ELSEVIER

Pattern Recognition Letters
journal homepage: www.elsevier.com

Deep Sparse Representation Classifier for Facial Recognition and Detection System

Eric-Juwei Cheng[1], Kuang-Pen Chou[2], Shantanu Rajora[3], Bo-Hao Jin[1], M. Tanveer[4], Chin-Teng Lin[5], Ku-Young Young[1], Wen-Chieh Lin[2], Mukesh Prasad[5]

[1]Department of Electrical Engineering, National Chiao Tung University, Hsinchu, Taiwan
[2]Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan
[3]Department of Information Technology, Delhi Technological University, New Delhi, India
[4]Discipline of Mathematics, Indian Institute of Technology Indore, Madhya Pradesh, India
[5]Centre for Artificial Intelligence, School of Software, FEIT, University of Technology Sydney, Australia

ABSTRACT

This paper proposes a two-layer Convolutional Neural Network (CNN) to learn the high-level features which utilizes to the face identification via sparse representation. Feature extraction plays a vital role in real-world pattern recognition and classification tasks. The details description of the given input face image, significantly improve the performance of the facial recognition system. Sparse Representation Classifier (SRC) is a popular face classifier that sparsely represents the face image by a subset of training data, which is known as insensitive to the choice of feature space. The proposed method shows the performance improvement of SRC via a precisely selected feature exactor. The experimental results show that the proposed method outperform other methods on given datasets.

Keywords: Face recognition, Deep learning, Feature extraction, Convolutional neural network, Sparse representation classifier

1. Introduction

In the past few year*s*, facial recognition system has been paid much attention due to its value for practical applications and theoretical challenges [1-4]. The technologies of face recognition have been widely used in various applications, such as public security, criminal identification, multimedia data management, etc. Moreover, various methods have been proposed and represented a great advantage in the field of facial and pattern recognition system. Despite of these achievements, face recognition still has significant challenges with respect to unconstrained conditions. The image of a face changes with variations, such as facial expression, pose, illumination conditions, noise, etc. All of these factors associated with uncontrolled environments, which degrade the recognition rate of facial recognition system. To handle these issues, robustness of the feature extracted from facial appearance descriptors should be seen as a crucial issue. Till date, numerous well-known methods for feature extraction have been introduced, including Local Binary Pattern (LBP) [5], Histogram of Oriented Gradients (HOG) [6], Scale Invariant Feature Transform (SIFT) [7], etc. Although, these handcrafted features lead to reasonable results in various applications, these pre-defined features are not tuned for the target object. For this reason, these features are only adaptive to particular data type and leads the results in poor performance on other unknown usages.

The architectures of deep learning attempt to learn multiple-level feature [32] in a hierarchical way, which makes highly invariant and discriminative representation [22] of the input data. Over the past several years, various deep learning techniques were proposed, e.g., Deep Belief Network (DBN) [16], Restricted Boltzmann Machines (RBM) [17], Deep Boltzmann Machine (DBM) [18], Deep Neural Networks (DNN) [9], Convolutional Neural networks (CNN) [19], etc. The deep learning methods have been demonstrated that its representation power achieves excellent performance on image classification [8]. The technologies of deep learning are successfully applied to variety of research areas, such as speech recognition [9], object detection [10], pedestrian detection [11], and face recognition [12-15, 25]. The Convolutional Neural Network (CNN) is a bio-inspired artificial neural network system, which learns high-level representation directly from raw pixel image. In general, CNN consists of several convolution layers, which are followed by a pooling layer, and then the output is being passed to a fully-connected network to perform the identification process. The benefits of CNN are that, It can extract shift-invariant local features from input images based on the concepts of local receptive field, shared weight, spatial subsampling; and more importantly, it can be efficiently trained on large images with a very small amount of training parameters. It has been shown that CNN achieves impressive performance on large-scale image recognition [8].

Recently, Sparse Representation Classifier (SRC) has attracted many researcher and engineers from the face and pattern recognition areas due to its impressive performance and robustness on occlusion and noise issues [20]. The principle of SRC is to find a sparse representation of the test samples as a linear combination of the whole set of training samples by solving a L1-minimization problem. Once L1-minimization computation is finished, SRC selects the subset of training samples, which most compactly expresses the test samples and rejects all other less compact representation. Furthermore, SRC does not has the training process for its classification; so, there is no need to train the SRC model again when a new face data is added into training set.

Although, SRC achieved considerable result in occlusion and illumination environment [20], it is sensitive to the misalignment of the cropped face image. Therefore, a CNN-based feature extractor is considered to alleviate the effect of misalignment by its shift-invariant property. In this paper, we propose a two-layer deep convolutional neural network (CNN) for feature extraction and sparse representation classification for identification. The remainder of the paper is organized as follows. Section 2 introduces the proposed system. Section 3 demonstrates the experiment results and conclusion is presented in Section 4.

2. Proposed Method

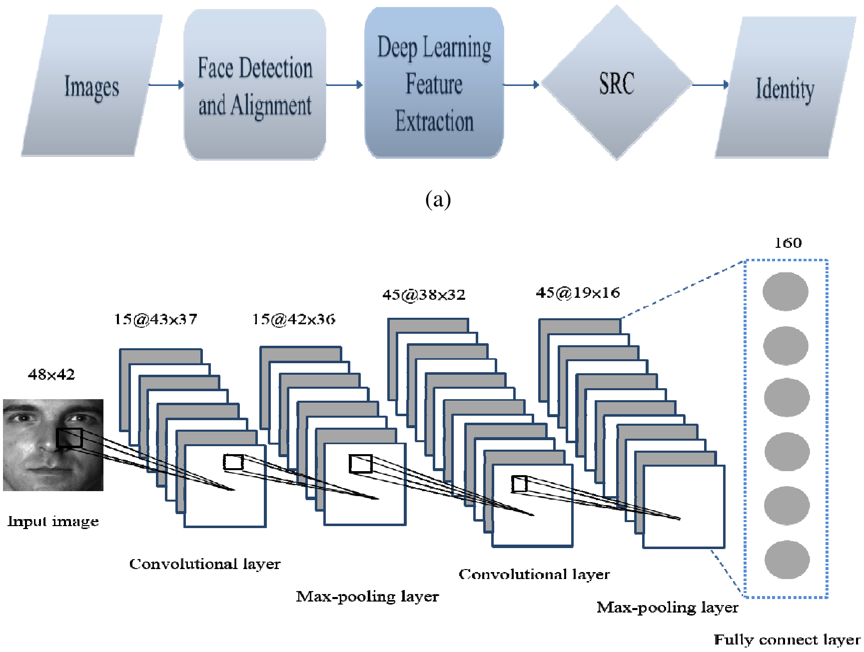
2.1. System Overview

Convolutional Neural Network (CNN) learns a hierarchical representations from the training image. Furthermore, the feature maps extracted by the CNN-based model is shown to be sparse and selective that effectively improve the discriminative power of face recognition system [14, 33-39]. The overall architecture of the proposed method is shown in Fig. 1. Fig. 1 (a) shows the flow chart of the proposed facial recognition model and fig. 1 (b) shows the proposed CNN based model for feature extraction. The proposed CNN model is composed of two convolution layers with max-pooling, and a fully connected layer which generates highly compact and predictive features for identification work. Once CNN model is trained, its output feature maps are used to perform the identification task via SRC.

The proposed CNN architecture is implemented with the open source deep learning framework called Caffe [21], which is widely-adopted recently in research associated with deep learning. The details architecture of proposed CNN is described in Fig. 2, which contains two convolution layers with max-pooling, followed by a fully-connected layer, and softmax output layer indicating identity classes in the training stage. In the test stage, the softmax layer is replaced with the SRC and the output of fully-connected layer is fed to the SRC.

2.2. Overfitting Issues

In spite of the significant success in large-scale image classification, one typical challenge to CNN is that, it can easily suffer from overfitting without a large amount of training data. As we train a model with excessive parameters and insufficient training data, the models get overfitting problem, which does not generalizes well to other unseen data. Thus, the overfitted model can almost perfectly predict training data, however fails when predicting test data. An averaging model approach is applied to train several different models on subsets of dataset then average the outputs of these separately trained networks. Averaging model is helpful to improve the performance of machine learning techniques; however, it is very expensive to train many different large networks. Moreover, the large networks generally need large amounts of training data and there may not be enough data available to train different networks on different subsets of the data.



(a)

(b)

Fig. 1. Proposed method. (a) flow chart of proposed face recognition method (b) proposed CNN for feature extraction

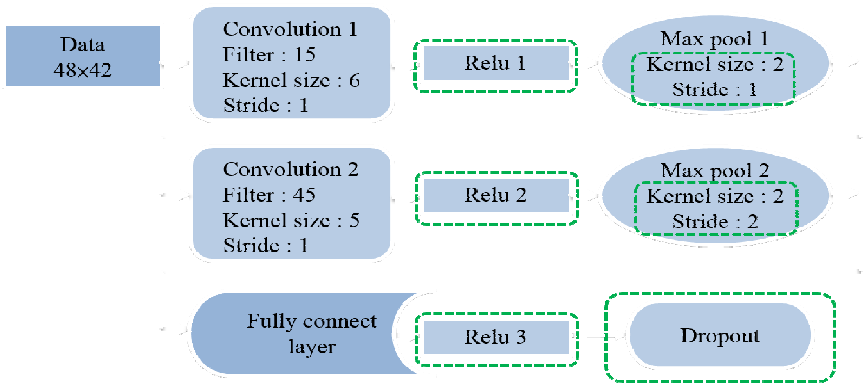


Fig. 2. Proposed architecture of CNN with parameters

Dropout is a powerful technique, which helps to reduce the generalization problem to large neural network model [20]. The concept of dropout jointly trains several models sharing subsets of parameters and input dimensions, which is similar to averaging model. This paper utilizes dropout in the fully-connected layer, as shown in fig. 2. The concept of dropout by comparing the dropout setting with the standard neural network is shown in fig. 3. During training time, dropout randomly removes some hidden units with the probability of 0.5. The output of the removed units is set to zero, that is, they neither contribute to the forward pass nor participate in backpropagation process. For a neural net with n units, dropout can be seen to create 2^n possible models by dropping some units in each epoch, and sampling from these models randomly. When, the model is being used at test stage, the dropout strategy at training time is replaced by a simple approximate averaging method that uses the network contains all of the hidden units. However, with their outgoing weights halved due to the fact that only half of them are used during training time. The results of dropout method shows that it is able to reduce complex co-adaptation of neuron, and mitigate overfitting in reasonable training time [8].

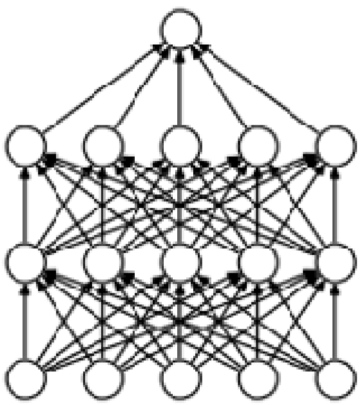
2.3. The Robustness of CNN Feature

The two important factors for the success of CNN in the large-scale face recognition [40-41] task are the sparsity of the feature extracted from the face image and the selectivity between different identities. Fig 4. displays an example of test image and the visualization result of the CNN model. It can be seen clearly that only around one half of the neurons in the hidden layers are activated, and the other half of the neurons are having zero output. In other words, only particular neurons are active with respect to the test face image. Such sparsity attribute of deep features can significantly improve the discriminative power of facial recognition system.

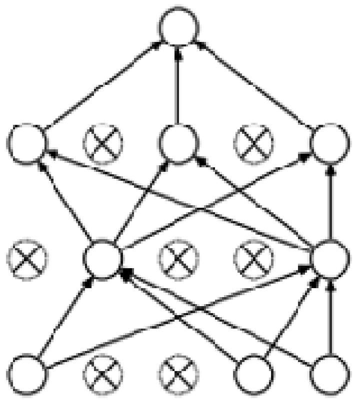
To demonstrate the selectivity of the CNN feature, two example of test images under the variant illumination condition are introduced, and the activation result of fully-connected layer corresponding to these test images are described in Fig 5. Two facts can be observed from fig. 5 that, first, both the face images excite a subset of neurons; however their activation pattern is totally different. Second, the same identity under different illumination conditions has very similar activation result. It shows that the neural activation is sparse and highly selective to the attribute of face images.

3. Experiment Results

To validate the proposed recognition framework, a series of experiments on four widely used face datasets, including Extended YALE B database [23], AR database [24], MIT faces database [30], and ORL faces database [31] were conducted. During these experimental investigations we had employed two evaluation protocols as suggested in the state of the art. The first evaluation protocol (P1) takes half of the images of each individual in the dataset as training data, and the remaining as the testing set. The second evaluation protocol (P2) adopts 10-fold cross-validation. For the Extended YALE B and the AR Database, the performance of the proposed scheme is duly compared with other canonical technologies, which adopt SRC as the classifier against both P1 and P2 evaluation protocols. For the MIT and ORL faces datasets, the performance of the proposed architecture is evaluated against the P2 evaluation protocol. Moreover, an elucidation of CNN architecture deployment for sparse feature extraction is presented in the following subsection.



(a)



(b)

Fig. 3. Dropout method. (a) A standard neural network. (b) A neural network with dropout

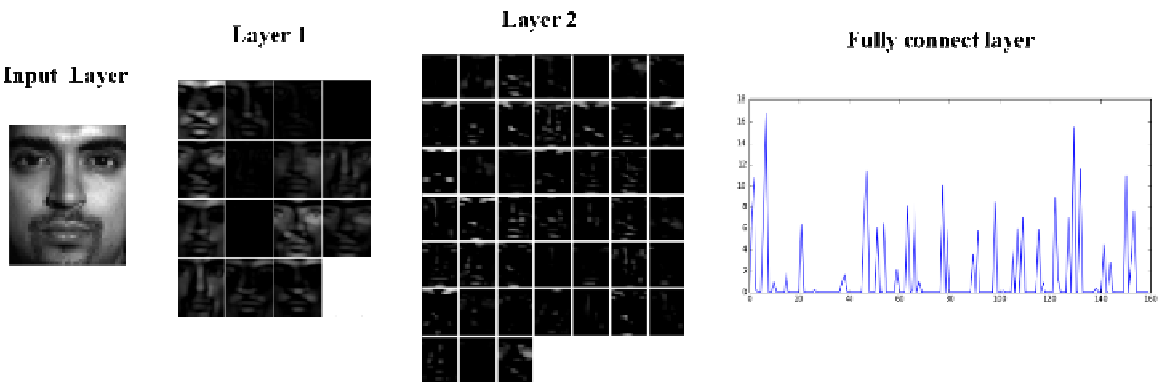


Fig. 4. Features in each layer of the proposed CNN

3.1. Different CNN Architectures for Feature Extraction

As elucidated in the previous sections, we use the former part of CNN, except the output layer, i.e the intermediate layer output, to produce sparse features as the input of SRC rather than other traditional compact features. The proposed feature extractor involves two convolution layers with 15 and 45 feature maps, respectively, and a fully connected layer with 160 neurons. This CNN architecture is named as CNN-15-45-160. In this section, we describe the two CNNs with different number of feature maps and convolution layers that are deployed during experimental investigations. The first one, referred to as CNN-6-16-160, maintains the same architecture but different feature maps used in each convolution layer. The other network, termed as CNN-20-40-60-80-160, has an increased depth with an addition of two convolution layers to extract more "deeper" features. The number of feature maps for each layer in the CNN-20-40-60-80 corresponds to 20, 40, 60 and 80, respectively. Fig. 6 depicts the architectures of these two CNN sparse feature extractors. The experiments reveal that deeper architecture reaps higher recognition rate, however, it needs to take the more computational cost to bring a slight improvement as shown in table 1.

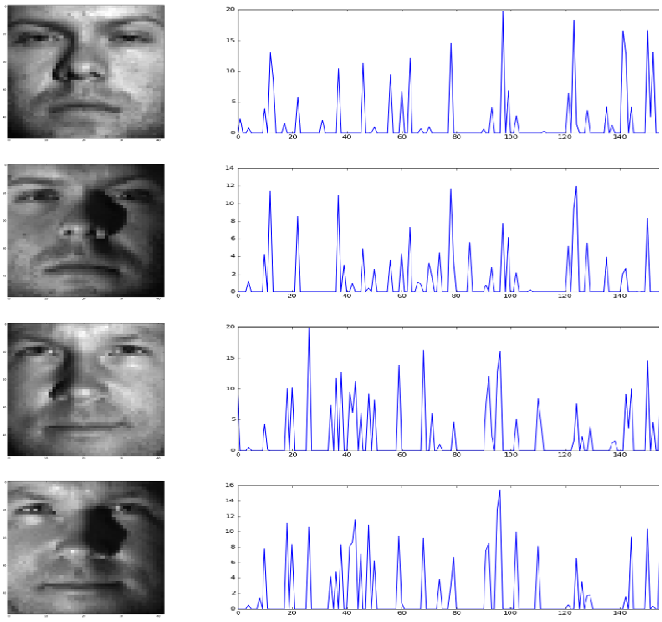
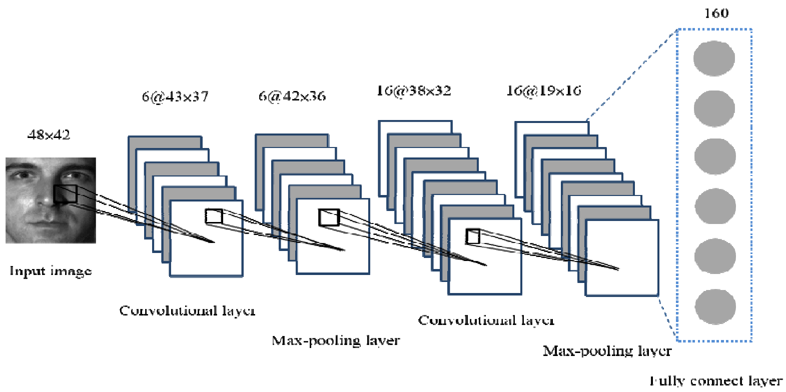
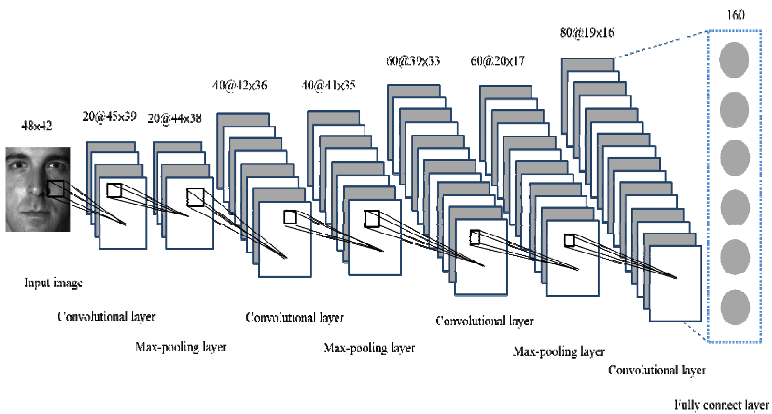


Fig. 5. Sparsity and selectivity of deep neural activations



(a)



(b)

Fig. 6. Different CNN architectures for comparison. (a) 6-16-160 (b) 20-40-60-80-160

3.2. Evaluation Protocol 1 (P1)

3.2.1. Extended Yale B Database

Two experiments are carried out using P1 evaluation under different dimensions to validate the performance of the proposed architecture. At first, we compare other existing methodologies and algorithms based on SRC with different features, including Eigenfaces [4], Laplacianfaces [26], Fisherfaces [27], randomfaces [28], and down-sampled images. The method of deep learning feature extraction is superior to the other feature extraction methods in each dimension as shown in fig. 7. Further, we compare the performance of the proposed method and the traditional CNN framework, which adopts the softmax layer as the classifier shown in table 2. Experimental results indicate that the combination of a decent CNN feature extractor and SRC outperform the standard CNN. The best recognition for each method under the experimental dimensions is given in table 3.

3.2.2. AR Database

Similar experiments in the previous section are performed on the AR face dataset. Fig. 8 shows the comparison of the fetched results across different SRC based methods on this dataset. It is clear that the proposed method is superior to the competitive methods on each dimension, except the Fisherfaces [29] method in low dimension cases. We also evaluate the proposed approach with the standard CNN as shown in table 4. The results show that our proposed architecture has better performance on each evaluation dimension. Table 5 lists the best performance of each compared method under various experimental dimensions.

Table 1. Best performance of all competitive methods under the testing dimension on the Extended YALE B database.

	CNN-6-16-160	CNN-15-45-60	CNN-20-40-60-80
Deep learning + softmax	95.41%	98.33%	98.83%
Deep learning + SRC	95.83%	98.58%	98.50%
Training Iterations	4500	5100	10000

Table 2. Recognition rate comparison of standard CNN and the proposed framework on Extended Yale B database.

Method	Dimensions				
	30	56	120	160	540
Deep learning + softmax	96.68 %	97.33 %	97.78 %	98.10 %	98.85 %
Deep learning + SRC	97.00 %	97.80 %	98.33 %	98.35 %	99.17 %

Table 3. Best performance of all competitive methods under the testing dimension on the Extended YALE B database.

Combinations	Dimensions	Recognition rate
Eigen + SRC	504	96.77%
Laplacian + SRC	504	96.52%
Random + SRC	504	98.09%
Down Sample + SRC	504	97.10%
Fisher + SRC	30	86.91%
Deep Learning + softmax	504	98.92%
Deep Learning + SRC	504	99.17%

Table 4. Recognition rate comparison of standard CNN and the proposed framework on AR database.

Method	Dimensions				
	30	54	130	160	540
Deep learning + softmax	77.40 %	85.84 %	92.56 %	93.42 %	94.42 %
Deep learning + SRC	78.68 %	85.98 %	93.99 %	94.56 %	95.85 %

Table 5. Best performance of all competitive methods under the testing dimension on the AR database

Combinations	Dimensions	Recognition rate
Eigen + SRC	504	91.99%
Laplacian + SRC	504	94.28%
Random + SRC	504	94.70%
Down Sample + SRC	504	93.85%
Fisher + SRC	54	92.27%
Deep Learning + softmax	504	94.42%
Deep Learning + SRC	504	95.85%

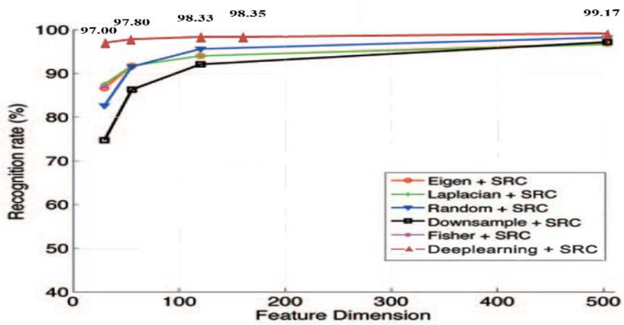


Fig. 7. Recognition rate comparison by using various feature extraction methods with SRC in the Extended Yale B database

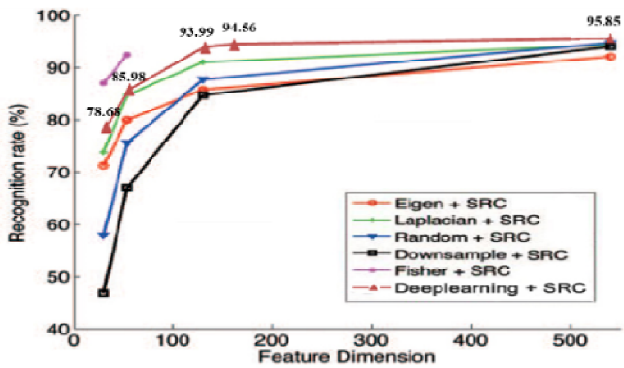


Fig. 8. Recognition rate comparison by using different feature extraction methods with SRC in the AR database

3.3. Evaluation Protocol 2 (P2)

3.3.1. Extended Yale B Database

In this experimental setting, a 10-fold cross-validation scheme is undertaken for evaluation of the proposed architecture. The Extended YaleB dataset involves a total 2146 samples of 26 individuals. The results across this dataset are pronounced in table 6. It achieves 0.9954, 0.9939 and 0.9947 for recall, precision and f1-score, respectively. Fig. 9 shows the confusion matrix of testing.

3.3.2. AR Database

All the 26 distinct faces of each person in the AR dataset participate for either training or testing phase according to the P2 protocol. The experiment results are described in table 6. It is to be noted that the performance reaches 1 for recall, precision and f1-score, which is 100 percentage recognition for each subject. Fig. 10 draws the confusion matrix of this experiment.

3.3.3. MIT Database

To further verify the robustness and effectiveness of the proposed framework, we conducted experimental investigations on the MIT face database [30] against the P2 validation protocol. There are a total of 3240 images of each subject adopted for evaluation. The proposed method operates phenomenally well in terms of recall, precision and f1-score as shown in table 6. Fig. 11 illustrates the confusion matrix for testing.

3.3.4. ORL Face Database

The ORL face database [31] by the Cambridge AT&T laboratory consists of 400 images for 40 individuals with frontal and slight tile head pose; with 10 images for each person. We evaluate the proposed method against the P2 validation protocol. The recognition results touch to 0.9925, 0.9963, and 0.9944, for recall, precision, and f1-score as shown in table 6. The confusion matrix for testing is given in fig. 12.

Table 6. Results for four face datasets with P2 validation (10 fold cross validation).

Dataset	Recall	Precision	F1-score
Extended YALE B	0.9954	0.9939	0.99465
AR	1	1	1
MIT	1	1	1
ORL	0.9925	0.9963	0.9943

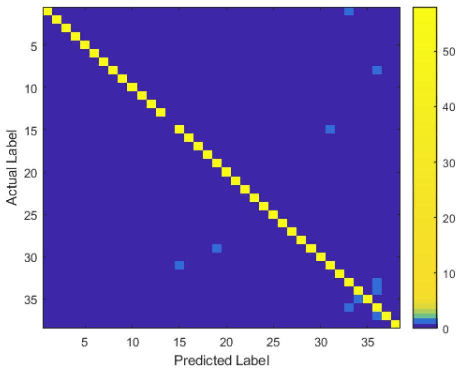


Fig. 9. Confusion matrix for Extended YALE B dataset.

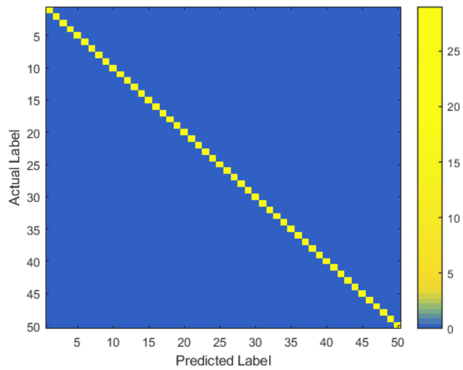


Fig. 10. Confusion matrix for AR dataset.

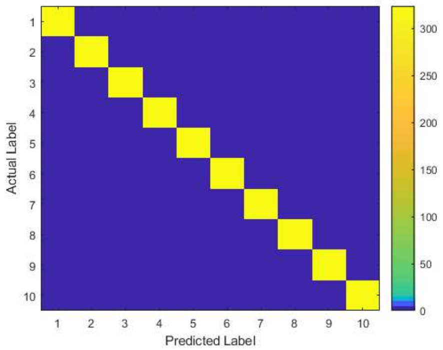


Fig. 11. Confusion matrix for MIT dataset

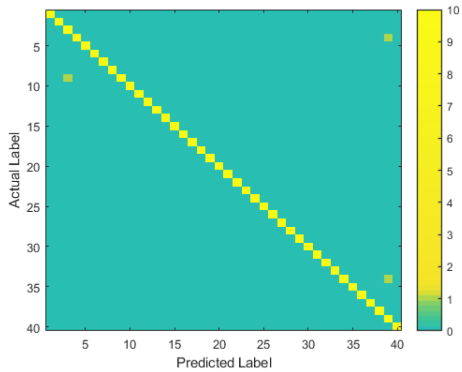


Fig. 12. Confusion matrix for ORL dataset

4. Conclusions and Future Work

This paper proposes a facial recognition model which is composed with a two-layer deep CNN for feature extraction and SRC for classification. SRC provides better classification result even if a simple feature extraction method is used. The proposed method shows that by choosing precise feature space can improve the performance of SRC. Also, the proposed system is highly resistant to variations of illumination and expression of the facial images.

Although CNN has shown superior performance in the image classification area, the huge amount of trainable parameters make it difficult to train when small dataset is used. Furthermore, SRC try to construct a training dictionary to sparsely represent the test image; that is, the performance of SRC is also influenced by the size of dataset. For future work, the performance of the proposed system would be evaluated on large scale dataset.

REFERENCES

1. 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.
2. W. Zhao and R. Chellappa, "Image-based Face Recognition: Issues and Methods", *Optical Engineering-New York-Marcel Dekker Incorporated*, vol. 78, pp. 375-402, 2002.
3. R. M. Ebied, "Feature Extraction using PCA and Kernel-PCA for Face Recognition", *8th International Conference on Informatics and Systems*, vol. 8, pp. 72-77, 2012.
4. M. A. Turk and A. P. Pentland, "Face Recognition Using Eigenfaces", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1991.
5. T. Ahonen, A. Hadid, and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037-2041, 2006.
6. N. Dalal and B. Triggs. "Histograms of Oriented Gradients for Human Detection", *CVPR*, 2005.
7. D. G. Lowe, "Distinctive Image Features from Scale-invariant Keypoints", *IJCV*, vol. 60, no. 2, pp. 91-110, 2004.
8. A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks", *NIPS*, 2012
9. G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury, "Deep Neural Networks for Acoustic Modeling in Speech Recognition", *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82-97, 2012.
10. R. Girshick, J. Donahue, T. Darrell, and J. Malik. "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation". *CVPR*, 2014.

11. W. Ouyang and X. Wang. "Joint Deep Learning for Pedestrian Detection", *ICCV*, 2013.
12. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701-1708, 2014
13. G. B. Huang, H. Lee, and E. L. Miller, "Learning Hierarchical Representations for Face Verification with Convolutional Deep Belief Networks", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2518-2525, 2012.
14. Y. Sun, X. Wang, and X. Tang, "Deep Learning Face Representation from Predicting 10,000 Classes", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1891-1898, 2014.
15. Y. Sun, X. Wang, and X. Tang, "Deep Learning Face Representation by Joint Identification-Verification", *Advances in Neural Information Processing Systems*, pp. 1988-1996, 2014.
16. G. E. Hinton, S. Osindero, and Y. W. Teh. "A Fast Learning Algorithm for Deep Belief Nets", *Neural Computation*, vol. 18, no. 7, pp. 527-1554, 2006.
17. R. Salakhutdinov, A. Mnih, and G. E. Hinton. "Restricted Boltzmann Machines for Collaborative Filtering", *ICML*, pp. 791-798, 2007.
18. R. Salakhutdinov and G. E. Hinton. "Deep Boltzmann Machines", *AISTATS*, pp. 448-455, 2009.
19. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based Learning Applied to Document Recognition", *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.
20. J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust Face Recognition via Sparse Representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210-227, 2009.
21. Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. "Caffe: Convolutional Architecture for Fast Feature Embedding", *Proceedings of the 22nd ACM international conference on Multimedia*, Orlando, Florida, USA, 2014.
22. Y. Wen, K. Zhang, Z. Li, and Yu Qiao, "A Discriminative Feature Learning Approach for Deep Face Recognition", *European Conference on Computer Vision, ECCV* 2016, pp. 499-515, 2016.
23. K. Lee, J. Ho, and D. Kriegman. "Acquiring linear subspaces for face recognition under variable lighting", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684-698, 2005.
24. A. Martinez and R. Benavente, "The AR Face Database", *CVC Technical Report* 24, 1998.
25. B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "Openface: A General-Purpose Face Recognition Library with Mobile Applications", *Technical report*, CMU-CS-16-118, CMU School of Computer Science, 2016.
26. X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang, "Face Recognition Using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328-340, 2005.
27. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.
28. S. Kaski, "Dimensionality Reduction by Computation for Clustering," *IEEE International Joint Conference on Neural Networks Proceedings*, vol. 1, pp. 4-9, 1998.
29. D. L. Li, M. Prasad, S. C. Hsu, C. T. Hong, C. T. Lin, "Face Recognition Using Nonparametric Weighted Fisherfaces," *EURASIP, Journal of Advance Signal Processing*, vol. 92, pp. 1-11, 2012.
30. MIT face database (n.d.). Retrieved June 6, 2003 from ftp://whitechapel.media.mit.umich.edu/pub/images/.
31. ORL face database. AT&T Laboratories, Cambridge, U.K. [Online]. Available: http://www.cam-orl.co.uk/facedatabase.html
32. R. Rajeev, V. M. Patel, and R. Chellappa, "Hyperface: A Deep Multi-task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 1, pp. 121-135, 2017.
33. S. Zeng, J. Gou, and X. Yang, "Improving Sparsity of Coefficients for Robust Sparse and Collaborative Representation-based Image Classification", *Neural Computing and Applications*, vol. 30, no. 10, pp. 2965–2978, 2018.
34. X. Wu, R. He, Z. Sun, and T. Tan, "A Light CNN for Deep Face Representation With Noisy Labels", *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2884-2896, 2018.
35. X. Liu, L. Lu, Z. Shen, and K. Lu, "A Novel Face Recognition Algorithm via Weighted Kernel Sparse Representation", *Future Generation Computer Systems*, vol. 80, pp. 653-663, 2018.
36. J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, and Y. Xu, "Nuclear Norm Based Matrix Regression with Applications to Face Recognition with Occlusion and Illumination Changes", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 1, pp. 156-171, 2016.
37. J. Yang, P. Ren, D. Zhang, D. Chen, F. Wen, H. Li, and G. Hua; "Neural Aggregation Network for Video Face Recognition" *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2017)*, pp. 4362-4371, 2017.
38. H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, "CosFace: Large Margin Cosine Loss for Deep Face Recognition", *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR-2018)*, pp. 5265-5274, 2018.
39. W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "SphereFace: Deep Hypersphere Embedding for Face Recognition", *In Conference on Computer Vision and Pattern Recognition (CVPR-2017)*, 2017.
40. Y. Wu, J. Li, Y. Kong, and Y. Fu, "Deep Convolutional Neural Network with Independent Softmax for Large Scale Face Recognition", *Proceedings of the 24th ACM international conference on Multimedia*, pp. 1063-1067, 2016.
41. L. Wan, N. Liu, H. Huo, T. Fang, "Face Recognition with Convolutional Neural Networks and Subspace Learning", *2nd International Conference on Image, Vision and Computing (ICIVC-2017)*, 2017.