Object Recognition with an Elastic Net-Regularized Hierarchical MAX Model of the Visual Cortex

Ali Alameer, *Student Member, IEEE,* Ghazal Ghazaei, *Student Member, IEEE,* Patrick Degenaar, *Member, IEEE,* Jonathon A. Chambers, *Fellow, IEEE,* and Kianoush Nazarpour, *Senior Member, IEEE*

***Abstract*—The human visual cortex has evolved to determine efficiently objects from within a scene. Hierarchical MAX (HMAX) is an object recognition model which has been inspired by the visual cortex, and sparse coding, which is a characteristic of neurons in the visual cortex, was previously integrated into the HMAX model for improved performance. In this work, in order to further enhance recognition accuracy, we have developed an elastic net-regularized dictionary learning approach for use in the HMAX model. We term this the En-HMAX model. With the En-HMAX model we can exploit the sparsity-grouping trade-off, such that correlated but informative features are preserved for object classification. Results show that the En-MAX model outperforms the original HMAX model in recognizing unseen objects by ~40% as well as the two special cases of the HMAX model, i.e., the least absolute shrinkage and selection operator (LASSO)-HMAX (~19%) and Ridge-HMAX (~9%) models.**

***Index Terms*—Elastic-net regularization, hierarchical MAX, dictionary learning, object recognition, sparsity**

I. Introduction

MACHINE vision has become an essential component of many human-computer interaction applications [1], [2]. By augmenting computers and robots with artificial vision, they have become capable of observing and (partially) understanding surrounding environments [3], [4]. Yet, reliably distinguishing objects and animals in arbitrary and cluttered backgrounds has remained challenging. This is because representations can differ considerably depending on position, orientation, and scale [5]. Therefore, a key challenge in machine vision is to represent the visual information, in such a way, that can allow recognition independent of physical conditions, such as size, occlusion, angle of view, and lighting. The recognition performance of many computer vision algorithms, however, declines when the object is rotated or shifted excessively [6].

In contrast, biological systems can recognize an object with different positions, orientations, and scales following a single observation [7]. In addition they can generalize to identify new objects, within the same category. Machine vision systems should therefore be able to similarly recognize and/or classify novel objects.

Manuscript received February 28, 2016.

The work of A. Alameer is supported by the HCED (Higher Committee for Education Development in Iraq). The work of K. Nazarpour is supported by the EPSRC, UK (grants: EP/M025977/1 and EP/M025594/1).

A. Alameer, G. Ghazaei and J.A. Chambers are with the School of Electrical and Electronic Engineering, Newcastle University, Newcastle NE1 7RU, UK.

P. Degenaar and K. Nazarpour are with the School of Electrical and Electronic Engineering and the Institute of Neuroscience, Newcastle University, Newcastle NE1 7RU, UK.

E-mail for correspondence: Kianoush.Nazarpour@newcastle.ac.uk.

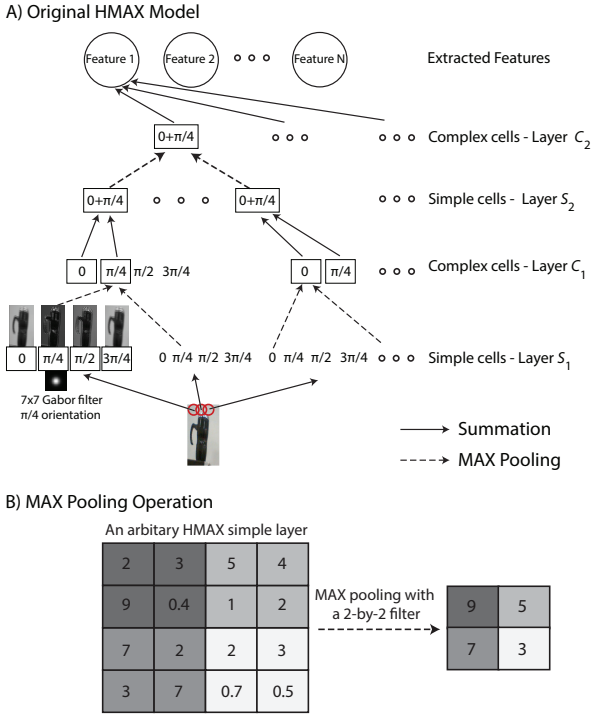

Fig. 1. A) Schematic of the HMAX model. The basic model consists of a hierarchy of two stages each having S and C layers with S_1 simple-cell like response properties to the C_2 layer with shape tuning and invariance properties [11]. B) MAX pooling operation over non-overlapping windows.

Histogram-based descriptors, such as the scale-invariant feature transform (SIFT) [7], the speeded up robust features (SURF) [6], and the Harris corner detector [8] have outperformed other approaches as they are robust to transformations of the previously seen objects. However, experimental results [9] have shown that such descriptors may not perform well on a generic object recognition task, due to the limited degree of invariance they provide. Many other histogram descriptors, such as the square patch of an image [10], are incapable of capturing discrepancies after object transformation.

Neuroscience experiments in rodents, e.g. [12], non-human primates, e.g. [11], and humans [13] support the hypothesis that the visual cortex can be approximated with a feedforward multi-layer structure. This architecture has inspired the multi-layer hierarchical MAX (HMAX) model [11] (Fig. 1A). The feedforward construct of the HMAX model can simulate the function of the early stages of the visual cortex in recognizing objects [14], [15]. In each stage of the HMAX model, two distinct groups of cortical cells are modeled [11]:

1) Simple cells S, to achieve selectivity;
2) Complex cells C, to offer invariance.

Recently, the HMAX model was implemented for use in real-time object classification applications [16], [17].

The primary visual cortex of the brain uses sparse coding to encode input data by strong activation of a relatively small set of neurons [5], [13], [18]. Sparse coding has been applied to the HMAX model previously to eliminate weak features in the higher layers and consequently to enhance classification performance [19]–[21]. For instance, in [21], sparse coding with a least absolute shrinkage and selection operator (LASSO) regularizer [22], was utilized in all S layers of the HMAX model. Whilst, this method improved classification performance, the LASSO regularizer discarded the grouping effect in the higher layers of the model [22].

Upon preliminary investigations [23], we observed that the higher layer features of the HMAX model may include highly correlated and grouped patches that may prove useful in classification. Our contribution is therefore to develop an elastic-net regularized [24] HMAX model, namely, En-HMAX, to balance efficiently the trade-off between the grouping effect of pixels and sparsity. We tested this new approach on a publicly available image database.

II. Method

A. Image Database

Seven image classes from the Caltech 101 dataset [25] were selected. These classes were: bass (54 images), binoculars (33 images), brontosaurus (50 images), camera (50 images), chair (62 images), gerenuk (34 images, also known as Waller's gazelle) and grand piano (99 images). Figure 2 shows two examples in four of these classes to reflect the richness of this dataset in terms of object size, orientation, position, and background. The rationale for choosing these classes was that an ample number of images per class was available, which allowed tuning the model parameters effectively; whilst keeping the computations to a reasonable level.

B. The original HMAX model

The original HMAX model (Fig. 1A) has two stages. A set of Gabor filters [26] forms the first stage and the second is a template matching mechanism. Each stage of the HMAX model has two sub-stages containing simple and complex cells, namely Simple 1 (S_1), Complex 1 (C_1), Simple 2 (S_2), and Complex 2 (C_2) [11]. The S_1 layer features are found by a bank of Gabor filters, resembling the cortical simple cell receptive fields. These filters can be represented with:

$$F(x,y) = \exp\left(-\frac{(x_0^2+\gamma^2 y_0^2)}{2\sigma^2}\right)\times\cos\left(\frac{2\pi}{\lambda}x_0\right) \qquad (1)$$

where

$$\begin{aligned} x_0 &= x\cos(\phi)+y\sin(\phi), \\ y_0 &= -x\sin(\phi)+y\cos(\phi). \end{aligned}$$

In (1), ϕ is the orientation of the stripes in a Gabor function, λ is the wavelength of the sinusoidal factor, γ is the spatial aspect ratio, and σ is the standard deviation of the Gaussian envelope.

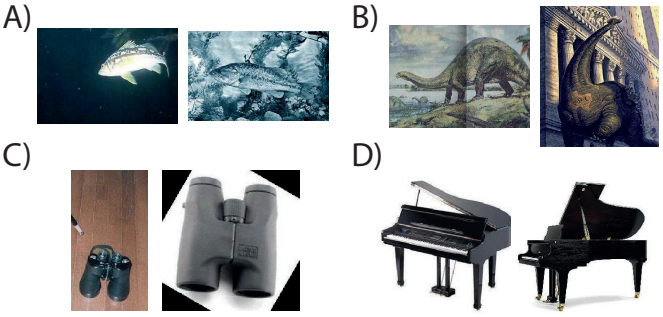

Fig. 2. Example of 4 (of 7) image classes, A) bass, B) brontosaurus, C) binoculars , and D) grand piano that were used in analysis. Samples illustrate the range of image sizes, orientations (portrait and landscape), and backgrounds.

In the HMAX model, an input image is first filtered with the above Gabor filters. This results in S_1 feature maps on which the MAX pooling operation is applied (Fig. 1B). MAX pooling is performed according to scale and orientation to achieve the sub-sampled layer C_1 feature maps. To build the S_2 layer, a set of prototype random patches is extracted from the C_1 layer. All patches from the C_1 layer are then compared with these prototypes using a radial basis or an Euclidean distance metric. The response of the comparison will be inversely proportional to the distance. Finally, the C_2 layer is generated by MAX pooling of S_2 to obtain position- and scale-invariant feature maps for classification. For more details, the reader is referred to [11], [27].

C. The proposed En-HMAX model

The proposed En-HMAX model differs from the original HMAX model in the following aspects:

1) Number of stages: The original HMAX model has only two stages (each comprising a simple and a complex layer) as shown in Fig. 1A. However, Serre et al. [14], among others, showed that an HMAX model with 3 stages is more appropriate to model rapid categorization. We therefore designed the En-HMAX model with three stages. Nevertheless for completeness, we compared both 2- and 3-stage En-HMAX models with the original 2-stage HMAX model.

2) Use of Elastic-Net Regularization: Hu et al. [21] proposed the use of sparse coding in the HMAX model to better represent the visual cortex. They adopted independent component analysis (ICA) [28] in the first simple layer of HMAX (S_1) followed by an ℓ_1-regularized dictionary learning structure in the following S layers. Here, we followed their approach and used ICA in S_1. We, inspired by [29], augmented the dictionary learning approach in S_2 and S_3 by using both ℓ_1 and ℓ_2 norms of the sparse coefficients matrix as penalizing terms.

Let $\mathbf{X} \in \mathbb{R}^{m\times n}$ contain m-dimensional image patches $\mathbf{x}$ in the S_2 or S_3 layers of the En-HMAX model, $\mathbf{D} \in \mathbb{R}^{m\times p}$ be a dictionary comprising p bases $\mathbf{d}$, and $\mathbf{S} \in \mathbb{R}^{p\times n}$ include n sparse vectors $\mathbf{s}$ in its columns. Then, in the matrix notation, sparse coding is formulated as $\mathbf{X} = \mathbf{DS}$. To learn the dictionary

TABLE I
PARAMETERS OF THE PROPOSED MODEL

Model parameters	Stage 1	Stage 2	Stage 3
Sparse coding	ICA	Elastic net	Elastic net
No. of bases	8	256	1024
Patch size	8×8	$4 \times 4 \times 8$	$2 \times 2 \times 256$
Sample size	25×10^4	25×10^4	25×10^4
Regularization	-	$\lambda_1 = 0.15$, $\lambda_2 = 0.15$	$\lambda_1 = 0.15$, $\lambda_2 = 0.15$
Pooling method	$\left(\sum_{r=1}^{n} \lvert q_r \rvert\right)^{\frac{1}{2}}$	$\left(\sum_{r=1}^{n} \lvert q_r \rvert\right)^{\frac{1}{2}}$	Max spatial pyramid
Pooling size	2×2	1×1	$\{1, 2, 4\}$

D and the sparse weighting matrix **S**, elastic-net regularization was used as the following

$$\begin{aligned} \underset{\mathbf{D},\mathbf{S}}{\text{minimize}} \quad & \|\mathbf{X} - \mathbf{DS}\|_F^2 + \lambda_1 \|\mathbf{S}\|_1 + \lambda_2 \|\mathbf{S}\|_F^2 \\ \text{subject to} \quad & \|\mathbf{d}_i\|_2 \leq 1, \quad i = 1, \cdots, p, \end{aligned} \tag{2}$$

where $\|.\|_F$ denotes the Frobenius norm and $\lambda_1, \lambda_2 \in \mathbb{R}_{\geq 0}$ are the regularization coefficients that regulate the trade-off between sparsity and the sensitivity of basis selection. When $\lambda_1 = 1$ and $\lambda_2 = 0$, (2) reduces to the ℓ_1 coding method described in [21], [22], hereafter called the LASSO-HMAX model and when $\lambda_1 = 0$ and $\lambda_2 = 1$, (2) reduces to another extreme case, which we call the Ridge-HMAX model. The notions of LASSO and Ridge regressions are borrowed from [22].

3) Pooling method: The C_1 and C_2 complex layers are partitioned into small non-overlapping square patches, termed **q** in a vector form. The ℓ_1 pooling is then applied such that from each patch the ℓ_1-norm, that is $\left(\sum_{r=1}^{n} \lvert q_r \rvert\right)^{1/2}$ is extracted. In addition, for C_3 we used the spatial pyramid [30] pooling method.

A full description of the parameters of the proposed En-HMAX model is presented in Table I. We used the same parameters and settings in both training and testing stages in all En-HMAX, Ridge-HMAX and LASSO-HMAX model setups.

D. Classification

Two classification scenarios were conducted: 15 or 30 images were selected randomly from each class to train the classifier. The remaining samples in each class were used for testing the classifier. The number of test images in each class was different, therefore to avoid bias, classification scores were adjusted according to the likelihoods. In addition, to ensure that the classification scores were not biased by the random choice of training samples, we repeated the classification for 20 independent runs in each condition (15 and 30 training samples). We report the average classification scores together with the standard deviations. A multi-class linear support vector machine (SVM) [31], [32] implemented within the LIBLINEAR library [32] was selected as the classifier due its computational simplicity.

E. Statistical Analysis

To test the statistical significance of using the En-HMAX model in improving the classification performance, we carried out a $3 \times 2 \times 2$ analysis of variance (ANOVA) with repeated measures. The main factors were the choice of model

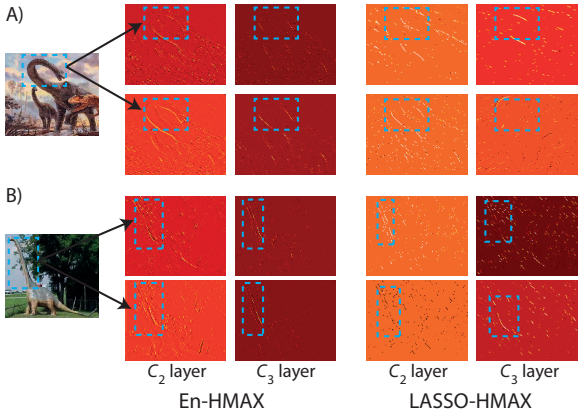

Fig. 3. Higher order correlation in representative feature maps extracted by using the En-HMAX model from the two example images A and B. Feature maps obtained by the En-HMAX model extract the neck of the brontosaurus very clearly. On the other hand, feature maps calculated with the LASSO-HMAX model are too sparse to reveal any determining feature of these image classes. Feature maps are gray scale. For visualization only, color scaling was used and feature maps were enlarged to counterbalance size shrinkage due to norm-pooling.

TABLE II
THE AVERAGE SPARSITY ACHIEVED WITH DIFFERENT MODELS

LASSO-HMAX		**En-HMAX**		**Ridge-HMAX**	
C_2	C_3	C_2	C_3	C_2	C_3
0.004	0.001	0.354	0.102	0.427	0.112

(LASSO-, En-, and Ridge-HMAX), whether classification was carried out at C_2 or C_3 layers, and finally the number of training samples, 15 versus 30. Following the main analysis, post-hoc comparisons were performed. Multiple comparisons were adjusted using Bonferroni correction.

III. RESULTS

A. Quantifying Sparsity

We hypothesized that using two penalty terms in (2), ℓ_1 and ℓ_2-norms of **S**, would lead to extraction of sparse C_2- and C_3-layer feature maps, which can retain second, and potentially higher, order correlation features. In support of this hypothesis, we provided representative examples of C_2- and C_3-layer feature maps, calculated with the En-HMAX and LASSO-HMAX ($\lambda_2 = 0$) model settings in Fig. 3. In this figure the responses of the C_2- and C_3-layers, calculated with the En-HMAX model, have several areas with class-specific strong activations that resemble the original image, e.g. the neck of the brontosaurus. The feature maps extracted by the LASSO-HMAX model are, however, too sparse and although they can correspond to some of the important features of the input images, many of the other important details are missed.

Table II reports the average sparsity achieved when all images of all classes were introduced to the En-, LASSO-, and Ridge-HMAX models. As predicted, using the En-HMAX model led to sparsity levels that fall between those achieved with the LASSO- and Ridge-HMAX models in both C_2 and C_3 layers.

TABLE III
AVERAGE CLASSIFICATION ACCURACY ± STANDARD DEVIATION (SD).

HMAX model configuration	2-layer Arrangement No. of training images 15	2-layer Arrangement No. of training images 30	3-layer Arrangement No. of training images 15	3-layer Arrangement No. of training images 30
HMAX [11]	35.014±0.09	40.587±0.08	-	-
LASSO-HMAX [21]	69.48±0.03	75.08±0.05	56.55 ±0.02	63.93±0.05
En-HMAX	75.14±0.02	80.37±0.04	78.71 ±0.01	82.72±0.04
Ridge-HMAX	66.14±0.02	71.45±0.05	67.27 ±0.02	73.30±0.06

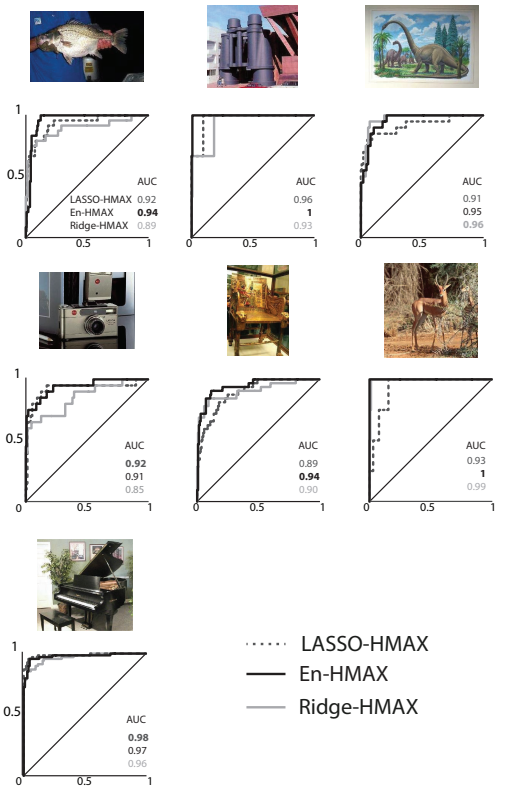

Fig. 4. Performance comparison of the En-, LASSO-, and Ridge-HMAX models with respect to the ROC and AUC measures; Top: Samples from images classes with different sizes and orientations; Bottom: The corresponding ROCs curves and the calculated AUC values for each image class. The highest AUC value is in a bold font. The vertical and horizontal axes denote the true positive and false positive rates, respectively.

B. Classification Scores

We compared the En-, LASSO-, and Ridge-HMAX models in terms of classification accuracy. For completeness, we included the classification scores achieved by the original 2-layer HMAX model [11]. Table III reports the classification results. Statistical analysis revealed the main effect of the model ($F_{2,18} = 266.59, p < 10^{-5}$), feature map selection ($F_{1,19} = 24.37, p < 10^{-5}$), and number of training data ($F_{1,19} = 115.83, p < 10^{-5}$). In both 2- and 3-layer structures and in both 15 and 30 training sample conditions, the En-HMAX model outperformed all other algorithms ($p < 10^{-5}$). The performance improvement in the 3-layer arrangement was considerably larger than that in the 2-layer setup ($p < 10^{-5}$). This is particularly interesting because in the experimental neuroscience literature, a 3-layer HMAX model setup is deemed more appropriate for modelling visual processing [14]. Finally, using 30 training images, instead of 15, improved classification scores significantly ($p < 10^{-5}$).

TABLE IV
F1-SCORES FOR 3-LAYER ARRANGEMENT WITH 30 TRAINING IMAGES

HMAX model configuration	F1-Score	Precision	Recall
LASSO-HMAX [21]	0.37	0.25	0.66
En-HMAX	0.63	0.51	0.83
Ridge-HMAX	0.49	0.39	0.66

Theoretical analysis indicated that all forms of ℓ_p norm pooling can offer invariance [33]. However, in practice, different pooling mechanisms could lead to stark differences in recognition performance. We found that the use of ℓ_1-norm pooling in the C_1 and C_2 layers offers much better performance than MAX (ℓ_∞-norm) pooling. The overall performance achieved by the use of ℓ_1- and ℓ_2-norm pooling in C_1 and C_2 were comparable.

Figure 4 shows the receiver operating characteristic curve [34] for all of the classes used in this experiment using a 3-layer En-HMAX model (30 training images). The area under the curve (AUC) was used to characterize the classification confidence in a specific binary classification task (e.g., camera versus not-camera) with a unity value reflecting a 100% accuracy. In 4 out of 7 classes, using the En-HMAX model led to the highest AUC. The performance of the En-HMAX model was only marginally lower than the LASSO-HMAX model in 2 classes and the Ridge-HMAX model in 1 class. Table IV reports the F1-scores [35], and the corresponding precisions and recalls, achieved with different models for a 3-layer En-HMAX model (30 training images). Results reflect the higher performance of the En-HMAX model when compared to the LASSO-HMAX and Ridge-HMAX models.

C. On real-time implementation

All models were implemented in Matlab on a dual-core i5 processor (3.4 GHz) PC with 32G RAM without GPU acceleration. Recently, the basic HMAX model was implemented in hardware to realise a biomimetic object recognition system [17]. It was shown that accurate performances in near real-time may be possible. We intend to implement the proposed the En-HMAX model in hardware. In a real-time setting, the dictionary learning stage of the proposed En-HMAX model remains a challenge. Future image classification systems may benefit significantly from bio-inspired vision constructs, such as En-HMAX model.

IV. CONCLUSIONS

The new En-HMAX model presented in this work provides two essential elements for image classification: selectivity and invariance. The main reason of using an elastic-net regularizer for the HMAX model was to encourage the grouping effect when the atoms in the dictionary are highly correlated. The key to the recognition performance of the En-HMAX model is the large number of automatically tunable units across its hierarchical architecture. Results show that our model outperforms the original HMAX model (by ~40%) as well as the two special cases of the En-HMAX model, i.e., the LASSO- and Ridge-HMAX models, by ~19% and ~9%, respectively.

REFERENCES

[1] J. R. Parker, *Algorithms for Image Processing and Computer Vision*. John Wiley & Sons, 2010.

[2] M. Hu, Z. Wei, M. Shao, and G. Zhang, "3-D object recognition via aspect graph aware 3-D object representation," *IEEE Sig. Process. Lett.*, vol. 22, no. 12, pp. 2359–2363, 2015.

[3] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *Int. J. Comp. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.

[4] L. Li, S. Jiang, and Q. Huang, "Learning hierarchical semantic description via mixed-norm regularization for image understanding," *IEEE Trans. Multimedia*, vol. 14, no. 5, pp. 1401–1413, 2012.

[5] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[6] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comp. Vis. Imag. Underst.*, vol. 110, no. 3, pp. 346–359, 2008.

[7] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Computer Vision*, vol. 2, 1999, pp. 1150–1157.

[8] C. Harris and M. Stephens, "A combined corner and edge detector," in *in Proc. 4th Alvey Vision Conference*, vol. 15, 1988, pp. 147–152.

[9] T. Serre, L. Wolf, and T. Poggio, "Object recognition with features inspired by visual cortex," in *Proc. CVPR*, 2005, pp. 994–1000.

[10] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-based object detection in images by components," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 4, pp. 349–361, 2001.

[11] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nat. Neurosci.*, vol. 2, no. 11, pp. 1019–1025, 1999.

[12] D. Zoccolan, N. Oertelt, J. J. DiCarlo, and D. D. Cox, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Proc. Nat. Aca. Sci.*, vol. 108, no. 21, pp. 8748–8753, 2009.

[13] R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, and I. Fried, "Invariant visual representation by single neurons in the human brain," *Nature*, vol. 435, no. 7045, pp. 1102–1107, 2005.

[14] T. Serre, A. Oliva, and T. Poggio, "A feedforward architecture accounts for rapid categorization," *PNAS*, vol. 104, no. 15, pp. 6424–6429, 2007.

[15] S. Hochstein and M. Ahissar, "View from the top: Hierarchies and reverse hierarchies in the visual system," *Neuron*, vol. 36, no. 5, pp. 791–804, 2002.

[16] A. Maashri, M. DeBole, M. Cotter, N. Chandramoorthy, Y. Xiao, V. Narayanan, and C. Chakrabarti, "Accelerating neuromorphic vision algorithms for recognition," in *Proc. 49th IEEE Design Autom. Conf.*, 2012, pp. 579–584.

[17] G. Orchard, J. G. Martin, R. J. Vogelstein, and R. Etienne-Cummings, "Fast neuromimetic object recognition using FPGA outperforms GPU implementations," *IEEE Trans. Neural Net. Learn. Sys.*, vol. 24, no. 8, pp. 1239–1252, 2013.

[18] E. T. Carlson, R. J. Rasquinha, K. Zhang, and C. E. Connor, "A sparse object coding scheme in area v4," *Current Biol.*, vol. 21, no. 4, pp. 288–293, 2011.

[19] J. Mutch and D. G. Lowe, "Object class recognition and localization using sparse features with limited receptive fields," *Int. J. Comp. Vis.*, vol. 80, no. 1, pp. 45–57, October 2008.

[20] Y. Huang, K. Huang, D. Tao, T. Tan, and X. Li, "Enhanced biologically inspired model for object recognition," *IEEE. Trans. Sys. Man Cyber. Part B*, vol. 41, no. 6, pp. 1668–1680, 2011.

[21] X. Hu, J. Zhang, J. Li, and B. Zhang, "Sparsity-regularized HMAX for visual recognition," *PLoS One*, vol. 9, no. 1, p. e81813, 2014.

[22] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Royal Stat. Soc. Series B*, vol. 58, pp. 267–288, 1994.

[23] A. Alameer, G. Ghazaei, P. Degenaar, and K. Nazarpour, "An elastic net-regularized HMAX model of visual processing," in *Proc. Intelligent Signal Processing (ISP), The 2^nd IET International Conference on*. IET, 2015.

[24] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Royal Stat. Soc. Series B*, vol. 67, pp. 301–320, 2005.

[25] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," *Comp. Vis. Imag. Underst.*, vol. 106, no. 1, pp. 59–70, 2007.

[26] S. Marcelja, "Mathematical description of the responses of simple cortical cells," *J. Opt. Soc. Am.*, vol. 70, no. 11, pp. 1297–1300, 1980.

[27] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 411–426, 2007.

[28] A. Hyvärinen, M. Gutmann, and P. O. Hoyer, "Statistical model of natural stimuli predicts edge-like pooling of spatial frequency channels in V2," *BMC Neurosci.*, vol. 6, no. 1, p. 12, 2005.

[29] B. Shen, B.-D. Liu, and Q. Wang, "Elastic net regularized dictionary learning for image classification," *Multimedia Tools App.*, pp. 1–14, 2014.

[30] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. IEEE Conf. Comp. Vis. Pattern Recog.*, 2009.

[31] V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[32] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, 2008.

[33] C. Gulcehre, K. Cho, R. Pascanu, and Y. Bengio, "Learned-norm pooling for deep feedforward and recurrent neural networks," in *Machine Learning and Knowledge Discovery in Databases*. Springer, 2014, pp. 530–546.

[34] X. Sun and W. Xu, "Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves," *IEEE Sig. Process. Lett.*, vol. 21, no. 11, pp. 1389–1393, 2014.

[35] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *J Machine Learning Tech.*, vol. 2, no. 1, pp. 37–63, 2011.