The Effects of Augmented Training Dataset on Performance of Convolutional Neural Networks in Face Recognition System

Mehmet Ali Kutlugün
Computer Science and Engineering
Istanbul Sabahattin Zaim University
Istanbul, Turkey
mehmet.kutlugun@std.izu.edu.tr

Asst.Prof. Yahya Şirin
Computer Science and Engineering
Istanbul Sabahattin Zaim University
Istanbul, Turkey
yahya.sirin@izu.edu.tr

MehmetAli Karakaya
Management Information Systems
Anadolu University
Eskişehir, Turkey
mehmetalikarakaya@aof.anadolu.edu.tr

Abstract— Nowadays, deep learning methods have been used in many areas such as big data analysis, speech and image processing with the increasing processing power and the development of graphics processors. In particular, face recognition systems have become one of the most important research topics in biometry. Light direction, reflection, emotional and physical changes in facial expression are the main factors in face recognition systems that make recognition difficult. Training of the system with the available data in small data sets is an important factor that negatively affects the performance. The Convolutional Neural Network (CNN) model is a deep learning architecture used for large amounts of training data. In this study, a small number of employee images set of a small-scale company has been increased by applying different filters. In addition, it has been tried to determine which data augmentation options have more effect on face recognition. Thus, non-real-time face recognition has been performed by training with new augmented dataset of each picture with many features.

Keywords— *deep learning; convolutional neural networks; image processing; face recognition; data augmentation*

I. INTRODUCTION

A. Deep Learning

Deep learning is a series of machine learning algorithms that try to model high-level abstractions of data with model architectures from multiple nonlinear transformations. Deep learning systems are used in many areas such as big data analysis, speech recognition, image classification, pedestrian recognition, generic visual recognition and face recognition along with improvements in power of processor and in graphics processors [1].

Deep learning, a class of machine learning uses many nonlinear processing unit layers for feature extraction and conversion. Each subsequent layer calculates the output of the previous layer as input [2].

Algorithms can operate supervised or unsupervised. Deep learning is based on learning with the representation of data. For example; when a representation is represented for an image, properties such as a vector or edge sets of density values per pixel may be considered. Some of these features represent data better, while others may represent worse. Deep learning methods are advantageous by using effective algorithms for hierarchical feature extraction that best represent data rather than manually extracted features [3,4].

The difficulty of generalizing the new examples becomes even more difficult when working with high dimensional data,

and the mechanisms used to achieve generalization in traditional machine learning are not sufficient to learn complex functions in high dimensional areas. Such areas also often require high computational costs. Deep learning is preferred to overcome these and other obstacles [5]. Particularly in the case of large amounts of training data, the deep learning method is an area of considerable work [6-8].

B. Convolutional Neural Networks (CNN)

In recent years, Convolutional Neural Network (CNN) models, which are deep learning models as an alternative to traditional feature extraction and artificial neural network methods, have started to be developed. CNN-based methods are preferred in order to reduce the effects of different species such as exposure, illumination, facial expressions and better features in image analysis [9].

The CNN has a convolution layer, nonlinearity layer and pooling layers and a flow from simple attributes to complex attributes. Class estimates are obtained as CNN outputs. The filters used in CNN's convolution layers update themselves to minimize the error throughout the training. As the filters are updated, the attributes obtained in the layer outputs also differ. This iterative improvement throughout the training phase provides more distinctive features [1]. The success of this method has been proven, especially in the case of processing large data [6,7,10].

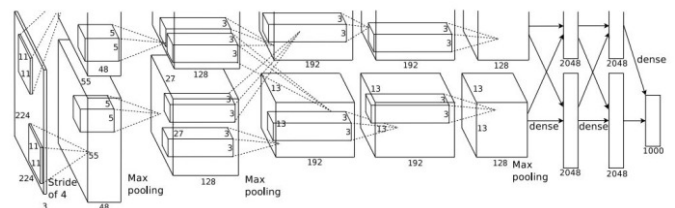Figure 1 shows a generally accepted CNN model AlexNet architecture.

Figure. 1. AlexNet Architecture [11].

II. PREVIOUS STUDIES

With the increasing amount of data, especially in the image processing area, problems such as image classification and object recognition stand out. Deep learning methods that prove themselves as a method that can cope with large data are frequently used and give the most successful results [4]. However, it is important to increase the amount of data without compromising the integrity of the data format for application areas with a limited set of data.

In their study, Si Liu et al. increased the image size to 1 and 1.2 times in order to increase the sensitivity rate in the cropped human-centered regions and to reduce the excessive compliance, with the filters such as horizontal reflection, 4 variations of each image. However, this study was carried out to include only the image area, and no facial separation was performed [12].

In their work, Vittorio et al. produced a vastly increased image collection to enrich the distinctive features of the painting and tried to eliminate the constraints such as lighting, exposure, facial expression, and low resolution [13]. In their study, Buslaev et al. performed a fast and efficient album creation process by performing multiple image conversion with image enhancement method [14].

Galdran et al. have added color coherence algorithms to the data sets of convolutional neural networks based on convolutional neural network studies for segmentation and classification of skin lesion analysis [15]. Cengil and Çınar performed image classification with CNN model on 8 different images selected from CIFAR-100 data set. In order to be used in the training process of the study, they trained the system they designed for 9 hours with the data increase process [16].

Doğan and Türkoğlu, with deep learning, designed the leaf classification for the model, AlexNet, Vgg16, Vgg19, ResNet50, GoogLeNet algorithms were used to compare the performance. They stated that if the amount of data increases, the performance will increase but the processing time will increase accordingly [17].

In order to increase classification performance in his work, Salman produced artificial data sets by adding different levels of noise through the existing training dataset [18].

In the above mentioned studies, it was emphasized that increasing the data set will increase the educational performance of the system but it is not explained how a performance difference is between the original data set and the increased data set. Increasing the data improperly depending on the type of data used and the application purpose may not yield positive results. The filters used to increase the data set and the proper size of these filters are very important for determining the characteristics.

In this study, on the data set containing few samples; firstly, the classification is made in its original form, then, applying different filters to the data set, the classification has made with the increased new data set. Thus, the performance of data augmentation on the face recognition system has revealed. In addition, it has been tried to determine which filters give more effective results in non-real-time face recognition.

III. PRESENTED STUDY

A. Aim

The purpose of the study is to obtain training data for use on the CNN model, which is designed to obtain a large number of samples from the face images containing a small number of samples in the source dataset without disturbing the integrity of the original image. This will reveal how the non-real-time face recognition system is affected by the classification performance. The process diagram of the proposed study is shown in Figure 2.
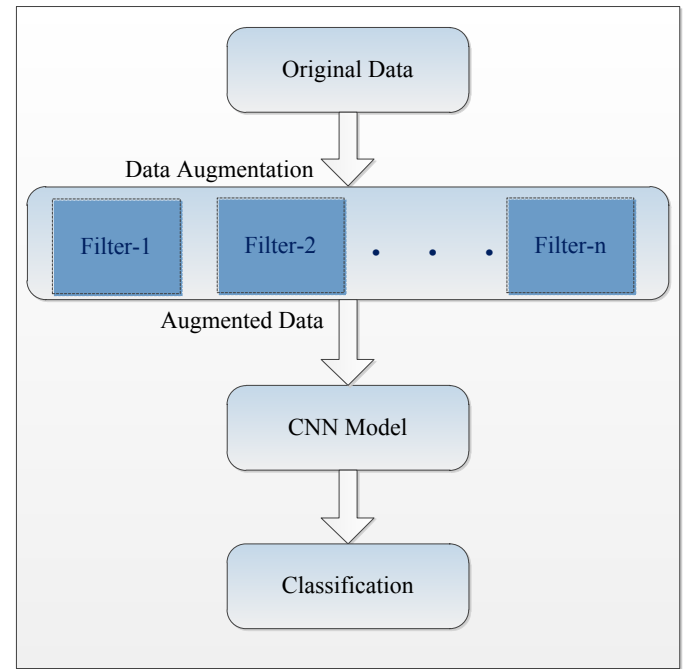

Figure. 2. Process Block Diagram

B. Method / Architecture

Deep learning architectures include CNN. CNN is a special type of feed forward neural network. The main characteristic of this method is that it can learn features at various levels by providing a more abstract representation of the input data. Thus, it can be applied to the network without attribute extraction. Therefore, a CNN is an end-to-end classifier. Another advantage is the structuring of feature extractors based on data used for training [19].

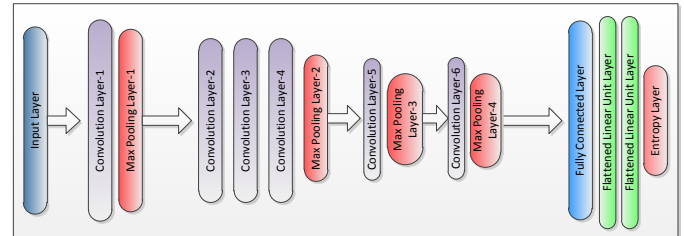The processes of the method applied in this study are shown in Figure 3.

Figure. 3. Applied CNN Model

IV. EXPERIMENTAL SETUP

A. Data Sets

150 face images of a small-scale company have been used for the data set. According to the principle of confidentiality of the data, the pictures are not used in their original form. In addition, we have compared the random picture of 150 people with the data set, known as Labeled Faces in the Wild (LFW) [20].

1) Data Preprocessing

All images have taken to 256x256 pixels and the image to be placed on the input layer has obtained in 256x256x3 size.

Data Augmentation

The best way to generate a machine learning model, perform better generalization is to train the model using more

data. However, the amount of data we have in practice is limited. One way to overcome this problem is to create artificial data and incorporate it into the training set. This approach is the easiest method for classification. The task of a classifier is to take a complex, high-dimensional input x and summarize it with a single category y. That is, the main task of a classifier is to be unchanged against a wide variety of transformations. By converting the x entries in the training set, new *(x, y)* binaries can be created easily [5]. In this way, the variety of data is increased by artificial methods and it affects the learning performance to a great extent. With this process, the pictures are subject to some distortion. On the image, it is attempted to increase the variety of pictures by performing operations such as angular rotation, changing perspective, scrolling and zooming [21].

Many processes, such as rotating or scaling the image, have been shown to be effective in enhancing educational performance. However, it is necessary to avoid making transformations that may perceive the correct class as false. For example, in an optical character recognition system, the letters b and d or numbers 6 and 9 may cause misclassification because they can be likened to each other. It is not appropriate to use data augmentation methods such as horizontal translation or rotation at 180-degrees angle for these systems [5].

B. Implementation Setup

The application has been implemented using the Keras library on the Python platform. To increase the data, Keras's ImageDataGenerator parameters have been used. Using the data enhancement techniques, the number of pictures with separate filters for each face image is gradually increased from 150 to 3300, 6300 and most recently to 9450.

These data sets; angular rotation, scrolling, cropping, zooming, turning filters have been applied and k-nearest neighbor algorithm has chosen to fill the resulting gaps.

TABLE I. AUGMENTED DATA SAMPLES

C. Calculation of Performance Ratio

The performance of the system has been measured by the accuracy percentage of the training set. The results are presented in Tables I and II. The data quantity and performance ratios obtained from the incremental increase of the original data set are presented in Table II. No data augmentation is applied to compare to 150 images randomly selected from the LFW dataset given in Figure 4. The performance graphs of the original dataset, which are incrementally augmented, are given in Figure 5 and Figure 6.

TABLE II.   SUCCESS EVALUATION TABLE

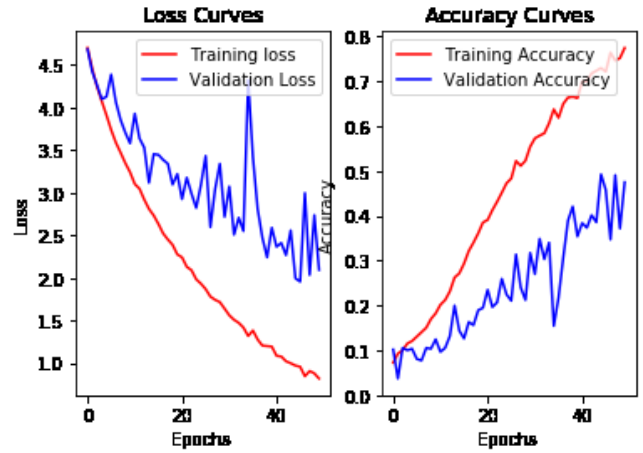| Total Number of Data	Training Time	Performance Rate (%)
150	58sec.	50
3300	20min.	62
6300	44min.	66
9450	1h. 4sec.	76

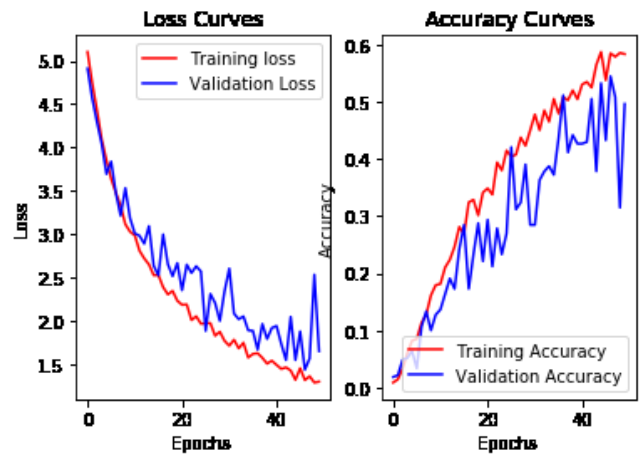Figure. 4.   Loss and Accuracy graphs of non- augmented LFW data set.

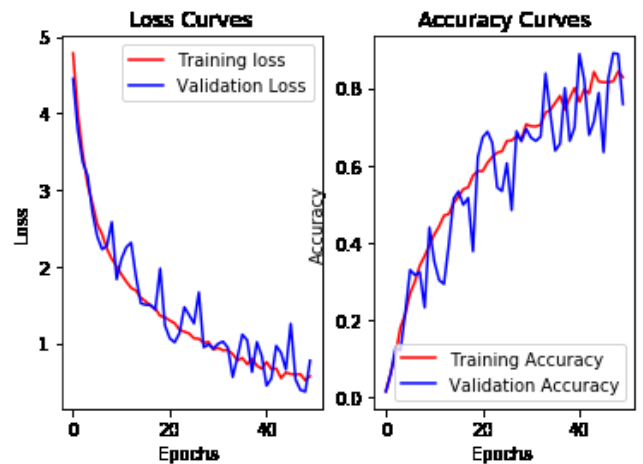Figure. 5.   Loss and Accuracy graphs of partial augmented dataset.

Figure. 6.   Loss and Accuracy graphs of comprehensive augmented dataset.

V. Conclusion and Evaluation

Deep learning systems are used in many areas such as big data analysis, speech recognition, image classification, pedestrian recognition, generic visual recognition and face recognition along with improvements in power of processor and in graphics processors. In particular, in order to extract better attributes in image analysis, traditional feature extraction and an alternative to artificial neural network methods, Convolutional Neural Network (CNN) models, which are deep learning models, have started to be developed.

It is important to increase the amount of data without disrupting the integrity of the data for application areas with limited data set. As the amount of data we have in practice is limited, one way to overcome this problem is to create artificial data and increase this data.

In this study, angular rotation, scrolling, cropping, zooming and turning filters are applied on a small number of data sets and the source data set is increased without disturbing the integrity of the original data. In addition, it has been tried to determine which data augmentation options have more effect on face recognition. Thus, non-real-time face recognition has been performed by training with new augmented dataset of each pictures with many features.

Experimental results show that the performance of the training is significantly increased depending on the amount of data. The filters used in this study, especially angular rotation and brightness filters have more effect on success. Because the LFW dataset contains more than one image for some people, no data augmentation is applied to this data set. This dataset has been used to compare with the augmented original dataset. According to these results, the LFW data set should also be increased for more performance. Although the original data set contains a sample for each person, face recognition is more efficient because only one person is in the pictures. However, the performance achieved with some filters is more effective than others. For example, horizontal flip or mirroring filters may be more suitable instead of vertical flip filter. At the same time, the parameter values used are also very effective. Excessive angular rotation for the data set used did not yield positive results.

VI. Future Studies

In order to further increase training performance, the appropriate filters can be designed to produce more data with reasonable parameter values and new models can be created with AlexNet, ZFNet, GoogLeNet, Microsoft RestNet, R-CNN architectures that can be considered successful in image classification. However, it should be taken into consideration that the training period will increase in proportion to this increase for very high data numbers. In addition, the effects of the filters can be discussed by using different filters than the filters used in this study.

References

[1] Bilgiç, A. et al., *"Face recognition classifier based on dimension reduction in deep learning properties."* Signal Processing and Communications Applications Conference (SIU), 2017 25th. IEEE, 2017.

[2] L. Deng and D. Yu, *"Deep Learning: Methods and Applications"* Found. Trends® Signal Process., vol. 7, no. 3–4, pp. 197–387, 2014.

[3] H. A. Song and S.-Y. Lee, *"Hierarchical Representation Using NMF"* in International Conference on Neural Information Processing, pp. 466–473, 2013.

[4] Şeker, A. et al., *"A Study on Deep Learning Methods and Applications",* Gazi Journal of Engineering Sciences, 3.3: 47-64,2017.

[5] Ian Goodfellow et al. *"Deep Learning"*, MIT Press, 2016.

[6] Çalık, N. et al., *"Signature recognition application based on deep learning"* Signal Processing and Communications Applications Conference (SIU), 2017 25th. IEEE, 2017.

[7] Ranzato, Y. M. et al., *"Sparse Feature Learning for Deep Belief Networks"*, Proc. Adv. Neural Inf. Process. Syst., vol. 20, pp. 1185-1192, 2007.

[8] Scherer, D. et al., *"Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition"*, International Conference on Artificial Neural Network, pp. 92-101, 2010.

[9] Yang, Hu et al., *"When Face Recognition Meets with Deep Learning: An Evaluation of Convolutional Neural Networks For Face Recognition"*, In Proceedings of the IEEE International Conference on Computer Vision Workshops, pp142-150, 2015.

[10] Chen, X. W. and Lin X., *"Big Data Deep Learning: Challenges and Perspectives"*, IEEE, vol. 2, pp. 514-525, 2014.

[11] Krizhevsky, A. et al., *"Imagenet classification with deep convolutional neural networks"* in Advances in Neural Information Processing Systems 25, pp. 1097–1105. Curran Associates, Inc.,2012.

[12] Liu, Si et al., *"Matching-cnn meets knn: Quasi-parametric human parsing"* Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.

[13] Vittorio, C. et al., "Robust Single-Sample Face Recognition by Sparsity-Driven Sub-Dictionary Learning Using Deep Features", *Sensors* 19.1:146, 2019.

[14] Buslaev, A. et al., *"Albumentations: fast and flexible image augmentations"* arXiv preprint arXiv:1809.06839, 2018.

[15] Galdran, A. et al., "Data-Driven Color Augmentation Techniques for Deep Skin Image Analysis", arXiv preprint arXiv:1703.03702, 2017.

[16] Cengil, E. and Çınar, A., *"A New Approach for Image Classification: Convolutional Neural Network"* European Journal of Technic 6.2,2016.

[17] Doğan, F. and Türkoğlu, İ., *"Comparison of Leaf Classification Performance of Deep Learning Algorithms",* Sakarya University Journal of Computer and Information Sciences 1.1: 10-21, 2018.

[18] Salman, M., *"Integration of Hyperspectral and Lidar Data in Attribute and Decision Levels and Classification with Deep-Curvilinear Neural Networks"* Master Thesis, Hacettepe University Institute of Science and Technology, 2018.

[19] Perlin, H. A. and Lopes, H. S., "Extracting Human Attributes Using A Convolutional Neural Network Approach", Pattern Recognition Letters, Vol. 68, pp. 250-259, 2015.

[20] Huang, G.B.; Ramesh, M.; Berg, T.; Learned-Miller, E. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments; Technical Report 07-49; University of Massachusetts: Amherst, MA, USA, 2007.

[21] Şahin, Ö., "*TL recognition for visually impaired people on iOs platform*", Master Thesis, Selçuk University Institute of Science and Technology, 2017.