



Gender classification from face images using central difference convolutional networks

Mohammadreza Sheikh Fathollahi¹ · Rezvan Heidari¹

Received: 18 July 2022 / Revised: 1 September 2022 / Accepted: 12 September 2022 / Published online: 22 September 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract

Nowadays gender classification which plays a vital role in face recognition systems is one of the main matters in computer vision. It is difficult to classify the gender from facial images when dealing with unconstrained images in a cross-dataset protocol. In this work, we propose two convolutional neural networks where one of the networks used the central difference convolution layer and another network used the vanilla convolution layer. The system was trained with the Casia WebFace dataset and tested on two cross-datasets, labeled faces in the wild (LFW) and FEI dataset. It is worth mentioning that the experimental results show the power and effectiveness of the proposed method. This method obtains a classification rate of 97.79% for the LFW dataset and 99.10% for the FEI dataset.

Keywords Gender classification · Central difference convolution · Face recognition · Convolutional neural networks

1 Introduction

Gender recognition from face images is a challenging problem with applications in various knowledge domains especially in computer vision. Classifying gender is easy for humans but very difficult for machines and it can be applied in various fields such as biometric authentication, security system, face anti-spoofing, criminology, and others. Many systems use user face images as input data because many personal characteristics such as age, gender, ethnicity, and identity can be extracted from them.

Automated face analysis of images has been extensively studied in recent decades due to its many applications such as human–computer interaction, surveillance systems, biometrics, augmented reality, and so on. Today, most purchases are made online and by automatically recognizing the gender of an image, the system can automatically select and recommend products that are of interest to the customers, like music recommender systems which are able to automatically detect the user’s musical preferences and create a playlist [1]. Existing overview-articles for algorithms related

to gender estimation include the works of Ng et al. [2], Khan et al. [3], and Bekios-Calfa et al. [4].

The difficulty of gender classification largely depends on the application context and the experimental protocol: a recognition model can be trained and tested on faces from the same dataset or different datasets (i.e. cross-dataset experiment), images of input faces can be taken under controlled or uncontrolled conditions and finally faces can be aligned before gender prediction or not [5]. Gender classification is not new in the field of computer vision. After a thorough review of the related works and according to [6], many useful algorithms for gender classification have already been developed. Generally, the extraction methods of image features are classified into two groups: global features-based methods (termed Holistic approaches [7]) and local features-based methods (termed Component-based methods [7] or Block Processing-based methods).

In general, Linear Discriminant Analysis (LDA) is implemented to classify patterns between two classes; however, it can be utilized for multiple classes. LDA gives class detachability by drawing a decision region between the various classes [8].

FaceHop [3] is recently proposed for gender classification on gray-scale face images. It uses PixelHop ++ for feature learning. For gray-scale face images of resolution 32×32 in the LFW and the CMU Multi-PIE datasets, FaceHop achieves correct gender classification rates of 94.63% and

✉ Mohammadreza Sheikh Fathollahi
mr.fathollahi73@gmail.com

¹ Department of Electrical and Computer Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

95.12% with model sizes of 16.9 K and 17.6 K parameters, respectively.

In [9] authors proposed a method that gender classification system, works for full-face images to half-face images, a Discrete Wavelet Transform (DWT) followed by MMDA is used for feature extraction. Their proposed approach uses DWT to gather the potential information from the face images. Support Vector Machine (SVM) and k-NN classifiers were used to find the features that can discriminate between males and females. Kaur K et al. [9] proposed method was evaluated on FERET and FEI databases, and the experimental result shows that the proposed technique achieves the gender classification target with more than 94% accuracy for both half-face and full-face images.

The authors in [10] used a local binary pattern (LBP) as a texture method. But LBP suffers from a big problem; it cannot show spatial relationships between local textures. Therefore, to increase the accuracy of gender classification, two LBP descriptors, based on (1) spatial relationships between neighbors with the distance parameter, and (2) spatial relationships between the reference pixel and its neighbor in the same direction, have been used. The authors used Gray Relationship Analysis (GRA), performed to identify gender through the extracted features. Using GRA with, and traditional LBP characteristics, the accuracy obtained is 97.14%, 93.33%, and 92.50%, respectively, on the FEI dataset.

Lian and Lu [11] utilized the facial texture information for gender classification. For this purpose, the authors divided the facial area into several small areas and used an LBP operator in each area to extract texture features. The extracted texture features are called LBP histograms and fed that feature vectors to SVM for classification. According to their results on the Chinese Academy of Sciences-Pose, Expression, Accessories, and Lighting (CAS-PEAL) database, the SVM classifier achieved an accuracy of 96.75%.

Sun et al. [12] proposed the LBP operator to extract features in facial images. The authors used the AdaBoost classifier, instead of using the SVM classifier, and achieved an accuracy of 95.75% in the FERET database. Lian et al. [11] and Sun et al. [12] used controlled images in their experiments. However, considering real-world applications, it is reasonable to design gender classification methods to perform well for images captured in uncontrolled conditions. Shan [13] used the Labeled Faces in the Wild (LFW) database, which contains real-world face images. The author trained an SVM classifier using an LBP histogram extracted from real face images. According to his results on the LFW database, the SVM classifier achieved an accuracy of approximately 94.8%.

Convolutional neural networks have shown significant performance in various image recognition problems. CNN-based methods are used to extract features and classification algorithms for automatic gender classification [14]. In the

early 1990s, researchers used neural networks to solve the problem of gender classification. Golomb A et al. [15] trained a neural network with two fully connected layers and achieved relatively good accuracy on a small training set.

Recently, instead of using hand-crafted features from grayscale images, some research groups have focused on a machine learning methodology to automatically learn features extracted from RGB facial images. In this approach, convolutional neural networks (CNNs) have been mainly used. For example, in [16], a CNN model was proposed to directly learn features from facial images. Islam et al. [17] applied three existing CNN models, namely GoogleNet [18], SqueezeNet [19], and ResNet-50 [20], to gender classification tasks. In addition, the authors demonstrated the feasibility of these models for classifying genders. Some studies have focused on image classification techniques for gender recognition. Zhang, et al. [21] propose a method for gender detection using images. First, they use CNN for feature extraction. Next, they apply a self-joint attention model for feature fusion. Finally, they use two fully connected neural network layers with ReLu and softmax activation functions, and one average pooling layer to predict the gender. AI applications are notable in using in the recruitment process because it will be the combination of human and computer collaboration that will work together to reach desired goals, in the concern of the facial identification is now to the built-up framework to recognize the individual face picture and video [22]. Hsu Y. C, et al. discuss a way for enhancing data. The proposed approach improves the quality of the input data by modeling probable barriers that may arise in real-world circumstances. The suggested method begins by randomly picking a fixed size region from an input image and then using one of the occlusion approaches. When utilizing blackout, random brightness, or blur occlusion techniques, faces are obscured, lightning is powerful, and resolution is restricted. A convolutional neural network and a VGG16 deep neural network [23]. Recently, various Bilinear CNN models have been proposed in the literature for fine-grained classification. Ben et al. [24] proposed a Bilinear CNN model combining a pre-trained VGG16 with a shallow CNN, as feature extractors, for facial emotion recognition. The proposed model was evaluated on CK + and FEI facial expression datasets and achieved accuracy rates of 86.98% and 85.35%, respectively.

Most gender detection methods rely on stacked convolutions and expert-designed network, which is weak in describing detailed information and easily being ineffective when the environment varies (e.g., different illumination), according to this we have utilized Central Difference Convolution (CDC), which is able to capture intrinsic detailed patterns via aggregating both intensity and gradient information. Moreover, CDC is able to extract the invariant detailed features (e.g., illumination & input camera).

The purpose in this paper is to improve the accuracy of gender classification using the CDC method, which has been used in face anti-spoofing [25] task.

In this paper, two parallel CNNs which concatenated at the first dense layer are used. One of the CNNs works with the vanilla convolution (VC) layer and another one is working with the central difference convolution (CDC) layer [25], the network built with VC is called VCNN, and the network built with CDC is called CDCN. Three datasets are used, one for training and two for testing the system. The difference between our system and the competitive algorithms is the CDC layer that will be described in a subsequent section.

For better comparison, we have tested and evaluated models individually and reported them in their relevant tables.

Paper is organized as follows. In Sect. 2, we described briefly about CDC and the gender classification system and its details. In Sect. 3, experimental setups and results are discussed. At last in Sect. 4, the paper is concluded.

2 Proposed method

In this paper, the unique and novel approach is to utilize two parallel CNNs which contains vanilla and central difference convolution layers. The main goal is to design an automatic gender classification system that is invariant detailed facial images. In this section, first is the introduction of CDC and then the introduction of VCCN and CDCN model.

2.1 Central difference convolution

The inputs after passing through a convolutional layer become abstracted to a feature map. Convolutional layers convolve the input and pass its result to the next layer; the convolution operation remains the same across the channel dimension. The convolutions are described in 2D while extension to 3D is straightforward.

There are two main steps in the 2D convolution: 1- sampling local receptive field region R over the input feature map x ; 2- aggregation of sampled values via weighted summation.

Hence, the output feature map Z can be formulated as:

$$Z(t_0) = \sum_{t_n \in R} x(t_0 + t_n) \cdot w(t_n) \tag{1}$$

where t_0 denotes current location on both input and output feature maps while t_n enumerates the locations in R .

Central difference into vanilla convolution is mentioned to enhance its representation and generalization capacity according to LBP [26] which describes local relations in a binary central difference way. Central difference convolution also consists of two steps, sampling, and aggregation. The sampling step is similar to that in vanilla convolution while the aggregation step is different as shown in Fig. 1.

$$Z(t_0) = \sum_{t \in R} (x(t_0 + t_n) - x(t_0)) \cdot w(t_n) \tag{2}$$

When $t_n = (0,0)$, the gradient value always equals to zero with respect to the central location t_0 itself. Therefore, generalization of central difference convolution as:

In order to efficiently implement CDC in modern deep learning framework, merging Eq. (3) into the vanilla convolution with additional central difference term:

$$Z(t_0) = \theta \cdot \sum_{t_n \in R} ((t_0 + t_n) - x(t_0)) \cdot w(t_n) + (1 - \theta) \sum_{t_n \in R} ((t_0 + t_n) - x(t_0)) \cdot w(t_n) \tag{3}$$

where hyper-parameter $\theta \in [0,1]$ tradeoffs the contribution between intensity-level and gradient-level information. The

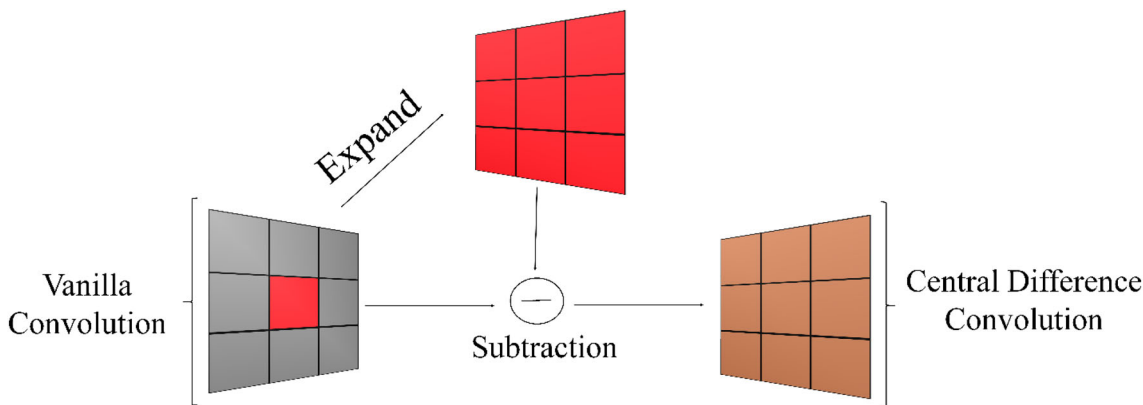


Fig. 1 Central difference convolution

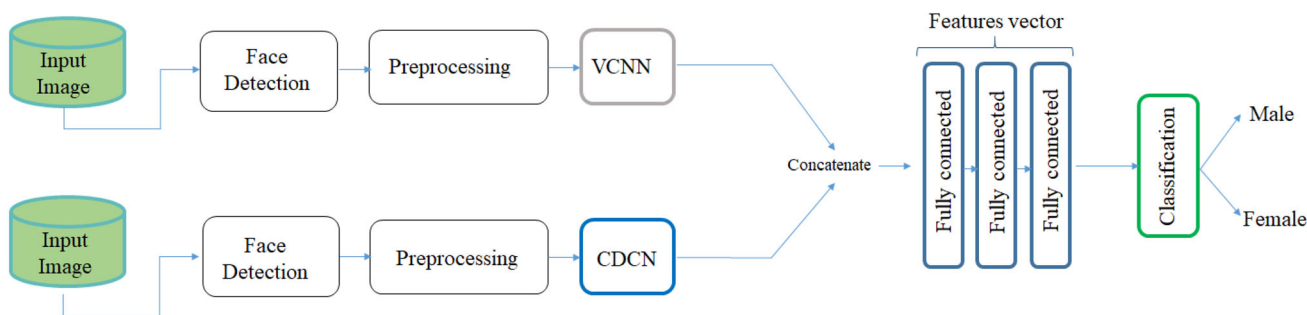


Fig. 2 The system architecture



Fig. 3 Face detection from raw images from Casia Web Face dataset

higher value of θ means the more importance of central difference gradient information.

$$\begin{aligned}
 Z(t_0) = & \underbrace{\sum_{t_n \in R} x(t_0 + t_n)}_{\text{Vanilla Convolution}} \cdot w(t_n) \\
 & + \theta \cdot \underbrace{-x(t_0) \cdot \sum_{t_n \in R} w(t_n)}_{\text{CDC Term}}
 \end{aligned}
 \tag{4}$$

According to Eq. (4), in PyTorch [27] and TensorFlow [28], CDC can be implemented by a few lines of code.

2.2 Models

As mentioned in the previous section, two parallel neural networks were used, as shown in Fig. 2, one neural network is with vanilla convolution layer and the other is with a central difference convolution layer.

According to Fig. 2, before feeding the inputs to neural networks, pre-processing is performed on all input images. At first, only the faces were detected and extracted from the raw images, according to Fig. 3. We did this task by using insight-face toolbox and zoo models [29] in preprocessing unit, we resized all input facial images into 128 * 128 for reducing process volume. After feeding the facial images into neural networks, we concatenated the two neural networks in the first dense layer and then classifying the gender of input images.

In this experiment, dropout regularization after every pooling layer to avoid over-fitting and feature co-adoption [30] is applied. For activation function, rectified linear units (ReLUs) [31] in all VC and CDC convolutional and dense layers are used.

$$\text{ReLU}(x) = \text{Max}\{0, x\}
 \tag{5}$$

The exception of using ReLUs is in the last layer, where the Softmax function [30] is applied instead.

$$S : R^k \rightarrow R^k$$

$$S(x)_j = \frac{e^{x_j}}{\sum_{j=1}^k e^{x_j}} \tag{6}$$

for $i = 1, \dots, k$
 $S : R^k \rightarrow R^k$

According to Fig. 4, the model that used is consists of four blocks and 17 layers including the input layer, the first block consists of three VC layers with (3,3) kernel size and Stride (1,1) and also a 2D Max-pooling layer with pool size (2,2). The second and third blocks are similar to the first block but the difference is in the max-pooling layer where the pool size is (4,4). The fourth block is the fully connected block that contains four dense layers. The first dense layer size is 150 and the last dense layer size is according to the output classes is equal to 2.

Another model that we design, which uses the CDC convolutional layer as shown in Fig. 5, has the same number of layers that uses in the VC model, as shown in Fig. 4. The model consists of four blocks and 17 layers including the input layer. The first block.

As mentioned in the previous section, we concatenated the two models at the first dense layer, and by using the softmax activation function, the gender images are classified, consisting of three CDC layers with (3,3) kernel size and Stride (1,1) and a 2D Max-pooling layer with pool size (2,2). The second, third, and fourth blocks are similar to the VCCN model.

Given input RGB facial image with size $128 \times 128 \times 3$, we use $\theta = 0.7$ as the default setting and ablation and study about θ is in [25].

3 Experiments and results

In this section, we reported datasets, experiments, setups, results, and the evaluation of the methodology which was described in the previous section.

3.1 Datasets

In this section, we presented the datasets which used in the experiments. Three publicly available face datasets: CASIA WebFace, Labeled Faces in the Wild (LFW) and FEI database, are used. The first one is used for training, and validation, whereas the second and third are used only for testing.

3.1.1 CASIA WebFace dataset

CASIA WebFace dataset was collected for face recognition purposes by Yi et al. [32]. The dataset contains photos

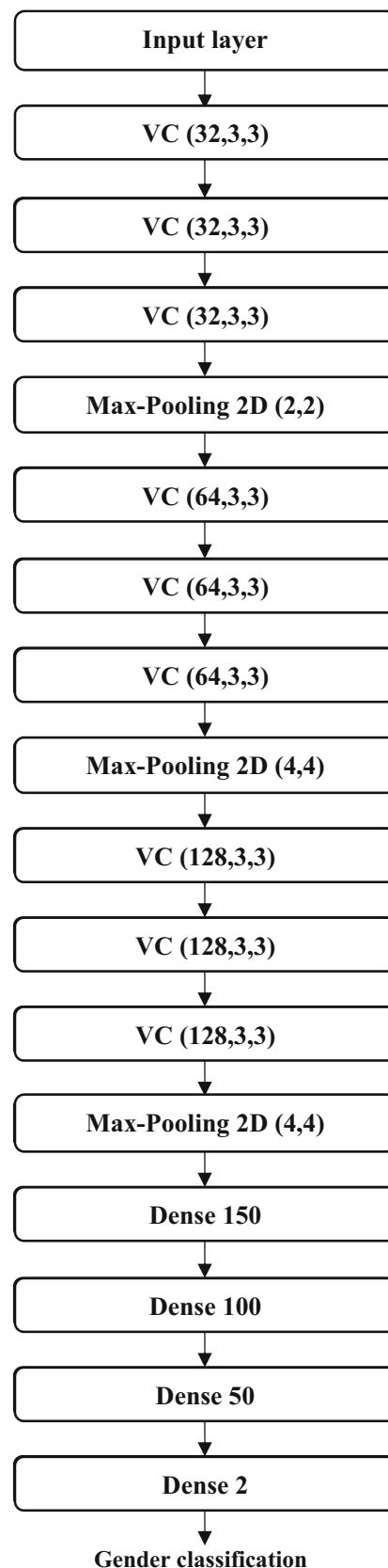


Fig. 4 The structure of the model with VC layers

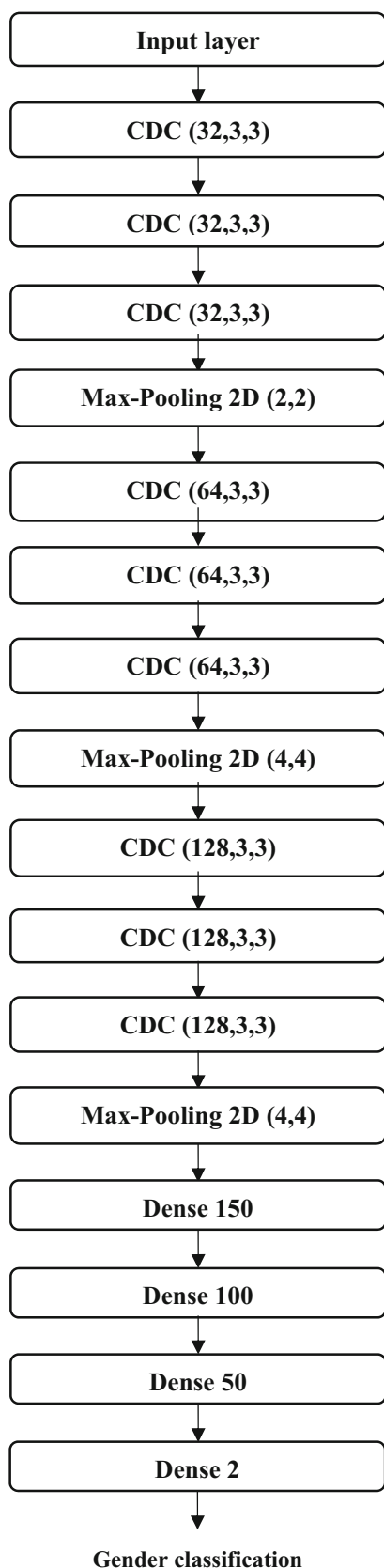


Fig. 5 The structure of the model with CDC layers

of actors and actresses born between 1940 and 2014 from the IMDb website. Images of the CASIA WebFace dataset include random variations of poses, illuminations, facial expressions, and image resolutions. In total, there are 494,414 face images of 10,575 subjects. In this work, CASIA WebFace have been used to train the networks. Authors of CASIA WebFace provide names of 10,575 subjects but not their genders.

3.1.2 Labeled faces in the wild (LFW) dataset

Being collected by Huang et al. [33], the LFW dataset has become a benchmark for face gender recognition in an unconstrained environment. It consists of 13,233 face images of 5749 celebrities. Contrary to CASIA WebFace, LFW contains photos of actors, actresses, politicians, sportsmen, and sportswomen.

3.1.3 FEI dataset

The FEI database [34] is a Brazilian face database that contains a set of face images taken between June 2005 and March 2006 at the Artificial Intelligence Laboratory of FEI in Sao Bernardo do Campo, Sao Paulo, Brazil. There are 14 images for each of 200 individuals, a total of 2800 images. All images are colorful and taken against a white homogenous background in an upright frontal position with profile rotation. The scale might vary about 10% and the original size of each image is 640×480 pixels. All faces are mainly represented by students and staff at FEI, between 19 and 40 years old with a distinct appearance, hairstyle, and adorns. The number of male and female subjects are exactly the same and equal to 100.

3.2 Experimental setup

CASIA WebFace dataset have been used to train the network, as mentioned in the previous section, the CASIA WebFace dataset has 494,414 face images of 10,575 subjects.

First, we detected and extracted only faces from raw images, then these samples are fed to the network for training and validating process. We used a total of 113,000 female and male images for training and validating the system according to Table 1.

It should be noted that validation samples and training samples have no similarities with each other, and in terms of validation, the samples were randomly selected and shuffled. To train the network, Adam optimizer [35], with the default learning rate of 0.1, was applied. The binary cross-entropy was used as a loss function. The batch size was 32 and the model was trained for 100 epochs.

Table 1 Splitting data into training and validation samples

Process	Dataset	Female images	Male images
Training	CASIA WebFace	56,500	56,500
Validation	CASIA WebFace	11,300	11,300

Table 2 The results of LFW dataset on both models separately

Models	Female images	Male images	Accuracy (%)
VCNN	3012	8471	95.23
CDCN	3012	8471	96.61

Table 3 The result accuracy in LFW dataset by fusing two models

	Female	Male	Total
Female	2899	140	–
Male	113	8331	–
Accuracy	96.24%	98.34%	97.79%

3.3 Results

As mentioned in the previous section, three databases to train and test the systems were used. Casia WebFace dataset for training and validating, LFW and FEI dataset for testing the system. First, we obtained the results without fusing the two networks and reported them; then results were reported by fusing two networks. We also compared the method with other approaches on LWF and FEI datasets as shown in Tables 4 and 7.

3.3.1 Results on LFW dataset

We tested the system with both models separately; first, we tested the system with the model that consists of VC layer according to Fig. 4; then we tested the system with the CDCN model according to Fig. 5; the results are reported in Table 2.

In the LFW dataset after detecting face images using the insight-face toolbox and zoo models, we got 11,483 images. Contain 3012 female images and 8471 male images.

In testing both models, the number of female images and male images are the same and tabulated in Table 2. The CDCN model, which consists of the CDC convolution layer, is able to extract the invariant detailed of images, shows slightly better results.

In Table 3, the results of the fusion model are reported;

Table 4 Gender classification results on LFW dataset

Studies	Algorithm	ACC (%)
Almabdy S et al. [36]	AlexNet	94.00
Almabdy S et al. [36]	ResNet-50	94.00
Rouhsedaghat M et al. [37]	FaceHop	94.63
Shan C [13]	LBP	94.80
Proposed method	VCNN	95.23
Han H et al. [38]	BIF	95.40
Almabdy S et al. [36]	AlexNet + ResNet-50	96.21
Proposed method	CDCN	96.61
Casas P et al. [39]	LBP + PCA	97.20
Antipov G et al. [40]	CNN	97.30
Proposed method	VCCN + CDCN	97.79

Table 5 The results of FEI dataset on both models separately

Models	Female images	Male images	Accuracy (%)
VCNN	1323	1369	96.23
CDCN	1323	1369	97.49

the number of test images in the fusion model is equal to individual models. As we can see, the accuracy increases in the fusion model. In the LFW dataset, the number of male images is much higher than the females, but the accuracy on male face images is higher, this due to the way the hair is covered or not covered or the makeups, children images, and so on. The accuracy that we achieved in total is 97.79%.

In Table 4, different results on the LFW dataset are compared.

In [32] the authors tested their system individually and reported the results, by using Alex-Net the 94% accuracy, and using ResNet-50 they achieved 94% accuracy either.

It can be seen in Table 4, the employed approach could produce significant performance compared to other studies in the literature. Achieved high accuracies are due to employing and fusing VCNN and CDCN which is a successful gender classification approach. The proposed method showed about 20% and 40% error reduction on LFW dataset in comparison with [36, 40], respectively.

3.3.2 Results on FEI dataset

In the FEI dataset, we tested the system with both models separately like the LFW dataset, first with VCNN then with CDCN. The result is tabulated in Table 5.

According to Table 5, the number of female images and male images are similar in both VCNN and CDCN models.

Table 6 The accuracy result in FEI dataset

	Female	Male	Total
Female	1316	17	–
Male	7	1352	–
Accuracy	99.47%	98.75%	99.1%

Table 7 Gender classification results on FEI dataset

Studies	Algorithm	ACC (%)
Jabra M et al. [24]	B-CNN	85.35
Liao M et al. [41]	LDF	88.20
Kaya Y et al. [10]	LBP	92.50
Wadhera A et al. [42]	PTGSP-CWP	93.30
Kaya Y et al. [10]	nLBP _d	93.33
Curtidor A et al. [7]	RLD	93.57
Kaur K et al. [9]	MMDA	94.00
Abdalali E et al. [8]	DCT	95.70
Proposed Method	VCNN	96.23
Kaya Y et al. [10]	dLBP _α	97.14
Proposed Method	CDCN	97.49
Almabdy S et al. [36]	AlexNet	97.50
Almabdy S et al. [36]	ResNet-50	98.50
Almabdy S et al. [36]	AlexNet + ResNet-50	99.00
Proposed method	VCCN + CDCN	99.10

Because of the ability of extract, the invariant detailed of images CDCN show slightly better accuracy on FEI dataset like LFW dataset.

In FEI dataset, all 2692 face images were detected via the using insight-face toolbox, the number of detected faces in female and male samples are tabulated in Table 5. According to Table 6, we tested the model on another cross-dataset. The accuracy of female images is slightly better than male images. The accuracy that we reach in the FEI dataset at total is 99.1%.

In Table 7, we compared different results on the FEI dataset. In [36] the authors tested their system individually on the FEI dataset like the LFW dataset, by using Alex-Net the accuracy reached 97.50%, and with using ResNet-50 they achieved 98.50% accuracy either.

In [36] authors fused two pre-trained convolutional neural networks Alex-Net and ResNet-50, respectively, to extract high-level features which are suitable for tasks like gender classification, and they could achieve a high accuracy on FEI dataset. It can be seen in Table 7 that the proposed method performed well on the FEI database compared to other studies. The proposed method showed 10% error reduction on

FEI dataset in comparison with the fusion of AlexNet and ResNet-50.

4 Conclusions

In this paper, we proposed a system that detects and classifies gender using two neural networks which fused in the first dense layer. One of the networks was used central difference convolution and another network was used standard convolution called vanilla convolution. The system trained with the Casia WebFace dataset then tested on two cross-datasets LFW and FEI datasets. According to the results the system was done well on the LFW dataset and FEI dataset, we achieved magnificent accuracies of about 97.79% and 99.1% On the LFW dataset and FEI dataset, respectively. According to the results, studying pre-trained convolutional neural networks and fusing them with CDCN can be our future work.

Data availability All data generated or analyzed during this study are included in this published article.

Declarations

Conflict of interest All authors declare that they have no conflict of interest.

References

1. SheikhFathollahi M, Razzazi F (2021) Music similarity measurement and recommendation system using convolutional neural networks. *Int J Multimed Inf Retr* 10(1):43–53
2. Ng CB, Yong HT, Bok MG (2012) “Vision-based human gender recognition: a survey,” *PRICAI 2012: trends in artificial intelligence. Lect Notes Comput Sci* 7458:335–346
3. Sajid A. K, Muhammad N, Sheeraz A, Naveed R, (2011) Gender classification using image processing techniques: a survey. In: *International Multi Topic Conference (INMIC)*.
4. Juan B, Jose MB, Luis B (2011) Revisiting linear discriminant techniques in gender recognition. *IEEE Trans Pattern Anal Mach Intell (TPAMI)* 33(4):858–864
5. Antipov G, Berrani S, Dugely J (2016) Minimalistic CNN-based ensemble model for gender prediction from face images. *Pattern Recognit Lett Elsevier* 70:59–65
6. Saha D, De D, Ghosh P, Sengupta S, Majumdar T (2020) An extensive analysis of human gender prophecy using computer vision. *IRJET* 7(5):1223–1229
7. Curtidor A, Baydyk T, Kussul E (2021) Analysis of random local descriptors in face recognition. *Electronics* 10(11):1–19
8. Abdalali O. E, Eltarhouni I. W, Abdo A. A, Almajbri I. S, (2021) Gender classification using discrete cosine transform and multi classifiers. In: “The 7th International Conference on Engineering & MIS.
9. Kaur K, Rai P and Khanna P, (2017) Gender classification system for half face images using multi manifold discriminant analysis. In: *7th International Conference on Cloud Computing, Data Science & Engineering - Confluence, 2017, pp 595–598*

10. Kaya Y, Ertuğrul ÖF (2017) Gender classification from facial images using gray relational analysis with novel local binary pattern descriptors. *Signal Image Video Process Springer* 11:769–776
11. Lian H, Lu B, (2006) Multi-view gender classification using local binary patterns and support vector machines. In: *Proc Int Symp Neural Netw (ISNN)*, pp 202–209.
12. Sun N, Zheng W, Sun C, Zou C and Zhao L, (2006) Gender classification based on boosting local binary pattern. In: *Proc Int Symp Neural Netw (ISNN)*, pp 194–201.
13. Shan C (2012) Learning local binary patterns for gender classification on real-world face images. *Pattern Recognit Lett* 33(4):431–437
14. Levi G, Hassner T, (2015) Age and gender classification using convolutional neural networks. In: *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Boston pp 34–42.
15. Golomb A, Lawrence T, Sejnowski J, (1990) Sexnet: A neural network identifies sex from human faces. In: *Proceedings of the conference on Advances in Neural Information Processing Systems*, Denver, USA.
16. Levi G, Hassner T, (2015) Age and gender classification using convolutional neural networks. In: *Proc IEEE Conf Comput Vis Pattern Recognit Workshops (CVPRW)*, pp 34–42.
17. Islam M, Tasnim N, Baek J (2020) Human gender classification using transfer learning via Pareto frontier CNN networks. *Inventions* 5(16):1–12
18. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. (2015) Going deeper with convolutions. In: *Proc IEEE Conf Comput Vis Pattern Recognit (CVPR)*, pp 1–9.
19. Iandola F, Han S, Moskewicz M, Ashraf K, Dally W, Keutzer K, (2017) SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5MB model size. In: *Proc Int Conf Learn Represent (ICLR)*, pp 207–212.
20. He K, Zhang X, Ren S and Sun J, (2016) Deep residual learning for image recognition. In: *Proc IEEE Conf Comput Vis Pattern Recognit (CVPR)*, pp 770–778.
21. Zhang X, Javid S, Dias J, Werghi N, (2021) Person gender classification on RGB-D data with self-joint attention. *IEEE*, Access, V. 9, <https://doi.org/10.1109/ACCESS.2021.3135428>.
22. Masud M, Muhammad G, Alhumyani H, Alshamrani SS, Cheikhrouhou O, Ibrahim S, Hossain MS (2020) Deep learning-based intelligent face recognition in IoT-cloud environment. *Comput Commun* 152:215–222
23. Hsu YC, Lin EL, Lin HC (2021) Age and gender recognition with random occluded data augmentation on facial images. *Multimed Tools Appl* 80(8):11631–11653
24. Jabra B. M, Guetari R, Chetouani A, Tabia H, Khelifa N, (2020) Facial expression recognition using the bilinear pooling. In: *15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory*.
25. Yu Z, Zhao C, Wang Z, Qin Y, Su Z, Li X, Zhou F, and Zhao G, (2020) Searching central difference convolutional networks for face anti-spoofing. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 5295–5305.
26. Boulkenafet Z, Komulainen J, Hadid A, (2015) Face anti-spoofing based on color texture analysis. In: *In IEEE international conference on image processing (ICIP)*, pp 2636–2640.
27. Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L, and Lerer A, (2017) Automatic differentiation in pytorch".
28. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Mawat S, Irving G, Isard M et al (2016) Tensorflow: a system for large-scale machine learning. *OSDI* 16:265–283
29. Kasinski A, Schmidt A (2010) The architecture and performance of the face and eyes detection system based on the Haar cascade classifiers. *Pattern Anal Appl* 13:197–211
30. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, and Salakhutdinov R, (2012) Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*
31. Zeiler M D, Ranzato M, Monga R, Mao M, Yang K, Le Q V, Hinton G E, (2013) On rectified linear units for speech processing. In: *International Conference on Acoustics, Speech and Signal Processing*. IEEE, pp 3517–3521
32. Yi D, Lei Z, Liao S, Li Z, (2014) Learning face representation from scratch. Technical report, [arXiv:1411.7923](https://arxiv.org/abs/1411.7923).
33. Huang G, Ramesh M, Berg T and Learned-Miller E, (2007) Labelled faces in the wild: a database for studying face recognition in unconstrained environments. *Univ Massachusetts Amherst, Massachusetts*, Tech Rep 07–49.
34. Thomaz C, (2012) FEI face database. 2012. [Online]. Available: <https://fei.edu.br/~cet/facedatabase.html>. [Accessed: 06 Apr 2019].
35. Kingma D, Ba J (2015) Adam: a method for stochastic optimization. In: *International conference on learning representations (ICLR)*, pp 232–235.
36. Almabdy S, Elrefaei L (2021) Feature extraction and fusion for face recognition systems using pre-trained convolutional neural networks. *Int J Comput Digit Syst* 10:465–461
37. Rouhsedaghat M, Wang Y, Ge X, Hu S, You S, and Kuo J, (2020) Facehop: a light-weight low-resolution face gender classification method. *arXiv preprint arXiv:2007.09510*.
38. Han H, Jain K, (2014) Age, gender and race estimation from unconstrained face images. In: *Tech. Rep. No. MSU-CSE-14-5*. East Lansing, Michigan: Department of Computer Science, Michigan State University.
39. Casas P, Jimnez, D, Yu L, Castro L, (2011) Single- and cross-database benchmarks for gender classification under unconstrained settings. In: *Computer vision workshops (iccv workshops)*, IEEE International Conference pp 2152–2159.
40. Antipov G, Baccouche M, Berrani A, Dugelay L, (2016) Apparent age estimation from face images combining general and children-specialized deep learning models. In: *IEEE CVPR Workshops*, pp 801–809.
41. Liao M, Gu X (2020) Face recognition approach by subspace extended sparse representation and discriminative feature learning. *Neurocomput Elsevier* 373:35–49
42. Wadhwa A, Agarwal M (2022) "Robust pattern for face recognition using combined weber and pentagonal-triangle graph structure pattern" *Optik*. Elsevier 259:168925

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.