Hindawi
Journal of Sensors
Volume 2018, Article ID 5034684, 10 pages
https://doi.org/10.1155/2018/5034684

Research Article

Landmark-Guided Local Deep Neural Networks for Age and Gender Classification

Yungang Zhang [1] **and Tianwei Xu** [2]

[1] *Department of Computer Science, Yunnan Normal University, Kunming, Yunnan 650500, China*
[2] *Graduate School, Yunnan Normal University, Kunming, Yunnan 650500, China*

Correspondence should be addressed to Yungang Zhang; yungang.zhang01@gmail.com

Received 22 February 2018; Accepted 30 April 2018; Published 9 July 2018

Academic Editor: Mucheol Kim

Copyright © 2018 Yungang Zhang and Tianwei Xu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Many types of deep neural networks have been proposed to address the problem of human biometric identification, especially in the areas of face detection and recognition. Local deep neural networks have been recently used in face-based age and gender classification, despite their improvement in performance, their costs on model training is rather expensive. In this paper, we propose to construct a local deep neural network for age and gender classification. In our proposed model, local image patches are selected based on the detected facial landmarks; the selected patches are then used for the network training. A holistical edge map for an entire image is also used for training a “global” network. The age and gender classification results are obtained by combining both the outputs from both the “global” and the local networks. Our proposed model is tested on two face image benchmark datasets; competitive performance is obtained compared to the state-of-the-art methods.

1. Introduction

Age estimation and gender distinction from face images play important roles in many computer vision-based applications, such as visual surveillance, security control, and human-computer interaction. Over the last decades, many methods have been proposed to tackle the age and gender classification task.

In early works, pixel intensity values are used directly as input to train a classifier such as neural network [1, 2] or support vector machine (SVM) [3]. However, when the resolutions of images increase, directly using intensity values dramatically increases the scales of image features as well. Therefore, some feature reduction techniques such as principal component analysis (PCA) are applied to reduce the dimensions of image features [4]. Some image descriptors which are more powerful for image representation have also been used in the area of age estimation and gender recognition tasks, such as local binary patterns (LBP) [5], shift-invariant feature transform (SIFT) [6], Gabor filters [7], histogram of oriented gradient (HOG) [8], and biologically inspired features (BIF) [9]. Although the tasks of age estimation and gender classification have been widely investigated over the last decades, the results obtained are still far away from real applications [10, 11].

In recent years, deep learning, especially convolutional neural networks (CNN) [12–15], have become an important tool in computer vision applications. In many vision-based areas, such as image classification, object detection, pose estimation, visual tracking, CNN have achieved superior results. [16]. More recently, CNN have been employed in face image-based age and gender classification tasks [17–19]. However, as face images vary in a wide range under the unconstrained conditions (namely, in the wild), the performances of CNN still need to be improved, especially in age estimation tasks. Moreover, the time cost on training CNN models is quite expensive in most proposed solutions.

In order to reduce the cost on CNN model training, a local deep neural network (LDNN) was proposed [20] for gender recognition; the LDNN model can achieve state-of-the-art performance while the training cost is considerably reduced. More recently, a modified version of LDNN is proposed by Liao et al. [21]; this modified LDNN shares the same network architecture with the one used in [20]. In [21], the number of image patches used for network training is further reduced; 9 fixed image patches are selected for network learning. The modified LDNN model can be used for both age and gender recognition. However, in this model, the local image patches used for training are fixed; this may not work well on the unconstrained images without carefully preprocessing. In [21], the authors find the eye areas and the mouth area are crucial parts for age estimation, while only the eye areas are important for gender classification.

The success of LDNN in age and gender classification and the relative discoveries from the former LDNN works inspire us to propose a LDNN model for age and gender estimation. In our proposed model, the local image patch selection is based on the detected facial landmarks, that is, the image patches used for network learning are dynamically generated. Therefore, the number of image patches can be greatly reduced while all the important information in a face image can be kept.

In [20], the Sobel edge detector is used for local feature extraction. However, in [21], it is illustrated that using other feature detectors can obtain different performances. In our proposed model, the holistically-nested edge detection (HED) [22] is used for global feature extraction. The age and gender classification results are obtained by combining both the outputs from the "global" and the local networks.

The remainder of the paper is listed as follows. In Section 2, a brief review of related work on age and gender classification using CNN is given. Section 3 introduces the proposed local deep neural network for age and gender estimation. Section 4 presents the experiment settings and the experimental results and analysis. Conclusions and future work are included in Section 5.

2. Related Work

The successful applications of CNN on many computer vision tasks have revealed that CNN is a powerful tool in image learning. If enough training data are given, CNN is able to learn a compact and discriminative image feature representation. Therefore, many researchers propose to use CNN in age and gender classification from face images. In this section, the related work on age and gender classification using CNN is briefly reviewed. The previous research on local deep neural networks for age and gender estimation is also introduced.

2.1. CNN for Age and Gender Estimation. An early CNN model used for age and gender estimation can be seen in [23], in which a multiscale convolution neural network model is proposed. In [18], the authors propose a convolutional net architecture that can be used even when the amount of learning data is limited. A chained CNN-based age and gender classification scheme is introduced in [24], where the age classifiers are trained for different genders. The apparent age estimation task is investigated in [25]; their proposed model fuses the real value-based regression and the Gaussian label distribution based GoogLeNet; the model was tested on LAP dataset. Later, their result is improved by Antipov et al. [26] by fusing the general model and the children model.

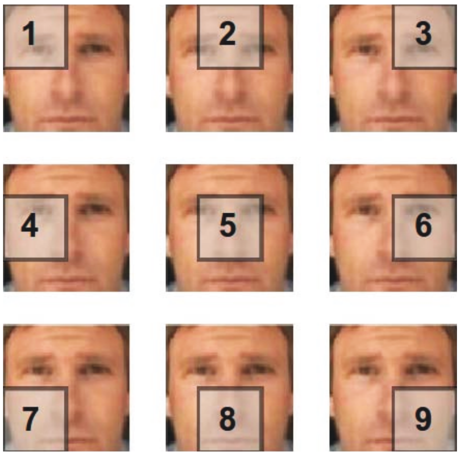


Figure 1: Illustration of the 9 image patches for local neural network training in [21].

Some researchers suggest using deeper networks for age and gender estimation. Yang et al. introduce the deep label distribution learning for apparent age estimation, where the distribution-based loss functions are used for training, which can exploit the uncertainty induced by manual labeling to learn a better model than using ages as the target [27]. The deep age distribution learning (DADL) is proposed in [28] for age prediction. Hou et al. [29] propose a deep CNN model similar with a VGG-16 net coupled with the smooth adaptive activation functions for age estimation. Their results were further improved by using the exact squared earth mover's distance in loss function [30]. In [31], convolutional neural networks are used for the extraction of deep features, then the standard support vector regression is used for gender and age prediction. Recently, the model in [31] is further improved by adding an expected value formulation after classification [32]. The directional age-primitive pattern (DAPP) is proposed in [33], which is a local face descriptor containing aging cue information; the model obtained state-of-the-art performance on Adience dataset.

2.2. Local Deep Neural Networks (LDNNs) for Age and Gender Estimation. Compare with CNN, LDNNs use a different training strategy: the small image patches around important regions of faces are extracted and used for network learning. An LDNN model for gender recognition is proposed in [20], where a feed-forward neural network without dropout is used. An edge detector is firstly used to obtain edges in face images, and small image patches are then selected around the obtained edges. All the image patches are fed into neural networks for training. The predictions of all the patches from the input test image are averaged for the final output. Using patches obtained in this way seldom leads to overfitting since the most redundant information has been removed during filtering.

Another LDNN model was proposed recently, which aims to further reduce the number of image patches used for training [21]. This model uses the same network architecture of [20]. In this modified version of LDNN, only 9 fixed image patches are used for the local network training, as presented in Figure 1. In addition, the authors split an image into five rows (2) and find that the rows containing eye regions and mouth region are important rows for age estimation. By using less training image patches, the model still achieve a competitive performance.

3. Methodology

In this section, we describe the proposed architecture for age and gender classification. Our methodology is essentially composed of three steps: (1) to implement face detection and facial landmark localization, (2) to select image patches based on the obtained facial landmarks, (3) and to construct LDNN model. In the following, the three parts are described in detail.

3.1. Facial Landmark Localization and Patch Selection. The first step of our proposed model is to detect a face in an image and to obtain the facial landmarks on the face, both are widely investigated areas [34–36]. Currently, the global spatial models are polularly used landmark localization methods, which are mainly based on local part detectors. Therefore, it is common to use mixtures of deformable part models or to use mixtures of trees for face detection and landmark estimation. Then the efficient dynamic programming algorithms can be applied to find globally optimal solutions. Without loss of generality, a mixture of trees model for face detection and landmark localization [37] is used here. A brief introduction of the method is given below.

The model is based on a mixture of trees with a shared pool of parts V. Every facial landmark is modeled as a part; the global mixtures are used to capture topological changes due to vi a viewpoint or deformable changes such as changes in expression.

Each tree-structured pictorial structure [38] is linearly parameterized and written as $T_m = (V_m, E_m)$, where m represents a mixture and $V_m \subseteq V$. For an image I, the location of a part i is denoted as $l_i = (x_i, y_i)$. All the image parts $L = \{l_i : i \in V\}$ are scored as

$$S(I, L, m) = \mathrm{App}_m(I, L) + \mathrm{SP}_m(L) + \alpha^m, \quad (1)$$

$$\mathrm{App}_m(I, L) = \sum_{i \in V_m} \omega_i^m \cdot \phi(I, l_i), \quad (2)$$

$$\mathrm{SP}_m(L) = \sum_{ij \in E_m} a_{ij}^m dx^2 + b_{ij}^m dx + c_{ij}^m dy^2 + d_{ij}^m dy. \quad (3)$$

In (2), for the feature vector $\phi(I, l_i)$ extracted from pixel location l_i of image I, the appearance scores of placing the template ω_i^m for part i at the location l_i tuned for mixture m are summed up.

Equation (3) computes the mixture-specific spatial arrangement of parts L. $dx = x_i - x_j$ and $dy = y_i - y_j$ represent the displacement of the ith part relative to the jth part. Each term in the sum can be interpreted as a spring that introduces spatial constraints between a pair of parts [37]. The parameters $(a, b, c, \text{and } d)$ specify the rest location and rigidity of each spring.

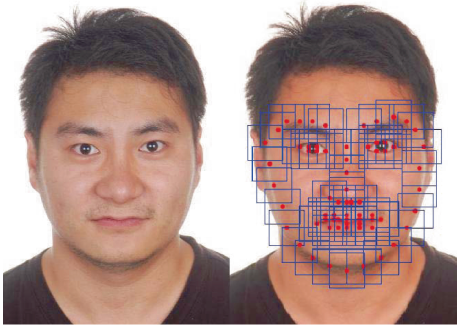


Figure 2: Landmark localization of a sample face image; the red points indicate the center of the detected landmark regions.

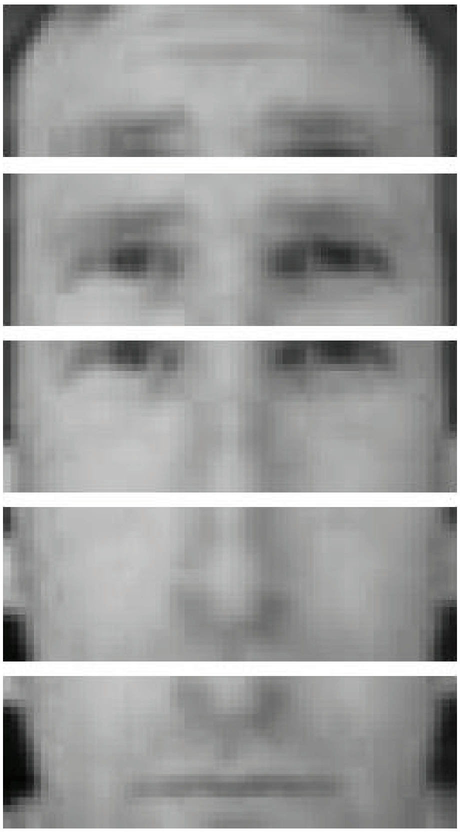


Figure 3: Illustration of the five rows split of a face image in [21].

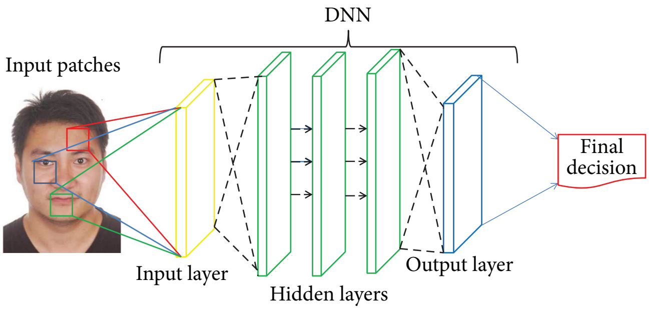


Figure 4: LDNN model used in [20, 21] is also used in this paper.

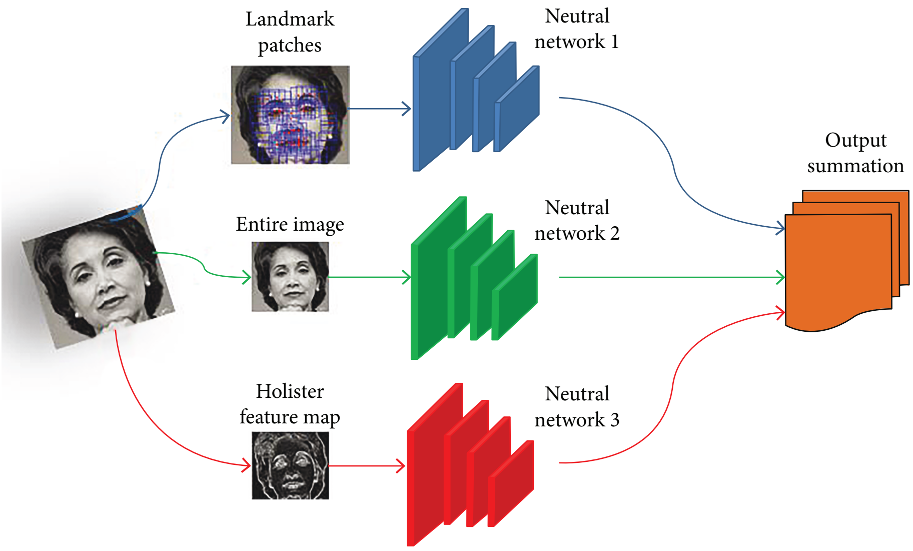


Figure 5: The proposed classification of the main architecture.

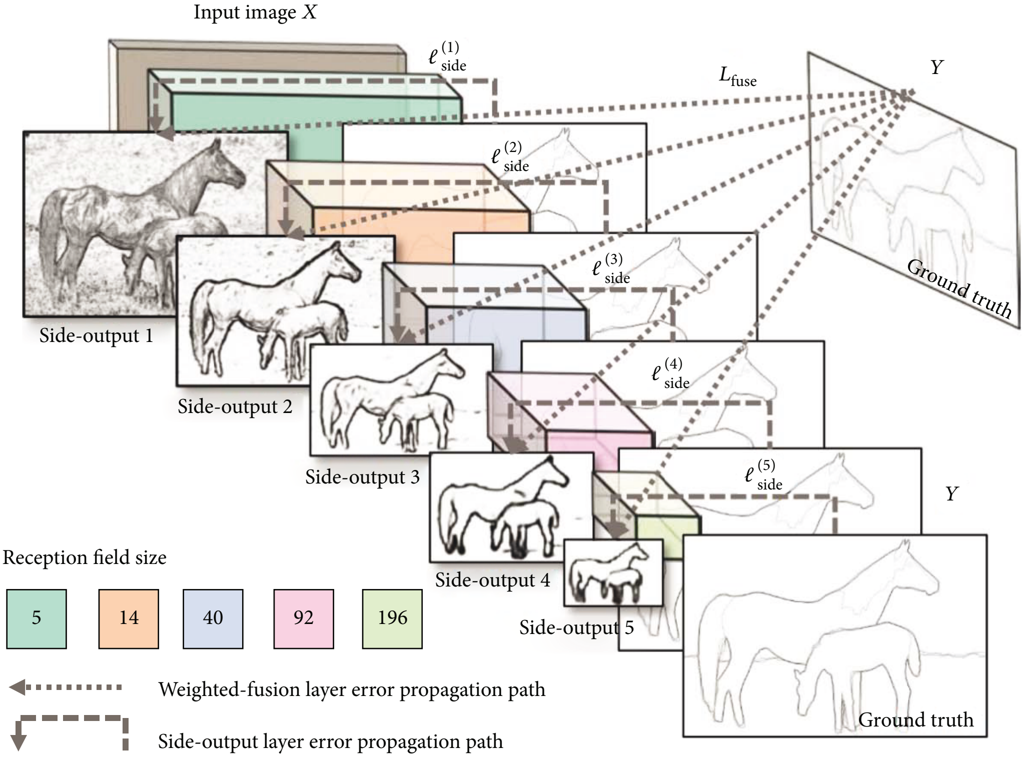

Input image X
$\ell^{(1)}_{\text{side}}$
$\ell^{(2)}_{\text{side}}$
$\ell^{(3)}_{\text{side}}$
$\ell^{(4)}_{\text{side}}$
$\ell^{(5)}_{\text{side}}$
L_{fuse}
Y
Ground truth
Side-output 1
Side-output 2
Side-output 3
Side-output 4
Side-output 5
Y
Ground truth
Reception field size
5
14
40
92
196
Weighted-fusion layer error propagation path
Side-output layer error propagation path

Figure 6: The HED architecture [22].

The model is trained in a fully supervised scheme, where the positive images with landmarks and mixture labels are provided, and the negative images without faces are also provided as well. The shape and appearance models are learned by using a structured predication framework. The Chow-Liu algorithm [39] is used to find the maximum likelihood tree structure which can give the best description of the landmarks in a given mixture. Figure 2 shows the landmark localization result from a sample image.

Once the facial landmarks are obtained, the local image patches can be determined. As shown in Figure 2, for the sample image, a total of 68 landmark points are detected; therefore, 68 image patches around the landmark centers are selected for the network learning. The size of the image patches in Figure 2 are 15 × 15 (blue squares); however, the size of an image patch can vary according to the size of the input images. In our experiments, the performances of different patch sizes are also compared.

Figure 7: Sample images from the LFW database.

The image patch selection used here is different with the two former LDNN-based methods used in [20, 21]. In [20], although the authors only keep the image patches whose center pixel is an active pixel in the binary mask image, there are still hundreds of image patches left for network training. In [21], only 9 fixed patches are used (Figure 3). However, in order to improve model performance, an image is divided into 5 rows, and the rows containing the eye regions and mouth region are used to assist the model output. The patch selection is more empirically decided in this scenario. Our landmark-based patch selection method can keep the most important information in a face image; moreover, it largely reduces the number of training patches.

3.2. Network Architecture. LDNNs are trained by using the image patches extracted from landmark regions of face images. As most of the redundant information has been discarded in our patch selection process, the left training patches cannot lead to the problem of overfitting. Therefore, it is reasonable to use a simple feed-forward neural network. The network architecture used in [20, 21] can also be directly used here for our tasks. Figure 4 shows the network architecture.

The whole procedure of our method is shown in Figure 5. For an input image, its landmark patches are detected and classified by the trained neural network. Then the outputs of the patches are averaged. Following the routine of [21], the entire image can be used to improve classification performance as well. Therefore, another neural network is trained by the entire image is also used here. Moreover, we employ the holistically-nested edge detection (HED) detector [22] to train a third neural network for further performance improvement.

The HED is a deep learning-based edge detection method; it aims to obtain a network that learns features from which it is possible to produce edge maps approaching the ground truth. HED uses multiscale and multilevel structure to generate 5 side-outputs which improve the final fusion result. The architecture of HED can be seen in Figure 6.

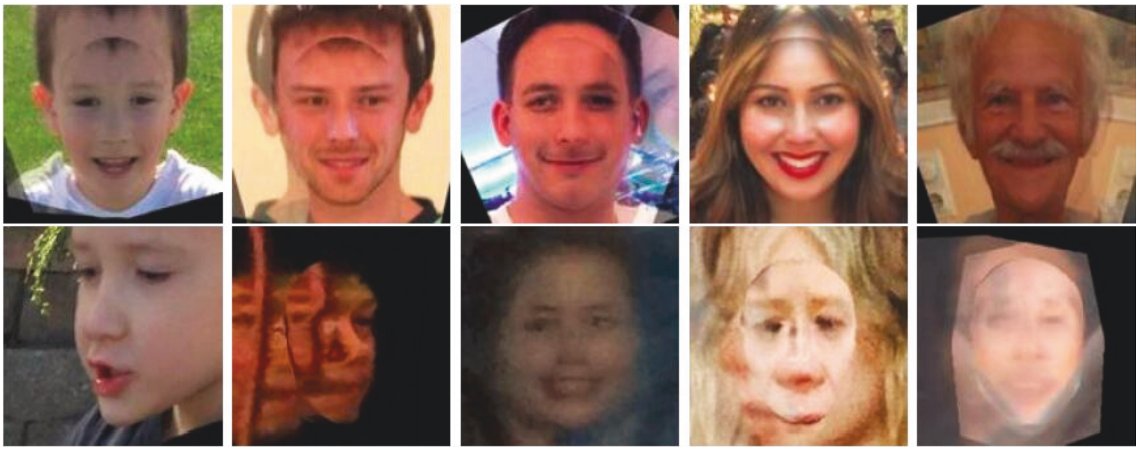


Figure 8: Sample images from Adience database.

Table 1: The label information of the Adience subset used in our experiments.

Age group	1	2	3	4	5	6	7	8	Total
Male	533	693	736	508	1635	1011	333	291	5740
Female	494	910	952	699	1867	875	296	308	6401
Total	1027	1603	1688	1207	3502	1886	629	599	12,141

In Figure 6, the side-output layers are inserted following the convolutional layers. Deep supervision is imposed at each side-output layer to guide the side-outputs toward edge predictions. The outputs of HED are multiscale and multilevel, with the side-output-plane size becoming smaller and the receptive field size is becoming larger. One weighted-fusion layer is added to automatically learn how to combine outputs from multiple scales. The entire network is trained with multiple error propagation paths (dashed lines). The details of HED can be seen in [22].

4. Experiments and Results

A series of experiments has been conducted on two popularly used face image datasets, the LFW database and the Adience database. In this section, the datasets used in our experiments are introduced firstly then the parameter settings of the experiments are introduced. Finally, the experimental results of gender and age estimation are given.

4.1. Face Image Datasets

4.1.1. Labeled Faces in the Wild (LFW). The labeled faces in the wild (LFW) database contains 13,233 face photographs labeled with the name and gender of the person pictured. Images of faces were collected from the web with the only constraint that they were detected by the Viola-Jones face detector [40]. The sample images from LFW database are shown in Figure 7.

There are four versions of LFW—the original version, funneled version, deep funneled version, and frontalized version (3D version). LFW is an imbalanced database including 10,256 images of men and 2977 images of women from 5749 subjects; 1680 of which have two or more images [40]. The 3D version is used in this work since the images are already cropped, aligned, and frontalized properly.

Table 2: The parameter settings in our experiments.

Learning algorithm	SGD + momentum
Dropout probability for input/hidden units	0.75/0.5
Initial learning rate	3
Learning rate update rule	$l_c = l_c * 0.997$ for each epoch
Initial/final momentum	0.5/0.99
Number of hidden units	512
Number of hidden layers	3
Activation function	ReLU

4.1.2. Adience Dataset. There are 26,580 face images from 2284 persons in the Adience dataset [41]. The images are with age and gender labels, which are collected from the Flickr albums and released by their authors under the Creative Commons (CC) license. The images are completely in the wild as the photos were taken under different variations in appearance, noise, pose, and lighting, and so on.

There are three versions of the Adience database, including the original version, aligned version, and frontalized version (3D version) with 26,580, 19,487, and 13,044 images, respectively. The 3D version is used in this work since most images are already frontalized and aligned to the centre of the image. However, images in the Adience database 3D version may be extremely blurry or frontalized incorrectly as shown in Figure 8. Additionally, people in the images

Table 3: The gender classification results on LFW dataset using different hidden layer numbers under the patch size 1313.

Methods compared	LDNN [20]	LDNN + locations [20]	LDNN-F [21]	Proposed	Proposed + locations
Accuracy (1 hidden layer)	91.66	92.64	94.03	94.26	94.32
Accuracy (2 hidden layers)	95.35	95.98	94.22	94.85	94.82
Accuracy (3 hidden layers)	95.81	96.04	95.64	95.53	96.02
Accuracy (4 hidden layers)	95.79	**96.25**	95.29	95.47	95.88

Table 4: Performance evaluation of different sizes of image patches.

Patch sizes	10 × 10	13 × 13	15 × 15	20 × 20	30 × 30
Accuracy (3 hidden layers)	95.78	**96.02**	95.63	95.86	95.54

Table 5: Classification results from different network combinations.

Network combinations	Entire image	Landmark patches	Entire image + landmark patches	Combined
Classification accuracy	92.86	95.06	95.87	**96.02**

could show emotions. Therefore, it is reported that patches extracted from those images may not always contain the same face region which may result in lower classification rates [21].

There are three subsets of the Adience dataset 3D version; this is because it is not necessary to label gender with age groups or vice versa. The first subset contains 12,194 images labeled with gender. The second subset comprises 12,991 images labeled with age. 12,141 images are included in the third subset, which is labeled as both gender and age. Our experiments are run on the third subset. The label information can be seen in Table 1.

4.2. Experimental Settings. In order to find appropriate parameters for the proposed method, a series of experiments has been conducted. The parameters listed in Table 2 produced good outcomes.

The experiments were run on a PC with an Intel i7 4 cores CPU, 16 G memory and an NVIDIA Geforce GTX 1080 GPU (8 G memory); the time cost for training the proposed model is around 10 hours.

4.3. Experimental Results on LFW. For comparison, we follow the routines in [20, 21] to carry out our experiments. Five cross-validations using the same five folds as [20, 21] are used. Around 67% of patches of men are randomly discarded in each fold to balance the data. We first set the size of the image patches as 13 × 13, which is the same as described in [20]. Table 3 lists the classification results of our method and the compared methods, where different numbers of hidden layers are also tested.

In Table 3, one can see that for the same model, besides the image patches themselves, if the center locations' coordinates are added to indicate where a patch is extracted (the "LDNN + location" and "proposed + location" columns), the classification performance can be improved. The method in [21] uses 9 fixed image patches; therefore, the location of patches are also fixed; the method is named as LDNN-F in Table 3.

Table 6: Gender classification results on LFW dataset from different methods.

Methods compared	Accuracy (%)
LDNN	96.25
LDNN-F	95.64
Compact CNN [26]	**97.03**
LBP + SVM [10]	95.6
Gabor + PCA + SVM [42]	94.01
Proposed	96.02

Table 7: Gender classification results on the Adience dataset.

Compared methods	Entire image	LDNN-F [21]	Proposed
Classification accuracy	77.84	78.63	**80.64**

In Table 3, the best gender classification is 96.25% from [20] with patch location. Due to the huge number of patches and the limited amount of memory, it is not feasible to train a neural network using all of the four training folds. In the same way in [21], only one fold was used for network training in this work. The smaller training set is a factor leads to a lower performance.

The effect of using different sizes of image patches are also evaluated. Three hidden layers are used in the network. The compared results can be seen in Table 4. The best performance among the tested patches sizes was obtained by 13 × 13, which is the same with the results in [20], where the authors explain the size of 13 × 13 was determined from a previous research. The best performance of [21] was obtained by using a larger size as 30 × 30. The reason is only 9 location-fixed patches are used for training; a larger patch is able to contain more useful information.

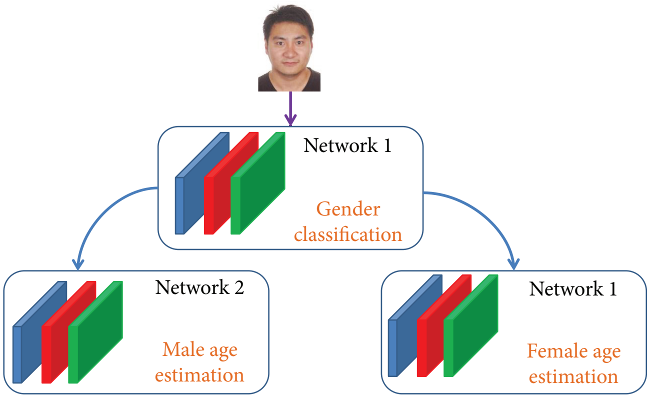


Figure 9: The age estimation scheme.

In our proposed model, besides the landmark-based image patches, the entire image and the holistic feature map extracted from the entire image are also used to further improve model performance. Table 5 lists the results of improvement bought by the holistic feature.

Some of the state-of-the-art methods work on the LFW dataset for gender classification are also compared in our experiments. The results are listed in Table 6. Among the compared methods, the best performance 97.03% was obtained by the method of “Compact CNN”; however, this method needs to construct an ensemble of learning models, which is much more complicated on model construction compared to our method.

4.4. Experimental Results on the Adience Dataset. The age and gender classification are run on the 3D version of the Adience dataset. We used the same routine in [21]; the networks are first trained separately for age and gender then the gender classification results are used to help age estimation.

The same parameters listed in Table 2 are also used here for model training. The performance of our model is shown in Table 7. Our proposed model achieves 80.64% correction rate on the data set, where the result in [21] is 78.63%. The main reason is the Adience dataset are not frontalized well; the location-fixed patches used in [21] may not always contain the same region of faces. In our method, by detecting facial landmarks in advance, the obtained landmark-based patches can relieve this problem much better.

The Aidence dataset contains 8 age groups and another 20 different age labels. Some folds even lack the images for some age groups; therefore, the age labels must be merged. We used the same merging scheme used in [21]; all the labels are merged into the 8 age groups. Please see their paper for details.

In the same way in [18, 21], the one-off classification rate is used for age estimation. That is due to the apparent similarity of persons in adjacent age groups; images which are categorized into adjacent age groups are considered to be correct classification.

For the age estimation, three sets of neural networks are constructed; each contains the model shown in Figure 3. The neural network 1 is used for gender classification, and neural network 2 and 3 are for male and female age estimation, respectively. If an input image is recognized as male then network 2 will be used for its age estimation; otherwise, network 3 will be activated. The whole process can be seen in Figure 9.

Table 8: Age estimation results from the two neural networks for men and women respectively.

Rate method	Neural network 1 (male’s age)		Neural network 2 (female’s age)	
	Exact	One-off	Exact	One-off
Entire image	38.94	77.76	36.68	75.12
LDNN-F [21]	39.90	80.32	41.27	77.14
Proposed	41.86	81.87	42.79	78.65

Table 9: Age estimation results from the proposed age estimation model compared with other CNN-based methods.

Methods	LDNN-F		CNN [41]		Proposed	
	Exact	One-off	Exact	One-off	Exact	One-off
Accuracy	41.82	77.98	45.1 ± 2.6	79.5 ± 1.4	**44.36**	**80.69**

It should be noted that the neural network 2 in Figure 9 is trained using 5740 face images of men, and the network 3 is trained using 6410 images of women. The individual performance of neural network 1 and neural network 2 on age estimation is shown in Table 8, and the results from the model in Figure 9 is given in Table 9.

5. Conclusion and Future Work

A modified version of local deep neural networks is proposed in this paper. Instead of using location-fixed patches, the facial landmarks are detected in advance, then the image patches around landmarks are selected for network training, which greatly reduces the training cost. Moreover, the experimental results show that the method proposed in this paper achieves competitive performance in the two tested datasets. The performance of the proposed model still can be improved by incorporating other schemes into current architecture, for example, to use a more efficient facial landmark detection method or to further optimize the network structure, these will be investigated in our future work.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Additional Points

The ownership of Figures 2, 4, and 9 in this paper belongs to the original author Yungang Zhang. Please do not reprint, duplicate, or use these pictures in any form without the permission from the author. Otherwise, the author will have the right to investigate for legal liability.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The project is funded by Natural Science Foundation China Grant no. 61462097 and Yunnan Provincial Education Department research Grant no. 2018JS143.

References

[1] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski, "Sexnet: a neural network identifies sex from human faces," in *Advances in Neural Information Processing Systems*, vol. 3, pp. 572–577, Denver, CO, USA, 1991.

[2] S. Tamura, H. Kawai, and H. Mitsumoto, "Male/female identification from 8 × 6 very low resolution face images by neural network," *Pattern Recognition*, vol. 29, no. 2, pp. 331–335, 1996.

[3] B. Moghaddam and Ming-Hsuan Yang, "Learning gender with support faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 707–711, 2002.

[4] J. Yang, D. Zhang, A. F. Frangi, and J.-Y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131–137, 2004.

[5] C. Shan, "Learning local binary patterns for gender classification on real-world face images," *Pattern Recognition Letters*, vol. 33, no. 4, pp. 431–437, 2012.

[6] J. G. Wang, J. Li, W. Y. Yau, and E. Sung, "Boosting dense sift descriptors and shape contexts of face images for gender recognition," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pp. 96–102, San Francisco, CA, USA, 2010.

[7] X. M. Leng and Y. D. Wang, "Improving generalization for gender classification," in *2008 15th IEEE International Conference on Image Processing*, pp. 1656–1659, San Diego, CA, USA, 2008.

[8] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pp. 886–893, San Diego, CA, USA, 2005.

[9] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 112–119, Miami, FL, USA, 2009.

[10] J. E. Tapia and C. E. Perez, "Gender classification based on fusion of different spatial scale features selected by mutual information from histogram of LBP, intensity, and shape," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 3, pp. 488–499, 2013.

[11] H. Han, C. Otto, and A. K. Jain, "Age estimation from face images: human vs. machine performance," in *2013 International Conference on Biometrics (ICB)*, pp. 1–8, Madrid, Spain, 2013.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional networks," in *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pp. 1097–1105, Lake Tahoe, NV, USA, 2012.

[13] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: integrated recognition, localization and detection using convolutional networks," 2013, https://arxiv.org/abs/1312.6229.

[14] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, "Fitnets: hints for thin deep nets," 2014, https://arxiv.org/abs/1412.6550.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, https://arxiv.org/abs/1409.1556.

[16] J. Gu, Z. Wang, J. Kuen et al., "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.

[17] S. Chen, C. Zhang, M. Dong, J. Le, and M. Rao, "Using ranking-cnn for age estimation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 742–751, Honolulu, HI, USA, 2017.

[18] G. Levi and T. Hassncer, "Age and gender classification using convolutional neural networks," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 34–42, Boston, MA, USA, 2015.

[19] J. van de Wolfshaar, M. F. Karaaba, and M. A. Wiering, "Deep convolutional neural networks and support vector machines for gender recognition," in *2015 IEEE Symposium Series on Computational Intelligence*, pp. 188–195, Cape Town, South Africa, 2015.

[20] J. Mansanet, A. Albiol, and R. Paredes, "Local deep neural networks for gender recognition," *Pattern Recognition Letters*, vol. 70, pp. 80–86, 2016.

[21] Z. Liao, S. Petridis, and M. Pantic, "Local deep neural networks for age and gender classification," 2017, https://arxiv.org/abs/1703.08497.

[22] S. Xie and Z. Tu, "Holistically-nested edge detection," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1395–1403, Santiago, Chile, 2015.

[23] D. Yi, Z. Lei, and S. Z. Li, "Age estimation by multi-scale convolutional network," in *Asian Conference on Computer Vision*, pp. 144–158, Singapore, 2014.

[24] A. Ekmekji, "Convolutional neural networks for age and gender classification," Technical Report, Stanford University, 2016.

[25] X. Liu, S. Li, M. Kan et al., "Agenet: deeply learned regressor and classifier for robust apparent age estimation," in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pp. 258–266, Santiago, Chile, 2015.

[26] G. Antipov, M. Baccouche, S.-A. Berrani, and J.-L. Dugelay, "Apparent age estimation from face images combining

general and children-specialized deep learning models," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 96–104, Las Vegas, NV, USA, 2016.

[27] X. Yang, B.-B. Gao, C. Xing et al., "Deep label distribution learning for apparent age estimation," in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pp. 344–350, Santiago, Chile, 2015.

[28] Z. Huo, X. Yang, C. Xing et al., "Deep age distribution learning for apparent age estimation," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 17–24, Las Vegas, NV, USA, 2016.

[29] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, and J. H. Saltz, "Neural networks with smooth adaptive activation functions for regression," 2016, https://arxiv.org/abs/1608.06557.

[30] L. Hou, C. P. Yu, and D. Samaras, "Squared earth mover's distance-based loss for training deep neural networks," 2016, https://arxiv.org/abs/1611.05916.

[31] R. Rothe, R. Timofte, and L. Van Gool, "Some like it hot - visual guidance for preference prediction," 2015, https://arxiv.org/abs/1510.07867.

[32] R. Rothe, R. Timofte, and L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," *International Journal of Computer Vision*, vol. 126, no. 2–4, pp. 144–157, 2018.

[33] M. T. B. Iqbal, M. Shoyaib, B. Ryu, M. Abdullah-Al-Wadud, and O. Chae, "Directional age-primitive pattern (DAPP) for human age group recognition and age estimation," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 11, pp. 2505–2517, 2017.

[34] M.-H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.

[35] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[36] Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3476–3483, Portland, OR, USA, 2013.

[37] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2879–2886, Providence, RI, USA, 2012.

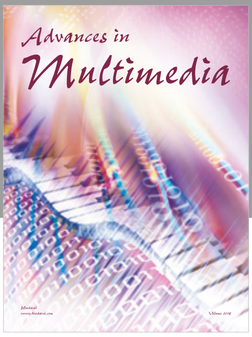
[38] Y. Yang and D. Ramanan, "Articulated pose estimation with flexible mixtures-of-parts," in *CVPR 2011*, pp. 1385–1392, Colorado Springs, CO, USA, 2011.

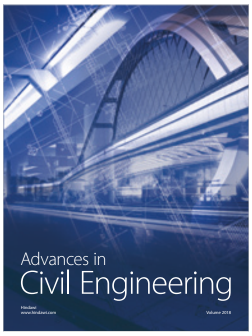
[39] C. Chow and C. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 462–467, 1968.

[40] 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," Technial Report 07-49, University of Massachusetts, Amherst, 2007.

[41] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170–2179, 2014.

[42] P. Dago-Casas, D. Gonzalez-Jimnez, L. L. Yu, and J. L. Alba-Castro, "Single- and cross- database benchmarks for gender classification under unconstrained settings," in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 2152–2159, Barcelona, Spain, 2011.





Hindawi

Submit your manuscripts at
www.hindawi.com