Multimed Tools Appl
https://doi.org/10.1007/s11042-017-5514-z

Image recommendation based on a novel biologically inspired hierarchical model

Yan-Feng Lu[1] **· Hong Qiao**[2,3] **· Yi Li**[4] **· Li-Hao Jia**[1]

Received: 31 October 2016 / Revised: 3 May 2017 / Accepted: 6 December 2017

Abstract Image recommendation has become an increasingly relevant problem recently, since strong demand to quickly find interested images from vast amounts of image library. We describe a biologically inspired hierarchical model for image recommendation. The biologically inspired model (BIM) for invariant feature representation has attracted widespread attention, which approximately follows the organization of cortex visuel. BIM is a computation architecture with four layers. With the image data size increases, the four-layer framework is prone to be overfitting, which limits its application. To address this issue, we propose a biologically inspired hierarchical model (BIHM) for feature representation, which adds two more discriminative layers upon the conventional four-layer framework. In contrast to the conventional BIM that mimics the inferior temporal cortex, which corresponds to the low level feature, the proposed BIHM adds two more layers upon the conventional framework to simulate inferotemporal cortex, exploring higher level feature invariance and selectivity. Furthermore, we firstly utilize the BIHM in the image recommendation. To demonstrate the effectiveness of proposed model, we use it to image classification and retrieval tasks and

✉ Yan-Feng Lu
yanfeng.lv@ia.ac.cn

Hong Qiao
hong.qiao@ia.ac.cn

Yi Li
littlepear@ncu.edu.cn

Li-Hao Jia
lihao.jia@ia.ac.cn

1 Research Center for Brain-inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, Beijing, China

2 The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China

3 CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai, China

4 School of Information Engineering, Nanchang University, Nanchang, China

Published online: 19 December 2017

perform experiments on CalTech5, Imagenet and CalTech256 datasets. The experiment results show that BIHM exhibits better performance than the conventional model in the tasks and is very comparable to existing architectures.

Keywords Image recommendation · Classification · Biologically inspired model · Image retrieval · Feature representation

1 Introduction

In recent years, there are a large number of digital images made every day in the internet. Effective image recommendation, and retrieval tools are strongly demanded by consumers, including fashion, shopping online, sensing, medicine, and so on [17, 44, 47, 50].

Efficient image retrieval techniques play important roles in the image recommendation system. Among them, content based image retrieval (CBIR) has drawn widespread research attentiveness in the last decade. Users express information requirements by importing a search query in the form of images. And then the retrieval system retrieves from the collected image library that is close to the user's query. In the CBIR, images are indexed by the image content, such as texture, spatial layout, color, and shapes [23, 34].

The retrieval performance of a CBIR system essentially depends on the feature representation, which has been carried out extensive research by researchers for decades [2, 8, 45]. In the visual system, humans trend to adopt high-level features (concepts) to represent images and measure their similarity. By contrast, While the most computer vision techniques extracted features with low-level features (color, shape, structure, texture and so on.) [11, 43].

To address this issue, several methods have been proposed. The spreading activation theory is used to give annotation for the untagged images [7, 52]. Low level feature extraction plays an important role in the CBIR system. Features are extracted from an entire image or a region [23, 31]. The Bayes point machines method used to give the soft annotation to the unlabeled image [16]. A tree structured self-organizing map is employed to deal with the large unlabeled database, working as hierarchical levels [3].

Although varieties of methods have been proposed, it remains a very challenging problems in the current system, which is mainly from the "semantic gap" from the difference between low level pixel values extracted by computers and human's high level concepts [2, 23, 34, 45]. From a high view, this issue is the essential challenge of Intelligence, which is how to teach machines to execute real-world like human.

Recently, important developments have been made in the research of brain science [14, 15, 35, 37]. The findings in the primary visual cortex V1 area are of significance. While researching the V1 area, Hubel & Wiesel discovered that the visual cortex analyzes features into various ways with different spatial orientations and frequencies [46]. The discovery gives an important support to early neuroscience theories. Based on these theories, Riesenhuber & Poggio described an original calculation framework for object recognition, called biologically inspired model (BIM) that tends to model the cognitive mechanism of the visual cortex [48]. Serre et al. upgraded the original BIM model and presented the standard BIM [41], which shows that the visual framework significantly improve the performance of object recognition.

The Standard BIM Model made a big step forward for extending neurobiological models to deal with real-world vision tasks, and many enhanced models have been proposed about on the basic framework of it. Mutch & Lowe [32]modified the model by constraining the number of feature input, inhibiting S1/C1 outputs, and increasing feature selection. Huang et al. improved the random processes in patch selection with a novel selecting strategy based on energy function which is combined with visual attention theory [13]. Qiao et al. developed a modified model by introducing association to the BIM architecture and showed competitive results on object recognition tasks [36]. Lu et al. proposed a novel receptive field in the S1 layers and upgraded the framework by novel patch selection and matching processes [27–29, 51]. These approaches all obtain better performances by incorporating some biologically motivated properties in addition. Besides, there are other types of models which are inspired by different properties of certain areas of visual cortex. The trainable COSFIRE filters [1], which model the selectivity for parts of contours or line segments based on the properties of shape-selective neurons in area V4, also achieve success in real applications. In conclusion, all these systems have shown that the relevant biological findings are helpful for constructing more robust computer vision algorithms.

BIM is a calculation model with S1, C1, S2, and C2 four layers, which concentrates on the invariance and selectivity of features [41]. With the image data size increases, the four-layer framework is prone to be overfitting in the big data cases, which limits its application. To address this weakness, motivated by biology, we describe a biologically inspired hierarchical model (BIHM) for image recommendation, which adds two more discriminative layers upon the conventional four-layer framework [30]. In contrast to the conventional BIM that successfully mimics the inferior temporal cortex (from V1 to V4) of the human visual system, which corresponds to the low level feature invariance and selectivity, the proposed BIHM adds two more layers based on the conventional BIM framework to simulate up to inferotemporal cortex (i.e., PIT and AIT), exploring higher level feature invariance and selectivity.

The remaining part of the article is organized as follows: in Section 2, we give an introduction about the conventional BIM; in Section 3, we propose the BIHM method and utilize the BIHM in image recommendation; in Section 4, we show experimental results based on three databases; finally, in Section 5, we concludes this paper.

2 Biologically inspired model review

The biologically inspired model (BIM) [41] presented by Serre et al. has attracted widespread attention recently. As an extension of the original model [37], it has been successfully applied in various recognition tasks: From single object recognition in the clutter condition to multi-categorization as well as the understanding of complex scene. BIM attempts to quantitatively simulate the information transfer process in the visual ventral pathway, which conducts the visual information from the lateral geniculate nucleus (LGN) through V1, extra striate visual areas II (V2), to IV (V4). It composed by four layers of calculating units in a hierarchical feed forward structure: S1, C1, S2 and C2 layers as the Fig. 1 shown.

The units in S1 layer are corresponding to the simple cells of the visual cortex, which compute the response for input image by Gabor filter bank. The C1 layer is corresponding to cortical complex cell layer and shows robustness to scale and shift transformations by pooling the afferent S1 units with the MAX operation in the same scale and orientation band. It tends to have larger receptive fields and increases the robustness to deformations from layer S1 to C1.

In the S2 stage, units pool C1 units from a local neighborhood across all orientations. A mass of patches are chose from the C1 layers of training images at random before that. Then the S2 units behave as radial basis function and pool over the C1 units in a Gaussian-like tuning way on the Euclidean distance between an input patch and a stored prototype. The C2 layer pools the S2 units over all scales and positions with a global maximum operation, which devotes to obtain a shift- and scale-invariant responses. Therefore, for the sampled k prototype patches, a k-dimensional feature vector is finally obtained after the four-stage features extraction, which has larger receptive fields and shows shift- invariant and scale-invariant properties (Figs. 1 and 2).

3 Biologically inspired hierarchical model

BIM has been demonstrated the strength on in various recognition tasks: From single object recognition in the clutter condition to multiclass categorization as well as the understanding of complex scene that depends on the recognition of both texture-based and shape-based objects. However, with the image data size increases, the conventional BIM four-layer framework is prone to be overfitting in the big data cases, which limits its application. To address this issue, motivated by biology, we propose a biologically inspired hierarchical model (BIHM) for image representation, which adds two more discriminative layers upon the conventional four-layer framework. In contrast to the conventional BIM that tends to involve over fitting in big data cases, BIHM is pond to describe more robust and representative features. Based on the proposed BIHM model, we build an image recommendation system. The flow chart of the BIHM is shown in Fig. 2.

Build upon the theories of Hubel et al. on visual cortex [14, 15, 35, 37, 46], the simple cells are not sensitive to illumination and need edge-like response at a particular phase, and position. The complex cells respond well to bars with a particular phase, while they are not sensitive to both phase and position of the bar in the receptive fields (RFs). At the upper layers the hypercomplex cells not only respond to bars in the phase and position invariant way, but also are selective to the bars with a specific length. Hubel and Wiesel suggested that this increasingly invariant and complex feature representations should be built by integrating the inputs from lower levels.

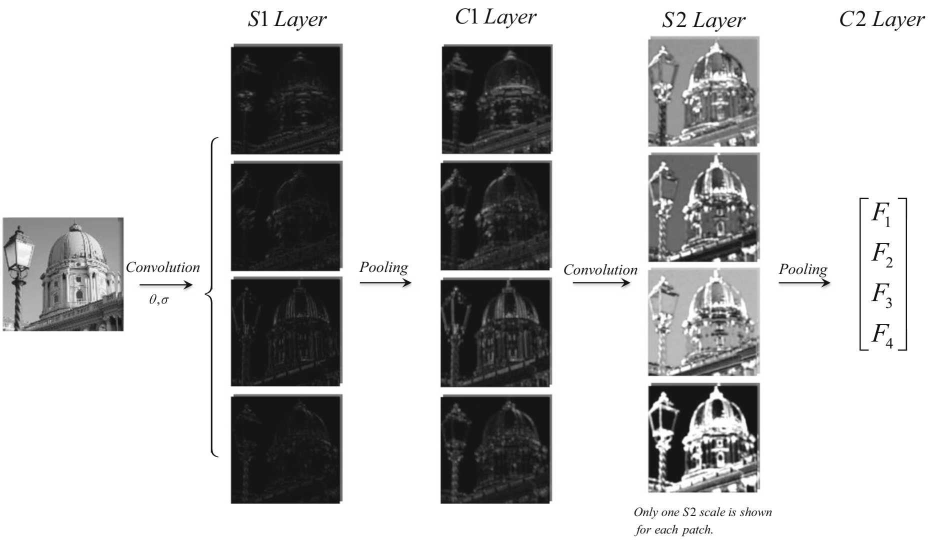

Fig. 1 BIM structure overview

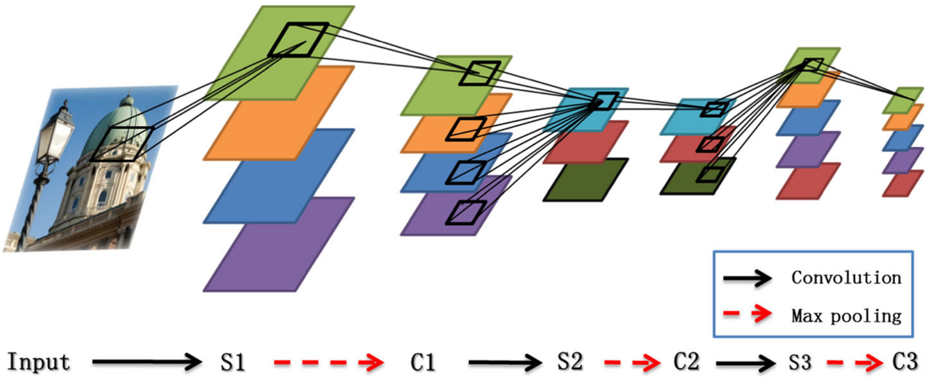

Fig. 2 Flow chart of the BIHM

Two sort of functional layers exists in the visual cortex framework: the S layers consisted of simple cells are interleaved with the C layers consisted of complex cells. In brief, along the hierarchy, each S layer increase the feature selectivity by tuning to features of increasing complexity, and each C layers increase the invariance to 2D transformations such as slight changes in position and scale by a max pooling operation [41].

Therefore, simple units in the S_k layer pool the units from a related local spatial neighborhood in the precedent C_{k-1} layer with various selectivity [40]. As a consequence, the complexity of the prior response of units increases from layer C_{k-1} to S_k. The pooling operation at the S stage is a Gaussian-like tuning function. The response of a simple unit can be defined as:

$$y = \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^{ns}\left(\mathrm{w}_j - \mathrm{x}_j\right)^2\right) \tag{1}$$

where σ denotes the sharpness of the tuning around the prior responses of the unit which corresponds to the weight vector $(w_1, \ldots, w_{ns})$. The maximal response of the unit is equal to one when the current input x execute an exact match.

Complex units in the C_k layer pool the prior S_k layer units with the same selectivity but at small different locations and scales to increase the robust to plane deformations from layer S_k to C_k. The pooling operation at the complex C stage is adopted a MAX operation [39]. The response of a complex unit is corresponding to the response of the strongest $(x_1, \ldots, x_{ns})$ from prior S_k layer. The mathematical expression of the complex unit operation is defined as:

$$y = \max_{j=1\ldots nc} x_j \tag{2}$$

In the next stages of the model, by interleaving these two operations, i.e., MAX over retinotopically organized inputs with the same prior response but small different scales and locations and TUNING over inputs with different preferred stimuli, an increasingly complex and invariant representation is built [18, 39]. From V1, the visual information is routed to V2, V4 and IT, which has been shown to be critical in the ability of primates to perform invariant recognition. This is done via two routes: a main route that follows the hierarchy of cortical stages strictly (i.e., step-by-step) as well as several by pass routes which skip some of the stages. It may help create a richer repertoire of features with various degrees of selectivity and invariances by pass routes.

S3 stage In the S3 stage, the same step is iterated one more time to increase the complexity of the prior response at the S3 level, where the response of S3 units with various selectivity are combined with a Gaussian-tuning operation to generate more complex selectivity [39].

The S3 units present the similarity between an stored prototype and the previous C2 layers and in a Gaussian-like tuning way by Euclidean distance. The mathematical expression of the corresponding S3 layers are given by:

$$S3 = \exp\left(-\beta^* \sum_{i=1}^{N} \left(C2(m,n) - F_i\right)^2\right), \tag{3}$$

where $C2(m, n)$ is the afferent C2 layer with a specific scale m and orientation n, and F_i denotes a sampled patch from the prior C2 layers, βdefines the sharpness of the exponential function, N is the number of the sampled patches.

C3 stage In the C3 stage, the C3 units acquired by pooling the S3 units with the same selectivity at adjacent scales and positions. The C3 units show the similar selectivity to complex features with the S3 units, but with a broader size of invariance [39, 40]. The S3 and C3 layers give a description of largely tuned shape.

The set of invariant C3 responses can be calculated by doing a global maximum value of inputting S3 units across all positions and scales. The responses of the C3 layers are given by:

$$C3 = \max_{(x,y,\sigma)} \left(S3(x, y, \sigma)\right). \tag{4}$$

where (x, y) denotes the position of S3 units and σ is the corresponding scale. The export is a feature vector with C3 values. The vector is used as the C3 features in the experiment tasks.

The S3 and C3 steps follow the processing modes in S2 and C2, their further inter-leaving and max-like pooling operations with the inferior layers bring in better feature selectivity and invariance. The deeper features with more discriminative information benefit the BIHM model in the cluttered recommendation tasks.

The content based recommendation system extracts the images from image library by analyzing their contents. The system mainly performs two steps, i.e. Feature representation and similarity measurement [25]. The performance of the system essentially relies on the method of feature representation. The described BIHM exhibits a balanced between invariance and selectivity, which fits well the recommendation task. In the recommendation system, the images in the dataset are described by BIHM as multi-dimensional feature vectors, which are stored in a feature dataset. To accurately get the required images, users should provide the system with example images, and the BIHM represents them as internal feature vectors. Then the similarities between the feature vectors of example images and the stored feature dataset are calculated and recommendation is performed with the help of an image indexing. We then could get the recommended images.

4 Experiments

In this part, we design some experiments to do the evaluation of BIHM in the image recommendation tasks. In Section 4.1, we compare the BIHM with BIM and SIFT on Caltech05 database [41]. In Section 4.2, we design a image classification experiment to evaluate the BIHM using Imagenet database [6]. In Section 4.3, we execute an image retrieval task to evaluate BIHM on Caltech256 [10] and compare it with some related methods. Given the various appearance transformation of the images, we applied the position-scale-invariant C3 features of BIHM, and passed the features to a classifier to execute classification. (In the experiments of this article, we select the linear Lib-SVM [4] as the classifier). We use two standard evaluation measures widely used in image recommendation tasks as performance evaluation metrics, including the classification precision and mean average precision (mAP).

4.1 Comparison between BIM and BIHM

The CalTech5 dataset contains the cars, frontal faces, aeroplanes, leaves, and motorcycles, as shown in Fig. 3. We applied this database to evaluate BIHM and make comparisons with the conventional BIM and the SIFT algorithm [26].

To make the experiment at a feature level and ensure a fair comparison between the methods, we neglected the position information from SIFT, because it was shown in [20, 41] that structural information does not appear to improve classification performance. We use the same experimental seting in [41]: We randomly choose 5000 reference patches from the training set. Given an input image, we calculated the minimum Euclidean distance between all its patches and the reference prototypes, thus we could obtain a feature vector with size 5000. We compared the position-scale-invariant *C2* features from the conversional BIM and the updated C3 features from BIHM with the SIFT features. We pass these features to an SVM and perform a present/absent classification task. We chose the classification precision for different numbers of features as the performance evaluation metric.

In this experiment, we chose 25 images at random from each category from the CaltTech5 database as positive training images and 25 different images from backgrounds as the negative training set. In the test stage, 100 different images (every category of the CaltTech5 database) and 100 other background images were chosen at random as a testing set. A various number of features (i.e., 10, 50, 100, 200, 500, 1000, 2000 and 4000) were obtained by selecting them from the 5000 available at random to train the models. The results in this part were tested with 10 times.

Fig. 3 Sample images of the CalTech5 datasets. The last one denotes a background image

Figure 4 shows the experimental results of the CalTech5 database with different numbers of features. Generally, it was demonstrated that BIHM exceeded BIM and SIFT in precision in most classes in this database. The SIFT features are adept at the detection of a transformed seen image, but they may lack discriminability in a more general classification task [28]. BIHM and BIM significantly exceeded SIFT for the faces, cars, leaves, and airplanes; in most cases, BIHM was clearly superior to BIM, especially, when the feature number is bigger and bigger, the improvement of BIM performance is limited due to the overfitting. In contrast, BIHM performance maintains growth trend with the increasing feature number. Therefore, in most image categories, BIHM shows superior performance.

4.2 BIHM based image classification

Imagenet is a huge and challenging database of object categories with millions of images [6]. On this database, We further evaluated BIHM in image classification task. To do a comparison experiment at the feature level, we employed the position-scale-invariant C3 features from BIHM. We conducted this experiment utilizing 2000 features. We selected 100 classes from Imagenet dataset randomly, and then we randomly selected 50 images from every class for training and 150 other images for testing. A multiclass linear SVM was choose as classifier. Table 1 gave the classification results on the Imagenet database.

As shown in Table 1, BIHM achieved 74.5% classification precision. It outperformed the related models, i.e., SIFT-variant +BoW, MKL, SIFT + GMM [9, 22, 38]. Even compared with the famed deep learning network CNN [19], BIHM still shows a competitive result. We should note that the CNN model use high-level features (20th layer), while BIHM employs C3 features that are relative low level (6th layer). In general, BIHM is effective and competitive computational model in the image classification task.

4.3 BIHM based image retrieval

The Caltech256 is a challenging dataset that contains 257 object categories and a background class including 30,607 images. Sample images are presented in Fig. 5. In this experiment, we evaluate the performance of BIHM in the retrieval task on the Caltech256 dataset. We did this experiment by 2000 features for the image retrieval procedure. We follow the same experimental setup in [45]. The employed subsets are set as 10, 20, 50 classes. The images from each category were randomly divided into a training set with 40 images and a test set with 25 image. We compared with GIST [33], BoW [5], conventional BIM, and SPM [21].To do this experiment at the feature level and perform a fair competition between the methods, we choose the Euclidean distance as the distance metric learning algorithm. The measuring results are mean average precision (mAP).

As shown in Table 2, we achieved 32.5% mAP using 10 classes retrieval, 22.3% using 20 classes and 18.75% using 50 classes. In the retrieval task, our proposed method exceeds the relevant methods, i.e., GIST, BoW, and SPM, in all three cases. When compared with the conventional BIM method, we note that the proposed model with newly added two layers works well, and has an improvement in the task. Our results were competitive with GIST, BoW, BIM, and SPM. As a whole, BIHM describes an effective computational model for feature representation, and show competitive results in the image recommendation task.

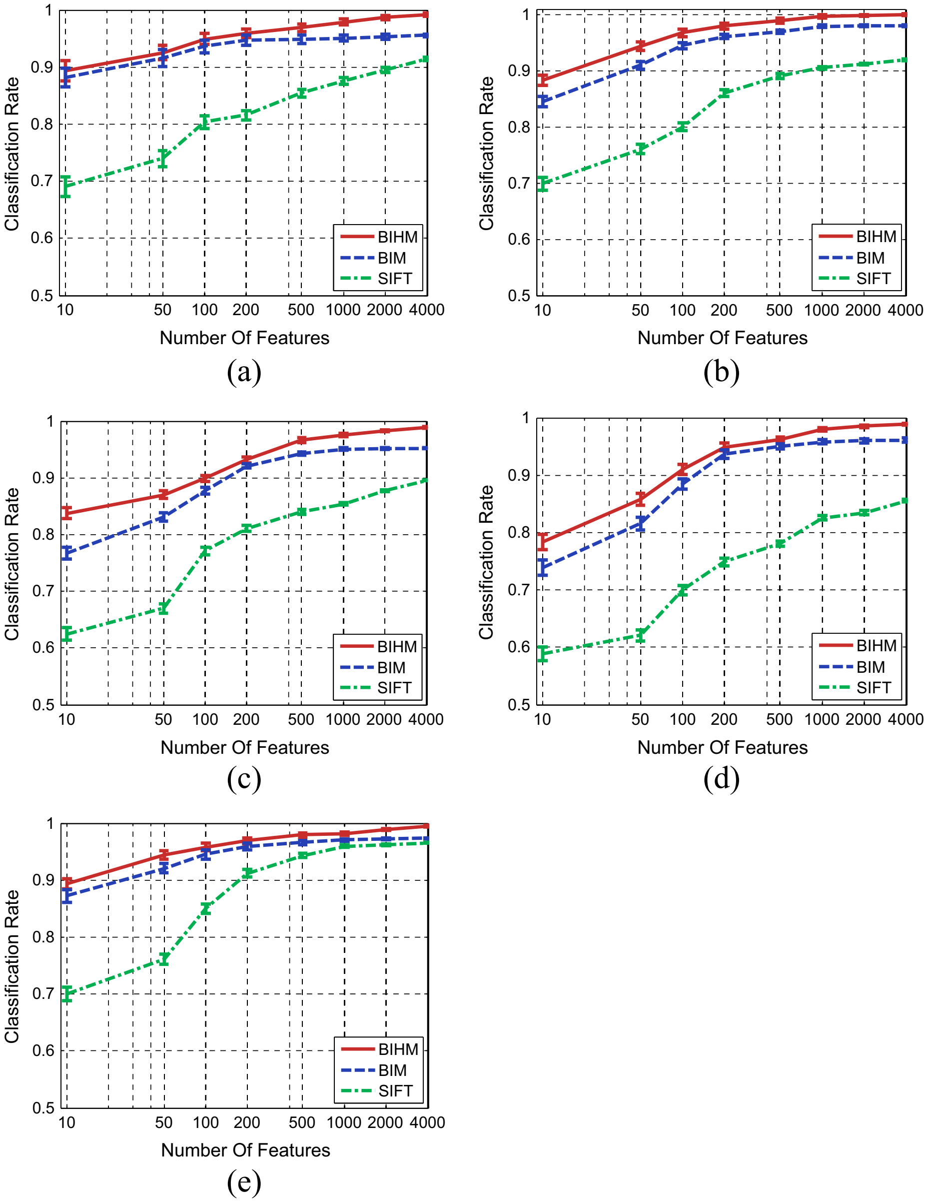

Fig. 4 Comparison of BIHM with standard BIM and SIFT on the CalTech5 dataset: **a** airplanes, **b** cars, **c** leaves, **d** faces, and **e** motorcycles

Table 1 Image Classification Performance on the Imagenet Dasebase

Method	Precision
SIFT-variant + BoW	0.560
MKL	0.658
SIFT + GMM	0.689
CNN	**0.787**
BIHM	0.745

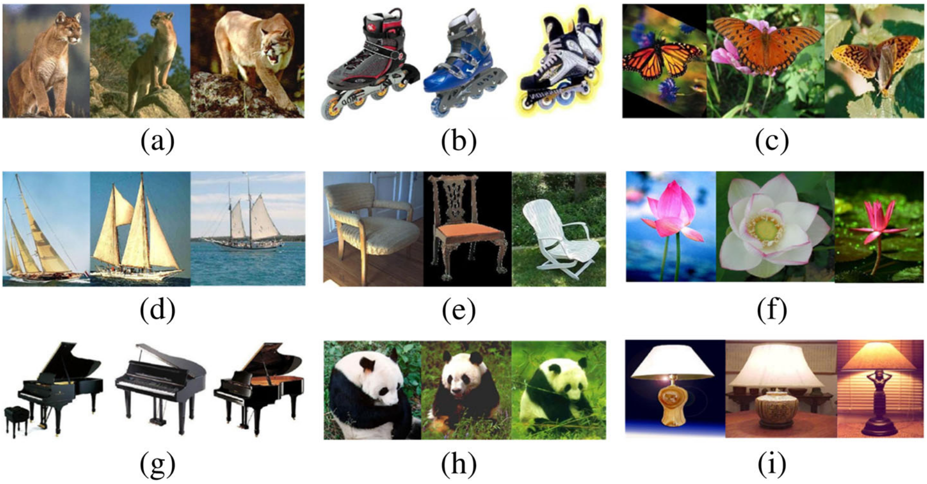

Fig. 5 Sampling images from the Caltech256 database: **a** cougar body, **b** inline skate, and **c** butterfly, **d** ketch, **e** chair, **f** lotus, **g** piano, **h** panda, **i** lamp

5 Discussion

The information processing mode is hierarchical in the primate visual cortex. The standard BIM is derived from the hierarchical theory of visual processing, i.e., the well-known Multi-Stage Hubel Wiesel Architecture [14, 15]. As described in the architecture, the hierarchical architecture of the visual cortex consists of three various levels cells, i.e., simple cells, complex cells, and hypercomplex cells. BIM mimics the ventral visual pathway from the stage of simple cells to that of complex cells. The simple cells are selective to the edge-like bars with particular orientations and locations in the receptive fields. The complex cells obtain the stimulations from various simple cells, so they can easily respond to the bars with different orientations and locations in the receptive fields. The higher level features represented by the hypercomplex cells could be regarded as the combination of that by the two types of lower level cells. The combination of these two kinds of cells extends the scale and position invariance properties. The feature representation is becoming increasingly abstraction from low level to high level. The higher abstraction level means better understandable and classifiable, which involves less guesswork and random component. Based on this consideration, we add more discriminative layers upon the conventional four-layer framework and enhance the understandable and classifiable of the biologically hierarchical inspired model. In contrast to the conventional BIM that mimics the inferior temporal cortex, which corresponds to the lower level feature, the proposed BIHM adds two more layers upon the conventional framework to simulate inferotemporal cortex, exploring higher level feature invariance and selectivity.

Table 2 Image retrieval performance on Caltech256

Method	10 classes	20 classes	50 classes
GIST	0.1850	0.1184	0.1184
BoW	0.2300	0.1400	0.0900
SPM	0.2700	0.1920	0.1355
BIM	0.2855	0.1732	0.1420
BIHM	**0.3250**	**0.2230**	**0.1875**

Recently, the Convolutional Neural Network (CNN), another form of hierarchical structure, has been obtained tremendous development, which has significantly improved the performance in visual recognition and detection, speech recognition, and many other domains. The CNN allows the hierarchical computational models to learn feature representations with various levels of abstraction. Since Hinton et al. firstly used 8-layer convolutional nets to claim top spot for object recognition in LSVRC2012 [19], the number of layers of convolutional nets is with continuous growth. In LSVRC2014, the GoogleNet got the champion with 22-layer network [42]. The ResNet from Microsoft is a convolutional network with 152 layers [12]. The SenseTime refreshed this record with an up to 1207-layer network and achieved state-of-the art performance (http://image-net.org/challenges/LSVRC/2016/results). It should be noted that we are not alleging the hierarchical structure is the deeper, the better, but based on the existing results, the deeper structure is with better representation using more data, which seems like one of ways to improve the performance for the hierarchical models. Hence we tend to improve the BIHM model with deeper network structure in the future based on the biological theories and some deep learning techniques.

In this article, the biologically hierarchical inspired model was proposed and successfully used in the recommendation task. We intend to extend our work into some other domains. For instance, it is a heavy workload for the medical workers to dispose a large number of clinical data every day, a powerful recommendation or retrieval tool is essential. This has been attracted keen interest, some effective retrieval approaches to the clinical data have been proposed [24, 49]. Therefore, it would be a meaningful work to extend the proposed BIHM in biomedical or clinical domain.

6 Conclusion

In the article, we describe a novel biologically inspired hierarchical model by adding S3 and C3 layers upon the conventional BIM framework to represent a high level invariance and selectivity of features, and successfully applied the proposed model in the image recommendation. The BIHM provides a balanced trade-off between the features' selectivity and invariance. The experiments on three various datasets showed our proposed model qualifies in image recommendation tasks. Our research by far has mainly concentrated on the performance upgrading of BIM. The improvement of calculating speed and practical applications will be our future work.

Acknowledgements The authors would like to thank the edtors and anonymous reviewers for their constructive comments. This work was supported by the National Science Foundation of China (Grant 61603389) and partially supported by National Natural Science Foundation of China (Grants 61502494, 61210009) and also by the Strategic Priority Research Program of the CAS (Grant XDB02080003) and the Development of Science and Technology of Guangdong Province Special Fund Project(Grant 2016B090910001).

References

1. Azzopardi G et al (2015) Trainable COSFIRE filters for vessel delineation with application to retinal images. Med Image Anal 19(1):46–57
2. Babenko A et al (2014) Neural codes for image retrieval. European Conference on Computer Vision. Springer International Publishing
3. Bhagyalakshmi A, Vijayachamundeeswan V (2014) A survey on content based image retrieval using various operators. 2014 International Conference on Computer Communication and Systems. IEEE

4. Chang C, Lin C (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2(3):27
5. Csurka G et al (2004) Visual categorization with bags of keypoints. Workshop on statistical learning in computer vision, ECCV. 1(1-22)
6. Deng J et al (2009) Imagenet: a large-scale hierarchical image database. IEEE Conf Comput Vis Pattern Recognit, 2009. CVPR 2009. IEEE
7. Dharani T, Laurence Aroquiaraj I (2013) A survey on content based image retrieval. 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME). IEEE
8. Donahue J et al (2015) Long-term recurrent convolutional networks for visual recognition and description. Proc IEEE Conf Comput Vis Pattern Recognit
9. Eitz M, Hays J, Alexa M (2012) How do humans sketch objects? ACM Trans Graph 31(4):44–41
10. Griffin G, Holub A, Perona P (2007) Caltech-256 object category dataset
11. Habibian A, van de Sande KEA, Snoek CGM (2013) Recommendations for video event recognition using concept vocabularies. Proceedings of the 3rd ACM conference on International conference on multimedia retrieval. ACM
12. He K et al (2016) Deep residual learning for image recognition. Proc IEEE Conf Comput Vis Pattern Recognit
13. Huang K et al (2011) Biologically inspired features for scene classification in video surveillance. IEEE Trans Syst Man Cybernet B Cybern 41(1):307–313
14. Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. J Physiol 148(3): 574–591
15. Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160(1):106–154
16. Juneja K et al (2015) A survey on recent image indexing and retrieval techniques for low-level feature extraction in CBIR systems. 2015 I.E. International Conference on Computational Intelligence & Communication Technology (CICT). IEEE
17. Kim M, Park SO (2013) Group affinity based social trust model for an intelligent movie recommender system. Multimed Tools Appl 64(2):505–516
18. Kobatake E, Wang G, Tanaka K (1998) Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. J Neurophysiol 80(1):324–330
19. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst
20. Lazebnik S, Schmid C, Ponce J (2005) A maximum entropy framework for part-based texture and object recognition. Int Conf Comp Vis 832–838
21. Lazebnik S, Schmid C, Ponce J (2006) Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. 2006 I.E. computer society conference on computer vision and pattern recognition. Vol. 2. IEEE
22. Li Y et al (2015) Free-hand sketch recognition by multi-kernel feature learning. Comput Vis Image Underst 137:1–11
23. Liu G-H, Yang J-Y (2013) Content-based image retrieval using color difference histogram. Pattern Recogn 46(1):188–198
24. Liu S, Wang Y, Hongfang et al (2016) MayoBMI at ImageCLEF 2016 handwritten document retrieval task. Proceedings of CLEF eHealth
25. Long F, Zhang H, Feng DD (2003) Fundamentals of content-based image retrieval. Multimedia Information Retrieval and Management. Springer Berlin Heidelberg, p 1-26
26. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
27. Lu Y, Zhang H et al (2014) Extended biologically inspired model for object recognition based on oriented Gaussian–Hermite moment. Neurocomputing 139:189–201
28. Lu Y, Zhang H et al (2015) Enhanced hierarchical model of object recognition based on a novel patch selection method in salient regions. IET Comput Vis 9(5):663–672
29. Lu Y, Zhang H et al (2016) Dominant orientation patch matching for HMAX. Neurocomputing 193:155–166
30. Lu Y et al (2017) A novel biologically inspired hierarchical model for image recommendation. International Symposium on Neural Networks. Springer International Publishing
31. Murala S, Maheshwari RP, Balasubramanian R (2012) Local tetra patterns: a new feature descriptor for content-based image retrieval. IEEE Trans Image Process 21(5):2874–2886
32. Mutch J, Lowe DG (2006) Multiclass object recognition with sparse, localized features. 2006 I.E. Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). Vol. 1. IEEE
33. Oliva A, Torralba A (2006) Building the gist of a scene: The role of global image features in recognition. Prog Brain Res 155:23–36
34. Plant W, Lumsden J, Nabney IT (2013) The Mosaic Test: measuring the effectiveness of colour-based image retrieval. Multimed Tools Appl 64(3):695–716

35. Poggio T, Girosi F (1990) Networks for approximation and learning. Proc IEEE 78(9):1481–1497
36. Qiao H et al (2014) Introducing memory and association mechanism into a biologically inspired visual model. IEEE Trans Cybernet 44(9):1485–1496
37. Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nat Neurosci 2(11):1019–1025
38. Schneider RG, Tuytelaars T (2014) Sketch classification and classification-driven analysis using fisher vectors. ACM Trans Graph 33(6):174
39. Serre T (2006) Learning a dictionary of shape-components in visual cortex: comparison with neurons, humans and machines
40. Serre T, Riesenhuber M (2004) Realistic modeling of simple and complex cell tuning in the HMAX model, and implications for invariant object recognition in cortex. Massachusetts Inst of Tech Cambridge Computer Science and Artificial Intelligence Lab
41. Serre T, Wolf L, Bileschi S et al (2007) Robust object recognition with cortex-like mechanisms. IEEE Trans Pattern Anal Mach Intell 29(3)
42. Szegedy C et al (2015) Going deeper with convolutions. Proc IEEE Conf Comput Vis Pattern Recognit
43. Tam K-P (2013) Concepts and measures related to connection to nature: Similarities and differences. J Environ Psychol 34:64–78
44. Viana W, Braga R, Lemos FD, de Souza JM, Carmo RA, Andrade RM, Martin H (2014) Mobile Photo Recommendation and Logbook Generation Using Context-Tagged Images. IEEE Multimedia 21(1):24–34
45. Wan J et al (2014) Deep learning for content-based image retrieval: a comprehensive study. Proceedings of the 22nd ACM international conference on Multimedia. ACM
46. Wang J (2015) Support recovery with orthogonal matching pursuit in the presence of noise. IEEE Trans Signal Process 63(21):5868–5877
47. Wang J, Shim B (2012) On the recovery limit of sparse signals using orthogonal matching pursuit. IEEE Trans Signal Process 60(9):4973–4976
48. Wang J, Kwon S, Shim B (2012) Generalized orthogonal matching pursuit. IEEE Trans Signal Process 60(12):6202–6216
49. Wang Y, Wu S et al (2016) A Part-Of-Speech term weighting scheme for biomedical information retrieval. J Biomed Inform 63:379–389
50. Wu L, Jin R, Jain AK (2013) Tag completion for image retrieval. IEEE Trans Pattern Anal Mach Intell 35(3):716–727
51. Zhang H, Lu Y et al (2016) B-HMAX: A fast binary biologically inspired model for object recognition. Neurocomputing 218:242–250
52. Zhu S, Wang B, Liu Y (2012) Spreading activation theory based image annotation. 2012 I.E. International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE

Yan-Feng Lu received his B.S. degree in Electrical Engineering from the Harbin Institute of Technology, China in 2010, and his Ph.D. degree from Korea University, Republic of Korea in 2015. He is currently an Associate Professor with the Institute of Automation, Chinese Academy of Sciences. His research interests include computer vision, robotics, biologically inspired intelligence, and machine learning. He is currently a member of the IEEE.

Hong Qiao received the B.Eng. degree in hydraulics and control and the M.Eng. degree in robotics and automation from Xi'an Jiaotong University, Xi'an, China, in 1986 and 1989, respectively, and the Ph.D. degree in robotics control from De Montfort University, Leicester, U.K., in 1995. She was an Assistant Professor with the City University of Hong Kong, Hong Kong, and a Lecturer with the University of Manchester, Manchester, U.K., from 1997 to 2004. She is currently a Professor with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. Her current research interests include robotics, machine learning, and pattern recognition. Dr. Qiao is currently a member of the Administrative Committee of the IEEE Robotics and Automation Society. She is the Editor in Chief of Assembly Automation, an Associate Editor of the IEEE transactions on cybernetics, the IEEE transactions on automation and sciences technology, and the IEEE/ASME transactions on mechatronics.

Yi Li received her B.S. degree in Electrical Engineering from the Harbin Institute of Technology, China in 2010, her M.S. degree from Ohio State University, United States in 2012, and her Ph.D. degree from Hanyang University, Republic of Korea in 2016. She is currently an Assistant Professor with the Nanchang University. Her research interests include electrical engineering, automation, and machine learning.

Li-Hao Jia received the B.Eng. degree and M.Eng. degree in mechanical engineering from Xi'an University of Science and Technology, Xi'an, China, in 2002 and 2005, respectively, and the Ph.D. degree in control science and engineering from Xi'an Jiaotong University, Xi'an, China, in 2011. He was a Postdoctoral fellow and an Assistant Professor with the Institute of Automation, Chinese Academy of Sciences, Beijing, China, from 2011 to 2015. He is currently an Associate Professor with the Research Center for Brain-inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His current research interests include robotics and machine learning. He is currently a member of the IEEE.