M5 Object detection

Week 4: Image retrieval

Group 6: José Manuel López Camuñas, Marcos Conde Osorio, Alex Martín Martínez

INDEX

- Image retrieval
- Task a: pre-trained image classification model
- Task b: Siamese network
- Task c: Triplet network
- Task d: Visualization of the learned image representation

Image retrieval

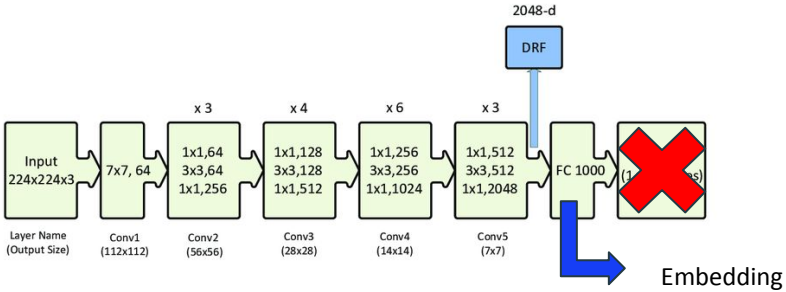
In this week's task, we performed image retrieval with 2 different approaches. On the one hand, we used a previously trained model on the classification task for the MIT split dataset, and on the other hand, we used the metric learning approach by testing how the Siamese and Triplet networks perform over the same dataset.

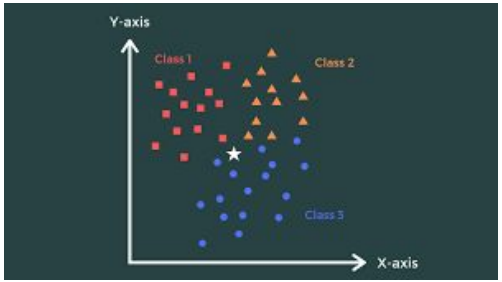
We used the previous models to perform the embedding, but to perform the final retrieval we used different algorithm such as KNN and FAIS.

Before performing the retrieval algorithms, we used PCA to reduce the dimensionality of the feature space in order to have a better performance of the models.

Task a: pre-trained image classification model

For this task, we used ResNet50 pre-trained on the MIT split for 5 epochs with cross entropy loss and SGD as the optimizer. We then removed the last layer of the fully connected layer to obtain an embedding of the images and performed KNN and FAIS to these vectors in order to obtain the retrieval.





Task a: pre-trained image classification model

Quantitative results

The metrics that we used to evaluate the performance of the models were the average precision (AP) with the top 1,3,5 and 10 retrieved results as well as the MAP. Also, the precision-recall curve and the confusion matrix.

Here we can observe that both algorithm perform very similar, but FAIS can achieve a little better performance.

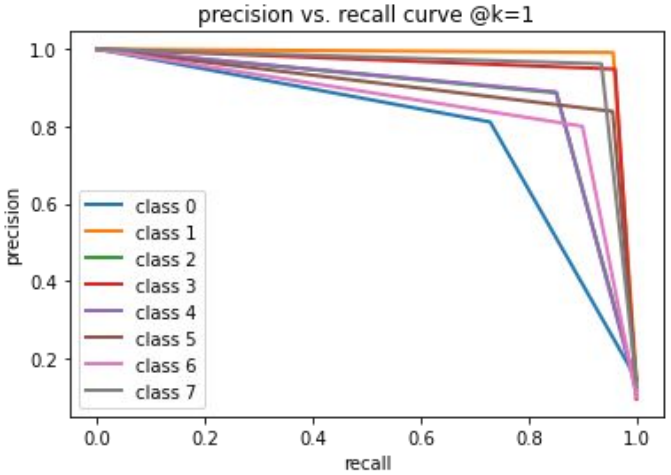
	AP@1	AP@3	AP@5	AP@10	MAP
FAIS	0.80	0.79	0.77	0.75	0.78
KNN	0.81	0.77	0.76	0.74	0.77

Task a: pre-trained image classification model

Quantitative results KNN

For all the top K results that we calculated the precision and recall curve, we obtained a similar plot.

We can observe that the class 1, 3 and 7 are well classified by the model approaching the ideal Pr-Re curve as they are near the (1,1) point in the plot. On the other hand, we observe that the class 0 is the worst one, followed by number 6. Probably those classes are being misclassified for other classes that have common features in the images.

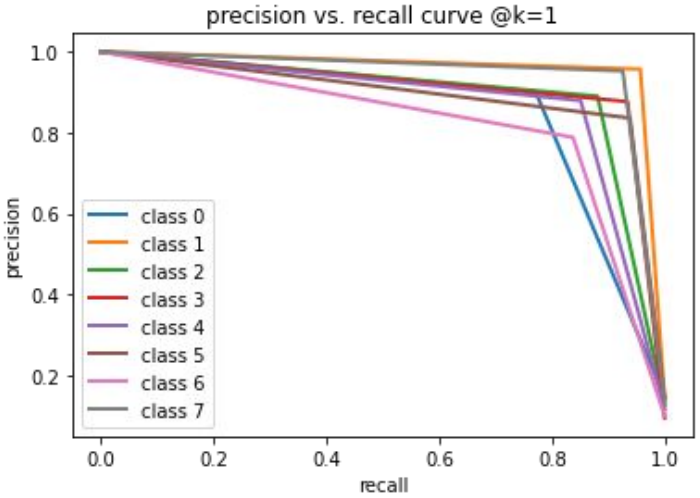


Task a: pre-trained image classification model

Quantitative results FAIS

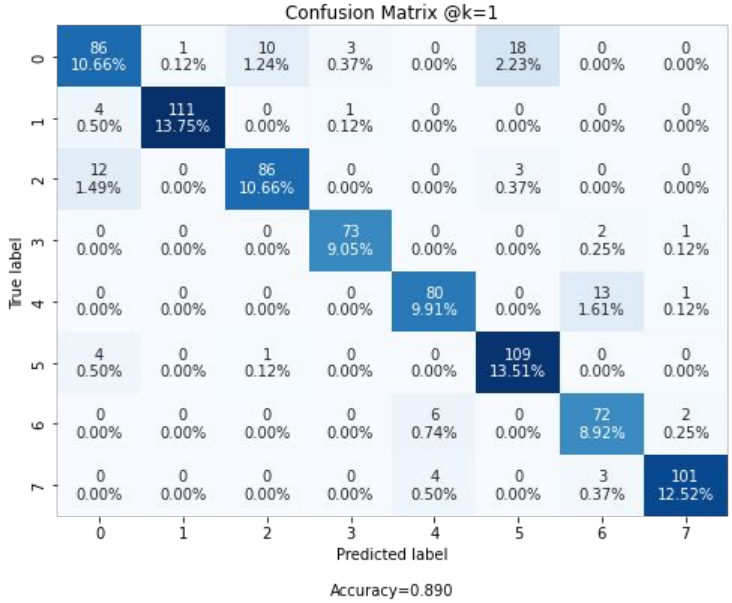
With the FAIS algorithm, we can observe that the class 1 and 7 are still the ones that perform the best, but the class 3 and 5 are had more similar curves than with the KNN.

We can also observe a substantial improvement in the performance of the class 0. So overall the curves are better than with the KNN.



Task a: pre-trained image classification model

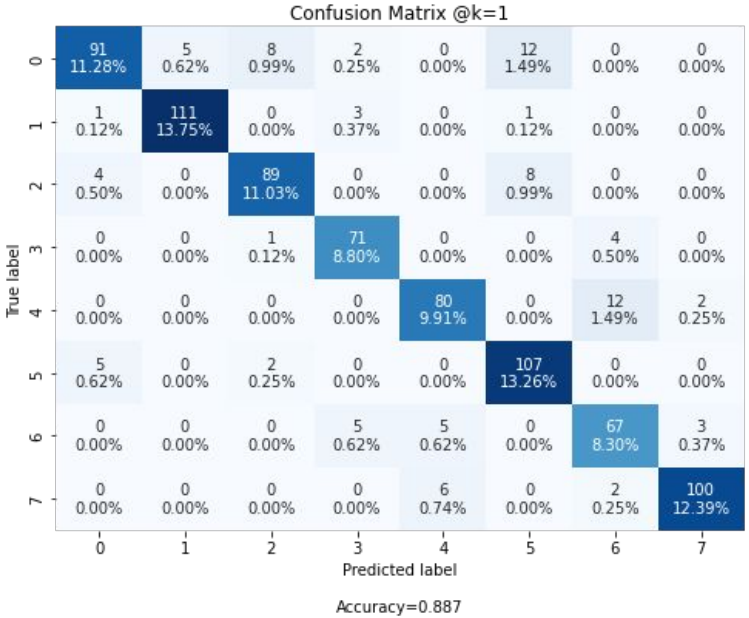
Quantitative results KNN



Here we can observe with more detail the Pe-Re curve behavior, as here we see the percentage of the labels being predicted of one class or of another one. In this confusion matrix we see that class 7, 3 and 1 have very few misclassifications. We can also see reflected that the class 5 haves a lot of misclassification with the class 0 being the classes mountain and open country respectively.

Task a: pre-trained image classification model

Quantitative results FAIS



In this confusion matrix, we observe the same behaviors as in the previous one but with a little more of precision than with the KNN algorithm

Task a: pre-trained image classification model

Qualitative results: KNN

top 3 predictions on train random sample

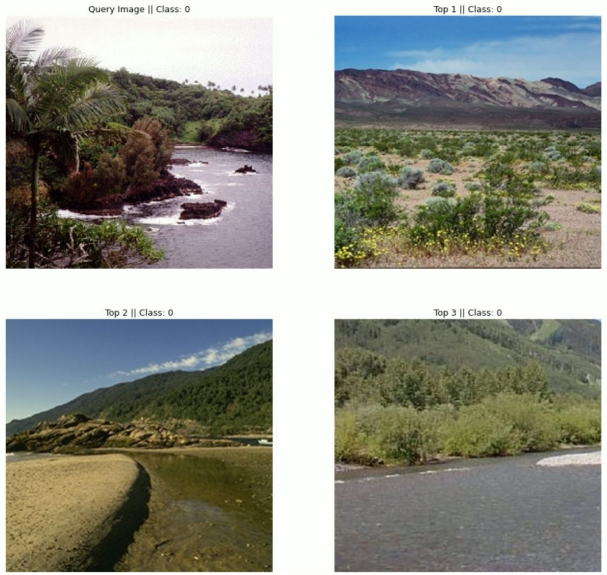
top 3 predictions on test random sample

Task a: pre-trained image classification model

Qualitative results: FAIS

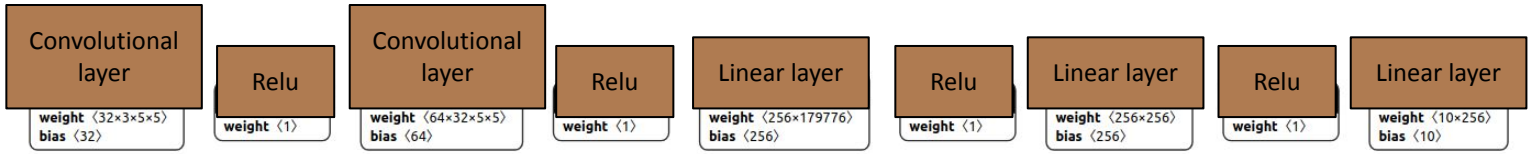
top 3 predictions on train random sample

top 3 predictions on test random sample



Task b: Siamese network

For this task we used a different model to perform the embedding as it is optimized to create different between images that are in different classes. Instead of the ResNet architecture that we used before, we now are using a very simple architecture with the following layers.



we then obtained a vector with 10 values. For this training we used the contrastive loss with margin 1, the Adam optimizer and a learning rate scheduler that reduces the learning rate by a factor of 10 every 8 epochs with an initial learning rate of 0.001.

Task b: Siamese network

To implement this, we mostly used code from the repo https://github.com/adambielski/siamese-triplet.

For the data loading, we took care of having balanced number of cases where the pair of images are equal and different. To do so, we used a coin flip before generating each pair to see if that pair would be equal or different. For the triplet, the approach was also this one.

Task b: Siamese network

Quantitative results

We can observe lower results compared to task A. This is mostly due to the usage of a deeper backbone (ResNet50) vs 2 convolutions blocks (siamese). Using ResNet50 would push our results close to task A.

Once again, FAIS is slightly better than KNN.

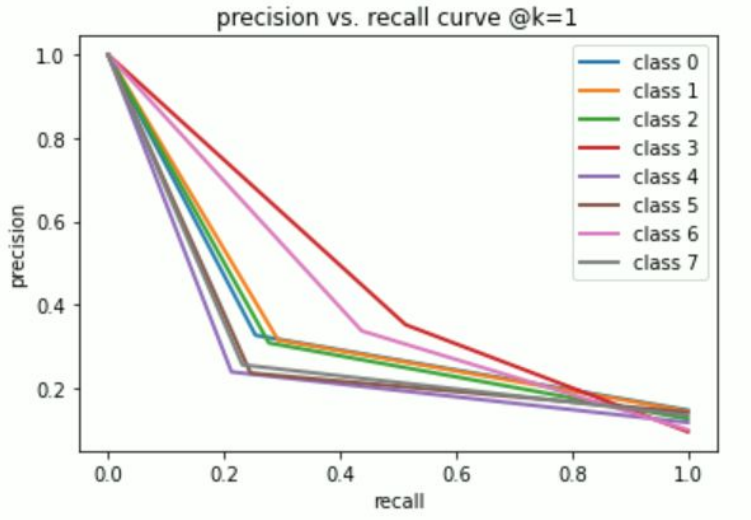
	AP@1	AP@3	AP@5	AP@10	MAP
FAIS	0.19	0.18	0.17	0.16	0.18
KNN	0.18	0.18	0.17	0.16	0.17

Task b: Siamese network

Quantitative results: KNN

In this case, we can see that all the classes are very far from the results with the previous model. We can see that the elbow of the curve doesn't approach the point (1,1).

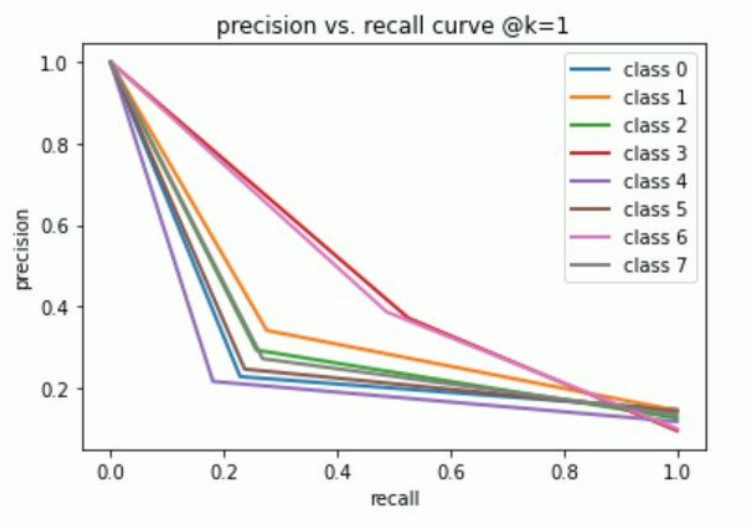
Observing the curves for the different classes seems like with this algorithm the better ones are class 6 and 3, although it can not be very



Task b: Siamese network

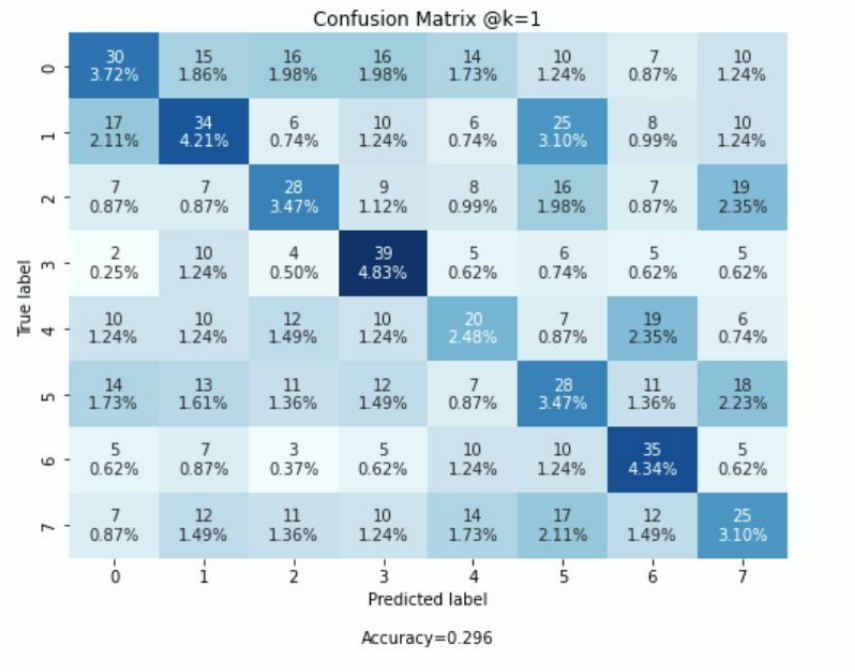
Quantitative results: FAIS

For the FAIS we also see that the classes 3 and 6 are the ones that perform the best and comparing with the previous one we can see an improvement in these two classes with respect to the KNN but the classes that had a relatively intermediate performance now perform worse giving two extremes.



Task b: Siamese network

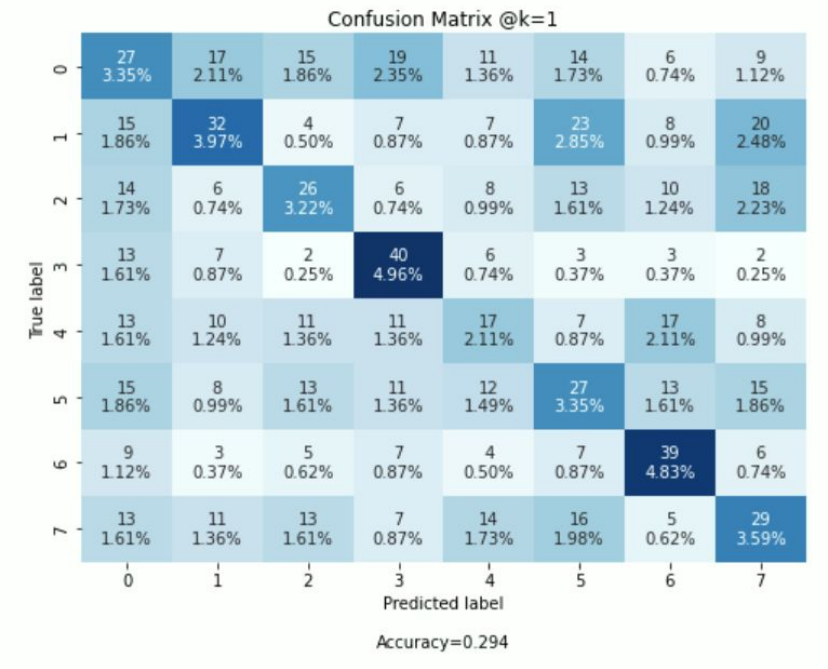
Quantitative results: KNN



Here we can not see a specific pattern to interpret as with task a due to the bad performance of this model. Overall we can see that for all the classes there are misclassifications, and in many of the classes these are evenly distributed. Although, we can see a bias of the predicted labels as from class 5 to be in reality from class 1 and the same for class 7 for class and 2.

Task b: Siamese network

Quantitative results: FAIS



With the FAIS we can observe the same patterns and behaviors but a little less accentuated as this models performs a little better.

Task b: Siamese network

Qualitative results: KNN

top 3 predictions on train random sample

top 3 predictions on test random sample

Task b: Siamese network

Qualitative results: FAIS

top 3 predictions on train random sample

top 3 predictions on test random sample

Task c: Triplet network

For this task we use the same architecture as with task b but in this case we use the Triplet network to perform the training. In this case, as well as a pair of images that may be from the same class or not, we also give the model an anchor image which is of the same class as one of the images. In this case, we used the triplet loss instead of the contrastive loss. The optimizer and the learning rate scheduler used in this task were the same as in the previous one.

This model was also trained for 20 epochs and gives a vector with 10 values.

Task c: Triplet network

Quantitative results

Similar to task B, we observe lower results compared to task A (but it's a big leap from task B). We suspect the difference in performance between A and B/C is the same commented in task B (the backbone problem). Triplet Network performs better than Siamese as expected.

Once again, FAIS is slightly better than KNN.

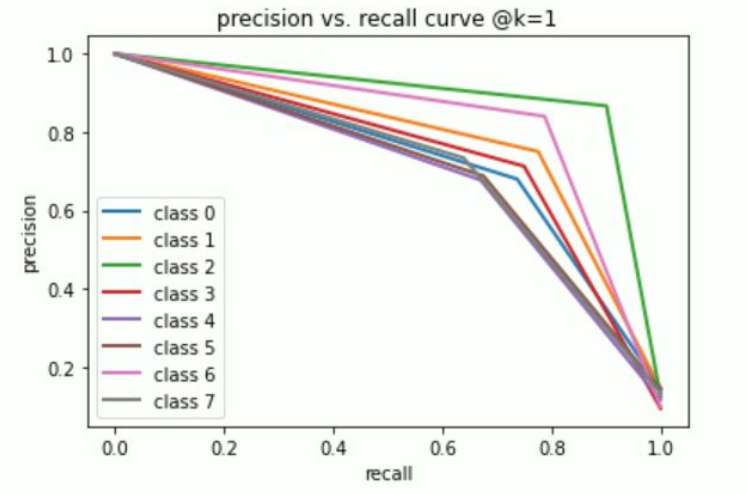
	AP@1	AP@3	AP@5	AP@10	MAP
FAIS	0.62	0.59	0.59	0.58	0.6
KNN	0.59	0.58	0.58	0.57	0.58

Task c: Triplet network

Quantitative results: KNN

For the KNN algorithm in this case we can observe that the elbow of the curve is approaching the (1,1) point although overall the curves are worse than the ones from task a.

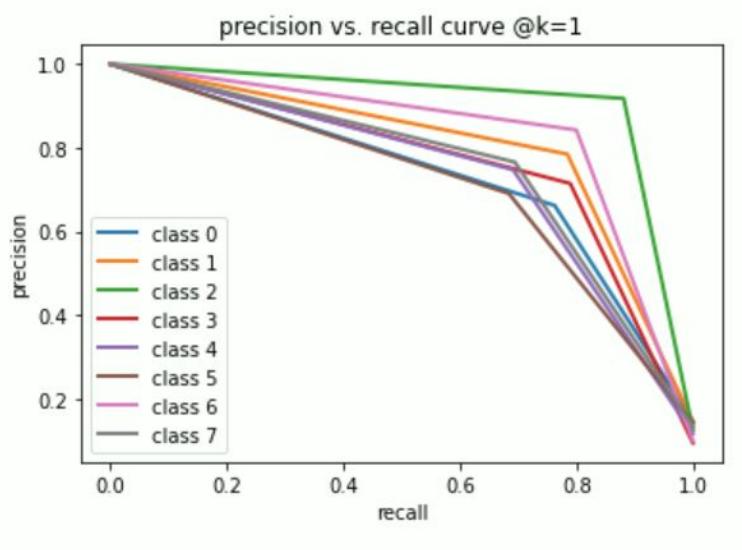
This time the class 2 is the best one an the class 7 which was the best one in task a with the KNN now it’s one of the worsts.



Task c: Triplet network

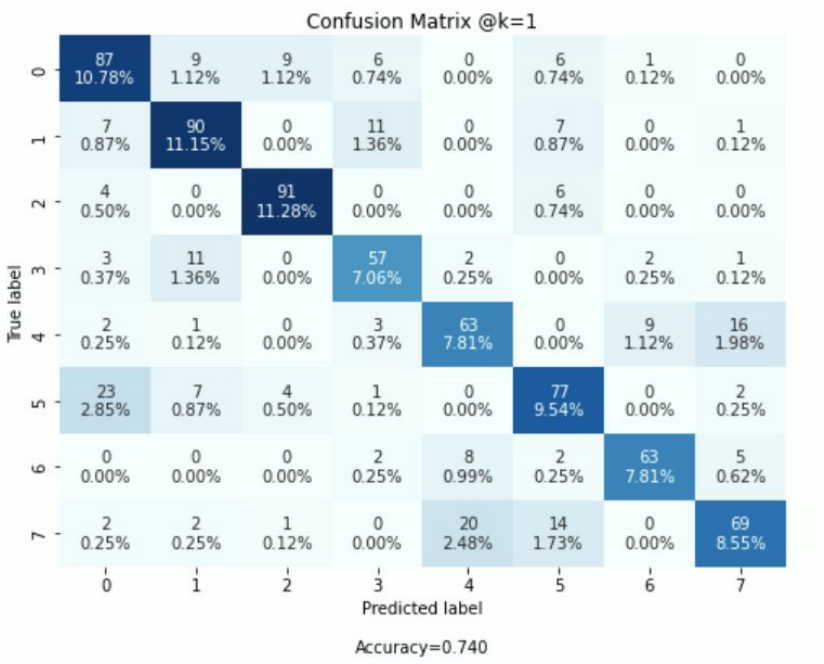
Quantitative results: FAIS

When using the FAIS we can see that overall the curves are near the ideal point and there are some changes in which classes are the worst ones with respect to the KNN. The worst retrieved classes are better and the best retrieved stayed more or less the same.



Task c: Triplet network

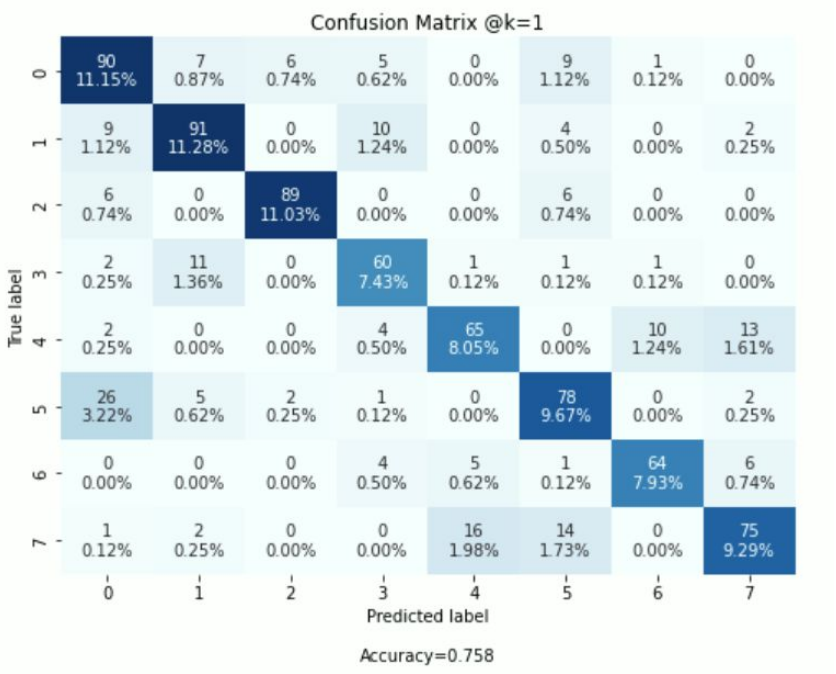
Quantitative results: KNN



In this confusion matrix, we see that for example when the label 3 is predicted from the model it could be wrongly classifying it with no particular bias, on the other hand when predicting class 7 there is a bias that the correct label was 4. The other way around also happens to be frequent. These 2 classes correspond to city and tall building photos, which share many features.

Task c: Triplet network

Quantitative results: FAIS

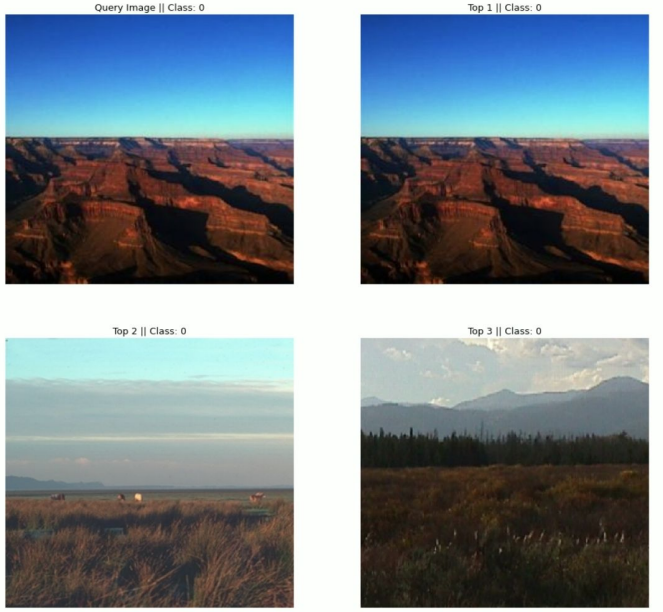


With the FAIS we see the same as with KNN, and in the case of predicting label 0 the bias towards the true label being 5 is even bigger and the bias between class 7 and 4 is smaller.

Task c: Triplet network

Qualitative results: KNN

top 3 predictions on train random sample



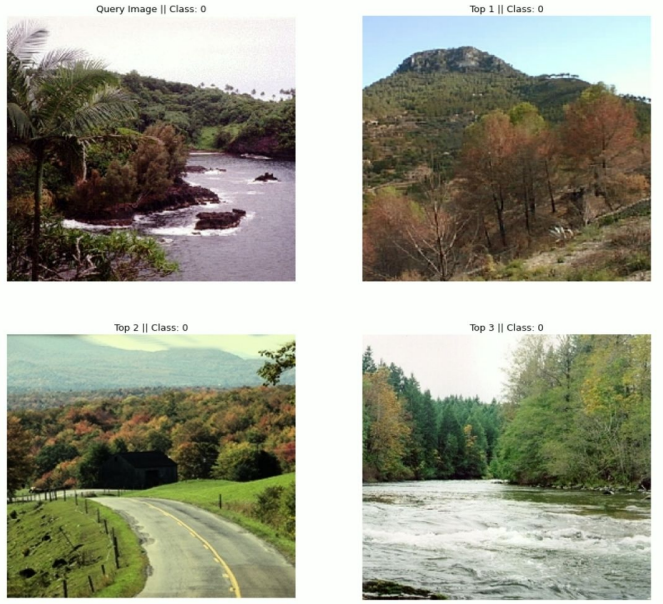
top 3 predictions on test random sample

Task c: Triplet network

Qualitative results: FAIS

top 3 predictions on train random sample

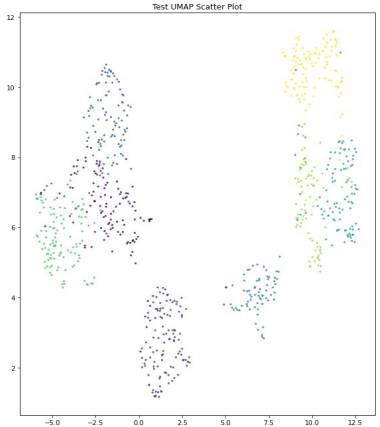
top 3 predictions on test random sample



Task d: Visualization of the learned image representation

Pre-trained model for classification

To interpret how the embedding was representing our images in the feature space, we used the UMAP and the TSNE methods to visualize it in a number of dimensions that can be graphically represented.

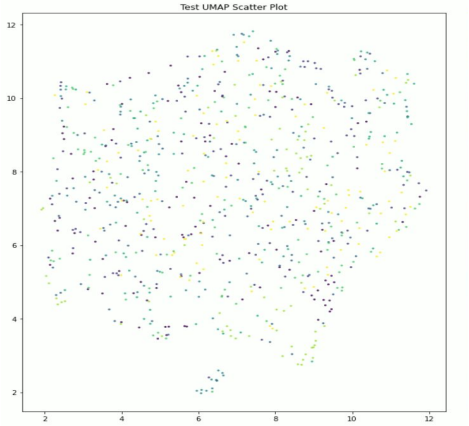


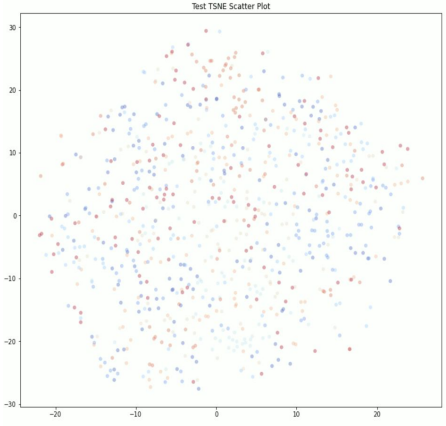
In both visualization we can see some differences, but they are overall similar. The points from the different classes seems to be distributed in 4 clusters instead of the 7 classes that there are in the dataset.

Task d: Visualization of the learned image representation

Siamese network

To interpret how the embedding was representing our images in the feature space, we used the UMAP and the TSNE methods to visualize it in a number of dimensions that can be graphically represented.



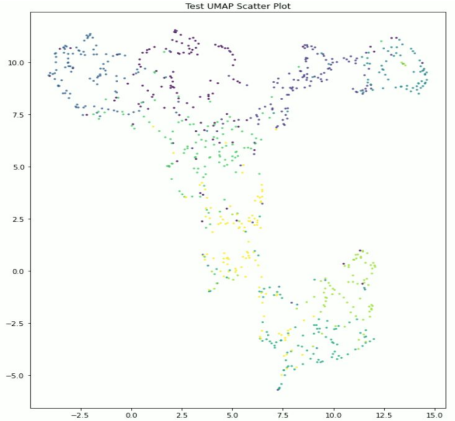


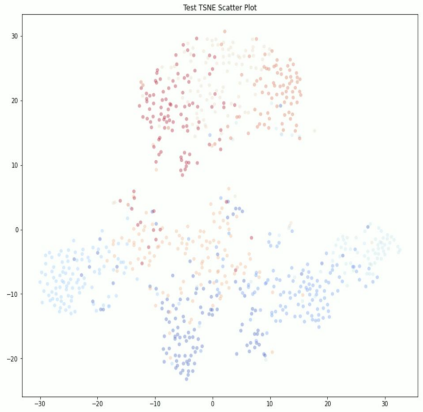
In both visualization we can see no cluster at all, our embedding representation is not good enough in siamese network (backbone problem).

Task d: Visualization of the learned image representation

Triplet network

To interpret how the embedding was representing our images in the feature space, we used the UMAP and the TSNE methods to visualize it in a number of dimensions that can be graphically represented.





Triple network representation is better than Siamese but worse than ResNet50. We can see some small clusters.

Task e: Interesting features to analyze

We implemented 2 out of the 5 options:

- Different retrieval methods on the same learned representation (for all tasks).
  - KNN
  - FAIS
- Different visualizations models with the same learned representation (for all tasks).
  - UMAP
  - TSNE

The results suggest that FAIS performs (overall) slightly better than KNN.

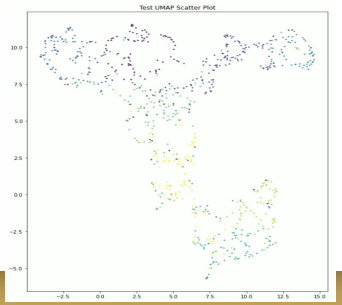
Similarly, TSNE presented clearer clusters than UMAP.

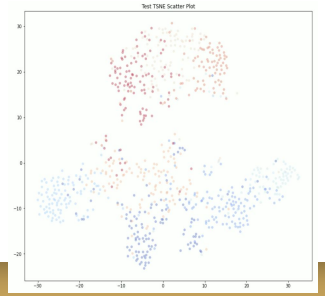
Final summary

Results on the different approaches tested

Changing the backbone for the embedding had a very negative impact in the performance. The Triplet network shows a great improvement with respect to the Siamese and having much fewer parameters than in the pretrained model the MAP isn't much lower.

Test	Pre-trained	Siamese	Triplet
MAP	0.78	0.18	0.6





Retrieval methods

The methods compared, KNN and FAIS, with all the embedding that were used show that FAIS is slightly better.

Visualization methods

In this case, we weren't able to observe a clear clusterization of the points in the feature space as the metric learning approaches didn't perform well enough, but we could see that the TSNE method gives better visualization qualitatively than the UMAP