Master MVA
Object Recognition and Computer Vision
Assignment 2: Image Classification- 08/11/2016
Achari Berrada Youssef

1 Training and testing an Image Classifier.

QA- Sparse features for matching specific objects in images

1- Tiling is used in order to add spatial information to the bag of features and can be used to perform spatial pyramid representation in several levels of spatial resolution.

QB- Train a classifier for images containing aeroplanes

1- See figure 1.

2- See figure 2. The most relevant visual words appear on the background also. Because, the background of a airoport (airstrip) is also associated with an airplane.

QC- Classify the test images and assess the performance

1- The bias term is not needed because it is a constant term added to all the images.

$$\text{Scores} = w^T * \text{Histograms} + \text{bias}$$

QD- Learn a classifier for the other classes and assess its performance

1- See figures 3 and 4.

2- The first image ranked on the test images is a false positive one. On the precision-recall curve, this correspond to the point with Precision 0 and recall 0.

QE- Vary the image representation

1- Benefits of spatial tiling: The performance is lower without spatial tiling. See figure 5.

2- Benefits of descriptor normalization: See table 2.

3- Self similarity of an L2 normalized histogram:

$$K(h,h) = \sum_i h_i^2 = 1$$

Hence, with Cauchy Schwarz inequality, we have for two different L2 normalized histograms:

$$K(h,h') \leq \sqrt{K(h,h).K(h',h')} = 1$$

And this is not true for L1 normalized or unnormalized histograms.

4- The classification performance is better when the histograms are L2 normalized.

QF- Vary the classifier: Hellinger Kernel:

1- The histograms should be L1 normalized. And this normalization should be applied before taken the square root.

2- L1 normalization: $K_H(h,h) = \sum_i \sqrt{h_i}.\sqrt{h_i} = \sum_i h_i = 1$
Square root: $\tilde{h} \leftarrow \sqrt{h}$
Self similarity:

$$K(\tilde{h}, \tilde{h}) = \sum_i \tilde{h}_i^2 = \sum_i h_i = 1$$

And

$$K(\tilde{h}, \tilde{h}') = \sum_i \tilde{h}_i \tilde{h}'_i = \sum_i \sqrt{h_i.h'_i} = K_H(h, h')$$

So it is equivalent to the Hellinger kernel.

3- Linear classifier is tractable and easy to solve.

4- See table 3. It is clear that the L1-Normalization is the best fit for the Kellinger Kernel. And we should do it before taking the square root of the histogram vector.

QG- Vary the number of training images

1- See figure 6.

2- No, in some cases, adding more training images could be useful because the slope is positive. When the slop attain 0, we can say that there is saturation. However, as the curve is convex, adding more training images will not improve significantly the performance.

2 Training an Image Classifier for Retrieval using Internet image search.

Stage P The performance improve when 10 images are used. See figure 7 and 8 to see the best ranked images in each class and for each number of training images used. To see the images used for training, see 9 and 10.

Table 1: Performance for Horse and Car classes.

	Training Dataset	5	10
Horses	mAP	0.10	0.16
	Precision at rank 36	3/36	6/36
Cars	mAP	0.49	0.61
	Precision at rank 36	21/36	33/36

3 Advanced Encoding methods:

QH- First order methods.

1- If we set K_{vlad} as the number of centroids in the VLAD representation, each centroids is in D dimension. And if we consider K_{BoVW} as the number of visual words in the BoVW representation. Then to obtain descriptors with the same dimension, we have the relation:

$$K_{vlad} \times D = K_{BoVW}$$

2- See table 4.

QI- Second order methods

1- See table 5.

2- The second order method is more efficient. In terms of memory, in the general case, Fisher Vector requires $K \times D + K \times D \times D$ memory if we want to store the entire covariance vector and it requires only $2K \times D$ if we suppose that the covariance matrix is diagonal. Also computation time is higher if we want to compute GMM components as it requires more operation for each component.

4 Figures

4.1 Section 1

- **Stage B**

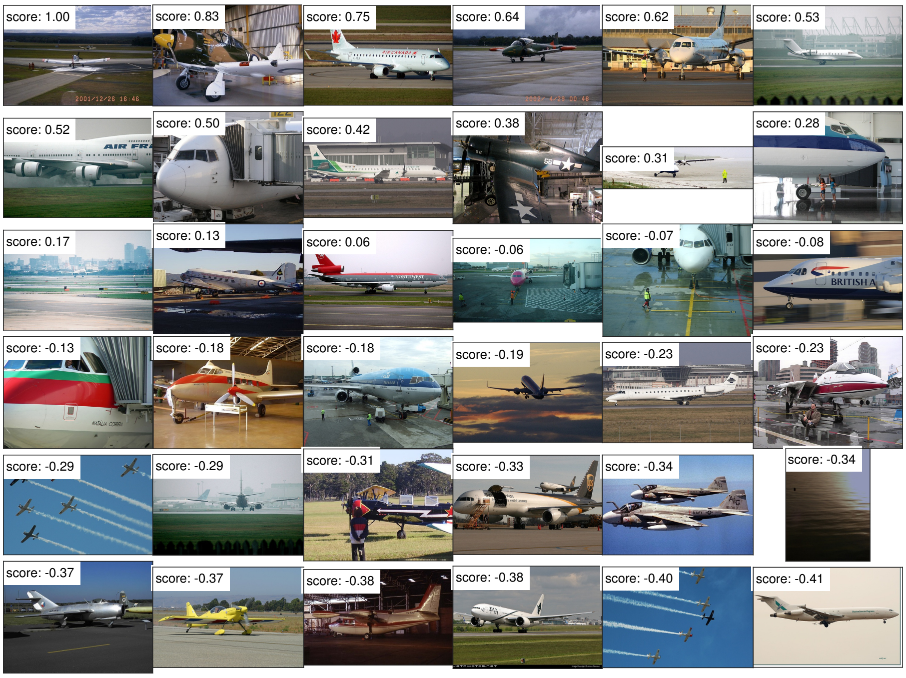


Figure 1: Ranked Images.

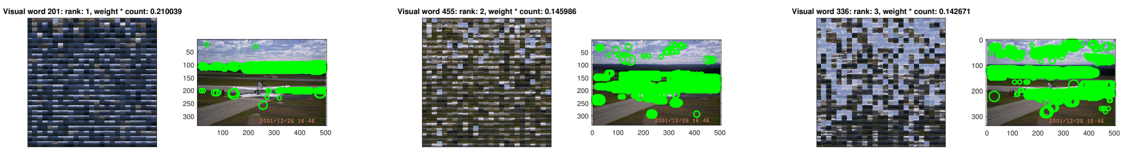
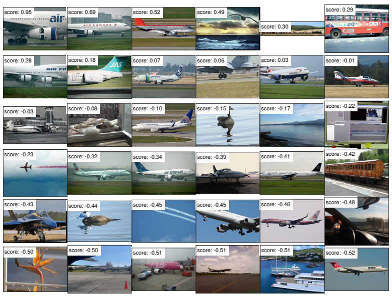
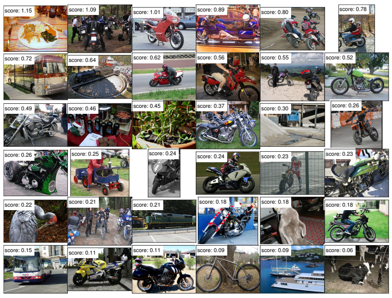


Figure 2: Ranked Visual Words on the best ranked image.

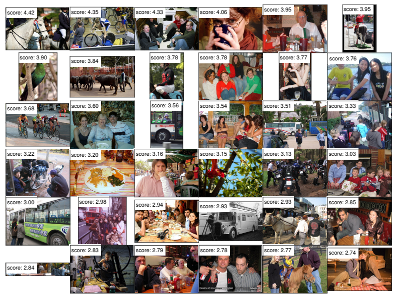
- **Stage D**



(a) Aeroplane,
First false positive = 4.

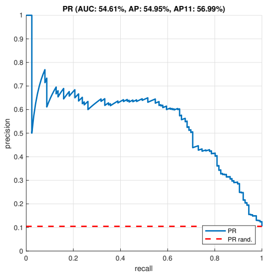


(b) Motorbike,
First false positive = 1.

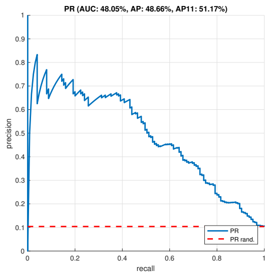


(c) Person,
First false positive = 7.

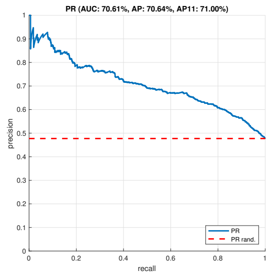
Figure 3: First Ranked images on the test set.



(a) Aeroplane AP = 0.55



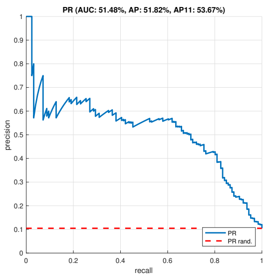
(b) Motorbike AP = 0.48



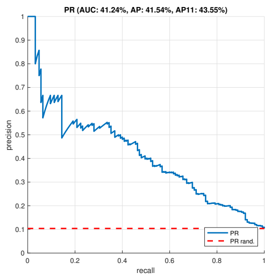
(c) Person AP = 0.71

Figure 4: Precision-Recall curves on the test images and their AP scores.

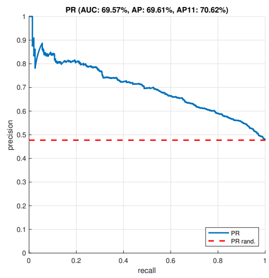
- **Stage E**
  QE1.



(a) Aeroplane AP = 0.51



(b) Motorbike AP = 0.41



(c) Person AP = 0.69

Figure 5: Precision-Recall curves on the test images and their AP scores with histograms without spatial tiling.

QE2.

Table 2: Benefits of descriptor normalization (mAP and Precision at rank-36).

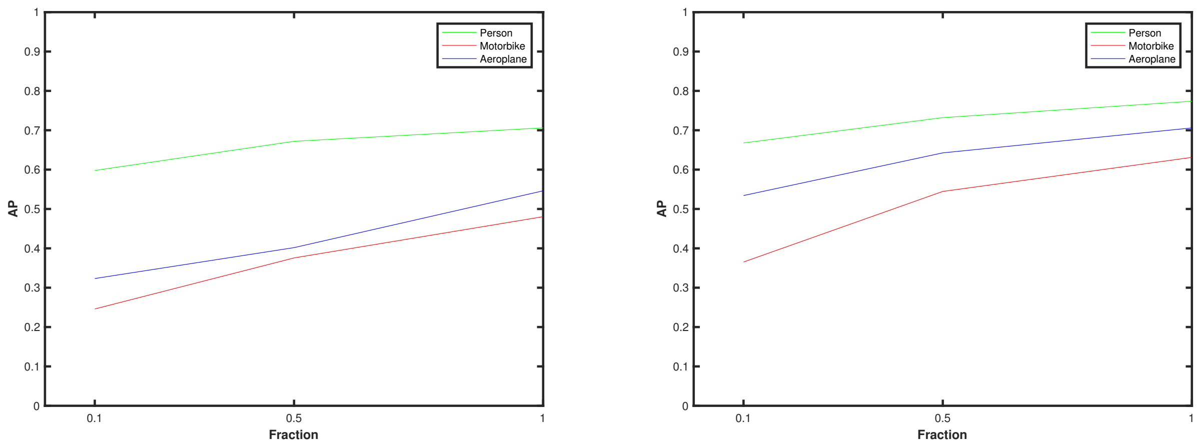
	Aeroplane	Motorbike	Person
L2-Normalization	0.55 24/36	0.48 24/36	0.71 33/36
No Normalization	0.62 29/36	0.48 27/36	0.67 28/36
L1-Normalization	0.52 31/36	0.25 14/36	0.56 20/36

- **Stage F**

Table 3: Normalization for Hellinger Kernel (mAP and Precision at rank-36).

	Aeroplane	Motorbike	Person
L1-Normalization	0.71 31/36	0.63 29/36	0.77 33/36
No Normalization	0.66 31/36	0.55 29/36	0.70 31/36
L2-Normalization	0.64 28/36	0.53 26/36	0.71 34/36

- **Stage G**



(a) Linear Kernel

(b) Kellinger Kernel

Figure 6: Varying the number of training images.

4.2 Section 2

- **Stage H:** Vlad representation.

Table 4: Vector of locally aggregated descriptors performance.

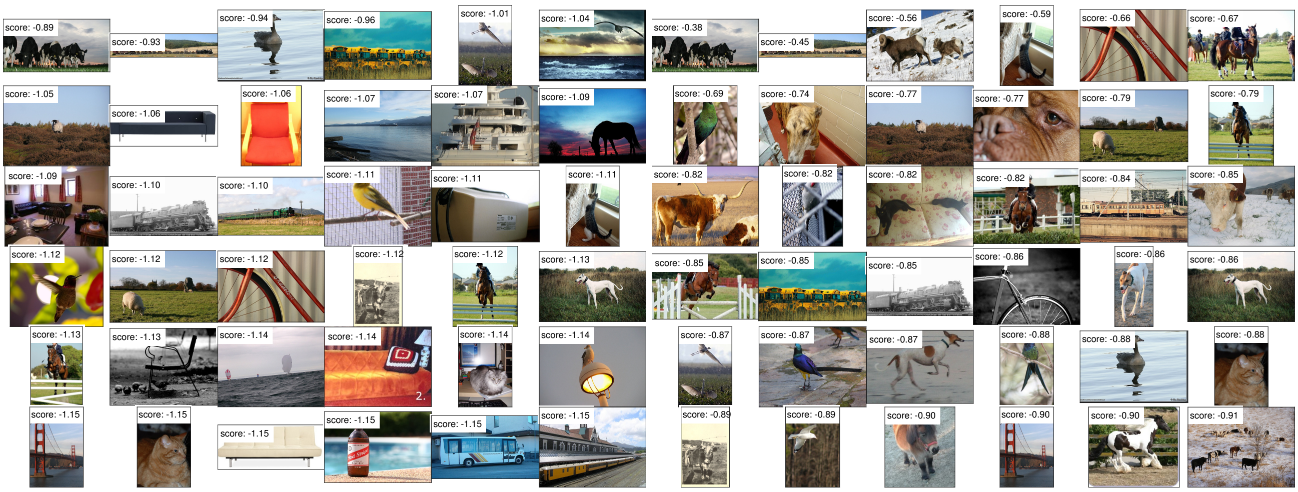
VLAD	Aeroplane	Motorbike	Person
Linear	0.75	0.69	0.76
	36/36	34/36	36/36
Hellinger	0.75	0.75	0.79
	35/36	36/36	35/36

- **Stage I:** FV representation.

Table 5: Fisher vector performance.

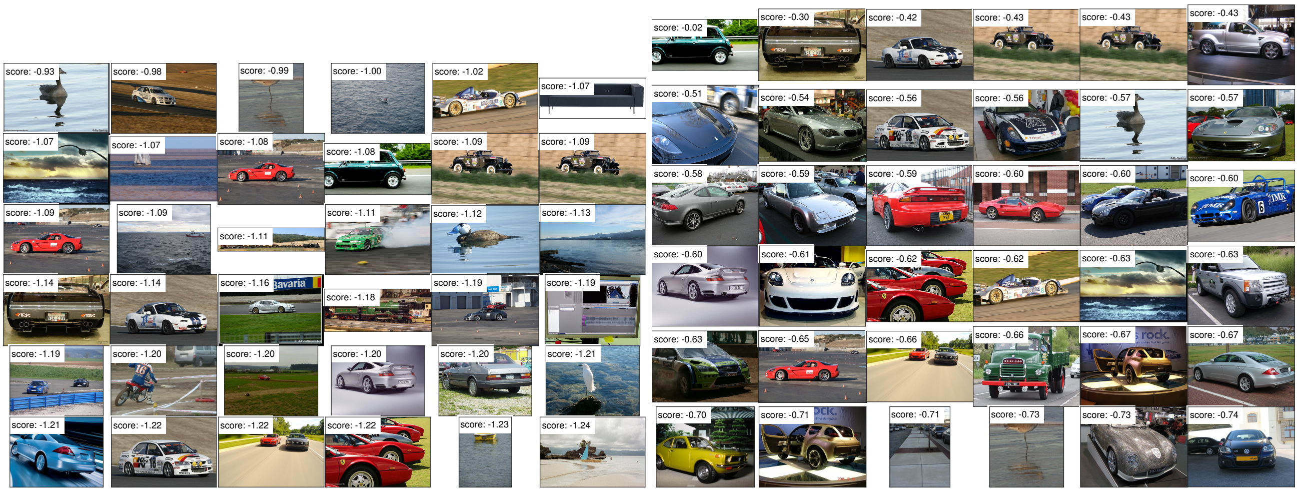
FV	Aeroplane	Motorbike	Person
Linear	0.70	0.73	0.77
	32/36	34/36	33/36
Hellinger	0.78	0.81	0.82
	35/36	36/36	35/36

- **Stage P**



(a) 5 Training Images. (b) 10 Training Images.

Figure 7: Top Ranked images for the class Horse.



(a) 5 Training Images. (b) 10 Training Images.

Figure 8: Top Ranked images for the class Car.

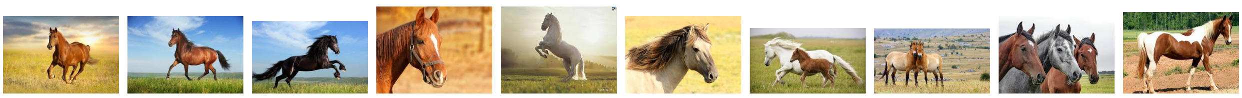


Figure 9: Training Images for horse class.

Figure 10: Training Images for car class.