



Novel patch selection based on object detection in HMAX for natural image classification

Mohammadesmaeil Akbarpour¹ · Mrinal Mandal² · M. Hashemi Kamangar¹

Received: 18 April 2021 / Revised: 23 July 2021 / Accepted: 18 October 2021 / Published online: 15 November 2021
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

Abstract

The human visual system (HVS) can effectively recognize objects in complex natural scenes with high speed and accuracy. Many models have been proposed based on HVS among which HMAX is one of the superior models. In HMAX, the random extraction of a large volume of training samples, called patches, has two drawbacks. First, patches from background, in addition to high computational cost, can produce wrong output. Second, patches with low information from objects may provide poor performance. In this paper, an optimum method, with two steps, is proposed to select patches with high discriminative information. First, a pool of patches is extracted from objects based on an unsupervised object detection method. Second, patches with high discriminative information were selected from the pool based on patch ranking. Further, complement of optimum patch for each class is considered as a new patch for other classes to increase the recognition rate. Experimental results with Caltech5, Caltech101 and Graz-01 databases show that the proposed model provides a significant performance improvement over the HMAX and other state-of-the-art models, in terms of speed, sensitivity, specificity and classification accuracy.

Keywords Classification · Unsupervised object detection · Human visual system · HMAX · Optimum patch

1 Introduction

Object recognition has proven to be a challenge for computer vision which has numerous range of applications, e.g., medical image, data Hiding, surveillance, agriculture and vehicle recognition [1–5]. It should respond invariably to objects from within-class and differently to objects from between-class [6]. The human visual system (HVS) is able to recognize objects easily in a cluttered scene in a fraction of second, while the most powerful computer systems are generally not capable of doing so. The HVS is very efficient in the object recognition and is the ultimate evaluator of any recognition performance. Because humans outperform the best machine

vision systems, building a system that emulates object recognition by the HVS has always been a desirable but elusive goal [7].

Due to the enormous complexity of the HVS and intricate connections in the visual pathway [8,9], computational modeling of HVS for object recognition directly from its overall anatomy and physiology is difficult. One way to overcome this limitation is modeling of subsystems [10] and their combination based on the HVS structure [11]. Based on experiment with cat's [12] and monkey's cortex [13], Hubel and Wiesel proposed a functional model of visual pathway in HVS. Riesenhuber et al. [14] and Serre et al. [7] extended Hubel and Wiesel's work and introduced a hierarchical model (HMAX) for object recognition with four modules: S1, C1, S2 and C2. The HMAX is one of the superior models in the field of object recognition inspired by hierarchical structure of the HVS with various applications, e.g., real-time visual tracking, face recognition, line segment perception and medical imaging [16–19]. Many recently proposed HVS models are improvements of the HMAX model [20–22].

Therault et al. [23] proposed an extended coding and pooling in the HMAX model. In this model, HMAX is improved by integrating the local filters at the first level and

✉ Mohammadesmaeil Akbarpour
smaeilakbarpour@gmail.com

Mrinal Mandal
mmandal@ualberta.ca

M. Hashemi Kamangar
mh.kamangar@shomal.ac.ir

¹ Electrical Engineering, Shomal University, Amol, Iran

² Electrical and Computer Engineering, University of Alberta, Edmonton, Canada

more complex filters at the last level. Lu et al. [24] proposed a dominant orientation patch matching for HMAX (henceforth referred to as the DHMAX model). In this model, HMAX model is improved by calculating the dominant orientation of the selected patches and implementing patch-to-patch matching.

Norizadeh et al. [25] proposed enhanced model with combination of SIFT algorithm and HMAX model (henceforth referred to as the SHMAX model). Their proposed model consists of two levels of improvement. The first level is increasing the speed of S2 module by comparing the extracted patches with only a few informative patches of input images (rather than the whole image). The second one is selecting the discriminative and distinctive patches in the training stage to increase classification accuracy.

Sufikarimi et al. [15] proposed a HVS-inspired model for object recognition in HMAX (henceforth referred to as the HHMAX model). Human vision intelligently extracts the useful features from information such as corners and edges. These parts of an image are very informative because if they are removed from the image, the HVS cannot recognize the object. They implemented this important information in the HHMAX model.

As HMAX is a hierarchical model with separated modules, some recent works have improved these modules [26,27]. A major problem with the HMAX model is the patch extraction in a random way which extracts two kinds of problematic patches. First, patches from background can produce wrong output. Second, patches with low information from object may provide poor performance.

In this paper, a novel patch selection method is introduced by combining unsupervised object detection and patch ranking to select optimum patch (**P**). Two main contributions are introduced to solve mentioned problems. First, a pool of patches is extracted from the objects based on unsupervised object detection. After C1 module, the produced images are invariant to scale, and patches are extracted from them. All images in within-class involve similar objects, and because of their invariance to the objects, extracted patches from objects have high similarity with images from within-class rather than images from between-classes. Comparison of these similarities is used to define an evaluation metric (*PN*) for extracting patch and producing pool of patches from objects. Second, patches with high discriminative information are selected from the pool based on patch ranking. Beside similarity of patch to images from within-class, its dissimilarity to images from between-class is used to define an evaluation metric (*PR*) for patch ranking. Selected optimum patch (**P**) has the best similarity with images from within-class; as a result, its complement (**I**–**P**) has the highest dissimilarity with images from within-class. It is considered as a new patch for other classes to produce maximum distance in matching and raise classification accuracy.

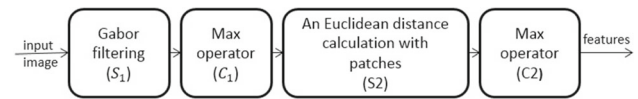


Fig. 1 Simplified schematic for original HMAX model

2 HMAX model

A schematic of the HMAX model is shown in Fig. 1. The HMAX model has four modules: S1, C1, S2 and C2. In the training stage, after the C1 module, patches with different sizes and orientations are extracted randomly. These stored patches are used in the testing stage.

2.1 S1 module

This module emulates the activity of simple cells in the cortex. It applies Gabor filters of different sizes and orientations. Let a set of Gabor filters $g_{\sigma,\lambda,\theta}$ be defined as EQ. 1:

$$g_{\sigma,\lambda,\theta}(x, y) = \exp\left(-\frac{x_0^2 + \gamma y_0^2}{2\delta^2}\right) \cos\left(\frac{2\pi x_0}{\lambda}\right) \quad (1)$$

$$x_0 = x \cos(\theta) + y \sin(\theta)$$

$$y_0 = y \cos(\theta) - x \sin(\theta)$$

where the parameters γ , σ , θ , and λ are the aspect ratio, effective width, orientation and wavelength of the Gabor filter, respectively. HMAX uses 16 Gabor filters with different sizes. Outputs of this module are obtained by convolving the input image (**I**) with a set of Gabor filters $g_{\sigma,\lambda,\theta}$ shown in Eq. (2):

$$S_{l,\theta} = g_{\sigma,\lambda,\theta} * I \quad (2)$$

where l refers to the number of outputs (from 1 to 16).

2.2 C1 module

This module emulates the activity of complex cells in the cortex. It takes the max over different scale outputs produced by the S1 module for scale invariance. These images are segmented into blocks of size $N_s \times N_s$ and with overlap of Δ_s (in each direction) between blocks. Each pixel of the output of C1 module is equal to the maximum of two corresponding blocks (in two consecutive $N_s \times N_s$ scale images). For every two adjacent scales, there is one band (B1 to B8). Therefore, the C1 module produces a total of 32 images (corresponding to four orientations and eight bands).


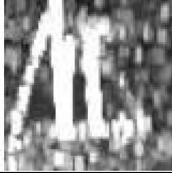
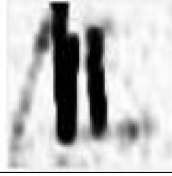
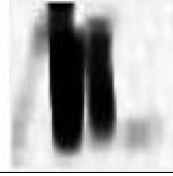
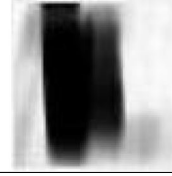
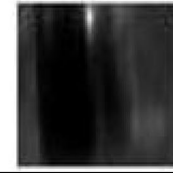

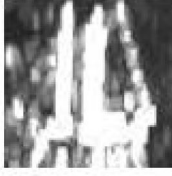

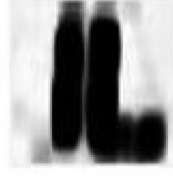
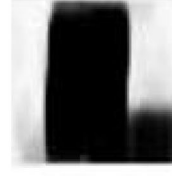
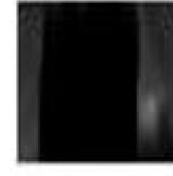




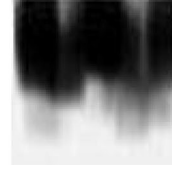

	Original images	C1 outputs	S_2 outputs for first image patches			
			4×4	8×8	12×12	16×16
Boat1						
Boat2						
car-side						

Fig. 2 Example of S_2 outputs. The first column shows three original images Boat1, Boat2 (from the schooner class) and Car-side (from car-side class) from the Caltech101 database. The second column shows

the C1 outputs corresponding to each given image. Columns 3–6 are Euclidean distance between local block of C1 image and slid patches from the Boat1 image with different sizes

2.3 S_2 module

This module calculates the similarity between $C1_{b,\theta}$ and the extracted patches. Each patch (P_i) is slid across an intermediate output image $C1_{b,\theta}$, and the similarity between the local $C1_{b,\theta}$ block (X) and P_i is calculated using Eq. (3):

$$S_2 = \exp(-\beta \|X - P_i\|^2) \quad (3)$$

where the parameter $\beta (>0)$ defines the sharpness of the exponential function and $\|\cdot\|$ is the Euclidian norm. A few example outputs of the S_2 module are shown in Fig. 2. In the first column, three original images from two classes of Caltech101 database are shown. In the second column, outputs of the C1 module for given images are shown. Other columns show Euclidean distance between local block of each C1 image and slid patches with different sizes from the Boat1 image with.

2.4 C2 module

In this module, to find the best matching, the maximum value of the S_2 module is calculated as features for each input image. By calculating this maximum, the similarity between extracted patch and input image is obtained as a C2 outputs. In Fig. 2, extracted patches are from Boat1. Maximum value of S_2 outputs (columns 3 to 6) is C2 outputs. C2 output for

Boat1 is 1 (exact matching) and for Boat2 is more than car-side. (Within-class is more than between-class.)

3 Proposed model

In the proposed model, a new module, Optimum Patch Selection (OPS), is introduced to select patches with high discriminative information from objects in HMAX (henceforth referred to as OMAX) for object recognition in natural images. Natural images consist of cluttered background that makes object recognition a challenge [28]. Two main contributions for patch selection in this proposed module are introduced: first, the patch extraction from objects based on unsupervised object detection to produce a pool of patches and second, patch selection with the best discriminative information from the pool based on patch ranking as an optimum patch (P). As an optimum patch, P has the best similarity with images from within-class and its complement has the highest dissimilarity with images from within-class. Therefore, its complement is considered as a new patch for other classes to produce maximum distance in matching and raise classification accuracy. A schematic of the proposed model is shown in Fig. 3.

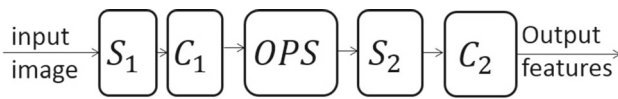


Fig. 3 Schematic of proposed model

3.1 OPS module

In natural images, detecting objects from the background and then extracting patches from them are difficult [29]. This module is introduced to select patches with two main properties: first, belonging to objects and second, involving high discriminative information. To define an evaluation metric based on a novel unsupervised object detection for the first property (i.e., extracting patches from the object), it is considered that all images from each class are involved the same object [30]. As mentioned before, $C1_{b,\theta}$ outputs are invariant to scale. Patches from the object have more similarity with images from within-class ($C2_{w-c}$) than with images from between-classes ($C2_{b-c}$). In this condition, a patch with high probability belongs to object; otherwise, that with high probability belongs to background and produces wrong outputs. Assume that the number of classes are NC and each class involve NT images and evaluation metric (PN_{score}) is calculated for extracted patches from nt th training image from nc th class. So there are $(NT-1)$ images in within-class and $(NC-1)NT$ images in between-classes to compare with patches extracted from selected image. For each patch, similarity with each image from within-class should be compared with similarity with all images from between-classes. Evaluation metric (PN_{score}) for each patch is defined as Eq. (4):

$$PN_{score}(k, m, n) = \begin{cases} 1; & C2_{w-c}(k) > C2_{b-c}(m, n) \\ 0; & \text{otherwise} \end{cases} \quad (4)$$

where index k is the image number from within-class (from 1 to NT , $k \neq nt$), index m is the number of between-classes (from 1 to NC , $m \neq nc$), and n is the number of images in between-classes (from 1 to NT). After calculating PN_{score} , the probability of belonging patches to object (PN) can be calculated. Let PN be the sum of all elements of PN_{score} matrix. The normalized PN is calculated using Eq. (5).

$$PN = \frac{\sum_{k=1, \neq nt}^{NT} \sum_{m=1, \neq nc}^{NC} \sum_{n=1}^{NT} PN_{score}(k, m, n)}{(NT-1)(NC-1)NT} \quad (5)$$

By defining a desirable threshold (thr), a pool of patches from an object is made based on unsupervised object detection. These patches have PN with a higher value than the threshold ($thr < PN \leq 1$) and with high probability are from objects.

To define an evaluation metric based on patch ranking for the second property (selecting patches with high discriminative information from the pool), beside PN , the distance

between similarity of patch with images from within-class and between-class is vital for classification [31]. More distance between these two similarities produces more distance between features and so produces higher classification accuracy rate. For each patch, evaluation metric to define this distance is comparison of its similarity to all images from within-class with similarity to all images from between-classes. This distance is defined as selection value (SV) obtained using Eq. (6).

$$SV = \frac{\sum_{k=1, \neq nt}^{NT} C2_{w-c}(k)}{\sum_{m=1, \neq nc}^{NC} \sum_{n=1}^{NT} C2_{b-c}(m, n)} \quad (6)$$

Finally, for calculating the patch ranking (PR) for discriminative information, both SV and PN are used in formula obtained using Eq. (7):

$$PR = 2^{\alpha(PN-1)}SV \quad (7)$$

where α is a constant (in our simulation $\alpha = 100$). By considering PR as a chance for selecting, the best patches are selected by using a Roulette wheel. All steps for producing pool of patches from objects based on unsupervised object detection and calculating patch ranking are given in Algorithm 1.

Algorithm 1: Calculating Selection Probability

Inputs: extracted patch (\mathbf{P}) from image (\mathbf{I}); all images from within-class except \mathbf{I} ; all images from between-classes; threshold (thr); NC (number of classes); NT (number of images in each class).

Output: discriminative information of patch as a Patch Ranking (PR) if it belongs to object; zero if it do not belongs to object.

```

1 : sum(PNscore) ← 0;
2 : sum(C2w-c) ← 0;
3 : sum(C2b-c) ← 0;
4 : for all images except I do
5 :   sum(C2w-c) ← sum(C2w-c) + C2w-c
6 :   sum(C2b-c) ← sum(C2b-c) + C2b-c
7 :   if C2w-c > C2b-c then
8 :     sum(PNscore) ← sum(PNscore) + 1
9 :   end
10 : end
11 : PN ← sum(PNscore) / ((NT-1)(NC-1)NT)
12 : if PN > thr then
13 :   PR ← 2α(PN-1) × sum(C2w-c) / sum(C2b-c)
14 : else
15 :   PR ← 0
16 : end
  
```

Assume \mathbf{P}_i is the best selected patch in the proposed model ($PN=1$ and SV is the highest), so it has the best similarity with all images from i -class and its complement has the highest dissimilarity with images from this class. In the proposed model, this complement is considered as a new patch for other classes to raise the distance between feature of classes.

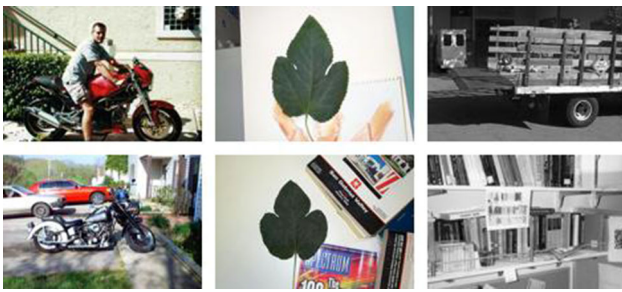


Fig. 4 Sample images from Caltech5 database

4 Experimental results

Three evaluation metrics of classification (sensitivity, specificity and classification accuracy) are used to evaluate proposed model (OMAX) against HMAX [7], and three recent models with the same framework with HMAX: DHMAX [24], HHMAX [15] and SHMAX [25]. Results are reported on three different image databases: Caltech5 [32], Caltech101 [33] and Graz-01 [34].

4.1 Image databases

In this paper, Caltech5, Caltech101 and Graz-01 databases are used for performance evaluation. All images are changed from RGB to grayscale with size 140×140 .

Caltech5: This database contains 3122 natural images with 5 object classes and a background: frontal-face, motor, rear-car, airplane, leaf and background. Leaf and motor are considered as positive images and background as negative images. Two sample images from each class are shown in Fig. 4.

Caltech101: This database contains 9144 natural images with 101 object classes and a background. Airplane and car-side are considered as positive images and background as negative images. Two sample images from each class are shown in Fig. 5.

Graz-01: This database contains 1103 images with different objects such as bikes, people, motor and shoes. Bike and people are considered as positive images and background images (*no_bike_no_people*) as negative images. Two sample images from each class are shown in Fig. 6.

4.2 Evaluation metrics

In this paper, support vector machine (SVM) is used to classify the test images. As explained in proposed model, many extracted patches are eliminated during producing pool. Besides, many patches with low discriminative information from the pool are eliminated. Furthermore, in some bands and orientations there are no desirable patches based on the proposed model and hence no patches are selected



Fig. 5 Sample images from Caltech101 database

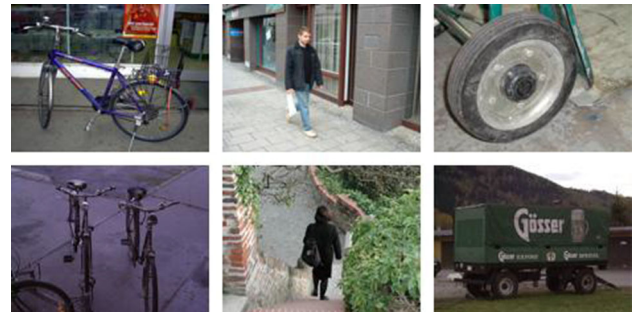


Fig. 6 Sample images from Graz-01 database

in these cases. Therefore, the number of selected patches in the proposed model is much less than the selected patches in the original HMAX and other recent models with the same framework. In other words, the proposed model increases the speed of classification.

For classification performance evaluation, the following quantitative metrics (sensitivity (TPR), specificity (TNR) and classification accuracy (A_c)) are used as Eq. (8):

$$\begin{aligned} \text{TPR} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{TNR} &= \frac{\text{TN}}{\text{FP} + \text{TN}} \\ A_c &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \end{aligned} \quad (8)$$

Confusion matrix [35] is also used to evaluate the proposed model. Besides confusion matrix, Kappa coefficient (κ) is calculated as Eq. (9).

$$\kappa = \frac{N \sum_{i=1}^{\text{NC}} n_{ii} - \sum_{i=1}^{\text{NC}} n_{i+n+i}}{N^2 - \sum_{i=1}^{\text{NC}} n_{i+n+i}} \quad (9)$$

where n_{ii} is diagonal element of matrix, n_{i+} is sum of i th row, n_{+i} is sum of i th column, and finally, N is equal to the total number of input images.

Fig. 7 Classification accuracy for each selected class. Top row: Caltech5 database classification accuracy (motor and leaf). Middle row: Caltech101 database classification accuracy (car-side and airplane). Bottom row: Graz-01 database classification accuracy (people and bike)

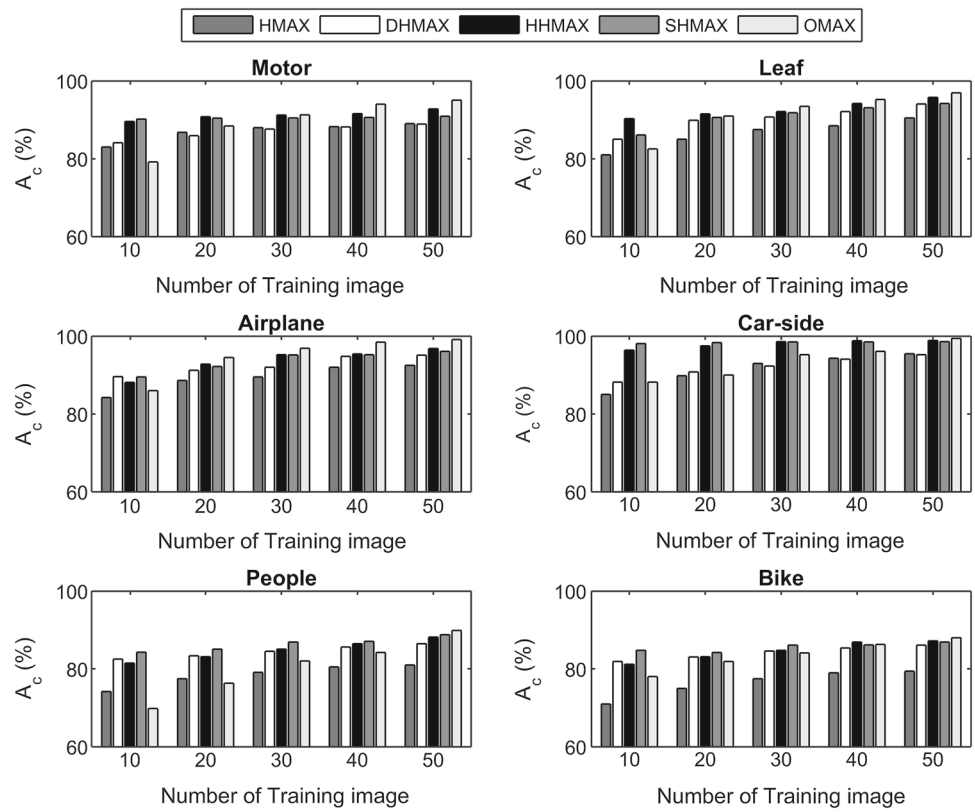


Table 1 C2 output for HMAX and OMAX for patches selected from Boat1

Model	Image class	B1	B2	B3	B4	B5	B6	B7	B8
HMAX	Boat2	0.853	0.901	0.863	0.832	0.906	0.828	0.875	0.842
	Car-side	0.812	0.751	0.801	0.803	0.868	0.819	0.815	0.806
	difference	0.041	0.15	0.062	0.029	0.038	0.009	0.06	0.036
OMAX	Boat2	0.912	0.946	0.942	0.942	0.978	0.943	0.922	0.949
	Car-side	0.643	0.593	0.632	0.591	0.547	0.587	0.701	0.666
	Difference	0.269	0.353	0.31	0.351	0.431	0.356	0.221	0.283

Table 2 Confusion matrix on Caltech5 databases for TN = 50

	Airplane	Motor	Rear-Car	Leaf	Faces
Airplanes	93.4	1.7	2.6	1.1	1.2
Motor	2.8	92.2	1.4	1.8	1.8
Rear-Cars	2.6	3.2	91.6	1.1	1.5
Leaf	2.1	2.6	4.3	89	2
Faces	2.9	2.1	3.5	3.1	88.4

4.3 Results

In order to determine the effect of the number of training images on the classification accuracy, two object classes from each database are selected (leaf and motor from Caltech5, airplane and car-side from Caltech101, bike and people from Graz-01). The object class is considered as positive, and the

background class is considered as negative. Proposed model is trained with 10, 20, 30, 40 and 50 images selected randomly from each class. 70 images are selected randomly from each class (except training image) as testing images. Classification accuracy from these test images is shown in Fig. 7 versus number of training images. All experiments are repeated ten times, and the mean value is reported as the result. In all simulations, *thr* has been set as 0.9. The proposed model provides a superior performance in higher number of training images.

Table 1 compares the C2 features of images shown in Fig. 2. Selected patches are from Boat1, So the similarity between these patches and other boat (image from within-class: Boat2) should be more than the similarity between them and car-side (image from between-class). For more classification accuracy, the similarity (C2 features) for Boat2 should be more than the C2 features for car-side. (Larger dif-

Table 3 Average performance comparison: proposed model with other models for TN = 50

	Caltech5 (motor and leaf)			Caltech101 (airplane and car-side)			Graz-01 (people and bike)		
	TPR	TNR	Ac	TPR	TNR	Ac	TPR	TNR	Ac
HMAX	92.65	86.85	89.75	96.45	91.55	94	82.45	77.95	80.2
DHMAX	94.23	88.73	91.48	97.45	92.85	95.15	89	83.6	86.3
HHMAX	96.9	91.7	94.3	99.5	96.2	97.85	90.95	84.45	87.7
SHMAX	95.3	89.8	92.55	98.6	96.1	97.35	90.95	84.75	87.85
OMAX	97.73	94.23	95.98	99.7	98.8	99.25	92.05	85.85	88.95

ference is better for the classifier.) As shown in Table 1, the difference between C2 features in proposed model is more than the difference between C2 features in original HMAX significantly.

Table 2 shows the confusion matrix for proposed model on all classes of caltech5 database. For each class, 50 and 70 images are chosen randomly as training and testing images, respectively. Results are for testing images. The best classification rate is for airplane. Because of their spatial shapes, the optimum patches with larger distance from other classes can be extracted. For evaluating Table 2, Kappa coefficient is calculated ($\kappa = 88.65\%$).

Table 3 lists the average performance of the proposed model (OMAX) as well as existing models, HMAX [7], DHMAX [24], HHMAX [15] and SHMAX [25]. Results are average of two selected classes from each database, and the number of training images is 50 (TN = 50). As shown, the proposed model is clearly superior.

5 Conclusion

In this paper, an enhanced HMAX model based on optimum patch selection is proposed to object recognition in natural images. A major limitation of the HMAX is its random patch extraction. In the proposed model, an optimum patch selection (OPS) module is introduced. The patch selection is done in two steps. First, a pool of patches is extracted using novel unsupervised object detection. Second, the optimal patches with high discriminative information are then selected using patch ranking. After selecting optimum patch, its complement is considered as a new patch for other classes. Experimental results on different databases show that the proposed model provides a superior performance over the state-of-the-art HMAX-based models.

References

- Ahmadvand, A., Kabiri, P.: Multispectral MRI image segmentation using Markov random field model. *SIViP* **10**(2), 251–258 (2016)

- Kurmi, Y., Gangwar, S., Agrawal, D., Kumar, S., Srivastava, H.S.: Leaf image analysis-based crop diseases classification. *SIViP* (2020)
- Sooksatra, S., Kondo, T., Bunnun, P., Yoshitaka, A.: Head-light recognition for night-time traffic surveillance using spatial-temporal information. *SIViP* **14**(1), 107–114 (2020)
- Soon, F.C., Khaw, H.Y., Chuah, J.H., Kanesan, J.: Vehicle logo recognition using whitening transformation and deep learning. *SIViP* **13**(1), 111–119 (2019)
- Varsaki, E.E., Fotopoulos, V., Skodras, A.N.: Data hiding based on image texture classification. *SIViP* **7**(2), 247–253 (2013)
- LeCun, Y., Huang, F.J., Bottou, L.: Learning methods for generic object recognition with invariance to pose and lighting. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. 2004, pp. II-104. IEEE
- Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., Poggio, T.: Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(3), 411–426 (2007)
- Provenzi, E.: Rudiments of human visual system (HVS) features. In: Computational Color Science, pp. 1–11 (2017)
- VanRullen, R.: The power of the feed-forward sweep. *Adv. Cogn. Psychol.* **3**(1–2), 167 (2007)
- Gupta, R., Mishra, A., Jain, S.: A semi-blind HVS based image watermarking scheme using elliptic curve cryptography. *Multimed. Tools Appl.* **77**(15), 19235–19260 (2018)
- Peelen, M.V., Downing, P.E.: Category selectivity in human visual cortex: beyond visual object recognition. *Neuropsychologia* **105**, 177–183 (2017)
- Hubel, D.H., Wiesel, T.N.: Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* **148**(3), 574–591 (1959)
- Hubel, D.H., Wiesel, T.N.: Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* **195**(1), 215–243 (1968)
- Riesenhuber, M., Poggio, T.: Hierarchical models of object recognition in cortex. *Nat. Neurosci.* **2**(11), 1019–1025 (1999)
- Sufikarimi, H., Mohammadi, K.: Role of the secondary visual cortex in HMAX model for object recognition. *Cogn. Syst. Res.* **64**, 15–28 (2020)
- Jazlaeiyan, M., Seyedin, S., Motamedi, S.A.: Enhanced Brain Inspired Model for Face Categorization Using Mutual Information Maximization. In: 2018 25th National and 3rd International Iranian Conference on Biomedical Engineering (ICBME) 2018, pp. 1–6. IEEE
- Bagheri, S., Saraf Esmaili, S.: An automatic model combining descriptors of gray-level co-occurrence matrix and HMAX model for adaptive detection of liver disease in CT images. *Signal Process. Renew. Energy* **3**(1), 1–21 (2019)
- Cai, B., Xu, X., Xing, X., Qing, C.: BIT: Bio-inspired tracker. In: 2015 IEEE International Conference on Image Processing (ICIP) 2015, pp. 2850–2854. IEEE

19. Liu, X., Cao, Z., Gu, N., Nahavandi, S., Zhou, C., Tan, M.: Intelligent line segment perception with cortex-like mechanisms. *IEEE Trans. Syst. Man Cybern. Syst.* **45**(12), 1522–1534 (2015)
20. Selvaraj, A., Russel, N.S.: Bimodal recognition of affective states with the features inspired from human visual and auditory perception system. *Int. J. Imaging Syst. Technol.* **29**(4), 584–598 (2019)
21. Akbarpour, M., Mehrshad, N., Razavi, S.-M.: Object recognition inspiring HVS. *Indones. J. Electr. Eng. Comput. Sci.* **12**(2), 783–793 (2018)
22. Zhang, H.-Z., Lu, Y.-F., Kang, T.-K., Lim, M.-T.: B-HMAX: A fast binary biologically inspired model for object recognition. *Neurocomputing* **218**, 242–250 (2016)
23. Theriault, C., Thome, N., Cord, M.: Extended coding and pooling in the HMAX model. *IEEE Trans. Image Process.* **22**(2), 764–777 (2012)
24. Lu, Y.-F., Zhang, H.-Z., Kang, T.-K., Lim, M.-T.: Dominant orientation patch matching for HMAX. *Neurocomputing* **193**, 155–166 (2016)
25. Cherloo, M.N., Shiri, M., Daliri, M.R.: An enhanced HMAX model in combination with SIFT algorithm for object recognition. *SIViP* **14**(2), 425–433 (2020)
26. Qiao, H., Xi, X., Li, Y., Wu, W., Li, F.: Biologically inspired visual model with preliminary cognition and active attention adjustment. *IEEE Trans. Cybern.* **45**(11), 2612–2624 (2014)
27. Zhang, Y., Zhang, L., Li, P.: A novel biologically inspired ELM-based network for image recognition. *Neurocomputing* **174**, 286–298 (2016)
28. Filali, J., Zghal, H.B., Martinet, J.: Ontology-based image classification and annotation. *Int. J. Pattern Recognit. Artif. Intell.* **34**(11), 2040002 (2020)
29. Xu, Q., Wang, F., Gong, Y., Wang, Z., Zeng, K., Li, Q., Luo, X.: A novel edge-oriented framework for saliency detection enhancement. *Image Vis. Comput.* **87**, 1–12 (2019)
30. Bai, S., Matsumoto, T., Kudo, H., Ohnishi, N., Takeuchi, Y.: Scene classification based on category-specific representations created through prototype feature selection. In: *Proceedings of the 27th Conference on Image and Vision Computing New Zealand*, pp. 174–179 (2012)
31. Cheng, G., Yang, C., Yao, X., Guo, L., Han, J.: When deep learning meets metric learning: remote sensing image scene classification via learning discriminative CNNs. *IEEE Trans. Geosci. Remote Sens.* **56**(5), 2811–2821 (2018)
32. Fergus, R., Perona, P., Zisserman, A.: Object class recognition by unsupervised scale-invariant learning. In: *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings*, pp. II–II. IEEE (2003)
33. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In: *2004 Conference on Computer Vision and Pattern Recognition Workshop*, p. 178. IEEE (2004)
34. Opelt, A., Fussenegger, M., Pinz, A., Auer, P.: Weak hypotheses and boosting for generic object detection and recognition. In: *European Conference on Computer Vision*, pp. 71–84. Springer (2004)
35. Stehman, S.V.: Selecting and interpreting measures of thematic classification accuracy. *Remote Sens. Environ.* **62**(1), 77–89 (1997)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.