[M2, MVA]

Object recognition and computer vision

Maha ELBAYAD
maha.elbayad@student.ecp.fr

Assignement 2

November 4, 2015

Part 1: Training and testing an Image Classifier

Stage A

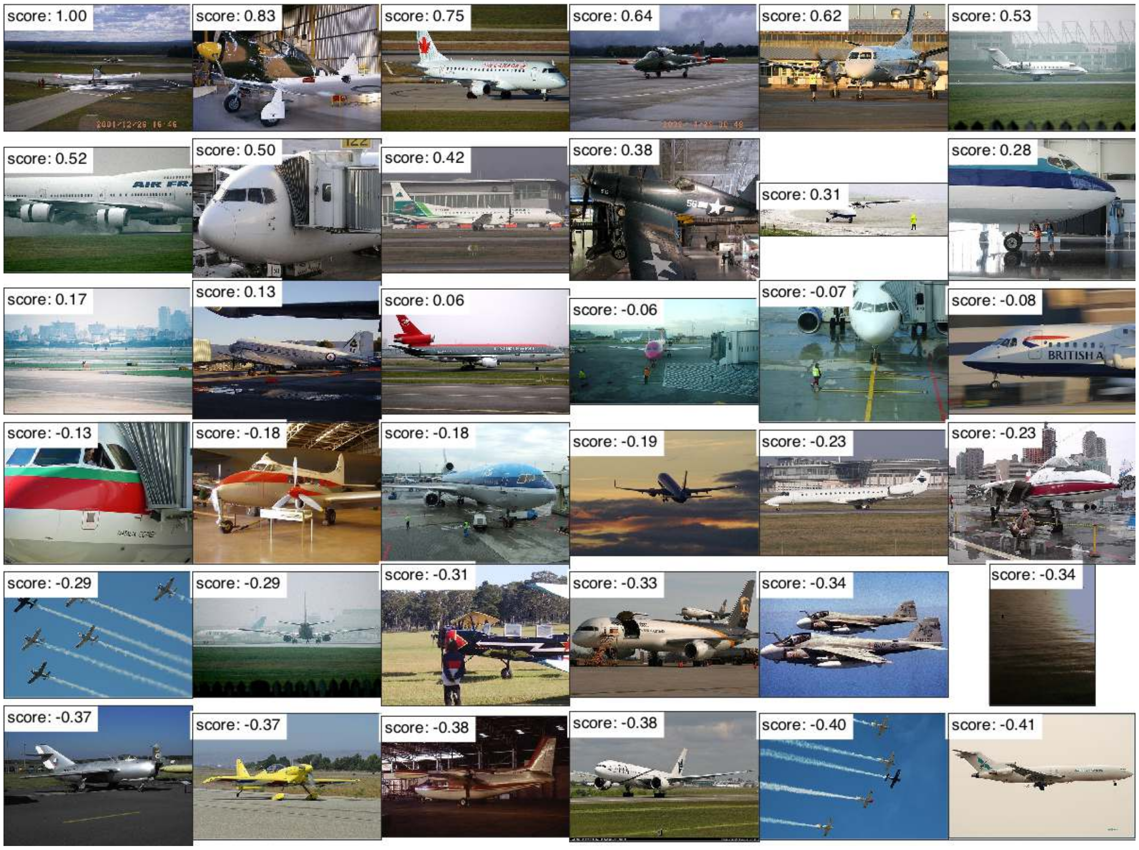
Q A1

The concatened histograms over spatial tiles represent a sort of semi-local features, in fact, if the histograms allow for more stability against image deformation, tiling neutralize the potential loss of spatial information.

Stage B

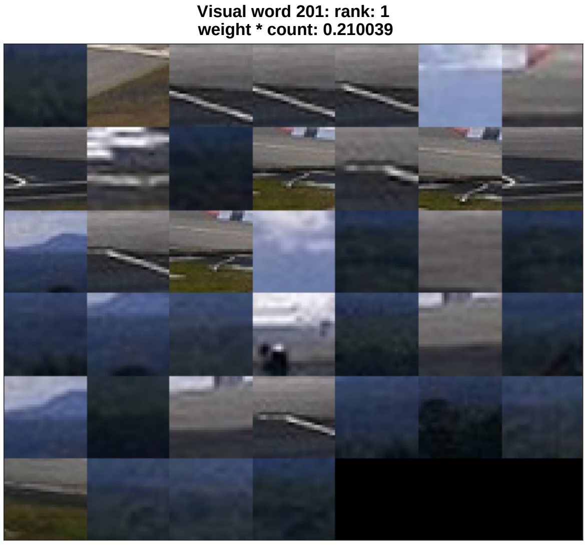
Q B1

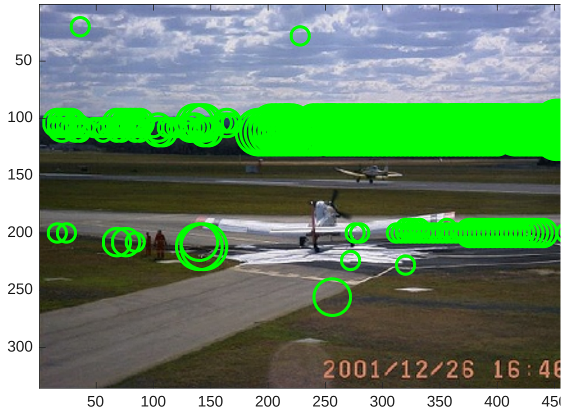
Top ranked images from the training set:

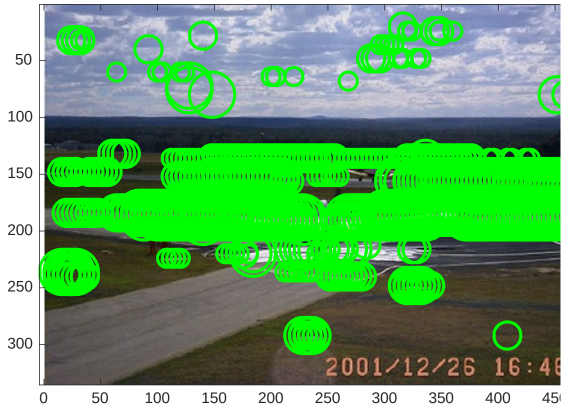


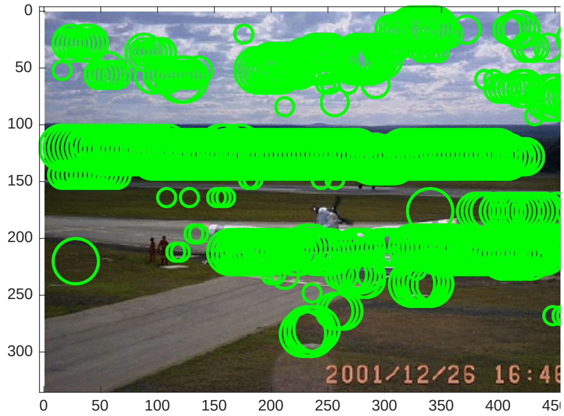
Q B2

Patches of the most relevant visual words on the top scored training image are shown in the figures below. As expected, the patches are for recurrent background elements {sky, horizon, mountain, clouds...} which are not exclusive to the positively labelled training images.









Stage C

Q C1

Since we're only comparing the test scores, the bias term would only translate them by a constant, not affecting the order.

Stage D

Q D1

The AP performance for the three categories is as exepected seeing the distibution of the datasets; in fact,

Table 1: Test performances

Category	Testing classes (+)/(-)	Test AP
Aeroplanes	126/1077	0.55
Motorbikes	125/1077	0.48
Person	983/1077	0.71

for the *person* category, even a random classifier can perform well as the training/testing sets are balanced compared to the unbalanced sets of *motorbike* and *aeroplane*.

Figure 1: Top ranked images - category *aeroplane*

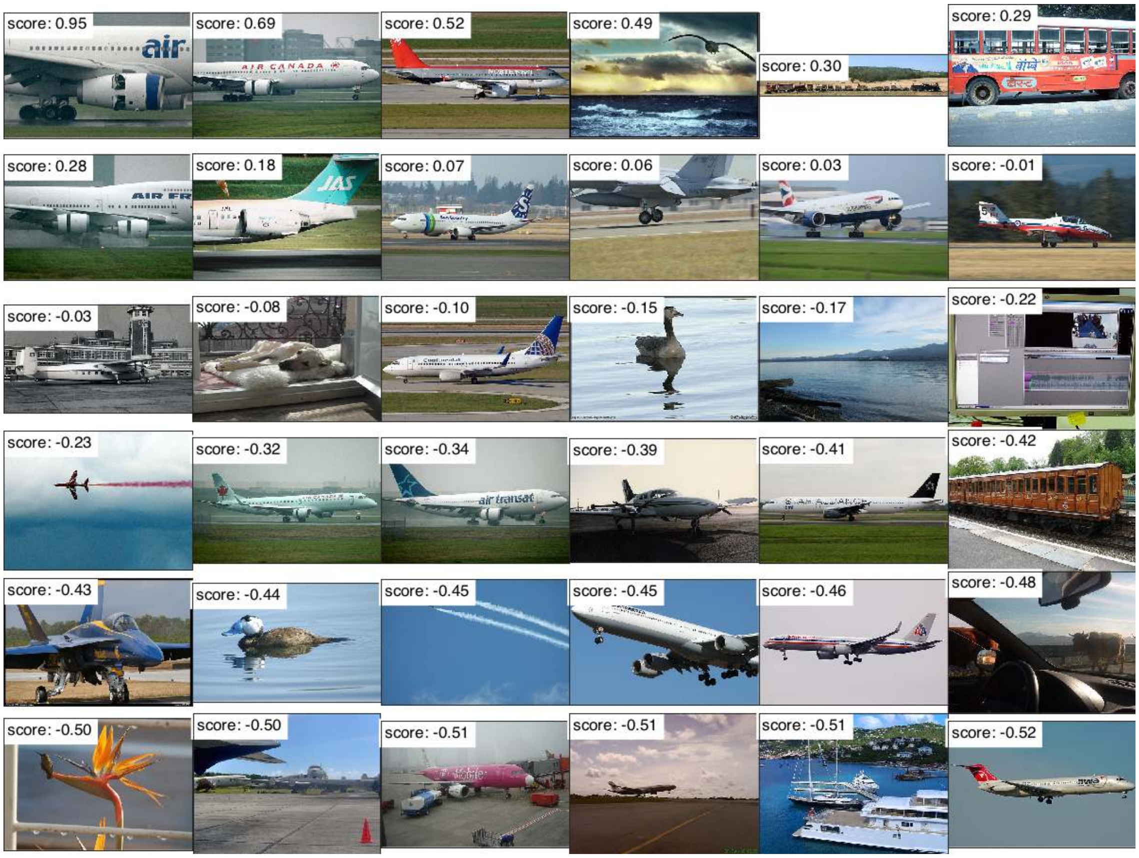


Figure 2: Top ranked images - category *motorbike*

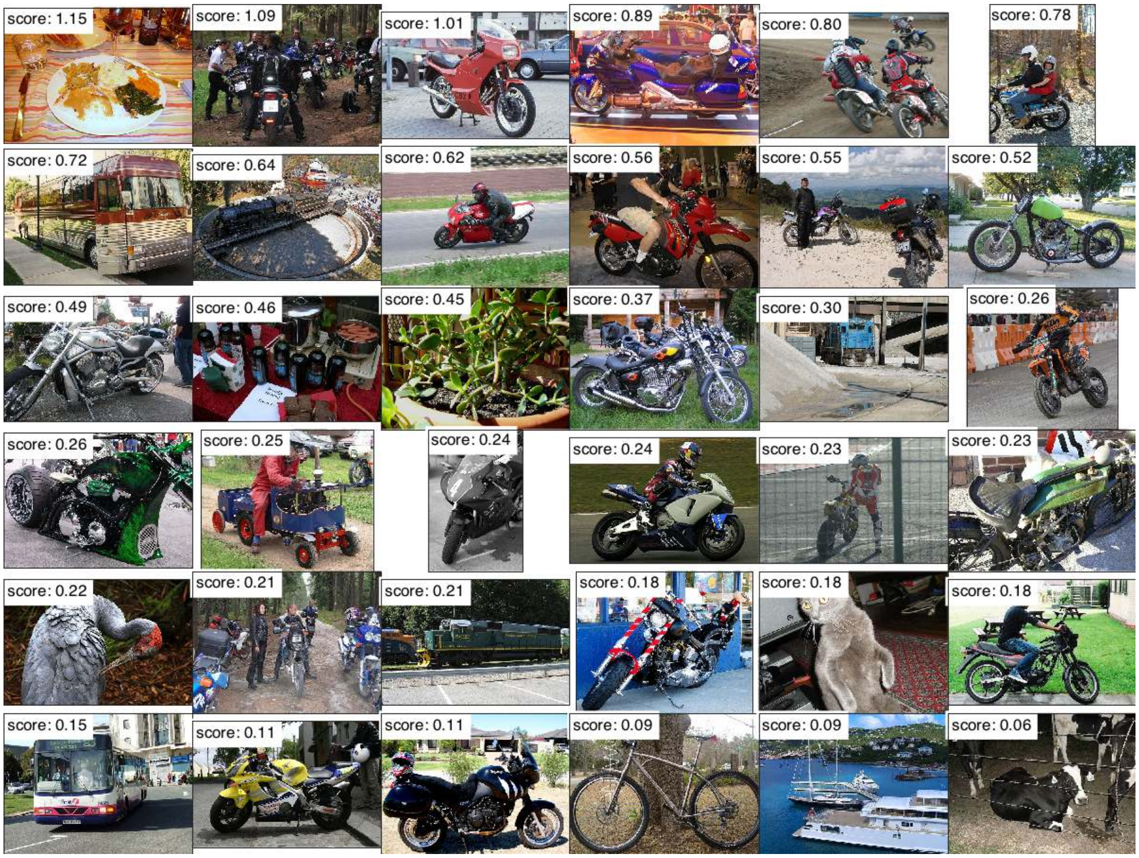


Figure 3: Top ranked images - category *person*

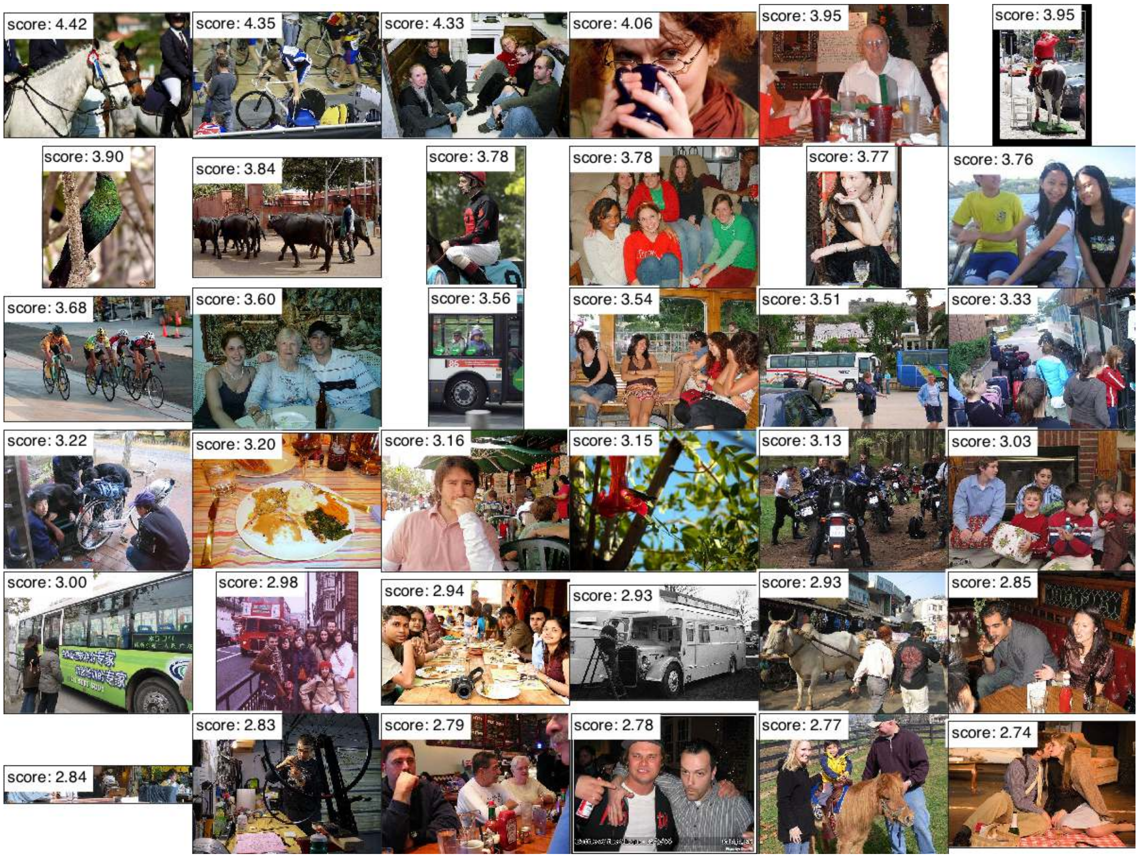
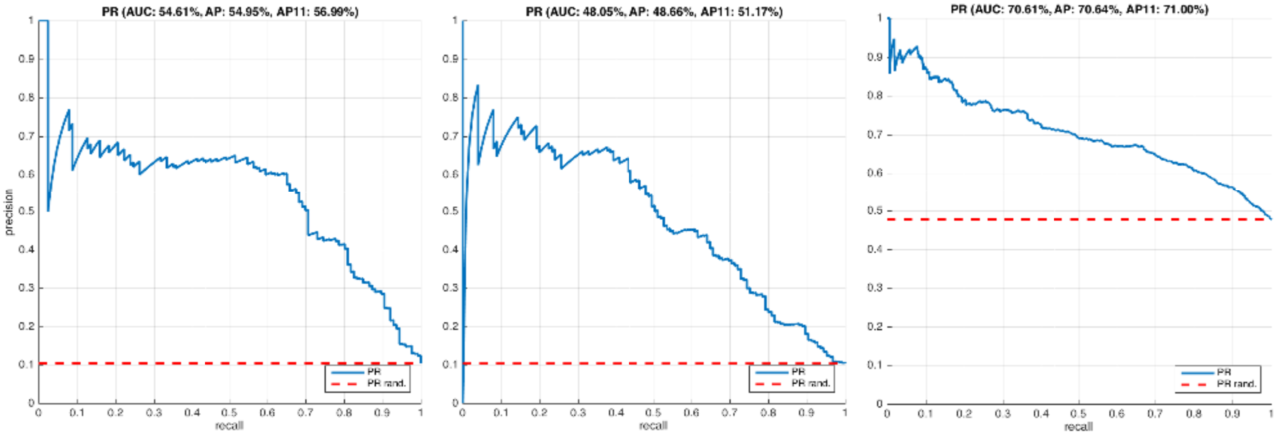


Figure 4: PR curves


(a) Aeroplane

(b) Motorbike

(c) Person

Q D2

The rank of the first false positive is **1**.
This false positive corresponds to the second point of the PR-curve at $(r = 0, p = 0)$ after the assumed starting point $(r = 0, p = 1)$

Stage E

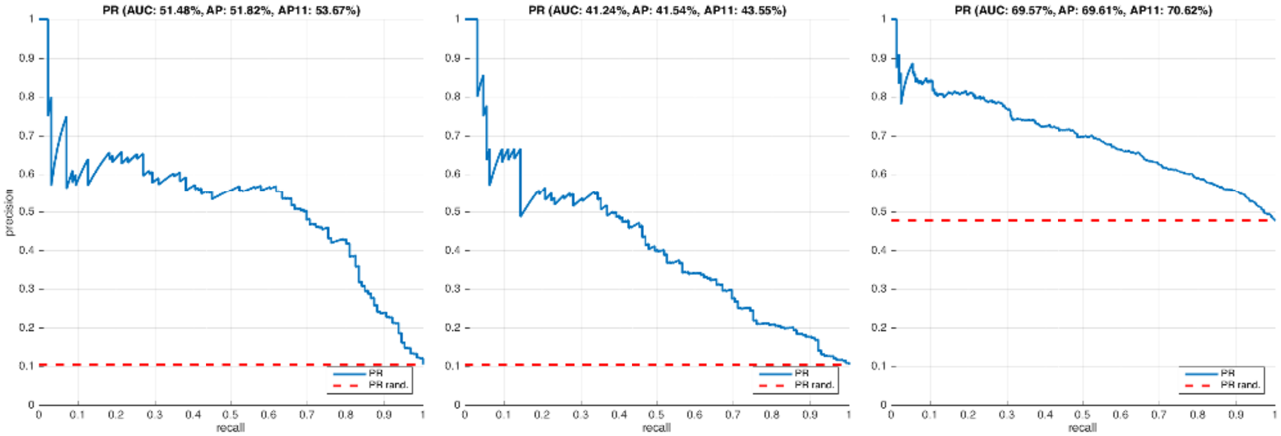
Q E1

The performance is deteriorating while the PR-curves keep on approximately the same shape. This lower performance is expected as we do not learn from on any global or semi-global feature.

Table 2: Test performances without spatial tiling

Category	w/o tiling AP	w/ tiling
Aeroplanes	0.51	0.55
Motorbikes	0.41	0.48
Person	0.70	0.71

Figure 5: PR curves



(a) Aeroplane (b) Motorbike (c) Person

Q E2

We change the norm from $L2$ to $L1$ then to no norm at all. The most performant norm for SVM classification on the given problem is $L2$.

Table 3: Impact of normalization

Category	$L2$	$L1$	$\varnothing$
Aeroplanes	0.55	0.52	0.62
Motorbikes	0.48	0.25	0.48
Person	0.71	0.56	0.67

Q E3

Using the linear kernel $K(h,h') = \sum_{i=1}^{d} h_i h'_i$ and for BoVW histograms h and h' that are $L2$ normalized:

$$K(h,h) = \|h\|_2^2 = 1 \tag{1}$$

And via C.S inequality $|K(h,h')| \leq \|h\|_2.\|h'\|_2 = 1$.
The equality 1 doesn't hold for $L1-$normalized or unnormalized histograms.

Q E4

With $L2$ normalization, the dot products used in SVM $K(h,h')$ are equal to $cosine(h,h')$ i.e they reflect similarities between histograms.

Stage F

Q F1

For a histogram h, the new histogram adapted for the Hellinger kernel would be $h := \sqrt{h}$. After this transformation we should $L2-$normalize the new h

Q F2

This procedure is equivalent to using the Hellinger kernel as we're explicitly mapping the features to the new space associated to the kernel instead of using the kernel trick.

Q F3

One advantage of using a linear classifier would be the ease to interpret the linear decision boundary. And from a computation point of view, solving the optimisation problem for a linear kernel is less costly.

Q F4

We compare the AP- test performance of the post-normalization (after root-squaring) to this of the alternative pre-normalization:

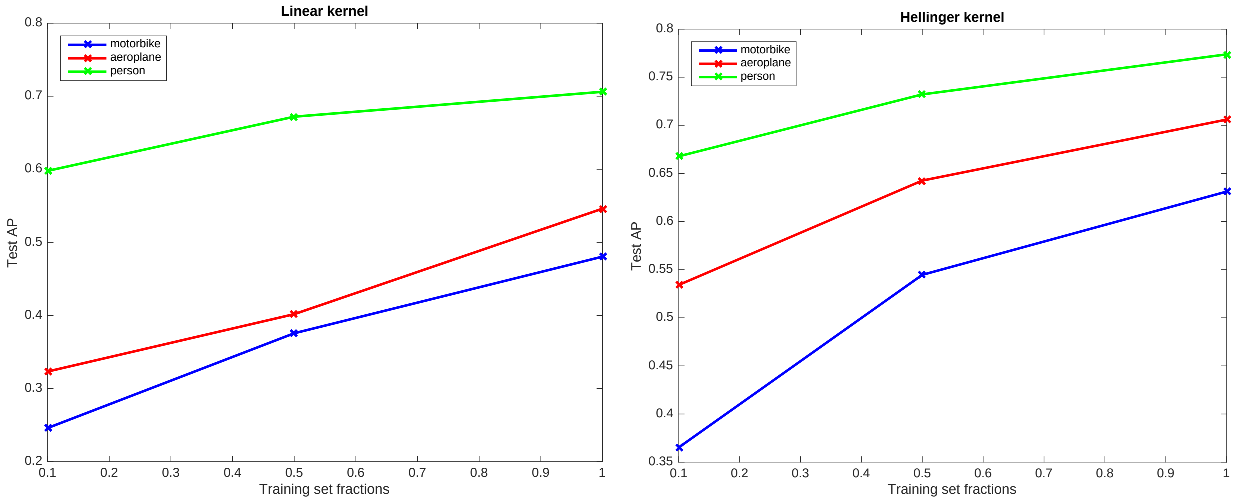
Table 4: AP - Test performances

Category	post-normalization ✓	pre-normalization
Aeroplanes	0.71	0.64
Motorbikes	0.63	0.53
Person	0.77	0.71

Stage G

Q G1

We assess our algorithm performance with regard to the size of the training set. The curves of AP on test set (same size) are shown on the figures below for the three categories using the linear kernel and the Hellinger kernel.



(d) Linear kernel

(e) Hellinger kernel

Q G2

From the tendancy of the curves on both kernels for all three categories, the performance doesn't seem near saturation. Thus we can assume that growing the training set wouldn't hurt.

Part 2: Training an Image Classifier for Retrieval using Internet image search

Q P2.1

For our *horses* classifier the 10 training images are shown below. The AP on test set with 5 then 10 images indicates that we can improve the classifier performance with more images.

Figure 6: First 5

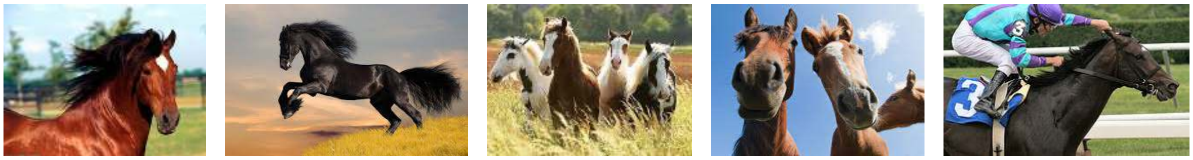
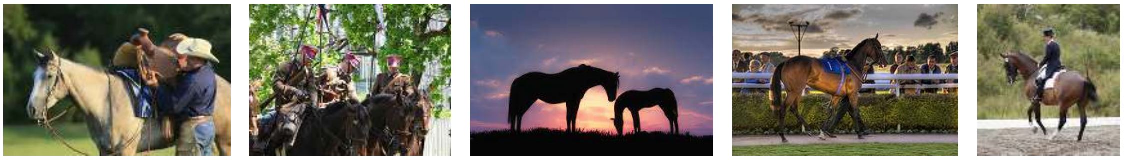


Figure 7: Second 5

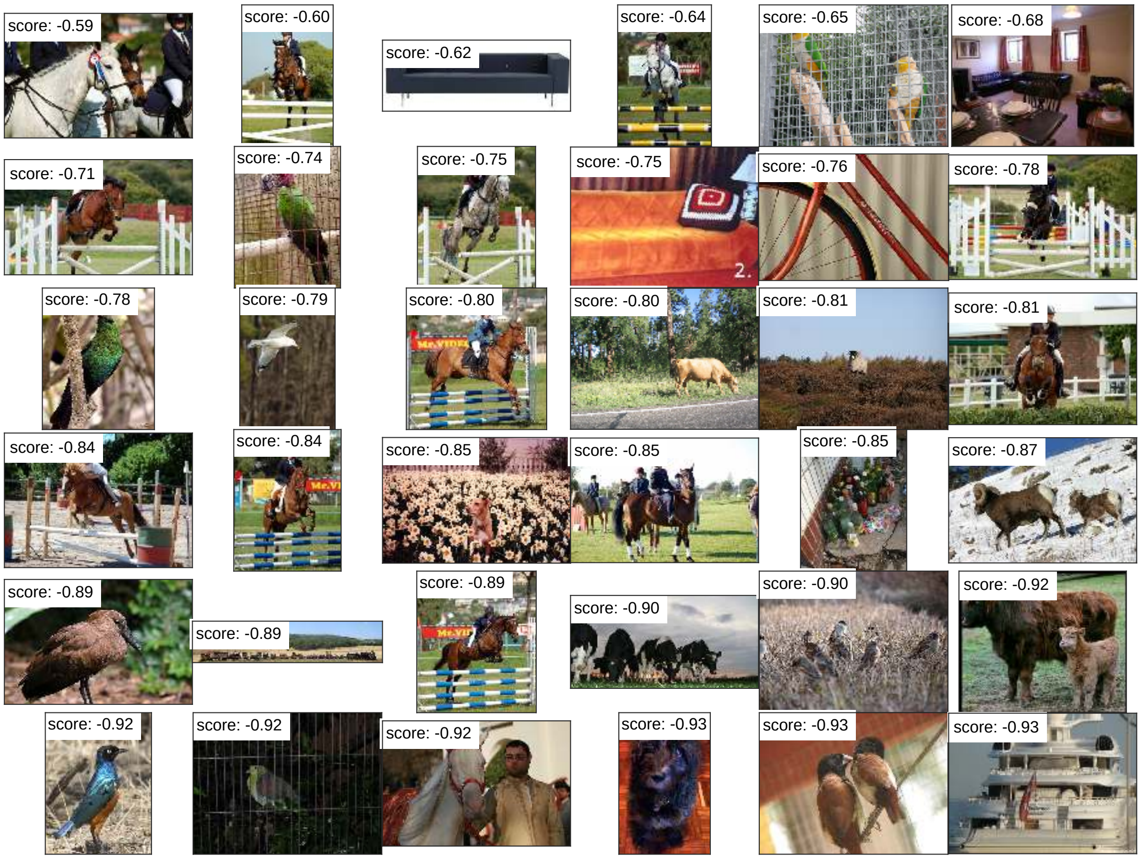


Training set size	Precision at rank-36	test AP
5	10	0.16
10	9	0.22
15	10	0.23
20	13	0.27
33	12	0.23

Q P2.2

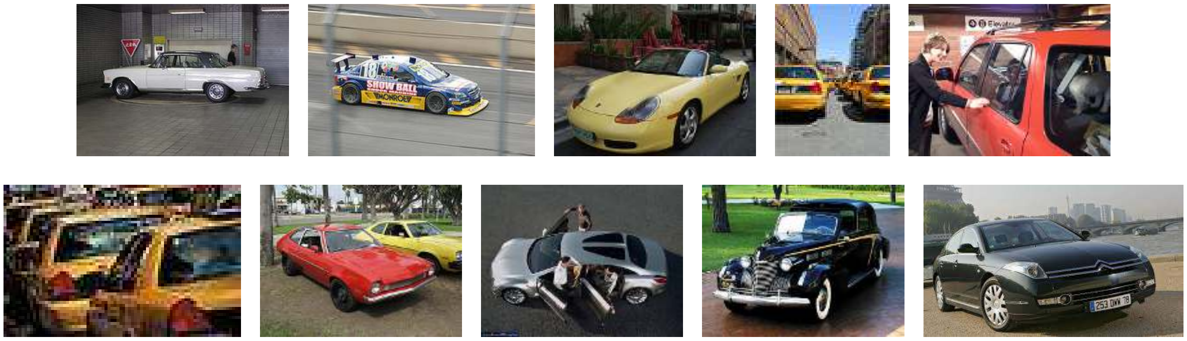
The best performane we achieved on the *horse* classifier was of Precision at rank-36=13 with 20 images at the training set. The top 36 scored images are shown below

Figure 8: TOP 36 scored images on the test set - *horse* classifier -20 images



For the *cars* category, some of the training images are shown in figure 9.

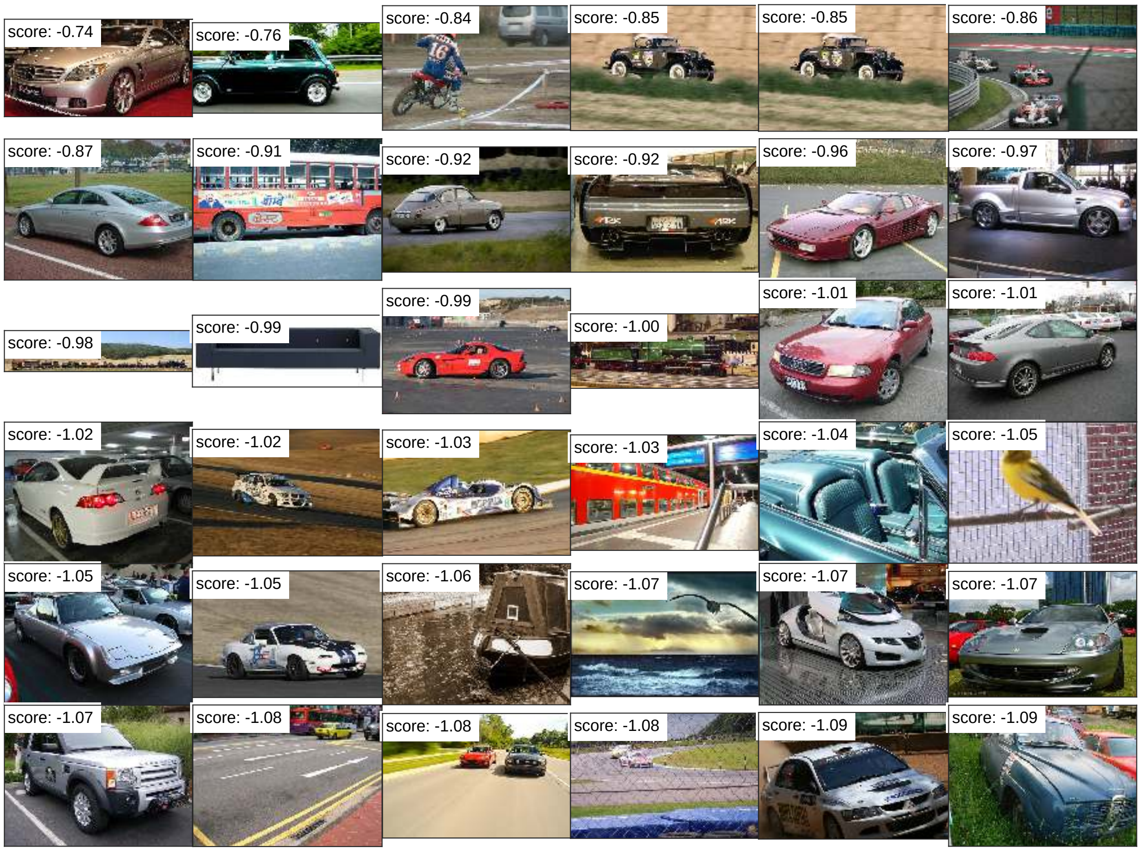
Figure 9: Training set - Cars



The task seems easier with cars, as the precision at rank 36 is promising. The top scored images are shown below. We can see from the top ranked images of both categories, that *horses* are more difficult to categorize as they're similar to other animal categories {cows, dogs, birds...} or even leather objects. On the other hand, *cars* have distinctive geometry and the misclassifications are not surprising.

Training set size	Precision at rank-36	test AP
5	20	0.44
10	28	0.47
15	26	0.47
20	27	0.48

Figure 10: TOP 36 scored images on the test set - *car* classifier -10 images



Stage H

Q H.1

The VLAD vector has dimension $K_{VLAD} \times D$ where D is the dimension of the SIFT descriptors and K the number of clusters. The BoVW vector has dimension K_{BoVW} (bins $\sim$ clusters) if the spatial tiling is omitted, otherwise it's $K_{BoVW} \times |tiles|$. For identical dimensions we use $K_{BoVW} = \dfrac{K_{VLAD} \times D}{|tiles|}$

Q H.2

We train the SVM classifier on the three different categories with the linear kernel then the Hellinger kernel.

Table 5: AP on test sets: VLAD vs. BoVW

		aeroplane	motorbike	person
Linear	BoVW	0.55	0.48	0.71
	VLAD	0.75	0.69	0.76
Hellinger	BoVW	0.71	0.63	0.77
	VLAD	0.07	0.07	0.55

The VLAD encoding outperfom BoVW with linear kernel SVM classification, nonetheless, the Hellinger kernel is obviously a bad choice of kernel on the VLAD space.

Stage I

Q I.1

In this part, we compare the FV encoding to the previous two:

Table 6: AP on test sets: VLAD vs. BoVW

		aeroplane	motorbike	person
Linear	BoVW	0.55	0.48	0.71
	VLAD	0.75	0.69	0.76
	FV	0.70	0.73	0.77
Hellinger	BoVW	0.71	0.63	0.77
	VLAD	0.07	0.07	0.55
	FV	0.07	0.10	0.52

The VF encoding surpasses the VLAD on the *motorbike* and *person* classes while it's slightly less performant then VLAD on the class *aeroplane*.
Same remark as for VLAD regarding the Hellinger kernel.

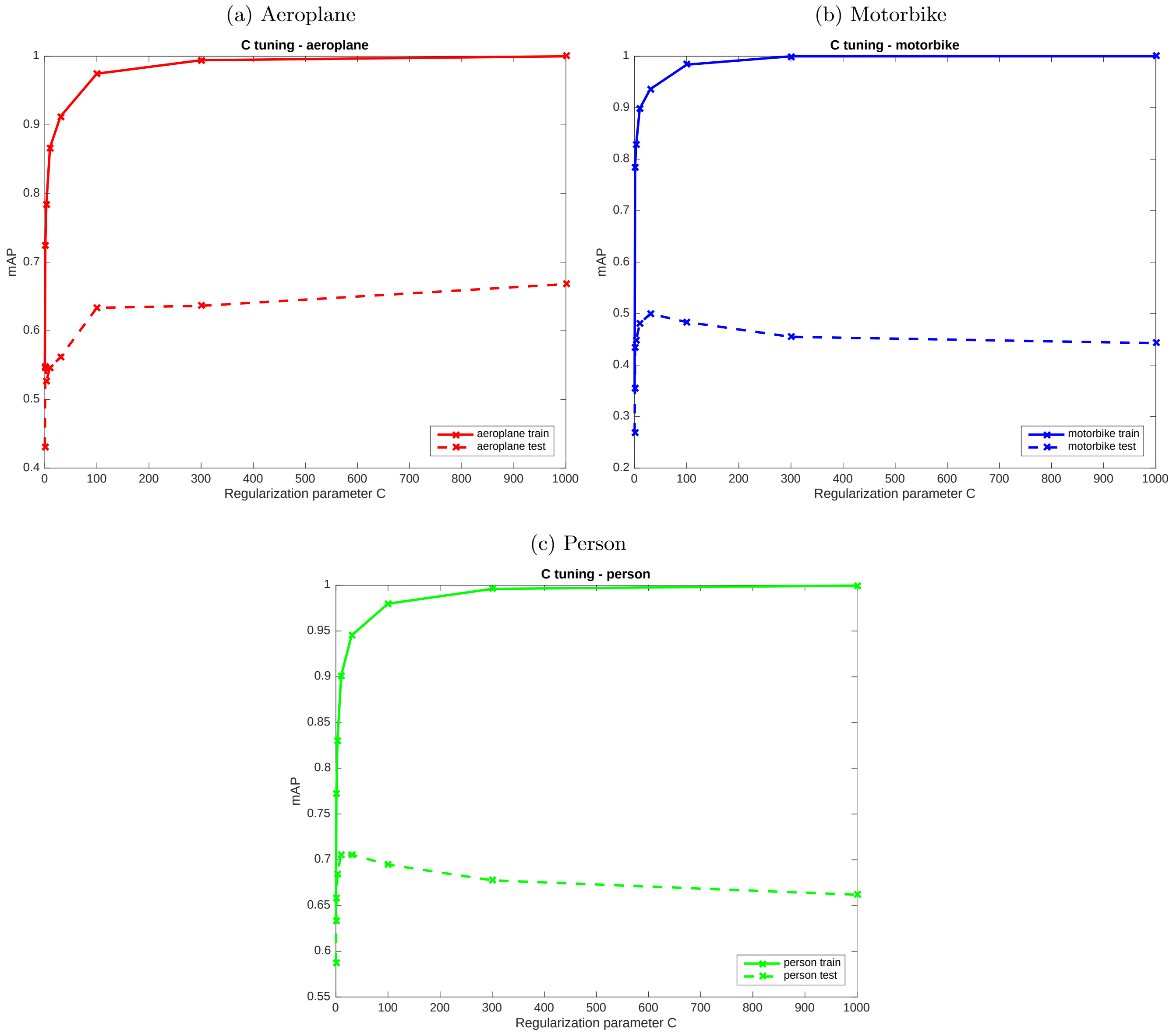
Q I.2

The VLAD vector is a non-probabilistic version of of the Fisher vector with half the dimension $((K \times D)$ instead of $(2D + 1) \times K1)$. Hence for a large number of images, or a large dimensionality we can favorise the VLAD over a small loss of AP. Besides, inlcuding a second order statitic is computation costly.

Q I.2

We vary the regularization parameter C (.1 1 3 10 30 100 300 1000) and assess the performance on both the training and test set to choose C yielding the best score on the test set (doesn't allow overfitting the training set). The results are shown on the curves below. We note that each category has a different optimal C and that even with tuning, VLAD and FV still outperform BoVW.

Figure 11: AP curves - train vs test



(a) Aeroplane

(b) Motorbike

(c) Person