Int. J. Nonlinear Anal. Appl. 14 (2023) 1, 2373–2381
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.26099.3661

Providing a hybrid method for face detection and gender recognition by a transfer learning and fine-tuning approach in deep convolutional neural networks and the Yolo algorithm

Peyman Jabraelzade, Asghar Charmin*, Mohsen Ebadpour

Department of Electrical Engineering, Ahar Branch, Islamic Azad University, Ahar, Iran

(Communicated by Ehsan Kozegar)

Abstract

The present study aims to assess a combined method for face detection and gender recognition using deep learning by evaluating the shortcomings of face detection methods accurately. Deep learning algorithms can learn high-level features and have attracted a lot of attention for use in the field of machine vision. The present study names a hybrid method called Hyper-Yolo-face and utilizes a clear image using deep Convolution Neural Networks (CNNs), Yolo algorithm, and local binary patterns (LBPs) to identify the face and recognize the gender. Reducing the number of parameters is regarded as an extremely important challenge in deep networks in terms of memory consumption and the amount of computing in the network. The proposed method is based on the AlexNet model and generalization in the loss function of version 3 of the Yolo algorithm, which leads to improved precision. The present study focuses on applying small filters in transfer learning and fine-tuning network layers and using a new regression loss function in the Yolo algorithm to make it more appropriate for multiscale face detection. The face images are detected and cut by the presented Yolo in the proposed method. Then, an LBP operator is applied so that richer information and images enter the AlexNet network to estimate other parameters including gender recognition. Based on the experiments on the AFLW, FDDB, and PASCAL datasets, the proposed method improves recognition precision significantly.

Keywords: Deep learning, face detection, Yolo, transfer learning, gender recognition, local binary pattern
2020 MSC: 68T07

1 Introduction

The ability of convolutional neural networks (CNN) to automatically extract image features using the concept of deep learning is considered as their most important advantage, which utilizes an extremely large database for network training [27]. A large number of well-known models have emerged for convolutional neural networks (CNNs) with recent advances in applying such networks in the field of computer vision, among which AlexNet and Yolo were assessed and selected as the most common ones. Using such models proposed a concept called transfer learning and fine-tuning, which involves utilizing pre-trained models in a new application separate from the trained one. A trained network with extremely appropriate results in the proposed image banks is selected during fine-tuning of a deep

*Corresponding author
Email addresses: p_jabraelzade@iau-ahar.ac.ir (Peyman Jabraelzade), a_charmin@sut.ac.ir (Asghar Charmin), m.ebadpour@ac.ir (Mohsen Ebadpour)

Received: April 2022 *Accepted:* June 2022

convolutional network. Then, the final fully connected layer is replaced with desired one which can be classified by the number of classes. The weights of the first layers in the main network enter the new one without change during the above-mentioned process and the final layers are re-trained so that the desired network can classify new images [29]. The present study aims to evaluate trained model of AlexNet and fine-tuning its hyperparameters in the proposed method. AlexNet is regarded as a deep CNN with five convolutional, three fully connected, and a number of activation and pooling layers, which won the annual photo classification competition by a large database during 2014 [11]. Local binary pattern (LBP) is utilized to feed the features to the network, upgrade the input dimensions of the images to six channels, and enter richer information into the network with the image in order to improve the detection precision because the aforementioned pattern applies both statistical and structural features of the texture and is considered as a powerful instrument for texture analysis. LBP which was first proposed by Ocalan et al. (1996) is among the most common descriptors utilized by the researchers in the most image processing studies to increase and improve precision due to its resistance to changes in brightness, low computational complexity, and ability to encode details [15]. The proposed method is based on the AlexNet model and applies the presented Yolo algorithm called Hyper-Yolo-face instead of the selective search one to generate propositions for face area and crop. The YOLOv3 architecture was selected as the face detection network structure with the proposed loss function for regression, which is regarded as a combination of the loss for Mean Square Error (MSE) and that for generalization Intersection over Union (GIoU). Then, faces were quickly detected and cut by the proposed Yolo algorithm and fed to the proposed network, along with additional information integrated through LBP. The rest of present study is organized as follows. The literature review is provided in Section 2. Details of the proposed method and implementation of Hyper-Yolo-face are given in Section 3. The results of the proposed approach on the AFLW, FDDB, and PASCAL datasets are described in Section 4, and the general conclusion is given with a brief discussion in Section 5.

2 Literature review

Machine learning methods have a wide range of applications [21, 22, 23]. Using pyramidal networks in face detection and adaptation is presented in [1]. In fact, CNN is utilized as a pyramid in training, which reduces performance to some extent and increases execution time with a high precision in face detection and adaptation. Face detection and adaptation is conducted by the deep learning approach and Recurrent Neural Network (RNN) algorithm in [33] which are among the latest studies. Applying the FDDB dataset is analyzed, the precision rate and face detection have a high index, and Receiver Operating Characteristic (ROC) indicates the improvement of the proposed approach. Weight restraint indices are tested to train facial images after the feature extraction and segmentation phase with the deep learning network in [32]. Using different data sets, especially FDDB, is among the most important results of the above-mentioned study, which has indicators of functional superiority in terms of evaluation criteria compared to previous methods. Utilizing the end-to-end training from a similar series of images in a dataset to detect faces from low-quality and resolution images is examined in [3]. The proposed approach is considered as a data-driven and test-based one with image processing techniques since it is called Face Super-Resolution Generative Adversarial Network (FSRGAN) for short. [6] applied a hybrid deep learning network to identify faces. In fact, a person's face is identified from the image and his/her emotional characteristics are detected to recognize emotional states from face images. Utilizing convolutional or torsional layers and applying RNN are studied. [4] used CNN technique from deep learning to identify faces from images in a field. Utilizing a dataset with 1553 images including those related to animals and human face is analyzed. Based on the results, the precision of 96.7% is observed in detecting faces from images and classifying human faces from other images. [8] used developed deep learning based on CNN technique to identify faces from images and create a system for recognizing facial expressions from images. Utilizing two datasets called Extended Cohn-Kanade (CK+) and (JAFFE) are evaluated, the results of which indicate an improvement in the proposed approach. A multifunctional approach with a deep learning approach is presented in [17] which estimates the tasks related to face detection, gender recognition, gesture estimation, and locating prominent points in the face with a relatively appropriate precision by combining the AlexNet architecture and RNNs. However, the aforementioned approach fails to act fast in detecting and identifying the face and its candidate areas due to applying selective search algorithm. The multi-layered neural network solution is used to confirm gender through face image in [26]. Combining Radial Basis Function (RBF) networks and Inductive Decision Trees (IDT) is among the methods utilized in gender recognition, as well.

3 The proposed method called Hyper-Yolo-Face

The present study seeks to examine a hybrid model of CNN for face detection and gender recognition, which includes two main parts. The first part applies the proposed Yolo algorithm, which quickly crops people's faces from images and scales them to the desired size of the network. The second part is regarded as a CNN, which receives and categorizes the cropped areas as face or non-face, and increases the dimensions of the images entering the network to six by the feature of LBPs. Yolo transformed the architecture of object and face detection systems, perceiving the recognition as a regression issue which goes directly from image pixels to box coordinates and class probabilities [18]. Darknet-53 is used as the backbone for Yolo network, which includes one feature extraction network and three detection networks. The feature extraction network is based on Darknet-53, which is considered as a combination of Darknet-19 and Residual Network (ResNet). To achieve multidimensional integration, low-level features are combined with high-level ones such as feature pyramid networks (FPN) [2, 9], which can make better use of the scales related to visual information and give more ideal performance to the multiscale detector. Yolo is regarded as highly generalizable and fails less in the face of new domains or unexpected input data compared to other systems. Yolo can be utilized to eliminate daily obstacles due to its promptness, despite having slightly weaker performance than RNNs [28]. As indicated, the transfer learning method is applied in the AlexNet model and fine-tuning the convolutional networks, in which the information and parameters stored in previous network trainings are used [20, 24], which makes it possible to utilize the pre-trained deep networks in a variety of applications. In fact, such kind of learning has a large number of applications in feature extraction, fine-tuning the network, and applying pre-trained models. Using transfer learning can be extremely useful and cost effective because training deep networks requires advanced hardware facilities such as Graphics Processing Unit (GPU) and takes time. Thus, the pre-trained model of AlexNet is fine-tuned to be utilized for appropriate application. Therefore, the first to seventh layers are applied without transformation, and the kernel size and the number of filters related to the eighth to eleventh layers are altered by converting a number of fully connected layers into a convolutional one. Proper fine-tuning the number of filters and the filter window affects the results obtained significantly. Convolution is regarded as the same type and stride is considered to be one in most layers. Using the same convolution allows the input dimensions to remain constant and deters information waste. The best values were tested for the appropriate application by investigating different ones. Utilizing smaller filters and more convolutional layers reduced the number of network parameters and fear of overfilling. Convolution layers can visualize and display the results obtained in the layers on the input data, which helps observe the results achieved by applying filters to the convolution layers. The inverse convolution method was used to display the results obtained from the convolution layers on the input data [25]. Displaying a heat-map of the input data to specify the objects in the image is considered as another ability of the convolutional network [34], which was conducted for convolutional layers so that the result of applying different filters to the layers can be observed and the filters can be utilized to detect different areas in the image. The first seven layers were transferred without transformation in the proposed model. Then, the layers were fine-tuned from the eighth one onwards so that the first fully connected layer was converted into convolutional ones and placed the remaining two fully connected layers after the eighth convolutional one. The final model includes six convolutional layers and two fully connected ones, in which the network learning parameters are reduced significantly. Figure 1 shows the block diagram related to the proposed method with the fine-tuned AlexNet network.

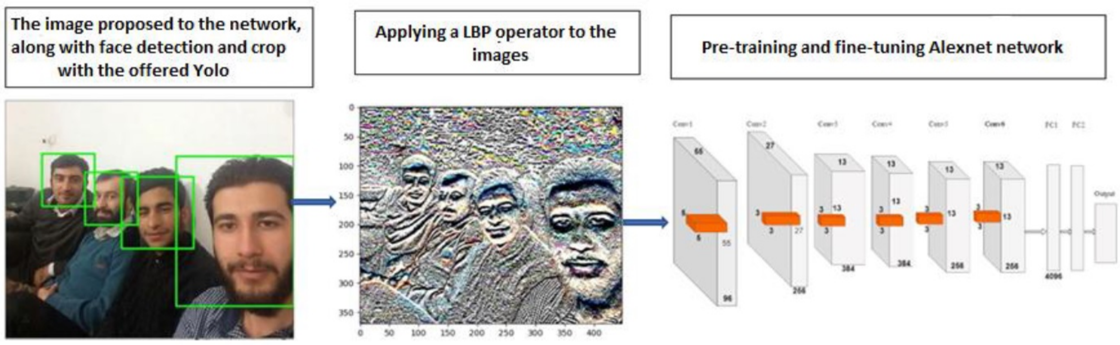

Figure 1: Block diagram of the proposed method

In order to visualize the results achieved from applying filters in convolutional layers, their effect on the input data are investigated by a method called reverse convolution, the output of which includes the input data in the first layer of the network, which is affected by the filters in the convolution layer. Identifying different filters for various applications based on their results on the input data is regarded as the advantage of such display [30], in which the network inputs include the undamaged images related to the dataset, along with applying a LBP operator to the images. A Rectified Linear Unit (ReLU) activator function is applied after each convolution or a fully connected layer. Then, task-specific loss functions are used to learn the weights of the network. The LBP operator is utilized separately for each RGB color channels related to the images entering the network to increase the channel size in each image to six. Figure 2 illustrates the output of applying the operator on the sample image.

Figure 2: LBP output on three RGB image channels

The network estimates the location of the face and gender information when an area is classified as a face. The face related to the images cropped by the proposed Yolo becomes a candidate to enter the main network to estimate the gender recognition parameter. Yolo optimizes a multicast loss function during training the model which includes the objective function of trust, classification, regression, and responsible for the absence of any object. However, face detection is considered as a binary classification obstacle. To optimize the total objective function in face detection, the weights are modified to 2: 1: 0.5: 0.5 experimentally. The final objective function is obtained as Equation (3.1).

$$L = 2 \cdot \sum L_{reg} + \sum L_{objconf} + 0.5 \cdot \sum L_{noobjconf} + 0.5 \cdot \sum L_{cld} \tag{3.1}$$

where L_{reg} indicates the loss of coordinate regression, $L_{narconf}$ is regarded as the loss of confidence of the box surrounded by objects, $L_{noobjconf}$ represents the loss of trust in boxes surrounded by no objects, and L_{cld} is considered as the loss of classification. Typically, the predicted location overlap and correct tags of the corresponding images are commonly applied as optimization estimates, while MSE function is used as the regression loss. However, there is a gap between optimization and MSE and maximizing the overlap threshold value. In particular, non-overlapping frames cannot be optimized, for which a generalization to the quality of the generated predictions is proposed as a new metric called GIoU, which creates a strong correlation between the optimization of MSE function and metric. The regression loss function was improved by combining the principal soft error l_n with the GIoU weight loss following [19]. The loss function used by YOLOV3 is a binary cross entropy loss (BCELoss), which is represented as:

$$\text{BCELoss} = -\frac{1}{n} \times \sum_i (t_i \times \log(o_i) + (1 - t_i) \times \log(1 - o_i)) \tag{3.2}$$

where o_i is the output value and t_i is the target value. Since the structure of the network layer needs to be changed after that, in order to prevent the predicted value from being too large, the negative predicted value causes the loss function to take too long to converge or have difficulty converging, so a sigmoid layer is added before the BCEloss loss function is used; the variable is mapped between zero and one; and then, the value is transferred to the loss function for calculation. Therefore, replace the loss function with the BCEWithLogitsLoss loss function with better numerical stability, as shown in the following formula:

$$\text{BCEWithLogitsLoss} = -\frac{1}{n} \times \sum_i (t_i \times \log(sigmoid(o_i)) + (1 - t_i) \times \log(1 - sigmoid(o_i))) \tag{3.3}$$

The BCEWithLogitsLoss loss function integrates the sigmoid layer into the BCELoss class and uses the log-sum-exp technique to achieve numerical stability. The performance of the trained model on the test dataset was better than that of the original YOLOV3 model. In the process of training with the AFLW dataset, the log information of each iteration of training of the improved Hyper-Yolo-Face model was collected, including the accuracy of face detection, the average IOU value, the accuracy of correct classification, the total number of detected faces, and the recall rate. By visualizing the information, as shown in Figure 3, the loss function converged steadily in the first 2000 iterations as the number of iterations increased.

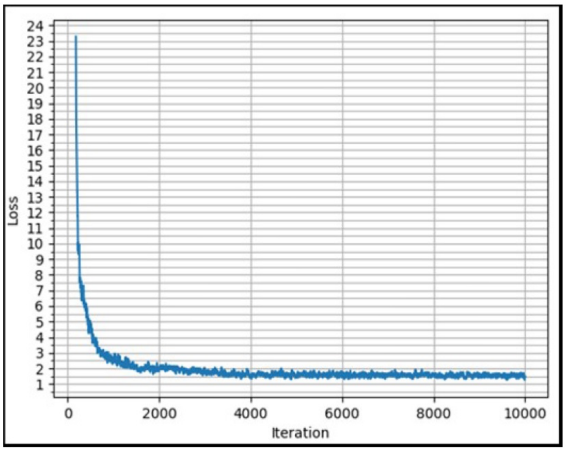

Figure 3: Loss curve of model training

3.1 Network training process

The AFLW dataset [10] which is utilized to train and test the proposed network includes 25,993 faces with 21,997 full-length real images related to different faces, facial expressions, races, ages, and genders, among which 2400 ones are randomly selected for testing and the rest for training. The proposed Yolo algorithm is applied for face detection and crop, which was discussed in Section 3 of the architecture and its loss functions. Gender recognition is considered as a two-pronged issue similar to face detection. A set of feature maps extracted during the network training process are regarded as distinguishing features of male and female classes. The Softmax loss is calculated by Equation (3.4) for a candidate area with an overlap of 0.5 with the target map.

$$loss_G = -(1-g)\cdot\log(1-p_g) - g\cdot\log(p_g) \tag{3.4}$$

where $g = 0$ when the gender is male, otherwise $g = 1$. Here, (p_0, p_1) is considered as the two-dimensional probability vector calculated from the network. The total loss is calculated as the weighted sum of each of the two losses as in Equation (3.5).

$$loss_{full} = \sum_{i=1}^{i=2} \lambda_{t_i} loss_{t_i} \tag{3.5}$$

where t_i is regarded as entry i of the tasks $T = \{D, G\}$ to detect the face and determine the gender. The values $(\lambda_D = 1, \lambda_G = 2)$ are selected for the experiments. The computational complexity in a deep learning algorithm is almost extremely high during network training. However, the above-mentioned complexity is reduced during network testing. In addition, the amount of data and appropriate precision are considered as important parameters, which determine how long or short the training time can be since the aforementioned algorithms are regarded as prompt and real-time.

4 Results

Taking the Precision rate (P) and Recall rate (R) as evaluation indexes, the method was compared with R-CNN, FAST-RCNN, and FASTER-RCNN with different improvement strategies. In order to accelerate the convergence speed of the network and avoid over-fitting, the impulse constant was set to 0.9, the weight attenuation coefficient to 0.0005, and the initial learning rate to 0.0005. Face detection results for the AFLW, PASCAL, and FDDB datasets are presented as follows. The FDDB database [7] contains 2,845 images including 5,171 images collected from news articles on the Yahoo Website. DP2MFD [19], Cascade CNN [13], and Hyper face [17] are among recently published methods compared in the evaluation. The FDDB dataset is extremely challenging for the proposed method and other R-CNN-based face detection methods due to its large number of dark and small faces. Some of the above-mentioned faces are not included in the candidate search areas. In addition, resizing small faces to 227×227 adds some distortions to the face, leading to lower detection points for enclosing frames. However, the performance of the proposed method has more precision compared to recently published face detection methods which are based on deep learning such as DP2MFD [19] and Faceness [28] on the FDDB dataset with a Mean Average Precision (MAP) of 99.5%. Further, adding convolutional layers results in extracting higher surface properties and increasing detection precision. The results obtained from fine-tuning the network indicate an improvement in network performance. Figures 4 demonstrate the precision-recall curves of different face detection methods related to the AFLW datasets, respectively.

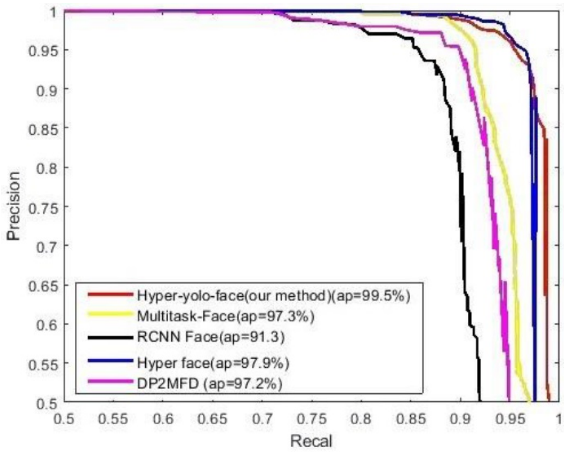

Figure 4: Testing the performance of the proposed method in face detection on the AFLW dataset (The numbers in the guide represent MAP for the relevant dataset)

Figures 4 show that Hyper-Yolo-face works with a wider margin than R-CNN-Face and other methods due to the proposal of utilizing the LBP operator, and the new regression loss function combined with the MSE loss and the GIOU loss, which can be observed from their map values in the AFLW dataset. The performance of gender recognition on the CelebA [14] and LFWA [5] datasets including gender information are tested here. The CelebA dataset includes 10,000 attributes with 200,000 images, while LFWA dataset includes 13,233 images with 5,749 attributes. The adopted approach is compared with FaceTracer [12], PANDA-w [31], PANDA-1 [31], and Hyper face [17]. We also compare our proposed Hyper-YOLO-face and the original YOLOv3 on the AfLW dataset, and the results are shown in Fig. 5. In the experiments, a proposal box is considered to be positive if the IoU between it and the ground truth bounding box is larger than 0.5; otherwise, it is labeled as negative one.

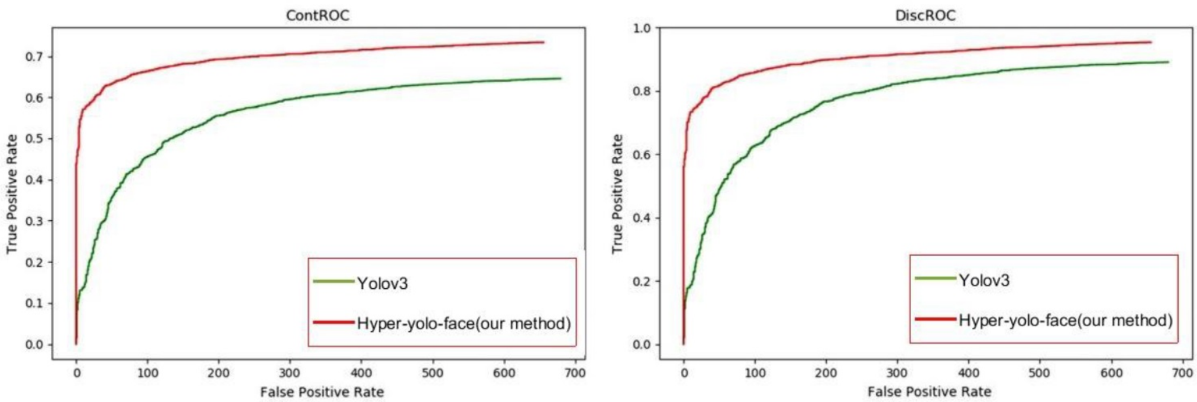

Figure 5: Detection results on the AFLW dataset. Left: ContROC. Right: DiscROC

Table 1 represents the performance of different gender recognition methods and Detection speed of Hyper-YOLO-face. The adopted method works better on both datasets than all of those listed in the table.

Table 1: Comparing the performance of the proposed model in gender detection (%) on CelebA and LFWA datasets

Method	**CelebA**		**LFWA**		**Detection Speed (fps) (1080 Ti Gpu)**
	Precision	Recall	Precision	Recall	
FaceTracer [28]	91	90.8	84	83.56	11
PANDA-W	93	92.6	86	85.1	13
PANDA-1 [16]	97	95.45	92	90.8	13
Hyper face [6]	97	96	94	92.7	38
Hyper-Yolo-face [our method]	98.2	98	98.25	98	48

5 Discussion and Conclusion

A combined two-task deep learning method called Hyper-Yolo-face was presented for face detection and gender recognition simultaneously. The combination of YOLOv3, transfer learning in AlexNet architecture, and LBP operator was applied as the main structure in the proposed face detection system, and a new loss function was offered to improve recognition.

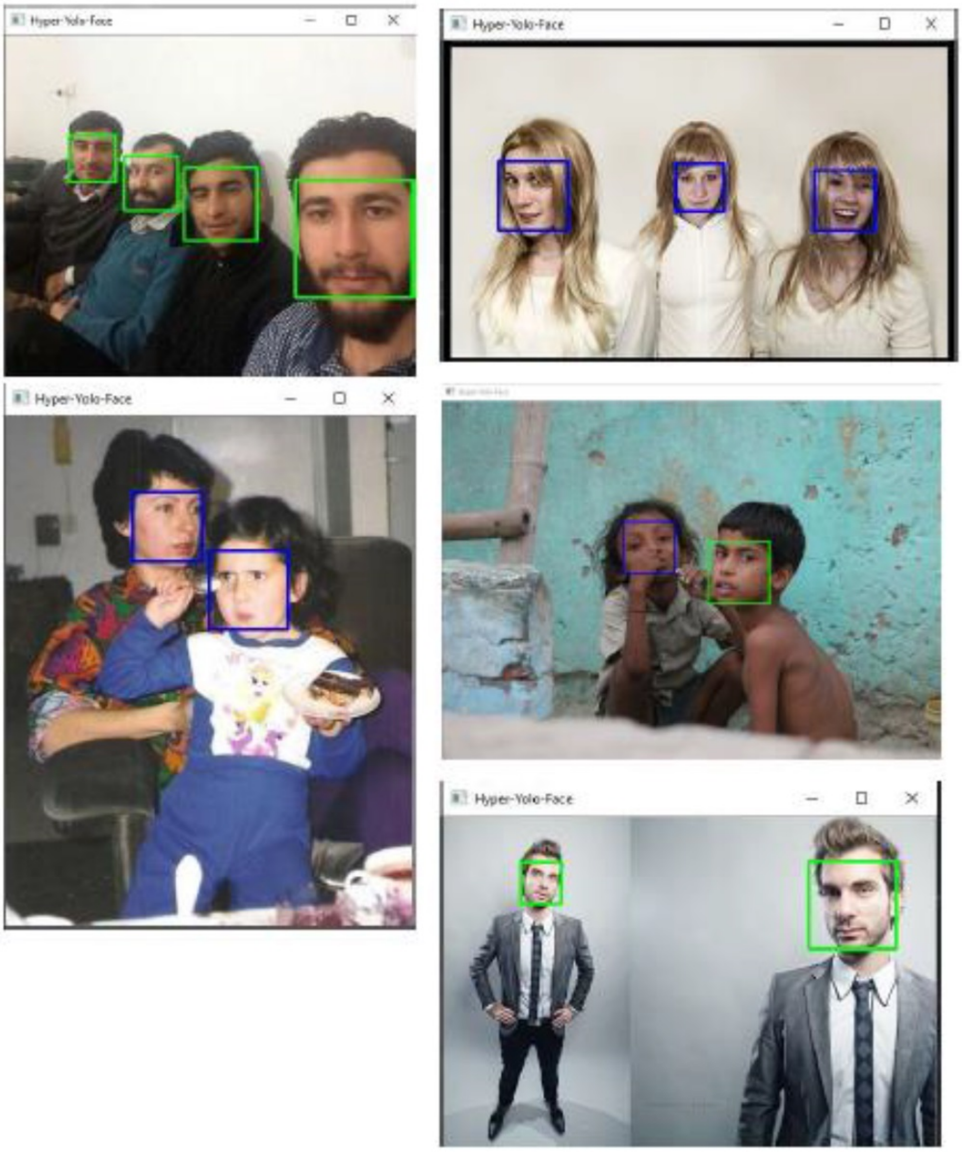

Figure 6: The output of the system from the proposed method in which the blue and green boxes represent the identified female and male faces, respectively

The results were presented based on fine-tuning and transfer learning of the pre-trained model on the database. Precise fine-tuning of AlexNet Convolution Network by converting fully connected layers to convolution ones and using appropriate filters for convolution layers was among the operations which improved the efficiency and precision of the network. Based on the results, the improved method can achieve a balance between performance and speed due to its adaptability and flexibility, which leads to more precise results by fine-tuning specific scenarios. Some improvements may be utilized more like a larger input image size, an appropriate anchor box scale for specific scenarios, and training data. Figure 6 illustrates five qualitative results related to the proposed method on the indicated datasets. As observed, the proposed method can perform face detection and gender recognition tasks on images including gestures, brightness, and sharp resolution changes with different backgrounds simultaneously and accurately.

References

[1] Z. Cai and N. Vasconcelos, *Cascade R-CNN: delving into high quality object detection*, IEEE Conf. Comput. Vision Pattern, 2018, pp. 6154–6162.

[2] W. Chen, H. Huang, S. Peng, C. Zhou and C. Zhang, *YOLO-face: a real-time face detector*, Visual Comput. **37** (2020), no. 4, 805–813.

[3] Y. Chen, Y. Tai, X. Liu, C. Shen and J. Yang, *FSRNet: end-to-end learning face super-resolution with facial priors*, Proc. IEEE Conf. Comput. Vision Pattern Recog., 2018, pp. 2492–2501.

[4] M.F. Hansen, M.L. Smith, L.N. Smith, M.G. Salter and B. Grieve, *Towards on-farm pig face recognition using convolutional neural networks*, Comput. Ind. **98** (2018), 145–152.

[5] G.B. Huang, M. Ramesh, T. Berg and E. Learned-Miller, *Labeledfaces in the wild: a database for studying face recognition in unconstrained environments*, Workshop on Faces in'Real-Life'Images: Detection Alignment Recognition, 2008.

[6] N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali and M. Zareapoor, *Hybrid deep neural networks for face emotion recognition*, Pattern Recognit. Lett. **115** (2018), 101–106.

[7] V. Jain and E. Learned-Miller, *Fddb: A benchmark for face detection in unconstrained settings*, UMass Amherst Tech. Rep. **2** (2010), no. 6.

[8] D.K. Jain, P. Shamsolmoali and P. Sehdev, *Extended deep neural network for facial emotion recognition*, Pattern Recog. Lett. **120** (2019), 69–74.

[9] M.R. Ju, H.B. Luo, Z.B. Wang, M. He, Z. Chang and B. Hui, *Improved YOLO v3 algorithm and its application in small target detection*, Acta Opt. Sin. **39** (2019), no. 7, 0715004.

[10] M. Kostinger, P. Wohlhart, P. Roth and H. Bischof, *Annotated facial landmarks in the wild: a large-scale, real-world database for facial landmark localization*, IEEE Int. Conf. Comput. Vision Workshops, 2011, pp. 2144–2151.

[11] A. Krizhevsky, I. Sutskever and G.E. Hinton, *Imagenet classification with deep convolutional neural networks*, In F. Pereira, C. Burges, L. Bottou and K. Weinberger (Eds.), Adv. Neural Inf. Process. Syst. **25** (2012), 1097–1105.

[12] N. Kumar, P.N. Belhumeur and S.K. Nayar, *FaceTracer: a search engine for large collections of images with faces*, Eur. Conf. Comput. Vision (ECCV), 2008, pp. 340–353.

[13] H. Li, Z. Lin, X. Shen, J. Brandt and G. Hua, *A convolutional neural network cascade for face detection*, In IEEE Conference on Computer Vision and Pattern Recognition, (2015), 5325–5334.

[14] Z. Liu, P. Luo, X. Wang and X. Tang, *Deep learning face attributes in the wild*, Proc. IEEE Int. Conf. Comput. Vision, 2015, pp. 3730–3738.

[15] T. Ojala, M. Pietikäinen and D. Harwood, *Performance evaluation of texture measures with classification based on Kullback discrimination of distributions*, Proc. 12th IAPR Int. Conf. Pattern Recog. (ICPR 1994), 1994, pp. 582–585.

[16] R. Ranjan, V.M. Patel and R. Chellappa, *A deep pyramid deformable part model for face detection*, Int. Conf. Biometrics Theory Appl. Syst. 2015, pp. 1–8.

[17] 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 Trans. Pattern Anal. Machine Intell. **41** (2016), no. 1, 121–135.

[18] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, *You only look once: unified, real-time object detection*, Proc. IEEE Conf. Comput. Vision Pattern Recog. 2016, pp. 779–788.

[19] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid and S. Savarese, *Generalized intersection over union: a metric and a loss for bounding box regression*, Proc. IEEE/CVF Conf. Comput. Vision Pattern Recog., 2019, pp. 658–666.

[20] Y. Savani, C. White and N.S. Govindarajulu, *Intra-processing methods for debiasing neural networks*, Proc. Adv. Neural Inf. Process. Syst. **33** (2020), 2798–2810.

[21] O.R. Seryasat and J. Haddadnia, *Assessment of a novel computer aided mass diagnosis system in mammograms*, Biomed. Res. **28** (2017), no. 7, 3129–3135.

[22] O.R. Seryasat and J. Haddadnia, *Evaluation of a new ensemble learning framework for mass classification in mammograms*, Clinical Breast Cancer **18** (2018), no. 3, 407–420.

[23] O.R. Seryasat, J. Haddadnia and H. Ghayoumi Zadeh, *Assessment of a novel computer aided mass diagnosis system in mammograms*, Iran. Quart. J. Breast Disease **9** (2016), no. 3, 31–41.

[24] R. Soundararajan and S. Biswas, *Machine vision quality assessment for robust face detection*, Signal Process.: Image Commun. **72** (2019), 92–104.

[25] X. Sun, P. Wu and S.C. Hoi, *Face detection using deep learning: an improved faster RCNN approach*, Neurocomput. **299** (2018), 42–50.

[26] M.N.A. Tawhid and E.K. Dey, *Gender recognition system from facial image*, Int. J. Comput. Appl. **180** (2018), no.23, 5–14.

[27] M.J. Uddin, P.C. Barman, K.T. Ahmed, S.A. Rahim, A.R. Refat and M. Abdullah-Al-Imran, *A convolutional neural network for real-time face detection and emotion & gender classification's*, SR J. Electr. Commun. Eng. (IOSR-JECE) **15** (2020), no. 3, 37–46.

[28] S. Yang, P. Luo, C.C. Loy and X. Tang, *Faceness-net: face detection through deep facial part responses*, IEEE Trans. Pattern Anal. Mach. Intell. **40** (2018), no. 8, 1845–1859.

[29] E. Zangeneh, M. Rahmati and Y. Mohsen Zadeh, *Low resolution face recognition using a two-branch deep convolutional neural network architecture*, Expert Syst. Appl. **139** (2020), 112854.

[30] D. Zeng, F. Zhao, S. Ge and W. Shen, *Fast cascade face detection with pyramid network*, Pattern Recognit. Lett. **119** (2019), 180–186.

[31] N. Zhang, M. Paluri, M. Ranzato, T. Darrell and L. Bourdev, *Panda: pose aligned networks for deep attribute modeling*, IEEE Conf. Comput. Vision Pattern Recogn., 2014, pp. 1637–1644.

[32] M.M. Zhang, K. Shang and H. Wu, *Learning deep discriminative face features by customized weighted constraint*, Neurocomput. **332** (2019), 71–79.

[33] C. Zhang, X. Xu and D. Tu, *Face detection using improved faster RCNN (2018)*, arXiv preprint arXiv:1802.02142, (2018).

[34] X. Zhao, X. Liang, C. Zhao, M. Tang and J. Wang, *Real-time multi-scale face detector on embedded devices*, Sensors **19** (2019), no. 9, 2158.