Gender Classification Using Deep Learning Techniques

Sahra Tilki
Department of Computer Engineering
Istanbul Sabahattin Zaim University
Istanbul, Turkey
sahra.tilki@izu.edu.tr

Hasibe Büşra Doğru
Department of Computer Engineering
Istanbul Sabahattin Zaim University
Istanbul, Turkey
hasibe.dogru@izu.edu.tr

Alaa Ali Hameed
Department of Computer Engineering
Istanbul Sabahattin Zaim University
Istanbul, Turkey
alaa.hameed@izu.edu.tr

Akhtar Jamil
Department of Computer Engineering
Istanbul Sabahattin Zaim University
Istanbul, Turkey
akhtar.jamil@izu.edu.tr

Jawad Rasheed
Department of Computer Engineering
Istanbul Sabahattin Zaim University
Istanbul, Turkey
jawad.rasheed@izu.edu.tr

Erdal Alimovski
Department of Computer Engineering
Istanbul Sabahattin Zaim University
Istanbul, Turkey
0000-0003-0909-2047

Abstract— Gender classification from face images is a challenging task due to presence of complex background, object occlusion, and variations in illumination conditions. Face images can be exploited for various applications such as expression analysis, recognition and tracking. In this paper, two deep learning-based methods are investigated for gender classification using face images. These methods include: convolutional neural network (CNN) and Alex Net. Experiments were performed to evaluate the performance of both models for identification of male and female classes from face images. Results show that both methods were effective for gender classification. Moreover, a comparative analysis was also performed between these two models and some of the popular methods for gender classification.

Keywords— Gender classification, gender recognition, AlexNet, CNN, deep learning.

I. INTRODUCTION

In people living in communities, gender is an important factor in the interaction between individuals. With the development of technology, the use of smart devices has increased and social media has started to attract attention by everyone. However, day-to-day studies on gender recognition have gained importance and number of applications using such techniques have increased [1]. Facial images are widely used in such applications as they provide useful information that can be exploited for extracting human interaction. Gender classification approach by using facial images generally consists of image processing, feature extraction and classification stages. These stages may vary depending on the purpose of the study and the features of the methods to be used. Therefore, the classifier, method and extracted features have a significant effect on the performance of the study.

Deep learning techniques are now widely used for various tasks such as classification, automatic feature extraction, object recognition etc., due to their high classification accuracy. Motivated from other fields, researchers have also exploited deep learning methods for gender prediction and classification from facial images. The following paragraphs present a summary of some studies.

Janahiraman and Subramaniam [2] aimed to make gender classification using different models of CNN architecture. A dataset was created from facial images from Malaysians and some Caucasians people. Their results achieved accuracy of 88% with the VGG-16 model, 85% with the ResNet-50 model and 49% with the Mobile Net model.

Akbulut et al [4] performed gender recognition from facial images using Local Recipient Areas-Excessive Learning Machine (LRA-ELM) and CNN architecture. The experiments were carried out using approximately 11 thousand images from the Adience dataset for age and gender recognition [3]. The proposed method resulted in 80% and 87.13% accuracy with LRA-ELM and CNN respectively.

In [5], a comparative analysis was perfomed among proposed method, Alex Net and VGG-16 models for gender classification including women, men, old, young, children, and babies. Experimental results show accuracy of 72.20% with the proposed CNN model, 99.41% with the VGG-16 model, and 65.63% with the AlexNet model.

Arora et al [6] proposed a CNN model for gender classification with facial images. In experimental studies, 1500 images were used for training and 1000 images were selected from the CASIA database and verified. As a result of the experiments, 98.5% accuracy was achieved.

In [7], a basic convolutional network architecture was proposed to increase the performance of automatic age and gender classification. This technique also produce optimal results even in presence of limited training data. High classification accuracy was obtained using this model on Adience database [3].

Ranjan et al [8] proposed a method called HyperFace for simultaneous facial recognition, pose prediction, and gender recognition using CNNs. This method combines the middle layers of a deep CNN using a separate CNN and then a multitasking learning algorithm was employed on the fused features. HyperFace-ResNet is based on the ResNet-101 model to increase algorithm speed and Fast HyperFace variants to create zone recommendations have been proposed.

Raza et al. proposed a deep learning method for predicting the gender of pedestrians. A preprocessing step was used to segment the pedestrian from the image. Then, stacked auto encoders with softmax classifier were used for classification. Accuracy rates of 82.9%, 81.8% and 82.4% have achieved in the anterior, posterior and mixed views in the MIT dataset, respectively, and approximately 91.5% in the PETA dataset [9].

In [10], Abdalrady and Aly presented the exchange of classical CNN models with the PCANet model for gender classification. In addition, the network architecture size was also reduced with PCANet in complex CNN models. this method achieved 89.65% accuracy for gender classification.

Yu et al [11] proposed a low-complexity CNN model with few layers. This method achieved 91.5% accuracy on a dataset consisting of 1496 whole body images.

In this study, a gender classification method is proposed using deep learning techniques. Specifically, CNN and AlexNet model will be employed for gender classification from images. Moreover, a comparative analysis will be performed between our method and state-of-the-art methods.

The rest of the paper is organized as follows. Section II provides a detailed account of the proposed method. Section III presents the obtained results. Finally, the paper will be completed with concluding remarks.

II. METHODOLOGY

A. Deep Learning

Deep learning is an artificial intelligence (AI) technique that aims to mimic human brain by learning from experience. In other words, it is a method that realizes the learning process by aiming to discover the hidden representation of the data [12]. These representations are learnt through a training process. For instance, to learn how to recognize an object, it is necessary to train the program with many object images that we label according to different classes. This training can take hours, days or even weeks. Generally, deep learning-based approaches need large amount of training data and take longer time for training compared to the conventional machine learning methods.

When trying to recognize any object or character on an image, finding unique properties is a time-consuming and difficult process as there are so many properties on the object or character. At this point, unlike classical machine learning, in which features are extracted manually, problems can be overcome with deep learning techniques that automatically extract the relevant features from the data. Deep learning is a neural network with many hidden layers, and these hidden layers can be thousands or millions. After an image is trained over the network, they can form complex concepts from simple concepts. When an image is trained in the network, it can learn objects such as characters, faces, cars by combining simple features such as shape, edges and corners. As the image passes through the layers, each layer learns a simple feature before moving on to the next layer one by one, as the layers increase, the network can learn more and more complex features and finally combine them to predict the image.

Deep Learning methods have proven their importance by achieving remarkable success in Natural Language Processing (NLP), Optical Character Recognition (OCR), Computer Vision (CV), Image Processing, Object Recognition and Classification.

B. Convolutional Neural Networks (CNN)

CNN, one of the deep learning techniques, is a powerful neural network. It is widely used as a solution to problems that may be encountered in areas such as Computer Vision and Image Processing. CNN can replace input data with trainable parameters in each layer and also make accurate assumptions about the nature of the images.

CNN architecture consists of five main types of neural layers; convolution layer, activation layer, pooling layer, fully connected layer and dropout. Each layer type plays a different role. Each layer of CNN converts the input volume into an output volume of neuron activation and eventually delivers it to fully connected layers. While simpler features such as edge information are obtained in the first layers, more complex features representing the image are obtained in deep layers. In the following section, the operations performed on CNN layers are explained in detail.

1)Convolutional Layer: This layer forms the basis of CNN. The transformation process is performed by circulating a filter that can have different sizes such as 3 * 3, 2 * 2 on the image. Filters apply a convolution process on the images coming from the previous layer in order to generate the output data and as a result of this process, the activation map is formed. We can explain the resulting activation map as the regions in which each filter has its own properties. During the training, the coefficients of the filters are updated for each learning in the training set, and thus it is determined which regions of the data are important to determine the features. Simple features of the image used such as edges are usually calculated in the first layers of the CNN model [13].

2)Activation Layer: The network has a linear structure due to the mathematical operations performed in the convolution layer. As a result of the application of the activation functions used in the activation layer, the network becomes a nonlinear structure. Thus, faster learning of the network is provided. It is very important to choose activation functions in a neural network architecture. The most commonly used of these are the sigmoid function, which is usually used in classification problems, Softmax, which is a generalization of the Sigmoid function for multiple categories, and the Rectified Linear Units Layer (ReLu), which is preferred as the activation function in most studies.

3)Pooling Layer: The layer used for size reduction in CNN architectures is the pooling layer. Information loss may occur as a result of the size reduction process, but these losses are beneficial for the network. Because the reduction in size provides less computational load for the upcoming layers of the network, it also works against network overfitting. There are two most commonly used types, average pooling and maximum pooling

4)Fully Connected Layer: Neurons in this layer are fully connected to all activations in the previous layer. As a result of these layers, two-dimensional feature maps are transformed into one-dimensional feature vector. The derived vector can be included in a certain number of categories for classification or used as a feature vector for further processing.

5) DropOut: It is one of the most used networking techniques in deep learning [14]. When training is done using big data in CNN, the network can overfitting. The basic logic based on removing some nodes in the network prevents memorization from occurring.

The ImageNet Large Scale Visual Recognition Competition (ILSVRC) is one of the largest competitions in the field of object recognition. The winning CNN models can be listed as AlexNet, Le Net, ZF Net, VGG-16, GoogLe Net and Microsoft RestNet. These models play an important role in understanding and enhancing deep learning and neural networks. Therefore, it is preferred in many studies.

Le Net: It is the model that gave the first successful result of the competition and was published in 1998. Postal numbers were created in order to read the numbers on bank checks [15].

Alex Net: It is the winning model of the competition held in 2012. Developed by Krizhevsky, Sutskever and Hinton. Successive convolution layers consist of fully connected layers using activation functions such as maximum pooling and Relu and Sigmoid [16].

ZF Net: This model, which is an improved version of the Alex Net architecture, won the competition in 2013. Unlike the Alex Net architecture, 7x7 size filters were used instead of 11x11 filters in the first layer and a 2-step slip amount in the pooling layer. Thanks to these differences, it is ensured that many original pixel information at the input size can be preserved [17].

VGG-16: It is a model developed for better results than the results obtained in previous years. The model, which has a very smooth architecture, was the second in the competition in 2014. It is similar to Alex Net and has many filters [18].

GoogLe Net: It is the model that won the competition in 2014. It has an architecture similar to Le Net, but differently it is the implementation of a new element called the starter module. Modules linked in parallel were used to reduce the probability of overfitting [19].

ResNet: The winning model of 2015 is the first network structure consisting of Residual blocks with 34 layers. It differs from all other architectures due to its design in a deeper structure [20].

III. Experiments and Result

In this study, a gender classification method is proposed using facial images with CNN models. The accuracy obtained by using the CNN model and Alex Net architecture were compared with the results of similar studies conducted in the literature.

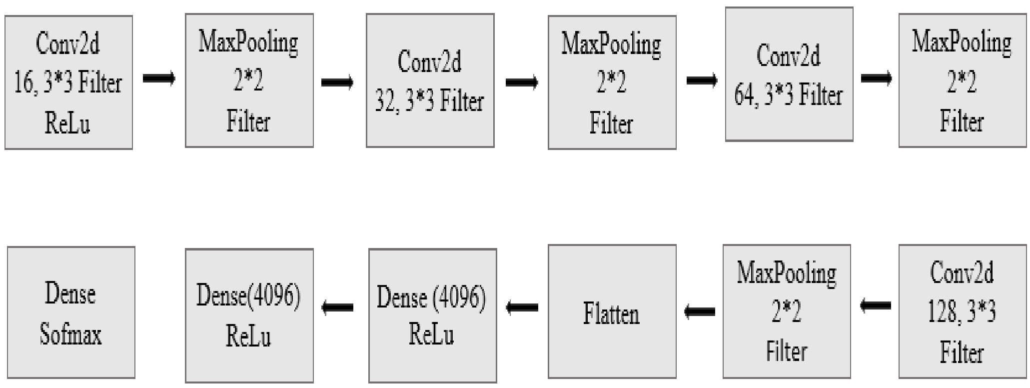

Fig. 1: Layer information of the created CNN model.

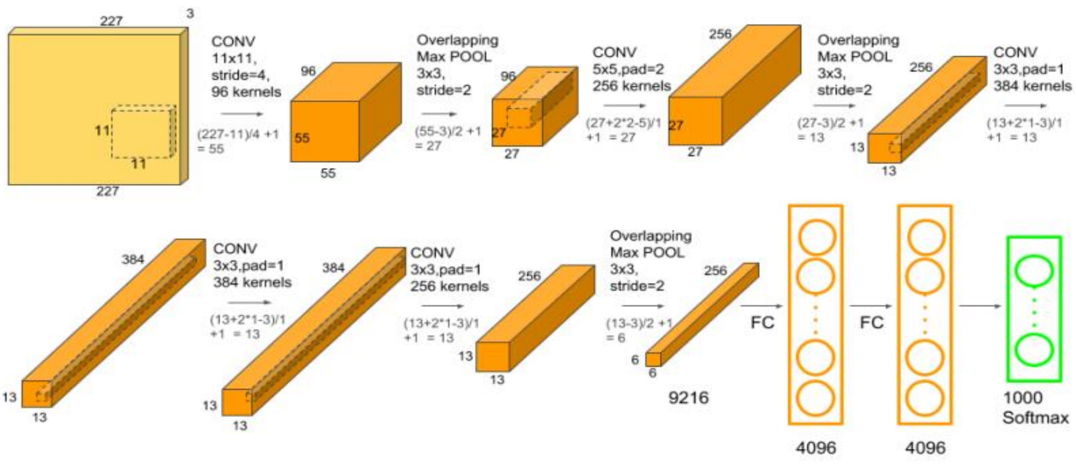

Fig. 2: Alex Net architecture layer information [16].

A. Dataset

Experimental studies have been carried out using a data set consisting of approximately 19,000 images, including female, male and child face images [21]. It is also worth mentioning that the dataset contained images with blurry faces of children and babies, which were not suitable for recognition, as their gender was ambiguous. These images were not included for training and a total of 5000 images were used, including 2500 female and 2500 male face images. Relatively, a smaller dataset has been used as provided in original Adience dataset, which contains 26 thousand face images of approximately 2,284 people obtained by smart phones and were used in previous gender classification studies [3]. Some sample images used in this study are shown in Fig. 3.

Fig. 3: Sample images included in the dataset.

B. CNN Models

In the study, the data set used was trained on two different CNN models which were implemented in Tensorflow and Keras. The first of these is a CNN model consisting of Conv2d, pooling, flatten and dense layers, as well as ReLu and Sigmoid activation functions. Detailed information about the layers of the CNN model created is shown in Fig. 1. The other model was Alex Net. Layer information of Alex Net architecture given in Fig. 2. For both models the dataset was divided into 80% training and 20% testing.

The proposed method was evaluated int terms of accuracy, precision, recall and F1-score. The metrics were obtained using true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Mathematically, these metrics are calculated as:

$$Accuracy = \frac{TN + TP}{TP + TN + FN + FP} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

Accuracy: The ratio of correctly predicted data in the model to the total dataset.
Precision: The metric that shows how much of the positively predicted data is actually positive.
Recall: It is the metric that shows how much of the data that should be predicted positively we predicted positively.
F1Score: Harmonic mean of Precision and Recall values. For this reason, it is a good indicator of success.

The overall results obtained for both CNN and AlexNet are summarized in Table I. In terms of F1-score, both models produced same results for classification of males and females. For each class, CNN produced better results (92%) compared to AlexNet (90%). Fig. 4 shows the test and train accuracy obtained for CNN and AlexNet models. Th overall behavior of both models look similar as also indicated by the quantitative results. Similarly, Fig. 5 shows the training and test loss for both models. The visuals indicate that overfitting did not occur in the training, due to the proximity of train and test accuracy values. The loss values express the sum of errors made for each sample in the training and test sets. It starts with high values at the beginning of the training and moves to low values as the model starts learning the data.

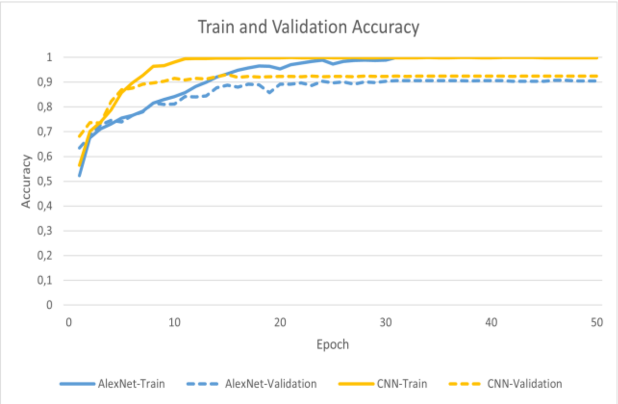

Fig. 4:Accuracy results of CNN and Alex Net models.

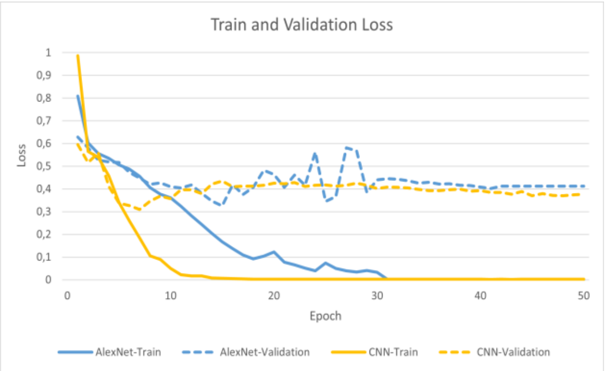

Fig. 5: Loss results of CNN and Alex Net model.

Additional experiments were performed to see the effect of the number of epochs on the validation accuracy of the classifiers. The results obtained for varying number of epochs are summarized in Table II. The results indicate that the test accuracy for both CNN and AlexNet models did not change after a certain number of (probably 20) epochs. However, increasing the number of epochs beyond 20, increase the processing time without any significant increase in the accuracy.

TABLE I. ACCURACY OBTAINED FOR ALEXNET AND CNN (%)

	Precision		Recall		F1-Score	
Classes	***AlexNet***	***CNN***	***Alex Net***	***CNN***	***AlexNet***	***CNN***
Male	%89	%91	%92	%93	%90	%92
Female	%92	%92	%89	%91	%90	%92

TABLE II. ACCURACY OBTAINED FOR DIFFERENT EPOCHS (%)

Epoch	*AlexNet*	*CNN*
10	81.10	90.40
20	90.20	92.20
30	89.80	92.30
40	90.50	92.40
50	**90.50**	**92.40**

Finally, we also conducted a comparative analysis of the CNN and AlexNet with some other popular networks for gender classification. The comparative results are presented in Table III. Compared to other methods, both CNN and AlexNet produced relatively higher classification accuracies for the gender classification problem.

TABLE III. COMPARATIVE ANALYSIS OF PROPOSED METHOD WITH OTHER MODELS IN TERMS OF ACCURACY (%)

Previous Works	*CNN*	*AlexNet*
Gündüz and Cedimoğlu [5]	72.20	65.63
Akbulut et al [4]	87.13	-
Levi and Hassner [6]	86.8	-
Yu at al [11]	91.50	-
Proposed	**92.40**	**90.50**

IV. CONCLUSION

In this study, Deep Learning models CNN and Alex Net are proposed for gender classification. Experiments were carried out on a dataset that is less in number than the images found in Adience, which was also used in previous gender recognition and classification studies. Gender classification was aimed by using the specified models together with the images in the dataset. The accuracy rates obtained as a result of experimental studies and the performance of the dataset in models were compared. With these procedures, the accuracy rates achieved were compared with the accuracy rates of similar studies in the literature, and better results were observed. In future studies, it is planned to make comparisons between different CNN models with the dataset containing more images.

REFERENCES

[1] A. Şeker, B. Diri and H. H. Balık, "Derin öğrenme yöntemleri ve uygulamaları hakkında bir inceleme," *Gazi Mühendislik Bilimleri Dergisi*, *3*(3), 47-64, 2017.

[2] T. V. Janahiraman and P. Subramaniam, "Gender Classification Based on Asian Faces using Deep Learning," In *2019 IEEE 9th International Conference on System Engineering and Technology (ICSET)* (pp. 84-89), October 2019, IEEE.

[3] E. Eidinger, R. Enbar and T. Hassner, "Age and gender estimation of unfiltered faces", *IEEE Transactions on Information Forensics and Security*, c. 9, sayı 12, ss. 2170– 2179, 2014.

[4] Y. Akbulut, A. Şengür and S. Ekici, "Gender recognition from face images with deep learning," In *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)* (pp. 1-4), September 2017, IEEE.

[5] G. Gündüz and İ. H. Cedimoğlu, "Derin Öğrenme Algoritmalarını Kullanarak Görüntüden Cinsiyet Tahmini," *Sakarya University Journal of Computer and Information Sciences*, *2*(1), 9-17.

[6] S. Arora and M. P. S. Bhatia, "A Robust Approach for Gender Recognition Using Deep Learning," In *2018 9th International Conference on Computing, Communication and Networking Technologies* (ICCCNT), pp. 1-6, 2018.

[7] G. Levi ve T. Hassner, "Age and gender classification using convolutional neural networks", *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops* (CVPRW), ss. 34–42, 2015.

[8] R. Ranjan, V. M. Patel and R. Chellappa, "Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(1), 121-135, 2017.

[9] M. Raza, M. Sharif, M. Yasmin, M. A. Khan, T. Saba and S. L. Fernandes, "Appearance based pedestrians' gender recognition by employing stacked auto encoders in deep learning," *Future Generation Computer Systems*, *88*, 28-39, 2018.

[10] N. A. Abdalrady, and S. Aly, "Fusion of Multiple Simple Convolutional Neural Networks for Gender Classification," In *2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE)* (pp. 251-256), 2020 February, IEEE.

[11] Z. Yu, C. Shen and L. Chen "Gender classification of full body images based on the convolutional neural network," In *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)* ,pp. 707-711, IEEE, 2017.

[12] I. Goodfellow, Y. Bengio, A. Courville and Y. Bengio, *Deep learning* (Vol. 1, No. 2). Cambridge: MIT press, 2016.

[13] R. C. Gonzalez, R. E. Woods and S. L. Eddins, " Digital image processing using MATLAB," Pearson Education India, 2004.

[14] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, 15(1), 1929-1958, 2014.

[15] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition*," Proceedings of the IEEE 86*(11), 2278–2324, 1998.

[16] A. Krizhevsky, I. Sutskever and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, 60(6), 84-90, 2017.

[17] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," In *European conference on computer vision* (pp. 818-833). Springer, Cham, 2014.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[19] C. Szegedy, W. Liu, Y. Q. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going Deeper with Convolutions," *2015 IEEE Conference on Computer Vision and Pattern Recognition* pp:1-9. IEEE.

[20] K. M. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp: 770-778, 2016.

[21] https://www.kaggle.com/ttungl/adience-benchmark-gender-and-age-classification.