Ontology and HMAX Features-based Image Classification using Merged Classifiers

Jalila Filali, Hajer Baazaoui Zghal, Jean Martinet

▶ **To cite this version:**

Jalila Filali, Hajer Baazaoui Zghal, Jean Martinet. Ontology and HMAX Features-based Image Classification using Merged Classifiers. International Conference on Computer Vision Theory and Applications 2019 (VISAPP'19), Feb 2019, Prague, Czech Republic. hal-02057494

HAL Id: hal-02057494
https://hal.science/hal-02057494

Submitted on 5 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Ontology and HMAX Features-based Image Classification using Merged Classifiers

Jalila Filali[1], Hajer Baazaoui Zghal[1] and Jean Martinet[2]
[1]*ENSI, RIADI Laboratory, University of Manouba, Tunisia*
[2]*Univ. Lille, CNRS, Centrale Lille, UMR 9189 – CRIStAL – Centre de Recherche en Informatique, Signal et Automatique de Lille, F-59000, Lille, France*
jalila.filali@ensi-uma.tn, hajer_bz@yahoo.fr, jean.martinet@univ-lille.fr

Keywords: Image Classification, HMAX Features, Ontology.

Abstract: Bag-of-Viusal-Words (BoVW) model has been widely used in the area of image classification, which rely on building visual vocabulary. Recently, attention has been shifted to the use of advanced architectures which are characterized by multilevel processing. HMAX model (Hierarchical Max-pooling model) has attracted a great deal of attention in image classification. Recent works, in image classification, consider the integration of ontologies and semantic structures is useful to improve image classification. In this paper, we propose an approach of image classification based on ontology and HMAX features using merged classifiers. Our contribution resides in exploiting ontological relationships between image categories in line with training visual-feature classifiers, and by merging the outputs of hypernym-hyponym classifiers to lead to a better discrimination between classes. Our purpose is to improve image classification by using ontologies. Several strategies have been experimented and the obtained results have shown that our proposal improves image classification. Results based our ontology outperform results obtained by baseline methods without ontology. Moreover, the deep learning network Inception-v3 is experimented and compared with our method, classification results obtained by our method outperform Inception-v3 for some image classes.

1 INTRODUCTION

Image classification consists in labeling images with one of a number of predefined categories. To achieve this goal, machine learning techniques are used. As the basis of image processing and computer vision, image representation is the key study content in this field because its performance directly affects the image classification results. In this context, the Bag-of-Visual-Words model (BoVW) proposed by (Sivic and Zisserman, 2003) is widely used in image classification and object recognition. In the literature, several works dealing with image classification revolve around BoVW method, which consist on building a visual vocabulary from image features (Wang and Huang, 2015), (Gao et al., 2013). The image features are quantified as visual words to express the image content through the distribution of visual words.

Recently, special attention has been shifted to the use of complex architectures which are characterized by multi-layers. Indeed, the biologically-inspired HMAX model was firstly proposed by (Riesenhuber and Poggio, 1999). 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 optimized in the work of (Serre et al., 2007) in order to add multi-scale representation as well as more complex visual features.

To improve image classification, several classification approaches based on ontologies have been proposed. The use of ontologies is generally motivated by the need to use semantic relations and describe data at a more semantic level for better classification. However, the accuracy of classification results remains far from authors' expectations due to several problems which still persist, such as the problem of ambiguity between classes. The ambiguity between classes is made explicit when similar visual features belonging to different image classes. An example of the ambiguity problem is presented in Figure 1.

In this paper, our objective is to propose an approach of image classification based on ontology and HMAX features using merged classifiers driven by taxonomic relationships in order to improve image classification. Our contribution consists in exploiting

Figure 1: Images correctly classified with positive (top) images that belong to *tiger* class and negative (bottom) images that, belong to *cat* class. The *cat* images appear in the *tiger* class due to their similar visual features to *tiger* images.

ontological relationships between classes in line with training visual-feature classifiers, and in merging the outputs of hypernym-hyponym classifiers to lead to a better discrimination between classes. The ontology is built to represent the semantic information associated with training images. The aim is to improve image classification using taxonomic relationships between image categories.
The remaining parts of the paper are organized in the following way. Section 2 explains the related works and our motivations. Section 3 details our proposed image classification approach. In section 4, we present the experimental setup and then, we present and discuss the image classification performance. The last section concludes and recommends possible areas for future works.

2 RELATED WORK AND MOTIVATIONS

Image classification has acquired the attention of researchers in computer vision and image processing. In image classification, several methods and approaches have been introduced and applied. In this section, we give a general overview of the main related works.

BoVW model was originally applied to classify images in the field of image processing and computer vision. Several studies have focused on the use of the BOVW-based methods for classifying and recognizing images. For instance, in (Singhal et al., 2017) a novel technique of image classification using BOVW model is proposed. The aim is to perform bi-linear classification of images, deciding between car and non-car images. The process involves feature detection of images using FAST features (Rosten and Drummond, 2006) and a supervised learning model is trained and then tested for image classification. The experiment shows that the proposed method is an efficient method of performing bi-linear classification.

Recently, several methods based on HMAX architecture have been used for improving image classification (Theriault et al., 2013), (Hu et al., 2014). In (Theriault et al., 2011), a method of feature learning based on HMAX architecture for image classification has been proposed. The purpose is to build complex features with richer information to improve image classification. Moreover, in (Zhang et al., 2016) a fast binary-based HMAX model (B-HMAX) is proposed for object recognition. The goal is to detect corner-based interest points and to extract few features with better distinctiveness. The idea is to use binary strings to describe the image patches extracted around detected corners, and then to use the Hamming distance for matching between two patches.

To improve image classification, several classification approaches based on ontologies have been proposed. For instance, (Su and Jurie, 2012) proposed to address two limitations of BOVW model: the lack of explicit meanings of visual words and the visual words are usually polysemous. Two novel methods have been proposed to improve the performance of the bag-of-words model for image classification. Both approaches consist in predicting a set of semantic attributes. One is combining bag-of-words histograms with semantic image descriptors. The other is embedding semantic information into the visual vocabulary. In addition, (Ristin et al., 2015) have addressed the problem of learning subcategory classifiers when only a fraction of the training data is labeled with fine labels while the rest only has labels of coarser categories. The aim is to adopt the framework of Random Forests (Breiman, 2001) and to propose a regularized objective function that takes into account relations between categories and subcategories. Moreover, in (Abdollahpour et al., 2015), a new visual vocabulary generation and feature representation method based on semantic taxonomies is proposed for image classification. The aim is, firstly, to leverage the semantic taxonomy to define visual words which are aware of contents and categories; secondly, to design a hierarchical classifier based on semantic taxonomies.

Other recent work tackle how coarse and fine labels can be used to improve image classification. In this context, (Dutt et al., 2017) address the problem of classification of coarse and fine grained categories by exploiting semantic relationships. In this work, the idea is to adjust the probabilities of classification according to the semantics of categories. An algorithm for doing such an adjustment is proposed to show the improvement for both coarse and fine grained classification. In (Lei et al., 2017), a weakly supervised

Table 1: Overview of the related image classification approaches.

Approaches	Model	Ontology	Dataset
(Al Chanti and Caplier, 2018)	BoVW	–	JAFFE DynEmo
(Durand et al., 2017)	FCN ResNet-101	–	VOC 2012
(Dutt et al., 2017)	CNN	✓	CIFAR-100
(Singhal et al., 2017)	BoVW	–	car images
(Lei et al., 2017)	CNN	✓	CIFAR-100
(Zhang et al., 2016)	HMAX	–	GRAZ01
(Szegedy et al., 2016)	Inception-v3	–	ILSVRC 2012
(Wang et al., 2016a)	CNN-RNN	–	NUS-WIDE MS-COCO
(Wang et al., 2016b)	BoVW	–	Caltech 101 VOC 2007
(Li et al., 2015)	HMAX	–	Caltech 101
(Abdollahpour et al., 2015)	BoVW	✓	CIFAR-10
(Hu et al., 2014)	HMAX	–	Caltech 101
(Gao et al., 2013)	BoVW	–	specific data
(Su and Jurie, 2012)	BoVW	✓	VOC 2007
(Theriault et al., 2011)	HMAX	–	Caltech 101

image classification method with coarse and fine labels has been proposed. The goal is to use weakly labeled data aiming at learning a classifier to predict the fine labels during testing. For this, authors have proposed a CNN-based approach to address this problem, where the commonalities between fine classes in the same coarse class are captured by min-pooling in the CNN architecture.

Recently, deep learning and Convolutional Neural Networks (CNNs) have been successfully applied in many vision tasks including image classification and object recognition. The quality of network architectures significantly improved by utilizing deeper and wider networks such as Inception-v3 (Szegedy et al., 2016) and AlexNet (Krizhevsky et al., 2012). (Wu et al., 2015) incorporated deep learning into a weakly supervised learning framework and demonstrated that their deep multiple instance learning system achieves convincing performance in both image classification and image annotation. Moreover, (Wang et al., 2016a) proposed a unified CNN-RNN framework for multi-label image classification, which effectively learns both the semantic redundancy and the co-occurrence dependency in an end-to-end way. The multi-label RNN model learns a joint low-dimensional image-label embedding to model the semantic relevance between images and labels. In addition, (Durand et al., 2017) introduced a deep learning method which jointly aims at aligning image regions for gaining spatial invariance and learning strongly localized features. The idea is to extend Convolution Neural Network at some levels for image classification.

To summarize the recent related work, we present in Table 1 a review of the related image classification approaches. It can clearly be seen that BoVW model has been applied by several work in both cases: without and with ontologies. But, in the literature, HMAX model is used only without ontologies for image classification. In addition, the use of ontologies is generally motivated by the need to improve image classification, however, the accuracy of classification results remain far from authors' expectations. Thus, despite this large number of the classification approaches in the literature, several problems persist, mainly the problem of ambiguity between classes that can degrade classification accuracy (cf.Figure 1).
In previous work, an image annotation approach based on visual features and ontologies has been proposed (Filali et al., 2017). To achieve better image annotation precision, an improvement of image classification is needed.

To overcome the related work limitations, we propose an ontology and HMAX features-based image classification method using merged classifiers to improve image classification. Our motivation is to exploit ontological relationships between image categories in line with training visual-feature classifiers, and to merge the outputs of hypernym-hyponym classi-

fiers in order to lead to a better discrimination between classes. The originality of our proposal concerns the integration of ontology in order to improve image classification and to decrease ambiguity between image classes.

In our method, firstly, we adopt HMAX model to extract visual features, and we build an ontology that can represent relationships between concepts (classes or categories) associated with training images. Secondly, visual-feature classifiers are trained according to taxonomic relationships between classes, in particular, for each super-category, images of subclasses are bagged in super-categories to train hypernym classifiers, and then for each sub-category, hyponym classifiers are trained using only images of the sub-classes in order to discriminate between the node's sub-categories. Finally, test images are classified using both hypernym and hyponym classifiers, then when the taxonomic relationship between the best hypernym class and the top hyponym classes are detected, output classifiers are merged to assign best classes to the test images. This enables to improve classification accuracy, and to decrease the ambiguity between classes.

3 ONTOLOGY AND HMAX FEATURES-BASED IMAGE CLASSIFICATION USING MERGED CLASSIFIERS

In this section, we describe our proposed image classification approach and detail their components. As depicted in Figure 2, our image classification approach is composed of three components: (1) feature extraction, (2) ontology building, and (3) image classification component. The different components are detailed below.

Firstly, visual features are extracted from the training set (cf.Figure 2: (1) feature extraction). In our work, we adopted the HMAX model because HMAX features are generic, dos not require hand-tuning, and can represent well complex features with richer information (a detailed description is given below). Secondly, image categories from the train set are used to build the ontology which consists in establishing relations between classes using taxonomic relationships found in WordNet (cf.Figure 2: (2) ontology building). Finally, the obtained HMAX features are used to train classifiers. We have selected a multi-class linear SVM in order to classify images (cf.Figure 2: (3) image classification). The classification method is detailed in Section 3.3.

3.1 Feature Extraction

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. We describe below the operations of each layer. 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 3 (Theriault et al., 2011).

- **Layer 1 (S1 Layer):** In this layer, each feature map is obtained from a convolution of the test 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:

$$L1_{s,o} = |g_{s,o} * I| \quad (1)$$

- **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 $\mathbf{U_{i,j}}$.

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

- **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.

$$L3_s^m = \alpha^m * L2_s \quad (3)$$

- **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.

$$L4 = \max_{(x,y),s} L3_s^1(x,y), ..., \max_{(x,y),s} L3_s^M \quad (4)$$

The obtained C2 features (HMAX features) are used as the input of the multi-class linear SVM model to train classifiers.

3.2 Ontology Building

As depicted in Figure 2 : (2) ontology building, the ontology building component consists of two main steps, namely, concept extraction and relation generation. Let us consider that categories of the training images consisting of the synset (synonym set or concept that is described by one or multiple words) IDs which are defined by an external lexical resource; we

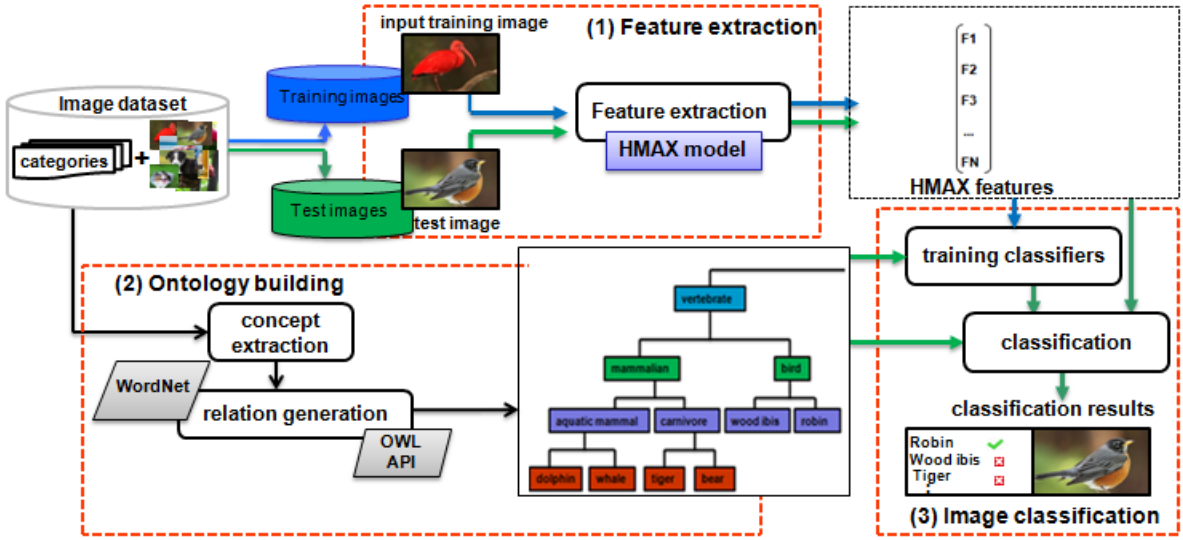

Figure 2: Ontology and HMAX features-Based Image Classification Using Merged Classifiers.

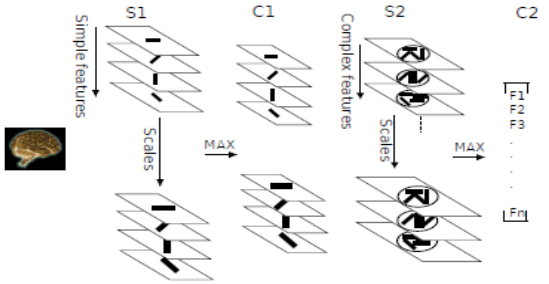

Figure 3: General architecture of the HMAX model (Theriault et al., 2011).

used WordNet [1]. To achieve this aim, firstly, we need to extract concepts (or classes) from the synsets. In particular, we extract the first concept that appears in this synset and relationships between concepts. We are interested in hypernymy and hyponymy relationships. Secondly, relationships between concepts are generated. The resulting sets of taxonomic relationships as well as the resulting concepts are the basis of our ontology. Finally, once the taxonomic relationships between concepts are extracted and created, the ontology is built. To successfully construct the ontology, we used the OWL API [2]. Some rules are applied in order to transform the extracted relationships in OWL language. Algorithm 1 is used for building our ontology.

3.3 Image Classification

In this section, we detail our image classification method. As depicted in Figure 2: (3) image classification,

[1]https://wordnet.princeton.edu/

[2]http://owlapi.sourceforge.net/

Algorithm 1: Building ontology.

Input : $Cat_I : categories, LR :$ *lexical resource*
Output: θ: ontology

$Initialization : \theta \leftarrow (root : animal)$
$Concepts \leftarrow extractConcepts(Cat_I)$
$SubC \leftarrow findHypoCon(root, concepts, LR)$
While $(|\mathbf{SubC}| > \mathbf{0})$**do**
  Foreach $(\mathbf{C} \in \mathbf{SubC}$ **do**
    $HyperC \leftarrow findHyperConcept(\theta, C, LR)$
    $TR \leftarrow createTaxonomicR(HyperC, C)$
    $AddTaxonomicR(\theta, TR)$
  EndForeach
  $SubC \leftarrow findHypoCon(SubC, concepts, LR)$
EndWhile
Return θ

the obtained visual features and the ontology are used to perform the image classification. Visual features are used as the input of SVM classifiers. In our work, the aim is to learn a discriminative model for each class in order 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 a non linear models is impractical. Given the visual features of the training images, we train a One-vs-All SVM classifier (Cortes and Vapnik, 1995) for each class to discriminate between this class and the other classes. Each classifier provide an output confidence value, which allows to determine the probability that a given image belongs to the related class.

In particular, the visual-feature classifiers are

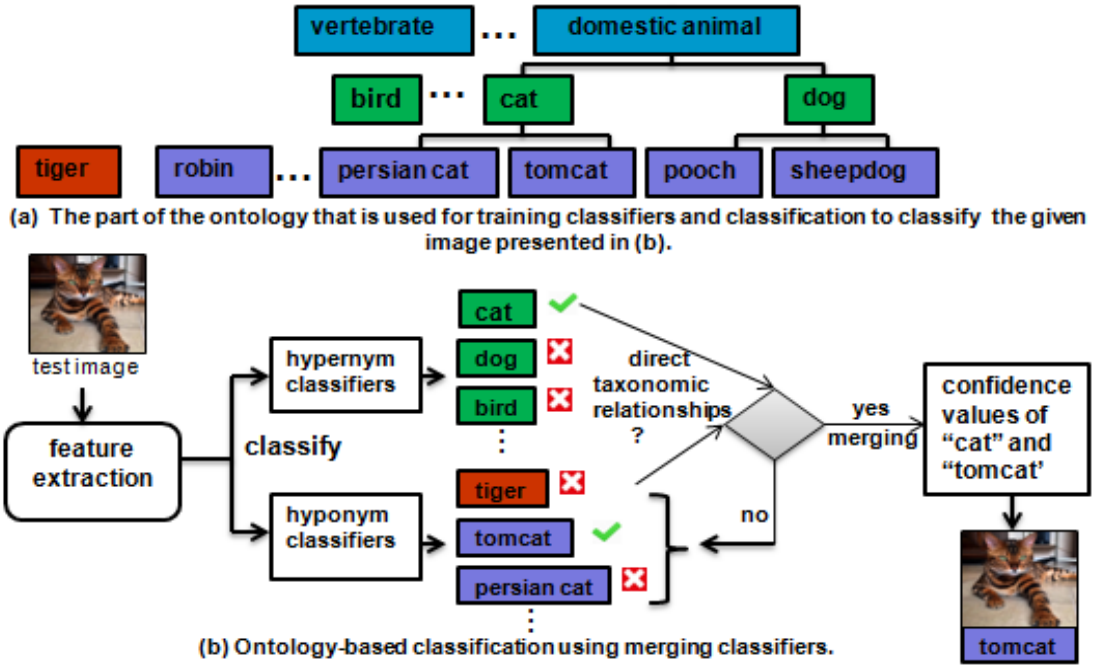

Figure 4: Proposed image classification method.

trained in two different strategies according to the taxonomic relationships between ontology concepts which represent image categories or classes. Firstly, at each node of the ontology, classes are bagged in super-categories based on the ontology in order to train hypernym classifiers. Specifically, based on semantic relationships, we can easily obtain training images for categories at each intermediate semantic level by grouping together images of their sub-categories or hyponym concepts. Secondly, for each hyponym concept in the ontology, visual-feature classifiers are learned using sub-categories as labels and their training images.

As depicted in Figure 4 (a), let us suppose that our ontology contains two hypernym concepts (super-categories) *dog* and *cat* and 4 hyponym concepts (sub-categories) *persian cat*, *tomcat* under *cat*, and *pooch*, *sheepdog* under *dog*. At the given node *domestic animal*, its intermediate children are regarded as the super-categories. For example, the training images of *cat* will include the training image of *persian cat* and *tomcat*, thus the hypernym classifier of this super-category is trained with these images. Then, for each hyponym concept (sub-category), we trained classifiers using their images separately. Finally, given, a test image, classification is done using both hypernym and hyponym classifiers as shown in Figure 4 (b). The test image is assigned to the class that has the best hypernym classifier that is *dog* in our example (because it gives the maximum confidence value). Also, it is assigned to the best hyponym classifier (*tiger* in our example) where classification is done using only the hyponym classifiers. As a result, top-k classes are obtained for each classification.

If the best hyponym class (*tiger*) has a direct relationship with the best hypernym class (*dog*), we merge the output of the hyponym-hypernym classifiers by combining their confidence values. For each selected hyponym class, a new confidence value is computed (is equal to the average of the confidence values of both hyponym and hypernym classes). The test image is assigned to the best hyponym class.

In the case where no direct relationship is detected between the best hyponym class and the best hypernym class, the next best hyponym classes from the top-k classes is considered and treated in the same manner. Therefore, *cat* and *tomcat* classifiers are selected because *tomcat* is the first hyponym class that appear in top k-classes and that has a direct relationship with *cat* class. The average of the confidence values obtained by the two selected classes is computed and is then assigned to *tomcat* class. This class becomes the correct class of the test image.

4 EXPERIMENTATION AND RESULTS ANALYSIS

Throughout this section, we illustrate the experimental results of our work. We start with the experimental setup, then, we present the evaluation of our method.

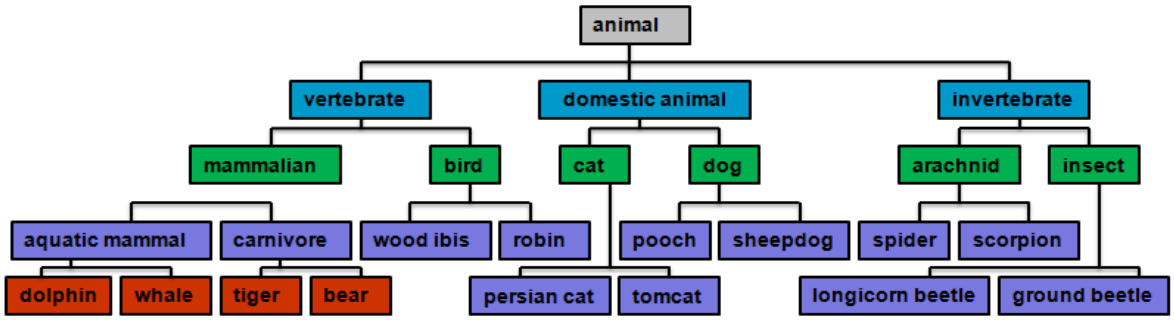

Figure 5: The ontology for the animal dataset.

4.1 Experimental Setup

The aim is to evaluate the classification accuracy of our proposed method. To achieve this goal, we evaluated our approach on an animal dataset from ImageNet (Deng et al., 2009). We selected 22 animal classes from ImageNet, including 6 super-classes (*cat*, *mammalian*, *bird*, *dog*, *insect* and *arachnid*), 12 subclasses and 4 other sub-classes of *carnivore* and *aquatic mammal* categories. For each class, we used 180 images as training data and 20 images as testing images, resulting in a total 4400 images. We use WordNet to generate a semantic ontology for the 22 animal categories. The resulting animal ontology is shown in Figure 5. We use accuracy as metric to evaluate the image classification results. In particular, for each class, the accuracy is computed as follows:

$$ACC = \frac{TP + TN}{N_t} \quad (5)$$

Where: TP and TN is the number of true positive and true negative of images that were correctly classified, respectively, and N_t is the total number of images.

4.2 Evaluation Results

The goal, in this experiment, is to show the effect and the advantages of using ontology to improve the image classification task. For this purpose, we compare our ontology-based image classification approach with the baseline method that consists in classifying images using SVM classification. Thus, to show better performance of our proposed image classification approach, we use different image classification strategies. We introduce the proposed strategies below:

1. **BoVW:** classical BoVW model is used with SVM.
2. **BoVW-ONTO:** same as above, plus ontology.
3. **HMAX:** HMAX features are extracted and classified with SVM.
4. **HMAX-ONTO:** same as above, plus ontology.

For the classification method based on BoVW model, SIFT features are extracted and quantized with KMeans and histograms of visual words are used to train SVM classifiers. In the case of the classification method that based on HMAX model, HMAX features are extracted as detailed in section 3.1.1 and they are used to train SVM classifiers. The size of the final features is, in the BoVW model, given by the size of the vocabulary. However, in the HMAX model, the size is given by the number of the C2 features. For both methods, multi-class classification is done using one-versus-all SVM.

For the classification strategies using ontology (HMAX-ONTO and BoVW-ONTO), features are extracted as previously. For classification, firstly, visual-feature classifiers are trained according to the ontology that describes the taxonomic relationships between image categories. Two ways of training classifiers are applied. In the first way, we trained classifiers which are associated with the super-categories (6 classes). Specifically, based on the taxonomic relationships, the training images of these categories include training images of their sub-classes. For example, training images for the super-category at the node *bird* include training images of *wood ibis* and *robin*. Thus, hypernyms classifiers are trained using images of both super and sub-categories. The second way consists in training classifiers which are associated with the sub-categories. In particular, for each class, we trained classifiers using their images separately. Thus, hyponyms classifiers are obtained using the sub-categories. If a class has children nodes in the ontology, the same process that is used in the first way is applied.

Given the test image, classification is done using both hypernym and hyponym classifiers. Firstly, the test image is assigned to the class that has the best hypernym classifier (that gives the maximum confidence value). Also, it is assigned to the best hyponym classifier where classification is done using only the hyponym classifiers. As a result, for each way, top

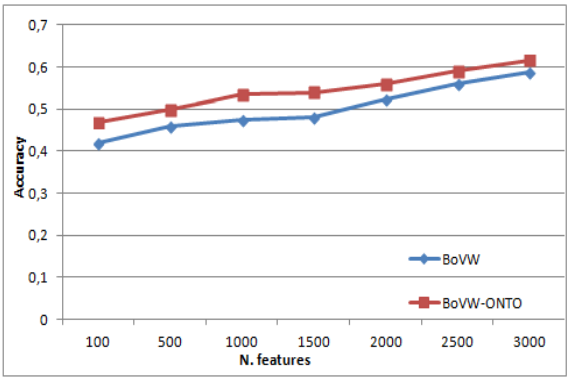

Figure 6: Comparison of accuracy depending on the number of features for BoVW and BoVW-ONTO methods.

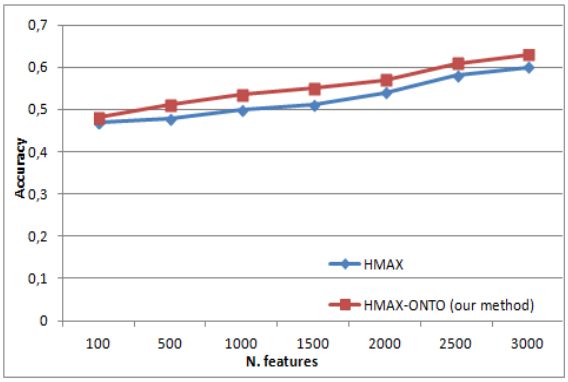

Figure 7: Comparison of accuracy depending on the number of features for HMAX and HMAX-ONTO methods.

k-classes are obtained. Secondly, if the best hyponym class has a direct relationship with the best hypernym class, we merge their confidence values and the test image is assigned to the best hyponym class. If no direct relationship is detected between the best hyponym class and the best hypernym class, the next best hyponym class is considered and treated in the same manner. Classification results of some test images are presented in Figure 11. Figure 6 shows the comparison results in term of accuracy depending on the number of features for the BoVW and BoVW-ONTO methods. The obtained classification results for the strategy based on our method using BoVW model and ontology (BoVW-ONTO) is clearly higher than the strategy based on the BoVW model without ontology. Figure 7 shows the same comparison for HMAX features. It highlights a similar increase in accuracy when the ontology is used. The best accuracies for BoVW-ONTO and HMAX-ONTO methods were performed with a dictionary of 3000 features. In fact, as depicted in Table 2, the best improvement obtained by HMAX-ONTO method is 8%. However, the improvement reaches 12,63% for BoVW-ONTO method. According to the results, we conclude that

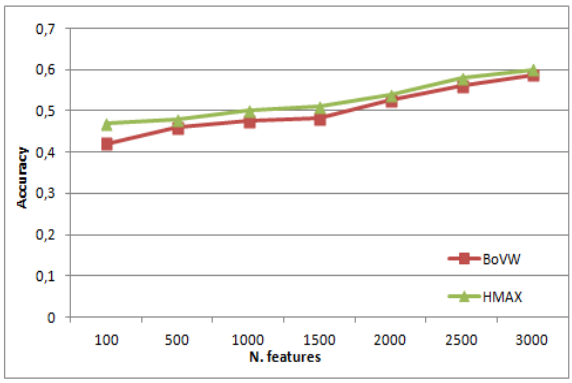

Figure 8: Comparison of accuracy depending on the number of features for BoVW and HMAX methods.

our ontology-based classification method with merged classifiers increases the classification accuracy for both HMAX and BoVW models. This explains that exploiting taxonomic relationships between images categories in line with training visual-feature classifiers and merging their out pouts classifiers, can improve the classification results.

We also focus on comparing the HMAX and BoVW models with and without ontology: HMAX versus BoVW (cf.Figure 8) and HMAX-ONTO versus BoVW-ONTO (cf.Figure 9). We observe in Figure 8 that, using SVM, classification method based on HMAX model provides a better performance than the classification method based on BoVW model. The best accuracy for HMAX method is obtained with a dictionary of 3000 features. When comparing the HMAX to BoVW method, the improvement reaches 11,66% (cf.Table 3). However, as depicted in Figure 9, we observe that, using the ontology, the difference in performance of classification results obtained by HMAX and BoVW models with ontology is much smaller and the best improvement reaches only 3,22% (cf.Table 3), and accuracy values are almost the same with a dictionary of 1000 features (cf.Figure 9). This results indicate that our proposed method, brings an increase in accuracy, independently of the selected features. Moreover, when comparing BoVW-ONTO and HMAX, we observe that the ontology helps the BoVW reach an accuracy (0.61) that is comparable to HMAX without ontology (0.60). Whereas without ontology, the accuracy of BoVW (0.42) is much lower than HMAX (0.46).

Finally, we compare our methods using ontology (BoVW-ONTO and HMAX-ONTO) with Inception-v3 model. The per-class accuracy is shown in Figure 10. Table 4 presents the accuracy comparison of some classes and the average accuracy of all used classes. We find that our methods with ontology outperform the Inception-v3 on some sub-classes such as *pooch*,

Table 2: Accuracy comparison for BoVW versus BoVW-ONTO and HMAX versus HMAX-ONTO methods.

Methods	Low-Acc	Best-Acc	Low-Improvement	Best-Improvement
BoVW	0.42	0.58	–	–
BoVW-ONTO	0.46	0.61	+4,94%	12,63%
HMAX	0.46	0.60	–	–
HMAX-ONTO	0.48	0.63	4,34%	8%

Table 3: Accuracy comparison for HMAX versus BoVW and HMAX-ONTO versus BoVW-ONTO.

Methods	Low-Acc	Best-Acc	Low-Improvement	Best-Improvement
BoVW	0.42	0.58	–	–
HMAX	0.46	0.60	2,21%	+11,66%
BoVW-ONTO	0.46	0.61	–	–
HMAX-ONTO	0.48	0.63	0,18%	3,22%

Table 4: Accuracy comparison of some classes for HMAX-ONTO, BoVW-ONTO and Inception-v3 methods.

Methods	whale	pooch	wood ibis	spider	cheep dog	tomcat	tiger	dolphin	average
HMAX-ONTO	0.65	0.65	0.7	0.7	0.7	0.6	0.7	0.75	0.638
BoVW-ONTO	0.6	0.55	0.65	0.65	0.65	0.6	0.7	0.6	0.611
Inception-v3	0.7	0.3	0.8	0.3	0.65	0.35	0.8	0.7	0.621

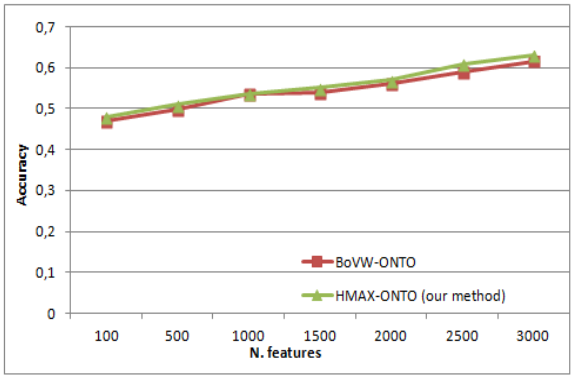

Figure 9: Comparison of accuracy depending on the number of features for BoVW-ONTO and HMAX-ONTO methods.

spider and *tomcat*. Accuracy values are almost the same for some super-classes such as *aquatic mammal* and *dog*. For all classes, the average accuracy obtained by HMAX-ONTO method is a little better than that is obtained by Inception-v3 model. This latter is also a little better than BoVW-ONTO method (cf.Table 4).

5 CONCLUSION

In this paper, an image classification approach has been defined. It relies on training visual-feature classifiers according to the taxonomic relationships between image categories. Firstly, visual features are extracted by adopting HMAX model. Then, concepts are extracted from image categories and taxonomic relationship between them are created to build the ontology, which represents the semantic information associated with the training images. Secondly, two ways of training feature classifiers are applied, the first one consists in training hypernym classifiers using training images of super-categories that included images of their sub-classes. The second way, is performed by training hyponym classifiers using only images of sub-categories. Finally, test images are classified using both hypernym and hyponym classifiers, then when taxonomic relationship between the best hypernym class and the best hyponym class (that appear in top-k hyponym classes) are detected, output classifiers are merged in order to assign best classes to test images. In this work, we have conducted an experimental study where we are focused on the improvement given by our approach. It is worth to be noted that our methods, using HMAX and BoVW models with ontology, achieve superior performance to the baseline methods. Also, using the ontology, the difference in classification performance between HMAX method and BoVW method is much smaller. Moreover, we compared our methods with the Inception-v3 model. Our methods outperform the Inception-v3 on some sub-classes and accuracy values are almost the same for some super-classes. For future work, the idea would be to evaluate our approach on a large

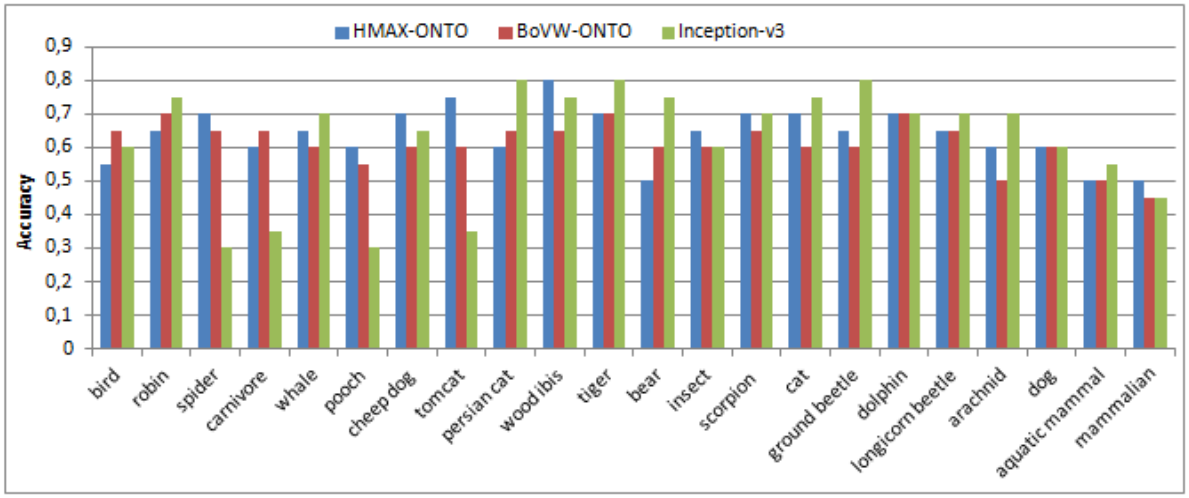

Figure 10: The per-class accuracy comparison of HMAX-ONTO, BoVW-ONTO and Inception-v3 model.

Figure 11: Classification results of some test images using HMAX and HMAX-ONTO methods.

image dataset, as well as to exploit other semantic relationships between classes.

REFERENCES

Abdollahpour, Z., Samani, Z. R., and Moghaddam, M. E. (2015). Image classification using ontology based improved visual words. In *Electrical Engineering (ICEE), 2015 23rd Iranian Conference on*, pages 694–698. IEEE.

Al Chanti, D. and 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*.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. Ieee.

Durand, T., Mordan, T., Thome, N., and Cord, M. (2017). Wildcat: Weakly supervised learning of deep convnets for image classification, pointwise localization and segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, volume 2.

Dutt, A., Pellerin, D., and Quenot, G. (2017). Improving image classification using coarse and fine labels. In *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, pages 438–442. ACM.

Filali, J., Zghal, H. B., and Martinet, J. (2017). Visually supporting image annotation based on visual features and ontologies. In *Information Visualisation (IV), 2017 21st International Conference*, pages 182–187. IEEE.

Gao, H., Dou, L., Chen, W., and Sun, J. (2013). Image classification with bag-of-words model based on improved sift algorithm. In *Control Conference (ASCC), 2013 9th Asian*, pages 1–6. IEEE.

Hu, X., Zhang, J., Li, J., and Zhang, B. (2014). Sparsity-

regularized hmax for visual recognition. *PloS one*, 9(1):e81813.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Lei, J., Guo, Z., and Wang, Y. (2017). Weakly supervised image classification with coarse and fine labels. In *The 14th Conference on Computer and Robot Vision (CRV), 2017*, pages 231–239. IEEE.

Li, Y., Wu, W., Zhang, B., and Li, F. (2015). Enhanced hmax model with feedforward feature learning for multiclass categorization. *Frontiers in computational neuroscience*, 9:123.

Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature neuroscience*, 2(11):1019.

Ristin, M., Gall, J., Guillaumin, M., and Van Gool, L. (2015). From categories to subcategories: large-scale image classification with partial class label refinement. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 231–239.

Rosten, E. and Drummond, T. (2006). Machine learning for high-speed corner detection. In *European conference on computer vision*, pages 430–443. Springer.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., and Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (3):411–426.

Singhal, N., Singhal, N., and 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*, pages 1–5. IEEE.

Sivic, J. and Zisserman, A. (2003). Video google: A text retrieval approach to object matching in videos. In *null*, page 1470. IEEE.

Su, Y. and Jurie, F. (2012). Improving image classification using semantic attributes. *International journal of computer vision*, 100(1):59–77.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826.

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

Theriault, C., Thome, N., and Cord, M. (2013). Extended coding and pooling in the hmax model. *IEEE Transactions on Image Processing*, 22(2):764–777.

Wang, C. and Huang, K. (2015). How to use bag-of-words model better for image classification. *Image and Vision Computing*, 38:65–74.

Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., and Xu, W. (2016a). Cnn-rnn: A unified framework for multi-label image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2285–2294.

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

Wu, J., Yu, Y., Huang, C., and Yu, K. (2015). Deep multiple instance learning for image classification and auto-annotation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3460–3469.

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