2017 IEEE 12th International Conference on Automatic Face & Gesture Recognition

Metric-Promoted Siamese Network for Gender Classification

Yipeng Huang, Shuying Liu, Jiani Hu, and Weihong Deng
Beijing University of Posts and Telecommunications, Beijing, China

***Abstract*— Gender classification is a fundamental and important application in computer vision, and it has become a research hotspot. Real-world applications require gender classification in unconstrained conditions where traditional methods are not appropriate. This paper proposes a Deep Convolutional Neural Network for feature extraction together with fully-connected layers for metric learning. A Siamese network is built for similarity measuring to promote the performance of classification. Extensive experiments on several databases demonstrate that a significant improvement can be obtained for gender classification tasks in both constrained and unconstrained conditions.**

I. INTRODUCTION

Gender classification plays an important role in many face recognition tasks. Lots of applications based on face recognition depend on a correct gender prediction. For instance, verification for payment, visual surveillance, and human-machine interaction require accurate gender classification. Besides these, other applications such as expression recognition and age classification will perform better if a face's gender can be determined. For above reasons, gender classification has attracted much attention in the past several years.

Traditional gender classification methods mainly consist of two processes, feature extraction and classifier training. There are various kinds of feature extraction methods proposed during past decades, classic features like Local Binary Pattern (LBP) [14], Histogram of Oriented (HOG) [1] and their improvement versions are wildly used in many works. Fisher Vectors (FV) [16] and Locality-constrained Linear Coding (LLC) [27] are more recent features which are also efficient in lots of cases. These features, regarded as encodings of descriptors, are seen as deeper features in contrast with features extracted directly from images. As for classifier, Support Vector Machine (SVM) [25] is the most popular one for its strong classification ability. Conventional methods achieve relatively accurate results in constrained conditions. However, few of them perform well in unconstrained conditions, where real-world photos are shot. Faces with more natural expression and uncertain shooting angle in unconstrained environments where there exists occlusions or blurs bring difficulty to gender classification, thus a more robust method is required to solve this problem.

Recently, with the proposition and development of Convolutional Neural Network (CNN), deep learning has made breakthroughs in the field of computer vision. Successful and famous models, especially Deep Convolutional Neural Network (DCNN), such as AlexNet [6], GoogleNet [22], and VGG [21] have been greatly promoting the development of the whole field. Deep features extracted by DCNN are more abstract and global because of the learning of convolution kernels' weights. Therefore, these features can represent an image more discriminatively. Moreover, compared with fully-connected layers, convolution layers reserve the spacial information of images. Based on the above reasons, CNN is used in this paper for feature extraction. For further improvement, we attempt to solve the problem of gender classification in unconstrained conditions using CNN with the promotion of metric learning. In our work, a CNN architecture is designed and a Siamese network is further built for metric learning. Extensive experiments show that this method significantly outperforms other state-of-the-art methods, proving that metric learning does make contributions to performance improvement.

The remainder of this paper is organized as follows. Section II introduces some related works. The proposed method is detailed in Section III. Experiments and result analysis are presented in Section IV. And finally, a conclusion is given in Section V.

II. RELATED WORK

A. Gender Classification

Many previous works made contributions to gender classification tasks. Discriminative feature extraction is an important factor in most works in the field of conventional computer vision. More recent methods are based on CNN, of which network architecture design is the focus.

Traditional methods consist of two main processes, feature extraction and classifier training, with some other efficient techniques such as image preprocessing and dimension reduction. For instance, Support Faces method applies SVM directly to image intensities to classify gender [11]. Multi-view Gender Classification extracts LBP features from various size of patches, with SVM used as classifier [8]. Multi-scale LBP features are extracted from patches in [19], and Adaboost is applied for feature selection. SVM is also used as the classifier. Fusion method utilizes mutual information measures for feature selection to fuse different kinds of features extracted from various scales [23]. A high accuracy is achieved on FERET database by applying Weber's Local Descriptors for gender classification [24]. Moreover, a survey of gender classification by traditional methods can be found in [9], and several prevalent methods are compared in [12].

With the development of deep learning, many gender classification methods based on DCNN with high classification rate are proposed in recent years. Age and Gender Classification based on CNN puts forward a CNN architecture with three convolutional layers and two fully-connected layers

978-1-5090-4023-0/17 $31.00 © 2017 IEEE
DOI 10.1109/FG.2017.119

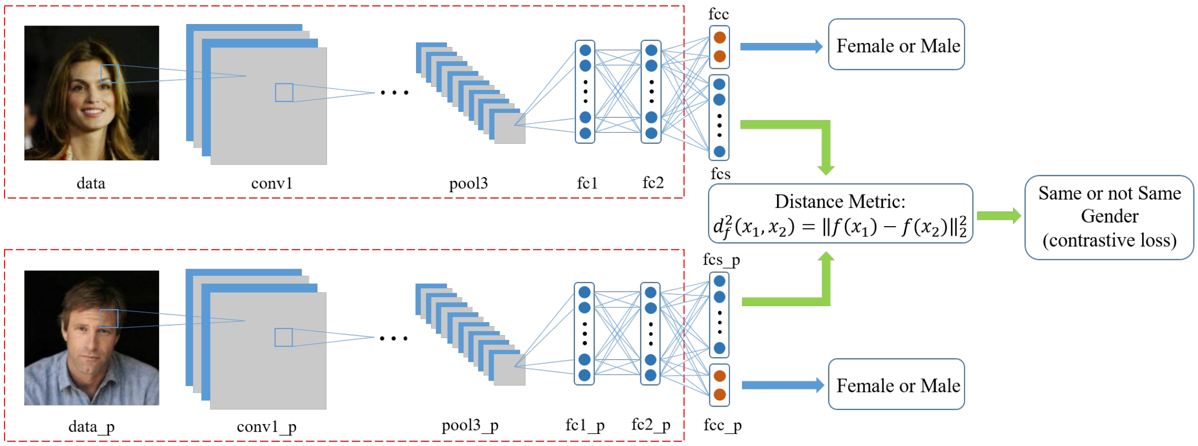

Fig. 1. The proposed network architecture of which the repeating parts are omitted for concision. This multi-task network is end-to-end with convolution layers for feature extraction and fully-connected layers for classification and similarity measuring.

[7], which performs well in real-world gender classification. Another DCNN architecture with the application of L2 Regularization is proposed in [26]. Local Deep Neural Networks (LDNN) method extracts local features from the input images and feeds these features to a network that learns to classify each local feature according to the label of the image to which it belongs [10]. DCNN-SVM method extracts features by fine-tuning network presented in [6] and uses SVM as a classifier, which achieves a state-of-the-art result. These methods based on DCNN perform better in unconstrained conditions for their robustness.

B. Metric Learning

Metric learning algorithms have been applied to the problem of similarity measuring in many cases such as classification, clustering, and matching. A common thought of metric learning is seeking a distance metric to enlarge the distance of negative pairs and reduce the distance of positive pairs. Mahalanobis Metric Learning for Clustering (MMC) minimizes the sum of distances of similar samples and maximizes the sum of distances of dissimilar samples [30]. Schultz et al. learn a distance metric from comparing the distances between similar samples and dissimilar samples [18]. Large-Margin Nearest Neighbors (LMNN) which aims at making one data point share the same label with its nearest neighborhood combines relative distance constraints and the regularizer [28].

The above algorithms only learn a linear transformation which may not be robust enough in unconstrained environments. To solve this limitation, many metric learning methods based on the deep neural network have been proposed in the past few years. Usually, a set of weights of a neural network instead of a Mahalanobis matrix is learned by training, making metric nonlinear due to nonlinear activation functions. Discriminative Deep Metric Learning (DDML) learns a set of hierarchical nonlinear transformations to project face pairs into the same feature subspace [4]. We will introduce more about DDML in Section III. MatchNet, including a convolutional network that extracts features from patches and a network of fully-connected layers that computes the similarity between features, jointly learns a deep network for patch representation as well as feature comparison [3].

III. METHODOLOGY

In this section, we will firstly discuss briefly about metric learning and propose a Siamese network architecture. Then, we will describe some details for network training and testing.

A. Metric-Promoted Siamese Network (MPSN)

In the theory of Mahalanobis distance metric learning, distance between pairs is denoted as form of

$$d_M(x_i, x_j) = \sqrt{(x_i - x_j)^T M (x_i - x_j)} \quad (1)$$

$$= \sqrt{(x_i - x_j)^T W^T W (x_i - x_j)} \quad (2)$$

$$= ||W x_i - W x_j||_2, \quad (3)$$

where M is a semi-definite matrix and W is its decomposition form that applies linear transformation to samples. The aim of this transformation is seeking a mapping method suitable for matching, clustering, or classification.

This transformation may not be powerful enough since it is linear. To improve its performance, DDML proposed a nonlinear mapping method using neural network. It defines the distance between two samples as

$$d_f^2(x_i, x_j) = ||f(x_i) - f(x_j)||_2^2, \quad (4)$$

where W is replaced by $f(\cdot)$ to represent the nonlinear transformation. A stochastic sub-gradient descent scheme is used to optimize parameters.

DDML projects face pairs into one feature space in a deep architecture, thus nonlinear manifold in unconstrained environments could be captured. Metric learning tends to

TABLE I

ONE SIDE OF METRIC-PROMOTED SIAMESE NETWORK ARCHITECTURE

Layer	Type	Size	Output
data	Data	—	$233 \times 233 \times 3$
conv1	Convolution	$7 \times 7 \times 64$, s=2	$114 \times 114 \times 64$
pool1	Pooling(max)	2×2, s=2	$57 \times 57 \times 64$
conv2	Convolution	$3 \times 3 \times 128$, s=2	$28 \times 28 \times 128$
pool2	Pooling(max)	2×2, s=2	$14 \times 14 \times 128$
conv3a	Convolution	$3 \times 3 \times 256$, s=1	$12 \times 12 \times 256$
conv3b	Convolution	$3 \times 3 \times 256$, s=1	$10 \times 10 \times 256$
conv3c	Convolution	$3 \times 3 \times 256$, s=1	$8 \times 8 \times 256$
pool3	Pooling(max)	2×2, s=2	$4 \times 4 \times 256$
fc1	Fully-Connected	—	2048
fc2	Fully-Connected	—	2048
fcc	Fully-Connected	—	2
fcs	Fully-Connected	—	512

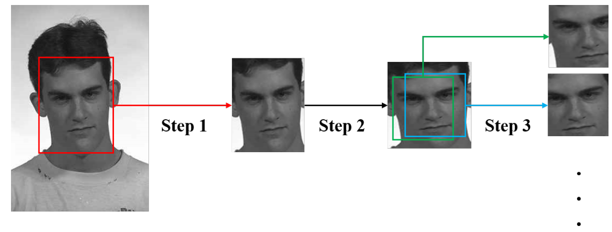

Fig. 2. An illustration for image preparation. Step 1 (for databases whose detected faces are not provided), detect the face with existing detector to wipe out the influence of the background. Step 2, resize the face to a certain size, thus all the detected faces are normalized to the same scale. Step 3, randomly crop patches of a certain size from the face.

enlarge the distance of negative samples and reduce the distance of positive samples, which makes extracted features more discriminative. Moreover, features that are useless to gender classification will be depressed since it makes no sense to similarity measuring. Inspired by DDML, we utilize metric learning as a way to promote the performance of gender classification in our work.

The proposed network architecture is illustrated in Fig. 1. We combine gender classification with metric learning in the network since a multi-task training can achieve mutual promotion. This network comprises five convolution layers with Batch Normalization (BN) and four fully-connected layers with dropout on one side. All of them are followed by Rectified Linear Units (ReLU). A Softmax layer is set after each feature network in order to assign the probable gender of images. The detailed structure of one side is demonstrated in Table I. It is worth mentioning that the layer named fcc is the output layer for classification and the layer named fcs is for computing distance between pairs. These two layers both follow fc2.

At the end of fully-connected layers, losses are defined and added with weights. The loss of cross entropy for classification can be defined as

$$L_c = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij}\cdot\log\hat{y_{ij}} + (1-y_{ij})\cdot\log(1-\hat{y_{ij}}), \tag{5}$$

where y is the groud truth label, $\hat{y}$ is the predicting probability of class i, C is the number of classes, and N is the batch size. We also use a contrastive loss layer to measure the similarity of two outputs. The mathematic form of contrastive loss is

$$L_s = \frac{1}{N}\sum_{i=1}^{N} {y_i}'\cdot d + (1-{y_i}')\cdot max(margin-d,0), \tag{6}$$

which aims at keeping the distance of different classes larger than $margin$, where d is the distance, usually Euclidean distance, between two features, and y' is the label of whether they belong to the same class. This contrastive loss is utilized as the loss of metric learning. We can notice that L_s will be small when the distance between similar samples is short or distance between dissimilar samples is long from (6). Therefore, through training, images are mapped to a feature space that is more suitable for gender classification with the promotion of metric learning. In other words, a similarity metric is learned for this task.

The final loss function adds the cross entropy loss and the contrastive loss, which is defined as

$$L = L_c + \lambda \cdot L_s, \tag{7}$$

where λ is a parameter for balancing.

B. Training and Testing Details

1) Data Preparation: Faces are detected and cropped out of images at first (if detected faces are not provided). Then these faces are resized to the resolution of 250×250. We use 233×233 patches randomly cropped from face images and their mirrors as the input of the network to avoid overfitting while training our network, and 233×233 patches cropped around face center is used to do the testing. This process is illustrated in Fig. 2. For metric learning, samples need to be prepared as pairs with a label of whether they belong to the same class.

2) Testing: After training a network model, testing should be applied to evaluate the validity of this network. Since metric learning for similarity measuring is a way to assist classification, one side of this Siamese network is enough for gender classification testing.

IV. EXPERIMENTS

The experiments are performed utilizing several gender classification methods including conventional computer vision and deep network on three databases, Labeled Faces in the Wild (LFW) [5], Face Recognition Technology (FERET) [17], and Adience benchmark [2]. In this section, we will introduce our training and testing sets, methods that are utilized, and the results of these experiments.

A. Databases

1) LFW: The LFW database is a set of celebrity images which are shot in unconstrained environments. Faces with uncertain expressions, uncertain shooting angles, and different illumination conditions are challenging to be recognized. This database consists of 13,233 images of 5,749 people, with 2,977 images of 1,483 females and 10,256 images of 4,266 males among them. We randomly select 2,000 female images and 2,000 male images as the training set to keep the balance of two classes and use the rest 9,233 images as the testing set.

2) FERET: The FERET database has 14,051 images in total, comprising 5,200 images of 499 females and 8,851 images of 705 males. Although images in FERET database are shot in constrained environments, fusing all the images including profile faces with different illumination to classify makes gender classification more difficult. Since the amount of samples in two classes are relatively balanced, we randomly select 6,159 images for training and 2,195 images for testing without considering the ratio of two classes in training set.

3) Adience: The Adience benchmark was constructed for age and gender classification in recent years. It is more challenging since its images are collected from users on social websites, where there exist more blurs and occlusions. Plenty of baby faces bring difficulty to gender classification as well. Adience benchmark includes more than 26 thousand images of 2,284 subjects. The images are divided into five folds for cross validation by the author.

Because images in the LFW database are gathered from news articles on the web and images in the FERET database are specifically shot in constrained environments, more and more researchers choose the Adience benchmark that is more challenging as the subject in their study. Also, the split of training and testing sets for classification are not given on the LFW and FERET databases, and the gender labels are even not complete, so the comparison between different researches makes no sense. For these reasons, we will compare different methods implemented by ourselves on the LFW and FERET databases to demonstrate the effectiveness of metric learning, and the proposed method will also be compared with other researches on the Adience benchmark to show the novel result of this method.

Images are preprocessed as described in Section III. LFW images aligned with commercial face alignment software and Adience of aligned version are used instead of the source images for simplicity. As for FERET images, we utilize FACE++ to detect face areas, and faces which are not detected are omitted.

B. Methods to be Compared

We compare this proposed method for gender classification with some other methods to explain the validity of our method.

1) LBP: LBP is a kind of texture feature. Face images are divided into tens of sub-regions. For each pixel in each region, the pixel is encoded by comparing its value with its neighbors' values, thus a patch can be described by a histogram. After computing and counting, an image can be represented by a feature made up of histograms of all the patches, which is called LBP feature. Histograms extracted from patches of different sizes, different sampling radius of circular neighborhoods, and different overlapping rates are fused to capture multi-scale features.

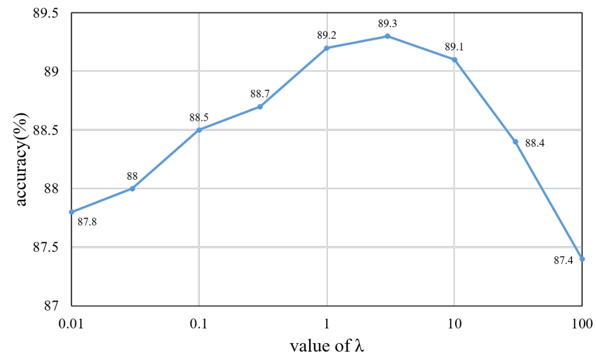

Fig. 3. The λ-accuracy curve. The horizontal axis in logarithm scale represents the value of λ and the vertical axis in linear scale represents the classification accuracy.

2) FV: FV is also a kind of feature. It can be seen as an approach to feature representation or feature encoding. Fisher vector can be encoded from SIFT descriptors by using a Gaussian Mixture Model as described in [20]. The purposes of this encoding are summarizing vectorial statistic local feature descriptors, similar to bag of visual words, and these representations also store the difference between dictionary and local features. Besides, we also fuse FV with LBP feature to generate a more representative and discriminative feature.

3) DDML: As is described in [4] and Section III, DDML is a metric learning method. In the experiments, we regard the trained network as a method of mapping features and reducing dimensions. LBP feature is utilized as the input of this network, and the output of it is seen as a representation of the feature, which is expected low-dimensional and more valid.

SVM classifier is applied to features above for classification, and Whitening Principle Component Analysis (WPCA) is used as the dimensional reduction method for high-dimensional features. In the experiment, we will train a DCNN whose architecture is the same with one side of the Siamese network proposed above to compare with it.

C. Determine the Value of λ

As is in (7), an appropriate λ value needs to be determined to ensure that both the two losses described in Section III converge. We train models with several λ values ranging from 0.01 to 100 to explore the relationship between λ and classification accuracy. Too large or too small λ values are not appropriate for training, since large values make classification ineffective and small values make metric learning nonsense. We choose images in the first fold of the Adience benchmark as testing samples and the rest images as training samples to conduct this experiment. All the parameters except λ are kept the same in different models.

TABLE II
AVERAGE ACCURACY ON THE LFW AND FERET DATABASES

Method	LFW(%)	FERET(%)
LBP	93.4	95.4
FV	93.8	95.9
LBP-FV-FUSION	94.3	96.2
LBP-DDML	94.2	95.9
DCNN	95.7	96.9
MPSN	96.4	97.8

The λ-accuracy curve is illustrated in Fig. 3. As we can observe, the relatively satisfying result can be obtained with the value λ being set to 3. Therefore, we set the value of λ to 3 for all of the following experiments.

D. Results

The results of the several experiments on LFW and FERET databases are presented in Table II. Among the methods in the table, LBP represents LBP feature with SVM classifier. FV represents Fisher vector encoded from dense SIFT and FUSION means the fusion of LBP feature and Fisher vector. LBP-DDML and FV-DDML are these two features transformed by DDML, classified by SVM classifier as well. DCNN represents classification by CNN whose architecture is one side of proposed Siamese network in this paper. MPSN is the proposed method.

Best accuracy with 96.4% on the LFW database and 97.8% on the FERET database is obtained by the proposed metric-promoted method. The results show that fusing features is an efficient approach to performance raise. Since feature selection is a process that selects the most discriminative dimensions, performance would be better, provided appropriate number of dimensions being set. DCNN is proved to be more effective compared with the conventional methods for its deeper representation of features. By utilizing DDML, accuracy on the LFW database is raised from 93.4% to 94.2%, and from 95.4% to 95.9% on the FERET database, which means that metric learning indeed makes contributions to a better performance. This can also be concluded by the comparison between DCNN and MPSN, where the accuracy is raised from 95.7% to 96.4% on the LFW database and from 96.9% to 97.8% on the FERET database.

TABLE III
AVERAGE ACCURACY (±STANDARD ERRORS) ON THE ADIENCE BENCHMARK

Method	Accuracy(%)
LBP [2]	75.3±0.9
FPLBP [2]	75.5±0.8
LBP+FPLBP+Dropout-0.5 [2]	77.8±1.3
LPQ+SIFT [13]	80.5±2.6
CoLBP [15]	83.9±1.9
CNN+single crop [7]	85.9±1.4
CNN+oversampling [7]	86.8±1.4
CNN+Fine-tuning [29]	86.2±0.7
CNN+Fine-tuning+oversampling [29]	87.2±0.7
MPSN(ours)	88.3±0.6

(a) Correctly classified examples. In the top row are some of female images and in the bottom row are some of male images.

(b) Misclassification examples. In the top row are some of misclassified female images and in the bottom row are some of misclassified male images.

Fig. 4. A few testing results on LFW database. In (a) are some correctly classified images and in (b) are some misclassified images.

Results on the Adience benchmark compared with results of other researches are depicted in Table III. We obtain an average accuracy of 88.3% on this database. Comparing with 77.8% by LBP+FPLBP+Dropout-0.5 in [2], 80.5% by LPQ+SIFT in [13], 83.9% by CoLBP in [15], 86.8% by CNN+oversampling in [7], and 87.2% by CNN+Fine-tuning+oversampling in [29], this result reaches the state-of-the-art performance. Some results of other methods for comparing in these papers are also listed in Table III. With deep features extracted by CNN and the promotion of metric learning, which aims at promoting discriminative feature extraction, our method performs well in both constrained and unconstrained conditions.

We present a few correctly classified images chosen from the LFW database in Fig. 4(a). These correct results implicate that the proposed method is robust enough in the condition of natural facial expressions, variable shooting angles, and real-world illumination. Also, several misclassifications are presented in Fig. 4(b). We could notice that images which are incorrectly classified are challenging enough for gender classification, especially when there exist occlusions or blurs in images. Moreover, classification of childrens gender still remains more difficult for their similar appearance.

V. CONCLUSION

In order to solve the problem of gender classification in unconstrained environments, we propose a metric-promoted Siamese network for feature extraction and similarity measuring. The main idea of this method is to promote the performance of classification by metric learning. This method achieves state-of-the-art performance on three standard databases, and we compare our results with other works on the most challenging one, the Adience benchmark, implicating that this method performs well in both constrained and unconstrained conditions. Deep feature and metric learning are proved to be conducive to performance improvement through comparative experiments.

VI. Acknowledgemets

This work was partially supported by the National Natural Science Foundation of China (Project 61471048, 61375031, and 61573068), Beijing Nova Program under Grant No. Z161100004916088, the Fundamental Research Funds for the Central Universities under Grant No. 2014ZD03-01, the Program for New Century Excellent Talents in University.

References

[1] 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)*, volume 1, pages 886–893. IEEE, 2005.

[2] E. Eidinger, R. Enbar, and T. Hassner. Age and gender estimation of unfiltered faces. *IEEE Transactions on Information Forensics and Security*, 9(12):2170–2179, 2014.

[3] X. Han, T. Leung, Y. Jia, R. Sukthankar, and A. C. Berg. Matchnet: Unifying feature and metric learning for patch-based matching. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3279–3286, June 2015.

[4] J. Hu, J. Lu, and Y. P. Tan. Discriminative deep metric learning for face verification in the wild. In *Computer Vision and Pattern Recognition*, pages 1875–1882, 2014.

[5] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. *Month*, 2008.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25(2):2012, 2012.

[7] G. Levi and T. Hassncer. Age and gender classification using convolutional neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 34–42, 2015.

[8] H. C. Lian and B. L. Lu. Multi-view gender classification using local binary patterns and support vector machines. *International Journal of Neural Systems*, 17(6):479–87, 2011.

[9] E. Makinen and R. Raisamo. Evaluation of gender classification methods with automatically detected and aligned faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):541–547, March 2008.

[10] J. Mansanet, A. Albiol, and R. Paredes. Local deep neural networks for gender recognition. *Pattern Recognition Letters*, 70(C):80–86, 2016.

[11] B. Moghaddam and M. H. Yang. Learning gender with support faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):707–711, 2002.

[12] E. Mkinen and R. Raisamo. An experimental comparison of gender classification methods. *Pattern Recognition Letters*, 29(10):1544–1556, 2008.

[13] H.-T. Nguyen and T. T. N. Huong. Unconstrained gender classification by multi-resolution lpq and sift. In *Information and Computer Science (NICS), 2016 3rd National Foundation for Science and Technology Development Conference on*, pages 212–217. IEEE, 2016.

[14] T. Ojala, M. Pietikainen, and D. Harwood. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, volume 1, pages 582–585. IEEE, 1994.

[15] B. Patel, R. Maheshwari, and B. Raman. Compass local binary patterns for gender recognition of facial photographs and sketches. *Neurocomputing*, 218:203–215, 2016.

[16] F. Perronnin, J. Sánchez, and T. Mensink. Improving the fisher kernel for large-scale image classification. In *European conference on computer vision*, pages 143–156. Springer, 2010.

[17] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

[18] M. Schultz and T. Joachims. Learning a distance metric from relative comparisons. In *International Conference on Neural Information Processing Systems*, pages 41–48, 2003.

[19] C. Shan. Learning local binary patterns for gender classification on real-world face images. *Pattern Recognition Letters*, 33(4):431–437, 2012.

[20] K. Simonyan, O. Parkhi, A. Vedaldi, and A. Zisserman. Fisher vector faces in the wild. In *British Machine Vision Conference*, pages 8.1–8.11, 2013.

[21] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *Computer Science*, 2014.

[22] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, June 2015.

[23] J. E. Tapia and C. A. 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*, 8(3):488–499, 2013.

[24] I. Ullah, M. Hussain, G. Muhammad, H. Aboalsamh, G. Bebis, and A. M. Mirza. Gender recognition from face images with local wld descriptor. In *2012 19th International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 417–420, April 2012.

[25] V. N. Vapnik. The nature of statistical learning theory. *Neural Networks IEEE Transactions on*, 8(6):988 – 999, 1995.

[26] A. Verma and L. Vig. Using convolutional neural networks to discover cogntively validated features for gender classification. In *2014 International Conference on Soft Computing and Machine Intelligence (ISCMI)*, pages 33–37, 2014.

[27] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3360–3367. IEEE, 2010.

[28] K. Q. Weinberger and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 10(1):207–244, 2009.

[29] J. V. D. Wolfshaar, M. F. Karaaba, and M. A. Wiering. Deep convolutional neural networks and support vector machines for gender recognition. In *Computational Intelligence, 2015 IEEE Symposium*, 2015.

[30] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell. Distance metric learning, with application to clustering with side-information. *Advances in Neural Information Processing Systems*, 15:505–512, 2003.