature. Thus, with the pre-trained model, we observed the effect of both feature extraction and data augmentation on accuracy. In addition, we used t-sne [9] to visualize the images and how the features extracted from the images in the dataset are distributed. In addition, the classification system that we propose in this study can be used for different problems in image classification. In this respect, the study contributes not only to the flower classification area but also to other areas where labeled data are scarce.

As seen in the figure 8, it can be said that, the performance rates decrease in all classifiers when the dataset grows. In this context, we found that the performance rate decrease in SVM is less than the decrease in the Multilayer Perceptron Classifier.

In all algorithms, it is seen that time consumption also increases as the data set grows. When the two algorithms with the highest performances are compared in terms of time consumption. At this point, we found that the performance decrease in the Multilayer Perceptron Classifier is less than that of SVM.

VII. Future Work

In the future, we will apply our approach in this study to a dataset containing different categories of images, discussing

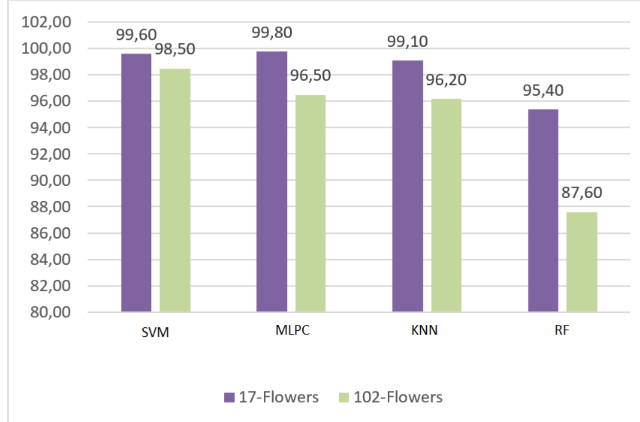


Fig. 8: Accuracy rates obtained by different classifiers for Oxford 17 and 102-Flowers datasets

the consistency of our results in detail. So we are to expand our findings.

References

[1] M.-E. Nilsback and A. Zisserman, "A visual vocabulary for flower classification," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2. IEEE, 2006, pp. 1447–1454.

[2] Z. Nilsback, Maria-Elena and Andrew, "Automated flower classification over a large number of classes," in *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*. IEEE, 2008, pp. 722–729.

[3] A. Angelova and S. Zhu, "Efficient object detection and segmentation for fine-grained recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 811–818.

[4] X. Xia, C. Xu, and B. Nan, "Inception-v3 for flower classification," in *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*. IEEE, 2017, pp. 783–787.

[5] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.

[6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[7] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[8] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[9] G. Hinton and Y. Bengio, "Visualizing data using t-sne," 2017.

[10] J. Kremer, K. Steenstrup Pedersen, and C. Igel, "Active learning with support vector machines," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 4, no. 4, pp. 313–326, 2014.