Deep Learning Face Representation by Joint Identification-Verification

Yi Sun[1] **Yuheng Chen**[2] **Xiaogang Wang**[3,4] **Xiaoou Tang**[1,4]
[1]Department of Information Engineering, The Chinese University of Hong Kong
[2]SenseTime Group
[3]Department of Electronic Engineering, The Chinese University of Hong Kong
[4]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
sy011@ie.cuhk.edu.hk chyh1990@gmail.com
xgwang@ee.cuhk.edu.hk xtang@ie.cuhk.edu.hk

Abstract

The key challenge of face recognition is to develop effective feature representations for reducing intra-personal variations while enlarging inter-personal differences. In this paper, we show that it can be well solved with deep learning and using both face identification and verification signals as supervision. The Deep IDentification-verification features (DeepID2) are learned with carefully designed deep convolutional networks. The face identification task increases the inter-personal variations by drawing DeepID2 features extracted from different identities apart, while the face verification task reduces the intra-personal variations by pulling DeepID2 features extracted from the same identity together, both of which are essential to face recognition. The learned DeepID2 features can be well generalized to new identities unseen in the training data. On the challenging LFW dataset [11], 99:15% face verification accuracy is achieved. Compared with the best previous deep learning result [20] on LFW, the error rate has been significantly reduced by 67%.

1 Introduction

Faces of the same identity could look much different when presented in different poses, illuminations, expressions, ages, and occlusions. Such variations within the same identity could overwhelm the variations due to identity differences and make face recognition challenging, especially in unconstrained conditions. Therefore, reducing the intra-personal variations while enlarging the inter-personal differences is a central topic in face recognition. It can be traced back to early subspace face recognition methods such as LDA [1], Bayesian face [16], and unified subspace s(J)15(oint)d whe-294(bd)1TJ002catise i1entityi02csacouldlook variations0 y

classes, while verification is to classify a pair of images as belonging to the same identity or not (i.e. binary classification). In the training stage, given an input face image with the identification signal, its DeepID2 features are extracted in the top hidden layer of the learned hierarchical nonlinear feature representation, and then mapped to one of a large number of identities through another function g(DeepID2). In the testing stage, the learned DeepID2 features can be generalized to other tasks (such as face verification) and new identities unseen in the training data. The identification supervisory signal tends to pull apart the DeepID2 features of different identities since they have to be classified into different classes. Therefore, the learned features would have rich identity-related or inter-personal variations. However, the identification signal has a relatively weak constraint on DeepID2 features extracted from the same identity, since dissimilar DeepID2 features could be mapped to the same identity through function $g(\ )$. This leads to problems when DeepID2 features are generalized to new tasks and new identities in test where g is not applicable anymore. We solve this by using an additional face verification signal, which requires that every two DeepID2 feature vectors extracted from the same identity are close to each other while those extracted from different identities are kept away. The strong per-element constraint on DeepID2 features can effectively reduce the intra-personal variations. On the other hand, using the verification signal alone (i.e. only distinguishing a pair of DeepID2 feature vectors at a time) is not as effective in extracting identity-related features as using the identification signal (i.e. distinguishing thousands of identities at a time). Therefore, the two supervisory signals emphasize different aspects in feature learning and should be employed together.

To characterize faces from different aspects, complementary DeepID2 features are extracted from various face regions and resolutions, and are concatenated to form the final feature representation after PCA dimension reduction. Since the learned DeepID2 features are diverse among different identities while consistent within the same identity, it makes the following face recognition easier. Using the learned feature representation and a recently proposed face verification model [3], we achieved the highest 99:15% face verification accuracy on the challenging and extensively studied LFW dataset [11]. This is the first time that a machine provided with only the face region achieves an accuracy on par with the 99:20% accuracy of human to whom the entire LFW face image including the face region and large background area are presented to verify.

In recent years, a great deal of efforts have been made for face recognition with deep learning [5, 10, 18, 26, 8, 21, 20, 27]. Among the deep learning works, [5, 18, 8] learned features or deep metrics with the verification signal, while DeepFace [21] and our previous work DeepID [20] learned features with the identification signal and achieved accuracies around 97:45% on LFW. Our approach significantly improves the state-of-the-art. The idea of jointly solving the classification and verification tasks was applied to general object recognition [15], with the focus on improving classification accuracy on fixed object classes instead of hidden feature representations. Our work targets on learning features which can be well generalized to new classes (identities) and the verification task.

2 Identification-verification guided deep feature learning

We learn features with variations of deep convolutional neural networks (deep ConvNets) [12]. The convolution and pooling operations in deep ConvNets are specially designed to extract visual features hierarchically, from local low-level features to global high-level ones. Our deep ConvNets take similar structures as in [20]. It contains four convolutional layers, with local weight sharing [10] in the third and fourth convolutional layers. The ConvNet extracts a 160-dimensional DeepID2 feature vector at its last layer (DeepID2 layer) of the feature extraction cascade. The DeepID2 layer to be learned are fully-connected to both the third and fourth convolutional layers. We use rectified linear units (ReLU) [17] for neurons in the convolutional layers and the DeepID2 layer. An illustration of the ConvNet structure used to extract DeepID2 features is showk.

Table 1: The DeepID2 feature learning algorithm.

input: training set $= f(x_i; l_i)g$, initialized parameters $_c$, $_{id}$, and $_{ve}$, hyperparameter , learning rate (t), t 0

while not converge **do**

t $t + 1$ sample two training samples $(x_i; l_i)$ and $(x_j; l_j)$ from

$f_i = \text{Conv}(x_i; _c)$ and $f_j = \text{Conv}(x_j; _c)$

$r_{id} = \frac{@\text{Ident}(f_i; l_i; _{id})}{@_{id}} + \frac{@\text{Ident}(f_j; l_j; _{id})}{@_{id}}$

$r_{ve} =$ $@\text{Verif}(f_i; f_j; y_{ij}; _{ve})$ [illegible]

4 Experiments

We report face verification results on the LFW dataset [11], which is the de facto standard test set for face verification in unconstrained conditions. It contains 13; 233 face images of 5749 identities

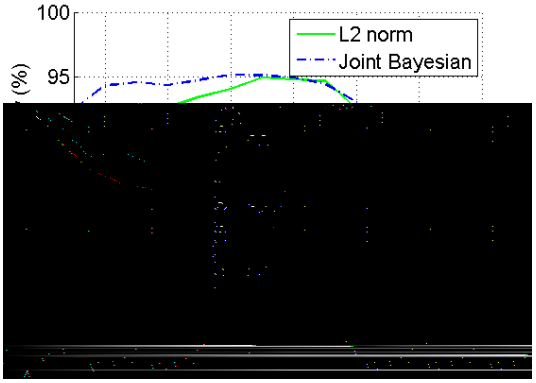

Figure 3: Face verification accuracy by varying the weighting parameter . is plotted in log scale.

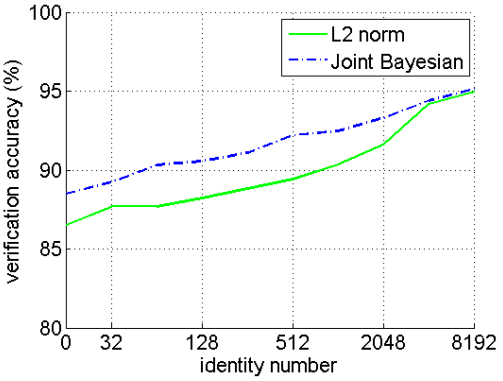

Figure 4: Face verification accuracy of DeepID2 features learned by both the the face identification and verification signals, where the number of training identities (shown in log scale) used for face identification varies. The result may be further improved with more than 8192 identities.

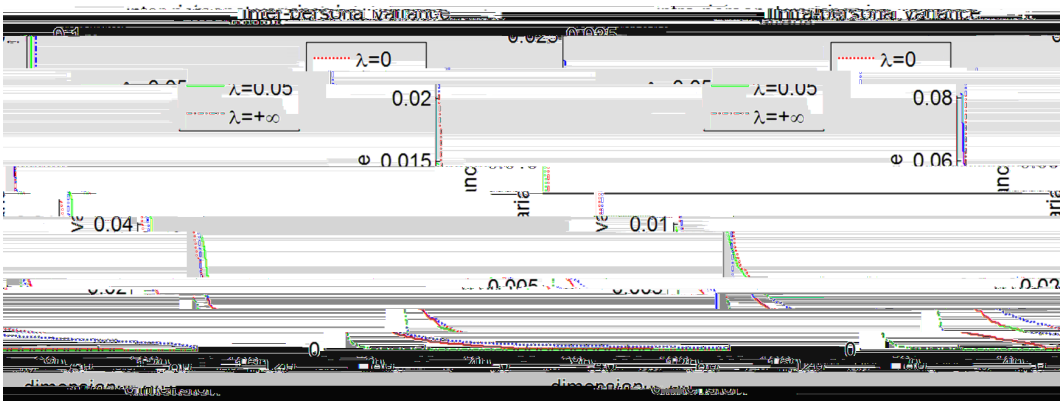

Figure 5: Spectrum of eigenvalues of the inter- and intra-personal scatter matrices. Best viewed in color.

personal variations, distinguishing different identities becomes difficult. Therefore the performance degrades significantly.

Figure 6 shows the first two PCA dimensions of features learned with = 0, 0:05, and + *1* , respectively. These features come from the six identities with the largest numbers of face images in LFW, and are marked by different colors. The figure further verifies our observations. When = 0 (left), different clusters are mixed together due to the large intra-personal variations, although the cluster centers are actually different. When increases to 0:05 (middle), intra-personal variations are significantly reduced and the clusters become distinguishable. When further increases towards infinity (right), although the intra-personal variations further decrease, the cluster centers also begin to collapse and some clusters become significantly overlapped (as the red, blue, and cyan clusters in Fig. 6 right), making it hard to distinguish again.

4.2 Rich identity information improves feature learning

We investigate how would the identity information contained in the identification supervisory signal influence the learned features. In particular, we experiment with an exponentially increasing number

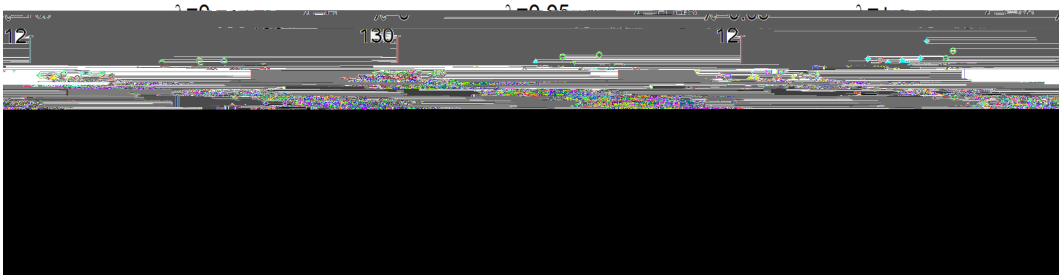

Figure 6: The first two PCA dimensions of DeepID2 features extracted from six identities in LFW.

Table 2: Comparison of different verification signals.

verification signal	L2	L2+	L2-	L1	cosine	none
L2 norm (%)	94.95	94.43	86.23	92.92	87.07	86.43
Joint Bayesian (%)	95.12	94.87	92.98	94.13	93.38	92.73

identifying a large number (e.g., 8192) of identities is key to learning effective DeepID2 feature representation. This observation is consistent with those in Sec. 4.1. The increasing number of identities provides richer identity information and helps to form DeepID2 features with diverse interpersonal variations, making the class centers of different identities more distinguishable.

4.3 Investigating the verification signals

As shown in Sec. 4.1, the verification signal with moderate intensity mainly takes the effect of

Table 3: Face verification accuracy with DeepID2 features extracted from an increasing number of face patches.

# patches	1	2	4	8	16	25
accuracy (%)	95.43	97.28	97.75	98.55	98.93	98.97
time (ms)	1.7	3.4	6.1	11	23	35

Table 4: Accuracy comparison with the previous best results on LFW.

method	accuracy (%)	
High-dim LBP [4]	95:17	1:13
TL Joint Bayesian [2]	96:33	1:08
DeepFace [21]	97:35	0:25
DeepID [20]	97:45	0:26
GaussianFace [13]	98:52	0:66
DeepID2	99:15	0:13

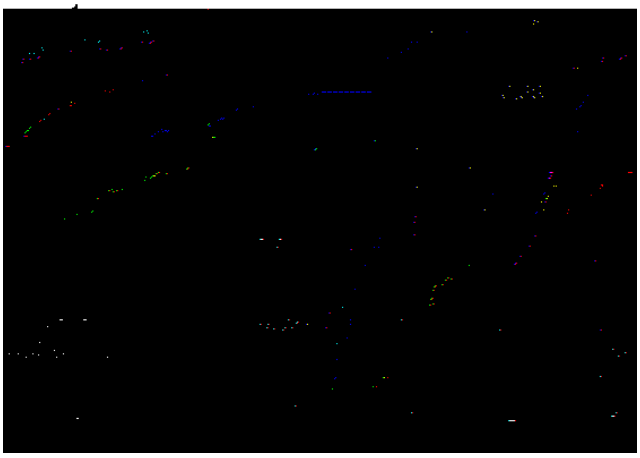

Figure 7: ROC comparison with the previous best results on LFW. Best viewed in color.

To further exploit the rich pool of DeepID2 features extracted from the large number of patches, we repeat the feature selection algorithm for another six times, each time choosing DeepID2 features from the patches that have not been selected by previous feature selection steps. Then we learn the Joint Bayesian model on each of the seven groups of selected features, respectively. We fuse the seven Joint Bayesian scores on each pair of compared faces by further learning an SVM. In this way, we achieve an even higher 99:15% face verification accuracy. The accuracy and ROC comparison with previous state-of-the-art methods on LFW are shown in Tab. 4 and Fig. 7, respectively. We achieve the best results and improve previous results with a large margin.

5 Conclusion

This paper have shown that the effect of the face identification and verification supervisory signals on deep feature representation coincide with the two aspects of constructing ideal features for face recognition, i.e., increasing inter-personal variations and reducing intra-personal variations, and the combination of the two supervisory signals lead to significantly better features than either one of them. When embedding the learned features to the traditional face verification pipeline, we achieved an extremely effective system with 99:15% face verification accuracy on LFW. The arXiv report of this paper was published in June 2014 [19].

References

[1] P. N. Belhumeur, J. a. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *PAMI*, 19:711–720, 1997.

[2] X. Cao, D. Wipf, F. Wen, G. Duan, and J. Sun. A practical transfer learning algorithm for face verification. In *Proc. ICCV*, 2013.

[3] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun. Bayesian face revisited: A joint formulation. In *Proc. ECCV*, 2012.

[4] D. Chen, X. Cao, F. Wen, and J. Sun. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In *Proc. CVPR*, 2013.

[5] S. Chopra, R. Hadsell, and Y. LeCun. Learning a similarity metric discriminatively, with application to face verification. In *Proc. CVPR*, 2005.

[6] M. Guillaumin, J. Verbeek, and C. Schmid. Is that you? Metric learning approaches for face identification. In *Proc. ICCV*, 2009.

[7] R. Hadsell, S. Chopra, and Y. LeCun. Dimensionality reduction by learning an invariant mapping. In *Proc. CVPR*, 2006.

[8] J. Hu, J. Lu, and Y.-P. Tan. Discriminative deep metric learning for face verification in the wild. In *Proc. CVPR*, 2014.

[9] C. Huang, S. Zhu, and K. Yu. Large scale strongly supervised ensemble metric learning, with applications to face verification and retrieval. *NEC Technical Report TR115*, 2011.

[10] G. B. Huang, H. Lee, and E. Learned-Miller. Learning hierarchical representations for face verification with convolutional deep belief networks. In *Proc. CVPR*, 2012.

[11] 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, 2007.

[12] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998.

[13] C. Lu and X. Tang. Surpassing human-level face verification performance on LFW with GaussianFace. Technical report, arXiv:1404.3840, 2014.

[14] A. Mignon and F. Jurie. PCCA: A new approach for distance learning from sparse pairwise constraints. In *Proc. CVPR*, 2012.

[15] H. Mobahi, R. Collobert, and J. Weston. Deep learning from temporal coherence in video. In *Proc. ICML*, 2009.

[16] B. Moghaddam, T. Jebara, and A. Pentland. Bayesian face recognition. *PR*, 33:1771–1782, 2000.

[17] V. Nair and G. E. Hinton. Rectified linear units improve restricted Boltzmann machines. In *Proc. ICML*, 2010.

[18] Y. Sun, X. Wang, and X. Tang. Hybrid deep learning for face verification. In *Proc. ICCV*, 2013.

[19] Y. Sun, X. Wang, and X. Tang. Deep learning face representation by joint identification-verification. Technical report, arXiv:1406.4773, 2014.

[20] Y. Sun, X. Wang, and X. Tang. Deep learning face representation from predicting 10,000 classes. In *Proc. CVPR*, 2014.

[21] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. DeepFace: Closing the gap to human-level performance in face verification. In *Proc. CVPR*, 2014.

[22] X. Wang and X. Tang. Unified subspace analysis for face recognition. In *Proc. ICCV*, 2003.

[23] X. Wang and X. Tang. A unified framework for subspace face recognition. *PAMI*, 26:1222–1228, 2004.

[24] X. Xiong and F. De la Torre Frade. Supervised descent method and its applications to face alignment. In *Proc. CVPR*, 2013.

[25] T. Zhang. Adaptive forward-backward greedy algorithm for learning sparse representations. *IEEE Trans. Inf. Theor.*, 57:4689–4708, 2011.

[26] Z. Zhu, P. Luo, X. Wang, and X. Tang. Deep learning identity-preserving face space. In *Proc. ICCV*, 2013.

[27] Z. Zhu, P. Luo, X. Wang, and X. Tang. Deep learning and disentangling face representation by multi-view perceptron. In *Proc. NIPS*, 2014.