

Machine Learning Framework for Image Classification

Sehla Loussaief, Afef Abdelkrim

LA.R.A, Ecole Nationale d'Ingénieurs de Tunis, Université Tunis El Manar. BP 32, le Belvédère 1002.

ENICarthege, Université de Carthage, 35 rue des Entrepreneurs. Charguia II.

Tunis, Tunisie.

sehla.loussaief@gmail.com, afef.a.abdelkrim@ieee.org

Abstract— Hereby in this paper, we are interested to extraction methods and classification in case of image classification and recognition application. We expose the performance of training models on varying classifier algorithms on Caltech 101 images categories. For feature extraction functions we evaluate the use of the classical SURF technique against global color feature extraction. The purpose of our work is to guess the best machine learning framework techniques to recognize the stop sign images. The trained model will be integrated into a robotic system in a future work.

Keywords— *Image feature extraction; Image classification; Image recognition; Training models; Classifier algorithms; SURF; Machine learning*

I. INTRODUCTION

The purpose of this paper is twofold. On one hand, it is an introduction to image classification paradigm. On the other hand, it attempts to give a comparison between different feature extraction and classification algorithms.

The rest of this paper is structured as follows: Section II provides back-ground information on machine learning. Section III presents a detailed description of the Bag of Features paradigm. It also exposes the SURF detector of image Region of interest (ROI) and highlights the unsupervised K-Means algorithm. In section IV we describe learning and recognition based on Bag of Words (BoW) models. Section V discusses experimentations carried out in the evaluation of different classifiers on image Caltech 101 dataset. In the conclusion we synthesize the obtained results and present the current direction of our research.

II. MACHINE LEARNING PARADIGM

Machine learning (ML) is an algorithm set especially suited to prediction. These ML methods are easier to implement and perform better than the classical statistical approaches [1].

Instead of starting with a data model, ML learns the relationship between the response and its predictors by the use of algorithms. During the learning phase ML algorithms observe inputs and responses in order to find dominant patterns.

The machine learning workflow is described in Fig 1:

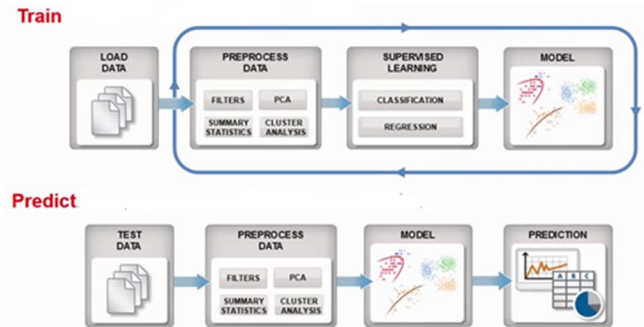


Fig 1: Machine learning workflow

Many fields of modern society use Machine-learning technologies: web searches, content filtering on social networks, recommendations on e-commerce websites, computer vision [2-4], etc.

In this work we are interested in computer vision. We deploy and test a machine learning based framework in image category classification. To achieve tests we use the Caltech¹ dataset.

As the main issue in image classification is image features extraction, we use in our research the Bag of Features (BoF) techniques described in section III.

III. BAG OF FEATURES PARADIGM FOR IMAGE CLASSIFICATION

Development of Bag of Features (BoF) model is inspired by that of Bag of words (BoW). In document classification fields (text documents), a BoW is a vector that represents the frequency of vocabulary words in a text document.

In computer vision, BoW model is used to classify image. In that case image features are considered as words. With the use of BoW model, an image is considered as a document. For "words" definition in images we use three stages: feature extraction, feature description (section III.A), and codebook generation (section III.B) [5-11].

A. Speed Up Robust Features (SURF) Technique

For features detection and extraction, we use the Speed Up Robust Features method. Salient features and descriptors are extracted from each image. This method is chosen over Scale-Invariant Feature Transform (SIFT) due to its concise descriptor length. In SURF, a descriptor vector of length 64 is generated using a histogram of gradient orientations in the local neighborhood around each key-point [12]. To analyze image and

¹ http://www.vision.caltech.edu/Image_Datasets/Caltech101/

extract features, SURF considers the processing of grey-level images only as they contain enough information [13].

In this paper, the SURF implementation is provided by the Matlab R2015a library.

B. Descriptors clustering: K-Means

After extracting descriptors from the training images, unsupervised learning algorithms, such as K-means, are used in order to group them into N clusters of visual words. The metric used to categorize a descriptor into its cluster centroid is the “Euclidean distance”. For this purpose, each image extracted descriptor is assigned to its closest cluster centroid.

In order to generate the histogram of counts, the cluster centroid's number of occupants is incremented each time a descriptor is mapped into it.

At the end of this process, each image is characterized by a histogram vector of length N. To ensure the invariance of this method with respect to the number of descriptors used, it is essential to normalize each histogram by its L2-norm.

To group the descriptors and construct the N visual words we use the K-means clustering. This approach is selected over Expectation Maximization (EM) as many experimental methods have confirmed the computational efficiency of K-means with respect to EM [14].

IV. LEARNING AND RECOGNITION BASED ON BoF MODEL

Research in computer vision field have led to many learning approaches to leverage the BoF model for the purpose of image recognition. For multiple label classification problems, the evaluation metric which is used is the confusion matrix.

A confusion matrix is defined as a particular table making it possible to visualize the accuracy of a supervised learning algorithm. Matrix columns symbolize the instances in a predicted class whereas rows represent the instances in an actual class (or vice-versa). The appellation is due the fact that it makes it simple to see if the system confuses two categories (i.e. mislabeling one as another) [15].

In this work we investigate many supervised learning algorithms such as: SVM [16], k-nearest neighbors [17] and Boosted Regression Trees [18,19] to classify an image. Each image in dataset is encoded by its “BoF” histogram vector as shown in Fig 2.

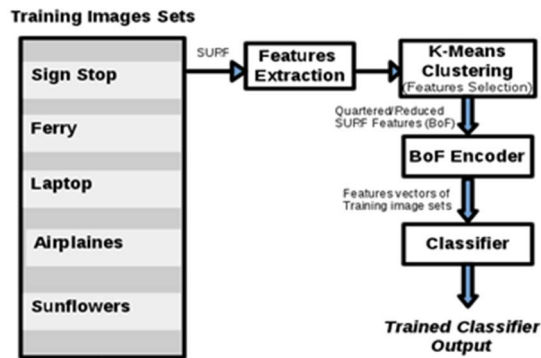


Fig 2: Image classification process

V. EXPERIMENTS AND EVALUATION

In the following we provide a summary of different experiments that we use to evaluate the performance of our image classification machine learning framework. Our results are reported on Calltech 101 image dataset to which we have added some new images of existing categories. We are interested in stop sign category recognition.

A. SURF Local Feature Extractor and Descriptor

In this experiment, we test the local feature extractor SURF and its robustness in matching features even after rotation and scaling image (Fig. 3, Fig. 4).



Fig. 3. SURF feature detection in stop sign image

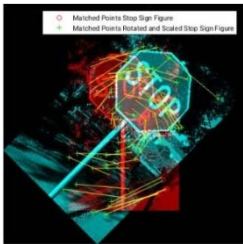


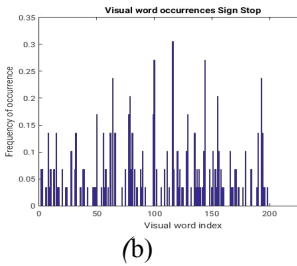
Fig. 4. SURF features points matching capabilities

B. Bag of Features Image Encoding

We use BoF to encode each image of the dataset into a vector feature which rep-rents the histogram of visual word occurrences contained in it (Fig. 5).



(a)



(b)

Fig. 5. Histogram of visual words occurrences on stop sign image

C. Classifier Training Process

The encoded training images are fed into a classifier training process to generate a predictive model. In this section, we are interested in measuring the classifier average accuracy and its confusion matrix.

Image categories from Calltech101 dataset used are described in TABLE I.

TABLE I. IMAGE DATASET CATEGORIES					
Set category	Stop Sign Images	Ferry Images	Laptop images	Airplanes images	Sunflower Images
Set size	69	67	81	800	85

1) Experiment 1: Classifier Evaluation Based on the Number of Image Categories.

For tests we use the SURF extractor and the Linear SVM classifier. We chose 30% among images for training and the remaining for validation. The obtained confusion matrixes are shown in TABLE II. , TABLE III. , TABLE IV. and TABLE V.

TABLE II. CONFUSION MATRIX FOR TWO IMAGE CATEGORIES CLASSIFICATION

Known	Predicted	
	Stop Sign	Ferry
Stop Sign	0.98	0.02
Ferry	0.00	1.00

TABLE III. CONFUSION MATRIX FOR THREE IMAGE CATEGORIES CLASSIFICATION

Known	Predicted		
	Stop Sign	Laptop	Ferry
Stop sign	0.93	0.04	0.03
Laptop	0.02	0.77	0.21
Ferry	0.00	0.02	0.98

TABLE IV. CONFUSION MATRIX FOR FOUR IMAGE CATEGORIES CLASSIFICATION

Known	Predicted			
	Stop Sign	Laptop	Ferry	Airplanes
Stop sign	0.95	0.05	0.00	0.00
Laptop	0.00	0.90	0.10	0.00
Ferry	0.00	0.00	0.85	0.15
Airplanes	0.00	0.05	0.10	0.85

TABLE V. CONFUSION MATRIX FOR FIVE IMAGE CATEGORIES CLASSIFICATION

Known	Predicted				
	Stop Sign	Laptop	Ferry	Airplanes	Sunflowers
Stop Sign	0.97	0.00	0.00	0.00	0.03
Laptop	0.07	0.49	0.14	0.10	0.20
Ferry	0.02	0.00	0.76	0.17	0.05
Airplanes	0.00	0.04	0.20	0.75	0.01
Sunflowers	0.00	0.00	0.02	0.00	0.98

As shown in Fig. 6, we notice that the average accuracy of the classifier is influenced by the number of categories in training dataset. This metric is lower when the numbers of sets increase.

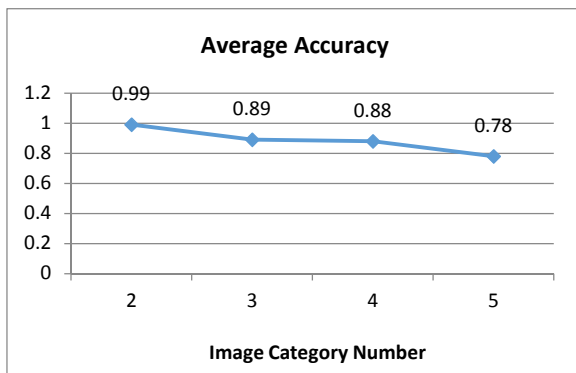


Fig. 6. Average accuracy variation based on the number of training dataset image categories

2) Experiment 2: Evaluating Image Category Classifier Using Color Extractor.

The Linear SVM classifier is applied on sign stop, ferry and laptop categories. We use a global color features extractor instead of the SURF technique.

It is reported that the achieved average accuracy is 0.76, as shown in TABLE VI.

We notice that in our approach is better to use a local features extractor (SURF) than a global features extractor. This result is expected as the global features extraction technique is better with scene categorization and not for object classification [20].

TABLE VI. CONFUSION MATRIX CLASSIFIER USING GLOBAL COLOR FEATURES EXTRACTOR

Known	Predicted		
	Stop Sign	Laptop	Ferry
Stop Sign	0.84	0.04	0.12
Laptop	0.10	0.67	0.23
Ferry	0.02	0.22	0.76

3) Experiment 3: Training Learner Evaluation.

We next fix the number of categories to 4, the feature extraction technique to SURF and evaluate models on varying the classifier algorithm: SVM, KNN and ensemble classifier categories. We then generate the histogram (Fig.7) of the average accuracy based on the training classifier.

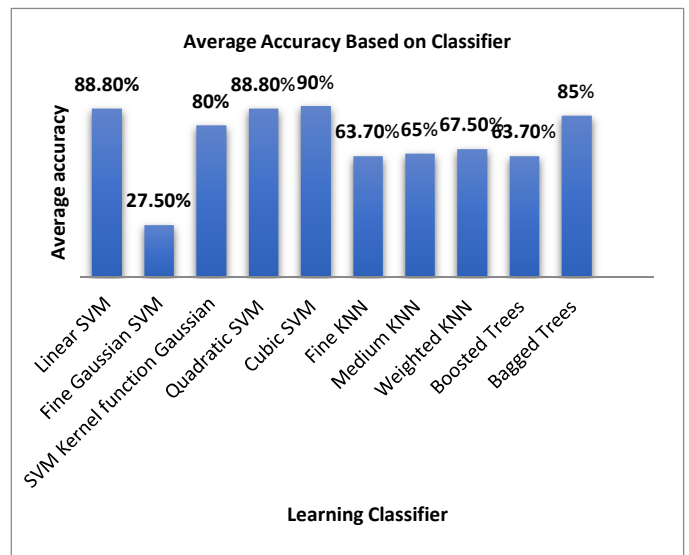


Fig. 7. Average accuracy based on the learning classifier

Measurements show that the image classification process performs better when we use a likelihood SVM. It's reported that the Cubic SVM yields average accuracy which reaches 90%. The KNN techniques offer an average accuracy around 65%. Among the ensemble classifier trainers (2 last tested algorithms) the bagged trees achieves the best accuracy.

VI. CONCLUSION

In this paper, we related the different techniques and algorithms used in our machine learning framework for image classification. We presented machine learning state-of-the-art applied to computer vision. We introduced the Bag of Features

paradigm and highlighted the SURF as its technique for image features extraction and description. Through experimentations we proved that using SURF local feature extractor method and a SVM (cubic SVM) training classifier performs best average accuracy. In test scenarios we focused on stop sign image as we project to apply the trained classifier in a robotic system.

ACKNOWLEDGMENT

We would like to thank Ehsen Zayen for helpful discussions and encouragement.

REFERENCES

- [1] L.Breiman, "Statistical modeling: the two cultures". Statistical Science, 16, 199–215, 2001.
- [2] M.Kef, L.Chergui, S.Chikhi, "Comparative study of the use of geometrical moments for Arabic handwriting recognition". SETIT'2012, Tunisie. DOI: 10.1109/SETIT.2012.6481933, 2012.
- [3] E.Zigh, M.Belbachir, "Hopfield network for stereo matching of panchromatic urban IKONOS images". SETIT'2012, Tunisie. DOI: 10.1109/SETIT.2012.6481934, 2012.
- [4] O.Bencharef, M.Fakir, B.Minaoui, A.Hajraoui, M.Oujaoura, "Color objects recognition system based on artificial neural network with Zernike, Hu & Geodesic descriptors". SETIT'2012, Tunisie. DOI: 10.1109/SETIT.2012.6481938, 2012.
- [5] L.Fei-Fei, P.Perona, "A Bayesian hierarchical model for learning natural scene categories". IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) 2: 524. doi:10.1109/CVPR.2005.16. ISBN 0-7695-2372-2, 2005.
- [6] T.Lindeberg, "Scale selection, computer vision: a reference guide, (K. Ikeuchi, ed.), Springer, pages 701-713, 2014.
- [7] E.Rosten, T.Drummond, "Machine learning for high-speed corner detection". European Conference on Computer Vision. Springer. pp. 430–443. doi: 10.1007 / 11744023_34 .CiteSeerX: 10.1.1.60.3991, 2006.
- [8] U.Vidal-Naquet, "Object recognition with informative features and linear classification". Proceedings Ninth IEEE International Conference on Computer Vision (PDF). pp. 1150–1157. doi:10.1109/ICCV.2003.1238356. ISBN 0-7695-1950-4, 1999.
- [9] R.Cuingnet, C.Rosso, M.Chupin, S.Lehéricy, D.Dormont, H.Benali, Y.Samson, O.Collot, "Spatial regularization of SVM for the detection of diffusion alterations associated with stroke outcome, Medical Image Analysis", 15 (5): 729–737, 2011.
- [10] A.Chergui, A.Bekhoucha, W.Sabbar, "Video scene segmentation using the shot transition detection by local characterization of the points of interest". SETIT'2012, Tunisie. DOI: 10.1109/SETIT.2012.6481949, 2012.
- [11] J.Hu, P.Guo, "Spatial local binary patterns for scene image classification". SETIT'2012, Tunisie. DOI: 10.1109/SETIT.2012.6481936, 2012.
- [12] D.Lowe, "Towards a computational model for object recognition in IT cortex". Proc. Biologically Motivated Computer Vision, pages 2031, 2000.
- [13] M. S.Nixon, A. S.Aguado, "Feature Extraction and Image Processing". Elsevier Ltd, ISBN: 978-0-12372-538-7, 2008.
- [14] D. G.Lowe, "Distinctive image features from scale-invariant keypoints". IJCV, 60(2):91–110, 2004.
- [15] M. W.David, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation (PDF)". Journal of Machine Learning Technologies 2 (1): 37–63, 2011.
- [16] R.Piyush, "Hyperplane based Classification: Perceptron and (Intro to) Support Vector Machines", CS5350/6350: Machine Learning, 2011.
- [17] C.Domeniconi, D.Gunopulos, J.Peng, "Large margin nearest neighbor classifiers". IEEE Transactions on Neural Networks, 16(4):899–909, 2005.
- [18] L.Rokach, O.Maimon, "Data mining with decision trees: theory and applications". World Scientific Pub Co Inc. ISBN 978-9812771711, 2008.
- [19] J. H.Friedman, "Greedy Function Approximation: A Gradient Boosting Machine", 1999.
- [20] Z.Jinyi, L.Wei, C.Chen, D.Qian, "Scene classification using local and global features with collaborative representation fusion". Information Sciences 348, 2016.