2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)

Face Verification Based on Feature Transfer via PCA-SVM Framework

Xiaofen Xing, Guicong Xu, Bolun Cai, Chunmei Qing*, and Xiangmin Xu
School of Electronic and Information Engineering
South China University of Technology, Guangzhou, China
Email: xfxing@scut.edu.cn, guicongxu@163.com, caibolun@gmail.com, {qchm,xmxu}@scut.edu.cn

***Abstract*—Face verification is a task to determine whether a pair of given facial images belong to the same person. In unconstrained real applications, inter and intra variations, including illumination, pose, occlusion, and expression, will seriously decrease the verification performance. Due to the lack of annotated data for face verification, extended datasets for face recognition with large samples are used to assist learning a robust feature representation generally. However, the extended data for face recognition is different from face verification on distribution and task. In this paper, a transfer learning based on PCA-SVM is proposed to alleviate above problem. The original feature representation is learnt from a deep convolutional neural network by face classification. Then a PCA-SVM based transfer method is used for feature reprojection from the source domain (face recognition) to the target domain (face verification), which reduces the divergence of feature distribution and task inconsistency. The proposed framework yields comparable results and the accuracy is 98.5% on LFW dataset.**

1. Introduction

Face verification is an important research area in computer vision and can be widely applied for facial automatic identity. The robust feature representation is crucial for face verification, and many hand-crafted features (e.g. Local Binary Pattern (LBP) [1], Haar-like feature [2], and Gabor wavelets [3]) can perform well in constrained applications. In real unconstrained environment, inter-class and intra-class variations will degrade the verification performance. Therefore, many researches are focused on learning invariant and discriminative facial feature representations during the past decade.

Convolutional neural networks (CNNs) have demonstrated state-of-the-art performance in many computer vision applications, such as image classification [4], object detection [5], salient detection [6], and image enhancement [7]. An evaluation [8] about automatic face recognition in unconstrained environments showed that the best performing system cannot rely on hand-crafted features only. Taigman et al. [9] employed a 3D face modeling for alignment and derived a face representation based on a nine-layer deep neural network. The robust feature representation learnt from four million labeled data belong to more than 4,000 identities, which achieves the accuracy closing to human-level on LFW dataset [10]. In [11], the training sample selecting mechanism and the network transferability is studied to improve feature representation. From above, the trend of using CNNs to learn robust feature representation from large-scale extended dataset is obvious.

Deep learning provides more powerful tools to learn invariant and discriminative feature representation, but large training data is needed to cover various variations. Due to the requirement of plenty diversity samples for network training, CNN based face verification methods employ additional training data and are trained as classification tasks generally. Therefore, these models may suffer the performance degradation because of the divergence of feature distribution and training task between the source domain (face recognition) and the target domain (face verification). In [12], a deep network called DeepID with four convolutional layers was proposed to extract features hierarchically. The generalization capability of DeepID is to increase as more face classes for recognition training. For face verification, joint Bayesian technique [13] was used to deal with the divergence of feature distribution. However, the training of Bayesian model also needs large annotated data to keep intra-personal invariant. To reduce intra-personal variations while large inter-personal differences, Sun et al. [14] introduced face recognition and verification as double supervision signals based on multi-task learning. Then more discriminative feature representation improvements (DeepID2+ [15], DeepID3 [16]) were made based on [14]. The hyper-parameters to balance the supervision signals between recognition and verification are hard to be selected in multi-task learning.

In this paper, a transfer learning framework is proposed for feature reprojection from face recognition to face verification. Firstly, Principal Component Analysis (PCA) [17] is used to minimize the feature distribution deviations between different datasets by mapping the high-dimensional features to a low-dimensional subspace. To transfer the face recognition task to the face verification task, Support Vector Machine (SVM) [18] is adopted to determine whether a pair of given facial images belong to the same person. The main contributions of this paper are:

- We propose a feature transfer framework based on

*. Chunmei Qing is the corresponding author.

978-1-5386-3066-2/17 $31.00 © 2017 IEEE
DOI 10.1109/iThings-GreenCom-CPSCom-SmartData.2017.166

PCA-SVM to deal with the divergence of feature distribution and task inconsistency between the source domain (face recognition) and the target domain (face verification).
- The feature representation learnt from the CASIA-WebFace dataset [19] achieves a comparable performance on LFW benchmark [10] with a single network.

The remainder of this paper is organized as follows. Section 2 describes a face recognition method based on Lightened CNN. Details of PCA-SVM framework proposed by us are given in Section 3 and the experiments are shown in Section 4. Finally, Section 5 makes a conclusion of our work.

2. Lightened CNN for Face Recognition

To reduce the waste of calculations and parameters, a lightened CNN architecture was proposed in [20], which can learn a compact embedding for face representation. The architecture of Lightened CNN is illustrated in Fig. 1. There are two kinds of deep models developed in [20]: the ModelA contains 4 convolutional layers, Max-Feature-Map (MFM) activation functions, 4 max-pooling layers, and a fully connected layers; differently to ModelA, the ModelB contains 5 convolutional layers and 4 Network in Network (NiN) [21] layers, which reduce the kernel size and accelerate calculation. In both ModelA and ModelB, the MFM activation function is defined as the maximum between two convolutional feature maps, and used to extract invariant and discriminative features. The representation feature vector learnt through lightened CNN is written as

$$f^{cnn} = F_{cnn}\left(I, \Theta\right), \tag{1}$$

where F_{cnn} is the feature extraction function defined by Lightened CNN, Θ is the model parameters, and I is an aligned facial image.

Lightened CNN is trained on the CASIA-WebFace dataset. A fully connected layers with softmax function is used to achieve the probability distribution p over the facial identities:

$$p = \sigma\left(W f^{cnn} + B\right). \tag{2}$$

$\sigma\left(\cdot\right)$ denotes the softmax function, and W, B are weight and bias parameters respectively. The network is trained to minimize the likelihood distance E on the face recognition task.

$$E\left(\Theta\right) = \frac{1}{MN}\sum_{n=1}^{N}\sum_{m=1}^{M} y_{n,m}\log\left(p_{n,m}\right), \tag{3}$$

where $y_{n,m}$ is the training sample label, and $p_{n,m}$ is the predicted probability. Here M is the number of facial identities and N is the total number of training samples in CASIA-WebFace dataset.

3. PCA-SVM Based Feature Transfer

Due to the data distribution and task divergence between the source domain and the target domain, the model trained on the face recognition task lacks a powerful generalization ability for face verification. In this section, a PCA-SVM based transfer learning framework from recognition to verification is proposed, and the framework is illustrated in Fig. 2. We adopt PCA to minimize the distribution deviations between the training dataset (CASIA-WebFace) and the test dataset (LFW) by reprojecting the high-dimensional features to a low-dimensional subspace. Then, to transfer the source task (face recognition) to the target task (face verification), SVM is used to determine whether the given facial images belong to the same person.

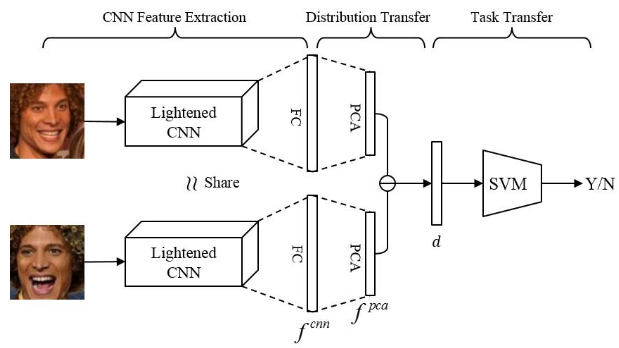

Figure 2. PCA-SVM framework for feature transfer.

3.1. Distribution Transfer Based on PCA

Known as domain transfer, the effort to bridge the gap between training and testing data distributions has been discussed under the context of deep learning [22], [23]. All of these methods face the same challenge of constructing the domain transfer function – a high-dimensional non-linear function. To narrow the distribution gap, PCA is adopted to reproject the original features to the principal subspace in a linear way, which maximizes the variance and minimizing the error. Therefore, we can easily transfer the trained model to a new domain by modulating the statistics by PCA. PCA based distribution transfer is straightforward to implement, has zero parameter to tune, and requires minimal computational resources. The feature after PCA projecting not only reduces the redundancy and noise, but also decreases the computational complexity. The feature transfer process is shown in Fig. 3.

To reduce the dimension of the original feature space, PCA is used to find the projection direction which is the most effective representation of the original data by the least mean square error. $\mathbf{F}^{cnn}$ is the feature matrix as $\mathbf{F}^{cnn} = \{f_1^{cnn}, f_2^{cnn}, ..., f_N^{cnn}\}$, where N is the sample number on the target dataset. The covariance matrix $\mathbf{C}$ is defined as

$$\mathbf{C} = \left(\mathbf{F}^{cnn} - \bar{f}^{cnn}\right)\left(\mathbf{F}^{cnn} - \bar{f}^{cnn}\right)^T, \tag{4}$$

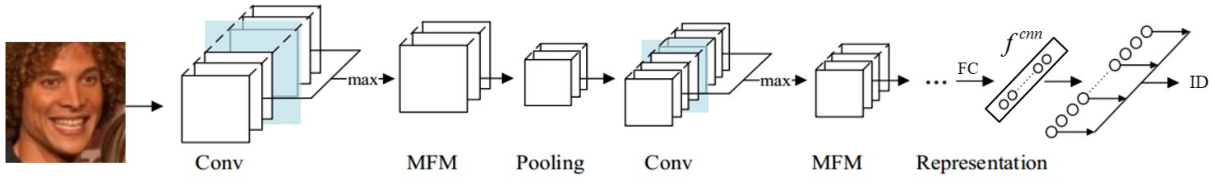

Figure 1. An illustration of the architecture of lightened CNN for face recognition.

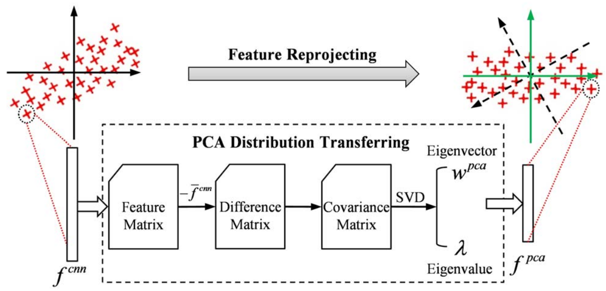

Figure 3. Distribution transfer based on PCA.

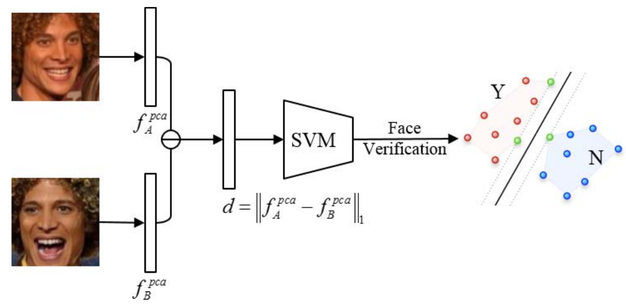

Figure 4. Task transfer based on SVM.

where $\bar{f}^{cnn}{=}1/N\sum_{n=1}^{N} f_n^{cnn}$ is the mean value of feature vectors. The covariance matrix $\mathbf{C}$ uses mean-value bias to perform data normalization for the training domain and the test domain. The projection weight w^{pca} can be obtained by Singular Value Decomposition (SVD) on the covariance matrix: $\mathbf{C} = \mathbf{W}\mathbf{\Sigma}\mathbf{W}^{-1}$. Here $\mathbf{W} = \{w^{pca}\}$ is a orthogonal eigenvector matrix, and $\mathbf{\Sigma}$ is a diagonal matrix with eigenvalues λ. Sorting the eigenvalues in descending order to select top-D principal components, the transferred feature f^{pca} is described as

$$f^{pca} = \{w_i^{pca}\}_{i=1}^{D} f^{cnn}, \tag{5}$$

where D is the dimension of f^{pca}.

3.2. Task Transfer Based on SVM

Face recognition is a multiple classification problem, which focuses on selecting the best matching from the candidate templates and determining the class identity of the image. While face verification is a one-to-one matching problem, which focuses on whether a pair of input faces coming from the same individual. SVM is adopted in our framework for task transfer (see Fig. 4). In addition, Radial Basis Function (RBF) kernel is used as a nonlinear transformation to extract more robust feature space. Therefore, SVM can achieve the global optimization in the process of task transfer and improve the accuracy of face verification.

The faces of the same individual may be very different suffering from the variance pose, illumination, expression, and occlusion. Therefore, it is crucial to reduce the intra-class variations while enlarging the inter-class differences for face verification. For the target domain, we use pairs to train the verification model, the L1 distance between a pair of facial images is described as

$$d = \left\| f_A^{pca} - f_B^{pca} \right\|_1, \tag{6}$$

where f_A^{pca} and f_B^{pca} are two random samples of the verification task. A hyperplane is trained to distinguish the label of an input pair $y \in \{+1, -1\}$, where $+1$ represents the input pair belonging to the same person, and -1 indicates the pair from the different persons. The kernel-based classification function is

$$f^{svm}(d) = w^{svm}\Phi(d) + b, \tag{7}$$

where $\Phi(\cdot)$ is the nonlinear mapping of feature distance d, and w^{svm} and b are the weights and bias respectively.

The parameters w^{svm} and b be solved by minimizing the criterion function:

$$J(w^{svm}, b) = \frac{1}{2}\|w^{svm}\|^2 + C\frac{1}{N}\sum_{n=1}^{N} L(y_n, f^{svm}(d_n)), \tag{8}$$

where L is the hinge-loss function and C is the penalty factor. By introducing the method of Lagrange equation to solve the above problem, we can obtain the following discriminant function:

$$f^{svm}(d) = \sum_{d_i \in \mathbf{V}} (\alpha_i - \alpha_i^*) K(d_i, d) + b, \tag{9}$$

where $\mathbf{V}$ is the support vector set, α is the Lagrange multiplier, and $K(d_i, d_j) = \Phi(d_i)\Phi(d_j)$ is the kernel function.

4. Experiments

In the experiment, we use CASIA-WebFace [19] dataset as the source domain and LFW [10] dataset as the target domain. Firstly, the CASIA-WebFace dataset is used to train the lightened CNN [20]. The facial samples are corrected to gray-scale image with 20 pixel pupil-distance and then cropped into 144×144 randomly for training. The *Caffe* [24] package is implemented to train the lightened CNN with Stochastic Gradient Descent (SGD) method in the source domain. We test the proposed framework on the target domain (LFW dataset), which is collected in unconstrained conditions including 13,233 images of 5,749 persons for face verification. The accuracy and Receiver Operating Characteristic (ROC) curve are adopted to evaluate the performance of our transfer framework.

4.1. Analysis of Distribution Transfer

To evaluate the effectiveness of distribution transfer based on PCA on the LFW database, we directly calculate the Euclidean distance of a pair and use a linear classification plane to whether a pair of given facial images belong to the same person.

The accuracy of different PCA dimensions is shown in Table 1. We can see that the highest accuracy can be obtained when the dimension is set to 70 for ModelA and 100 for ModelB. In addition, the verification accuracy can also be improved by feature reprojection without dimensionality reduction. The ROC curves and the average accuracy on LFW dataset are shown in Figure 5, which demonstrates that the PCA based distribution transfer achieves the better overall performance compared to original ModelA and ModelB. From the figure, we can see that ModelA-PCA outperforms original ModelA by 0.23% in mean accuracy, and ModelB-PCA is 98.40% compared to ModelB 98.13% – a difference of 0.27%.

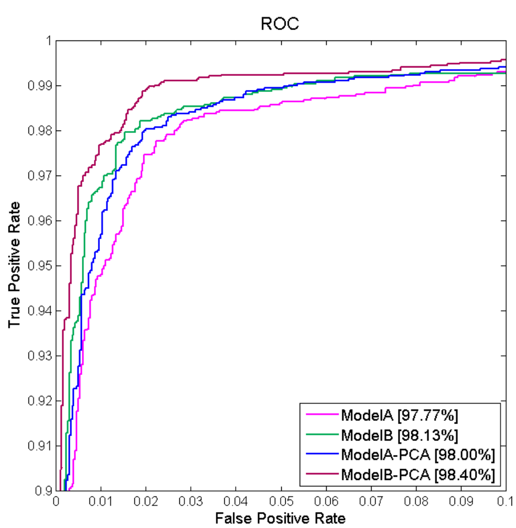

Figure 5. The ROC curves compared on distribution transfer.

TABLE 2. Accuracy Comparisons on task transfer

Methods	Average Accuracy (%)
ModelA [20]	97.77
ModelB [20]	98.13
ModelA-SVM (ours)	**97.79**
ModelB-SVM (ours)	**98.18**

4.2. Analysis of Task Transfer

We use the verification accuracy to evaluate the effectiveness and feasibility of SVM task transfer. In LFW dataset, 90% of samples are randomly selected to train SVM with RBF kernel, and the rest is used to test. The average accuracy of the ten-fold cross validation is used as a criterion for comparison. The hyper-parameters of SVM are chosen by grid search (the penalty factor $C = 128$ and the RBF parameter $\sigma = 2^{-11}$ in this experiment). The result of experiments comparison is shown in Table 2. In this paper, we fuse CNN feature and SVM classifier to solve task transfer problem. The introduction of SVM reduces the influence of noise samples near the classification hyperplane. In addition, nonlinear kernel (RBF) increases the inter-class distance and reduces the inner-class distance. The mixture of recognition and verification task learning extracts a more robust feature than single task. The accuracy of ModelA-SVM and ModelB-SVM are 97.79% and 98.18% respectively, and both are better than the original models.

4.3. Comparison of PCA-SVM framework on LFW

In this section, the fusion of PCA distribution transfer and SVM task transfer evaluates the verification accuracy in the LFW dataset. We compare the proposed framework with deep learning based methods in LFW dataset, shown in Table 3. The PCA-SVM framework proposed in this paper solves the domain adaptation problem effectively.

The results of our framework on LFW verification outperform those of DeepFace [9], WebScale [11], DeepID [12], DeepID2 [14], WebFace [19], VGGFace [25], M-S_UTR [26], MMN [27], and Lightened-CNN [20] for **single-network**. In this paper, PCA-SVM framework proposed for advanced learning combines the generative and discriminative model: the distribution transfer based on PCA is a generative model and the task transfer based on SVM as a discriminative model. Clearly, the hybrid-model achieved excellent performances in comparison to single-network. Besides, our dataset is inferior to Facebook [9], CUHK [12] and VGG [25]. Their training sets contain millions of images while CASIA-WebFace only includes 0.5M images.

5. Conclusion

In this paper, we propose a feature transfer framework based on PCA-SVM for face verification to deal with the gap of data distribution and training task. Firstly, PCA is adopted to narrow the gap of data distribution: the model

TABLE 1. ACCURACY COMPARISONS ON DISTRIBUTION TRANSFER WITH DIFFERENT FEATURE DIMENSIONS (%)

Dimensions	30	50	70	85	100	180	256	Non-tranfer
ModelA	96.87	97.53	**98.00**	97.80	97.77	97.87	97.90	97.77
ModelB	97.63	98.17	98.27	98.30	**98.40**	98.30	98.30	98.13

TABLE 3. ACCURACY COMPARISONS WITH THE STATE-OF-THE-ART METHODS

Methods	Average Accuracy (%)
DeepFace - single [9]	95.92
DeepFace - 7models [9]	97.35
WebScale - single [11]	98.00
WebScale - 4models [11]	98.37
DeepID [12]	95.35
DeepID + Joint Bayes [12]	96.05
DeepID2 - single [14]	95.43
DeepID2 - 4models [14]	97.75
WebFace [19]	96.13
WebFace + PCA [19]	96.30
WebFace + Joint Bayes [19]	97.30
WebFace + unrestricted [19]	97.73
VGGFace [25]	97.27
MS_UTR [26]	96.95
MMN1 [27]	97.32
MMN2 [27]	98.12
ModelA [20]	97.77
ModelA + PCA (ours)	**98.00**
ModelA + SVM (ours)	**97.79**
ModelA + PCA-SVM (ours)	**98.10**
ModelB [20]	98.13
ModelB + PCA (ours)	**98.40**
ModelB + SVM (ours)	**98.18**
ModelB + PCA-SVM (ours)	**98.50**

is trained on the source dataset (CASIA-WebFace) and is tested on the target dataset (LFW). The verification accuracy of PCA based distribution transfer is increased by about 0.3% compared with the original model. Secondly, due to the difference of source task (face recognition) and target task (face verification), we introduce an SVM classifier with nonlinear kernel for task transfer. Finally, the PCA-SVM framework performs the feature transfer of distribution and task, which achieves the better validation accuracy than the existing single-network based algorithms on the LFW.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (#61171142, #61401163), the Science and technology Planning Project of Guangdong Province of China (#2014B010111003, #2014B010111006), Guangzhou Key Lab of Body Data Science (#201605030011).

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[2] P. I. Wilson and J. Fernandez, "Facial feature detection using haar classifiers," *Journal of Computing Sciences in Colleges*, vol. 21, no. 4, pp. 127–133, 2006.

[3] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image processing*, vol. 11, no. 4, pp. 467–476, 2002.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[5] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[6] J. Han, D. Zhang, X. Hu, L. Guo, J. Ren, and F. Wu, "Background prior-based salient object detection via deep reconstruction residual," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 8, pp. 1309–1321, 2015.

[7] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[8] M. Günther, A. Costa-Pazo, C. Ding, E. Boutellaa, G. Chiachia, H. Zhang, M. de Assis Angeloni, V. Štruc, E. Khoury, E. Vazquez-Fernandez *et al.*, "The 2013 face recognition evaluation in mobile environment," in *Biometrics (ICB), 2013 International Conference on*. IEEE, 2013, pp. 1–7.

[9] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1701–1708.

[10] 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, Tech. Rep., 2007.

[11] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "Web-scale training for face identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2746–2754.

[12] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1891–1898.

[13] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun, "Bayesian face revisited: A joint formulation," *Computer Vision–ECCV 2012*, pp. 566–579, 2012.

[14] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Advances in neural information processing systems*, 2014, pp. 1988–1996.

[15] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2892–2900.

[16] Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," *arXiv preprint arXiv:1502.00873*, 2015.

[17] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[18] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

[19] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *arXiv preprint arXiv:1411.7923*, 2014.

[20] X. Wu, R. He, and Z. Sun, "A lightened cnn for deep face representation," in *2015 IEEE Conference on IEEE Computer Vision and Pattern Recognition (CVPR)*, 2015.

[21] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.

[22] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," *arXiv preprint arXiv:1412.3474*, 2014.

[23] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *International Conference on Machine Learning*, 2015, pp. 97–105.

[24] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.

[25] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition." in *BMVC*, vol. 1, no. 3, 2015, p. 6.

[26] D. Wang, C. Otto, and A. K. Jain, "Face search at scale: 80 million gallery," *arXiv preprint arXiv:1507.07242*, 2015.

[27] C. Ding and D. Tao, "Robust face recognition via multimodal deep face representation," *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 2049–2058, 2015.