Processing Occlusions Using Elastic-net Hierarchical MAX Model of The Visual Cortex

Ali Alameer
Newcastle University, UK
School of Electrical and Electronic Engineering
A.m.a.Alameer@newcastle.ac.uk

Patrick Degenaar
Newcastle University, UK
School of Electrical and Electronic Engineering and the Institute of Neuroscience
patrick.degenaar@newcastle.ac.uk

Kianoush Nazarpour
Newcastle University, UK
School of Electrical and Electronic Engineering and the Institute of Neuroscience
kianoush.nazarpour@newcastle.ac.uk

***Abstract*—Humans can recognise objects under partial occlusion. Machine-based approaches cannot reliably recognise objects and scenes in the presence of occlusion. This paper investigates the use of the elastic net hierarchical MAX (En-HMAX) model to handle occlusions. Our experiments show that the En-HMAX model achieves an accuracy of $\sim$ 70%, when $\sim$ 50% artificial occlusions are applied to the centre of the visual object-field. Furthermore, when the same percentage of occlusion is applied to the peripheral, the model reports higher accuracies. A similar degree of robustness has been observed when recognising scenes. The results suggest that cortex-like models, such as the En-HMAX are reliable for solving the occlusion challenge.**

***Keywords*—Elastic-net regularization, hierarchical MAX, dictionary learning, object recognition, sparsity, occlusion, regions of vision.**

I. Introduction

Recent years have witnessed huge progress in machine vision, making it a cornerstone of pattern recognition and visual processing. Within machine vision, one of the most challenging problems is recognising scenes and objects under partial occlusion. Artificial vision has developed significantly during the last decades [1]. It includes interpreting high-dimensional data from the real world into a numerical representation to form a particular decision [2], [3].

Despite progress towards more accurate object recognition, partial occlusion remains the main challenge to state-of-the-art object recognition models. Recognising partially occluded objects is an essential problem in visual processing. Researchers have introduced different types of solutions to address this problem [4], [5]. However, the common factor in their methods was to model the presence of occlusion. Typically, these methods depend on the occluder, where the training data comprises both the occluder and the occludee. The key limitation of these methods is that it can not be generalised as they are application specific. For example, recognising a car parked in a road, where the visible part of the car together with the occluder form a pattern that continuously repeat itself. These patterns were considered enough to perform correct classification.

One other approach to handle occlusion is to utilise the statistical inconsistency that the occluder generates when applied to a scene. Girshick et al. [6] developed a grammar model to detect occlusions. It is based on representing objects using deformable parts within a variable structure, to perform recognition. In [7], 3D sensors have been utilised to determine the depth inconsistency to locate and isolate occlusions. Similarly, In [8], [9], local similarity has been used to decorrelate partial occlusions. However, these approaches may not be practical in all situations as the performance decreases drastically whenever an image feature leaps out from the corresponding models' features, even if the rest of features introduce an accurate match with its corresponding template. A growing body of evidence support the proposition that biological systems are able to recognise an object under partial occlusion [10]. Based on the above observations, and because the occluder does not have a geometrically fixed pattern as shown in Fig. 1, we decided to explicitly model occluder patterns using the 2-D Gaussian formula. Specifically, we applied the Gaussian function to the interior of the image at different scales (from occlusion-free to full image occlusion).

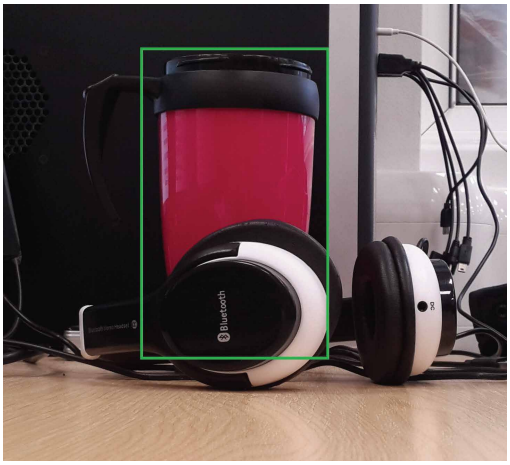

Fig. 1. Example detection and recognition of a cup under partial occlusion.

To that end, we evaluate the En-HMAX [11] model on the following proposed scenarios:

1) Central Gaussian-shaped occlusions as shown in Fig. 3.
2) Peripheral Gaussian-shaped occlusions.

The remainder of the paper is organised as follows. Section II presents the methods of the classical HMAX model, elastic net regularizer, the En-HMAX model, dataset, and a description of the experiments. Section III gives supporting experimental results to the datasets. Section IV present a discussion of the paper. Finally, section V concludes the paper.

978-1-5090-5795-5/17/$31.00 © 2017 IEEE

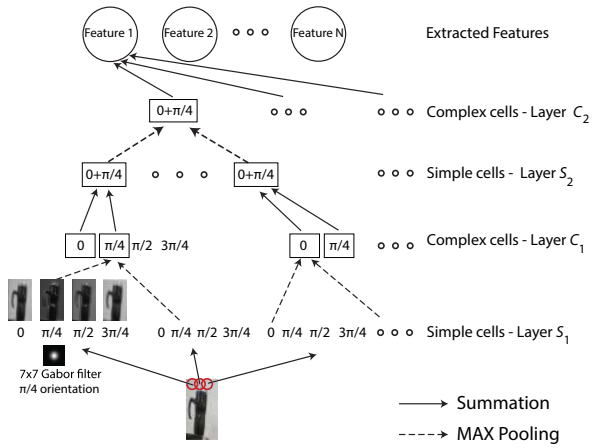

Fig. 2. A schematic of the standard HMAX model configuration.

II. Method

A. The Classic HMAX Model

The HMAX model uses the classic scheme of convolution/pooling as reported in [12]. The convolutional layers generate selective feature maps, and the pooling layers provide invariance.

The HMAX model consists of four layers which contains a combination of simple and complex cells, namely S_1, C_1, S_2, and C_2 as show in Fig.2. S_1 layer is a set of Gabor filters, resembling the cortical simple cell receptive fields. The Gabor filters are defined as:

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

where

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

where λ is the wavelength of the sinusoidal factor and γ is the spatial aspect ratio.

S_1 layer feature maps are calculated by convolving a set of Gabor filters $F(x,y)$ (scales σ and orientations ϕ) with the input images. The Gabor filters mimic the simple cell activation in the primary visual cortex (V1). The C_1 feature map is calculated by obtaining only the maxima of a neighbouring square patch. This efficiently increases the invariance to transformations and specifically to translation [13].

Patches from the C_1 layer are then compared with prototypes using a radial basis function or Euclidean distance metric; the smaller the distance, the higher the response. Finally, the C_2 layer response is generated by max-pooling of S_2 to obtain position- and scale-invariant feature maps for classification [14].

B. Elastic-Net Regularizer

Let $\mathbf{x_i} \in R^m$ be an image patch. Given a set of bases $\mathbf{d_i} \in R^m$, sparse coding searches for the sparse coefficients s_j such that $\mathbf{x_i} = \sum_{j=1}^{p} \mathbf{d_i}\ s_j$, where p denotes the size of dictionary. In matrix notation, the equation becomes:

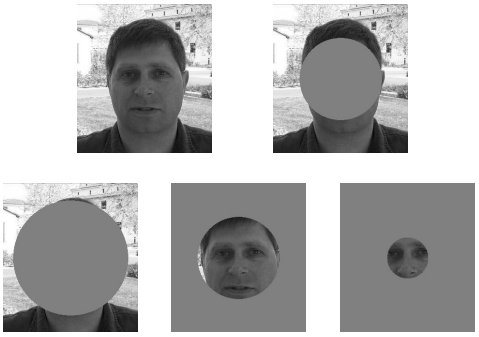

Fig. 3. A sample of class-A occlusions

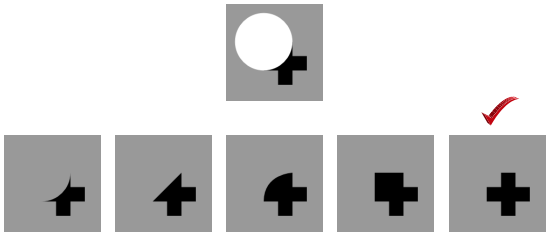

Fig. 4. A smaple of Class-B occlusions. Image above is occluded by a circle shape occluder. Images below are the potentials of original image.

$$\mathbf{X} = \mathbf{DS}, \tag{2}$$

where $\mathbf{X}$ is an n-dimensional local descriptor extracted from the input images and $\mathbf{D}$ is a p-dimensional dictionary matrix with each column defining a basis $\mathbf{d_i}$. Each column of $\mathbf{S}$ is a vector $\mathbf{s_i} \in R^p$ holding the sparse coefficients of the p bases for reconstructing $\mathbf{x_i}$. An elastic net formulation is:

$$\begin{aligned} \text{minimize} \quad & \|\mathbf{X} - \mathbf{DS}\|_F^2 + \lambda_1 \|\mathbf{S}\|_1 + \lambda_2 \|\mathbf{S}\|_F^2 \\ \text{subjected to} \quad & \|\mathbf{d}_i\|_2 \leq 1, \forall i = 1, ..., p. \end{aligned} \tag{3}$$

where $\|.\|_F$ denotes the Frobenius norm. For every input image patch $\mathbf{x_i}$ in $\mathbf{X} \in R^{m\times n}$, a vector $\mathbf{s_i}$ in $\mathbf{S} \in R^{p\times n}$ is reproduced, corresponding to a basis $\mathbf{d_i}$ in the dictionary $\mathbf{D} \in R^{m\times p}$. λ_1 and λ_2 are the regularization parameters which control the trade off between sparsity and goodness of fit. The sparsity of the coefficients is controlled by λ_1, while λ_2 controls the sensitivity of basis selection from the dictionary.

C. The En-HMAX Model

In-line with [12], [14], [15], the En-HMAX is a hierarchical feed-forward model. It attempts to mimic the first 100 milliseconds of the ventral stream processing of the primate visual cortex. Generally, this model contains four or six layers, divided into simple S and complex C units. The S layers combine independent component analysis (ICA) and Elastic net to increase specificity. In the C layers, the L_P norm pooling is used to provide invariance. Using the elastic-net-regularizer in S_2 and S_3 layers reinforces the model sparsity and grouping effect simultaneously. In particular, in the higher layers of the hierarchy [16], [17].

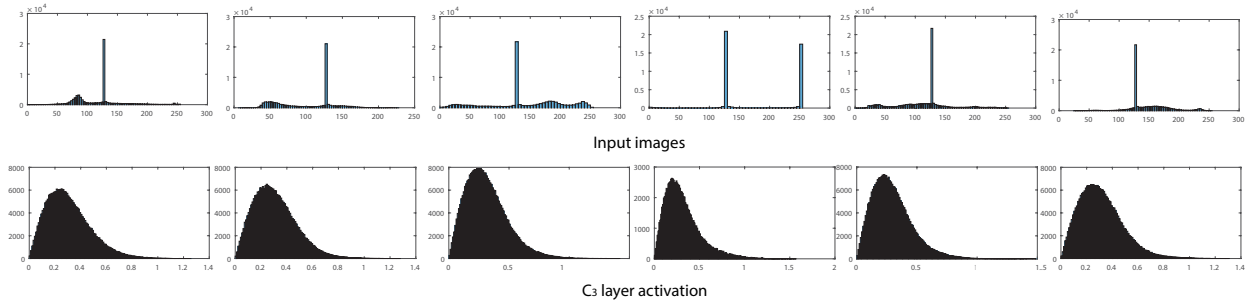

Fig. 5. Histogram representing the overlap between input data. First row: histogram of some selected examples of the object images with 50% central occlusion. It is clear that overlap exists in the input data, located in the centre of the images. Second row: histogram of the non-zero coefficients of En-HMAX activation.

D. Dataset

1) Objects dataset: To investigate the performance of the En-HMAX model under partial occlusion, eleven categories from Caltech 101 and Caltech 256 [18], [19] databases have been selected. The categories were: car sides (123 images), dollar bills (52 images), faces easy (435 images), Garfield (34 images), inline skates (31 images), motorbikes (798 images), pagodas (47 images), pandas (38 images), scissors (39 images), trilobites (86 images) and Windsor chairs (56 images). Class-A occlusion type was applied on the dataset, as described below. The occlusions have been applied to the centre and the peripheral of the images, creating two types of occlusion as shown in Fig. 3.

2) Scene dataset: Fifteen scene categories [20] dataset has been selected to investigate the model robustness against occlusions on scenes. The dataset contains a plethora of scene images that belongs to 15 categories. Each category consists of 200 to 400 images, with an average image size of 300×250 pixels. The classes were: bedroom, CAL suburb, industrial, kitchen, living room, MIT coast, MIT forest, MIT highway, MIT inside city, MIT mountain, MIT open country, MIT street, MIT tall building, PAR office, and store. This dataset is considered as one of the most complete scene category datasets in the literature thus far. Class-A occlusion type has also been applied, to evaluate the model robustness.

Class-A Occlusion: To derive the importance of the diagnostic regions, artificial occlusions have been applied on both datasets. Peripheral and central regions of the images were occluded. The generated datasets have different percentages of occlusion. The elliptical Gaussian function corresponding to the occlusion is given by:

$$f(x,y) = \frac{1}{2\pi\sigma^2} e^{-\left[(x-\mu_x)^2+(y-\mu_y)^2\right]/2\sigma^2}, \tag{4}$$

where the variance σ^2 determines the size of the occlusion. We have used the same σ values for horizontal and vertical axes of the images, therefore the occlusion was reduced to a circle.

3) klab Dataset: A set of various partially occluded stimuli was created by klab [21]. This dataset is considered challenging. Sensitive parts of the stimulus were occluded, making the recognition process relatively difficult. Therefore, in order to solve this type of occlusions, the recognition algorithm has to intelligently predict the original shape, as shown in Fig. 4.

Class-B Occlusion: Models of rapid categorisation of the visual cortex do not involve mechanisms for perceptual grouping and top-down processing. Such mechanisms are responsible of driving the completion of the occlusion in the human brain. Class-B occlusion requires the mentioned mechanisms to solve the task. The ground truth of this dataset is based on human decision. An example of Class-B occlusions is shown in Fig 4. Human subjects have chosen the shape on the far right corresponding to the original shape of the above image. However, most computer-based models select the shape on the far left corresponding the (+ shaped) image.

E. Statistical Regularities

In experiments 1, the input data is extremely overlapped as shown in Fig. 5. Clearly, overlapping (i.e., observations that are spatially concentrated) is not visible at the level of C_3 layer. Rather, we see that the observations are uniformly distributed among the different values in the spectrum.

The combination of norm pooling and dictionary learning using elastic net has enabled the En-HMAX to introduce a prominent performance against the highly overlapped input data. Nevertheless, the filters in the S_1 layer are learned from natural images, enabling the model to be adapted to the natural environment statistics, creating filters that are tuned to new image features. In particular, the learning in the S layers of the model is unsupervised and sequential, and the learning process develops through the hierarchy of the model.

F. Experiments

1) Experiment 1 - Robustness against occlusion: We tested our model on different sizes and types of class-A occlusions. The En-HMAX model was used to process the data and extract features. We use the classical HMAX model to compare performance. The models were trained using only 15 samples per category, where the training samples were occlusion-free images. The testing samples were class-A occlusion with different scales of occlusion, as shown in Fig. 3. A fixed number of images per category (15 training / 15 testing) was used in this experiment to report the overall error rates, as recommended in [18]. This experiment has also been used to study the field of attention of a limited area of the visual field, for both objects and scenes.

TABLE I

CLASSIFICATION ACCURACY IN A PERCENTAGE OF DIFFERENT TYPES AND SIZES OF OCCLUSIONS ON THE OBJECT DATASET.

Objects			
Occlusion type	Occlusion size	HMAX [12]	En-HMAX [16]
Central	~25%	54.090 ± 0.17	99.818 ± 0.003
Central	~50%	43.272 ± 0.07	70.636 ± 0.05
Central	~75%	28.500 ± 0.04	29.000 ± 0.03
Peripheral	~25%	73.454 ± 0.11	99.454 ± 0.006
Peripheral	~50%	65.681 ± 0.09	74.545 ± 0.06
Peripheral	~75%	54.181 ± 0.13	24.772 ± 0.03

TABLE II

CLASSIFICATION ACCURACY IN A PERCENTAGE OF DIFFERENT TYPES AND SIZES OF OCCLUSIONS ON THE SCENE DATASET.

Scenes			
Occlusion type	Occlusion size	HMAX [12]	En-HMAX [16]
Central	~25%	25.100 ± 0.13	99.166 ± 0.005
Central	~50%	17.466 ± 0.07	69.766 ± 0.13
Central	~75%	14.666 ± 0.03	20.833 ± 0.06
Peripheral	~25%	31.400 ± 0.17	100.000 ± 0
Peripheral	~50%	19.700 ± 0.08	53.133 ± 0.15
Peripheral	~75%	18.833 ± 0.06	17.400 ± 0.08

2) Experiment 2 - Processing class-B occlusions: In this experiment, we tested the model using class-B occlusion. The euclidean distant is used to measure the similarity of the different feature maps. The distant is calculated between each of the spatially pooled feature maps (SPFMs). That is, for a target SPFM $\mathbf{z}$, the distant r of the corresponding SPFM $\mathbf{p}$ is given by $r = \|\mathbf{z} - \mathbf{p}\|$.

III. RESULTS

A. Experiments 1

The results of classification are shown in Table (I, II). Various percentages of occlusion are used to test both versions of the HMAX model. class-A occlusions block the models from encountering implicit features in different image regions, making the task more difficult. It can be noticed that when an equal viewable area is presented, central vision outperforms peripheral vision for recognising objects. This is by producing slightly higher accuracies, and vice versa for scene images.

Table III shows the confusion matrix of both paradigms of occlusion. It can be noticed that our model shows more robustness towards central occlusions for the scene recognition. However, the model is more sensitive towards the peripheral details to recognise scenes.

B. Experiments 2

State of the art biological models have been used to detect class-B occlusions. The highest accuracy achieved was 30% as shown in Table IV. En-HMAX outperforms other models of object recognition with a performance of 33.333% to understand class-B occlusion. En-HMAX model achieved acceptable scores comparing to human scores in the low-variation stimuli. Conversely, in the high-variation stimuli, all of the algorithms used fail to match the human performance by a large margin. It is not surprising that "ventral stream pathway" inspired models are not nearly as effective as human performance, but it is instructive to note that the task is sufficiently difficult that the models perform in such way.

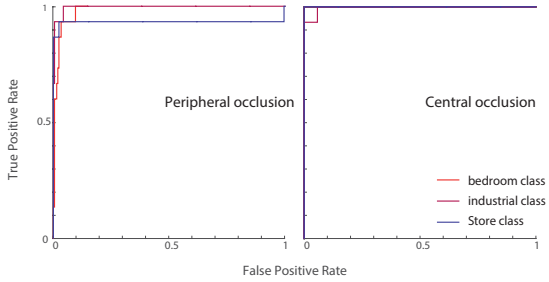

Fig. 6. ROC curve of the En-HMAX model with respect to the scene dataset (central and peripheral occlusion of 50% size). All fifteen classes are included in this analysis and only classes with the lowest AUCs are denoted above (red, purple, and blue). The vertical and horizontal axes denote the true positive and false positive rates, respectively.

IV. DISCUSSION

The model was presented with images that contain information from everything but foveal, parafoveal vision for both objects and scenes images. The results show robustness against occlusions, where an accuracy of 99% has been scored when 25% of the visual scene is occluded. Furthermore, from Table (I, II), we conclude that peripheral information is more useful for recognizing scenes and the central information is effiecient for object recognition.

Cortex models have shown substantial robustness against occlusion. The En-HMAX has robustly separated the occluded dataset classes, as shown in Fig.6; where most of the classes have scored a maximum area under the curve (AUC). Although the En-HMAX model has outperformed (in the small margin) other models in the literature on class-B occlusion, feed-forward models inspired by the ventral stream of the visual cortex were insufficient (as expected) to solve complex occlusions. Class-B occlusion requires attention and top-down processing to solve. In the future, it is constructive to add layers to the En-HMAX model that resemble top-down connections and memory association to solve such type of occlusions.

V. CONCLUSION

Our experimental results show that hierarchical structures, such as the En-HMAX model, offer substantial robustness in recognizing objects under partial occlusion. The model provides two elements essential for image classification: selectivity and invariance. Additionally, our results show that the En-HMAX model gives more weight to objects centred in the image rather than their surroundings, in-line with the focal attributes of biological vision.

In this study, all occlusions had the same shape, location, and pixel value; producing consistent areas of statistical regularity. Using an elastic-net dictionary learning in the HMAX model scheme encouraged the grouping effect when atoms in the dictionary were highly correlated. As a result, En-HMAX showed higher performance when encountering highly correlated data, for example in the class-A occlusion condition.

TABLE III

CONFUSION MATRIX FOR SCENE DATASET WITHIN AN OCCLUSION SIZE OF 50%; A) CENTRAL OCCLUSION, THE MODEL SHOWS MORE ROBUSTNESS TO CENTRAL OCCLUSION FOR THE SCENE RECOGNITION; B PERIPHERAL OCCLUSION, THE MODEL IS MORE SENSITIVE TO PERIPHERAL DATA. BLANK SPOTS REPRESENT 0.

A 50% Central Occlusion

	suburb	coast	forest	highway	insidecity	mountain	country	street	building	office	bedroom	industrial	kitchen	livingroom	store
suburb	**100.00**														
coast		**100.00**													
forest		33.33	**66.67**												
highway		100.00		**0.00**											
insidecity		40.00			**60.00**										
mountain		13.33				**86.67**									
country		40.00				6.67	**53.33**								
street		66.67						**20.00**			13.33				
building		6.67							**93.33**						
office										**100.00**					
bedroom		6.67									**93.33**				
industrial		6.67								6.67		**86.67**			
kitchen		13.33								26.67	40.00		**20.00**		
livingroom		13.33												**86.67**	
store		60.00													**40.00**

B 50% Peripheral Occlusion

	suburb	coast	forest	highway	insidecity	mountain	country	street	building	office	bedroom	industrial	kitchen	livingroom	store
suburb	**0.00**					**100.00**									
coast		**6.67**				**93.33**									
forest			**0.00**	**6.67**		**93.33**									
highway				**93.33**		**6.67**									
insidecity					**93.33**	**6.67**									
mountain						**100.00**									
country						**93.33**	**0.00**								
street								**100.00**							
building						**33.33**			**66.67**						
office						**93.33**				**6.67**					
bedroom						**33.33**					**66.67**				
industrial				**13.33**		**53.33**						**33.33**			
kitchen						**33.33**							**66.67**		
livingroom						**93.33**								**6.67**	
store				**6.67**		**73.33**			**6.67**						**13.33**

TABLE IV

CLASSIFICATION ACCURACY IN PERCENTAGE ON THE PARTIALLY OCCLUDED KLAB DATASET [21].

Model Architecture	Total performance in percentage
Our model	33.333
HMO [22]	30
GaborJet [23]	30
HMAX [12]	30

Biologically-inspired vision systems that can handle occlusion may enhance the reliability and accuracy of robotic grasp systems, such as those developed for prosthetic applications [24]

VI. ACKNOWLEDGEMENT

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).

REFERENCES

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

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

[3] G. Ghazaei, A. Alameer, P. Degenaar, G. Morgan, and K. Nazarpour, "An exploratory study on the use of convolutional neural networks for object grasp classification," in *IET Intelligent Signal Processing Conference*, 2015, p. 5.

[4] Z. Ying and D. Castanon, "Statistical model for occluded object recognition," in *Information Intelligence and Systems, 1999. Proceedings. 1999 International Conference on*. IEEE, 1999, pp. 324–327.

[5] B. Pepikj, M. Stark, P. Gehler, and B. Schiele, "Occlusion patterns for object class detection," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 3286–3293.

[6] R. B. Girshick, P. F. Felzenszwalb, and D. A. McAllester, "Object detection with grammar models," in *Advances in Neural Information Processing Systems 24*, J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, Eds., 2011, pp. 442–450.

[7] D. Meger, C. Wojek, J. Little, and B. Schiele, "Explicit occlusion reasoning for 3D object detection," in *Proceedings of the British Machine Vision Conference*. BMVA Press, 2011, pp. 113.1–113.11.

[8] C. F. Olson and D. P. Huttenlocher, "Automatic target recognition by matching oriented edge pixels," *Image Processing, IEEE Transactions on*, vol. 6, no. 1, pp. 103–113, 1997.

[9] S. Z. Der and R. Chellappa, "Probe-based automatic target recognition in infrared imagery," *Image Processing, IEEE Transactions on*, vol. 6, no. 1, pp. 92–102, 1997.

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

[11] A. Alameer, G. Ghazaei, P. Degenaar, and K. Nazarpour, "An elastic net-regularized H-MAX model of visual processing," in *IET Intelligent Signal Processing Conference*, 2015, p. 5.

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

[13] K. Fukushima and S. Miyake, "Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position," *Pattern Recognition*, vol. 15, no. 6, pp. 455–469, 1982.

[14] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 3, pp. 411–426, 2007.

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

[16] A. Alameer, G. Ghazaei, P. Degenaar, J. A. Chambers, and K. Nazarpour, "Object recognition with an elastic net-regularized hierarchical MAX model of the visual cortex," *IEEE Signal Processing Letters*, vol. 23, no. 8, pp. 1062–1066, 2016.

[17] A. Alameer, P. Degenaar, and K. Nazarpour, "Biologically-inspired object recognition system for recognizing natural scene categories," in *Students on Applied Engineering (ISCAE), International Conference for*. IEEE, 2016, pp. 129–132.

[18] 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.

[19] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," 2007.

[20] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2. IEEE, 2006, pp. 2169–2178.

[21] J. Kubilius, "Partially occluded figures," 2016. [Online]. Available: 10.6084/m9.figshare.2114191.v1

[22] D. L. K. Yamins, H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert, and J. J. DiCarlo, "Performance-optimized hierarchical models predict neural responses in higher visual cortex." *Proceedings of the National Academy of Sciences of the United States of America*, 2014.

[23] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *Computers, IEEE Transactions on*, vol. 42, no. 3, pp. 300–311, 1993.

[24] G. Ghazaei, A. Alameer, P. Degenaar, G. Morgan, and K. Nazarpour, "Deep learning-based artificial vision for grasp classification in myoelectric hands," *Journal of Neural Engineering*, vol. 17, no. 3, p. 036025, 2017.