Deep Sparse Representation Classifier for Facial Recognition and Detection System

Eric-Juwei Cheng[1], Mukesh Prasad[2], Kuang-Pen Chou[3], Bo-Hao Jin[1], Deepak Puthal[2], Ku-Young Young[1], Wen-Chieh Lin[3], Chin-Teng Lin[2]

[1]Department of Electrical Engineering, National Chiao Tung University, Hsinchu, Taiwan
[2]Centre for Artificial Intelligence, School of Software, FEIT, University of Technology Sydney, Australia
[3]Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan

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. In the experimental results we compare the CNN feature to various feature extraction methods and it can be easily seen that the proposed method outperform other methods.

Keywords: Face recognition, Deep learning, feature extraction, Convolutional Neural Network, Sparse Representation Classifier

1. Introduction

In the past few years, facial recognition system has been paid much attention due to its value for practical applications and theoretical challenges [1-4]. 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 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, the 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, they are only adaptive to particular data type and leads the results in poor performance on other unknown usages.

Deep learning architectures attempt to learn multiple-level feature in a hierarchical way that makes highly invariant and discriminative representation of the input data. Over the past several years, several 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. 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 a variety of research areas such as speech recognition [9], object detection [10], pedestrian detection [11], and face recognition [12-15]. Convolutional Neural networks (CNN) is a bio-inspired artificial neural network system which learn high-level representation directly from raw pixel image. In general, CNN consist 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 they can extract shift-invariant local features from input images based on the concepts of local receptive field, shared weight, spatial subsampling; and more importantly, CNN 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. When the 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 have 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.

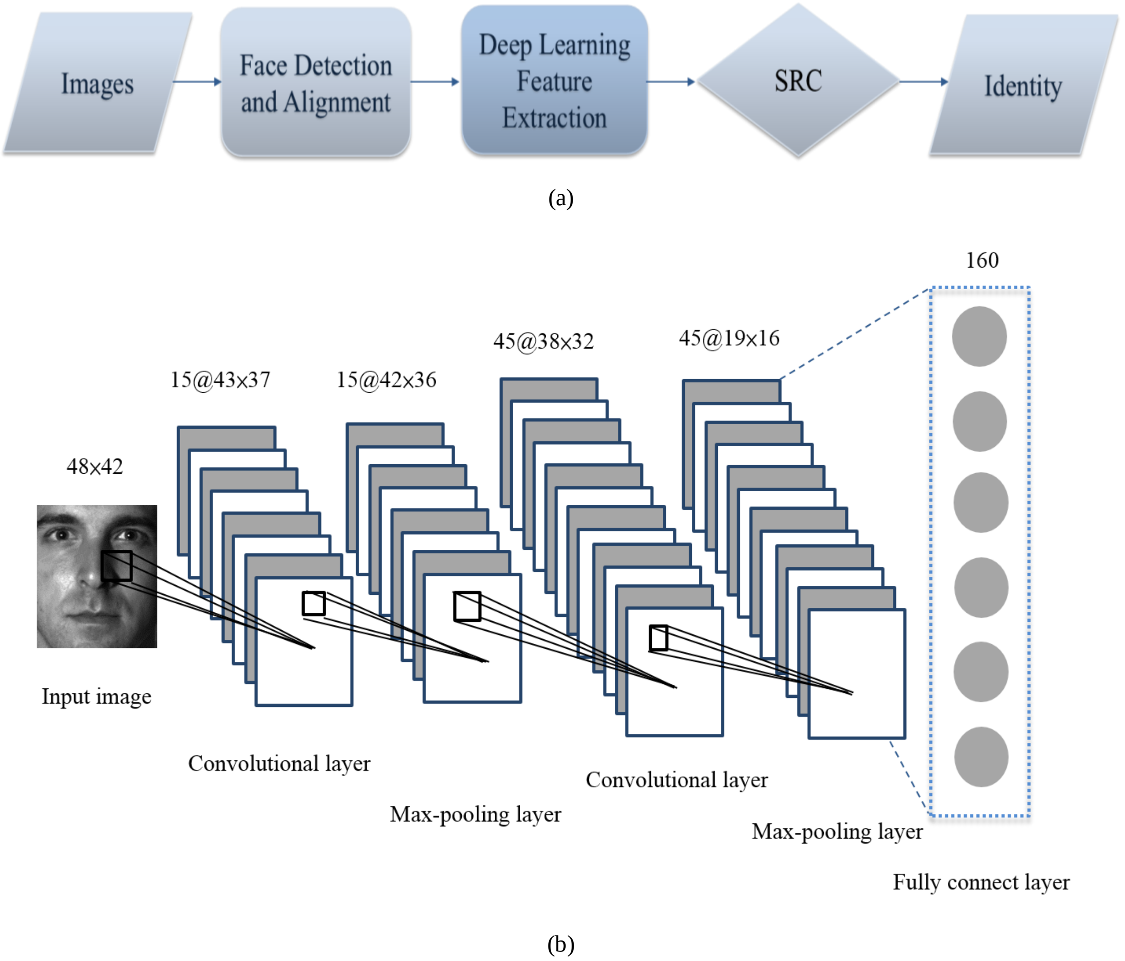
2. Proposed Method

2.1. System Overview

Convolutional neural networks (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]. 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. When CNN model is trained, its output feature maps are used to perform the identification task via sparse representation classifier (SRC).

2.2. Proposed CNN Architecture

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.



(a)

(b)

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

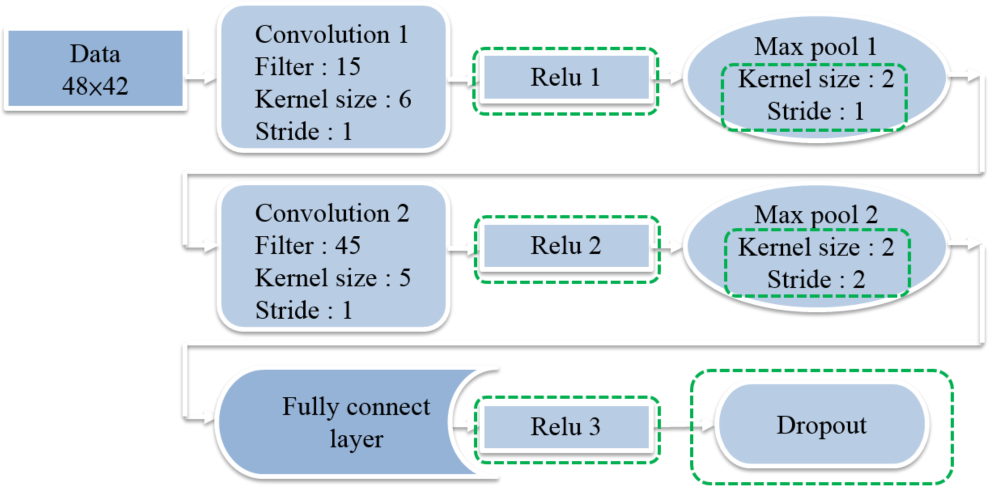


Fig. 2. Proposed architecture of CNN with parameters

2.3. Overfitting issues

In spite of the significant success in large-scale image classification, one typical challenge to CNN is that they 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, but 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, 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.

Dropout is a powerful technique that 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. Fig 3. shows the concept of dropout by comparing the dropout setting with the standard neural network. 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 we are 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 use the network contains all of the hidden units, but 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]. This paper utilizes dropout in the fully-connected layer, as described in Fig 2.

2.4. The robustness of CNN feature

The two important factors for the success of CNN in the large-scale face recognition 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. 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.

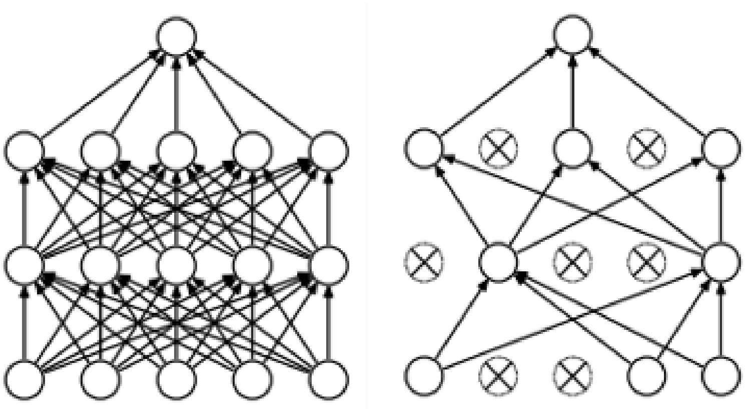


Fig. 3. Dropout method. (a) A standard neural network. (b) A neural network with dropout

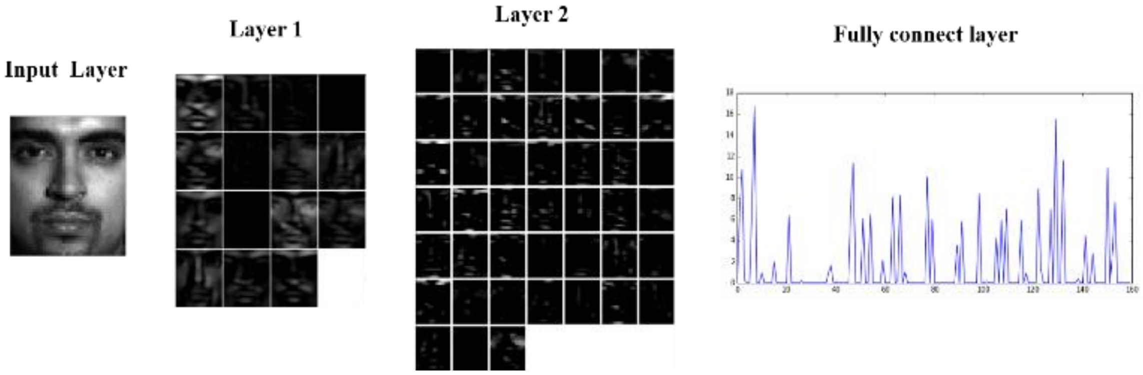


Fig. 4. Features in each layer of the proposed CNN

3. Experiment Results

3.1. Databases

For experiment two benchmark databases namely; the Extended Yale B database [23] and the AR face database [24] are considered. The Extended Yale B database contains 2,414 frontal-face images of 38 individuals. The cropped and normalized 192×168 face images are captured under extremely various lighting conditions. We randomly select half of the images for training and the other half for testing with respect to each individual. The AR face database was created by Martinez and Benavente that consists of over 4,000 frontal images for 126 individuals and includes more facial variations, such as various facial expressions, illumination conditions, and occlusions comparing to the Extended Yale B database. The facial images of most people are taken in two sessions. Each session includes 13 color images and 120 individuals (65 men and 55 women) participated in both sessions. A subset of the dataset which consists of 50 male subjects and 50 female subjects is chosen for the experiments of the proposed method. For each session, only 7 images with illumination and facial expressions change are considered. The seven images from session 1 for training and the other seven images from session 2 for testing.

It should be noted that due to the long training time of CNN model, we don't apply any pre-training process in our experiment. Only the training images in the benchmark database are used to train the proposed CNN model. The following sections evaluate the proposed system in terms of three measures: (1) different dimensions of extracted feature, (2) the comparison between the softmax classifier and the sparse representation classifier with deep features, (3) different feature extraction methods with the sparse representation classifier.

3.2. Results of Extended Yale B database

The architecture of CNN requires a fixed input size for its input images, all the training face image are resized to 48×42 to match the input size of the proposed CNN architecture. The training and testing process repeat five times with randomly selected samples, and the final recognition result is computed by averaging the recognition rate of each testing results. The strategy of training the proposed facial recognition system via randomly chosen samples ensures that the performance does not depend on any special selection of the training data. Furthermore, the proposed method also evaluates the classification performance of SRC with different dimensions of CNN feature, which is chosen to be 30, 56, 120, 160 and 504, as described in [20]. The evaluation results demonstrated in Table 1 shows that SRC with deep convolutional feature has better performance than the softmax classifier in any feature dimension. The combination of SRC and CNN achieve a maximum recognition rate of 99.17% for 504-dimension feature spaces.

Further, the proposed method evaluates the classification performance of SRC with various feature extraction algorithm. John Wright [20] has claimed that the choice of feature space is no longer a critical issue for recovering the sparse representation of target face image, and the classification performance of SRC mainly depends on the dimension of feature space. However, SRC attempts to recover a sparse representation of the test samples via finding a linear combination of the training samples, which means that the computation time of SRC becomes very long by selecting a large feature dimension. In addition, SRC doesn’t perform well in the small feature dimension case due to the lack of information gathering from the target face image. Thus, to circumvent these problems, the choice of feature extraction method is still a concern. Fig 6 displays the recognition performance for the SRC in conjunction with several different feature extraction methods including Eigenfaces [4], Laplacianfaces [26], Fisherfaces [27], randomfaces [28], and down-sampled images. It can be observed that the Fisherfaces is not available for all feature dimensions as discussed in [20]. Obviously, CNN achieves great classification result than other feature extraction method in the case of low feature dimension. The proposed CNN architecture outperforms other methods by at least 1%, which can be seen in Table 2.

Table 1 Recognition rate comparison of soft-max and SRC with deep features in Extended Yale B database

	Dimensions				
Method	30	54	130	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 2 Performance of the proposed CNN architecture with other methods 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 + SRC	504	99.17%
Deep Learning + softmax	504	98.92%

3.3. Results of AR database

Similar to the case of the Extended Yale B database, the face images are cropped and resized to 50×45 in order to match the input size of the CNN architecture with randomly chosen the training and testing samples in the evaluation process. The feature space dimensions are selected to 30, 54, 130, 160 and 540, which are slightly different from the experiment of Extended Yale B database. The classification results of SRC in different feature dimensions are summarized in Table 3. The recognition rate of the proposed approach is over 90% with 130-dimention features. Fig 7 shows the result of SRC with the feature extraction methods described in the previous section. It can be seen that CNN feature outperforms other methods. Although Fisherface has better recognition rate than the proposed method in lower dimension case, its maximum recognition rate is 92.17% which is lower than the proposed technique 95.85%, as demonstrated in Table 4. In the 540 dimension, the maximum rate except deep features is 94.7 percent which is achieved by the random faces. CNN outperforms other methods by at least 1% that is same as the evaluation result of Extended Yale B database.

3.4. Comparison of Different Architectures of CNN

This section shows the experimental results on three different CNN architectures. These CNN architectures are consisted of various numbers of convolutional layers and feature maps, which are described in Fig 8. The proposed method evaluates these models on the Extended Yale B database, and the result is summarized in Table 5. Noticeably, the two-layer model displays strong performance. Although the CNN model described in Fig 8(c) uses more convolutional and pooling layer than the other two models, it doesn’t have better recognition rate than the case of Fig 8(b).

Table 3. Recognition rate comparison of soft-max and SRC with deep features in 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 4. Performance comparison of the proposed CNN with others methods 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 +	504	95.85%

SRC		
Deep Learning + softmax	504	94.42%

4. Conclusions and Future Works

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.

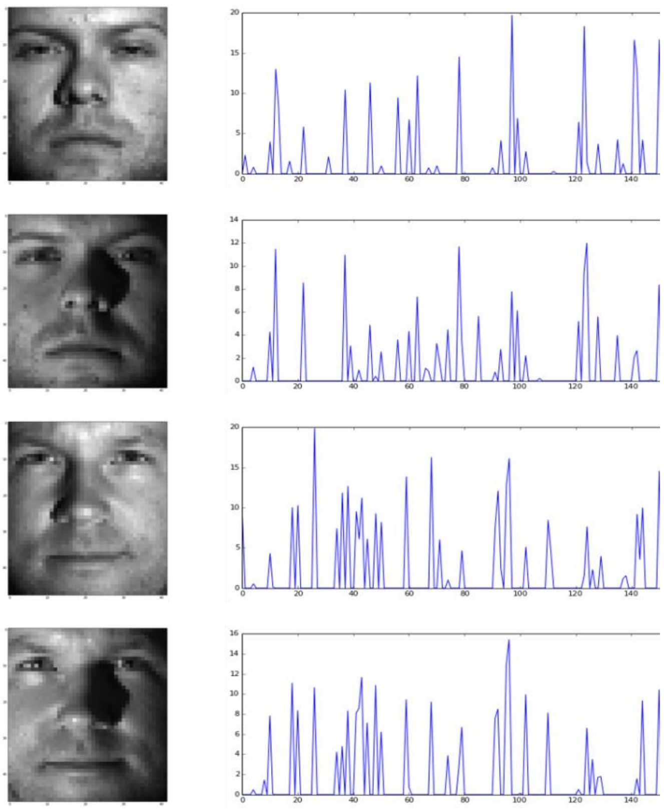


Fig. 5. Sparsity and selectivity of deep neural activations

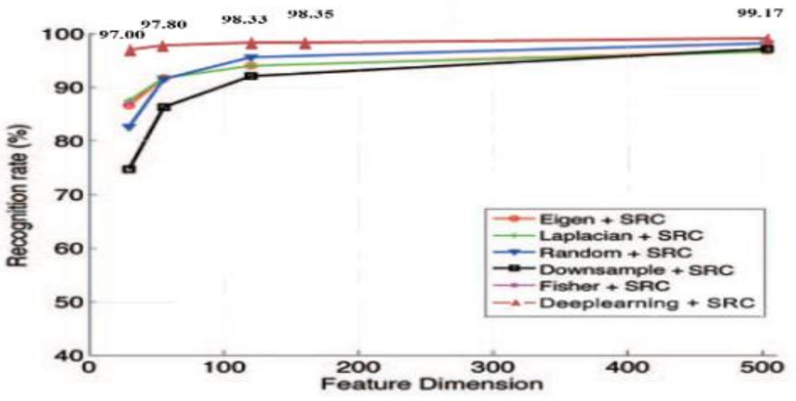


Fig. 6. Recognition rate comparison by using various feature extraction methods with SRC in the Extended Yale B database

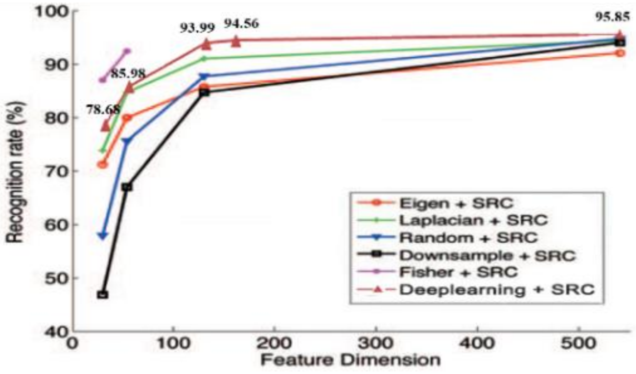
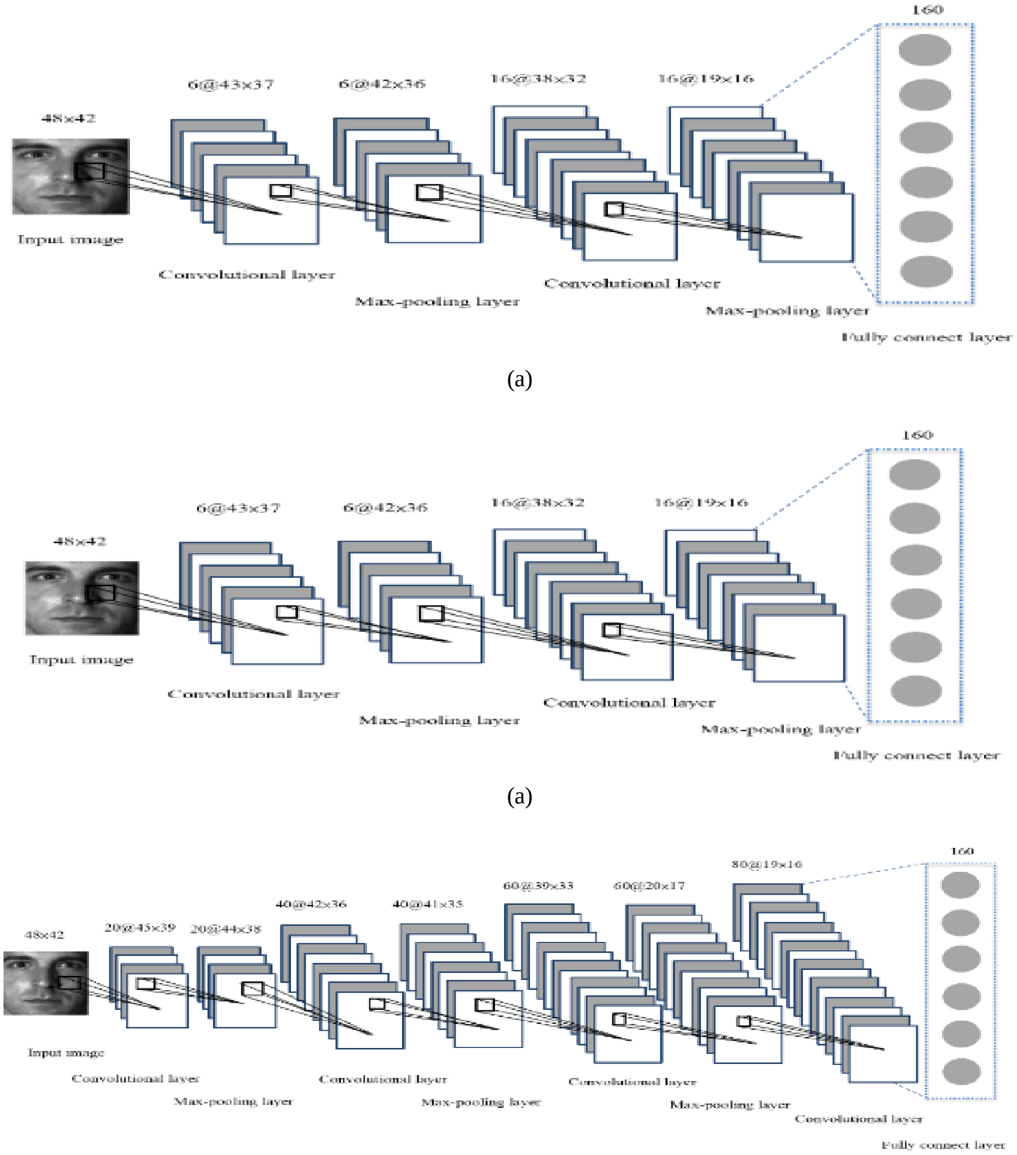


Fig. 7. Recognition rate comparison by using different feature extraction methods with SRC in the AR database



(a)

(a)

(a)

Fig. 8. Different CNN architectures for comparison. a) 6-16-160 (b) 15-45-160 (c) 20-40-60-80-160

REFERENCES

1. W. Zhao, R. Chellappa, P. J. Phillips, and A.Rosenfeld, “Face Recognition: A Literature Survey”, *ACM Computing Surveys*, vol. 35, no. 4, 2003, pp. 399-458

2. W. Zhao and R. Chellappa, “Image-based Face Recognition: Issues and Methods”, *Optical Engineering-New York-Marcel Dekker Incorporated*, vol. 78, 2002, pp. 375-402

3. R. M. Ebied, “Feature Extraction using PCA and Kernel-PCA for Face Recognition”, *8th International Conference on Informatics and Systems*, vol. 8,2012, pp. 72-77

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, 2006, pp. 2037-2041

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*, 2004, 60(2):91-110

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 (6), 2012, pp. 82-97

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*, 2014, pp. 1701-1708

13. G. B. Huang, H. Lee, and E. Learned-Miller, “Learning Hierarchical Representations for Face Verification with Convolutional Deep Belief Networks”, *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2518-2525

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*, 2014, pp. 1891-1898

15. Y. Sun, X. Wang, and X. Tang, “Deep Learning Face Representation by Joint Identification-Verification”, *Advances in Neural Information Processing Systems*, 2014, pp. 1988-1996

16. G. E. Hinton, S. Osindero, and Y. W. Teh. “A fast learning algorithm for deep belief nets”. *Neural Computation*, 2006, 18(7):1527-1554

17. R. Salakhutdinov, A. Mnih, and G. E. Hinton. “Restricted boltzmann machines for collaborative filtering”, *ICML*, 2007, pp. 791-798

18. R. Salakhutdinov and G. E. Hinton. “Deep Boltzmann machines”, *AISTATS*, 2009, pp. 448-455

19. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition”, *Proceedings of the IEEE*, 1998, 86(11):2278-2324

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, 2009, pp. 210-227

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”. *arXiv preprint*, 2014, arXiv:1408.5093

22. G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors”, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2012

23. K. Lee, J. Ho, and D. Kriegman. “Acquiring linear subspaces for face recognition under variable lighting”, *IEEE TPAMI*, 2005, 27(5):684-698

24. A. Martinez and R. Benavente, “The AR Face Database”, *CVC Technical Report 24*, 1998

25. Q. Le, M. Ranzato, R. Monga, M. Devin, K. Chen, G. Corrado, J. Dean, and A. Ng. “Building high-level features using large scale unsupervised learning”, *ICML*, 2012

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, 2005, pp. 328-340

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, 1997, pp. 711-720

28. S. Kaski, “Dimensionality Reduction by Computation for Clustering,” *IEEE International Joint Conference on Neural Networks Proceedings*, vol. 1, 1998, pp. 4-9