Deep Learning Networks for Off-Line Handwritten Signature Recognition

Bernardete Ribeiro[1], Ivo Gonçalves[1], and Sérgio Santos[1], and Alexander Kovacec[2]

[1] CISUC, Department of Informatics Engineering
[2] Department of Mathematics,
University of Coimbra, Portugal
{bribeiro,icpg,sdsantos}@dei.uc.pt, kovacec@mat.uc.pt

Abstract. Reliable identification and verification of off-line handwritten signatures from images is a difficult problem with many practical applications. This task is a difficult vision problem within the field of biometrics because a signature may change depending on psychological factors of the individual. Motivated by advances in brain science which describe how objects are represented in the visual cortex, advanced research on deep neural networks has been shown to work reliably on large image data sets. In this paper, we present a deep learning model for off-line handwritten signature recognition which is able to extract high-level representations. We also propose a two-step hybrid model for signature identification and verification improving the misclassification rate in the well-known GPDS database.

Keywords: Deep Learning, Generative Models, Signature Recognition.

1 Introduction

The robustness and efficiency by which humans can recognize objects has since ever been intriguing for researchers and a trigger challenge in computational intelligence. Motivated by the extreme efficiency of the visual recognition system recent studies in brain science fields show that this is largely due to the expressive deep architecture employed by human visual cortex systems [15]. Research in brain science has recently traced the respective roles of the perceptual and visuo-motor skills on letter shape learning and handwriting movement execution [12]. In the scope of biometric analysis, an important problem is to distinguish between genuine and forged signature which is a hard task. The continued motivation to investigate this problem may be attributed in part to its challenging nature which depends on various factors such as behavioral characteristics like mood, fatigue, energy, etc.. Feature extraction and pattern recognition undoubtedly constitute essential components of a signature verification system. Research has been very intensive in the last years and many approaches have been devised mainly using discriminative techniques [2,6,5,7]. This kind of solutions plays an important role, with many applications in different fields, namely

C. San Martin and S.-W. Kim (Eds.): CIARP 2011, LNCS 7042, pp. 523–532, 2011.

in many official documents, such as detecting whether a person is misusing a citizen ID, or to verify if a bank check was really signed by the owner, or even accelerating the legal process of authenticating documents.

In this paper, we use instead a generative model broadly construed on a deep neural architecture trained by the contrastive divergence method introduced by Hinton [10]. The dataset is the GPDS("Grupo de Procesado Digital de Senales") signatures image database[1] which provided 300 signatures folders, 24 genuine and 30 faked for each folder. First, the feature extraction is performed implementing the algorithms described in literature [2] yet novel features were extracted. Second, we propose a two-step hybrid model, for signatures identification and verification, with good performance for all the dataset. For the sake of results comparison [2], one important part of the tests considered 39 and 44 folders of signatures. Third, we put forward a deep learning architecture which made possible to set up a model with representational layers working out as the human mental representation ability.

The paper is organized as follows. Section 2 describes both the Restricted Boltzmann Machine (RBM) model and the deep learning algorithm. We introduce the signature verification problem in Section 3 starting by describing the GPDS database, and proceeding with the preprocessing and the feature extraction stages. In Section 4 we introduce the experimental setup, present the results, and discuss the proposal regarding the two-step hybrid identification and verification model and the deep learning methodology. Finally, in Section 5, we summarize the conclusions and point out further lines for future work.

2 Deep Learning

Theoretical results suggest that deep learning architectures with multiple levels of non-linear operations provide high-level abstractions for object recognition similar to those found in the human brain. Deep Belief Networks have recently been proposed with notable success excelling the state-of-the-art in visual recognition and AI areas. Bengio [3] gives an overview of the learning algorithms for deep architectures, in particular those exploiting Restricted Boltzmann Machines, which are used to construct deeper models such as Deep Belief Networks.

2.1 Restricted Boltzmann Machine

A Restricted Boltzmann Machine (RBM) is an energy-based generative model that consists of a layer of binary visible units ($\mathbf{v}$, whose states are observed) and a layer of binary hidden units ($\mathbf{h}$, whose states cannot be observed)(Hintom, 2006 [10]), [11]. The hidden units with no pairwise connections act as latent variables (features) that allow the RBM to model distributions over state vectors (see Figure 1). With these restrictions, the hidden units are conditionally independent given visible units (i.e. a visible vector). Given an energy function

[1] Offline GPDS signature database http://www.gpds.ulpgc.es/download/

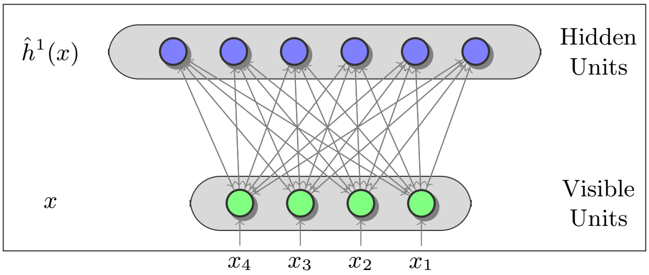


Fig. 1. Restricted Boltzmann Machine (RBM) with ($V = 4$, $H = 6$)

$E(\mathbf{v}, \mathbf{h})$ on the whole set of visible and hidden units, the joint probability is given by:

$$p(\mathbf{v}, \mathbf{h}) = \frac{e^{-E(\mathbf{v},\mathbf{h})}}{Z} \tag{1}$$

where Z is a normalizing partition function i.e., ensures that $p(\mathbf{v}, \mathbf{h})$ is a valid distribution.

For the binary units $h_i \in \{0, 1\}$ and $v_i \in \{0, 1\}$ the energy function of the whole network is:

$$\begin{aligned} E(\mathbf{v}, \mathbf{h}) &= -\mathbf{h}^T \mathbf{W} \mathbf{v} - c^T \mathbf{v} - b\mathbf{h}^T \\ &= -\sum_{jk} W_{jk} v_k h_j - \sum_k c_k v_k - \sum_j b_j h_j \end{aligned} \tag{2}$$

The marginal distribution over $\mathbf{v}$ is:

$$p(\mathbf{v}) = \sum_h p(\mathbf{v}, \mathbf{h}) = \sum_h p(\mathbf{v}|\mathbf{h}) p(\mathbf{h}) \tag{3}$$

With H hidden units the hidden vector $\mathbf{h}$ can take 2^H possible values, thus 2^H distributions $p(\mathbf{v}|\mathbf{h})$. Therefore, computing the marginal for a large H is impractical. A good estimator of the log-likelihood gradient is the Contrastive Divergence (CD) algorithm ([10]).

A good property of the RBM is that the posterior of one layer given the other is easy to compute.

$$\begin{aligned} p(\mathbf{v}|\mathbf{h}) &= \prod_k p(v_k|\mathbf{h}) \text{ where } p(v_k = 1|\mathbf{h}) = sigm(c_k + \sum_j W_{jk} h_j) \\ p(\mathbf{h}|\mathbf{v}) &= \prod_j p(h_j|\mathbf{v}) \text{ where } p(h_j = 1|\mathbf{v}) = sigm(b_j + \sum_k W_{jk} v_k) \end{aligned} \tag{4}$$

where $sigm$ is the sigmoid function $\frac{1}{(1+e^{-z_i})}$ with $z_i = b_i + \sum_j W_{ji} s_j$ where s is the state of the unit i and b the bias. Inference of hidden factor $\mathbf{h}$ given the observed $\mathbf{v}$ can be done easily because $\mathbf{h}$ are conditionally independent given $\mathbf{v}$.

2.2 Learning in Deep Neural Networks

Definition 1. *Deep Neural Network: A deep neural network contains an input layer and an output layer, separated by l layers of hidden units.*

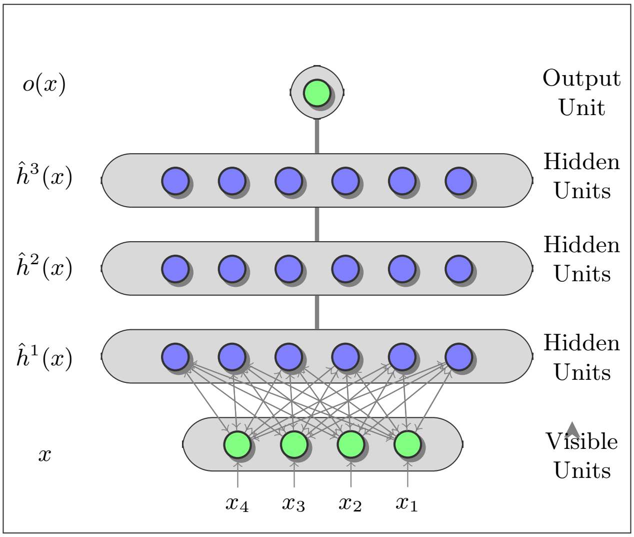


Fig. 2. Deep Belief Network with three hidden layers

The learning algorithm in Boltzmann Machines [1] allows to discover interesting features that may represent complex regularities in the training data. The algorithm can be very slow in networks with many layers, but it is fast in an RBM that has a single layer of feature detectors. The composed neural network can efficiently be trained by composing RBMs using the feature activations of one layer as the training data for the next (see Figure 2). The rationale is that the whole network can be viewed as a single, multilayer generative model and each additional layer improves a lower bound on the probability that the multilayer model would generate the training data (Hinton, 2006 [9]). Learning one hidden layer at a time is much more effective given their size which can be very large (MM of weights). Besides, highest level features are much more useful for classification (or dimension reduction) than raw data vectors.

An energy-based model of RBMs can be learnt by performing (stochastic) gradient descent on the empirical negative log-likelihood of the training data with respect to the RBM parameters.

$$\frac{\partial}{\partial\theta}(-\log p(\mathbf{v}_0)) = E_{p(\mathbf{h}|\mathbf{v}_0)}\left[\frac{\partial E(\mathbf{v}_0,\mathbf{h})}{\partial\theta}\right] - E_{p(\mathbf{v},\mathbf{h})}\left[\frac{\partial E(\mathbf{v},\mathbf{h})}{\partial\theta}\right] \tag{5}$$

where θ are the model parameters. This gradient is difficult to compute analytically. Markov Chain Monte Carlo methods are well-suited for RBM models. One iteration of the Markov Chain works well in practice.

$$\mathbf{v}_0 \xrightarrow{p(\mathbf{h}_0|\mathbf{v}_0)} \mathbf{h}_0 \xrightarrow{p(\mathbf{v}_1|\mathbf{h}_0)} \mathbf{v}_1 \xrightarrow{p(\mathbf{h}_1|\mathbf{v}_1)} \mathbf{h}_1 \tag{6}$$

where the operations of sampling are schematically indicated. Estimation of the gradient using the above procedure is denoted by CD-1, where CD-k represents the Contrastive Divergence algorithm [10,4] for performing k iterations of the Markov Chain up to $\mathbf{v}_k$.

Given a training set of state vectors (data) learning consists of finding weights and bias that define a Boltzmann distribution in which the training vectors have high probability.

3 Signature Verification: Problem Statement

Our main task is to develop an off-line signature verification system able to distinguish faked signatures from genuine ones. To achieve this goal, the images pre-processing, feature extraction and classifiers design steps need to be performed in the GPDS database of digitalized signatures.

3.1 GPDS Signature Data Base

The GPDS database was downloaded from `http://www.gpds.ulpgc.es/download/` under a license agreement. The database contains data from 300 individuals: 24 genuine signatures for each individual plus 30 forgeries of his/her signature. Detailed information on how the GPDS dataset was built is given in [6] where it is also described how the data images were acquired (and pre-processed) prior to its completion (and organization) in the dataset. In [14] an interesting discussion on the different types of existing forgeries can be found.

3.2 Feature Extraction for the GPDS Signature Data Base

Feature extraction from image signatures is a crucial component of the verification rate system. Generally, an image feature is a distinctive primitive characteristic of a particular signature. More specifically, certain features are defined by the visual appearance of an image, while other result from image specific manipulations. The challenge is to find the optimal set of features able to perform forgery detection since it is not feasible to use the whole raw image. By using adequate algorithms one can extract features able to isolate characteristic regions within an image (image segmentation) and subsequent identification or labeling of such regions (image classification).

Fourteen different features have been extracted from the GPDS database to allow for signature classification using the methods described in the literature. The features (Width, Height, Tri-Surface, Six-Fold Surface, Best Fit) were described in [2]; the features (Geometric Parameters (Polar and Cartesian)) in [6] and the Modified Direction Feature (MDF) in [5]. The remaining four novel extracted features (K-Means, Histogram of frequencies, Discrete Cosine Transform and Wavelet Transform) are briefly described next.

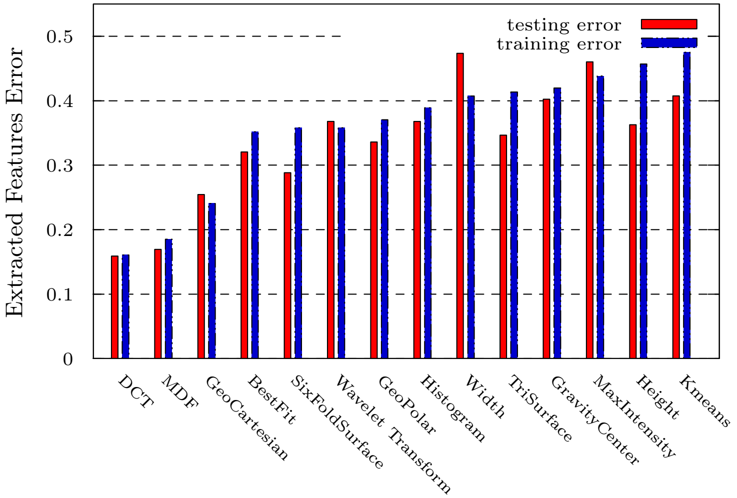


Fig. 3. Feature Ranking Analysis

1. **K-Means:** The k-means clustering algorithm was applied to the images pixels with the goal to identify the positions of the signature's main elements. The algorithm was applied to each image individually, with k set to 5 by empirical experimentation. The clusters' centroids Cartesian coordinates are the feature's values.
2. **Histogram Frequencies:** In order to evaluate signatures' intensity variations along the perpendicular axes, the frequencies of each image's horizontal and vertical histogram were calculated. The histogram frequencies are obtained using the Discrete Fourier Transform. To characterize the frequencies obtained, the three distribution quartiles values were saved.
3. **Discrete Cosine Transform Frequencies (DCT):** This feature evaluates the crispness of the signatures, whether specific frequency intervals occur more along the vertical (or the horizontal) axes. The two-dimensional Discrete Cosine Transform was applied to each image individually. The resultant frequencies are divided into N frequency intervals with the same length, N=5 by empirical experimentation. Each frequency band is separated by a diagonal axis, in order to compare frequencies with mainly a vertical orientation from those with mainly an horizontal orientation. For each interval, the proportion of the frequency amplitudes between the vertical and horizontal regions is calculated and used as a feature value.
4. **Discrete Wavelet Transform Frequencies (DWT):** A space-frequency analysis using the two-dimensional Discrete Wavelet Transform (DWT) is applied to the signatures, to evaluate the horizontal, vertical and diagonal pixels variations. The DWT is recurrently applied, with the *haar* wavelet, to inferior frequency levels. For each orientation (and each level of decomposition), the gravity centers of the frequency amplitudes are calculated. Those values represent the signatures' regions where those frequency intervals are most present.

A global analysis of their discriminative power is illustrated in Figure 3 where MDF and DCT features show the strongest influence on the classifier's performance. An example of the MDF feature extraction is given in Figure 4. Several values of max transitions from black to white were used in the picture and the

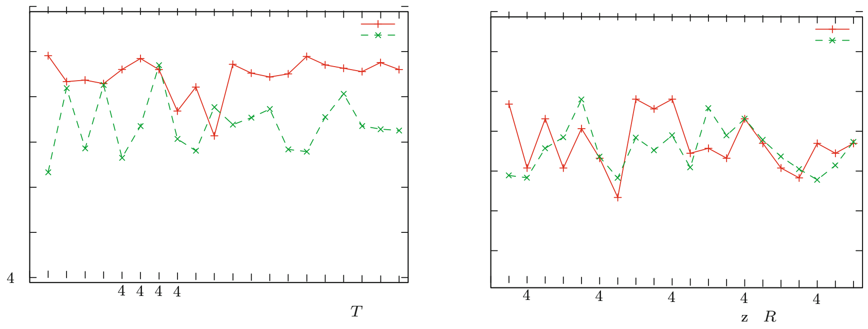


Fig. 4. Performance and Error for (a) Transition (T)/ (b) Resampling (R) for MDF Feature. Best pair found$(T = 4; R = 5)$

image resampling. The best results w.r.t. error performance were found with the pair $(T = 4, R = 5)$.

4 Experimental Setup and Results

The database was split in two parts for training and testing. We followed the procedure in [2] i.e., for each signature we used from the genuine set, 20 samples for training and 4 for testing. As for the forged set, 25 samples were used for training and 5 for testing. Overall we come up with 658 attribute values for the whole set of extracted features. We tested out a number of configurations [8] with variable size in number of features' combinations (and corresponding attributes).

4.1 Two-Step Signature Verification Model

We put forward a hybrid model consisting of two steps, the first, identifies the owners of the signatures while, the second, determines its authenticity, i.e., accepts or rejects a signature. This architecture could mostly be used to verify the signature of a check or a signed document. This approach requires a classifier that can identify any signature (*identification classifier*) and several classifiers that given signatures of only one individual can determine its authenticity (*specific classifiers*). This entails the existence of a multi-class classifier for the identification classifier and N binary specific classifiers, one for each individual. The identification classifier will be trained with all the signatures in a standard way. Each specific classifier will be trained with both authentic and forged signatures from only one individual. The specific classifiers are expected to achieve higher accuracy than a *general classifier*, i.e. a binary classifier trained with signatures from different individuals. Intuitively, this originates from the idea that it is easier to find an authentic/forged pattern from a single individual than it is to find the same pattern for every individual possible. The generic classifier is used

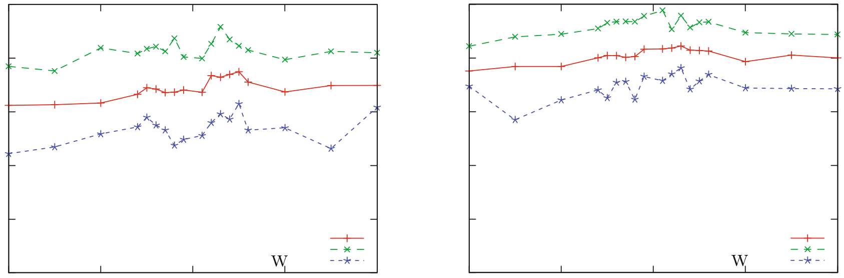


Fig. 5. Specific and (Generic) Classifiers Performance versus the Ratio threshold

whenever its predicted accuracy (based on the cross-validation error) is higher than that of the specific classifier. To fine tune the approach we included a prevalence *ratio* threshold that explicitly favors the specific (or generic) classifier if it is greater (or smaller) than 1, respectively. As an example, if we want to choose the specific classifier even if its predicted accuracy is up to 5% lower than the general classifier, a ratio of 1.05 should be set. The results illustrated in Figure 5 show how this works out with changing values of the ratio. It is also observed a higher value of sensitivity than specificity since the number of forged signatures is greater than the genuine ones.

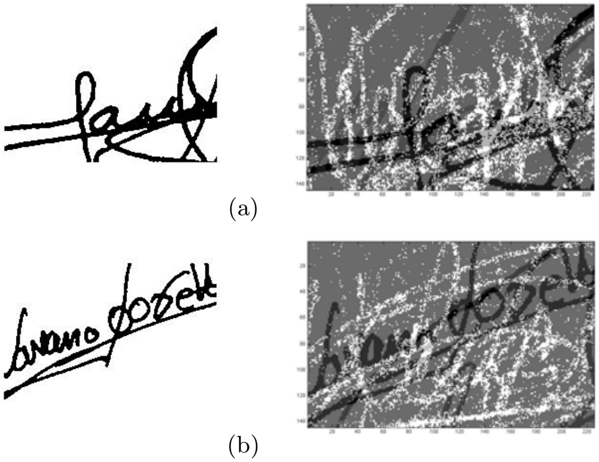
We tried out several algorithms for the design of the classifiers, namely,(Fisher Linear Discriminant, Feed Forward Neural Network, Radial Basis Neural Network, Naive Bayes and Support Vector Machines (SVMs)) whose study is available elsewhere [8]. We choose SVMs to present the results (see Table 1) of an experimental analysis of the performance of the specific classifier with the number of folders varying from 10 to 300. The best feature configuration attained (MDF, Width, Six-Fold Surface and Wavelet Transform) consisting of 179 values for the images signatures was used. Moreover, we perform 30 runs for each number of folders in the Table 1 and averaged the results (including standard deviations). The metrics for performance evaluation were evaluated from the confusion matrix with True Positive (TP), True Negative (TN), False Positive (FP) and False Negative and are indicated as follows: Recall, Precision, Training Accuracy, Testing Accuracy and F1 measure. The latter measures the trade-off between the Recall and Precision and is a good indicator in skewed distributions as in the case of the GPDS database. In Biometrics, the error of type I (i.e. False Positive Rate) is the False Reject Rate (FRR). It means a false alarm of the positive class (forged signature). The error type II (i.e. False Negative Rate) is the False Accept Rate (FAR) which relates to missing to detect a forged signature. The results compare well with those presented in [2] and [13] in particular w.r.t the FAR, which has the lowest value for 44 folders and is a good indicator for the system's performance in the case of 300 folders which contain 16200 signatures.

Table 1. Signature Verification Performance (%)

Folders	Recall	Precision	FRR	FAR	Trn Acc	Test Acc	F1
10	86.67 ± 23.46	84.09 ± 15.31	28.33	13.33	91.26 ± 8.19	80.00 ± 16.22	82.75 ± 16.34
20	84.86 ± 19.85	85.42 ± 13.83	22.86	15.14	95.30 ± 4.20	81.43 ± 14.27	83.65 ± 13.29
39	87.28 ± 9.77	89.01 ± 15.30	17.82	12.72	97.36 ± 3.70	85.01 ± 14.30	85.97 ± 10.32
44	90.76 ± 16.17	84.58 ± 14.78	26.42	**9.24**	93.21 ± 5.88	83.12 ± 14.45	**86.15** ± 12.32
60	80.17 ± 21.99	91.32 ± 12.57	11.77	19.83	97.64 ± 2.54	83.75 ± 13.66	83.19 ± 15.74
120	80.22 ± 22.99	90.45 ± 14.26	11.94	19.78	98.12 ± 1.85	83.70 ± 14.51	83.18 ± 16.86
200	82.25 ± 21.64	91.25 ± 13.56	11.50	17.75	99.46 ± 1.10	85.03 ± 14.25	84.63 ± 16.29
240	82.83 ± 21.14	90.54 ± 14.28	12.71	17.17	99.86 ± 0.56	84.81 ± 14.72	84.75 ± 16.10
300	85.33 ± 20.67	86.23 ± 15.79	20.25	**14.67**	94.10 ± 5.10	82.85 ± 15.11	**84.37** ± 15.05

4.2 Deep Learning Model for Signature Recognition

Figure 6 shows the learned weights of two signatures of different owners, where the white dots appear as noise. The deep neural network architecture with 100 visible units and two layers with 100 hidden units each was able to extract layer-by-layer high-level representations of the images. The learning rate is $\eta = 0.08$ and the momentum $\alpha = 0.4$. The number of epochs was varied from 100, 500, 1000 and 5000. We clamped into the network 10 signature folders after cropping the images to the reasonable size of (144×225). The cost of training was very high increasing with the number of epochs and with the number of hidden units. Therefore it was only possible to test with 10 folders of signatures although there is room for improvements. We show that the architecture learns the relevant features at hand given very limited prior knowledge.


(a)

(b)

Fig. 6. Signature (a) original (b) learning weights

5 Conclusions and Future Work

We presented a verification system for off-line signature recognition proposing a two-step hybrid classifier system with overall good performance in the GPDS database. In addition, the preliminary results with a deep learning architecture are promising and raise interest regarding the application of this kind of models in this problem. We were able to extract a high-representation of the signature images through multi-layers in a deep hierarchical structure that allows non-local generalization and comprehensibility in this specific domain. Despite the great prospect of deep learning technologies future work will perform an extensive study to cope with the millions of parameters that need to be adjusted, in particular, with the use of Graphics Processing Units (GPU).

References

1. Ackley, D., Hinton, G., Sejnowski, T.: A learning algorithm for Boltzmann Machines. Science 9(1), 147–169 (1985)
2. Armand, S., Blumenstein, M., Muthukkumarasamy, V.: Off-line signature verification using an enhanced modified direction feature with single and multi-classifier approaches. IEEE Computational Intelligence Magazine 2(2), 18–25 (2007)
3. Bengio, Y.: Learning deep architectures for AI. Foundations and Trends in Machine Learning 2(1), 1–127 (2009)
4. Bengio, Y., Delalleau, O.: Justifying and generalizing contrastive divergence. Neural Computation 21(6), 1601–1621 (2009)
5. Bluemenstein, M., Liu, X.Y.: A modified direction feature for cursive character recognition. In: IEEE-IJCNN, pp. 2983–2987 (2004)
6. Ferrer, M.A., Alonso, J.B., Travieso, C.M.: Off-line geometric parameters for automatic signature verification using fixed-point arithmetic. IEEE Trans. on Pattern Analysis and Intelligence 27(6), 993–997 (2005)
7. Gader, P.D., Mohamed, M., Chiang, J.H.: Handwritten word recognition with character and inter-character neural networks. IEEE Trans. on Sys., Man and Cyber. - Part B: Cybernetics 27(1), 158–164 (1997)
8. Gonçalves, I., Santos, S.: Off-line signatures verification system. Tech. Rep. trp-#10/11, University of Coimbra, Portugal (2011)
9. Hinton, G.E.: Learning multiple layers of representation. Trends in Cognitive Sciences 11, 428–434 (2007)
10. Hinton, G.E., Osindero, S., Teh, Y.: A fast learning algorithm for deep belief nets. Neural Computation 18, 1527–1554 (2006)
11. Larochelle, H., Bengio, Y., Louradour, J., Lamblin, P.: Exploring strategies for training deep neural networks. J. of Machine Learning Research 10, 1–40 (2009)
12. Longcamp, M., Boucard, C., Gilhodes, J.C., Anton, J.L., Roth, M., Nazarian, B., Velav, J.L.: Learning through hand- or typewriting influences visual recognition of new graphic shapes: Behavioral and functional imaging evidence. Science Journal of Cognitive Neuroscience 20, 802–815 (2008)
13. Lv, H., Wang, W., Wang, C., Zhuo, Q.: Off-line chinese signature verification based on support vector machines. Pattern Recognition Letters 26, 2390–2399 (2005)
14. Nguyen, V., Blumenstein, M., Vallipuram, M., Leedham, G.: Off-line signature verification using enhanced modified direction features in conjunction with neural classifiers and support vector machines. In: IEEE-ICDAR, pp. 1300–1304 (2009)
15. Yu, K., Xu, W., Gong, Y.: Deep learning with kernel regularization for visual recognition. In: Neural Information Processing Systems, NIPS (2009)