Lab 8: PHOW

Cristian M. Amaya
Universidad de los Andes
cm.amaya10@uniandes.edu.co

Abstract: In this laboratory, a PHOW-based classifier will be train with the VLfeat library to obtain the best result possible in the ImageNet200 dataset, to be able to do this, a method that works with the Caltech 101 dataset will be adapted and optimized.

Keywords: Classification, dataset, SIFT, SVM

1.. Introduction

Caltech 101 is one important dataset in the artificial vision history, as it made the classification problem the main focus of researchers for a large period of time. Many datasets proceeded Caltech101, these new ones possess more images, multiple classification classes and consider more class variation parameters as the image are more natural, one of these datasets is Imagenet200.

In this practice, an Oriented Histograms based classification method will be adapted to run on the Imagenet200 database, the results will be compare to the results of the original method in Caltech101 to observe the difference between both datasets.

2.. Methods and Materials

2.1.. Datasets

ImageNet200: Classification dataset that consist of 200 categories and 100 images of training and another 100 images of test for each category. The intention of ImageNet was to create a image ontology database in which the factor of scale and diversity in each class was taken into account to create each set. [2]

Caltech 101: Dataset that consist of 101 categories, it was based in a subset of categories from the dictionary and the early Google image search engine. The images were rotated in an arbitrary angle, resize in a specific scale and faced in the same direction.[3]

2.2.. Classification method and library

The classification method was created by Andrea Vedaldi and uses the VLfeat library functions, the method uses dense SIFTs, spatial histograms of visual words and SVMs to classify the images in Caltech101. The principal feature of this method is the Pyramid Histogram Of Visual Words (PHOW) features are a variant of dense SIFT descriptors, that can describe orientations at multiple scales of the image.[1]

The method was change to be able to use the Imagenet200's training and test sets, the predefine parameters will be also change to obtain better results taking into account the difference between the datasets.

2.3.. Parameters

The most relevant parameters for this method are: the number of images in the training and test sets that will be considered to build the classifier and test it, the number of categories to take into account, the number of visual words to be used, the C parameter for the SVM and the spatial partitioning in X and Y. Other parameters will be take tested, but this are the more pertinent ones.

To optimize the classifier's accuracy the best set of parameters will be obtained through experimentation, changing one parameter at a time will keeping the others unchanged will be a good approach assuming that the parameters are not heavily codependent, at the end, the best parameters will be tested all at ones to see how.

3.. Results and Discussions

3.1.. Datasets comparison

In figure 1 and 2, the ACA for the Caltech101 classifier apply on similar sets of Caltech and imageNet can be seen. The ACA for imageNet is far lower that Caltech101, this could be as Caltech images are from catalog, most of them have a similar orientation and background, and only 102 are consider.

On the other hand, the images in imageNet200 are more natural, which can lead to more interclass variations, different kinds of orientation and backgrounds. Which can be a disadvantage to a classifier based in orientations mainly.

3.2.. Parameter testing

The results for the parameter testing can be seen in figure 3 through 5, first the relation between the number of test and train images was tested, the initial method begins with 15 images for each set, the final decision was to use all 100 images of each class to train the models, as more interclass variations are taken into account.

Other parameters were also tested, such as the spatial partitions in X and Y, in which the use of [3 5] for both coordinates produce the best result. The C parameter for the SVM was also tested, with larger C the margin of the hyperplane becomes thinner and can produce better results, the number of visual words was one of the most important parameters, as more features can be consider for the classification. Other parameters were tested but didn't improve the results, therefore they weren't consider.

Finally, the best parameters were tested one a time with the default parameters and 100 images of test, and training, as it can be appreciated in figure 5.

3.3.. Classification results

The final method's confusion matriz can be seen in figure 6, the best result for the PHOW classifier so far is 26,46%, this can be a consequence of the limitation of bag of words and orientation descriptors, other methods like neural network or random forest can be used, but this could require a great effort just to optimize the parameters and prevent overfitting, in the case of neural networks.

In the case of best and worst classes, that can be seen in figure 7, the most important insight is that dogs are difficult to classify, as 4 of the worst categories are of different dog races. In the best cases, web sites and zebras are categories that one can expect to be easily classified, also Schooner is the best class, this could be as the sea can be quite different from a boat and it is also a semi uniform background.

4.. Conclusions

The difference between Caltech101 and ImageNet200 can be consider a milestone in the evolution of the classification study, as more distortion and variation factors are taking into account went one studies natural images. Which proves the difficulty of this branch of study when one uses natural images instead of catalog ones, the introduction of ImageNet was an important step and lead to a new era of investigation with its winning method.

It is also apparent the limitation of the PHOW classifier and their visual words went testing its different parameters, which shows the need of new techniques that can overcome some of this limitations, which now exist and it is well known that neural networks won over ImageNet. But for the limitations and limited testing, the ACA is larger that simple random probability of classification (0,5%) and the initial ACA was increased by a factor of 186%.

References

[1] A. Bosch, A. Zisserman, and X. Munoz. Image classification using random forests and ferns. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.

[2] J. Deng, K. Li, M. Do, H. Su, and L. Fei-Fei. Construction and Analysis of a Large Scale Image Ontology. Vision Sciences Society, 2009.

[3] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer vision and Image understanding*, 106(1):59–70, 2007.

Images

Dataset comparison

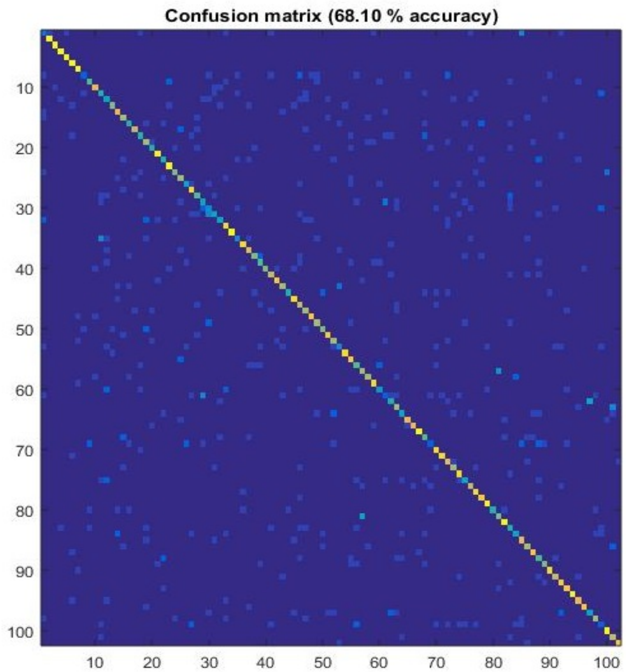

Figure 1. Confusion matrix for the Caltech101 Dataset. (ACA=68.1%)

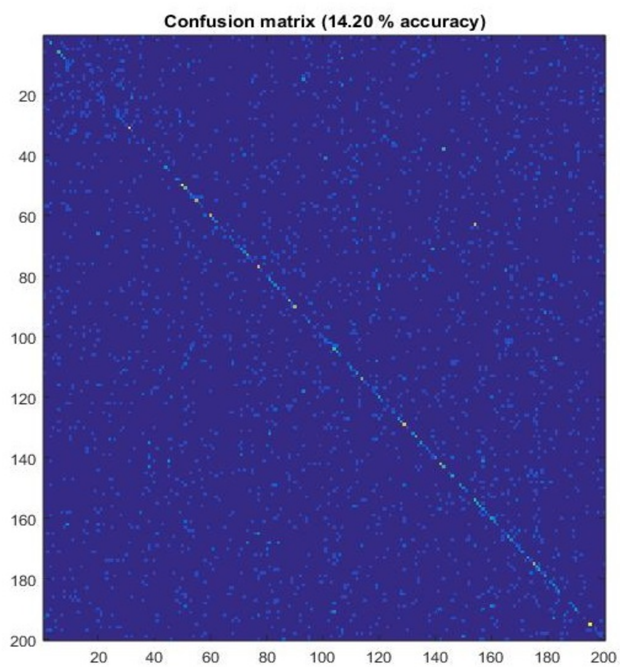

Figure 2. Confusion matrix for the imageNet200 Dataset. (ACA=14.2%)

Parameters testing

Images		ACA
Train	Test	
10	10	12,4%
15	15	14,2%
15	30	14,6%
30	15	19,2%
30	30	18,9%
50	30	21,2%
50	50	21,3%
100	50	24,0%
100	100	**24,3%**

Figure 3. Evaluation with different train and test sets sizes

Parameter	Value	ACA
Spatial Partition	X[2 4] Y[2 4]	14,2%
	X[3 5] Y[3 5]	**14,53%**
	X[4 6] Y[4 6]	13,70%
Number of Words	600	14,2%
	700	15,17%
	800	14,93%
	1000	**15,63%**
Bias Multiplier	1	**14,2%**
	3	1,93%
SVM C	10	14,2%
	15	14,20%
	30	**14,60%**
Max Comparisons	50	14,20%
	100	14,20%

Figure 4. Evaluation with different parameters variations

Parameters		Classifier			
		1	2	3	4
Images	Train	100	100	100	100
	Test	100	100	100	100
Number of words		600	1000	1000	1000
Spatial Partition		XY[2 4]	XY[2 4]	XY[3 5]	XY[3 5]
SVM C		10	10	10	30
ACA		24,30%	**26,46%**	26,38%	26,44%

Figure 5. Evaluation of classifiers with best parameters

Classification results

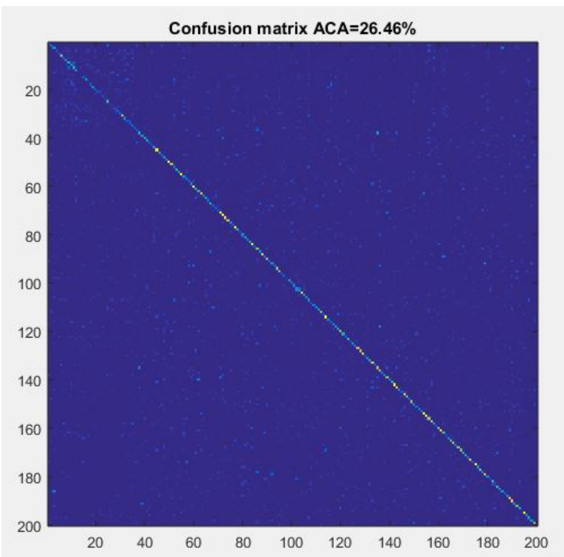

Figure 6. Final imageNet200 classifier (ACA=26.46%)

Best Scores		Worst Scores	
Class	ACA	Class	ACA
Schooner	69,9%	Labrador retriever	3,6%
Web site	66,9%	Weasel	2,7%
Zebra	62,6%	Lakeland terrier	2,4%
Oscilloscope	60,0%	Chesapeake Bay retriever	1,5%
Convertible	59,2%	Chihuahua	1,2%

Figure 7. Best and worst classes for the classifier