Providing a hybrid method for face detection and gender recognition using deep learning to improve accuracy

Peyman Jebraelzadeh[1], Asghar Charmin[2*], Hamid Vahdati[3], Mohsen Ebadpour[4]
[1]Department of Electrical Engineering,Ahar Branch, Islamic Azad University, Ahar, Iran
[2]Department of Electrical Engineering,Ahar Branch, Islamic Azad University, Ahar, Iran
[3]Department of Electrical Engineering,Ahar Branch, Islamic Azad University,Ahar, Iran
[4]Department of Electrical Engineering, Ahar Branch, Islamic Azad University, Ahar, Iran

Article Info

Article history:

ReceivedJun
Revised Nov
Accepted Jan

Keywords:

Deep learning
Face detection
Local binary model
Yolo model
Gender recognition

ABSTRACT (10 PT)

In general, identifying and locating faces in images or videos is considered as the first step in face recognition. It is quite clear that an accurate detection algorithm can significantly benefit system performance and vice versa. Therefore, face recognition is one of the key steps in the application of face recognition systems. In deep learning algorithms are able to learn high-level features, which have been highly regarded by researchers for use in the field of machine vision, as well as in a variety of fields such as image classification, and human gesture estimation, which are the key activities for image perception. In this paper, we present a hybrid method called Hyper-Yolo-face to identify faces and recognize the gender of a given image using deep convolutional neural networks, the Yolo algorithm, and local binary patterns. The proposed network architecture is based on the AlexNet model and the integration of the binary pattern operator and Yolov3, which results in increasing performance and accuracy. Yolo changes the architecture of face recognition systems and looks at the problem of recognition as a regression problem which goes directly from the pixels of the image to the coordinates of the box and the probability of the classes. Experiments on the AFLW and FDDB datasets indicated that the proposed model performs significantly better than other algorithms and methods and improves detection accuracy.

Corresponding Author:

Asghar Charmin,
Department of Electrical Engineering, Islamic Azad University, Ahar, Iran
Email: a_charmin@sut.ac.ir

1. INTRODUCTION

Convolutional Neural Networks are considered as one of the most important -deep learning methods in which several layers are taught in a powerful way. This method is very efficient and is one of the most common methods in various applications of machine vision. Despite advances in deep learning theory, it needs better understanding and optimization of neural network architecture to improve desirable features such as invariance and class recognition. However, further development of techniques for creating or collecting more comprehensive training data which enables networks to better learn features which are robust and subject to changes such as geometric transformations and blocking is promising. With recent advances in the use of CNN in the field of computer vision, popular models of convolutional neural networks have emerged. The present study aimed to evaluate the most common models and then chose the AlexNet and Yolo models in this study.AlexNet architecture is the first deep architecture introduced by Geoffrey et al. as the pioneers of deep learning. This model is a simple but powerful architecture that paved the way for great research without which there would be no deep learning of what it is now.

This architecture consists of five layers of convolution and three fully connected layers. The convolution layers with different sizes are the same filters which are responsible for extracting the properties from the input image.

Basically, each convolution layer consists of a large number of isometric kernels. For example, the first layer of AlexNet including 96 kernels with a size of 11 × 11 × 3 [1] Face recognition and analysis are challenging problems in machine vision, which has been studied for some applications such as face verification, face tracking, and person identification [2, 3 and 4]. In this paper, a new CNN-based method is presented for recognizing faces and gender from a specific image. Binary pattern is considered as a powerful tool for tissue analysis since it uses both statistical and structural characteristics of the texture. In addition, this operator is used while feeding features to the network and increase the input dimensions to 6 channels, which will improve the accuracy of detection.

In the operator of local binary pattern, local texture patterns are extracted by comparing the value of adjacent pixels with the value of the central pixel and are represented by binary codes. The local binary model was first proposed by Ocalan et al (1996)as one of the most common descriptors due to its resistance to brightness changes, low computational complexity, and ability to encode details for the purpose ofimproving accuracy. The information in the attributes is hierarchically distributed across the network. The deeper layers are category-dependent and suitable for learning complex tasks such as face and gender recognition. On the other hand, the lower layers estimate edges and corners and have better locating characteristics. These layers are more suitable for finding the landmarks of the image and estimating the gesture [5]. Obviously, all the middle layers of a deep CNN should be used to teach various tasks such as face and gender recognition.By considering features and information in this paper, the following ideas are presented. The new CNN architecture is suggested, which performs face and gender recognition by combining the middle layers of the network. This method is based on the architecture of AlexNet model [1] and uses a proposed algorithm [6] called “Hyper-Yolo-face” instead of a selective search algorithm to generate area suggestions and face crop.

In this method, the faces of the images are quickly detected and cut by the proposed Yolo algorithm and fed to the proposed network along with additional information added via local binary patterns (LBP). Further, the YOLOv3 architecture is selected as the structure of our face recognition network, which is developed and modified in several respects including a proposal for a new regression loss function which combines MSE loss and GIOU loss, as well as providing the training of more appropriate enclosed face recognition boxes with k-clustering. The rest of this paper is organized as follows. Section 2 reviews the previous studies. Section 3 describes the details of the proposed method, as well as the implementation of the proposed deep CNN AlexNet and Hyper-Yolo-face approaches. The results of the proposed approach on the AFLW and FDDB datasets are described in Section 4. Finally, the conclusions of this paper are summarized and discussed in Section 5.

2. Reviewing the previous studies

Ramaiah et al. [7] proposed a convolutional neural network to solve the problem of changes in brightness and head position in face recognition. In the proposed method, individuals are distinguished from each other using local patterns on their faces. The accuracy of this method improved by 96.4% compared to the previous methods, and the detection accuracy of 95.99% was obtained by evaluating this method on the Yale database. Gao et al. [8] suggested a new type of building block for deep architecture called an automatic encoder with an observer to identify the face with a training sample from each individual. In this method, all the different faces are first modeled for mapping with the normal face of each person, then the features related to the same person are extracted, and finally, a self-encoder is used with the feature supervisor. In addition, the extracts which are resistant to light scattering and opacity are extracted, making face recognition easier. By evaluating this method, the identification accuracy in AR datasets, Extended Yale-B, CMU-PIE, and Multi-PIE was 21.85, 22.82, 79.82, and 97.93%, respectively. Zhang et al. [9] used Spars coding and Softmax classification neural networks to solve the problem of changes in brightness, low state, and image quality in face recognition. Face image preprocessing was used to build and train a deep network hierarchy. The deep neural network was trained by a recursive algorithm and optimized using two different schemes. Based on the result, ORL, Yale, Yale-B, and PERET datasets indicated a recognition accuracy of 5.97, 67.94, 82, and 78.92%, respectively.Viola-Jones Detector [10] is a traditional method which uses Haar-like feature classifications for instant face recognition and works well for full faces with enough light. In addition, face recognition methods based on the Modifiable Parts Model (DPM) [11] are presented in which a face is defined as a set of parts [12, 13]. Some features such as HOG or Haar wavelets do not receive distinct face information in different gestures or brightness changes in unrestricted face recognition. To overcome these limitations, some studies have focused on using various CNN-based face recognition methods [14-18], which produced new results on many of the existing challenging face recognition datasets. Other face recognition methods include NPDF Faces [19], Adapt [20], and [21].

3. Proposed approach and architecture

In this section, it is explained the results of research and at the same time is given the comprehensive discussion. Results can be presented in figures, graphs, tables and others that make the reader understand easily [2, 5]. The discussion can be made in several sub-chapters.In this study, a hybrid CNN model is presented for face and gender recognition, which includes two main parts. In the first part, the proposed Yolo algorithm called Yolo_face crops people's faces from images and scales them to 227 × 227 pixels. Regarding the second part, a CNN receiving these resized cropped areas and categorizing them as face or non-face is used, which applied the LBP feature to increase the size of the input images to 6 color channels.

The present proposed network consists of five convolutional layers with three fully connected layers, the network structure of which is shown in Figure 1. In addition, layers are combined and a separate grid is used from AlexNet. In general, the direct combination of these features is considered as one of the simplest methods. The different dimensions for these layers are in order, and one of the disadvantages of this feature is that they cannot be easily combined with each other. Therefore, convolutional layers are added to obtain feature maps which are compatible based on the dimension at the output. Then, the output of these layers is combined together to form a feature map with dimensions. These dimensions continue to train a more dual-task detection system. As a result, a layer of nuclear convolution () is added to reduce these dimensions. Further, a fully connected layer () including 3072 feature vector outputs is added. At this point, the network is divided into two separate branches including face recognition and gender recognition tasks.

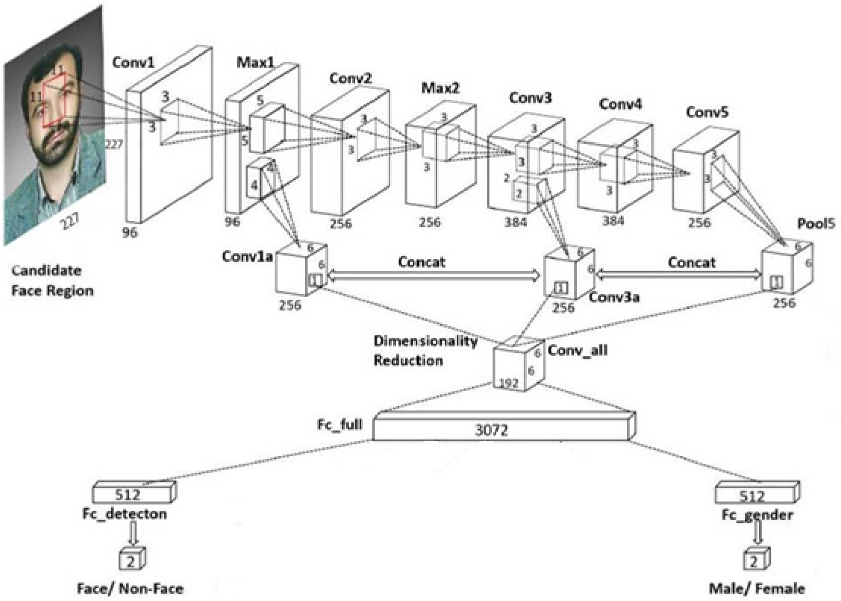

Figure 1. The proposed network structure based on AlexNet

In the next procedure, fully connected layers with the dimensions of 512 are added. Finally, a fully connected layer is added to each branch to take advantage of the face tags and gender identification. Then, an activation function or ReLU is used after each convolution or a fully connected layer. In addition, task-specific loss functions are used to learn the weights of this network, which generates a binary number for each pixel according to the3 × 3 pixel label. Further, labels are obtained by thresholding the amount of neighboring pixels with the center pixel value. In this case, label 1 is placed for the pixels with a value greater than or equal to the value of the central pixel, while label 0 is used for those with the values less than the value of the central pixel. Finally, the labels are rotated side by side to form an 8-bit number. Figure 2 shows the way the operator works.

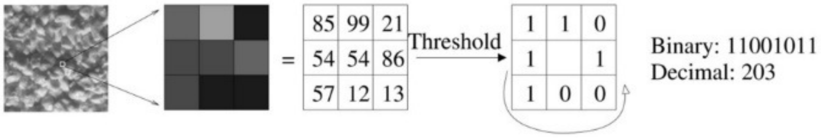

Figure 2. Local binary pattern operator [22]

During the recent years, LBP method has experienced a lot of changes in order to improve performance in various applications such as improving separation strength, increasing tolerance to changes, selecting a neighbor, and combining with other methods. In addition, for each RGB color channel, the LBP operator is separately used on each of the color channels of the input images to the network to increase the channel size of each image to six dimensions. The output of this operator on the sample image is shown in Figure 3.

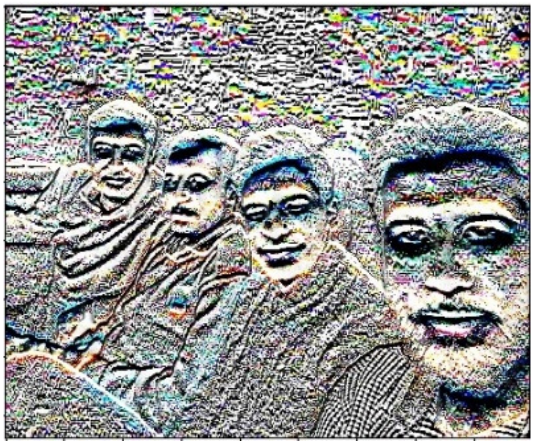

Figure 3. LBP output on three RGB image channels

The network estimates the location of the face and gender when an area is classified as a face. The cropped face of the images by the proposed Yolo method can be an appropriate alternative for entering the main network to estimate the gender determination parameter. In this case, Darknet-53 is used as the structure of the face recognition network. Adjusting the number of searches for clustered boxes experimentally with the goal of a proper face box for face recognition is considered as the most important step. In the procedure, enclosed k boxes are randomly selected as the initial clustering centers, and then the IoU of enclosed k boxes and all other enclosed boxes are calculated. In addition, all face labels are classified into k class using the IoU as the criterion for overlapping enclosed boxes. Further, the mean values of the enclosed box size of class k are considered as the centers of the new cluster. Additionally, this process repeated until convergence. Finally, horizontally enclosed boxes are moved vertically for face recognition.

During the training of this model, Yolo optimizes a multi-part loss function which consists of objective function of reliability, classification, regression, and the absence of any object. However, face recognition is a binary classification problem. To optimize the total objective function in face recognition, the weights were experimentally modified to 2: 1: 0.5: 0.5.

The final objective function is obtained as Equation (1):

$$L = 2 \cdot \sum L_{reg} + \sum L_{objconf} + 0/5 \cdot \sum L_{noobjconf} + 0/5 \cdot \sum L_{cld} 1 \qquad (1$$

where L_{reg} indicates the coordinate regression loss, $L_{narconf}$ shows the confidence loss of the enclosed box with objects, and $L_{noobjconf}$is the loss of trust in enclosed boxes with no objects, and $L_{cls}$is considered as the classification loss. Traditionally, the IoU predicted location and the corresponding tags of the corresponding monitored data are commonly used as optimization estimates, and the MSE function is used as the regression loss. However, there is a gap between optimizing MSE and maximizing IoU. In most cases, it is almost impossible to optimize for non-overlapping boxes. To solve this problem, a generalization to the IoU as a new metric was proposed called GIoU. The new metric has a strong correlation between optimizing the MSE function and the metric itself. Using [23], the regression loss function was improved by combining the principal soft error $l_n$with the GIoU weight loss. New regression losses can be calculated as Equations (2-4).

$$GIoU = IoU - \frac{A_{C-U}}{A_C} \qquad (2$$

$$L_{GIoU} = 1 - GIoU \qquad (3$$

$$L_{reg} = \sum_{c=x.y.w.h} \sum \left(\left| \Delta \mathrm{c}_{pred} - \Delta c_{truth} \right| + \alpha \cdot L_{GIoU} \right)^2$$

$$= \sum \left(\left| \Delta \mathrm{x}_{pred} - \Delta x_{truth} \right| + \alpha \cdot L_{GIoU} \right)^2$$

$$+ \sum \left(\left| \Delta \mathrm{y}_{pred} - \Delta y_{truth} \right| + \alpha \cdot L_{GIoU} \right)^2$$

$$+ \sum \left(\left| \Delta \mathrm{w}_{pred} - \Delta w_{truth} \right| + \alpha \cdot L_{GIoU} \right)^2$$

$$+ \sum \left(\left| \Delta \mathrm{h}_{pred} - \Delta h_{truth} \right| + \alpha \cdot L_{GIoU} \right)^2 \quad (4$$

where A_c indicates the smallest convex set enclosing the predicted location, correct labels α are considered as a real value factor, and x, y, w, and h are the locations and the size of the binding boxes, respectively. The α factor was set to 0.1 in this model.

3.1. Training and testing process of the network

AFLW dataset [24] is used for training and testing the proposed network. This dataset contains 25,993 faces with 21,997 full-face real-time images, face shapes, race, age and gender, among which 2400 images were randomly selected for testing and the rest were used for training. In addition, various harm functions were used for training facial and gender recognition tasks. Further, the proposed Yolo algorithm was used for face recognition and crop, which was discussed in Section 3 of the architecture and its loss functions. Gender recognition is a two-pronged issue similar to face recognition. Regarding a candidate area with 0.5 overlap with the target map, the softmax loss was calculated from Equation (5).

$$\mathrm{loss_G} = -(1 - \mathrm{g}) \cdot \log(1 - \mathrm{p_g}) - \mathrm{g} \cdot \log(\mathrm{p_g}) \quad (5$$

where $g = 0$ when the gender is male; otherwise, $g = 1$.

Here (p_0, p_1) is the two-dimensional probability vector calculated from this network. The total loss is calculated as the weighted sum of each of the two losses as in Equation (6).

$$\mathrm{loss_{full}} = \sum_{\mathrm{i=1}}^{\mathrm{i=3}} \lambda_{\mathrm{t_i}} \mathrm{loss_{t_i}} \qquad (6$$

where t_i is the first of the tasks $T = \{D \text{و} G\}$ to identify the face and determine the gender.

The weight parameter is decided based on the importance of this task in the overall loss. Finally, values $\left(\lambda_D = 1 \text{و} \lambda_G = 2\right)$ were selected for our experiments.

4. Data analysis and results

The results of face recognition for the AFW and FDDB datasets are presented in this section. The AFW [25] dataset was compiled by Flicker, and the images in this dataset include many changes in appearance and viewpoint. There are a total of 468 faces in this dataset. The FDDB database [26] of 2,845 images contains 5,171 images collected from news articles on the Yahoo Web site. Some recently published methods which were compared in this evaluation include DP2MFD [27], CascadeCNN [28], and Hyperface [29].

The FDDB dataset is very challenging for this proposed method and other R-CNN-based face recognition methods due to its large number of dark and small faces.

First, some of these figures are not in the candidate search areas. Second, resizing small faces to the size of the input adds distortion to the face, leading to a low detection score. Despite these issues, the performance of our proposed model in comparison with recently published deep learning face recognition methods such as DP2MFD [27] and Faceness [30] on the FDDB dataset with map is equal to 99.4%. The accuracy-recall curves of the various detectors associated with the AFW and PASCAL face datasets are shown in Figures 4 and 5, respectively. The performance of various detectors is compared using the receiver performance characteristic curves (ROCs) on the FDDB dataset in Figure 6. As shown, our proposed method works better than all of the commercial and academic diagnostics reported on the AFW and PASCAL datasets. Hyper-yolo-face has an average accuracy of 99.4% and 98.34% for AFW and PASCAL datasets, respectively.

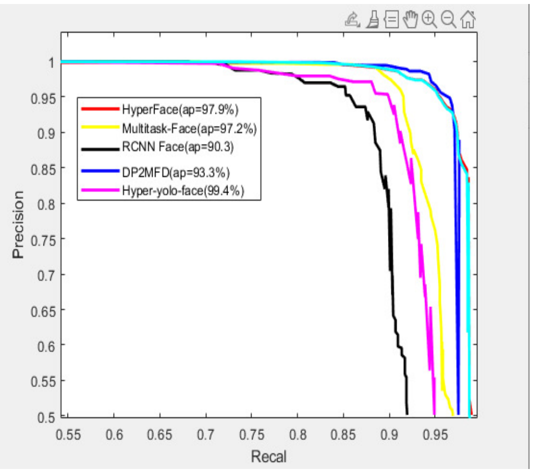

Figure 4. Testing the performance of the proposed method for face recognition on the AFW dataset. The numbers in the guide represent the average accuracy (map) for the relevant dataset.

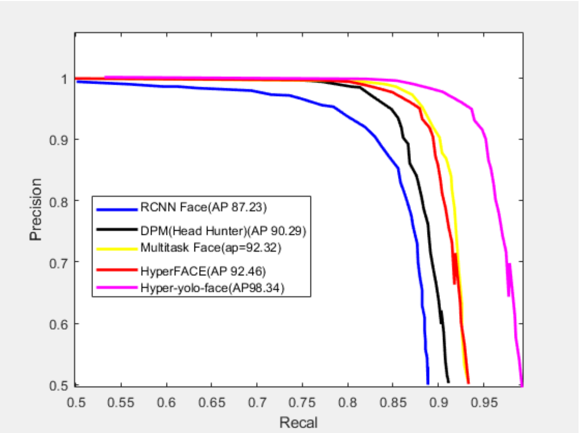

Figure 5. Testing the performance of the proposed method for face recognition on the PASCAL face dataset. The numbers in the guide represent the average accuracy (map) for the relevant dataset.

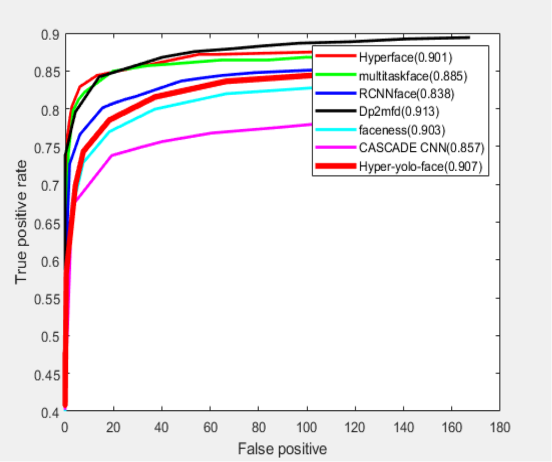

Figure 6. Testing the performance of the proposed method for face recognition on the FDDB dataset. The numbers in the guide represent the average accuracy.

Figures 5 and 6 clearly show that Hyper-Yolo-face works with a wider margin than R-CNN-Face and other methods. This productivity increase is mainly related to the proposal, the use of the LBP operator, and the new regression loss function, which we combined with the MSE loss and the GIOU loss, which can be seen from their map values in the AFLW dataset.

In this paper, the function of gender recognition on the CelebA [31] and LFWA [32] datasets containing gender information was tested. The CelebA dataset contains 10,000 attributes and 200,000 images.The LFWA dataset includes 13,233 images with 5,749 attributes. The present approach was compared with FaceTracer [33], PANDA-w [34], PANDA-1 [34] and Hyper face [28]. The performance of different gender determination methods is shown in Table 2. The present method works better on both datasets than all of the methods listed in the table.

Table 1. Comparison of performance (%) based on gender recognition on CelebA and LFWA Datasets

Method	CelebA	LFWA
FaceTracer[31]	91	84
PANDA-[64]	93	86
PANDA-1[64]	97	92
Hyper face	97	94
Hyper-Yolo-Face[ours method]	99	100

5. Discussion and conclusion

In this paper, a new double-task deep learning method called Hyper-yolo-face was presented for face and gender recognition simultaneously. A combination of YOLOv3, client architecture, and LBP binary pattern operator was used as the main structure of the proposed face recognition system, and a new loss function was introduced for improving recognition. The system performed similarly to the ResNet-101, but the speed was almost twice more than the ResNet-101. To achieve multidimensional integration, low-level features were combined with high-level features such as pyramidal networks. This design was able to make better use of the scales of a lot of visual information and thus give better performance to the multi-scale face recognition system. Comprehensive experiments were performed to compare the proposed method with some popular face recognition systems. The results showed that our improved method can achieve a balance between performance and speed. The proposed method is adaptive as well as flexible, which may be achieved by adjusting specific scenarios to achieve more accurate results. Some improvements such as a larger input image size, a suitable anchor box scale for specific scenarios, and more training data may be used. Figure 7 displays several qualitative results of our method on the AFW, AFLW, and FDDB datasets. As shown, our method can simultaneously perform both face recognition and gender determination on images including gestures, brightness, and sharp resolution changes with different backgrounds.

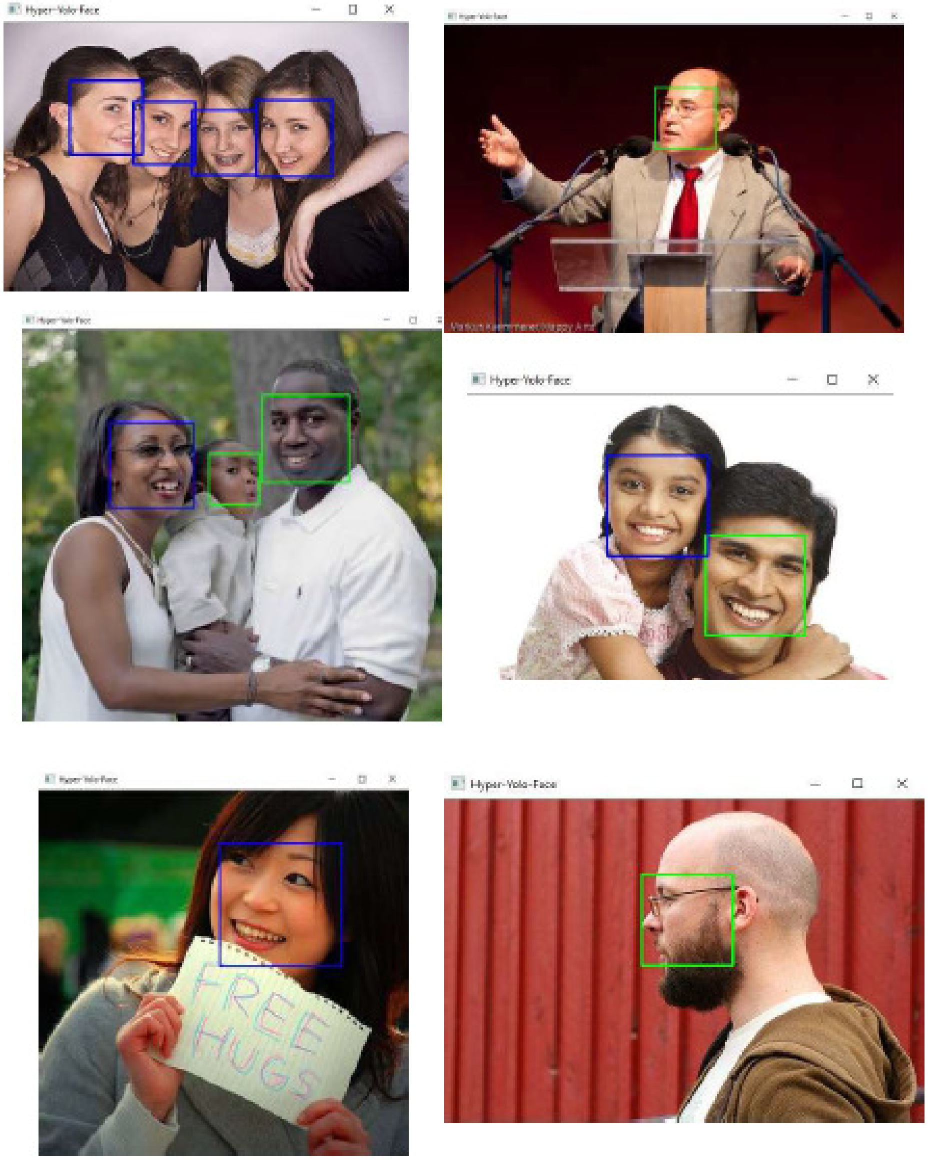

Figure 7. Qualitative results of the proposed method in which the blue boxes represent the identified female faces and the green boxes represent the male identified faces.

REFERENCES

[1] 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, editors, Advances in Neural Information Processing Systems 25, pages 1097–1105. Curran Associates, Inc., 2012.

[2] S. S. Farfade, M. Saberian, and L.-J. Li. Multi-view face detection using deep convolutional neural networks. In International Conferenc on Multimedia Retrieval, 2015.

[3] R. Ranjan, V. M. Patel, and R. Chellappa. A deep pyramid deformable part model for face detection. In International Conference on Biometrics Theory, Applications and Systems, 2015.

[4] S. Yang, P. Luo, C. C. Loy, and X. Tang. From facial parts responses to face detection: A deep learning approach. In IEEE International Conference on Computer Vision, 2015.

[5] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. CoRR, abs/1311.2901, 2013.

[6] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 779–788.

[7] Ramaiah, N. P., Ijjina, E. P., & Mohan, C. K. (2015, February). Illumination invariant face recognition using convolutional neural networks. In Signal Processing, Informatics, Communication and Energy Systems (SPICES), 2015 IEEE International Conference on (pp. 1-4). IEEE.

[8] Gao, S., Zhang, Y., Jia, K., Lu, J., & Zhang, Y. (2015). Single sample face recognition via learning deep supervised autoencoders. IEEE Transactions on Information Forensics and Security, 10(10), 2108-2118.

[9] Zhang, Z., Li, J., & Zhu, R. (2015, October). Deep neural network for face recognition based on sparse autoencoder. In Image and Signal Processing (CISP), 2015 8th International Congresson (pp. 594-598). IEEE.

[10] P. A. Viola and M. J. Jones. Robust real-time face Detection.International Journal of Computer Vision, 57(2):137–154, 2004.

[11] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan.Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence,32(9):1627–1645, Sept 2010.

[12] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2879–2886, June 2012.

[13] M. Mathias, R. Benenson, M. Pedersoli, and L. Van Gool. Facedetection without bells and whistles. In European Conference on Computer Vision, volume 8692, pages 720–735. 2014.

[14] R. Ranjan, V. M. Patel, and R. Chellappa. A deep pyramid deformablepart model for face detection. In International Conference on Biometrics Theory, Applications and Systems, 2015.

[15] 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, pages 5325–5334, June 2015.

[16] S. Yang, P. Luo, C. C. Loy, and X. Tang. From facial parts responses to face detection: A deep learning approach. In IEEE International Conference on Computer Vision, 2015.

[17] S. S. Farfade, M. Saberian, and L.-J. Li. Multi-view face detection using deep convolutional neural networks. In International Conference on Multimedia Retrieval, 2015.

[18] B. Yang, J. Yan, Z. Lei, and S. Z. Li. Convolutional channel features.In IEEE International Conference on Computer Vision, 2015.

[19] S. Liao, A. Jain, and S. Li. A fast and accurate unconstrained face detector. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015.

[20] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang. Probabilistic elastic part model for unsupervised face detector adaptation. In IEEE International Conference on Computer Vision, pages 793–800, Dec 2013.

[21] D. Chen, S. Ren, Y. Wei, X. Cao, and J. Sun. Joint cascade face detection and alignment. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, European Conference on Computer Vision, volume 8694, pages 109–122. 2014.

[22] Ojala, T., Pietikäinen, M., and Mäenpää, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. Pattern Analysis and Machine Intelligence, IEEE Transactions on 24, 7 (2002), 971–987.

[23] Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S.: Generalized intersection over union: a metric anda loss for bounding box regression. In: The IEEE Conference onComputer Vision and Pattern Recognition (CVPR) (2019)

[24] M. Kostinger, P. Wohlhart, P. Roth, and H. Bischof. Annotatedfacial landmarks in the wild: A large-scale, real-world databasefor facial landmark localization. In IEEE International Conference onComputer Vision Workshops, pages 2144–2151, Nov 2011.

[25] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2879–2886, June 2012.

[26] V. Jain and E. Learned-Miller. Fddb: A benchmark for face detection in unconstrained settings. Technical Report UM-CS-2010-009, University of Massachusetts, Amherst, 2010.

[27] R. Ranjan, V. M. Patel, and R. Chellappa. A deep pyramid deformable part model for face detection. In International Conference on Biometrics Theory, Applications and Systems, 2015.

[28] 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, pages 5325–5334, June 2015.

[29] R. Ranjan, V. M. Patel, and R. Chellappa. HyperFace: A Deep Multi-Task LearningFramework for Face Detection, LandmarkLocalization, Pose Estimation, and GenderRecognition. IEEE RANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 2016.

[30] S. Yang, P. Luo, C. C. Loy, and X. Tang. From facial parts responsesto face detection: A deep learning approach. In IEEE InternationalConference on Computer Vision, 2015.

[31] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributesin the wild. In International Conference on Computer Vision, Dec.2015.

[32] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, Oct. 2007.

[33] N. Kumar, P. N. Belhumeur, and S. K. Nayar. aceTracer: A Search Engine for Large Collections of Images with Faces. In European Conference on computer Vision (ECCV), pages 340–353, Oct 2008.

[34] N. Zhang, M. Paluri, M. Ranzato, T. Darrell, and L. Bourdev.Panda: Pose aligned networks for deep attribute modeling. In IEEE Conference on Computer Vision and Pattern Recognition, pages 1637–1644, 2014.