PHOW Image Classification - Lab 8 - Computer Vision

Yesith Juez
University of the Andes
jy.juez@uniandes.edu.co

Nicolas Florez
University of the Andes
n.florez228@uniandes.edu.co

1. Introduction

Since the real world is made up of a wide variety of objects (visually classified by humans), where each one varies in its position and intrinsic behavior within a category (for example, a great variety of dogs), the need arises of the recognition of objects by means of computer vision [1]. Due to the large area of interest that has developed in this field, two databases were created that are now of great importance for the recognition of objects (Caltech 101 and ImageNet). The goal of this laboratory is to perform the classification of a set of images for each databases, using the method of Pyramidal Histograms of Visual Words (PHOW).

2. Materials and Methods

2.1. Database

In this laboratory we used two (2) databases to better understand how is the impact of the database with the behavior of the classifier.

2.1.1 Caltech-101

Caltech 101 is a database that contains images of objects divided into 101 categories. Each category has approximately 40 to 800 images. This database was created by Fei-Fei-Li, Marco Andreetto and Marc Aurelio Ranzato, who formed the images with a size of 300x200 pixels [2].

2.1.2 ImageNet

ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Currently they have an average of over five hundred images per node. ImageNet contrary to Caltech101 has differentiated the sets of images as Train and Test. Caltech101 separates train and test randomly in the code from the whole database.

2.1.3 PHOW

The PHOW allows to explore spatial position information based on a spatial pyramid model, such as SIFT. The characteristics of the representation of a spatial pyramid are the following: Extension of a bag of features and Locally orderless representation at several levels of resolution [3].

For this study we used different sets of images from the database in order to acknowledge how the database affects directly in the average of the classification within the descriptor used in this laboratory.

We changed the number of categories from 5 to 30 to see just the effect of incrementing the categories.

We also changed the number of images within a category in order to observe the effect of reducing the number of images of a whole category.

Changing those variables results in various sets of experimentation that we used to compare the performance of the algorithm and the descriptor itself.

The sets of variables were used for both train and test sets of the ImageNet as Caltech101 does not differentiate the train and test sets.

2.2. Image features

The image features used in the algorithms are from the SIFT descriptor. Sift generates a Bag of Words representation of the images. The Sift implementation used was an algorithm created by Andrea Vedaldi available in the VLfeat library that is used in the example script phow_caltech101.m.

For the database Caltech101 it wasn't necessary to do any change of the code as the example was created purely for that database. In order to make it work for ImageNet some lines of code were changed to make it work.

2.3. Classification

The step for the classification of the images from the databases was based again in the Vedaldi script phow_caltech101.m. In the example its used a variant from the Supervised machine learning algorithm SVM (Support Vector Machine) using the Chi squared distance metric. The

Variant of the algorithm used is found in the library used VLFeat.

2.3.1 Evaluation

From the function phow_Caltech101.m, for the Caltecht database the number of classes was varied between 5, 10, 15 where for each number of classes the parameters of SIFT descrptors x (Sx), SIFT descriptor y (Sy) and the number of images were varied, obtaining for each case the Accuracy. (see figure 19). The number of images was only varied up to 20 images, because in some categories the maximum was 40 images (20 for test and 20 for training).

For the ImageNet database the number of classes was varied between 5, 10, 15 and 30, where for each number of classes the parameters of SIFT descrptors x (Sx), SIFT descriptor y (Sy) and the number of images were varied, obtaining for each case the Accuracy. (see figure 44).

3. Results

The confusion matrices were calculated and the performance(Accuracy) of the algorithm for every set described in the previous section for each database. For Caltech101 the confusion matrices can be found in the Figures 1-18 with its respective accuracies. For ImageNet database the same matrices can be found in the Figures 20-43. Additional to the confusion matrices in the tables found in the Figures 19 and 44 we resume the different scenarios used with its respective accuracies that inform the performance of the algorithm with each set.

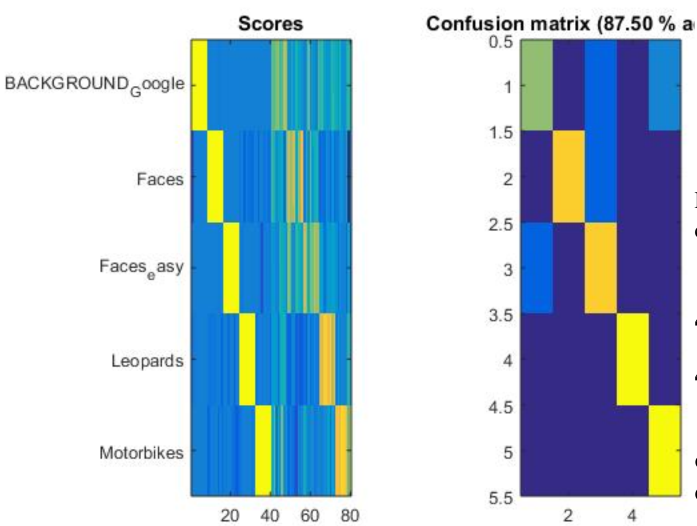

Figure 1. Confusion Matrix of the set 5 categories 8 images per category Sx=Sy=2 for Caltech101

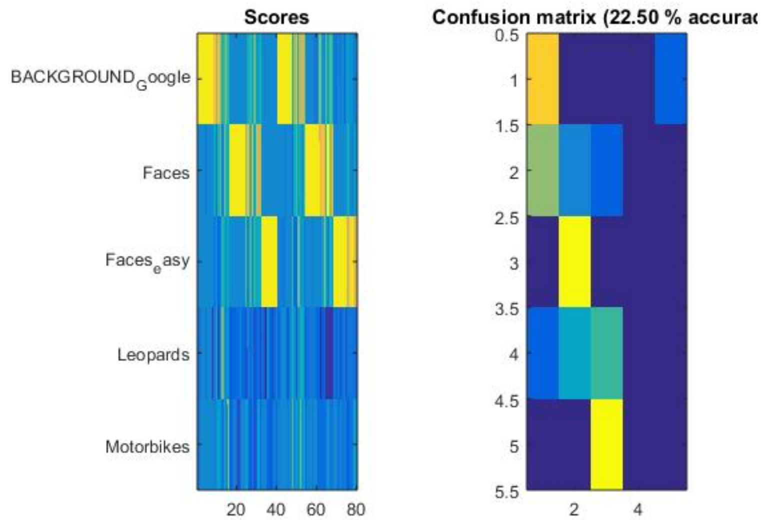

Figure 2. Confusion Matrix of the set 5 categories 15 images per category Sx=Sy=2 for Caltech101

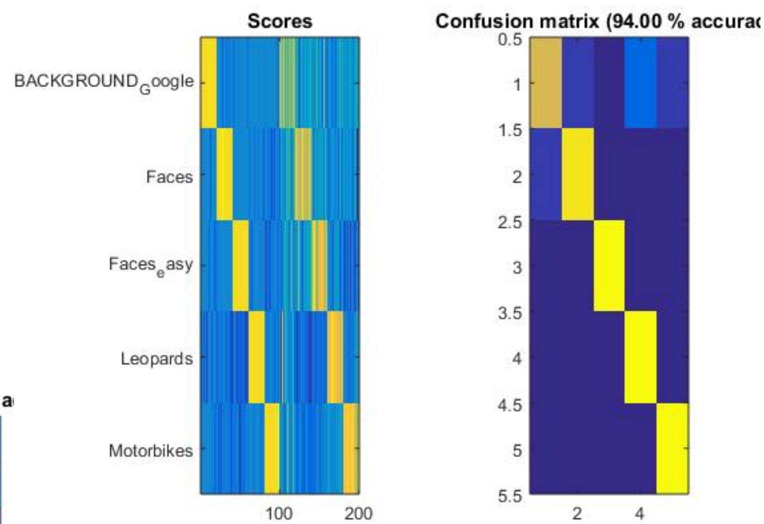

Figure 3. Confusion Matrix of the set 5 categories 30 images per category Sx=Sy=2 for Caltech101

4. Discussion

4.1. Caltech101

For Caltech 101 using the table in the Figure 19 can be observed that the high accuracy from all the sets is 96% corresponding to the scenario with 5 Classes, SX=SY=8, 15 Images for train and 15 images for test. The minimum accuracy rate is from the set of 5 classes, SX=SY=2 and 15 images for both train and set with an accuracy of 23%. This informs us in first instance that belonging to the same classes and the same amount of images SX and SY affects

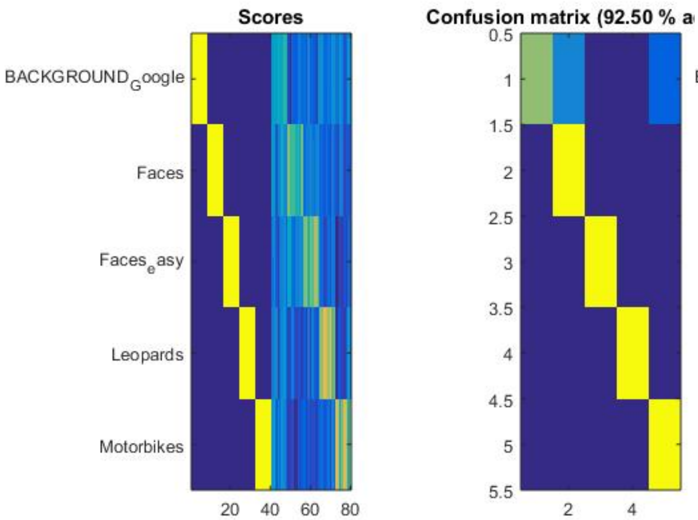

Figure 4. Confusion Matrix of the set 5 categories 8 images per category Sx=Sy=8 for Caltech101

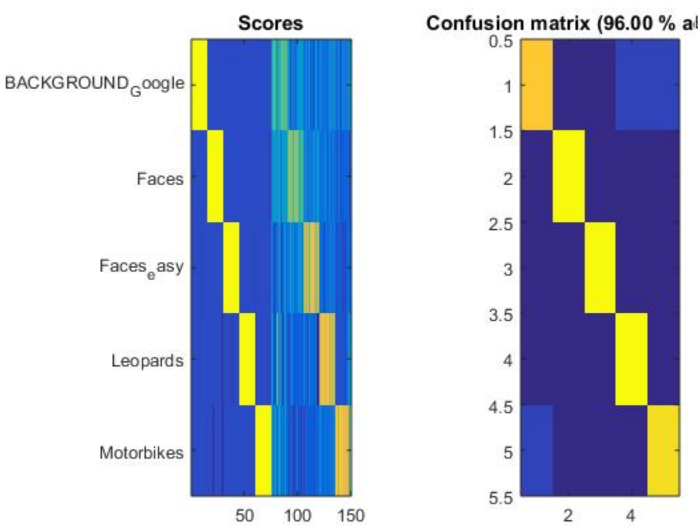

Figure 5. Confusion Matrix of the set 5 categories 15 images per category Sx=Sy=8 for Caltech101

in a high impact the performance of the overall algorithm.

Also analyzing the table can be found that putting more classes to the algorithm affects negatively to the accuracy of the algorithms reducing the performance and the accuracy percentage. In the same way using more images per class increments the Algorithm performance.

With those key finding the ideal scenario is formed with a few classes, high SX and SY, and high amount of images.

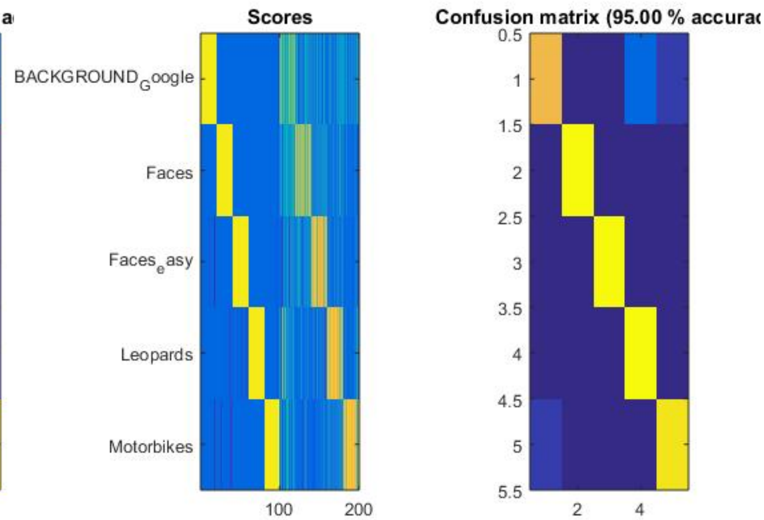

Figure 6. Confusion Matrix of the set 5 categories 30 images per category Sx=Sy=8 for Caltech101

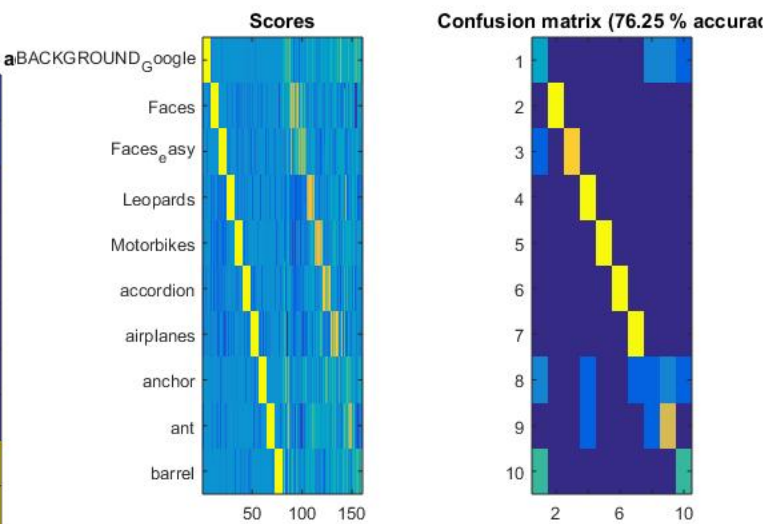

Figure 7. Confusion Matrix of the set 10 categories 8 images per category Sx=Sy=2 for Caltech101

4.2. ImageNet

For ImageNet using the table in the Figure 44 can be observed that the high accuracy from all the sets is 54% corresponding to the scenario with 5 Classes, SX=SY=2, 30 Images for train and 30 images for test. The minimum accuracy rate is from the set of 30 classes, SX=SY=8 and 8 images for both train and set with an accuracy of 10%.

The findings from this database is that incrementing the number of images also increments the accuracy rate of the algorithm and following Caltech101 using few classes also increments the accuracy rate.

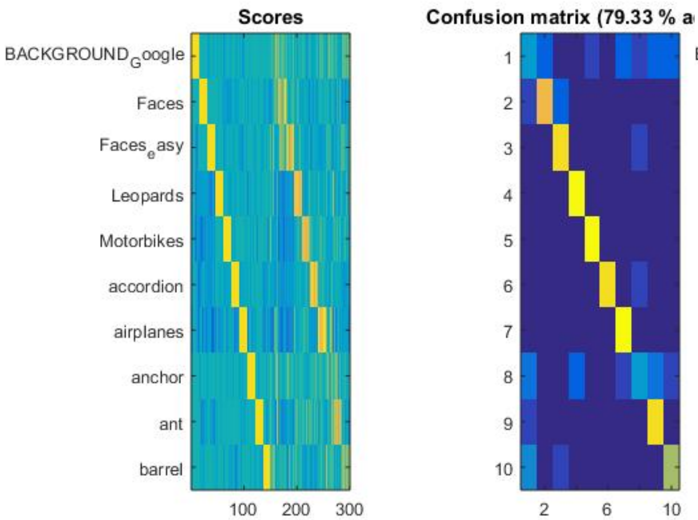

Figure 8. Confusion Matrix of the set 10 categories 15 images per category Sx=Sy=2 for Caltech101

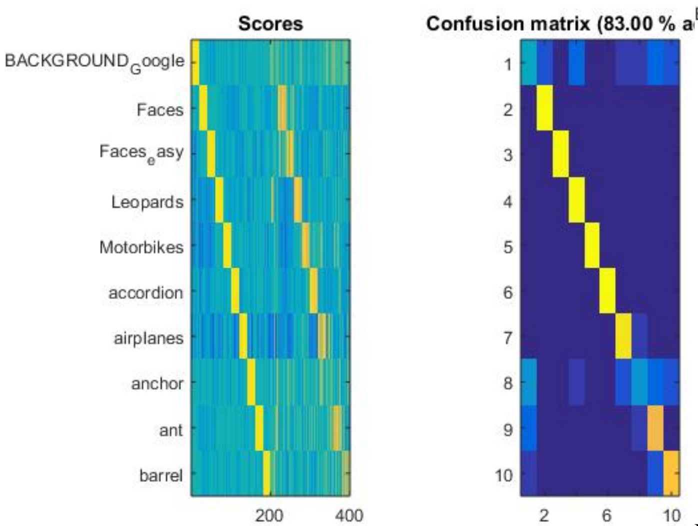

Figure 9. Confusion Matrix of the set 10 categories 20 images per category Sx=Sy=2 for Caltech101

The numbers form ImageNet database differs a lot from CALTECH database that tells us that the use of the database affects a lot the performance of the algorithm an with that the scenario so the algorithms must be created thinking in the attacked problem.

With the findings of both database with all the scenarios used we can think that the ideal scenario is formed with a few classes, high SX and SY, and high amount of images.

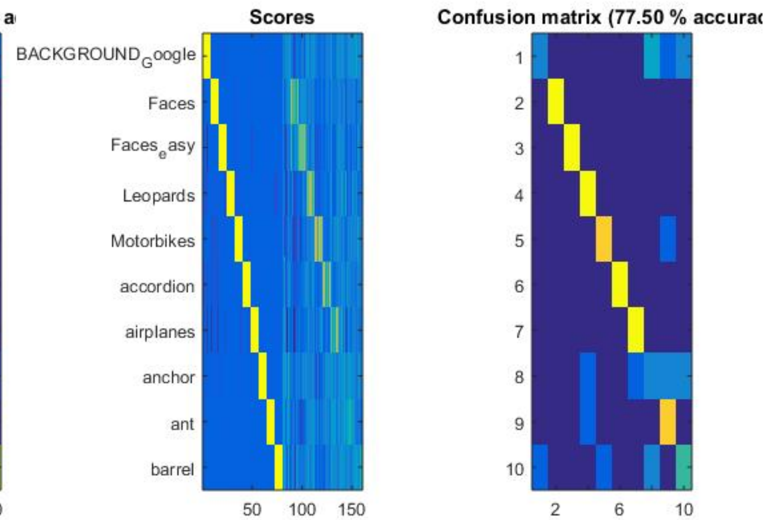

Figure 10. Confusion Matrix of the set 10 categories 8 images per category Sx=Sy=8 for Caltech101

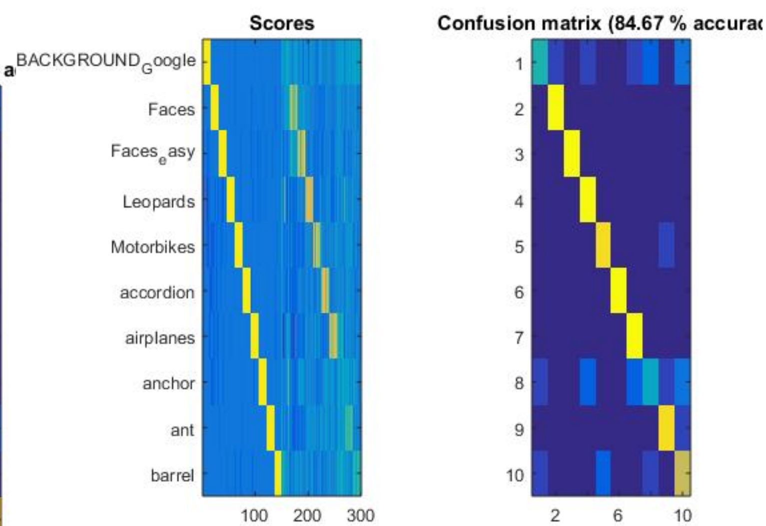

Figure 11. Confusion Matrix of the set 10 categories 15 images per category Sx=Sy=8 for Caltech101

5. Conclusions

The recognition of images based only on descriptors is a relatively complex task when it comes to natural images, since it has a great variety in the real world (it can be represented in different ways), so it is important to extract the descriptors for this class taking into account the other descriptors of SIFT.

This method of classification, unlike others, has a very low computational cost, since it divides the image into different portions.

Contrary to what one would expect, the number of

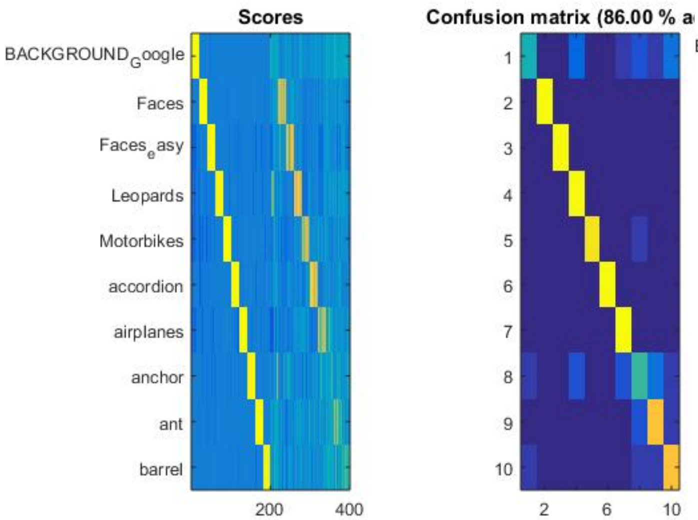

Figure 12. Confusion Matrix of the set 10 categories 20 images per category Sx=Sy=8 for Caltech101

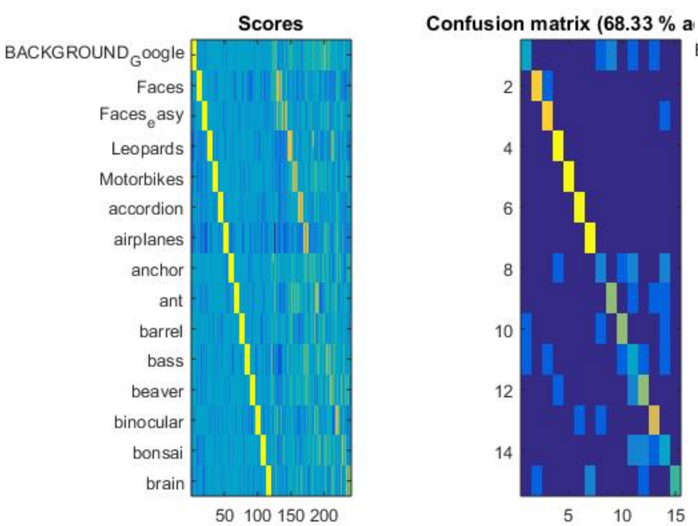

Figure 13. Confusion Matrix of the set 15 categories 8 images per category Sx=Sy=2 for Caltech101

classes affects the performance of the classifier as it increases.

Comparing the two databases, the one with better performance is caltech101, because it has a greater number of images for each category.

References

[1] R. Szeliski, "Computer Vision: Algorithms and Applications", 2011.

[2] Caltech 101, "http://www.vision.caltech.edu/Image_Datasets/Caltech101/", 2006.

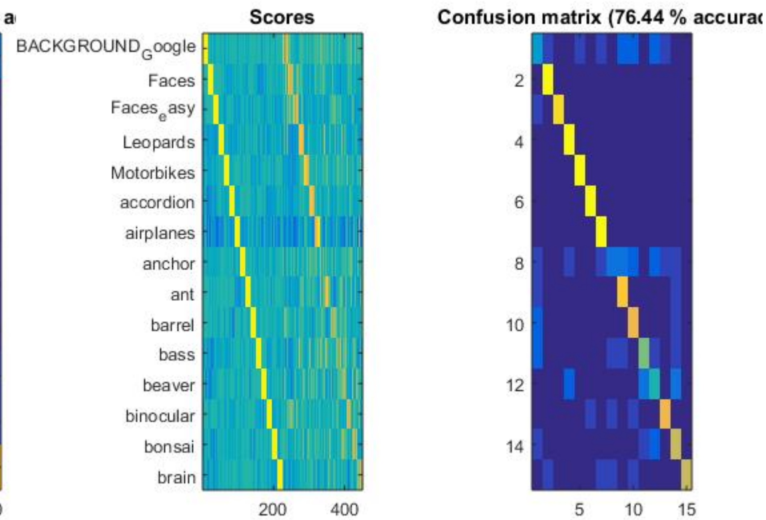

Figure 14. Confusion Matrix of the set 15 categories 15 images per category Sx=Sy=2 for Caltech101

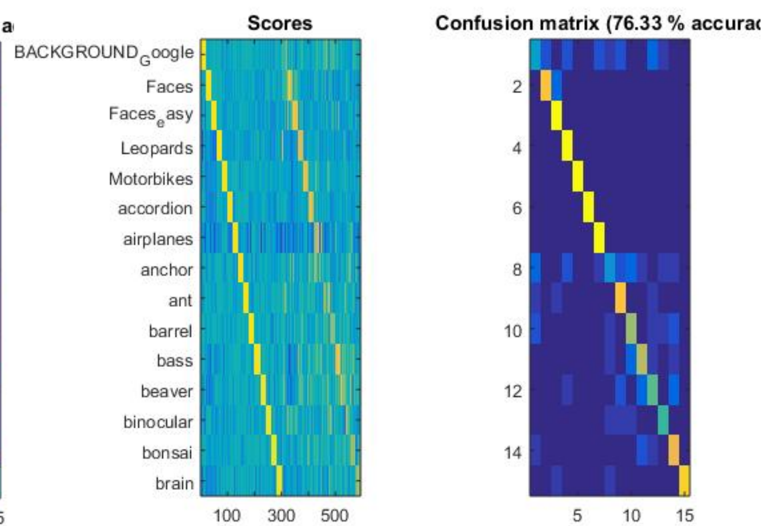

Figure 15. Confusion Matrix of the set 15 categories 20 images per category Sx=Sy=2 for Caltech101

[3] Pablo Arbelaez, "Computer Vision: Recognition 01", 2018.

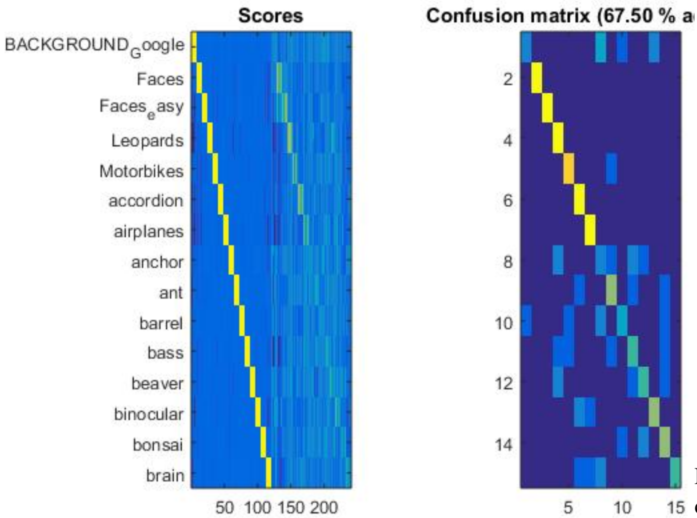

Figure 16. Confusion Matrix of the set 15 categories 8 images per category Sx=Sy=8 for Caltech101

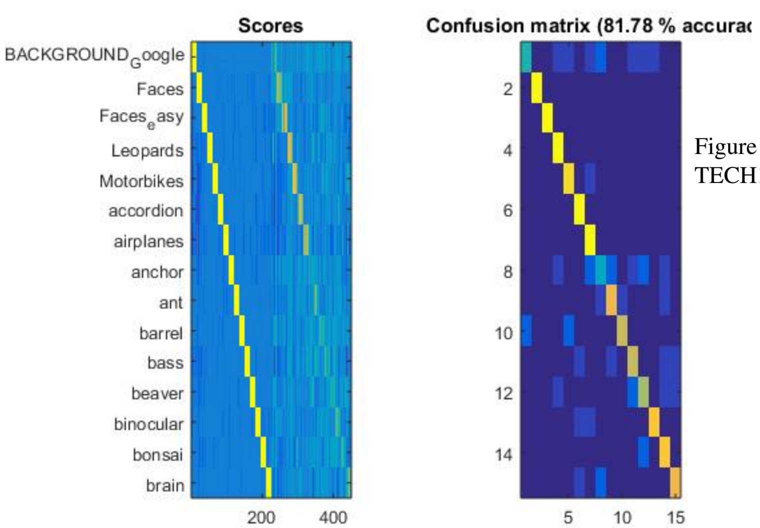

Figure 17. Confusion Matrix of the set 15 categories 15 images per category Sx=Sy=8 for Caltech101

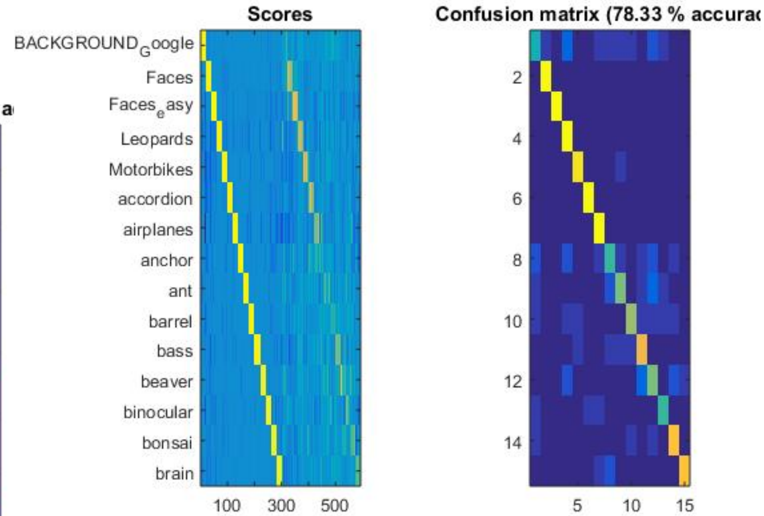

Figure 18. Confusion Matrix of the set 15 categories 20 images per category Sx=Sy=8 for Caltech101

Num Clases	Sx, Sy	Images	Accuracy
5	2	8	88%
		15	22.50%
		20	94%
	8	8	93%
		15	96.00%
		20	95%
10	2	8	76%
		15	79.33%
		20	83%
	8	8	78%
		15	84.67%
		20	86.00%
15	2	8	68.33%
		15	76.44%
		20	76.33%
	8	8	67.50%
		15	81.78%
		20	78.33%

Figure 19. Accuracy Table of the different scenarios for CALTECH101 database

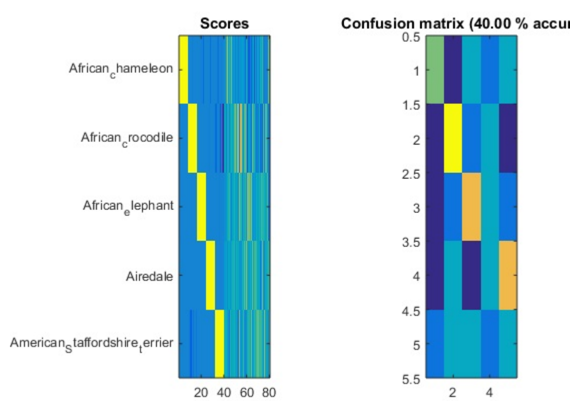

Figure 20. Confusion Matrix of the set 5 categories 8 images per category Sx=Sy=2 for ImageNet

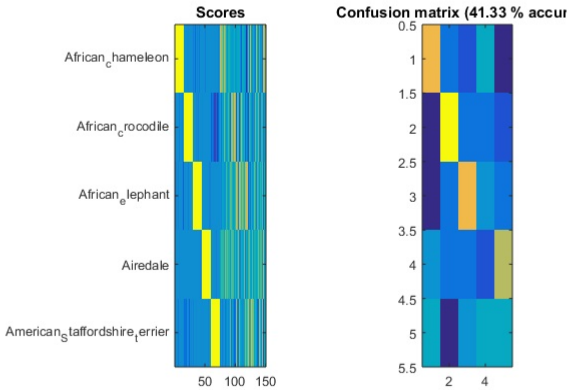

Figure 21. Confusion Matrix of the set 5 categories 15 images per category Sx=Sy=2 for ImageNet

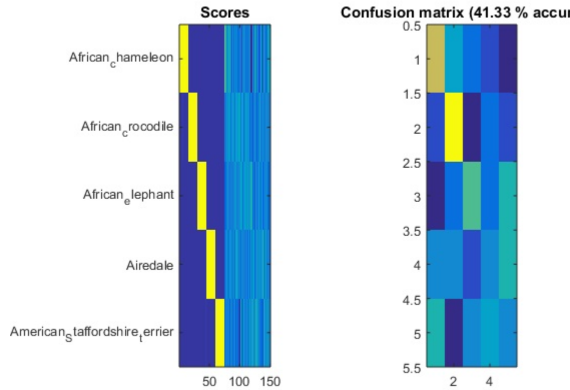

Figure 24. Confusion Matrix of the set 5 categories 15 images per category Sx=Sy=8 for ImageNet

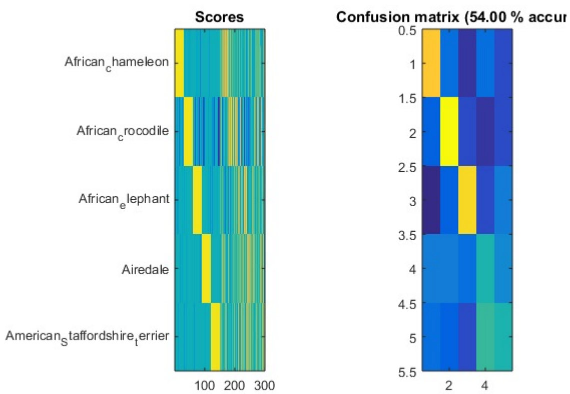

Figure 22. Confusion Matrix of the set 5 categories 30 images per category Sx=Sy=2 for ImageNet

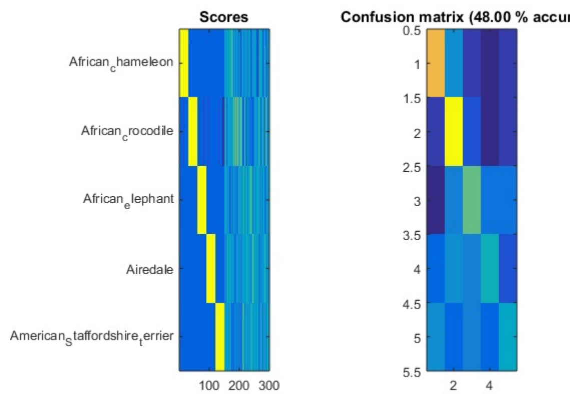

Figure 25. Confusion Matrix of the set 5 categories 30 images per category Sx=Sy=8 for ImageNet

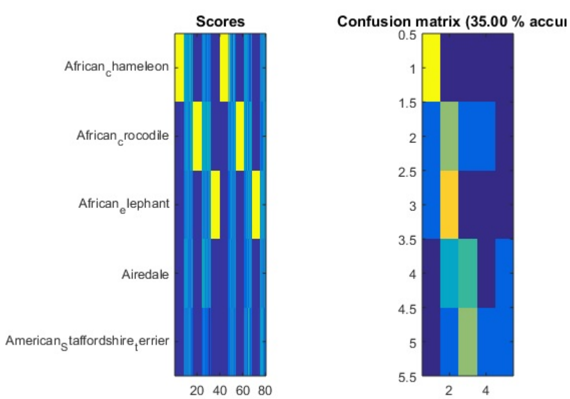

Figure 23. Confusion Matrix of the set 5 categories 8 images per category Sx=Sy=8 for ImageNet

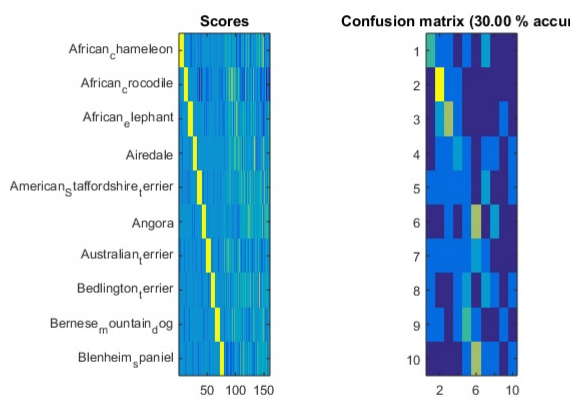

Figure 26. Confusion Matrix of the set 10 categories 8 images per category Sx=Sy=2 for ImageNet

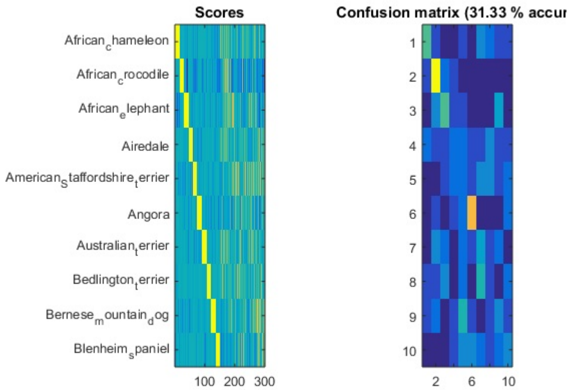

Figure 27. Confusion Matrix of the set 10 categories 15 images per category Sx=Sy=2 for ImageNet

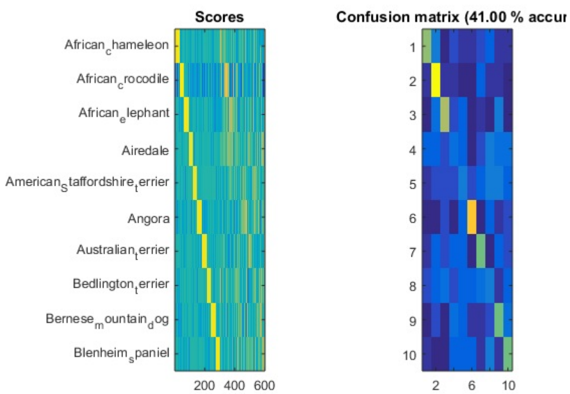

Figure 28. Confusion Matrix of the set 10 categories 30 images per category Sx=Sy=2 for ImageNet

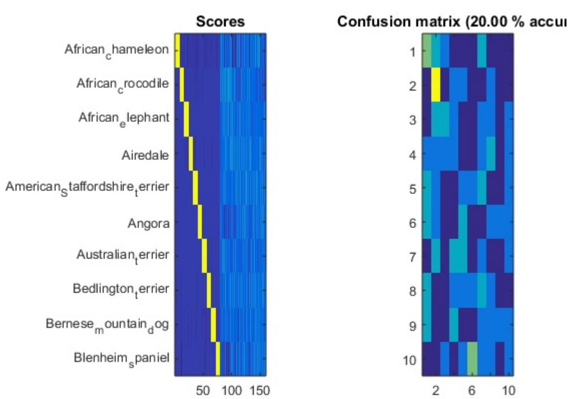

Figure 29. Confusion Matrix of the set 10 categories 8 images per category Sx=Sy=8 for ImageNet

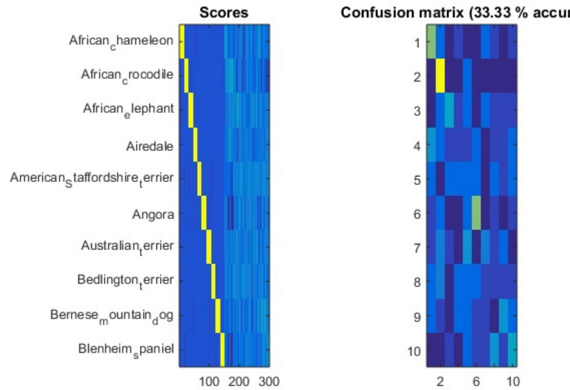

Figure 30. Confusion Matrix of the set 10 categories 15 images per category Sx=Sy=8 for ImageNet

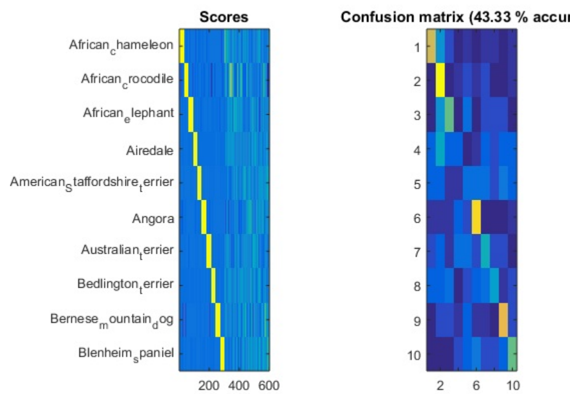

Figure 31. Confusion Matrix of the set 10 categories 30 images per category Sx=Sy=8 for ImageNet

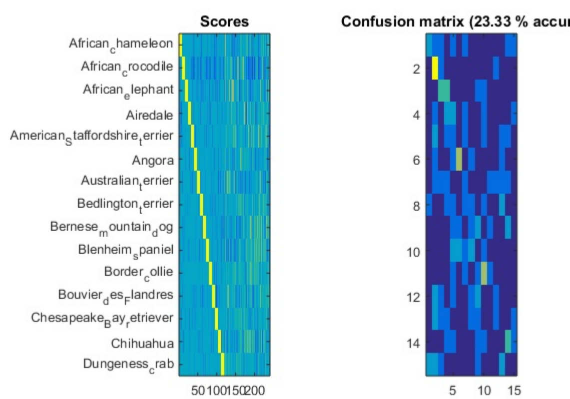

Figure 32. Confusion Matrix of the set 15 categories 8 images per category Sx=Sy=2 for ImageNet

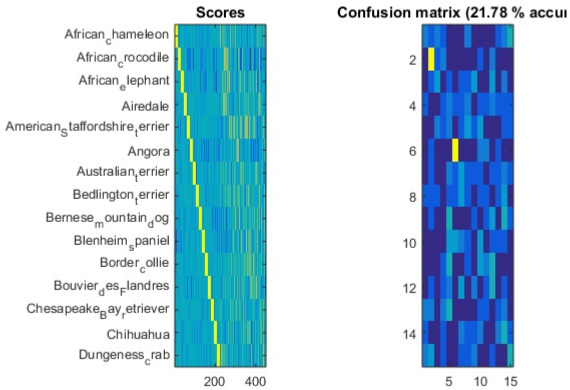

Figure 33. Confusion Matrix of the set 15 categories 15 images per category Sx=Sy=2 for ImageNet

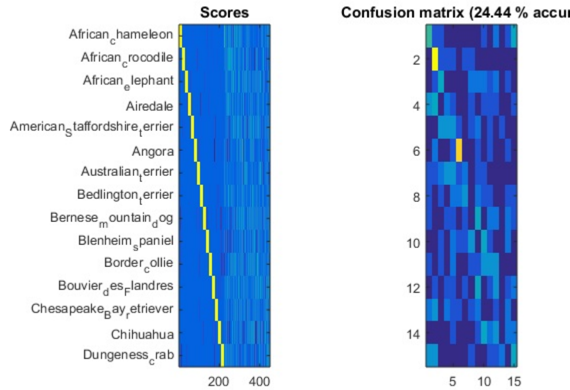

Figure 36. Confusion Matrix of the set 15 categories 15 images per category Sx=Sy=8 for ImageNet

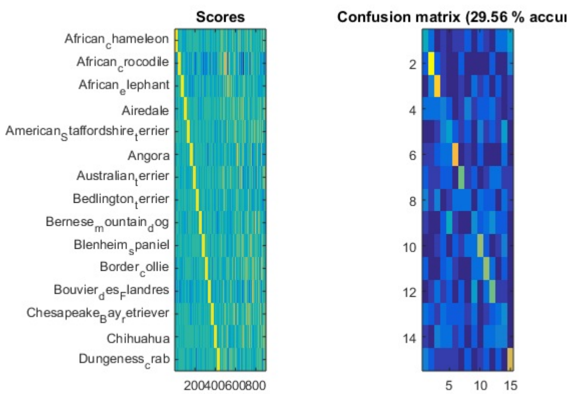

Figure 34. Confusion Matrix of the set 15 categories 30 images per category Sx=Sy=2 for ImageNet

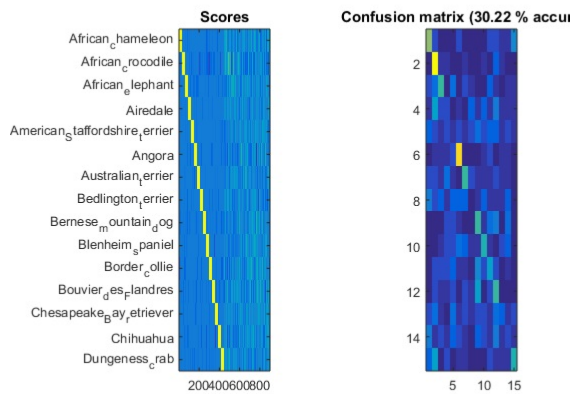

Figure 37. Confusion Matrix of the set 15 categories 30 images per category Sx=Sy=8 for ImageNet

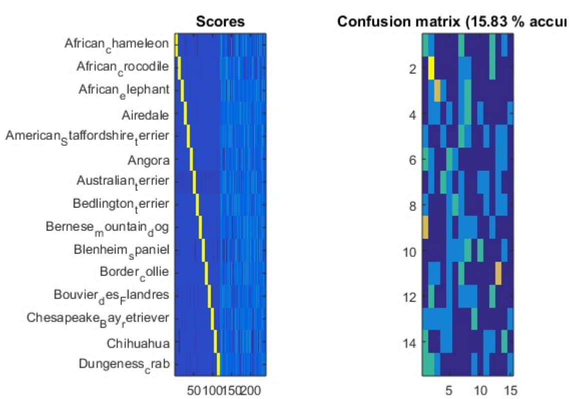

Figure 35. Confusion Matrix of the set 15 categories 8 images per category Sx=Sy=8 for ImageNet

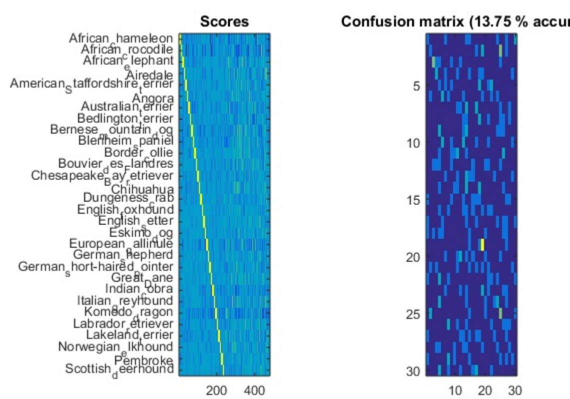

Figure 38. Confusion Matrix of the set 30 categories 8 images per category Sx=Sy=2 for ImageNet

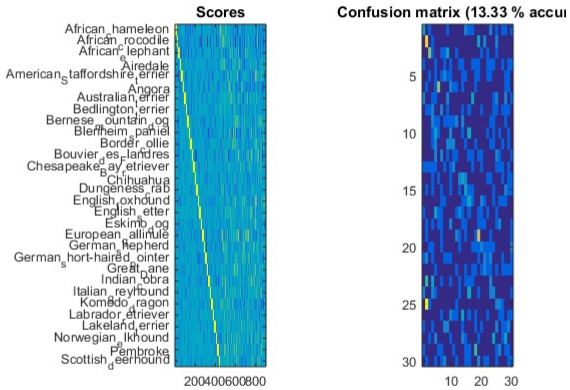

Figure 39. Confusion Matrix of the set 30 categories 15 images per category Sx=Sy=2 for ImageNet

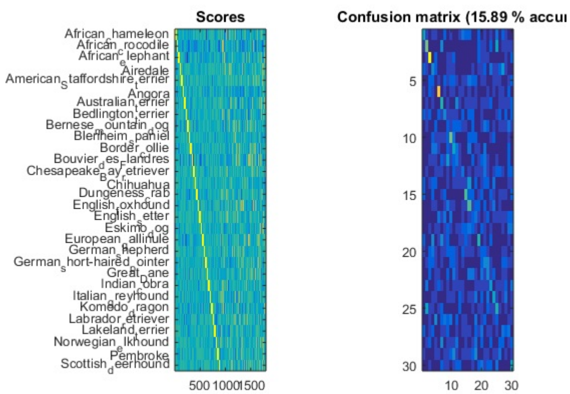

Figure 40. Confusion Matrix of the set 30 categories 30 images per category Sx=Sy=2 for ImageNet

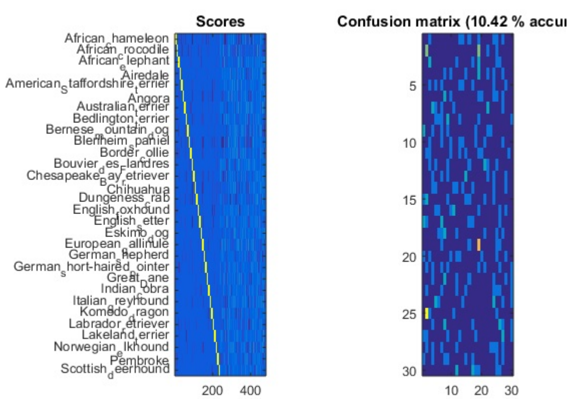

Figure 41. Confusion Matrix of the set 30 categories 8 images per category Sx=Sy=8 for ImageNet

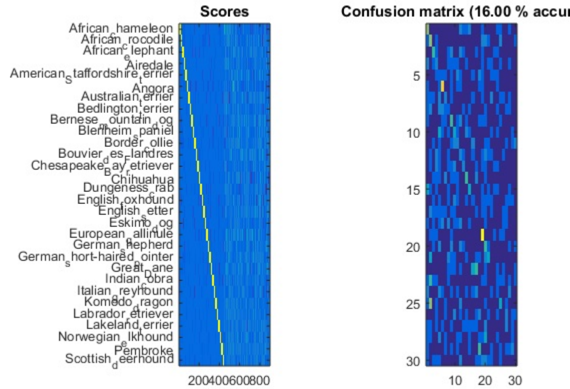

Figure 42. Confusion Matrix of the set 30 categories 15 images per category Sx=Sy=8 for ImageNet

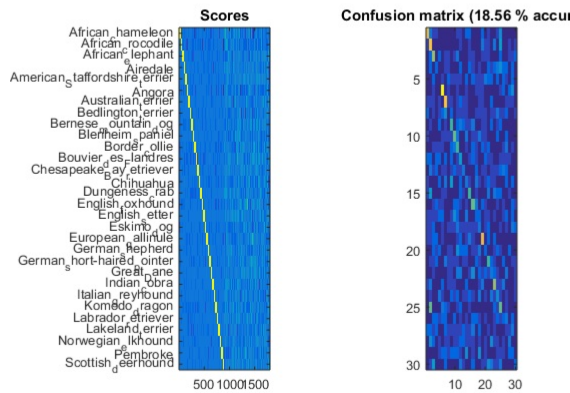

Figure 43. Confusion Matrix of the set 30 categories 30 images per category Sx=Sy=8 for ImageNet

Num Clases	Sx, Sy	Images	Accuracy
5	2	8	40%
		15	41.33%
		30	54%
	8	8	35%
		15	41.33%
		30	48%
10	2	8	30%
		15	31.33%
		30	41%
	8	8	20%
		15	33.33%
		30	43.33%
15	2	8	23.33%
		15	21.78%
		30	29.56%
	8	8	15.83%
		15	24.44%
		30	30.22%
30	2	8	13.75%
		15	13.33%
		30	15.89%
	8	8	10.42%
		15	16%
		30	18.56%

Figure 44. Accuracy Table of the different scenarios for ImageNet database