HMAX Model: A Survey

Chang Liu
Department of Computer Science
Tsinghua University
Beijing, China
cliu13@mails.tsinghua.edu.cn

Fuchun Sun
Department of Computer Science
Tsinghua University
Beijing, China
fcsun@mail.tsinghua.edu.cn

***Abstract*—HMAX model is a bio-inspired feedforward architecture for object recognition, which is derived from the simple and complex cells model in cortex proposed by Hubel and Wiesel. As a hierarchical bio-based recognition model, HMAX captures the properties of primate cortex with alternated S layers and C layers, corresponding to simple cells and complex cells respectively. Although constrained by biological factors, HMAX shows satisfying performance in different fields when competing with other state-of-the-art computer vision algorithms. Insightful ideas and methods have been developed for this hierarchical model, which advances the progress of HMAX model. This paper reviews the origin of this model, as well as the improvements and modifications based on this model.**

Keywords*—*bio-inspired, max pooling, hierarchical model, object recognition

I. Introduction & Background

Within recent decades, lots of efforts have been made for object recognition, yet few clues have been found regarding to the essence of how human visual cortex works in such tasks. One major issue during the recognition procedure is to understand how the human recognition system gives outstanding performance in terms of selectivity, the ability to select specific object from backgrounds or clutter, and invariance, the ability to remain selectivity despite the changes in scale and rotation. Despite the fact that human recognition system outperforms all the state-of-the-art machine algorithms or systems, some valuable ideas present inspiring explorations of the human recognition mechanism.

HMAX model [4] belongs to one of these biologically based object recognition frameworks, which is derived from the simple and complex cells model in cortex proposed by Hubel and Wiesel [5,6]. It should be noted that unlike other object recognition models or feature extraction models proposed and optimized for specific tasks, HMAX model differs from them in its creativeness to incorporate biological findings into computer algorithms and systems. As a hierarchical bio-based recognition model, HMAX captures the property of primate cortex: the first hundreds milliseconds of visual processing follow a feedforward hierarchical mechanism, and the receptive fields of the neurons along the pathway tend to get larger with respect to their optimal stimuli [7,37,38,39,40]. Selectivity and invariance are achieved by alternatively increased in its hierarchical layers with different methods. The core of the proposed model [4] is to extend the classical models of complex cells built from simple cells by introducing a MAX operation, which pools the maximum response from afferents. Serre et al. extends the original HMAX model by learning a vocabulary of visual features from images in the higher layers, as well as applying it to the real-world object categories [8, 9]. Constrained by biological factors, the extended HMAX model performs quite well when competing with the state-of-the-art models. Based on these models, many enhanced models emerge with improvements or novel ideas in different application areas, and have achieved satisfying results in each domain.

This article aims at providing a thorough survey of the algorithms, models and applications that improved and enhanced the HMAX model, as well as developments and future explorations of this area, which has not been done in the literature to our knowledge. The paper is structured as follows: Section II will present the general idea of the original HMAX model [4], which mainly reviews the biological-feasible features of the proposed model. Section III gives a detailed introduction and implementation of the enhanced HMAX model [8,9], which founded the basis of this research

978-1-4799-1959-8/15/$31.00

area. Algorithms and models, derived from the enhanced HMAX model, will be categorized and summarized in Section IV in terms of the vertical impartments and adjustments of the model as well as the horizontal applications and expansion in different areas. Section V concludes the paper and points out further possible exploration directions of the literature.

II. ORIGINAL HMAX MODEL

Although it is the fact that computer recognition systems are task-oriented and shall be evaluated in terms of precision, real-time capability, robustness and so on, the implementation of such algorithms shall not neglect the factor of the biological feasibility. The former algorithms and models are basically data-driven, while the latter emphasizes the recognition-driven perspective of the system. When jumped out of the task circle, interesting and inspiring ideas can enlighten us to improve existed models.

Accomplishments have been made in the area of how human visual cortex works in real-time recognition tasks [2,3]. Even though the natural mechanism of how the human recognition system works is still under cover, possible presumptions have provided reasonable explanations and perspectives. In the model of simple cells, which responds primarily to oriented edges and bars of particular orientations, and complex cells, which responds similarly with simple cells yet have a degree of spatial invariance [5], it is proposed that simple cells with neighboring receptive fields tend to feed into the same complex cell, which causes the phase-invariance of the complex cell. Intuitively derived from this idea, a simple hierarchical feedforward architecture is created with layered simple cells and complex cells, where the responses of the simple cells are pooled over into the receptive fields of complex cells [4]. Figure 1 gives a sketch of this original HMAX model. Two different methods are applied to achieve both selectivity and invariance, which are template matching (solid lines) and MAX pooling (dash lines). Given an input pattern, it is firstly filtered with S1 layer, which performs scanning and matching with handcrafted templates to mimic the cell-like receptive fields. After which a MAX operation is performed on an array of S1 responses with a softmax approximation in (1), where performs linear summation when $p = 0$ and MAX operation when $p \to \infty$.

$$c_i^1 = \sum_j \frac{\exp(p \cdot |s_j|)}{\sum_k \exp(p \cdot |s_k|)} s_j \tag{1}$$

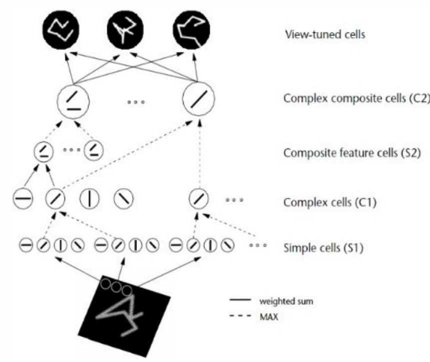

Fig.1. Origin HMAX model structure, originally presented in [4]

In spite of other optional pooling mechanisms in hierarchical models, MAX pooling method yields more convincible explanation. It is discussed in detail about the advantages of MAX pooling method in [4], and the reasons of such a method can be summarized as: MAX pooling method ensures that the strongest afferent determines the postsynaptic response, thus could signal the best match of any part of the stimulus to the afferent's preferred feature, avoiding the case in which signals may be mixed up by different stimuli with combining afferents [4]. Theoretical and experimental supports can be found in [11-13].

III. ENHANCED HMAX MODEL

Firstly introduced in the literature, MAX pooling method proves to preserve good performance as well as biological supports. However, the features used in the standard model (original HMAX model) are simple and static. The higher features of the model should be learnt from visual experience instead, which is the core idea of the enhanced HMAX model [8, 9].

Similar with the standard model, the enhanced HMAX model adapts a four-layered hierarchical structure, with simple S-units layer alternated with complex C-unit layer. Yet, several adjustments have been made to suit the model for real-world object recognition.

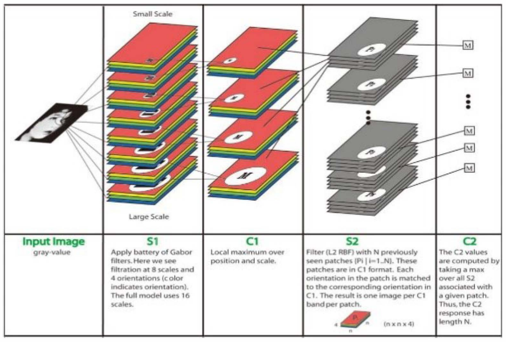

Fig.2. enhanced HMAX model, originally presented in [8]

1) S1 layer: Given an input image, it is firstly filtered in S1 layer with Gabor functions [18] with the following form (2):

$$F(x,y) = \exp\left(-\frac{(x_0^2+\gamma^2 y_0^2)}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda}x_0\right)$$
$$\text{s.t. } x_0 = x\cos\theta + y\sin\theta$$
$$\text{and } y_0 = -x\sin\theta + y\cos\theta \qquad (2)$$

The parameters are defined as: σ the effective width of Gaussian function, $\theta$the orientation, and λ the wavelength. Parameters are tuned and determined with regard to the properties of V1 simple cells. The S1 responses form the 8 bands, within each has four orientations and two scales [9]. Each band has two filter sizes, thus leads to a total number of 64 different S1 responses, with four orientations ($\theta = 0°, 45°, 90°, 135°$).

2) C1 layer: After Gabor filtering in S1 layer, 64 filtered images are obtained corresponding to the 64 distinct receptive fields, and thus achieves selectivity. The responses of C1 layer are pooled from S1 layers within the same orientation and same band, but among the two scales, which leads to a total number of 32 C1 responses. The pooling method still adopts the local maximum operation, and thus increases the invariance of size and rotation. Note that the pooling grid and overlap increase with the filter size.

3) S2 layer: S2 layer is where the learning stage occurs. Before computing the S2 responses, a set of N prototypes $\boldsymbol{P}_i$(dictionaries) are learned from a positive training set. These prototypes are obtained via sampling patches from the responses of C1 layer of all the four orientations. After the prototypes are constructed, the response r of a S2 unit can be defined as:

$$r = \exp(-\beta\|\mathbf{X} - \mathbf{P_i}\|^2) \qquad (3)$$

where parameter β determines the sharpness of the tuning and $\mathbf{X}$ is the input image patch.

4) C2 layer: The responses of C2 units are computed by pooling the maximum from S2 responses in all bands and all orientations, which leads to scale and shift invariant features. The final feature is a vector with length N, where N is the number of the learned prototypes in the learning stage.

The final C2 features possess selectivity as well as invariance to rotation and scale, and can be fed into further classifiers such as SVM. Experiments of real-world object recognition in clutter and without clutter are conducted, as well as object recognition of texture-based objects [9]. Results show satisfying performance against benchmark models of the period.

The contribution of the enhanced HMAX model lies in that it not only inherits the biological features from the origin HMAX model, but also incorporates a learning process to obtain scale and shift invariance C2 features, which are proved feasible and competitive in real-world object recognition.

Some issues or feasible future explorations remain in the enhanced HMAX model. For example, in terms of the model architecture, the depth of the hierarchical structure can be modified as either lessened to C1 layer, or lengthened to higher levels such as S3 and C3 based on specific tasks. Beyond feedforward structure, some feedback mechanism can also be applied in the higher layer of the model to gain better performance, since such mechanism has been proved existent along the visual pathway [14-17]. Moreover, not all the prototypes obtained in the learning procedure provide effective information since those prototypes are sampled randomly from the C1 responses, and chances of selecting background patches are relatively high. The massive sampling procedure causes the high-cost computation issue as well, which makes the model unlikely practicable for real-time application.

IV. MODIFICATION & APPLICATION

Proved to be competing and well-performed in real-world object recognition tasks [9], this hierarchical MAX pooling architecture has been widely used in different areas. Either other methods are combined and incorporated with the HAMX model, or modifications are made based on the original architecture. This section reviews the most popular and useful methods, classifying them with respect to the modifications made by these methods.

A. Sparse-HMAX Models

One major issue of the enhanced HMAX model is the calculation cost. Because of the large number of the prototypes selected in the learning stage, for each input image, a template matching needs to be computed for each prototype, leading to the incapability of the model in real-time applications. To avoid the massive computation problem, one plausible way is to reduce the inputs for the layers within the architecture. Some biology research works have provided assistance with this cost-reduction issue.

Recent biology observations have found that neurons in visual pathway do not fire for the most of time, but fire only occasionally, based on which sparse coding is implemented to learn the properties of the receptive fields of V1 cells and V4 cells[20, 21, 22, 31]. To introduce the sparsity constraint is important for learning models derived from biology in terms of the statistics of natural images [19]. It is natural and biological plausible to combine the sparse-based technique with the bio-inspired HMAX model [10,24,25,26,29,30].

Mutch et al. proposed a similar model like HMAX model where sparsity constrain is adopted [10,30]. The inputs for the S2 layer are sparsified to reduce the computation cost. In the original HMAX model, the S2 features are calculated by computing the responses using all the corresponding prototypes selected in C1 layer, which contains all the four orientations $(0°, 45°, 90°, 135°)$ in equation (4). The inputs for S2 are sparsified by selecting the most preferred orientation of one position, discarding all the other three orientations. Thus the calculation is cut down to 1/4, where the C1 prototype contains 16 C1 unit values, rather than 64$(0°, 45°, 90°, 135°)$ C1 unit values yet preserving the stimulus feature at the same time.

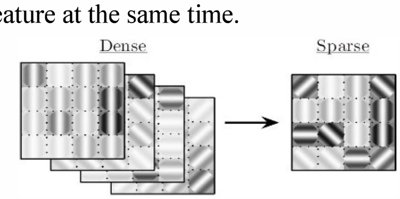

Fig.3. Dense and Sparse S2 features, originally presented in [10]

Though the method proposed in [10,30] achieves good performance with some bio-inspired improvements, the sparsity constraint used in the model lacks the incorporation with sparse coding mechanism, and also lacks the learning ability to extract higher level features. Hu et al. proposed a model which combines the standard sparse coding (SSC)/independent component learning (ICA) with HMAX model. For each S layer in the model, the feature maps are learnt using SSC/ICA based on the outputs of previous C layer, after which the MAX pooling method is implemented to obtain the C layer feature maps [24]. The sparse regularized architecture is shown in figure 4.

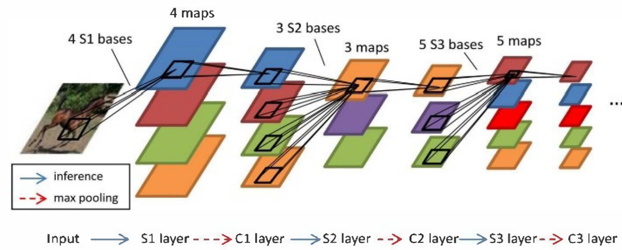

Fig.4. Sparse HMAX with six layers, originally presented in [24]

Another enlightening finding of biology in the visual pathway is the lateral inhibition phenomenon [27,28], where the activated neurons tend to suppress the firing of their neighbor neurons. Such phenomenon offers a new perspective of reduce the inputs for each layer. Within an image patch of size $n \times n$ of some layer in the HMAX architecture, the maximum response R_{max} and the minimum response $R_{\min}$ can be found to determine the inhibition level along with an inhibition parameter h, as showed in equation (4) [10,29,30,32]. Any response value that satisfying the condition below will be set to zero to reduce the amount of inputs.

$$R < R_{min} + h(R_{max} - R_{\min}) \quad (4)$$

The lateral inhibition can also be implemented in the way of gradient filtration [26]. The horizontal and vertical gradients over the original S1 features are computed and further used in equation (5) as sparsification constraint. Similar as equation (4), only points that satisfy this condition will be retained.

$$|F_{x(i)}| + |F_{y(i)}| \geq \frac{\alpha}{n} \sum_{k=1}^{n} \left(|F_{x(k)}| + |F_{y(k)}| \right) \quad (5)$$

where F_x and F_y denotes the horizontal and vertical gradient respectively.

With the incorporation of sparsity, HMAX model can achieve better and faster performance with the reduction in the inputs of each layer, while maintaining the support from the biology evidence at the same time.

B. Architecture Enhanced Models

The HMAX model has a feedforward four-layer architecture with alternated S and C layers. Some algorithms and models try to improve the HMAX model in terms of its architecture, either the feedforward architecture, or the four-layer architecture. Naturally, incorporating a feedback mechanism to the HMAX model can provide better performance since feedback has proved existence in biology [14-17], and several methods have carried out the feedback operation into the HMAX model [26,33,34,35,36]. Though the recognition tasks of human being are accomplished mostly in the very first 100-200 ms of the visual perception [37], incorporating a feedback mechanism can to some extent help with filtering the learning prototypes to focus on the informative patches, rather than some uninformative background patches, since the learning procedure causes huge computation cost with blindly selecting prototypes.

Motivated by these ideas, Huang et al. proposed a feedback scheme to replace the random patch selection stage, which is derived from the idea that different patches should play different roles in classifying objects [26]. This feedback procedure can find the most discriminative patches, along with the different contributions of each patch. A Boosting method is adopted with weak SVM classifiers to enhance the previously misclassified samples by adjusting the Boosting weights. The architecture of the model with feedback is shown in figure 5.

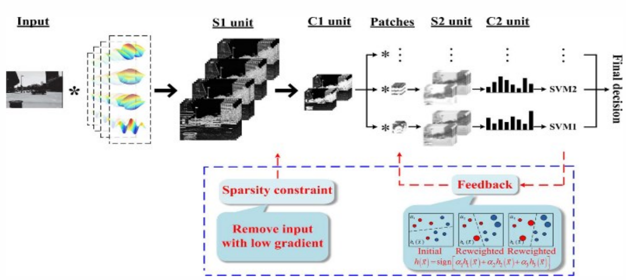

Fig.5. Framework with feedback, originally presented in [26]

Other methods which incorporate feedback mechanism in HMAX model exist in the literature. Salvador et al. constructed a HMAX-like architecture with feedback information from multiple parents by including high-level connections that reproduce modulatory effects, where Bayesian methods were also combined [35]. Rajaei et al. simulated the feedback from complex cells to simple cells to learn informative intermediate-level visual features based on HMAX model [33]. Yu et al. placed C2 coefficients around the objects, after which a clustering algorithm (K-means) to discard the uninformative information on the irrelevant object [34]. Salvador et al. employs three different types of units, of which the feedback predictive signal fS and fC units are encoded [41].

Extended from the layer number of the HMAX model, some methods increase the layers from the original four (two S layers and two C layers) to more layers with higher-level features. Hu et al. introduced a high-layered architecture to learn high level features, in which S3 and C3 stage were included, and noted that such layers can be further increased with demands [24]. The high-layered architecture can refer to figure 4. Similarly, Jhuang et al. also extended the model with S3 and C3 stages to capture the feature of motion processing in visual cortex [25]. Whereas some methods use only the earlier stages of HMAX model, yet achieves good or better performance. Guo et al. employed only S1 and C1 features for human age estimation [42], while similar methods are adopted in [43], [44] in the field of scene classification and face verification, respectively.

C. Feature Selection & Filtering Modification

It has been noted that the learning stage in [8] brings in the issue of uninformative prototypes, and as noted in [9], the usage of C1 features or C2 features can vary with respect to different tasks. Obviously, implemented a feature selection operation or filtering modification, the computation cost and the uninformative patches would reduce. Christian et al. provides a flexible description of object regions to improve the way the local filters at the first level are integrated [1]. The filters in [1] are not limited to a single scale, and thus can represent mid-level structures containing multiple scales inside the same spatial neighborhood. The model structure is shown in figure 6, where L1, L2 corresponds to S1 and C1 layer respectively.

The S2 filters can also be defined as a normalized dot product in the following form:

$$Y_i = \frac{\langle P_i|X\rangle}{||P_i||||X||} \tag{6}$$

Adopted in [45], such method not only speeds up the computation but also improves the classification scores since a normalized dot product is invariant to light intensity

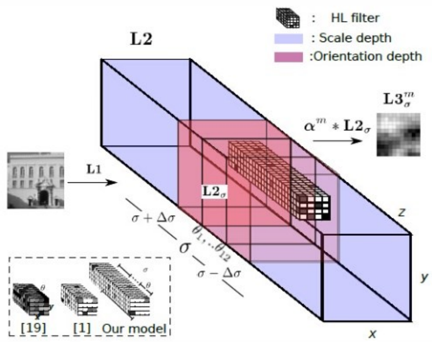

Fig.6. Multiple scales filters, originally presented in [1]

changes, whereas in [8,9] is not. Other filtering or feature selection techniques include sorting the variances of C2 coefficients to simply conduct patch selection [34], using a feature selection technique based on SVM normal while fitting SVM hyperplanes [30], or computing C2 vectors of a small subset of the training frames by matching them to all the motion prototypes, after which a identifying stage is performed to select features [25] and so on [42,46].

V. Conclusion

Biologically inspired, HMAX proved to be capable of competing with state-of-the-art computer vision algorithms. First proposed by Riesenhuber et al.[4], the MAX-pooling operation came into literature with plausible biology evidences; enhanced by Serre et al.[8,9] and Mutch et al. [10,30], HMAX proved to be useful in real-world object recognition. Lots efforts and attempts have been made based on this cortex-like recognition mechanism, and have achieved satisfying results.

This paper gives a thorough review in the literature of HMAX model, landmark algorithms and models are detailed, with brief summary and classification of the works based on this model. Future work of this hierarchical model will focus on filtering methods, feature selection mechanisms and enhancement of pooling method. Given further biology findings in the future, it is believed this bio-inspired model could explore the essence of visual recognition of human being in a new level.

References

[1] Theriault, Christian, Nicolas Thome, and Matthieu Cord. "Extended coding and pooling in the HMAX model." Image Processing, IEEE Transactions on 22.2 (2013): 764-777.

[2] Orban, Guy A. "Higher order visual processing in macaque extrastriate cortex." Physiological Reviews 88.1 (2008): 59-89.

[3] DiCarlo, James J., Davide Zoccolan, and Nicole C. Rust. "How does the brain solve visual object recognition?." Neuron 73.3 (2012): 415-434.

[4] Riesenhuber, Maximilian, and Tomaso Poggio. "Hierarchical models of object recognition in cortex." Nature neuroscience 2.11 (1999): 1019-1025.

[5] Hubel, David H., and Torsten N. Wiesel. "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex." The Journal of physiology 160.1 (1962): 106.

[6] Hubel, David H., and Torsten N. Wiesel. "Receptive fields and functional architecture of monkey striate cortex." The Journal of physiology 195.1 (1968): 215-243.

[7] Kandel, Eric R., James H. Schwartz, and Thomas M. Jessell, eds. Principles of neural science. Vol. 4. New York: McGraw-Hill, 2000.

[8] Serre, Thomas, Lior Wolf, and Tomaso Poggio. "Object recognition with features inspired by visual cortex." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 2. IEEE, 2005.

[9] Serre, Thomas, et al. "Robust object recognition with cortex-like mechanisms." Pattern Analysis and Machine Intelligence, IEEE Transactions on 29.3 (2007): 411-426.

[10] Mutch, Jim, and David G. Lowe. "Multiclass object recognition with sparse, localized features." Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. Vol. 1. IEEE, 2006.

[11] Riesenhuber, M. & Poggio, T. in Advances in Neural Information Processing Systems Vol. 10 (eds. Jordan, M., Kearns, M. & Solla, S.) 215–221 (MIT Press, Cambridge, Massachusetts, 1998).

[12] Wang, Gang, Manabu Tanifuji, and Keiji Tanaka. "Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging." Neuroscience research 32.1 (1998): 33-46.

[13] Logethetis, Nikos. "Object vision and visual awareness." Current opinion in neurobiology 8.4 (1998): 536-544.

[14] Casagrande, V. A. "A third parallel visual pathway to primate area V1." Trends in neurosciences 17.7 (1994): 305-310.

[15] Markov, Nikola T., et al. "Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex." Journal of Comparative Neurology 522.1 (2014): 225-259.

[16] Murphy, P. C., and A. M. Sillito. "Corticofugal feedback influences the generation of length tuning in the visual pathway." (1987): 727-729.

[17] Bullier, Jean, et al. "The role of feedback connections in shaping the responses of visual cortical neurons." Progress in brain research 134 (2001): 193-204.

[18] Jones, Judson P., and Larry A. Palmer. "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex." Journal of neurophysiology 58.6 (1987): 1233-1258.

[19] Olshausen, Bruno A. "Emergence of simple-cell receptive field properties by learning a sparse code for natural images." Nature 381.6583 (1996): 607-609.

[20] Olshausen, Bruno A., and David J. Field. "Sparse coding with an overcomplete basis set: A strategy employed by V1?." Vision research 37.23 (1997): 3311-3325.

[21] Baddeley, Roland, et al. "Responses of neurons in primary and inferior temporal visual cortices to natural scenes." Proceedings of the Royal Society of London. Series B: Biological Sciences 264.1389 (1997): 1775-1783.

[22] Carlson, Eric T., et al. "A sparse object coding scheme in area V4." Current Biology 21.4 (2011): 288-293.

[23] Waydo, Stephen, and Christof Koch. "Unsupervised learning of individuals and categories from images." Neural computation 20.5 (2008): 1165-1178.

[24] Hu, Xiaolin, et al. "Sparsity-Regularized HMAX for Visual Recognition." PloS one 9.1 (2014): e81813.

[25] Jhuang, Hueihan, et al. "A biologically inspired system for action recognition." Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. Ieee, 2007.

[26] Huang, Yongzhen, et al. "Enhanced biologically inspired model for object recognition." Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 41.6 (2011): 1668-1680.

[27] Amari, Shun-ichi. "Dynamics of pattern formation in lateral-inhibition type neural fields." Biological cybernetics 27.2 (1977): 77-87.

[28] Blakemore, Colin, and Elisabeth A. Tobin. "Lateral inhibition between orientation detectors in the cat's visual cortex." Experimental Brain Research 15.4 (1972): 439-440.

[29] Hamidi, Mandana, and Ali Borji. "Invariance analysis of modified C2 features: case study—handwritten digit recognition." Machine Vision and Applications 21.6 (2010): 969-979.

[30] Mutch, Jim, and David G. Lowe. "Object class recognition and localization using sparse features with limited receptive fields." International Journal of Computer Vision 80.1 (2008): 45-57.

[31] Barth AL, Poulet JF (2012) Experimental evidence for sparse firing in the neocortex. Trends in Neurosciences 35: 345–355.

[32] Woodbeck, Kris, Gerhard Roth, and Huiqiong Chen. "Visual cortex on the GPU: Biologically inspired classifier and feature descriptor for rapid recognition." Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on. IEEE, 2008.

[33] Rajaei, Karim, et al. "A stable biologically motivated learning mechanism for visual feature extraction to handle facial categorization." PloS one 7.6 (2012): e38478.

[34] Yu, Guoshen, and J-J. Slotine. "Fastwavelet-based visual classification." Pattern Recognition, 2008. ICPR 2008. 19th International Conference on. IEEE, 2008.

[35] Dura-Bernal, Salvador, Thomas Wennekers, and Susan L. Denham. "Modelling object perception in cortex: Hierarchical bayesian networks and belief propagation." Information Sciences and Systems (CISS), 2011 45th Annual Conference on. IEEE, 2011.

[36] Dura-Bernal, Salvador, Thomas Wennekers, and Susan L. Denham. "The role of feedback in a hierarchical model of object perception." From Brains to Systems. Springer New York, 2011. 165-179.

[37] Thorpe, Simon, Denis Fize, and Catherine Marlot. "Speed of processing in the human visual system." nature 381.6582 (1996): 520-522.

[38] Thorpe, Simon J., and Michèle Fabre-Thorpe. "Seeking categories in the brain." SCIENCE-NEW YORK THEN WASHINGTON- (2001): 260-262.

[39] VanRullen, Rufin, and Christof Koch. "Visual selective behavior can be triggered by a feed-forward process." Journal of Cognitive Neuroscience 15.2 (2003): 209-217.

[40] Keysers, Christian, et al. "The speed of sight." Cognitive Neuroscience, Journal of 13.1 (2001): 90-101.

[41] Dura-Bernal, Salvador, Thomas Wennekers, and Susan L. Denham. "The role of feedback in a hierarchical model of object perception." From Brains to Systems. Springer New York, 2011. 165-179.

[42] Guo, Guodong, et al. "Human age estimation using bio-inspired features." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.

[43] Song, Dongjin, and Dacheng Tao. "Biologically inspired feature manifold for scene classification." Image Processing, IEEE Transactions on 19.1 (2010): 174-184.

[44] Ma, Bingpeng, Yu Su, and Frederic Jurie. "Covariance descriptor based on bio-inspired features for person re-identification and face verification." Image and Vision Computing 32.6 (2014): 379-390.

[45] Theriault, Christian, Nicolas Thome, and Matthieu Cord. "HMAX-S: Deep scale representation for biologically inspired image categorization." Image Processing (ICIP), 2011 18th IEEE International Conference on. IEEE, 2011.

[46] Ning, Huazhong, et al. "Hierarchical space-time model enabling efficient search for human actions." Circuits and Systems for Video Technology, IEEE Transactions on 19.6 (2009): 808-820.