Computer Vision 1 Final project part 1 - BoW

Anne Chel id:10727477, Bobbie van Gorp id:11161108,
Marvin Lau id:12364282 and Milan Klaasman id:11431474

March 2019

1 Introduction

In this assignment, a Bag of Visual words approach is taken as image representation, which will be used for binary classification with SVM for 5 classes. These classes are images of airplanes, birds, cars, horses and ships. They will be tested on a test set of 800 images of each class. Their mean average precision and accuracy are computed. In addition, the top 5 images that are ranked highest by a classifier with respect to the corresponding class probabilities as well as the worst 5 images are shown.

2 Design and experimental setup

First, a vocabulary is constructed with 150 images from each class out of the five, which is 750 images in total , for all cases to stay consistent. Descriptors are extracted from each image, for dense sampling, a step size of 10 is chosen, which is not too big or not too small. This value was especially chosen in respect to the limited computational budget, mainly memory (and run-time). In addition, a bin size of 8 is used in contrast to the default 3 as done in an example in the VL documentation of a function and worked better under first impressions. For handling RGB and opponent-RGB (oRGB), (key)points were extracted first as was done in the grayscale case. Those points (or frames) are then extracted from each channel to compute the descriptors. Collected descriptors are then clusters using k-means. Then, train and test data is quantized by constructing the histogram based on the found clusters, each descriptor is assigned to a cluster with k-nearest neighbour. For training, the remaining 350 images from each class are used to train a binary SVM model for each class out of the five. The SVM is tested on the full test set of 800 images from each class out of the five with accuracy and mean average precision (mAP) as metrics. Many cases are then measured including different combinations for the SIFT method (keypoints, dense), colorspace (gray, RGB, oRGB) and vocabulary size (400, 1000, 1000). This leads to 2x3x3x5 = 90 cases. Each of those cases were run with seed number 1.

Lastly, class probabilities are used for each classifier to rank images, which were labeled with the prediction that it is an image of the corresponding class that the classifier should recognize. The top and worst 5 are saved.

3 Results

The results of each case for each class can be observed in Tables 1,2,3,4 and 5 and the mean results in Table 6. In addition, the accuracy and mAP of all settings and classes are visualized in bar charts in Figure 1 and 2 respectively and the mean accuracy and mAP of the five classes in Figures 3 and 4. Different settings will be compared and observations will be elaborated in the sections that will follow.

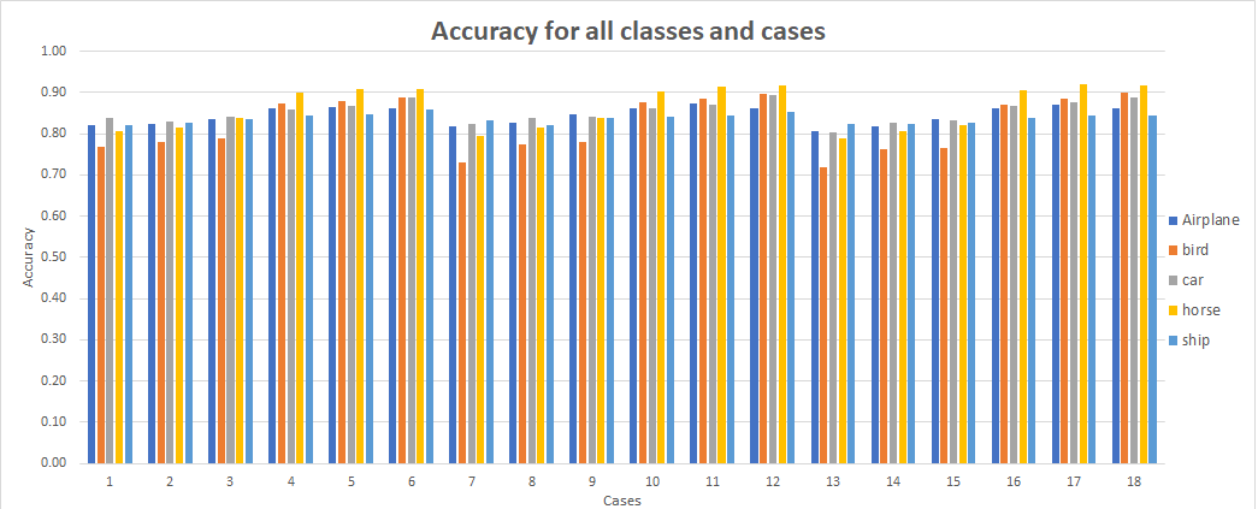

Figure 1: Bar chart of accuracy of all classes and cases

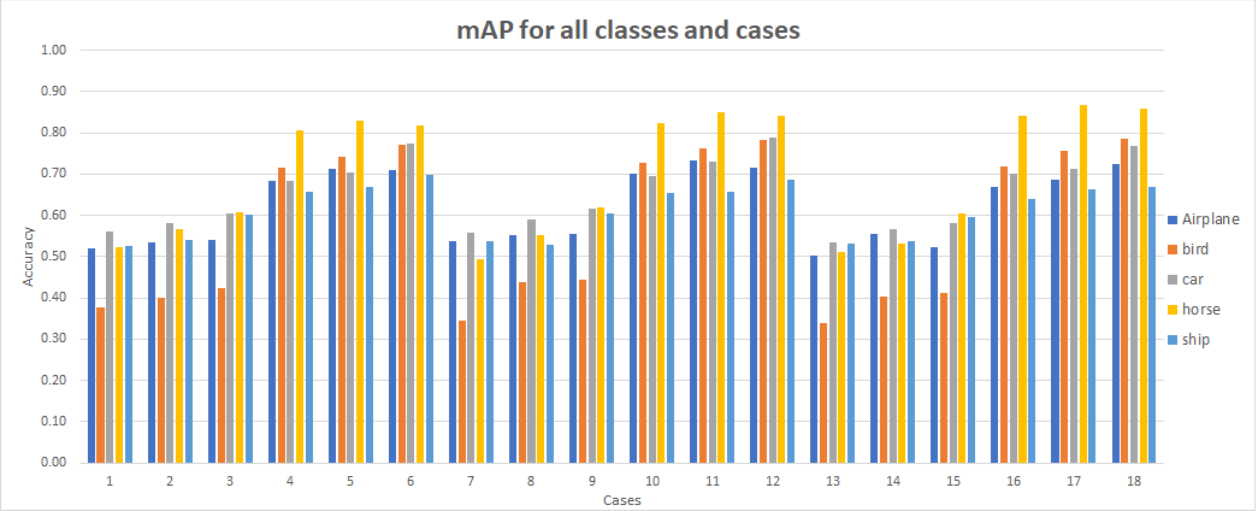

Figure 2: Bar chart of mAP of all classes and cases

3.1 SIFT methods

Looking at the accuracy and mAP results for each class, it can observed that dense sampling leads to greater performance than sparse sampling of keypoints. This is also supported by the four bar charts and is most prominent in the difference in mAP scores as visualized in Figure 2 and 4. Dense sampling leads to mAP scores approximately between 0.7 to 0.76 when looking at the mean results in Table 6. In most cases, dense sampling leads to an 40 procent increase on the mAP score of sparse sampling keypoints which lies approximately in the range of 0.48 to 0.56. Dense sampling SIFT descriptors leads to an consistent and possibly greater amount of descriptors depending on parameters. In this case, default parameters are used except for a binsize of 8 instead of 3 and a stepsize of 10. These settings ensures reasonable amount of descriptors are extracted from each image and is possibly greater than the approach of keypoints in most cases. Extracing more descriptors from images leads to more information that can be beneficial in classification, which is proved by the results in this case. Extracting descriptors based on keypoints can differ per image and can lead to less amount of descriptors and is therefore sparse. In addition, the parameters such as stepsize in dense sampling can be adjusted to a lower value for example to see if it will improve the performance even more as the amount descriptors extracted from an image increases even more. Dense sampling does come at the cost that it has more overhead and can be memory expensive when extracting more descriptors.

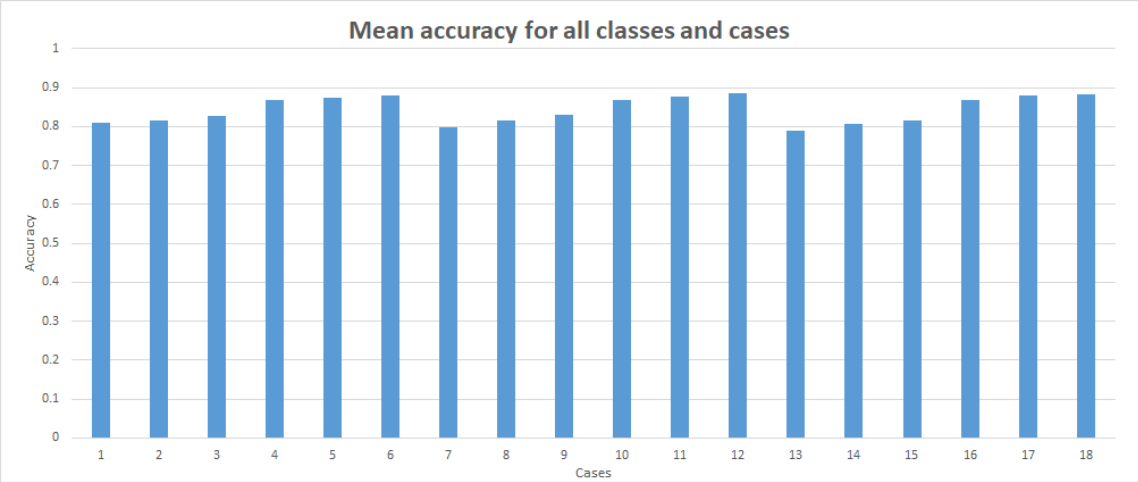

Figure 3: Bar chart of mean accuracy of the five classes

3.2 Color spaces

Looking at the three color spaces, gray performs the worst out of the three in both sparse and dense sampling cases. Making use of the color channels increases the performances a bit, not as prominent as with dense sampling but still an increase of 0 to 0.06 in mAP score in all cases. This makes sense as color channels definitely can contain use full information and contribute to classification. Humans recognize objects also partially on color. We know for example, that grass is green or that snow is white normally or in this case that horses are usually brown, black or white. A observation of an red horse is rare for example unless the image is manipulated. However, the

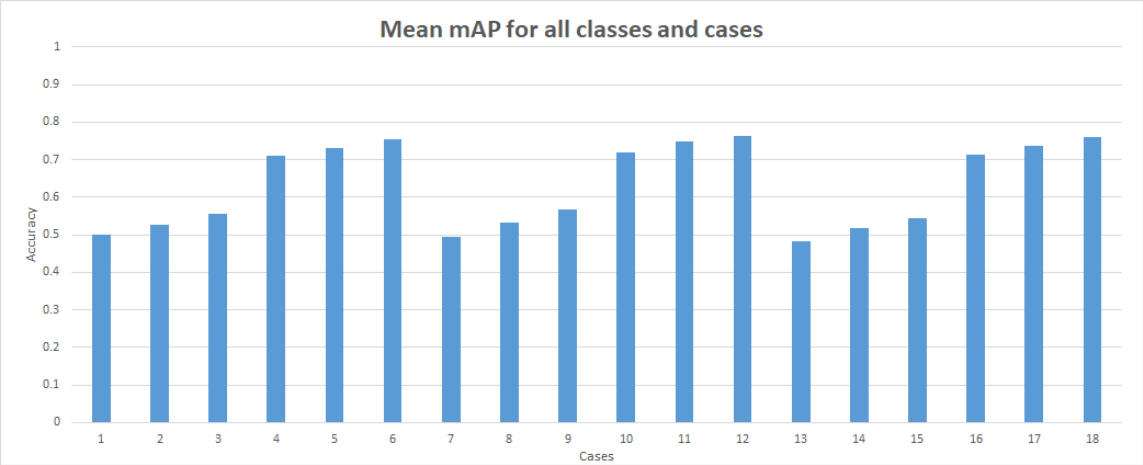

Figure 4: Bar chart of mean mAP of the five classes

Setting	# Clusters	Color space	SIFT method	Accuracy	mAP	# Correct top5
1	400	gray	keypoints	0.8200	0.5192	5
2	400	RGB	keypoints	0.8245	0.5345	3
3	400	oRGB	keypoints	0.8350	0.5396	2
4	400	gray	Dense	0.8612	0.6849	5
5	400	RGB	Dense	0.8640	0.7133	4
6	400	oRGB	Dense	0.8618	0.7096	5
7	1000	gray	keypoints	0.8167	0.5375	4
8	1000	RGB	keypoints	0.8265	0.5509	5
9	1000	oRGB	keypoints	0.8458	0.5552	2
10	1000	gray	Dense	0.8618	0.7017	5
11	1000	RGB	Dense	0.8728	0.7325	5
12	1000	oRGB	Dense	0.8612	0.7167	5
13	4000	gray	keypoints	0.8065	0.5039	5
14	4000	RGB	keypoints	0.8175	0.5549	5
15	4000	oRGB	keypoints	0.8340	0.5226	3
16	4000	gray	Dense	0.8622	0.6685	4
17	4000	RGB	Dense	0.8702	0.6856	4
18	4000	oRGB	Dense	0.8630	0.7240	5

Table 1: Results for airplane class

images that are used in this experiment are classification of real life objects with their corresponding (normal) color so color can contribute to unique property of an class. Furthermore, in most cases, the opponent-RGB also performs slightly greater than the RGB case with rest of the settings staying the same. Only in some cases of the airplane and horse class does the RGB class perform greater than the opponent-RGB case. The opponent-RGB space consist of luminance component, red-green channel and blue-yellow channel. As mentionend before, horses are mostly brown, which consist of more distinctive values in the red and green channel to create that color. When those 2 channel values are combined into 1 value, some information may be lost to represent those. In addition, the blue and yellow value have less impact as well with brown. Therefore, this representation may be more sparse for certain color as some information is lost and this would explain this mainly for the horse class and possibly also for some cases of the airplane class. Lastly, when looking at the mean

mAP scores, in all cases the oRGB results are slightly better with max 0.03 compared to the RGB cases.

Setting	# Clusters	Color space	SIFT method	Accuracy	mAP	# Correct top5
1	400	gray	keypoints	0.7672	0.3756	1
2	400	RGB	keypoints	0.7812	0.4010	2
3	400	oRGB	keypoints	0.7883	0.4242	3
4	400	gray	Dense	0.8740	0.7169	3
5	400	RGB	Dense	0.8805	0.7427	5
6	400	oRGB	Dense	0.8868	0.7698	4
7	1000	gray	keypoints	0.7312	0.3457	2
8	1000	RGB	keypoints	0.7745	0.4375	4
9	1000	oRGB	keypoints	0.7802	0.4454	5
10	1000	gray	Dense	0.8752	0.7285	5
11	1000	RGB	Dense	0.8838	0.7630	5
12	1000	oRGB	Dense	0.8958	0.7835	4
13	4000	gray	keypoints	0.7195	0.3376	0
14	4000	RGB	keypoints	0.7620	0.4037	1
15	4000	oRGB	keypoints	0.7660	0.4111	1
16	4000	gray	Dense	0.8690	0.7198	5
17	4000	RGB	Dense	0.8852	0.7568	5
18	4000	oRGB	Dense	0.8998	0.7863	5

Table 2: Results for bird class

Setting	# Clusters	Color space	SIFT method	Accuracy	mAP	# Correct top5
1	400	gray	keypoints	0.8395	0.5595	4
2	400	RGB	keypoints	0.8305	0.5823	5
3	400	oRGB	keypoints	0.8403	0.6053	5
4	400	gray	Dense	0.8590	0.6821	4
5	400	RGB	Dense	0.8688	0.7044	5
6	400	oRGB	Dense	0.8875	0.7726	5
7	1000	gray	keypoints	0.8223	0.5584	4
8	1000	RGB	keypoints	0.8377	0.5901	4
9	1000	oRGB	keypoints	0.8413	0.6171	5
10	1000	gray	Dense	0.8610	0.6937	5
11	1000	RGB	Dense	0.8698	0.7303	5
12	1000	oRGB	Dense	0.8950	0.7893	5
13	4000	gray	keypoints	0.8037	0.5346	5
14	4000	RGB	keypoints	0.8265	0.5653	4
15	4000	oRGB	keypoints	0.8333	0.5822	2
16	4000	gray	Dense	0.8668	0.6999	5
17	4000	RGB	Dense	0.8748	0.7123	5
18	4000	oRGB	Dense	0.888	0.7674	5

Table 3: Results for car class

Setting	# Clusters	Color space	SIFT method	Accuracy	mAP	# Correct top5
1	400	gray	keypoints	0.8057	0.5226	3
2	400	RGB	keypoints	0.8155	0.5676	5
3	400	oRGB	keypoints	0.8395	0.6086	3
4	400	gray	Dense	0.8995	0.8049	5
5	400	RGB	Dense	0.9087	0.8282	5
6	400	oRGB	Dense	0.9070	0.8180	4
7	1000	gray	keypoints	0.7943	0.4941	4
8	1000	RGB	keypoints	0.8157	0.5521	5
9	1000	oRGB	keypoints	0.8380	0.6188	5
10	1000	gray	Dense	0.9012	0.8240	5
11	1000	RGB	Dense	0.9145	0.8513	5
12	1000	oRGB	Dense	0.9167	0.8405	4
13	4000	gray	keypoints	0.789	0.5118	5
14	4000	RGB	keypoints	0.8057	0.532	4
15	4000	oRGB	keypoints	0.8217	0.6055	3
16	4000	gray	Dense	0.9063	0.8408	5
17	4000	RGB	Dense	0.9193	0.866	5
18	4000	oRGB	Dense	0.916	0.8591	5

Table 4: Results for horse class

Setting	# Clusters	Color space	SIFT method	Accuracy	mAP	# Correct top5
1	400	gray	keypoints	0.8215	0.5253	4
2	400	RGB	keypoints	0.8253	0.5416	5
3	400	oRGB	keypoints	0.8350	0.6018	5
4	400	gray	Dense	0.8450	0.6581	5
5	400	RGB	Dense	0.8462	0.6684	5
6	400	oRGB	Dense	0.8588	0.6970	5
7	1000	gray	keypoints	0.8310	0.5382	5
8	1000	RGB	keypoints	0.8215	0.5276	2
9	1000	oRGB	keypoints	0.8390	0.6032	5
10	1000	gray	Dense	0.8420	0.6544	5
11	1000	RGB	Dense	0.8430	0.6586	2
12	1000	oRGB	Dense	0.8532	0.6871	5
13	4000	gray	keypoints	0.8223	0.5306	5
14	4000	RGB	keypoints	0.8245	0.5376	5
15	4000	oRGB	keypoints	0.8267	0.5958	5
16	4000	gray	Dense	0.8375	0.6411	5
17	4000	RGB	Dense	0.8448	0.664	5
18	4000	oRGB	Dense	0.8442	0.6693	5

Table 5: Results for ship class

3.3 Vocabulary size

Then after observing results of different vocabulary size, increasing the vocabulary from 400 to 1000 can slightly increase the performance in most cases. The mean results also supports this, mainly in the dense sampling cases. On the other hand, increasing the vocabulary size to 4000 decreases the overall performance in most cases compared to either one of the two other cases. This applies to all color spaces. It could be the case that increasing the size to 1000 leads to more distinctive clusters, while increasing to 4000 all adds distinctive clusters, but also introduces too many clusters acting as noise. Thus some descriptors may not be clustered in the right one for example as with more clusters, there is a higher chance that clusters are close to each other and a slight position change of a descriptor in that space can lead to a miss classification of cluster so to

Setting	# Clusters	Color space	SIFT method	Accuracy	mAP
1	400	gray	keypoints	0.8108	0.5004
2	400	RGB	keypoints	0.8200	0.5254
3	400	oRGB	keypoints	0.8276	0.5559
4	400	gray	Dense	0.8677	0.7094
5	400	RGB	Dense	0.8736	0.7314
6	400	oRGB	Dense	0.8804	0.7534
7	1000	gray	keypoints	0.7991	0.4948
8	1000	RGB	keypoints	0.8152	0.5316
9	1000	oRGB	keypoints	0.8289	0.5679
10	1000	gray	Dense	0.8682	0.7205
11	1000	RGB	Dense	0.8768	0.7471
12	1000	oRGB	Dense	0.8844	0.7634
13	4000	gray	keypoints	0.7882	0.4837
14	4000	RGB	keypoints	0.8072	0.5187
15	4000	oRGB	keypoints	0.8163	0.5434
16	4000	gray	Dense	0.8684	0.7140
17	4000	RGB	Dense	0.8789	0.7369
18	4000	oRGB	Dense	0.8822	0.7612

Table 6: Mean results of the five classes

speak.

3.4 Comparing classes

Starting with the airplane class, it has reasonable results that is a bit of in the middle of all results. Sparse sampling leads to mAP scores between 0.50 to 0.55 and for dense sampling this increases from 0.66 to 0.73. The bird class has got the lowest mAP scores for keypoints sampling between the range of 0.34 to 0.44, while dense sampling leads to similar result as the airplane classifier with mAP scores between 0.71 to 0.78. It could be the case that keypoints cannot extract enough descriptors and thus cannot fully "analyse" the texture and shape of a bird well as well as its environment, which is achieved with dense sampling as it samples at every interval. Then the car class has got similiar mAP scores with dense sampling (between 0.53 to 0.61), but got higher mAP scores (from 0.68 to 0.78) for keypoints approach compared to the airplane class. Thus, a car may have more distinctive keypoints that can be detected and extracted than airplanes. The mAP scores for keypoints smapling for the horse classifier are similiar to the one of the car, however the mAP scores for dense sampling for this classifier are the highest for all classes, namely values between 0.8 to 0.86. This could be due to the reasoning that was given for dense sampling, RGB color space and the properties of a horse as mentioned in a previous section. Lastly, the ship classifier got similiar mAP scores for the keypoints sampling compared to the car, but got the lowest mAP scores (between 0.64 to 0.69) for dense sampling of all classes. Summarizing the best mAP score from each classifier:

- Airplane: mAP of 0.7325 (and 0.8728 accuracy) comes from airplane class is 1000 clusters, dense sampling on RGB channels.
- Bird: mAP of 0.7863 (and accuracy 0.8998) comes from the bird class with 4000 clusters with dense sampling on oRGB channels.
- Car: mAP of 0.7893 (and accuracy 0.8950) comes from the car class with 1000 clusters with dense sampling on oRGB channels.

- Horse: mAP of 0.866 (and accuracy 0.9193) comes from the horse class with 4000 clusters with dense sampling on RGB channels
- Ship: mAP of 0.6970 (and accuracy 0.8588) comes from the ship class with 400 clusters with dense sampling on oRGB channels.

Lastly, the top5 and worst 5 images are shown for each setting for each classifier. A small part of this are shown below, but the rest can be found in the appendix. Looking at a example in Figure 28, the top 5 images based on class probabilities of the airplane classifier for setting 1 are shown and are ranked from left to right with left being the most confident and the most right being the fifth most confident one. Similarly, the worst 5 images can be seen in Figure 29 with left being the worst one and the most right is the fifth worse one. The top 5 images consist all of airplanes and the worst 5 ones contains some classification of three times a bird and two times a ship surprisingly. It could be the case that a ship is also made out of metal and has some similar shape at some corners and that birds may have an similar background in images such as the (blue) sky which leads to these misclassifications. The airplane class has some deviation in the amount of correct images in the top5. The car,horse and ship classifiers mainly has 4 or 5 correct images in the top5 with rarely a 2 or a 3. This is a bit surprising for the ship classifier as it got the lowest mAP scores for dense sampling, but can still be consistent in getting the top5 images correct. The lower mAP score may then be explained by either misclassifications in the lower ranks or that it has less positive assignments. Lastly, the bird classifier got the lowest mAP scores and also has the lowest amount of correct images in the top5 in most cases. It is also include a case, which has the lowest mAP score and is the only one which has 0 correct images in the top5. The top 5 and worst 5 images for this case, which is setting13, are shown in Figure 7 with the top 5 images at the top row and the worst 5 at the bottom row. From these images, it can be observed that the top 5 images consist of 3 airplanes, a ship and a horse. Earlier, we saw that birds were present in the worst 5 images for setting 1 of airplane classifier. For the same reason, it could be that the background such as the sky may have played a role into misclassfying the airplane as a bird. An explanation for the horse could be due to the environment as well. Furthermore, it can be observed that the worst 5 images does contain three images of a bird. Thus, the low mAP score of the bird classifier is not entirely due to misclassification, but also due to low ranking of correct classifications.

Figure 5: Top 5 images of setting 1 for airplane class

Figure 6: Worst 5 images of setting 1 for airplane class

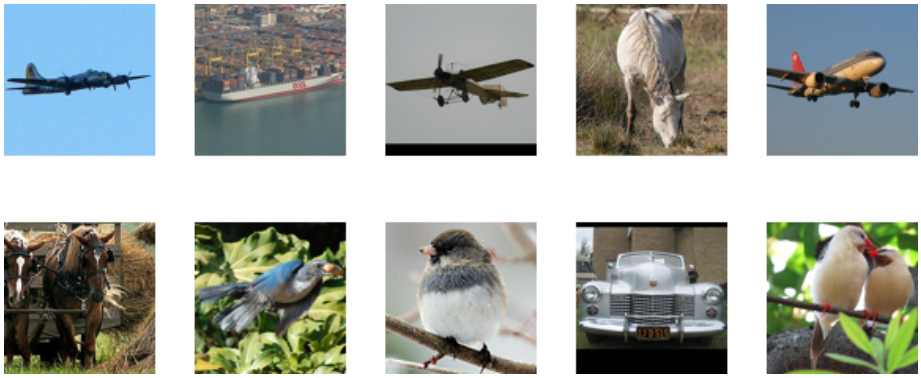

Figure 7: Top and worst (bottom) 5 images of setting 13 for bird classifier

4 Conclusion

In this assignment, a Bag of Visual words approach is taken for image representation used as features for five binary SVM classifiers, which in turn are used in 18 settings of experiments for each class. Out of the experiments, it can be concluded that dense sampling can significantly increase the overall performance, although it does with the cost of a lot memory and overhead. In addition, using color channels also slightly increases the performance, in particularly oRGB, which is slightly better than RGB in most cases with some exceptions such as for the horse classifier. Furthermore, a vocabulary size of 1000 results into a general good performance compared to 400 and 4000 as the former does not contain enough information and the latter consist of too much and creates noise. Lastly, the horse classifier result into the highest mAP score coming from setting 17 and the bird classifier has the lowest mAP score coming from setting 13. The distribution of workload of this assignment is as follows: The programming part was done by Marvin. The writing of the report was done by Marvin.

APPENDIX

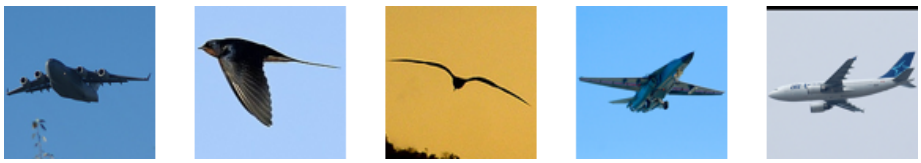

Figure 8: Top 5 images of setting 2 for airplane class

Figure 9: Worst 5 images of setting 2 for airplane class

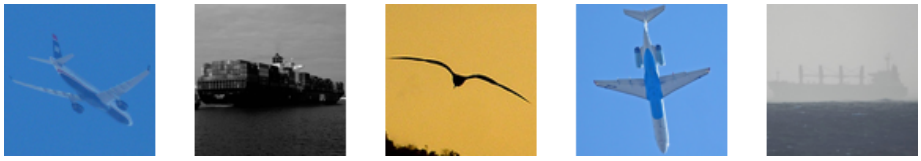

Figure 10: Top 5 images of setting 3 for airplane class

Figure 11: Worst 5 images of setting 3 for airplane class

Figure 12: Top 5 images of setting 4 for airplane class

Figure 13: Worst 5 images of setting 4 for airplane class

Figure 14: Top 5 images of setting 5 for airplane class

Figure 15: Worst 5 images of setting 5 for airplane class

Figure 16: Top 5 images of setting 6 for airplane class

Figure 17: Worst 5 images of setting 6 for airplane class

Figure 18: Top 5 images of setting 7 for airplane class

Figure 19: Worst 5 images of setting 7 for airplane class

Figure 20: Top 5 images of setting 8 for airplane class

Figure 21: Worst 5 images of setting 8 for airplane class

Figure 22: Top 5 images of setting 9 for airplane class

Figure 23: Worst 5 images of setting 9 for airplane class

Figure 24: Top 5 images of setting 10 for airplane class

Figure 25: Worst 5 images of setting 10 for airplane class

Figure 26: Top 5 images of setting 11 for airplane class

Figure 27: Worst 5 images of setting 11 for airplane class

Figure 28: Top 5 images of setting 12 for airplane class

Figure 29: Worst 5 images of setting 12 for airplane class

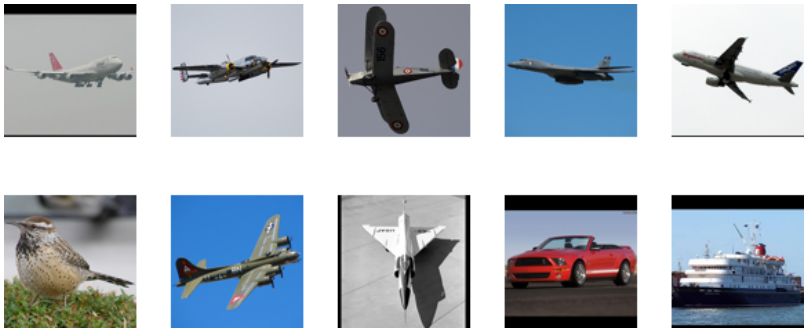

Figure 30: Top and worst 5 images of setting 13 for airplane class

Figure 31: Top and worst 5 images of setting 14 for airplane class

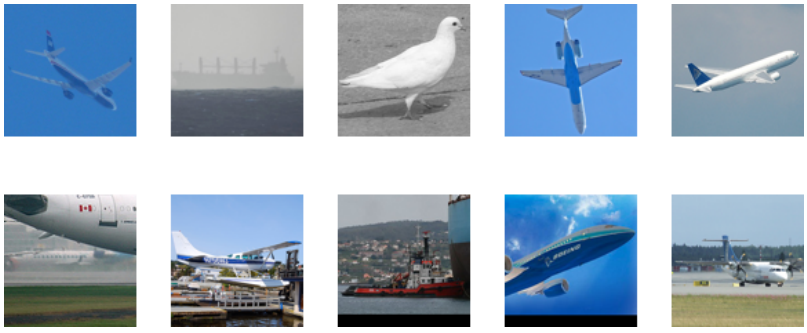

Figure 32: Top and worst 5 images of setting 15 for airplane class

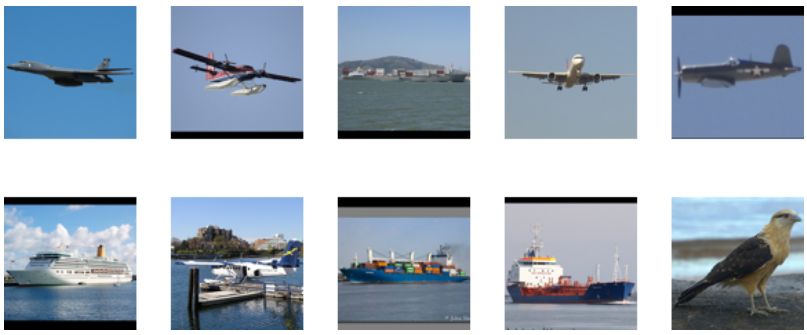

Figure 33: Top and worst 5 images of setting 16 for airplane class

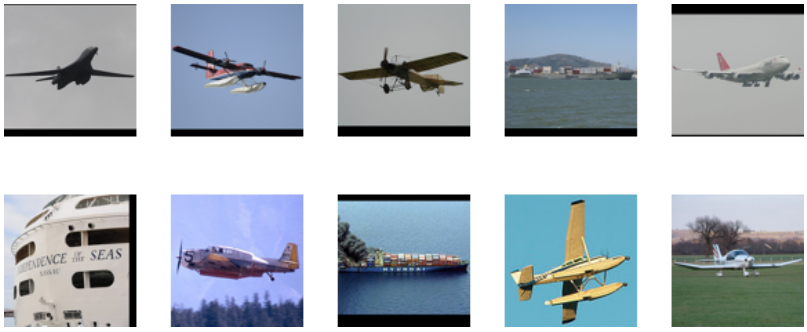

Figure 34: Top and worst 5 images of setting 17 for airplane class

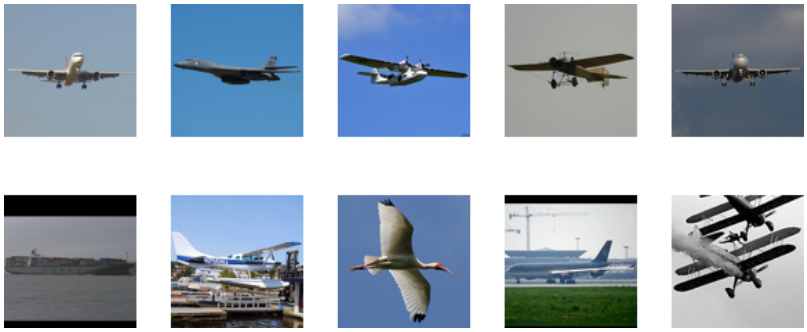

Figure 35: Top and worst 5 images of setting 18 for airplane class

Figure 36: Top 5 images of setting 1 for bird class

Figure 37: Worst 5 images of setting 1 for bird class

Figure 38: Top and worst 5 images of setting 2 for bird class

Figure 39: Top and worst 5 images of setting 3 for bird class

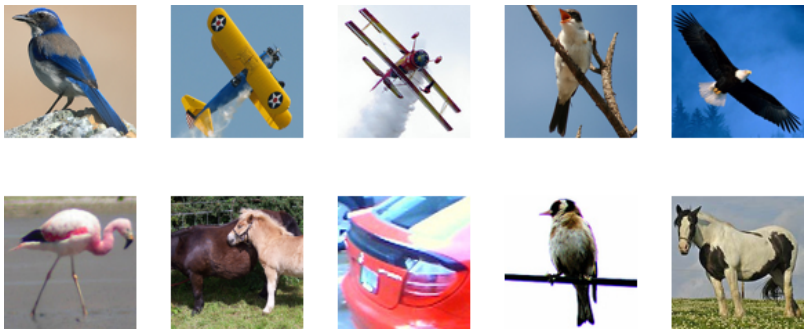

Figure 40: Top and worst 5 images of setting 4 for bird class

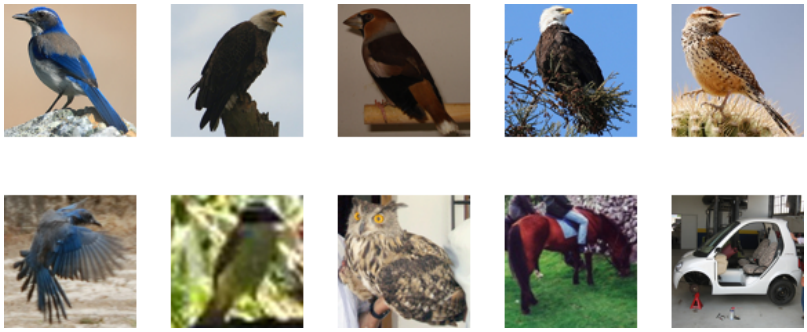

Figure 41: Top and worst 5 images of setting 5 for bird class

Figure 42: Top and worst 5 images of setting 6 for bird class

Figure 43: Top and worst 5 images of setting 7 for bird class

Figure 44: Top and worst 5 images of setting 8 for bird class

Figure 45: Top and worst 5 images of setting 9 for bird class

Figure 46: Top and worst 5 images of setting 10 for bird class

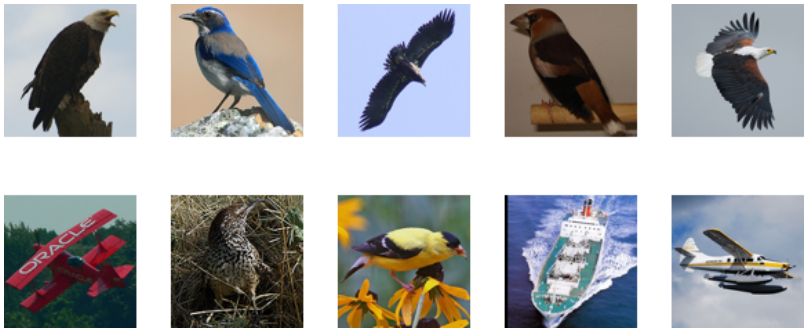

Figure 47: Top and worst 5 images of setting 11 for bird class

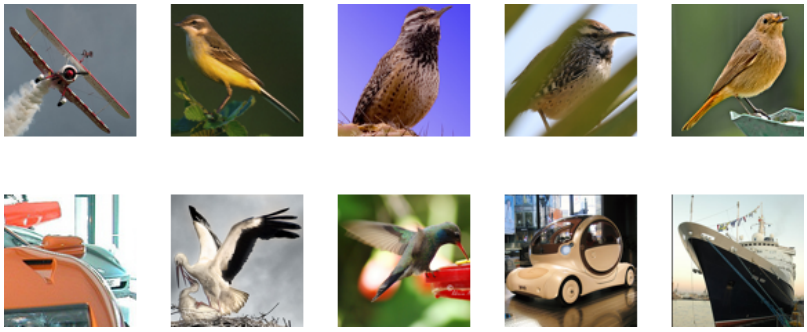

Figure 48: Top and worst 5 images of setting 12 for bird class

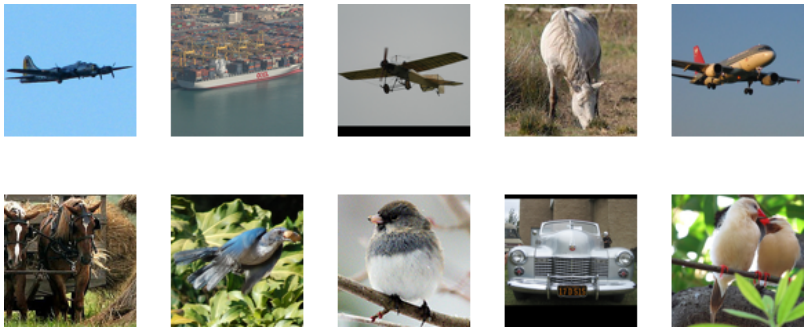

Figure 49: Top and worst 5 images of setting 13 for bird class

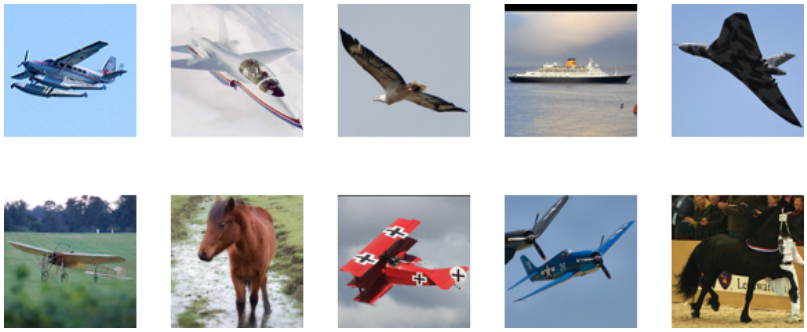

Figure 50: Top and worst 5 images of setting 14 for bird class

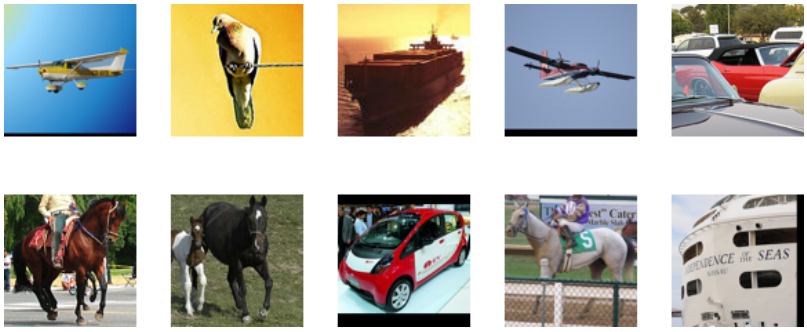

Figure 51: Top and worst 5 images of setting 15 for bird class

Figure 52: Top and worst 5 images of setting 16 for bird class

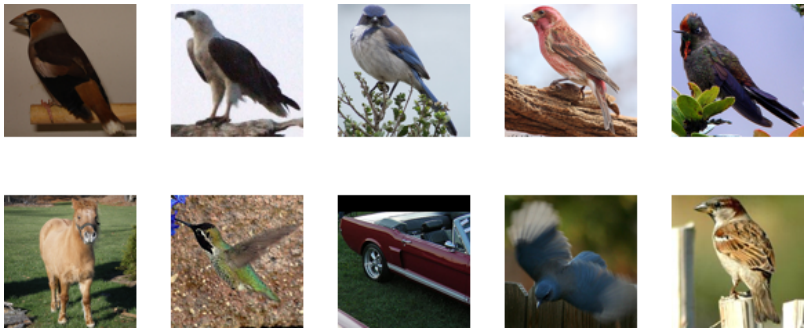

Figure 53: Top and worst 5 images of setting 17 for bird class

Figure 54: Top and worst 5 images of setting 18 for bird class

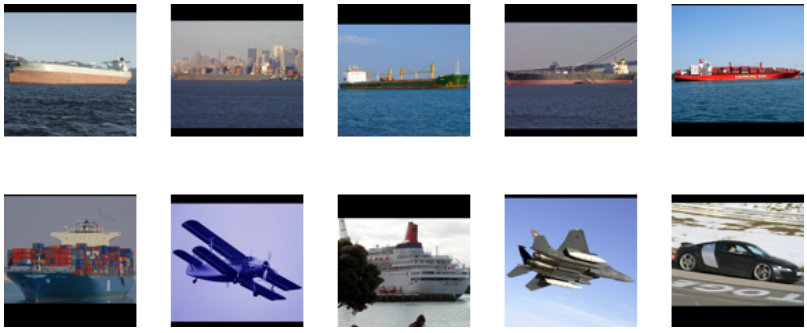

Figure 55: Top and worst 5 images of setting 18 for ship class