A Semi-supervised Pipelined Deep Learner for an Adaptive and Efficient Biometric System

Abhijit Das, *Student Member, IEEE,*

Supervisor: Javier Ortega-Garcia, *Senior Member, IEEE* and Aythami M. Moreno, *Member, IEEE*

Abstract—**This work proposes a semi-supervised pipelined deep learner for efficient and adaptive biometric authentication. A semi-supervises based schema namely Error Driven Incremental Model (EDIM) is used for the online learning of the semi-supervised learner. Deep generative network, based on Convolutional Neural Network has been used for the deep learning phase. Coupled with back propagation based Markov Chaining (MC) to take care of the error calculation for the neural net for judging the incremental learning. In order to access the effectiveness of the conceived learning mechanism for adaptive biometrics (biometric characteristics can change either temporarily or permanently, perhaps due to ageing, diseases or treatment for diseases or maybe accidents, adaptability of the system with such changes is termed as adaptive biometrics), we have considered the multimodal ocular trait (combination of iris and sclera) to assess its applicability in biometric domain. Initial investigation using an in-house data solicits appreciable result of the learning mechanism in biometric scenario.**

Index Terms— **Convolutional Neural Network, biometrics, Adaptive biometric, Markov Chaining, Error Driven Incremental Model, Semi-supervised learning.**

I. INTRODUCTION

The biometric is a scientific field of research which establishes identity of individual employing physiological or behavioural characteristics. Due to availability of inexpensive biometric sensors, increasing computing power equipped processor, and low-cost memory devices, it is expected that biometric technology will have broader application in the upcoming near future. Therefore broader scope of future research is required to solve upcoming challenges and also to push the scope of boundaries.

This work was done while the author visited Biometric Recognition Group - ATVS, at Universidad Autonoma de Madrid, Madrid, Spain.

A. Das is a Ph. D Research Fellow from, Institute for Integrated and Intelligent Systems, Griffith University, Queensland, Australia (email-abhijit.das@griffithuni.edu.au).

J. Ortega-Garcia, is a Full Professor from, Universidad Autonoma de Madrid, Madrid, Spain.

A. M. Moreno, is a Post Doctorate Fellow, from, Universidad Autonoma de Madrid, Madrid, Spain.

Among the scope of biometrics that requires real attention through research is the adaptability of biometric systems [1]. Adaptability in the context of biometrics can be defined as the ability of biometric traits to adjust with the change in information acquired in terms of environment, ageing and posture [2]. Change in biometric traits or variation in the traits is a key challenge for the biometrics community as it can lead to misidentification [12, 13]. The main reasons for misidentification or rejection of the correct individual by biometric systems are scarcity of training samples, presence of substantial intra-class variation during testing and lack of standardization of the data acquisition environment [3].

Moreover biometric characteristics can change either temporarily or permanently, perhaps due to ageing, diseases or treatment for diseases or maybe accidents. In order to handle these issues, it would be preferable to have a biometric system which adapts well to the changing problem [5, 8] rather inter class variability. These drawbacks of the present biometric systems have stimulated the interest of researchers in this domain. Recent developments in the adaptive system research area have opened up a new research field. The ideal case of an adaptive biometric system is expected to handle the intra-class variations which changes with time (in many cases). These changes can happen for various reasons like ageing, variations in pose and lack of standardization of data acquisition rules [4]. The advantages of such a system are: learner need not get trained from scratch every time new data is available (never ending learning paradigm) and no need to store old data. This aspect of learning will significantly reduce the maintenance cost of biometric system. These are the characteristics which makes this research area so attractive and suitable for real-time scenario.

It can be inferred from the literature that an adaptive biometric system can operate in a supervised or semi-supervised mode [2, 6, and 7] in which biometric samples are labelled and updated manually and automatically depending on the mode. The supervised method represents the best-case performance as all the available positive (genuine) samples are used for adaptation. However, manual intervention makes this process time consuming, expensive and also diminishes its applicability in real life. Therefore it is generally infeasible to automatically update references regularly by considering a semi-supervised learning tool. The feasibility of such adaptive biometric system are been proposed by several work in the literature. Despite these investigations, several gaps reside in this context. One of the main gaps is the low accuracy achieved by the proposed adaptation technique [9, 10 and 11]. Therefore considering the above illustrated issued and after reviewing the literature that limited work has been undertaken on adaptive biometrics it can be easily assume that this research needs more systematic and structured research attention. If we perceive biometrics as a research field with broad economic and scientific impact, then designing efficient algorithms and systems and breaking these highlighted challenges of adaptive biometric systems will require a multidisciplinary effort in signal processing, pattern recognition, machine learning, sensor design, embedded systems, and information fusion. Recent advances in machine learning have seen widespread development of algorithms in specific area of deep learning technique [15-16]. Deep learning methods have been extremely successful recently in the field of computer vision and machine intelligence, in particular in the areas of speech recognition, object recognition and language modelling. Deep representations are information presentations at multiple levels of abstraction, of increasing non-linearity.

Assuming deep learning as a strong approach and also an up-and-coming and promising machine learning tool, it can be a solution to mitigate the problem of low accuracy of adaptive biometric systems. On the other hand it will be also interesting to note the behaviour of this machine learning tool in this new paradigm of biometric. Considering all this above fact and mitigate the problem of low accuracy in adaptive biometric system this research proposal is conceived.

In general the success of machine learning algorithms generally depends on data representation. It can be hypothesized that variation in data can entangle and hide the more or less different explanatory factors that are employed to classify these information. Although specific domain knowledge can be used to help design representations, even learning with generic priors can also be used to quest the need for these artificial intelligence. Furthermore the motivating for designing more powerful representation of learning algorithms can another way to solve this gap. Therefore, various recent advances in learning technique are proposed. One such example is training by deep supervised networks. Recent work in the area of unsupervised feature learning and deep learning of generative models [17-19] focusing on advances in understanding the probabilistic and geometric (manifold) aspects of regularized auto-encoders can be found in the recent literature. This motivates longer-term unanswered questions about the appropriate objectives for learning good representations, for computing representations [20] (i.e., inference), and the geometrical connections between representation learning, density estimation and manifold learning is solved by deep learning. It can also be assume that the supervised and semi-supervised deep learning may boost the performances of challenging non-ideal data representations like adaptive biometric system [21 and 22]. In general face verification and cross-age face recognition are already investigated in this aspect of research of deep learning.

Due to the aforementioned discussed facts for personal identification a new deep learning pipeline network is proposed as new architecture for deep learning. Then a new type of deep features, which is robust to image local translation and scaling, is proposed face analysis parsing task, namely parsing the human image into semantic regions, like hair, glasses, face, upper-body clothes, bag, etc. [23] is proposed. Further these local semantic features can be used for robust human re-identification task. It can also be observed that de-noising auto-encoders and their deep recurrent and stochastic generalization (called Deep Generative Networks or GSNs) can be associated with a Markov chain (MC) whose stationary distribution is a consistent estimator of the data generating distribution. This circumvents is one of the challenges of probabilistic models (especially deep ones), i.e., the need for approximate inference and MC in the middle of the training loop and GSNs can actually be trained by back-prorogation algorithm.

Therefore this can be easily assumed from the above mention fact that this deep learning pipelined tool can be employed as online learner to efficiently classify and outbreak pitfall of low accuracy rate in adaptive biometrics. Therefore in this research as way foreword a pipeline deep learner based semi-supervised biometric algorithm is proposed.

The organization of the rest of the manuscript is: in section II proposed learner is described, in III the experimental setup are enumerated and in IV the conclusion of the work is drawn.

II. PROPOSED ONLINE DEEP LEARNER

This section is divided into two sub section. First subsection describes the proposed deep learner for ocular trait and the second subsection proposes the online method of the deep learner.

A. Proposed deep architecture for ocular trait

Current approaches to object recognition make essential use of machine learning methods. To improve their performance, we can collect larger datasets, train more powerful models, and use better techniques for preventing overfitting. Until recently, datasets of labelled images were relatively small. Simple recognition tasks can be solved quite well with datasets of this size, especially if they are augmented with label-preserving transformations. For example, the current best error rate on the MNIST digit-recognition task have achieved high recognition accuracy of (<0.3%). This overwhelming success of pattern recognition was possible due to emerging of Deep architecture based Convolutional Neural Network (CNN).

CNN is a Neural Network (NN) variant that consists of a number of convolutional layers alternating with subsampling layers and ends with one or more fully connected layers in the standard multilayer perceptron (MLP). CNN combines image feature extraction and classifier in one trainable module. It accepts a two dimensional (2-D) raw image with minimal pre-processing and retains the 2-D topology throughout its training. Classification is performed during training, and at the end of training, the final weights obtained behave as a feature extractor to classify the query input sample. CNN have been applied to various case studies of biometrics and pattern recognition such as face recognition [24], handwriting recognition [25], license plate recognition [26] and several other applications.

A well-known CNN architecture called LeNet-5 was proposed by [24]. The LeNet-5 CNN, which was first applied in a handwriting recognition problem, consisted of seven layers performing convolution alternated with subsampling operations. The first convolutional layer, which convolves the input with a convolution kernel, essentially acts an edge detector that extracts salient features of the input samples. The kernel, which is of size 5×5 in this case, consists of weighting coefficients that creates blurring (low pass filter), sharpening (high pass filter) or edge enhancement effect. The convolution process is performed by moving a flipped kernel through the images and the resulting output is placed as a new pixel of a feature map at the succeeding layer. The second layer performs subsampling, that is, a local averaging on a non-overlapping small window size of 2×2 in this case. This operation reduces the resolution of the feature maps from the previous layer, 7 essentially adding robustness against small distortions in translation, rotation and scaling. The final two layers of this architecture are MLPs that act as classifiers.

The CNN architecture proposed in this paper is based on the work by [27], in which convolution and subsampling layers are fused into one layer. As shown in Figure 1, this fusion significantly reduces the total number of layers in the CNN from seven to four. Simard's CNN was also applied in handwriting recognition, and a higher accuracy than that of LeNet-5 was achieved.

The proposed CNN architecture is based on the design that is consisted of fused convolutional and subsampling layers as first proposed by Simard et al. in [27]. The method dispenses away with complex operations, such as momentum, weight decay, structure-dependent learning rates, averaging layers, extra padding on the input, tangent prop and even the fine tuning of the architecture, which are otherwise needed in the conventional CNN solutions. The proposed deep CNN is shown in Figure 1.

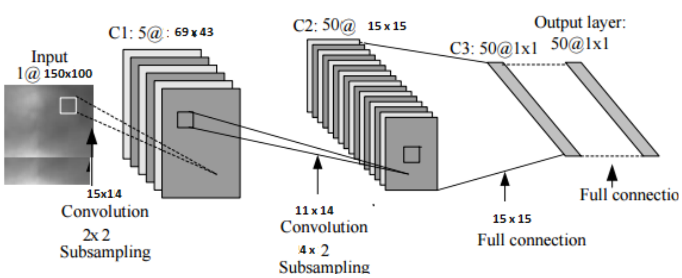


Fig: 1 The proposed CNN based deep architecture

The proposed CNN consists of four layers (namely C1, C2, C3 and C4 convolutional layers), and this does not include the input layer. We refer to this architecture as the 5-50-50 model, implying that there are 5, 50 and 50 feature maps in layers C1, C2, and C3 respectively. Layer 4 (i.e. the output layer) is fixed at 50 neurons since the target number of categories to classify is 50 subjects; hence this information is implicit, and therefore unnecessary to include in the model referencing name. Cross-validation technique, a popular empirical method for estimating the generalization accuracy of a NN is applied to find the best model configuration, in which we determine the best number of feature maps and the best connection schemes to implement.

In our CNN design, the input image size is set to 150×100 (resized for the ease of the computation) pixels and a 15×14convolution kernel size is used for C1 and a 11×14convolution kernel size is used for C2. For subsampling operations, the kernel sizes are different for different layers - the kernel size in subsampling in layer C1 is 2×2, and in C2 it is fixed at 4×2. The selection of convolution and subsampling sizes determine the sizes of the feature maps at layer C3 for our purpose it is 15 x15. We are targeting for 1×1 neurons at the final two layers of MLP classifier. Each layer performs feature extraction tasks and dimensionality reduction process at the same time.

B. Proposed online incremental model

A crude definition of incremental learning is that learning is a continuous process with new data. Prior work has addressed two major scenarios, out of which the first one is relevant to this study. The first is concept drift in the training dataflow, and therefore the classifier learns in a non-stationary environment. The second is when there are existing classifiers that are related to the new classes to be learned. In order to solve such problem deep learner can be the solution.

Supervised learning using deep CNN has shown its promise in large-scale image classification task. As a building block, it is now well positioned to be part of a larger system that tackles real-life multimedia tasks. An unresolved issue is that such model is trained on a static snapshot of data. Instead, the training should be as continuous learning processes as new type of sample data may arrive form a same class which is very much different from the training sample used during enrollment. A system with capability to adjust to these changes is useful in practical scenarios, as it gradually expands its capacity to predict increasing number of new verity of sample. It is also our attempt to address the more fundamental issue: a good learning system must deal with new knowledge that it is exposed to, much as how human do.

We developed a training algorithm that grows a network incrementally. Each class are grouped according to similarities as cluster and self-organized into 1 level. The model of new sample arrival class is cloned from existing ones and can be trained in parallel. These models inherit features from existing ones and thus further speed up the learning.

In our incremental learning process, we assume there is a model (named M) representation for each classes that is already trained on n classes with s number of samples (M_{nk} represents model for class n having k number of train sample during enrollment). The goal of incremental learning is to evolve from M_{nk} to M_{nk+1}, to train M_n classes, in which a new sample arrived k+1 which is quite different from the other samples. Obviously, the model must not increase its capacity to accommodate more samples with new variation. The simplest way to grow is to make a transfer learning, to clone the model and update it by new weight. In other words, M_{nk} share the same structure as M_{nk+1} except it has W_{nk} should be updated. We can make the model possible by injecting units in the fully-connected layers or rather than having more feature maps in the convolutional layers. We can update the weight by calculating the error occurred for a particular sample and calculate the weight difference and update it by back propagation till the fully connected layer. Where E signifies error, w_{ij} the weight associated and alpha the learning rate.

$$\Delta w_{ij} = -\alpha \frac{\partial E}{\partial w_{ij}} \qquad (1)$$

The back propagation learning was performed by the transition operator of a MC as proposed in [28]. MC's stationary distribution estimates the data distribution. Therefore, the correction the weight due to the newly introduced data/ distribution affected by it can be talked.

The transition distribution of the Markov chain is conditional on the previous state, generally involving a small move, so this conditional distribution has fewer dominant modes, being unimodal in the limit of small moves. Thus, it is easier to learn because it is easier to approximate its partition function, more like learning to perform supervised function approximation, with gradients that can be obtained by backpropagation.

III. EXPERIMENTAL RESULTS AND DISCUSSION

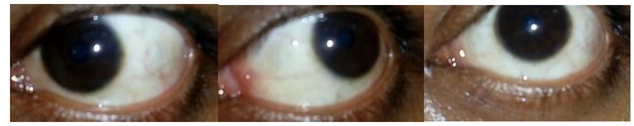
The experimental results and the dataset employed for the experimentation are explained in the below subsections.

A. Data Set

We have prepared an in-house populated from 25 individuals. Data was collected from both the eyes of each individual, so we have 50 different classes. Video of eye was captured form each individual, which was further scrutinized to get 1000 frames i.e. images for each eye. While providing the eye data the volunteers were asked to move there eye ball from left to right, right to left, to get different view angle of sclera pattern.

The individual comprise of both male and female and different skin color, few of them were wearing contact lens and images were taken in the different time in the day. The database contains blurred images and images with blinking eyes, closed eye and blurred eye images. High resolution images are provided in the database (300 dpi resolution and 750 x 500 dimensions). All the images are in JPEG format. Here for each individual image in different multi angle is considered. For each angle several images are considered. For each individual both left and right eye is utilized. For the ease ness of the experiment the images were resized to a dimension of 150x 100 pixels.

Different lighting conditions are considered during the image accusing. A Lenovo K3 note mobile camera and its lightening flash with a diffuse shade were used for image capturing.We have used different angle images and some of the sample images are shown below in Figure 2(a). A frame work of the image capturing is in figure 2(d).



(a) (b) (c)

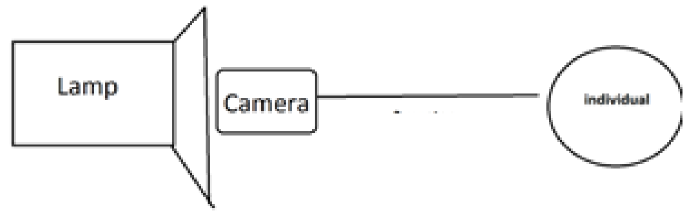


Figure 2: (a-c) Different type of eye images used in the experiments. (d) A framework of the image capturing

Images in the data set are different in quality. Some of them are not occluded having good quality of sclera regions. Images with medium quality and poor quality with respect to sclera region visibility are also present in the database. Some closed eye images were also taken for the experiments. The lighting condition was also considered. For example 1st half of the images were taken in a dark room so that the noise factors such as reflection, luminosity, and contrast were minimized. In the second half, the images were taken under natural illumination conditions with spontaneous user participation in order to introduce natural luminosity and add more noise factors than the first half. The database contains blurred images and images with blinking eyes as shown in Figure 3.

(a) (b)

Figure 3: Examples of closed and blurred eyes..

All the simulation experiments performed here were developed in Matlab 2013a on the Windows7 operating system along with a core I5 processor and 8GB RAM as the hardware configuration.

B. Image pre-processing

For the ease of the algorithm, the images were down sampled by image resizing techniques. The down sampling was undertaken for low resolution images also.

In order to get the segmented iris, Daugmans integro–differential method, was used to calculate the center of the iris. The iris image was cropped considering its radius length, along the iris center as shown in Figure 4a.

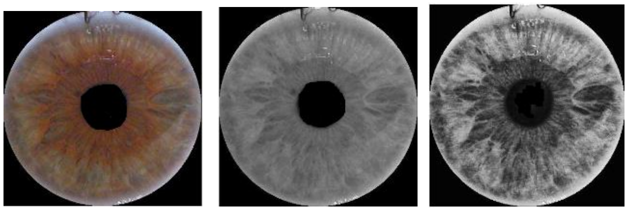
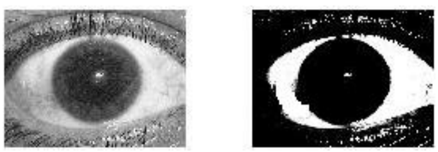


Figure 4: (a) Segmented iris image, (b) Red channel of the iris image, (c) Adaptive histogram equalization iris image

The patterns in the iris are not prominent, so in order to make them clearly visible, image enhancement is required. Adaptive histogram equalization is performed with a window size of 2 x 2 at a clip limit of 0.01, with full range and distribution, exponential to get the best result on the red channel of the iris image (as the iris patterns are most prominent in the red channel as shown in Figure 4b). The vessel structure becomes more prominent after the histogram equalization as shown in Figure 4c. The enhanced image is used further for feature extraction and identification.

Sclera segmentation was performed by the fuzzy c-means method.

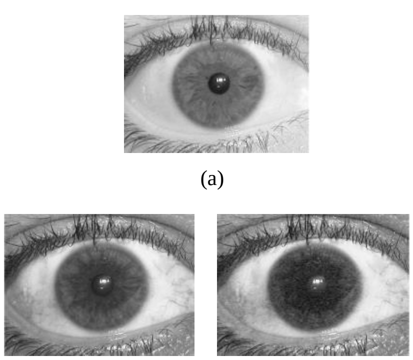


(a) (b)

Figure 5: (a) grey image of 1(a). (b) shows the Fuzzy C means-based sclera segmentation of (a).

The vessels in the sclera are not prominent, so in order to make them clearly visible, image enhancement is required. Adaptive histogram equalization was performed with a window size of 42 x 42 in the green channel of the sclera image (as the sclera vessel patterns are most prominent in the green channel, which is shown in Figure 6(b)) to make the vessel structure more prominent as shown in Figure 7(a) and performed.

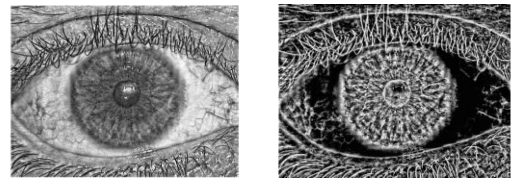
Furthermore, the Discrete Meyer wavelet was used to enhance the vessel patterns. A low pass reconstruction of the above-mentioned filter was used to enhance the image. Figure 7(a) shows the enhanced vessel image after applying the filter.



(a)

(c) (d)

Figure 6: (a) The red channel component, (b) The green channel component of , and (c) blue channel component of the original image,



(a) (b)

Figure 7: (a) Adaptive histogram equalization of the sclera image. (b) the vessel enhanced image after using a wavelet filter.

As proposed, the feature extraction approach uses images as an input therefore the image level fusion-based technique was used for combining both traits.

C. Result and discussion

The result and the discussion of the various experiment performed are summarized in this subsection. In the table 1 results achieved on the proposed schema is summarized.

Table 1: Result on the proposed schema.

Trait	**Accuracy (in %)**		
	50% of samples from each classes as training and 50% testing	**50% of samples from each classes as training, 10 % online learning and 40% testing**	**50% of samples from each classes as training, 20 % online learning and 30% testing**
Sclera	58.33	61.51	67.32
Iris	63.01	65.31	70.02
Combination of iris and sclera	66.55	68.23	71.78

The result obtained are far from the performances of the most competitive state-of-the-art approaches. However, it can be inferred from the above table that promising improvement is achieved in the incremental version in contrast to the general scenario. These solicit the successful implementation of the schema conceived for adaptive biometrics.

IV. CONCLUSIONS AND FUTURE SCOPE

This work proposes a semi-supervised pipelined deep learner for efficient and adaptive biometric. A semi-supervises based incremental learning model, Error Driven Incremental Model (EDIM) is used for the online leaning of the semi-supervised learner. Deep generative network, based on Convolutional Neural Network has been used for the deep learning phase, coupled with Markov Chaining (MC) to take care of the error calculation for the neural net for judging the incremental learning by back propagation technique. In order to access the effectiveness of the conceived learning mechanism for adaptability of biometrics (changes to biometric characteristics that can change either temporarily or permanently, perhaps due to ageing, diseases or treatment for diseases or maybe accidents, adaptability of the system with such changes is termed as adaptive biometrics), we have considered the multimodal ocular trait (combination of iris and sclera) to assess its applicability in biometric domain. Initial investigation using an in-house data solicits appreciable result of the learning mechanism in biometric scenario.

An initial result is achieved in the investigation so, future scope will focuses on fine tuning the parameter to get more effective accuracy.

ACKNOWLEDGMENT

This research has been partially supported by IEEE Computational Intelligence Society Graduate Student Research Grant 2015.

Reference

1. N. Poh, A. Rattani and F. Roli, Critical Analysis of Adaptive Biometric Systems, IET Biometrics, 1(4): 179 - 187, 2012.
2. A. Rattani, G. L. Marcialis, E. Granger, F. Roli, A Dual-staged Classification-Selection Approach for Automated Update of Biometric Templates, Intl Conf. on Pattern Recognition (ICPR), Japan: 2972 - 2975, 2012.
3. A. Rattani,: Adaptive biometric system based on template update procedures. PhD thesis, University of Cagliari, Italy, 2010
4. A. Rattani, B. Freni, G. L. Marcialis, F. Roli: Template update methods in adaptive biometric systems: A critical review. Proc. Third Int. Conf. Biometrics, Sardinia, Alghero, 847–856, 2009.
5. N. Poh, J. Kittler, S. Marcel, D. Matrouf, J. F. Bonastre.: Model and score adaptation for biometric systems: coping with device interoperability and changing acquisition conditions. Proc. Int. Conf.on Pattern Recognition, Istambul, Turkey, 1229–1232, 2010.
6. U. Uludag, A. Ross, A. Jain: Biometric template selection and update: A case study in fingerprints, Pattern Recognition., 37 (7): 1533–1542, 2004.
7. X. Jiang, W. Ser: ‘Online fingerprint template improvement’, IEEE Trans. Pattern Anal. Mach. Intell., 24 (8):1121–1126, 2008.
8. C. Ryu, K. Hakil, A. K. Jain: ‘Template adaptation based finger print verification’. Proc. 18th Int. Conf. on Pattern Recognition, HongKong, 582–585, 2006.
9. X. Liu, T. Chen, S. M. Thornton: Eigenspace updating for nonstationaryprocess and its application to face recognition’, PatternRecognit., 36 (9), 1945–1959, 2003.
10. F, Roli, L. Didaci, G. L. Marcialis:Template co-update in multimodalbiometric systems. Proc. IEEE/IAPR Int. Conf. on Biometrics, Seoul,Korea, , 1194–1202, 2007.
11. F. Roli, G. L. Marcialis: Semi-supervised PCA-based face recognition using self-training. Proc. Joint IAPR Int. Workshop on S+SSPR06,HongKong, China, 560–568, 2006.
12. A. Rattani, G. L. Marcialis, F. Roli: Biometric template update usingthe graph mincut: a case study in face verification. Proc. Sixth IEEE Biometric Symp., Tampa, USA, 23–28, 2008.
13. L. Deng and D. Yu Deep Learning: Methods and Applications, 2014
14. http://research.microsoft.com/pubs/209355/DeepLearningNowPublishing-Vol7-SIG-039.pdf
15. Y. Bengio, . Learning Deep Architectures for AI". Foundations and Trends in Machine Learning **2** (1). 2009.
16. Y. Bengio, A. Courville, and P. Vincent., Representation Learning: A Review and New Perspectives ,IEEE Trans. PAMI, special issue Learning Deep Architectures, 1798-1828 ,2013.
17. J. Schmidhuber, Deep Learning in Neural Networks: An Overview, http://arxiv.org/abs/1404.7828, 2014
18. S Hyun Ah, and S-Young Lee. Hierarchical Representation Using NMF. Neural Information Processing. Springer Berlin Heidelberg, 2013.
19. A. Olshausen, Emergence of simple-cell receptive field properties by learning a sparse code for natural images .Nature 381.6583. 607-609, 1996.
20. R. Collobert . Deep Learning for Efficient Discriminative Parsing". videolectures.net. Ca. 7:45. 2011
21. L. Gomes, Machine-Learning Maestro Michael Jordan on the Delusions of Big Data and Other Huge Engineering Efforts". IEEE Spectrum, 2014.
22. K. Fukushima, Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position". Biol. Cybern **36**: 193–202. doi:10.1007/bf00344251.
23. P. Werbos., Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences, PhD thesis, Harvard University, 1974.
24. B. Cheung, Convolutional Neural Networks Applied to Human Face Classification " in 11th International Conference on Machine Learning and Applications (ICMLA), 580-583, 2012.
25. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-Based Learning Applied to Document Recognition, in Proceedings of the IEEE, , 2278-2324, 1998.
26. Y. N. Chen, C. C. Han, C. T. Wang, B. S. Jeng, and K. C. Fan, The Application of a Convolution Neural Network on Face and License Plate Detection, presented at the 18th International Conference on Pattern Recognition, 552-555 , 2006.
27. P. Y.Simard, D. Steinkraus, and J. C.Platt, Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis," in Seventh International Conference on Document Analysis and Recognition,958-963, 2003.
28. Y. Bengio, E Thibodeau-Laufer, G Alain, J Yosinski. Deep generative stochastic networks trainable by backprop.:1306.1091. 2013.