Comparing HMAX and BoVW Models for Large-Scale Image Classification

Jalila Filali, Hajer Baazaoui Zghal, Jean Martinet

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2019) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Comparing HMAX and BoVW Models for Large-Scale Image Classification

Jalila Filali[a], Hajer Baazaoui Zghal[a,b], Jean Martinet[c]

[a] *University of Manouba, ENSI, RIADI Laboratory,Tunisia*
[b] *ETIS UMR 8051, CY University, ENSEA, CNRS, F-95000, Cergy, France*
[c] *Université Côte d'Azur/I3S/CNRS, Polytech Nice Sophia Campus SophiaTech, Sophia-Antipolis, France*

Abstract

Image classification is one of the most important topics in computer vision. It became crucial for large image datasets. In the literature, several image classification approaches are proposed. In this context, Bag-of-Visual Words (BoVW) model has been widely used. The BoVW model relies on building visual vocabulary and images are represented as histograms of visual words. However, recently, attention has been shifted to the use of complex architectures which are characterized by multilevel processing. HMAX (Hierarchical Max-pooling model) model has attracted a great deal of attention in image classification, due to its architecture, which alternates layers of feature extraction with layers of pooling. This paper aims at comparing bags of visual words model to HMAX model for image classification using large datasets. To achieve this goal, we study the use of image features obtained by BoVW model with SIFT (Scale-Invariant Feature Transform) descriptors, and we compare them to HMAX features. Image classification is performed by using the support vector machine (SVM) classifiers. Experiments are achieved using accuracy as an evaluation metric and ImageNet and OpenImages datasets. Results have shown that the classification performance obtained by HMAX model outperforms the classification using large image databases.

Keywords: Image classification ; feature extraction; HMAX model; Bag of Visual Words model

1. Introduction

Recently, image classification has become a popular field of interest in image processing and computer vision. To understand what is in a dataset, we need to classify data. However, when large image datasets are used, classifying complex data became a difficult task. Image classification generally consists in labeling images by one of predefined image categories. To this end, various methods of image classification have been proposed. Machine learning methods are successfully applied. As the basis of image processing and computer vision, image representation is the key study content in this field, as its performance directly affects image classification results.
In this context, BoVW model proposed by [16] is widely used in image classification and objects recognition. The model is originally inspired by ideas in the domain of text analysis which has been used in text mining. Representing

images by vectors of visual words is based on analogies between text and image.
In the literature, several works on image classification field revolve around the bag of words methods which consist on building a visual vocabulary from image features [20], [21]. The image features are quantized as visual words to express the image content through the distribution of visual words. Thus, the image features are essentially obtained in one step using a flat architecture.
However, recently, attention has been shifted to the use complex architectures which are characterized by multilevel processing. In this context, the biologically inspired HMAX model was firstly proposed by [12]. The HMAX model has attracted a great deal of attention in image classification, due to its architecture which alternates layers of feature extraction with layers of maximum pooling. The HMAX model was recently optimized by the works of [13] and [14] in order to add multi scale representation as well as more complex visual features.
We attempt to address a comparative study issue of feature extraction models for image classification—the effect of feature extraction model on image classification performance . In doing this, we compare the bag of visual words model to HMAX model for image classification on large image datasets. To achieve our aims, we study the use of features obtained by BoVW model with SIFT (Scale-Invariant Feature Transform) descriptors, and we compare them to HMAX model where features is based on computing Gabor filters of multiple orientations and scales. Image classification process is performed by the use of the support vector machine (SVM) classifiers with linear function. In particular, we compare the image classification results obtained by both BoVW and HMAX models on large image datasets. Our goal is to determine which extraction feature model is well suited for image classification using large image databases.
The remaining parts of the paper are organized in the following way. Section 2 explains the related works and our motivations. Section 3 details the image classification method using both BoVW and HMAX models with SVM classifiers. In section 4, we present the experimental setup and then, we compare and discuss the image classification performance of the two models. The last section concludes and recommends possible areas for future works.

2. Related work and motivation

In computer vision and image processing field, when large image databases are used, image classification and categorization became difficult tasks. To facilitate these tasks on large databases, several image categorization and classification approaches have been introduced and applied. In this context, many feature extraction models have been introduced. BoVW model has been widely used in the area of image classification, which rely on building visual vocabulary. Several studies have focused on the use of the BoVW model based methods for classifying and recognizing images. For instance, in [21], an improved bag of words model was proposed for image classification. The goal of this work is to combine salient region extraction and spatial geometry structure in order to integrate the global and local information which not only can produce more discriminative visual words but also avoid the disturbance of complex background information and the changing location to a certain extent. The experiment demonstrates that the proposed method provide a higher classification accuracy.
In addition, in [15] a novel technique of image classification using BOVW model is proposed. The goal is to perform bi-linear classification of images, deciding between car and non-car images. The process involves feature detection and extraction, K-means clustering is then applied to make the bag of visual words and images are expressed as histograms of visual words. A supervised learning model is trained and is then tested for classification of images into respective classes. In [2] introduced an improved BoVW approach for automatic emotion recognition. The purpose is to study several aspects of the BoVW approach that are linked directly to gain classification performance, speed and scalability. Also, the goal is to improve the codebook generation process through employing k-means as a clustering method, wherein they gain speed and accuracy. A learning process is performed by using histogram intersection kernel by SVM to learn a discriminative classifier.
Several methods based on HMAX architecture have been used for improving image classification [4], [19] and [5]. In [18], a method of feature learning based on HMAX architecture for image classification has been proposed. The purpose of this work is to build complex features with richer information to improve image classification. The classification results show significant improvements over previous architecture using the same framework.
In [23] a fast binary-based HMAX model (B-HMAX) is proposed for object recognition . This model allow to detect corner-based interest points and to extract few features with better distinctiveness. The idea is to use binary

Table 1. Overview of image classification approaches based on BoVW and HMAX models

Approaches	Model	Task	Dataset	Images number
[2]	BoVW	classification	JAFFE DynEmo	143 120
[3]	BoVW	classification	specific data	120
[23]	HMAX	object recognition	GRAZ01	150
[15]	BoVW	classification	car images	1000
[4]	HMAX	classification	Caltech101	30
[11]	HMAX	classification	Caltech101	
[6]	HMAX	classification	Caltech101	—
[21]	BoVW	classification	Caltech 101 VOC 2007	300 300
[1]	BoVW	classification	CIFAR-10	
[17]	BoVW	classification	VOC 2007	

strings to describe the image patches extracted around detected corners, and then to use the Hamming distance for matching between two patches. The experimental results demonstrate that the B-HMAX model can significantly improve the accuracy performance. To resume the recent related works, we present in table 1, a review of image classification approaches based on BoVW and HMAX models. In the literature, there are few studies that have focused on comparing the bag of words method to HMAX model for image classification field. There are only two works: the first one is proposed by [10],it is based on a comparative study of local descriptors for object category recognition. The aim is to evaluate the performance of the two descriptors SIFT and HMAX, applied to the task of visual object detection; and the second work [22], it consist to study the comparison of feature matching for visual object categorization.

In the first work [10], SIFT descriptors are compared to C1 features of HMAX as local descriptors for object recognition. Experiments, which are performed using 200 images from 9 classes of Caltech database, showed that the C1 features of HMAX model are better than SIFT features for object detection.

In [22], the goal is compare two feature matching techniques which can be integrated in the HMAX framework for object categorization: MAX technique and histogram technique originating from Bag-of-Words method. The experiments showed that the histogram technique outperforms the MAX technique on a small dataset. For both works, we conclude that there are many weaknesses: firstly, we see that the comparatives are performed using a small dataset (only 200 images for the first work and only 5 classes for the second work) for two interested tasks: object detection and categorization. That tends to reduce the size of dictionary of features, and thus influences directly on performance that are evaluated. Consequently, this tends to decrease the effectiveness of detection and classification. Analysis and results obtained using a very small dictionary of features cannot confirm a concrete comparative. Secondly, in the first work, only features of the second layer of HMAX model (C1 features) are used to perform the comparison to SIFT descriptors. This study serve to compare only the SIFT features to C1 features. Thus, due to use only the C1 level of HMAX model and SIFT descriptors, BoVW features and HMAX model are not exploited in sustainable manner. This can directly affect object detection performance. Finally, we notice that, in [22], the goal is to use the proposed techniques for feature vector generation, but this letter remains limited to use a small size of features.

Our motivations are to study and to compare the bags of visual words method to HMAX model on image classification on large datasets. For this, we exploit the features obtained by BoVW model with SIFT descriptors, and all HMAX features which is based on computing Gabor filters of multiple orientations and scales.

Our purpose is to provide a concrete comparative of BoVW and HMAX models for image classification on large image datasets. To perform our comparison study, we focus on three main questions:

- What are the effects of orientation and scale variations on classification accuracy for HMAX model?

- How many features can be used to give a better classification for both BoVW and HMAX models?
- What model is more efficient for image classification using large image datasets?

3. BoVW and HMAX models for image classification

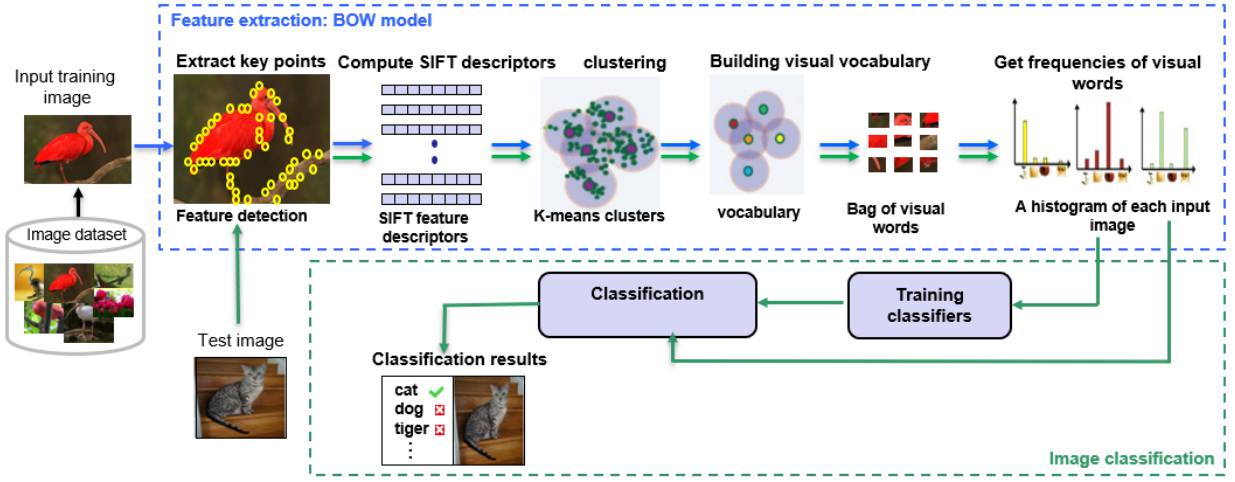

Fig. 1. Image classification method based on BoVW model

We propose two different image classification methods: 1) image classification method based on BoVW model, and 2) image classification method based on HMAX model. For both methods we adopt SVM classifiers to classify images. In the next subsections we detail the main steps of the two methods, and we introduce a comparison between BoVW and HMAX models.

3.1. Image classification based on BoVW model

In this subsection, we describe an image classification method based on bag of visual words and we detail their components. As depicted in figure 1, the image classification approach is composed of two components: (1) feature extraction and (2) image classification component.
Firstly, visual words are generated from the training set by adopting the BoVW model (cf. figure 1: (1) Feature extraction: BoVW model). Secondly, the images are represented by histograms that are built by getting frequencies of the obtained visual words. Finally, the obtained histograms are used to train classifiers. We have selected a multi-class linear SVM in order to classify images (cf. figure 1:(2) Image classification). To generate visual vocabulary using BoVW model, three steps are applied: 1) interest points detection, 2) computing descriptors and 3) clustering descriptors step. The clustering step allows to build visual words or visual vocabulary:

1. **Interest points detection:** Several detectors of interest points are developed in the literature. To extract a large number of interest points from images, we used SIFT detector which is proposed by David Lowe [7].
2. **Computing descriptors**: This step consists in extracting features by computing descriptors for interest points which are detected in the previous step. To perform this step, we compute SIFT (Scale-invariant feature transform) descriptors [8]. SIFT converts each patch to 128-dimensional vector and each image represented by a collection of vectors of the same dimension (128).
3. **Clustering:** This step allows to cluster local descriptors which are computed in the previous step, the goal is to represent each feature by the centroid of the cluster it belongs. To perform this goal, we used the *K-Means* algorithm [9] which is the most widely used clustering algorithm for visual vocabulary generation. Visual vocab-

ulary (or *codebook*) is then defined as the centers of the learned clusters. The number of the clusters is the visual vocabulary size and each cluster represents a visual words.

To summarise, visual vocabulary generation process starts by interest points detection of training images using SIFT detector. Then, SIFT descriptors are computed for each obtained key-points. K-Means clustering is then applied in order to generate visual words. Finally, each image is represented by a histogram which reflects the frequency of visual words occurrence. The obtained histograms are used to train classifiers.

3.2. Image classification based on HMAX model

In this subsection, we describe an image classification method based on HMAX model and we detail their components. As depicted in figure 3, the image classification method is composed of two components: (1) feature extraction and (2) image classification component.
To extract visual features from training images, we use the HMAX model. In particular, we adopt the HMAX model to provide complex and invariant visual information and to improve the discrimination of features. The HMAX model follows a general 4-layer architecture. Simple ("S") layers apply local filters that compute higher-order features and complex ("C") layers increase invariance by pooling units. A general architecture of the HMAX model is presented in figure 2 [18].
We describe below the operations of each layer of the HMAX model:

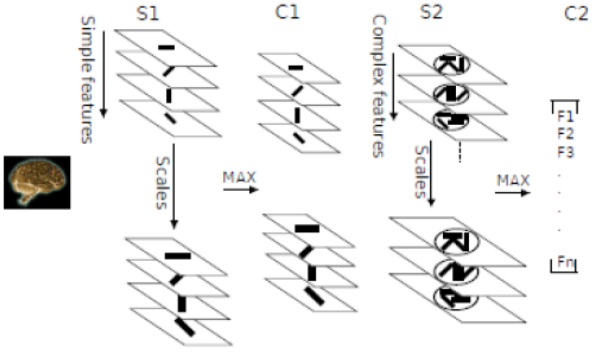

Fig. 2. General architecture of the HMAX model [18]

- **Layer 1 (S1 Layer):** In this layer, each feature map is obtained by convolution of the input image with a set of Gabor filters $\mathbf{g_{s,o}}$ with orientations **o** and scales **s**. In particular S1 Layer, at orientation **o** and scale **s**, is obtained by the absolute value of the convolution product given an image I:

$$L1_{s,o} = |g_{s,o} * I| \tag{1}$$

- **Layer 2 (C1 Layer):** The C1 layer consists in selecting the local maximum value of each S1 orientation over two adjacent scales. In particular, this layer partitions each $\mathbf{L1_{s,o}}$ features into small neighborhoods $\mathbf{U_{i,j}}$, and then selects the maximum value inside each $\mathbf{U_{i,j}}$.

$$L2_{s,o} = \max_{U_{i,j} \in L1_{s,o}} * U_{i,j} \tag{2}$$

- **Layer 3 (S2 Layer):** S2 layer is obtained by convolving filters α^m, which combine low-level Gabor filters of multiple orientations at a given scale.

$$L3_{s,m} = \alpha_m * L2_s \tag{3}$$

- **Layer 4 (C2 Layer):** In this layer, L4 features are computed by selecting the maximum output of $L3_s^m$ across all positions and scales.

$$L4 = \max_{(x,y),s} L3_S^1(x,y), ..., \max_{(x,y),s} L3_S^M \tag{4}$$

The feature extraction process is summarized as follows: given the input image, the S1 layer corresponds to compute Gabor filters of different orientations and scales. Then, the C1 features describe the complex cells by taking the max S1 over scales and positions. The third layer S2 consists to combine the C1 features into more complex features for multiple orientations at a given scale. Finally, C2 features are obtained by selecting the maximum output of each S2 feature across all positions and scales. The C2 features are used as the input of the linear SVM model to train classifiers. Table 2 introduce a comparison study of BoVW and HMAX models based on four criteria: architecture type, selection of image points, feature dictionary and how to store obtained visual features.

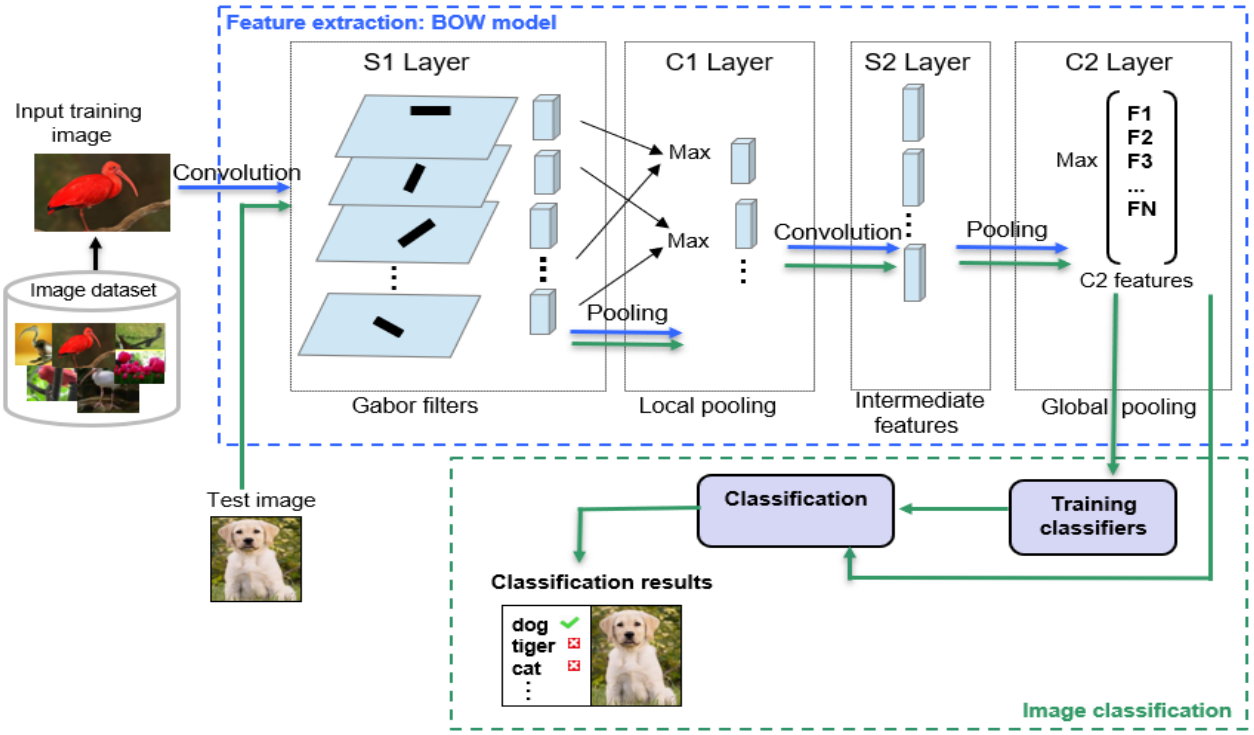

Fig. 3. Image classification method based on HMAX model

3.3. Training classifiers and image classification

In this subsection, we detail image classification component for both classification methods that are introduced in the previous sections.
To classify images, for both methods, we used SVM classifiers. The aim is to learn a discriminative model for each

Table 2. Comparison of BoVW and HMAX models.

Comparison criteria	BoVW Model	HMAX Model
Architecture	flat architecture	4 layers
Image points selection	only points detected by SIFT detector are selected	all image points are selected
Dictionary of features	obtained by clustering SIFT descriptors: clusters or visual words	Obtained by the C2 layer: C2 features
Storing features	histograms of occurrences of the dictionary features	Storing only the maximum output (best response) of each S2 filters in C2 layer

class to predict the visual features membership. To achieve this goal, we focus on linear SVM classifiers since the diversity of image categories makes that using non linear models is impractical. In particular, given the visual features of the training images, we train a One-vs-All SVM classifier for each class to discriminate between this class and the other classes. During the classification phase, when a test image is introduced, visual feature are extracted according to the model that is used (BoVW or HMAX model). After that, classification is done using classifiers that are trained.

4. Experimentation and results analysis

In this section, we illustrate the experimental results of our work. We start with the experimental setup, then, we present the evaluation of the two classification methods by introducing image classification performance.

4.1. Experimental setup

The aim is to evaluate the classification performance using both BoVW and HMAX model. To achieve this goal, we used ImageNet [1] and OpenImages datasets [2]:

- ImageNet: images are organized according to the WordNet hierarchy. This database contains about 1,281,167 images from 1000 synsets. The number of images for each synset (category) ranges from 732 to 1300 and all images are in JPEG format. Images are heterogeneous and represent diverse themes. In our work, we used about 200K images from 1000 categories or synset, there are 190K as a training set and 10K images as a testing set.

- OpenImages: where images have been annotated with labels spanning over 600 categories, there are 1,743,042 images as a training set and 125,436 as a testing set. In our work, we used about 200 images for each categories.

We used accuracy as metric to evaluate the image classification results.

4.2. Evaluation results

In this section, we compared the classification performance of two proposed classification methods: 1) the classification method based on BoVW model using SIFT descriptors, and 2) the classification method based on HMAX model. Both methods have many parameters that influence classification performance: dictionary size, number of orientations and scales for HMAX model. Thus, we focus the experimental evaluation on the following three questions:

[1] http://www.image-net.org/
[2] https://storage.googleapis.com/openimages/web/download.html

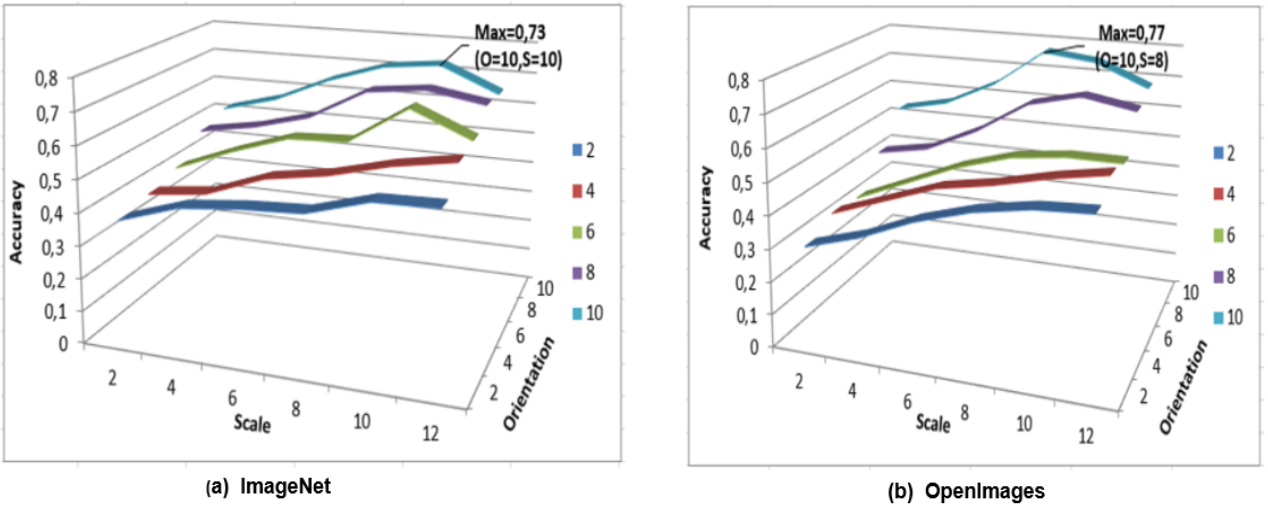

Fig. 4. Comparison of performance classification depending on orientation and scale for HMAX model using ImageNet and openImages.

1. What are the effects of orientation and scale variations on classification accuracy for HMAX model?
2. How many features can be used to perform an efficient classification for both BoVW and HMAX models?
3. What model is more efficient for image classification on large image databases?

To answer the three questions, firstly we focus on evaluating the HMAX model. For this purpose, we study the variation of orientations and scales number that are used to convolve images with Gabor filters. Secondly, we report performance of image classification depending on the number of features for each classification method. Finally we compare image classification accuracy obtained by both BoVW and HMAX models on ImageNet and OpenImages databases.

To evaluate image accuracy of classification method based on HMAX model, depending on the variation of the number of orientations and scales, we tested HMAX model using six scales 2,4,6,8,10, 12 and five orientations 2,4,6,8,10. The obtained classification results using ImageNet and openImages datasets are presented in figure 4.

As illustrated in figure 4, we notice that the best classification performance (**0.73**) is obtained with *10 orientations* and *10 scales* for ImageNet. According to this figure, it can be seen that the best accuracy achieved **0,73** with *10 orientations* and *10 scales* for ImageNet and **0,77** with *10 orientations* and *8 scales* for OpenImages. We notice that the accuracy value increases with the increase of the scale until reaching 10 scales for ImageNet and 8 scales for OpenImages. The increase of the classification accuracy could be explained by the impact of the amount of data that are extracted. However, when the number of scales tends to 12, the accuracy value decreases to 0,65 for both ImageNet and OpenImages datasets. The degradation in the classification accuracy can be explained by the lack of additional data to be extracted. There is no more data to be exploited for computing response Gabor filters. Thus, when the scale number tends to 12, the redundancy of the same information that are extracted with 10 scales for ImageNet and 8 scales for OpenImages can decreases the accuracy value.

To study, the second proposed question, we focus on the influence of the number of features on classification performance for both models using ImageNet and OpenImages. For this purpose, different sizes 500, 1000, 1500,2500, 3000, 3500, 4000 are applied to experiment and a comparison of classification performance is shown in figure 5.

As illustrated in figure 5, we observe that the HMAX model gives better performance than the BoVW model for both ImageNet and OpenImages.

Table 3 details the comparison of classification results for BoVW and HMAX models using ImageNet database. We notice that the best gain obtained by the HMAX model is **13.55%**. Likewise, for OpenImages, the HMAX model gives better results with a better improvement which reaches **19.14 %** (cf. Table 4).

Table 3 shows that the best classification, for the HMAX model, is obtained with a dictionary of 3500 features using

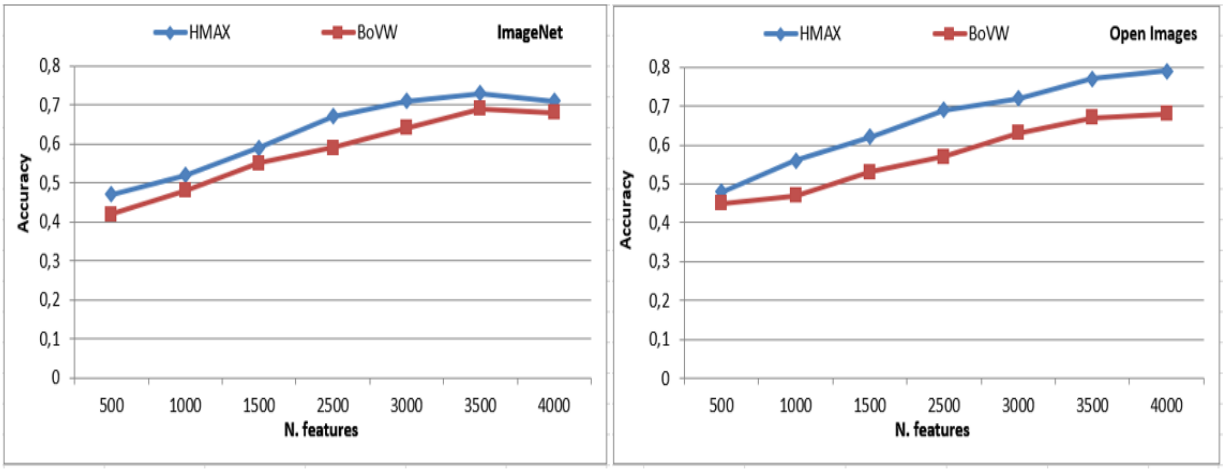

Fig. 5. Accuracy results comparison for HMAX and BoVW model using ImageNet and OpenImages

Table 3. Accuracy results for HMAX and BoVW model using ImageNet

Models \ $N.features$	500	1000	1500	2500	3000	3500	4000	Low-Improvement	Best-Improvement
BoVW	0,42	0,48	0,55	0,59	0,64	**0,69**	0,68	–	–
HMAX	0,47	0,52	0,59	0,67	0,71	**0,73**	0,71	+4,41%	+13,55%

Table 4. Accuracy results for HMAX and BoVW model using OpenImages

Models \ $N.features$	500	1000	1500	2500	3000	3500	4000	Low-Improvement	Best-Improvement
BoVW	0,45	0,47	0,53	0,57	0,63	0,67	**0,68**	–	–
HMAX	0,48	0,56	0,62	0,69	0,72	0,77	**0,79**	+6,66%	+19,14%

ImageNet (**0.73**). However, for OpenImages, using a dictionary of 4000 features, HMAX model achieves the best performance (**0.79**) (cf. Table 4).
The differences in classification performance between HMAX and BoVW models using ImageNet and OpenImages datasets, can be explained by considering that the HMAX model build complexes visual features with richer information using multiple orientations and scales of image structures. However, BOW model select only the interest points that are detected by SIFT detectors, and represent images by only the distribution of features, as histogram which reflects the cluster's frequency of occurrence.
According to our experiments and results, we conclude that the classification method based on HMAX model provides a better performance than the classification method using BoVW model on large image databases.

5. Conclusion

In this paper, we carry out a comparative study of two classification methods: 1) classification based on BoVW model with SIFT descriptors; and 2) classification based on HMAX model. We aim to perform a fair and concrete comparison, using the same SVM classifiers, the same training and test sets and the same size of dictionary of features. We perform our experiments on ImageNet and OpenImages that considered as large databases.
Firstly, we evaluate the influence of variation of orientations and scales number on classification accuracy for HMAX model, our results showed that that HMAX model work better with 10 orientations and 8 scales. Secondly, we report the classification performance of the BoVW and HMAX models depending on the numbers of features. Experiments is shown that the performance increased as a function of the number of features for both models, and the HMAX

model outperforms the BoVW model for all dictionary sizes. Finally, we conclude that the image classification based on HMAX model using large databases, is more efficient than the method which is based on BoVW model.
For future work, the idea would be to study both models for image annotation task to help users to understand and explore images from large image databases .

References

[1] Abdollahpour, Z., Samani, Z.R., Moghaddam, M.E., 2015. Image classification using ontology based improved visual words, in: 23rd Iranian Conference on Electrical Engineering, Tehran (2015), IEEE. pp. 694–698.

[2] Al Chanti, D., Caplier, A., 2018. Improving bag-of-visual-words towards effective facial expressive image classification, in: VISIGRAPP, the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications.

[3] Gao, H., Dou, L., Chen, W., Sun, J., 2013. Image classification with bag-of-words model based on improved sift algorithm, in: Control Conference (ASCC), 2013 9th Asian, IEEE. pp. 1–6.

[4] Hu, X., Zhang, J., Li, J., Zhang, B., 2014. Sparsity-regularized hmax for visual recognition. PloS one 9, e81813.

[5] Lau, K.H., Tay, Y.H., Lo, F.L., 2015. A HMAX with LLC for visual recognition. CoRR abs/1502.02772.

[6] Li, Y., Wu, W., Zhang, B., Li, F., 2015. Enhanced hmax model with feedforward feature learning for multiclass categorization. Frontiers in computational neuroscience 9, 123.

[7] Lowe, D.G., 1999. Object recognition from local scale-invariant features, in: Proceedings of the International Conference on Computer Vision, Kerkyra, Corfu, Greece, September 20-25, 1999, IEEE Computer Society. pp. 1150–1157.

[8] Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. International journal of computer vision 60, 91–110.

[9] MacQueen, J., et al., 1967. Some methods for classification and analysis of multivariate observations, in: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Oakland, CA, USA. pp. 281–297.

[10] Moreno, P., Marín-Jiménez, M.J., Bernardino, A., Santos-Victor, J., de la Blanca, N.P., 2007. A comparative study of local descriptors for object category recognition: Sift vs hmax, in: Iberian Conference on Pattern Recognition and Image Analysis, Springer. pp. 515–522.

[11] Mutch, J., Lowe, D.G., 2006. Multiclass object recognition with sparse, localized features, in: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, IEEE. pp. 11–18.

[12] Riesenhuber, M., Poggio, T., 1999. Hierarchical models of object recognition in cortex. Nature neuroscience 2, 1019.

[13] Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., Poggio, T., 2007. Robust object recognition with cortex-like mechanisms. IEEE Transactions on Pattern Analysis & Machine Intelligence , 411–426.

[14] Serre, T., Wolf, L., Poggio, T., 2005. Object recognition with features inspired by visual cortex, in: Conference on Computer Vision and Pattern Recognition (CVPR'05) on, IEEE.

[15] Singhal, N., Singhal, N., Kalaichelvi, V., 2017. Image classification using bag of visual words model with fast and freak, in: Electrical, Computer and Communication Technologies (ICECCT), 2017 Second International Conference on, IEEE. pp. 1–5.

[16] Sivic, J., Zisserman, A., 2003. Video google: A text retrieval approach to object matching in videos, in: 9th IEEE International Conference on Computer Vision (ICCV 2003), 14-17 October 2003, Nice, France, IEEE Computer Society. pp. 1470–1477.

[17] Su, Y., Jurie, F., 2012. Improving image classification using semantic attributes. International journal of computer vision 100, 59–77.

[18] Theriault, C., Thome, N., Cord, M., 2011. Hmax-s: deep scale representation for biologically inspired image categorization, in: Image Processing (ICIP), 2011 18th IEEE International Conference on, IEEE. pp. 1261–1264.

[19] Theriault, C., Thome, N., Cord, M., 2013. Extended coding and pooling in the hmax model. IEEE Transactions on Image Processing 22, 764–777.

[20] Wang, C., Huang, K., 2015. How to use bag-of-words model better for image classification. Image and Vision Computing 38, 65–74.

[21] Wang, R., Ding, K., Yang, J., Xue, L., 2016. A novel method for image classification based on bag of visual words. Journal of Visual Communication and Image Representation 40, 24–33.

[22] Wijnhoven, R., de With, P.H., 2009. Comparing feature matching for visual object categorization: Max vs. bag-of-words, in: Proceedings of the 30th Symposium on Information Theory in the Benelux, May 28-29, 2009, Eindhoven, The Netherlands, Technische Universiteit Eindhoven. pp. 169–176.

[23] Zhang, H.Z., Lu, Y.F., Kang, T.K., Lim, M.T., 2016. B-hmax: A fast binary biologically inspired model for object recognition. Neurocomputing 218, 242–250.