A Novel Hybrid CNN-AIS Visual Pattern Recognition Engine

Vandna Bhalla(✉), Santanu Chaudhury, and Arihant Jain

Indian Institute of Technology Delhi, New Delhi, India
{vbhalla.du,arihant.jms}@gmail.com, santanuc@ee.iitd.ernet.in
http://www.iitd.ac.in

Abstract. Machine learning methods are used today mostly for recognition problems. Convolutional Neural Networks (CNN) have time and again proved successful for many image processing tasks primarily for their architecture. In this paper we propose to apply CNN to small data sets like for example, personal photo albums or other similar environs where the size of training dataset is a limitation, within the framework of a proposed hybrid CNN-AIS model. We use Artificial Immune System Principles to enhance the small size of training data set. A layer of Clonal Selection is added to the local filtering and max pooling of CNN Architecture. The proposed Architecture is evaluated using the standard MNIST dataset by limiting the data size and also with a small personal data sample belonging to two different classes. Experimental results show that the proposed hybrid CNN-AIS based recognition engine works well when the size of training data is limited in size.

Keywords: CNN · Clonal Selection (CS) · Artificial Immune Systems (AIS) · Small data size · Diversity

1 Introduction

Today all object recognition approaches use machine-learning methods. Larger the dataset better is the performance. Labeled datasets like NORB, Caltech-101/256 and CIFAR-10/100 with tens of thousands of images are in todays scenario considered small and LabelMe and Image Net with millions of images are preferred. A simple recognition task also requires datasets of size of the order of tens of thousands of images. It is always assumed that objects in realistic settings show a lot of variability. Hence it is essential to have larger training sets to learn to recognize them with almost all current technologies. To learn from thousands of objects from millions of images, a model with a large learning capacity with powerful processing is required. CNN have shown very good image classification performance which can be attributed to their ability to learn rich image representation compared to hand crafted low level features used in other image classification techniques. Recent publications indicate that deep hierarchical NN improve pattern classification. In fact Deep NN are fully exploited to

M. Kryszkiewicz et al. (Eds.): PReMI 2015, LNCS 9124, pp. 215–224, 2015.
DOI: 10.1007/978-3-319-19941-2_21

their best potential when they are wide with many maps per layer and deep with many layers. We too saw through our experiments that properly trained wide and deep CNNs can outperform all previous methods. Learning CNNs requires a huge number of annotated training images. This property prevents application of CNNs to scenarios with limited training data. We present an innovative, adaptive, self-learning, and self-evolving hybrid recognition engine, which works well with small sized training data. The model uses the intelligent information processing mechanism of Artificial Immune System (AIS) and helps Convolutional Neural Network (CNN) generate a robust feature set taking the small set of input training images as seeds. Our model performs visual pattern learning using a heterogeneous combination of supervised CNN and Clonal Selection (CS) principles of AIS. It can be extended to perform classification tasks with limited training data particularly in the context of personal photo collections where for each training sample different points of view are gathered in parallel using clonal selection. This is very different from populating datasets with artificially generated training examples [9] by randomly distorting the original training images with randomly picked distortion parameters. The many shortcomings of small size data sets have been widely recognized by Pinto [14]. Small training data has not been given much prominence in recent research.

Specific contribution of this paper is as follows: A hybrid Convolutional Neural Network-Artificial Immune System (CNN-AIS) Recognition Engine Architecture designed to work with modest sized training data. This is detailed in Sect. 3. The model was tested on well-known MNIST digit database and showed remarkable success. The current best error rate of 0.21 % on the MNIST digit recognition task approaches human performance. We have got very good results with considerably smaller number of training samples. In addition we get comparable accuracy with the state of the art methods when the data size is increased to large numbers. The results are presented in Sect. 4. For completeness we assessed our CNN-AIS model by applying it on a personal photo collection and successfully accomplished classification for two categories of classes. Section 5 gives an application of this system in real world with results.

2 Related Work

Our model is inspired by many related works on image classification and deep learning which we briefly discuss here. The general structure of the deep convolutional neural network (CNN) was introduced in early nineties [6]. This deep convolutional neural network architecture, called LeNet, is still being used today with a lot of consistent improvements to the individual components within the architecture. An important idea of the CNN is that the feature extraction and classifier were unified in a single structure. The model was proposed for handwritten digit recognition and achieved a very high success rate on MNIST dataset [7]. But it demands substantial amount of labeled data for training (60,000 for MNIST). Though the results are promising and exciting but the bothersome part is that millions of annotated images are needed for each new

visual recognition task. Also the size of input is very small (28×28) with no background clutter, illumination change etc. which is an integral part of normal pictures/images. In fact for most realistic vision applications this is not the case. The multiple processing layers of machine learning systems extract more abstract, invariant features of data and have higher classification accuracy than the traditional shallower classifiers. These deep architectures have shown promising performances in image [7] language [19] and speech [10]. Ranzato et al. [15] trained a large CNN for object detection (Caltech 101 dataset) but obtained poor results though it achieved perfect classification on the training set. The weak generalization power of CNN when the number of training data is small and the number of free parameter is large is a case of over fitting or over parametrization. The success of object recognition algorithm to a large extent depends on features detected. The features should have the most distinct characteristics among different classes while retaining invariant characteristics within a class. Other biologically inspired models like HMAX [16] use hardwired filter and use hard Max functions to compute the responses in the pooling layer. The problem was that it was unable to adapt to different problem settings.

Transfer learning is one technique to conquer the shortfall of training samples for some categories by adapting classifiers trained for other categories. One such method [12] proposes to transfer image representations learned with CNNs on large datasets to other tasks with limited training datasets. This method fails to recognize spatially co occurring objects. The false positives in their results correspond to samples closely resembling target object classes. Recognition of very small or very large objects could also fail.

Successful algorithms have been built on top of handcrafted gradient response features such as SIFT and histograms of oriented gradients (HOG). These are fixed features and are unable to adjust to model the intricacies of a problem. Traditional hand designed feature extraction is laborious and moreover cannot process raw images while the automatic extraction mechanism can fetch features directly. In [1,2] supervised classifiers such as CNNs, MLPs, SVMs and K-nearest Neighbors are combined in a Mixture of Experts approach where the output of parallel classifiers is used to produce the final result. One such recognition system, CNSVM [11], is a classifier built as a single model with SVM and CNN. The CNN is trained using the back-propagation algorithm and the SVM is trained using a non-linear regression approach. The work, again, requires large training data. A comparison of Support Vector Machine, Neural Network, and CART algorithms using limited training data points was done though the data was for the land-cover classification [17]. SVM generated overall accuracies ranging from 77 % to 80 % for training sample sizes from 20 to 800 pixels per class, compared to 67–76 % and 6273 % for NN and CART, respectively. CNNs though efficient at learning invariant features from images, do not always produce optimal classification and SVMs with their fixed kernel function are unable to learn complicated invariance. Our approach is different and we propose a single coupled architecture for training and testing using deep CNN and AIS principles.

3 Convolutional Neural Network-Artificial Immune System (CNN-AIS) Model

We use deep convolutional neural networks as wide and deep trained networks are better than most other methods. Our proposed Architecture integrates Clonal Selection (CS) principles from Artificial Immune System (AIS) with deep Convolutional Neural Networks (CNN) in a novel way. We will briefly introduce the AIS theory and the basic CNN structure that we have used in our model. Subsequently the architecture of the hybrid CNN-AIS trainable recognition engine is presented.

Artificial Immune Systems (AIS). Artificial Immune System use Clonal Selection and Negative Selection principles imitating the Human Biological Immune System. The main task of the immune system is to defend the organism against pathogens. In the human body the B-cells with different receptor shapes try to bind to antigens. The best fit cells proliferate and produce clones which mutate at very high rates. The process is repeated and it is likely that a better B-cell (better solution) might emerge. This is called Clonal Selection. These clones have mutated from the original cell at a rate inversely proportional to the match strength. Two main concepts are particularly relevant for our framework. (i) Generation of Diversity: The B cells produce antibodies for specific antigens. Each B cell makes a specific antibody, which is expressed from the genes in its gene library. The gene library does not contain genes that define antibodies for every possible antigen. Gene fragments in the gene library randomly combine and recombine and produce a huge diverse range of antibodies. This helps the immune system to make the precise antibody for an antigen it may never have encountered previously. (ii) Avidity: Refers to the accrued strength of various diverse affinities of individual binding interaction. Avidity (functional affinity) is the collective strength of multiple affinities of an antigen with various antibodies. Based on this biological process, quite a few Artificial Immune System, (AIS), [4,5], have been developed in the past. Castro developed the Clonal Selection Algorithm (CLONALG) [3] on the basis of Clonal Selection theory of the immune system. It was proved that it can perform pattern recognition. The CLONALG algorithm can be described as follows: 1. Randomly initialize a population of individual (M); 2. For each pattern of P, present it to the population M and determine its affinity with each element of the population M; 3. Select n of the best highest affinity elements of M and generate copies/clones of these individuals proportionally to their affinity with the antigen which is the pattern P. The higher the affinity, the higher the number of clones, and vice-versa; 4. Mutate all these copies with a rate proportional to their affinity with the input pattern: the higher the affinity, the smaller the mutation rate; 5. Add these mutated individuals to the population M and reselect m of these maturated individuals to be kept as memories of the systems; 6. Repeat steps 2 to 5 until a predefined optimal criterion is met.

Convolutional Neural Network (CNN). A Convolutional Neural Network [9] is a multilayer feed forward artificial neural network with a deep supervised learning architecture. The ordered architectures of MLPs progressively learn the higher level features with the last layer giving classification. Two operations of convolutional filtering and down sampling alternate to learn the features from the raw images and constitute the feature map layers. The weights are trained by a back propagation algorithm using gradient descent approaches for minimizing the training error. We have used Stochastic Gradient Approach as it prevents getting stuck in poor local minima. A simplified CNN was presented in [13] which we have used for our work instead of using the rather complicated LeNet-5 [8]. The model has five layers.

CNN-AIS Model. The architecture of our hybrid CNN-AIS model was designed by adding an additional layer of Artificial Immune System (AIS) based Clonal Selection (CS) in the traditional Convolutional Neural Network (CNN) structure, Fig. 1. The model is explained layer wise. **Convolutional Layer:** A 2D filtering between input images n, and a matrix of kernels/weights K produces the output I where $I_k = \Sigma_{i,j,k}$ M $(n_i * K_j)$ where M is a table of input output relationships. The kernel responses from the inputs connected to the same output are linearly combined. As with MLPs a scaled hyperbolic tangent function is applied to every I. **Sub sampling Layer:** Small invariance to translation and distortion is accomplished with the Max-Pooling operation. This is for faster convergence and improves generalization as well. **Fully Connected Layer - I:** The input to this layer is a set of feature maps from the lower layer which are

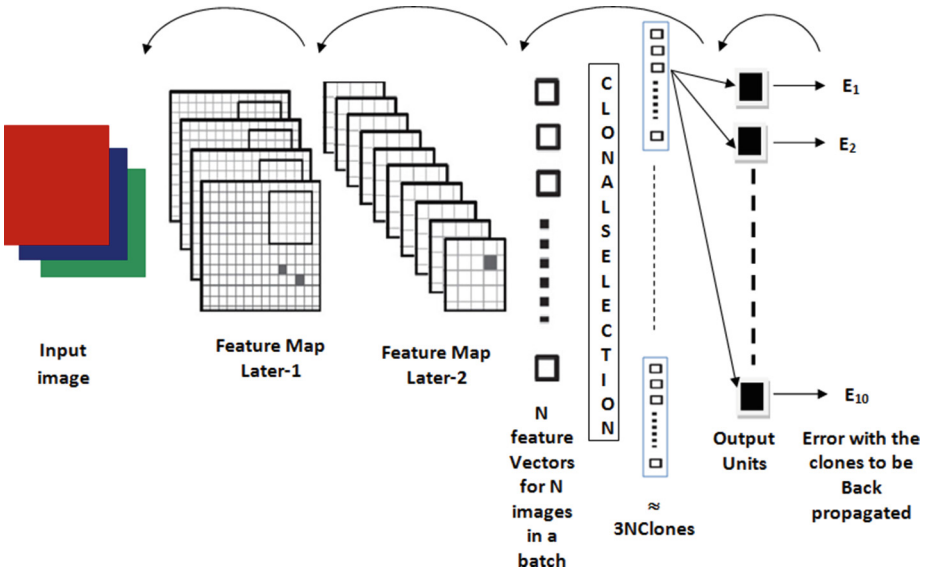

Fig. 1. The architecture of the hybrid model

combined into a 1-dimensional feature vector and subsequently passed through an activation function. **Clonal selection Layer:** This is the new additional layer that we propose in our architecture and it is the second last layer. The purpose of this layer is to generate additional input data for the final MLP layer. Additional data helps the MLP to train better which in turn leads to better trained kernels at the convolutional layer. This layer receives its input from the fully connected layer-I in the form of 1-D feature vector for all the images (n) in the current running batch. Each feature vector in the Feature set undergoes Cloning, Mutation and Crossover according to the rules of Clonal Selection to generate additional features that satisfy the minimum threshold criteria and resemble the particular class. The number of clones is calculated by

$$\text{CNum} = \eta \text{ x affinity (Feature Vector1, Feature Vector2) ... (i)},$$

where η is the cloning constant. Higher the affinity of match the greater the clone stimulus gets, the more the cloning number is. On the contrary, the number is less, which is consistent with biological immune response mechanism. Mutation frequency is defined as Rate, which is calculated by

$$\text{Rate} = \alpha \; 1/ \text{ affinity (Feature Vector1, Feature Vector2) ... (ii)}$$

where α is mutation constant. In accordance with (ii), the higher the affinity of match, the smaller the clone stimulus gets, the lower the mutation frequency is. On the contrary, the mutation frequency is higher. Hence from n initial feature sets we now have (n $\times$ CNum) features. These newly generated feature vectors are grouped into batches and individually fed to the output layer and the subsequent error is back propagated to train the kernels of the CNN. Hence from the seeds of a few representative images of each class a bigger set is evolved using Clonal Selection principles of Artificial Immune System. End of training phase yields a set of representative features, which we call antibodies, from each class of size much larger than the original dataset and a trained CNN. Though we start with random values of feature sets (antibodies) for each class but eventually they converge to their optimal values. **Output Layer (Fully Connected Layer-II):** This layer has one output neuron per class label and acts as linear classifier operating on the 1-dimensional feature vector set computed from the CS layer

4 Results

We performed tests on MNIST dataset. The results are tabulated in Table 1 which show the remarkable improvement that our hybrid model achieves when the training dataset is limited. We observe around 10–15 % improvement over traditional CNN when applied for small data size. When the data set is large, then too, CNN+ AIS achieves 0.66 % error rate giving an improvement of 0.04 %. Table 2 compares our performance with some other distortion techniques used to increase the data size found in literature. Very recently error rate upto

0.21 % has been achieved using DropConnect [18]. So for large datasets too our model approaches the achieved accuracy. Our model enhances accuracy by extending data set at feature level using CS principles rather than at the input level like for example in affine distortions. Our model improves accuracy for applications where the training data set is scarce. This inspired us to apply it for personal photo album where the training data is small. This application is discussed in the next section. We reiterate that our model gives very good results for small as well as large training data sizes unlike other models in literature.

Table 1. Test results

Training data size	300	500	1000	2000	5000	15000	30000	60000
CNN	70 (%)	89 (%)	91.6 (%)	94 (%)	96.4 (%)	98.3 (%)	98.9 (%)	99.3 (%)
CNN+AIS	85 (%)	91.9 (%)	94 (%)	96.02 (%)	97.9 (%)	98.9 (%)	99 (%)	99.34 (%)
Improvement	**15 (%)**	**3 (%)**	**2.4 (%)**	**2 (%)**	**1.5 (%)**	**0.6 (%)**	**0.1 (%)**	**0.04 (%)**

Table 2. Comparison of results with some other techniques

S.No	Algorithms	Technique	Error
1	2-layer MLP(MSE)	Affine Distortion	1.6 (%)
2	SVM	Affine + Thickness	1.4 (%)
3	Tangent Dist.	Affine Distortion	1.1 (%)
4	LeNet5(MSE)	Affine Distortion	0.8 (%)
5	Boost.LeNet4(MSE)	Affine Distortion	0.7 (%)
6	**CNN+AIS**	**Clonal Selection Principle**	**0.66(%)**

5 Application: Personal Photo Album

The CNN-AIS generates a robust and diverse pool of feature vectors and a trained CNN for any class. We tested this model for a personal collection of photos for two classes Picnic (A) and Conference (N). For every testing image (the antigen), the trained CNN-AIS model computes the feature vector and compares this with the feature set pool of that class, Fig. 2. If the number of matches of the test image with the various feature sets of that class and the combined affinities exceed the threshold then the testing image is placed in that class. These emulate the antibodies in a human body recognizing an antigen.

A two phase testing mechanism is used for classification. The first phase matches the test image feature with the 3N feature sets (antibodies) of each of the classes. The total number of antibodies lying above the threshold for matching is counted (C) for each class. All classes providing a minimum number of C are

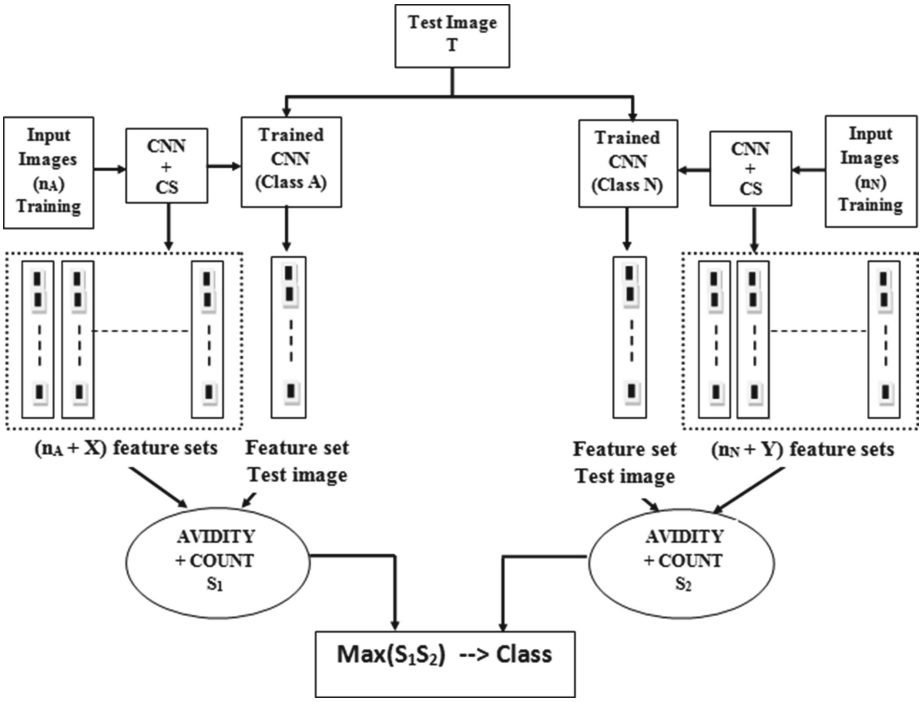

Fig. 2. Application: hybrid model used for personal photo classification

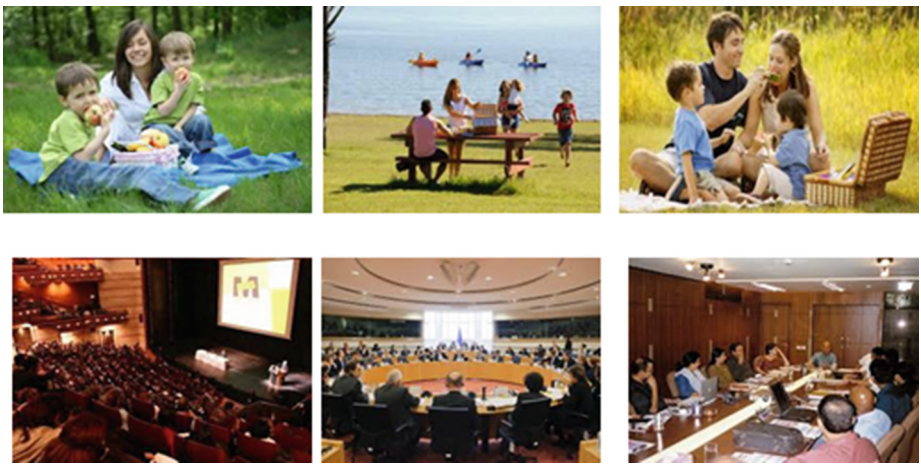

Fig. 3. Sample images from application dataset

qualified for phase 2 testing. The second phase calculates Avidity for each class which is the mean strength of multiple affinities of all qualified antibodies in C with the testing image antigen. It is calculated by taking the mean of individual scores (calculated using inner product measure) of matching of test image with each antibody above the threshold for each qualified class. This score is labeled avidity. The class is eventually decided on the basis of the combined scores S = (Count + Avidity). The experimental results are summarized in Table 3. Here TA is the True Acceptance, TR is True Rejection and Num is the number of images used. The sample dataset used for our experiments is shown in Fig. 3. Despite the diversity in the dataset and the small size of training data set, our model gives good results.

Table 3. Result analysis of personal photo album

Category	Training Num	Clones	TA(Num)	TR (Num)	Aggregate
Picnic (A)	40	160	86 (%)(50)	90 (%)(150)	88 (%)
Conference(N)	40	160	83 (%)(50)	88 (%)(150)	85 (%)

6 Conclusion

The AIS layer shows a marked improvement in recognition when training data is limited. Unlike other methods where additional data was generated at the input level, here the artificial data is generated at the feature level which is computationally fast and more accurate. The proposed model is showing promising results on personal photo albums and this can be extended to other applications where availability of data is scarce. A new class can be added to the existing set of classes dynamically replicating the behavioral aspects of self-learning and self evolving of human immune system. Even when one has an apparently enormous data set, the number of data for some particular cases of interest can be small. In fact, data sometimes exhibits a property known as the long tail, which means that a few things (e.g., words) are very common, but most things are quite rare. For example, 20 % of Google searches each day have never been seen before. So the problem of addressing small sample sizes is very relevant even in the big data era.

References

1. Abdelazeem, S.: A greedy approach for building classification cascades. In: Seventh International Conference on Machine Learning and Applications, ICMLA 2008, pp. 115–120. IEEE (2008)
2. Borji, A.: Combining heterogeneous classifiers for network intrusion detection. In: Cervesato, I. (ed.) ASIAN 2007. LNCS, vol. 4846, pp. 254–260. Springer, Heidelberg (2007)

3. De Castro, L.N., Von Zuben, F.J.: Learning and optimization using the clonal selection principle. IEEE Trans. Evol. Comput. **6**(3), 239–251 (2002)
4. Dudek, G.: An artificial immune system for classification with local feature selection. IEEE Trans. Evol. Comput. **16**(6), 847–860 (2012)
5. Hart, E., Timmis, J.: Application areas of ais: the past, the present and the future. Appl. Soft Comput. **8**(1), 191–201 (2008)
6. Knerr, S., Personnaz, L., Dreyfus, G.: Handwritten digit recognition by neural networks with single-layer training. IEEE Trans. Neural Netw. **3**(6), 962–968 (1992)
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Pereira, F., Burges, C., Bottou, L., Weinberger, K. (eds.) Advances in Neural Information Processing Systems. Curran Associates Inc., Red Hook (2012)
8. Lauer, F., Suen, C.Y., Bloch, G.: A trainable feature extractor for handwritten digit recognition. Pattern Recognit. **40**(6), 1816–1824 (2007)
9. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
10. Mohamed, A.-R., Sainath, T.N., Dahl, G., Ramabhadran, B., Hinton, G.E., Picheny, M.A.: Deep belief networks using discriminative features for phone recognition. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5060–5063. IEEE (2011)
11. Nagi, J., Di Caro, G.A., Giusti, A., Nagi, F., Gambardella, L.M.: Convolutional neural support vector machines: hybrid visual pattern classifiers for multi-robotsystems. In: 2012 11th InternationalConference on Machine Learning and Applications (ICMLA), vol. 1, pp. 27–32. IEEE (2012)
12. Oquab, M., Bottou, L., Laptev, I., Sivic, J.: Learning and transferring mid-level image representations using convolutional neural networks. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1717–1724. IEEE (2014)
13. Pan, W., Bui, T.D., Suen, C.Y.: Isolated handwritten farsi numerals recognition using sparse and over-complete representations. In: 10th International Conference on Document Analysis and Recognition, ICDAR 2009, pp. 586–590. IEEE (2009)
14. Pinto, N., Cox, D.D., DiCarlo, J.J.: Why is real-world visual object recognition hard? PLoS Comput. Biol. **4**(1), e27 (2008)
15. Ranzato, M., Huang, F.J., Boureau, Y.L., LeCun, Y.: Unsupervised learning of invariant feature hierarchies with applications to object recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2007, pp. 1–8. IEEE (2007)
16. Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., Poggio, T.: Robust object recognition with cortex-like mechanisms. IEEE Trans. Pattern Anal. Mach. Intell. **29**(3), 411–426 (2007)
17. Shao, Y., Lunetta, R.S.: Comparison of support vector machine, neural network, and cart algorithms for the land-cover classification using limited training data points. ISPRS J. Photogram. Remote Sens. **70**, 78–87 (2012)
18. Wan, L., Zeiler, M., Zhang, S., Cun, Y.L., Fergus, R.: Regularization of neural networks using dropconnect. In: Proceedings of the 30th International Conference on Machine Learning (ICML-2013), pp. 1058–1066 (2013)
19. Yu, D., Wang, S., Karam, Z., Deng, L.: Language recognition using deep-structured conditional random fields. In: 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 5030–5033. IEEE (2010)