Autoencoder versus PCA in face recognition

Krzysztof Siwek
Institute of Theory of Electrical Engineering, Measurement and Information Systems, Warsaw, Poland

Stanislaw Osowski
Institute of Theory of Electrical Engineering, Measurement and Information Systems, Warsaw University of Technology and Electronic Faculty, Military University of Technology, Warsaw, Poland

***Abstract*—The paper presents the comparative analysis of the computer systems for face recognition. Autoencoder, the typical representative of deep learning is compared with the classical PCA transformation. Both, autoencoder and PCA serve as the tools for feature generation and selection. However, the important difference is the nonlinearity and multilayer structure applied in autoencoder. Final task of recognition is done by the support vector machine or softmax circuit. The numerical results performed on the multiclass base of faces have shown superiority of autoencoding principle, especially when the number of recognized classes is very high.**

Keywords—deep learning; autoencoder;PCA; face recognition

I. INTRODUCTION

The problem of face recognition is very important in non-invasive and non-expensive way of person identification and verification. Many different methods used for solving this problem have been proposed and tested [1,2,3,4,5]. The important point is extraction of the numerical descriptors of the image forming the diagnostic features used as the input attributes to the final classifier, responsible for face recognition.

The most often used is the linear principal component analysis (PCA) applied either directly or indirectly in the recognition system. This paper will investigate and compare the new approach to this problem by applying the deep learning solution in the form of an autoencoder [6]. Autoencoder is a nonlinear multilayer structure extracting the features of the image, step by step, in an auto-associative unsupervised way. The succeeding hidden layers are responsible for the more and more compressed coding of the analyzed image. The last layer is the reduced size characterization of the input image, representing the diagnostic features, that are used as the input attributes to the final recognizing system. The image classification is built on the basis of the support vector machine (SVM) or softmax classifier.

The deep learning approach will be compared to the classical PCA, relying on the linear transformation of the vector **x** into other vector **y** of the reduced dimension. The elements of vector **y** serve as the input signals to the classifier.

Both approaches will be applied to the set of images representing faces of 51 different persons, photographed in various positions of face and in different lighting conditions. The results of investigations in the form of accuracy of the person recognition have shown superiority of deep learning.

II. AUTOENCODER FOR FEATURE EXTRACTION

Autoencoder is the neural network of multilayer structure which tries to map the input signals into their smaller size representation, that allows reconstructing their original values with an acceptable accuracy. In other words, it tries to learn the parameters of the structure allowing its output vector to be approximately equal input. Learning process is performed layer after layer. In the first step the structure containing only one hidden layer is trained to obtain good reconstruction of input signals on the output size. It is an auto-associative way of learning [6]. An example of this first step is shown in Fig. 1.

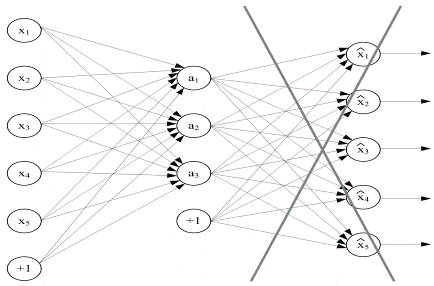

Fig. 1. The first step of learning one hidden layer autoencoder.

After finishing this step the output layer is removed and hidden layer serves as an input in the next steps in forming the following layers. This process is illustrated in Fig. 2.

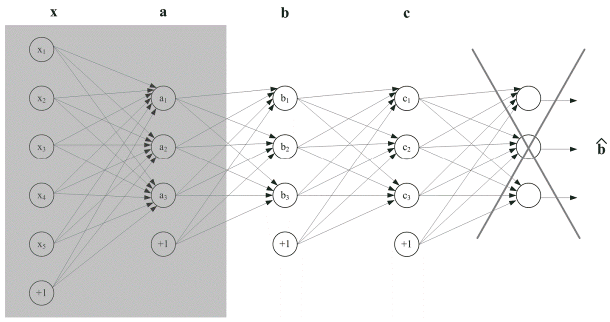

Fig. 2 The illustration of the process of forming multilayer structure of autoencoder.

The succeeding layers: a, b, c, etc. are the hidden layers formed in the learning process in the same way as shown in Fig. 1. The signals of succeeding layers represent the diagnostic features of the analyzed data, which are subsequently extracted in an unsupervised way from the input vector. The last layer (in our case c) represents signals that will be used as the final set of the diagnostic features serving as the input attributes to the output classifier, which is responsible for recognition of faces.

III. PCA FOR FEATURE EXTRACTION

PCA is an orthogonal transformation of data converting the observations represented by vectors **x** into a set of vectors **y** of smaller dimension. The elements of vector **y** form the set of linearly uncorrelated variables, called principal components [3,4,5]. This transformation is defined by the linear relation **y=Wx**, in which matrix **W** is defined on the basis of the cross covariance of the input data. The first principal component represents the largest variance of the data and is equal to the largest eigenvalue of the covariance matrix. The next components of this vector represent the next variances, all arranged in the decreasing order with respect to eigenvalues of the covariance matrix. The following directions defined by the rows of the transformation matrix **W** are orthogonal to the preceding ones. The PCA components arranged in the order according to their decreasing variance represent the most important statistical information contained in the original set of data.

Limiting the number of principal components to the most important eigenvalues the reduction of dimensionality of data can be achieved. The signals of the hidden layer are formed as the weighted sum of elements of the original input vector. These signals create the patterns which are characteristic for the particular class of data and represent coding of the image. They will serve as the input attributes to the output classifier.

IV. NUMERICAL EXPERIMENTS OF FACE RECOGNITION

A. Data base

The data base used in numerical experiments contained the faces of 51 persons, which are treated as 51 classes. Each class was composed of 20 photographs of the same person made in different poses at varying lighting conditions. The size of the original images in all cases was the same and equal 25×25.

Some chosen set of typical examples of original images in visual imageries, which are subject to recognition, are presented in Fig. 3. As we can see the same person is photographed in different poses and at varying illumination. Some images show the face with glasses and some without glasses. The faces are shown in different scale, representing either full face or only some part of it. There are evident differences among different representatives of the same family of persons.

The input images applied to the encoding or PCA transformation stages were vectored. It means, that each 25×25 image is represented by the vector containing 625 elements. These vectors were subject to further processing according to the applied method (either PCA or encoding procedure).

Fig. 3 The exemplary 5 representatives of faces for eight different persons taking part in experiments.

B. Results of Experiments

In these experiments we have applied autoencoder of 2 nonlinear layers of sigmoidal nonlinearity. The structure of autoencoder in recognition of the face images is presented in Fig. 4.

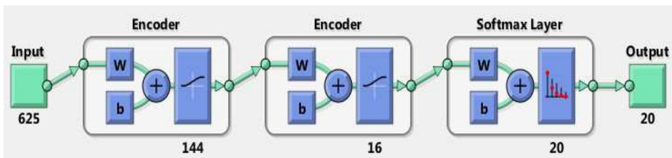

Fig. 4 The structure of the applied deep net autoencoder in classification system.

Fig. 5 illustrates the representation of the 16 features used as the input signal to the softmax classifiers [6] in the final layer of autoencoder. All of them correspond to eight different images of the same person. The experiments have been conducted using Matlab implementation [9]. For comparison the PCA based feature representation supplying SVM as the classifier has been also investigated.

The key to the success in face recognition is insensitivity of the diagnostic features to different poses and lighting conditions of the face. The autoencoding principle to feature generation is very efficient in this respect. This is well illustrated in Fig. 5. It shows the distribution of the diagnostic features created by the autoencoder and the linear PCA for the same set of 8 images belonging to the same class. In both cases the number of features was the same and equal 16. The features of autoencoder were created by 2-layer nonlinear transformation with sigmoidal function (circuit structure of Fig.4). In the case of PCA 1-layer linear transformation was used (circuit structure: 625-16).

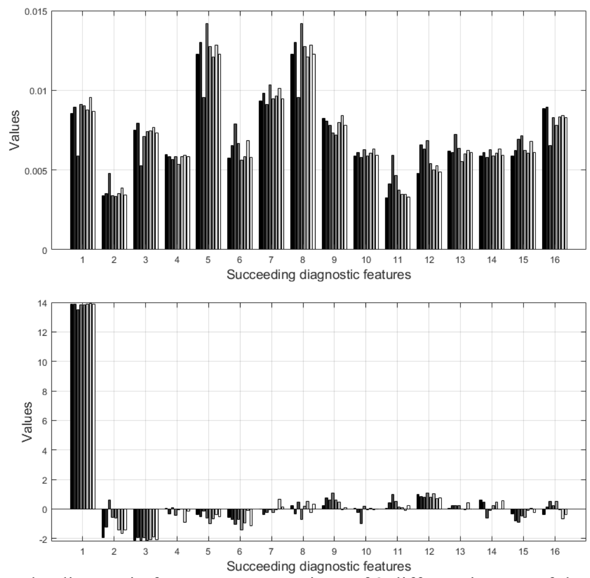

Fig. 5 The diagnostic feature representations of 8 different images of the same person: upper row – the distribution of features created by an autoencoder, bottom row – the corresponding PCA feature distribution.

We have calculated the average value and standard deviation of these features created by using both methods and then defined the ratio of standard deviation (std) to their mean. The lower the ratio std/mean for the feature the better quality of such feature.

TABLE I THE STANDARD DEVIATION AND THE RATIO STD/MEAN OF THE DIAGNOSTIC FEATURES EXTRACTED FOR THE FACE IMAGES BELONGING TO THE SAME CLASS.

	AUTOENCODER		PCA	
	std	std/mean	std	std/mean
feature 1	0.0011	0.1320	0.1276	0.0092
feature 2	0.0005	0.1343	0.8140	-0.7896
feature 3	0.0008	0.1129	0.0979	-0.0480
feature 4	0.0002	0.0324	0.3285	-1.5064
feature 5	0.0013	0.1063	0.2457	-0.4693
feature 6	0.0008	0.1215	0.3919	-0.4680
feature 7	0.0004	0.0430	0.3182	-17.2887
feature 8	0.0013	0.1063	0.4376	6.7499
feature 9	0.0004	0.0531	0.3620	0.7639
feature 10	0.0002	0.0327	0.3626	-2.5299
feature 11	0.0009	0.2261	0.3420	1.2016
feature 12	0.0008	0.1445	0.1417	0.1612
feature 13	0.0005	0.0762	0.1633	1.2055
feature 14	0.0002	0.0327	0.4240	2.1052
feature 15	0.0005	0.0729	0.3310	-0.7819
feature 16	0.0008	0.0919	0.4435	-35.9751

The statistical results of std and std/mean concerning the diagnostic features of Fig. 5 for one chosen face family are presented in Table I.

The advantage of autoencoding principle over PCA is evident. The standard deviation of the feature values within the same class of faces is much smaller and more balanced for all features. Very high discrepancy is visible also in the rate std/mean. This rate is small and well balanced for all features in the case of autoencoder. It is not true in the case of PCA application.

Final numerical experiments have been performed to check the class recognition ability of the developed diagnostic features. The support vector machine (SVM) with Gaussian kernel has been used as the classifier in the case of PCA. The width of Gaussian function and regularization constant C have been adjusted using additional experiments on the validation data set. Their optimal values found in experiments were as follows C=1000 and σ=0.8. In the case of autoencoder softmax classifier was used [7]. The number of diagnostic features in both solutions was the same and equal 16.

Fig. 6 The face images representing typical results after different angle rotation and of sharpening procedure applied to chosen set of faces.

The experiments have been repeated ten times at random choice of learning and testing data. 60% of samples have been used in learning and 40% in testing. Due to relatively small number of faces in data base we have increased their population by including rotation (the angle 45^o, $\pm 90^o$ and 180^o) and including sharpening operation of the images (some sort of the noise). The typical results of such transformations are presented in Fig. 6.

The statistical results of testing in the form of average error and standard deviation of face recognition for 20 and 51 classes are presented in Table II. These data correspond to the samples not taking part in learning. The results of PCA and autoencoder applications have been obtained for the same data base [4] and represent the average values of errors followed by standard deviations obtained in all runs.

TABLE II THE AVERAGE ERROR OF FACE RECOGNITION FOR 20 AND 51 CLASSES OF DATA (PERSONS) OBTAINED AT APPLICATION OF AUTOENCODER AND PCA PREPROCESSING.

Number of classes	Autoencoder	PCA
20	3.23%±0.99%	3.20%±1.55%
51	9.53%±2.10%	13.5%±1.53%

This is well seen for recognition of 51 classes, where the autoencoder achieved the error rate 9.53% while PCA 13.5%. Table III presents the confusion matrix for one run of the autoencoder solution in recognition of 20 classes of data. We can see only scarce off-diagonal elements which are different from zero.

TABLE III THE CONFUSION MATRIX OF 20 CLASSES IN FACE RECOGNION AT APPLICATION OF AUTOENCODER.

25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	31	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1
0	0	1	0	34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	30	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	27	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0	2	0	0
0	0	0	0	0	0	0	0	0	14	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	1	0	36	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0	0
0	0	2	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	1	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	37	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18

At small number of classes both methods of feature generation are similar, although PCA was slightly better. However, when the number of classes was increased the autoencoding principle was unbeatable.

V. CONCLUSIONS

The paper compares two methods of diagnostic feature definition in face recognition. They include the nonlinear autoencoding principle and linear PCA application. Both belong to the unsupervised approaches. The main advantage of PCA is its simplicity and low cost of processing. The encoder is a multilayer structure applying nonlinearity in signal processing. Its main advantage is relatively higher insensitivity to the natural variability in the images forming the same class and significant resistance to the noise present in the images forming the data base.

The results of numerical experiments of face recognition using SVM and softmax classifiers have proved the superiority of an autoencoding principle of feature extraction. The variability of the values of diagnostic features within the same class is significantly smaller for autoencoding way of feature extraction. This leads to higher efficiency of autoencoder generated features in face recognition. As a result the average error of classification is lower. The advantage of an autoencoder is especially well seen when the number of recognized classes is very high and at the same time the population of data is sufficiently large.

REFERENCES

[1] P. Belhumeur, J. Hespanha and D. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, *IEEE Trans. PAMI*, vol. 19, pp. 711–720, 1997.

[2] P. J. Phillips, H. Moon, S. Rizvi and P. Rauss, The FERET evaluation methodology for face recognition algorithms, *IEEE Trans. PAMI*, 2000, vol. 22, No 10, pp. 1090-1104.

[3] S. Kong, J. Heo, B. R. Abidi, J. Paik and M. A. Abidi, Recent advances in visual and infrared face recognition – a review, *Comput. Vision Image Understanding*, vol. 97, pp. 103-135, 2005.

[4] K. Siwek and S. Osowski, Comparison of methods of feature generation for face recognition, *Przegląd Elektrotechniczny*, vol. 90, pp. 206-209, 2014.

[5] K. Siwek, S. Osowski and J. Jakubowski, Face recognition in visible and infra-red imagery - comparison of methods, *Conf. Proc. "Computational Problems of Electrical Engineering*", Zilina, 2014, pp. 1-3

[6] I. Goodfellow, Y. Bengio and A Courville, *Deep learning*, MIT Press, 2016.

[7] R. Gonzalez and R. Woods, *Digital image processing,* New Jersey: Prentice Hall, 2008.

[8] B. Schölkopf and A. Smola, *Learning with kernels,* Cambridge, MIT Press, MA, 2002.

[9] *Matlab user manual,* MathWorks, Natick, USA, 2017a.