A Computational Visual Neuroscience Model for Object Recognition

Sahar Seifzadeh[1*], Mohammad Rezaei[2], Omid Farahbakhsh[3]

1. Young Researchers and Elite Club, Qazvin Branch, Islamic Azad University, Qazvin, Iran.
2. Sleep Disorders Research Center, Kermanshah University of Medical Sciences, Kermanshah, Iran.
3. Master Mind Researchers Corporation, Shiraz, Iran.

Use your device to scan and read the article online

Citation: Seifzadeh S, Rezaei M, Farahbakhsh O. A Computational Visual Neuroscience Model for Object Recognition. Journal of Advanced Medical Sciences and Applied Technologies (JAMSAT). 2016; 2(4):313-320. http://dx.crossref.org/10.18869/nrip.jamsat.2.4.313

http://dx.crossref.org/10.18869/nrip.jamsat.2.4.313

Article info:
Received: 31 Aug. 2016
Accepted: 18 Oct. 2016

Keywords:
Cortex-like model, Biological object recognition, Biologically inspired neural network, HMAX, ELM

ABSTRACT

In this study with the inspirations from both neuroscience and computer science, a combinatorial framework for object recognition was proposed having benefited from the advantages of both biologically-inspired HMAX_S architecture model for feature extraction and Extreme Learning Machine (ELM) as a classifier. HMAX model is a feed-forward hierarchical structure resembling the ventral pathway in the visual cortex of the brain and ELM is a powerful neural network, which randomly chooses hidden nodes and specifies analytically the single-hidden layer. ELM theories conjecture that this randomness may be true for biological learning in animal brains. It should be noted that the principle reason of using ELM is mainly as a result of its biological structure in order to imitate the biological object recognition system of mammalians and partly for its incredible speed which drastically lessens the runtime. Classification results are reported in Caltech101 dataset, at the focal point with its combinatorial framework serving considerable improvements over latest studies in both classification rate (96.39%) and the low runtime (0.417s).

1. Introduction

The mammalian's visual system outperforms the best computer vision systems that have been explored so far. The human vision is unique in many aspects including flexibility, speed, scalability and accuracy. On the oter hand one of the most challengeable problems in machine vision systems is invariance, which could have changed in illumination, location and scale. Another challenge is the speed in classification and the power of object recognition for which the human vision system is a great example. Hence constructing the system which could do as well as human visual system and imitate the processing flow and network structure of the visual cortex has always been regarded as an ultimate goal for machine vision systems.

The brain's visual cortex is composed of several regions which tend to be hierarchically organized. The information streams divide into two path ways through visual cortex, which called ventral and dorsal streams. The object recognition is performed in visual cortex. The information provided from the retina passes through the

*** Corresponding Author:**
Sahar Seifzadeh, MSc.
Address: *Young Researchers and Elite Club, Qazvin Branch, Islamic Azad University, Qazvin, Iran.*
Tel: *+98 (71) 32305471*
E-mail: *seifzadeh.sahar5@yahoo.com*

lateral geniculate nucleus and relayed through Thalamus to reach the visual cortex (V1,V2, V4) and the inferior temporal gyrus (IT). IT has a key role in recognizing invariant objects and to provide a major source of input to the prefrontal cortex (PFC), that connects perception to actions and memory. The popular hierarchical model in object recognition task is the HMAX model which is neural network model for image classification. This model mimics the hierarchical structure of the visual cortex in the feed-forward path of the ventral stream, starting from V1 (primary visual cortex), through V2, and V4 to IT. In the V1 area, simple features like oriented lines are tune, while V4 is for intermediate complexity features including geometric shapes like circle, rectangle, etc. Finally IT is tuned for complex object features like faces [1].The Visual cortex topography is illustrated in Figure 1.

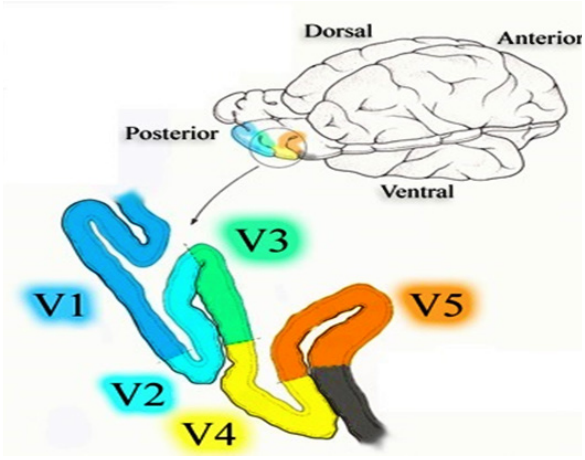

Figure 1. Visual cortex topography adapted from http://hubel.med.harvard.edu/book/b18.htm, CC attribution 3.0 license.

The two computational units defined in HMAX include simple and complex units which rely on the tuning properties of simple and complex cells and neural circuits found in V1.

The HMAX model

This neuroscientifically-inspired model was first introduced by Riesenhuber and Poggio [2]. The original version of HMAX describes a feed-forward hierarchical structure resembling the ventral pathway in the visual cortex. Another pathway of the visual cortex named 'dorsal' pathway, deals more with the visual motor aspects (i.e. where to reach for an object), which may code an object's location. It should be noted that such a hierarchical model follows a bottom-up approach in data

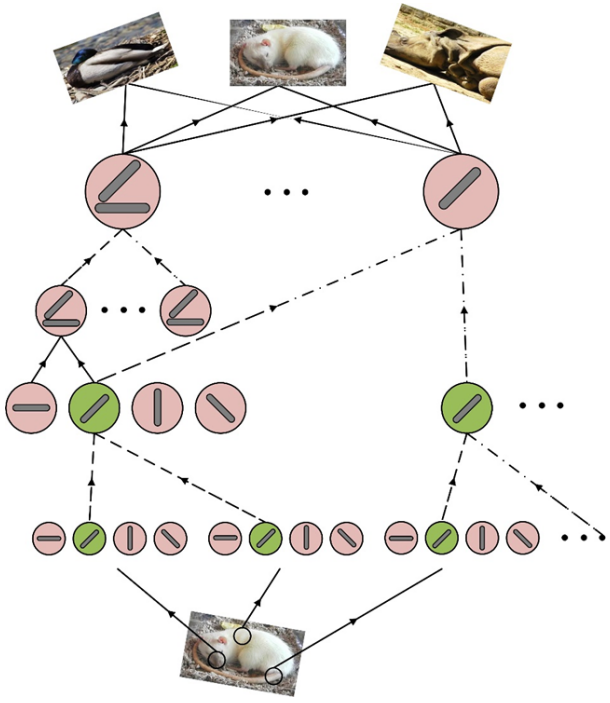

Figure 2. The basic HMAX model consists of a hierarchy of five levels. S1 refers to the classic cells in the primary visual cortex. The C1 unit models complex cells, which incorporate tolerance to shift and size. The S2 unit counter-pose the input with patterns learned upon training. In fact, C2 responses are computed by taking global maximum over all scales positioned for each S2 type over the entire S2 lattice.

processing using low-pass filters over images with different orientations, then it uses pooling operations over the filtered images. Figure 2 demonstrates the original HMAX model adapted by Riesenhuber and Poggio [2].

Serre et al. extended the original HMAX model. They applied two different recognition scenarios, primarily showing that an application to the semi supervised object recognition problem in clutter did not involve image scanning [3]. They also used StreetScences database to describe that scene understanding involved different rigid objects as well as texture-based ones.

Mutch et al. applied Gabor filters in all positions and scales, feature complexity and position/scale invariance built up by alternating template matching and maximum pooling operations, through which they refined the approach in several biologically-plausible ways [4]. Figure 3 illustrates the Mutch's model. Owing to the shallowness of Mutch's prototype, his method could not optimally tune the local image structures.

In another study however, authors proposed a framework for rapid object recognition and presented a feature-selective hashing scheme to model the memory association in IT cortex [5]. They examineed their experimental results on l000-class Amsterdam Library of Object Images (ALOI) dataset. Their framework was based on Mutch's improved HMAX model and they use the feature-selective hashing scheme with the Nearest Neighborhood (NN) as a classifier. In addition, in our previous work [6], a biologically-inspired model with feature selective hashing was proposed to recognize animals which was applied on KTH database containing 1239 images in 13 classes with photos taken from the animals' wildlife.

Theriault et al. presented a new extension of the HMAX model named "HMAX-S" [7]. Their model was based on the previous (S1-C1-S2-C2) architecture with two major improvements. First, against the RBF model which was used in a recent report [3] they redefined the S2 filter with normalized dot product and focused on increasing the complexity variable of the network by building these filters with richer information. Their second novelty was in the training phase of the prototype.

Extreme Learning Machine (ELM)

The learning speed is prominent criteria in classifiers while feed-forward neural networks retain slower learning speed than required. This has been regarded as the fundamental drawback of these systems in many applications over the past decades. The above might partly be due to using slow gradient-based learning algorithms for the training and tuning parameters with such slow-learning algorithms. To overcome these drawbacks, a fast-learning neural network called Extreme Learning Machine (ELM) was proposed by Guang-Bin et al. in 2005 [8] for Single-hidden Layer (or the so-called feature mapping) Feed-forward Neural networks (SLFN) which randomly chose hidden nodes and analytically specified the output weights. Contrary to the basics in neural networks in which all the hidden nodes in SLFN's need to be tuned, all the hidden nodes (or neurons) parameters are self-determining from the target functions

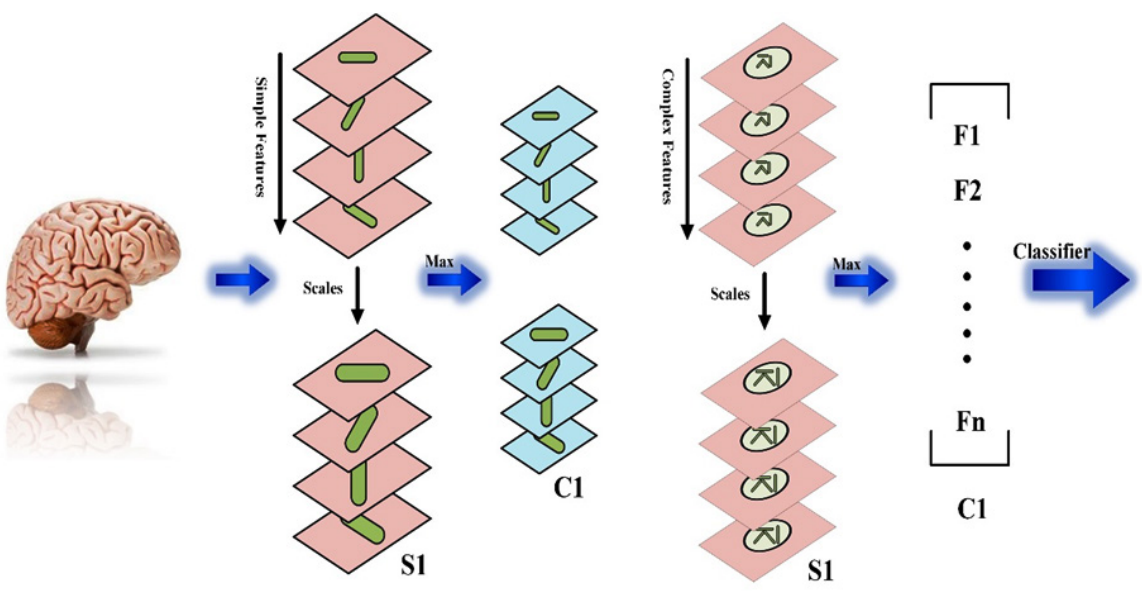

Figure 3. The Architecture of the network used in Mutch's method.

JAMSAT

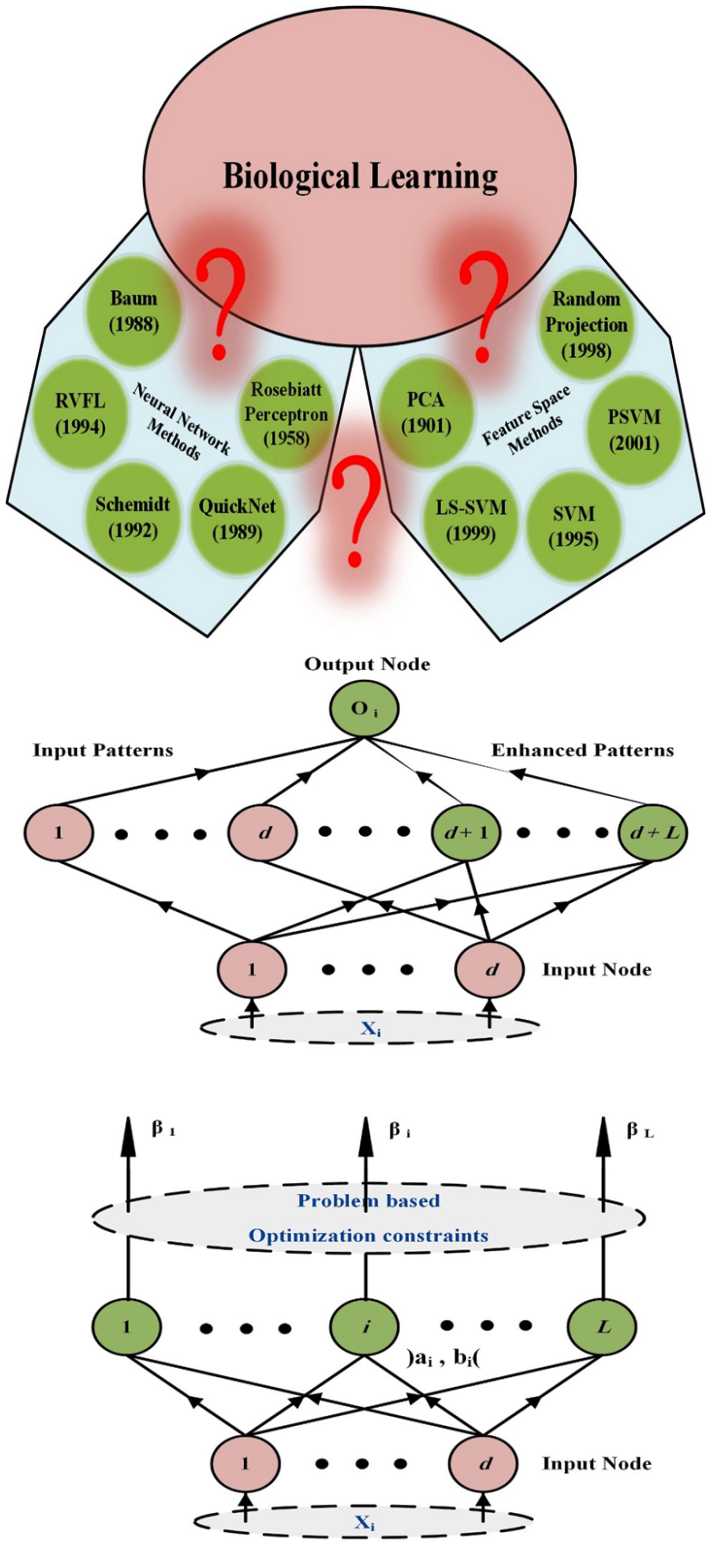

Figure 4. a) The missing relationship among artificial neural networks, feature space methods and biological learning mechanisms, b) ELM Feature Mapping. Consider each input data is a d-dimensional vector x ¼ ½x1; …; xi; …; xd T, through a single hidden layer feed-forward neural network. The ELM tends to map the data into L-dimensional ELM feature space (hidden layer feature space). H and L are the number of the hidden nodes used in the feature mapping process.

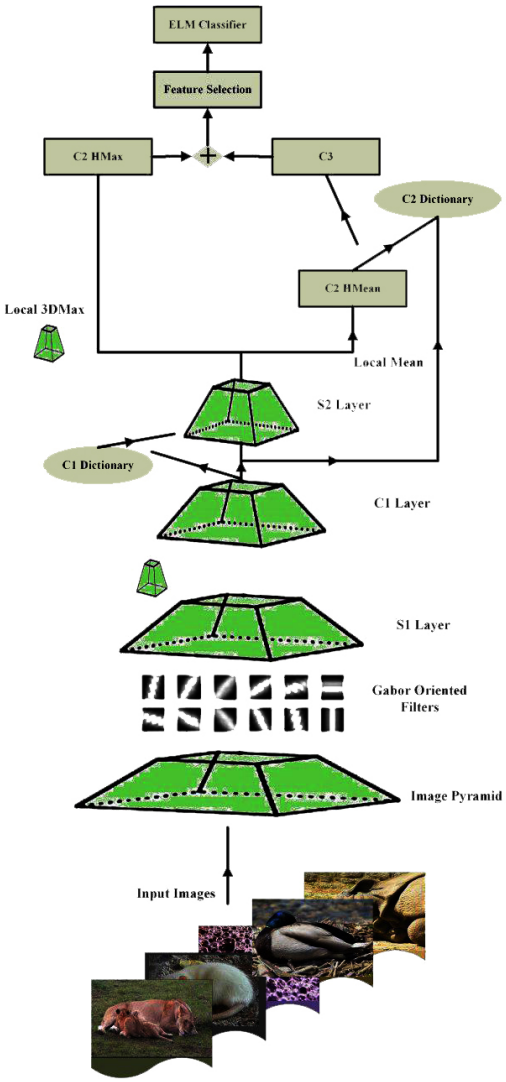

Figure 5. Our proposed model.

or the training datasets. The ELM theories surmise that this random action might be true for biological learning in animal brains. ELM is also efficient in batch, sequential and incremental learning. Moreover, ELM has been successfully used in image processing, signal processing, brain-computer interface, biometrics, etc.

From the mathematical standpoint, researches on the approximation capabilities of feed-forward neural networks has focused on two major aspects i.e. the global approximation on compact input sets and the finite set of training samples. An earlier research [9] demonstrated that if the Activation Function (AF) is bounded, no matter constant and continuous, the continuous mappings can be approximated in measure by neural networks over the compact input sets. In 1993 Leshno et al. [10] improved this proposed model and proved that feed-forward network with a non-polynomial activation function could approximate continuous functions.

In Guang-Bin et al's recent paper [11], they investigated ELMs in three aspects i.e. random neurons, random features and kernels. They demonstrated that in theory, ELMs (with the same kernels) tend to improve Support Vector Machine (SVM) and its variants in both regression and classification applications with much easier implementation. This paper showed that before ELM has been proposed, the relationship among different learning methods was not clear and ELM aims to provide a biologically-inspired simple and efficient conjunct learning framework to fill the gap between artificial learning methods and biological learning mechanism (Figures 4a and 4b).

2. Materials and Methods

One of the most capable approaches to indicate the efficiency of the specific classifier is by displaying results in confusion matrix. The main flowchart of proposed work is depicted in Figure 5.

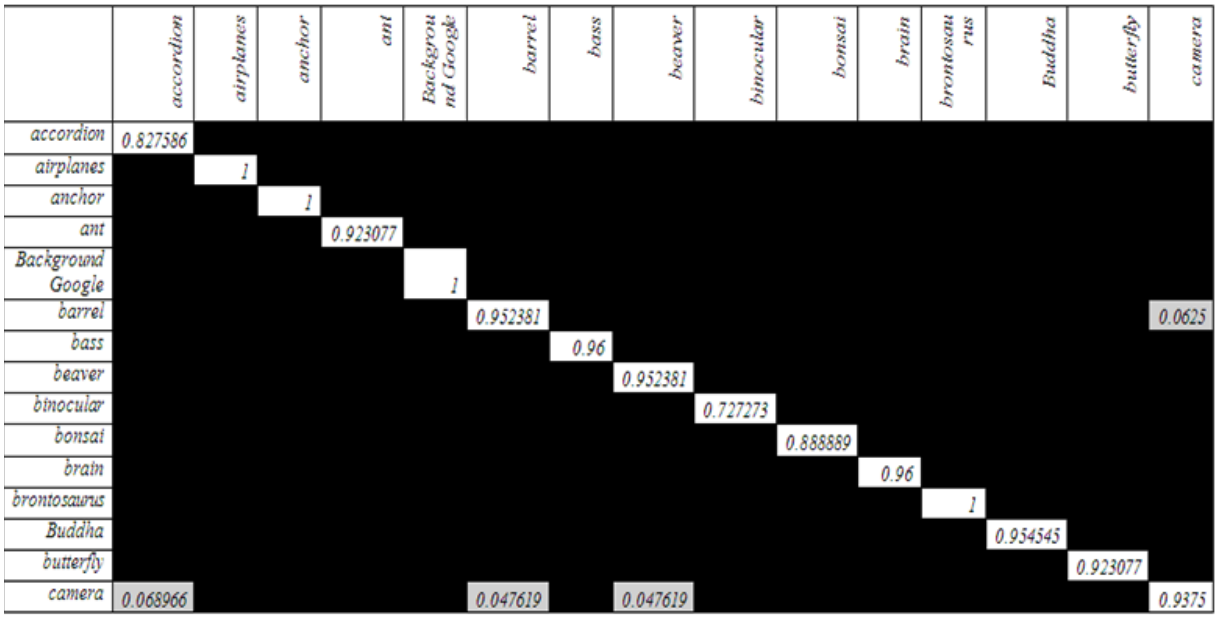

Figure 6. Our confusion matrix in 15 example classes.

As demonstrated in Figure 6, calculated confusion matrix from our result is shown more closely, where 15 classes are selected as examples. Some images assigned to their classes exactly and some of them classified with high accuracy to their relevant classes. Figure 7 demonstrates image categories which were exactly assigned to their respective classes.

In the present investigation, we employed a Cori-5 2.5-GH PC and MATLAB software. The total runtime was 0.417318 seconds which represented the high speed classification of ELM in [3], while they used an 8-core PC at 3.47 MHz and the runtime was approximately 1 hour for Caltech101. The number of our hidden neurons was 500 and the implemented activation function was sigmoid. The classification rate to test our data was 96.39%. The above rate for our train data was 90.667%. Table 1

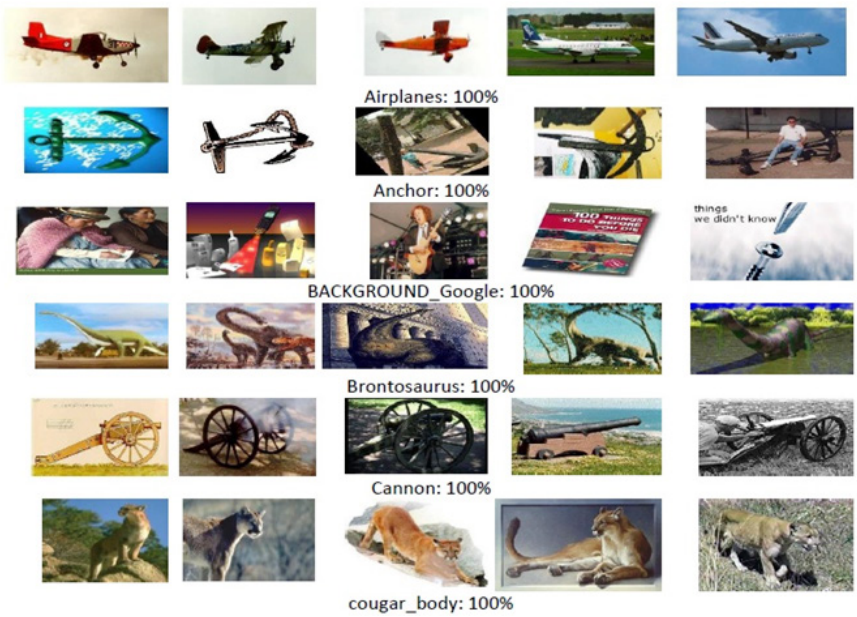

Figure 7. Our best six classification accuracies on Caltech101.

Table 1. Classification results in average precision on caltech101.

HMAX Model	30 Images
Our Model	96.39%
Serre et al.	42%
Mutch and Lowe	54%
Lecun et al.	54±1.0%
Lee et al.	65.4±0.5%
Jarret et al.	65.6±1.0%
Zeiler et al.	66.9±1.1%
Fidler et al.	66.5%
Zeiler et al.	71.0±1.0%
Theriault et al.	61±0.5%
Theriault et al.	0.97±76.32

represents the present work methodology as compared to other conducted recent models. There was a classification percentage based on 30 images with the same dataset (caltech101) suggesting that our model acquired the most favorable accuracy among all recent studies.

3. Results

Unlike superficial learning models, deep learning represent learning at multiple levels of representation and abstraction which helps with interpretation of the data. Further to basic neuroscience insights, some theoretical analyses from machine-learning provide support for the argument that deep models are more compact and stentorian than the superficial ones in representing most learning functions.

4. Discussion

Our proposed HMAX model was inspired by HMAX_S [7] where 4 layer architecture of HMAX_S was used for the purpose of feature extraction. This feature extraction model mimics basic alternating convolution/pooling scheme. The S1 units are the first processing step. This step corresponds to the classic cells in the primary visual cortex (V1). Their receptive field and summation behavior is modeled by Gabor functions. Their input is a grayscale image, and their output image is convolved using their specific Gabor filter. The C1 unit models complex cells, which incorporate tolerance to shift and size. Complex cells with larger receptive fields tend to respond to oriented bars or edges anywhere within their receptive field (tolerance to position). Meanwhile, the S2 unit counter-pose the input with patterns learned during the training. In the ideal case, such patterns are characteristic parts describing an object, e.g. the hand and eye of a human. Finally, C2 responses are computed by taking a global maximum overall scales and positions for each S2 type over the entire S2 lattice [3].

Classification is considered as the final step whereby image classification task is combination of extracted features from the images as well as classifying them into relevant classes. The prominent goal of this task is assigning images to their related classes. It is worth noting that selecting the appropriate feature extraction and classification algorithms is among the key and substantial tasks. In this paper, HMAX-s feature extractor for specific features from the images was used as outlined above. Additionally, ELM classifier was employed to classify our images from Caltech101 dataset consisting of 101 classes from different objects in the environment.

5. Conclusion

The present investigation proposed a novel architecture of HMAX model which acquired more accuracy for biological feature extraction. In addition, using the ELM as a classifier for assigning each picture to its appropriate class, yielded notable results both for the accuracy

and runtime on Caltech101 dataset. Taken together, according to our classification results, it can be inferred that HMAX model and ELM neural network, which are collectively regarded as biologically-inspired models, work effectively in serving object recognition purposes.

Acknowledgements

The authors wish to thank Dr. Mohammad Nami for his comments that greatly improved the manuscript and the Research Council of Kermanshah University of Medical Sciences (Grant Number: 94191) for the financial support.

Conflict of Interest

The authors declared no conflict of interests.

References

[1] Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. Journal of Physiology. 1962; 160(1):106-54. doi: 10.1113/jphysiol.1962.sp006837

[2] Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. Nature Neuroscience. 1999; 2(11):1019–25. doi: 10.1038/14819

[3] Serre T, Wolf L, Bileschi S, Riesenhuber M, Poggio T. Robust object recognition with cortex-like mechanisms. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2007; 29(3):411–26. doi: 10.1109/tpami.2007.56

[4] Mutch J, Lowe DG. Object class recognition and localization using sparse features with limited receptive fields. International Journal of Computer Vision. 2008; 80(1):45–57. doi: 10.1007/s11263-007-0118-0

[5] Lee YJ, Tsai CY, Chen LG. A cortex-like model for rapid object recognition using feature-selective hashing. Paper presented at: The 2011 International Joint Conference; 2011 31 Jul - 5 Aug; San Jose, CA, USA.

[6] Seifzadeh S, Faez K. A cortex-like model for animal recognition based on texture using feature-selective hashing. Paper presented at: The 2014 Iranian Conference on Intelligent Systems (ICIS); 2014 Feb 4-6; Bam, Iran.

[7] Theriault C, Thome N, Cord M. Extended Coding and Pooling in the HMAX Model. IEEE Transactions on Image Processing. 2013; 22(2):764–77. doi: 10.1109/tip.2012.2222900

[8] Huang GB. An insight into extreme learning machines: random neurons, random features and kernels. Cognitive Computation. 2014; 6(3):376–90. doi: 10.1007/s12559-014-9255-2

[9] Lee H, Grosse R, Ranganath R, Ng AY. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. Paper presented at: The 26th Annual International Conference on Machine Learning; 2009 June 14-18; Montreal, Canada.

[10] Leshno M, Lin VY, Pinkus A, Schocken S. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. Neural Networks. 1993; 6(6):861–7. doi: 10.1016/s0893-6080(05)80131-5

[11] Guang-Bin Huang, Hongming Zhou, Xiaojian Ding, Rui Zhang. Extreme learning machine for regression and multiclass classification. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics). 2012; 42(2):513–29. doi: 10.1109/tsmcb.2011.2168604