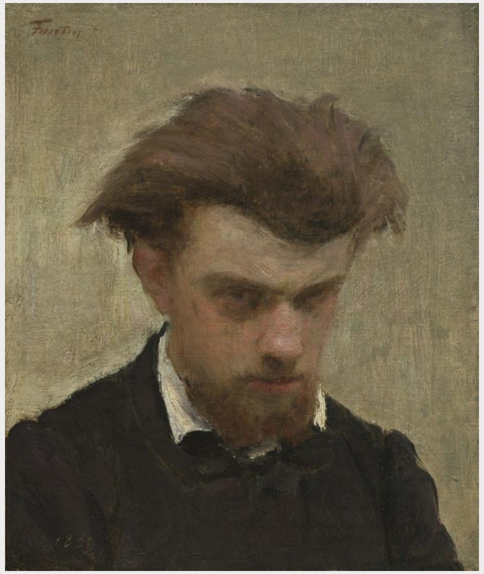
REDUCE PAINTINGS’ FEATURES DIMENSIONALITY WITH AUTOENCODERS

Introduction

- The visual arts field concerns all the kinds of arts that produce a visible item, especially paintings.
- Discovering artistic similarities between paintings made by different artists is an important task for research purposes, but also for recommending new items to art enjoyers.
- We look for similarities in the drawing and painting style, the colour usage, the subject of the painting etc. are likely to have influenced each other.

On the right, an example of two look-alike paintings: the first is from Durand, the second from Fantin



The need of deep learning

- The human brain is able to find similarities and analogies between two images relatively easily, but it is a time consuming operation, which becomes infeasible when dealing with large sets of data.
- We need an automatic way to extract salient characteristics from image and let machines do the comparison.

Deep learning techniques such as convolutional neural networks come in aid in these situations.

Convolutional neural networks

- Convolutional neural networks (CNN) are able to find the features of an image thanks to the convolution operation.
- They are composed of a stack of convolutional layers, each of which applies a different set of filters which are convolved on the input data to obtain a feature map. The latter is then subsampled by a pooling layer and fed to the next convolutional layer.
- As the number of layers grows, so does the complexity of the retrieved features.
- On top of the stack of convolutional layer there is a fully connected structure to perform classification.

The features are too large

After developing and training a CNN, we can let it compute the prediction for an image in order to extract from some (or all) of the layers the computed weights, which will be the features. This will lead to obtain a vectorized representation of the image (embedding).

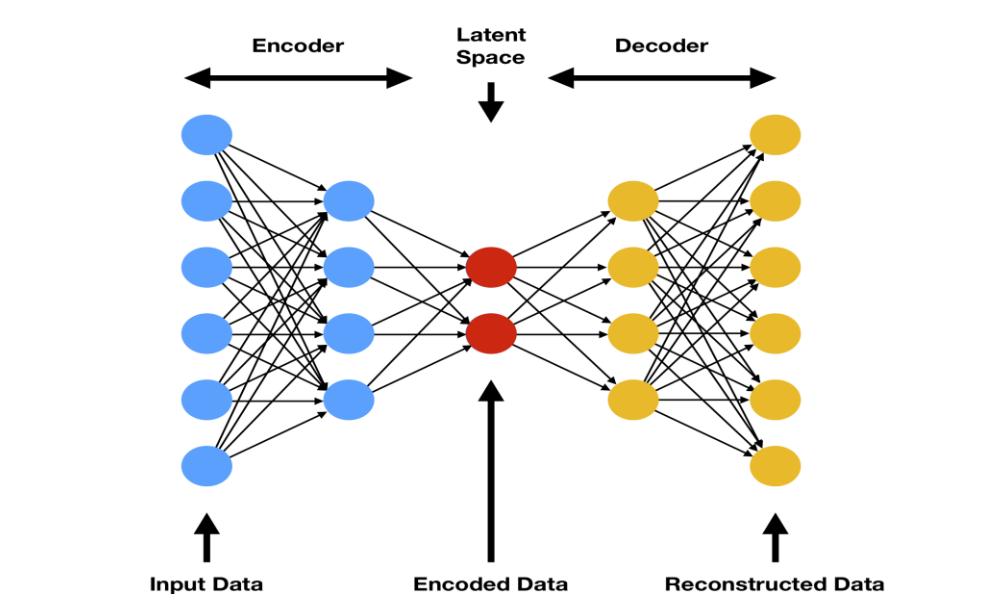
To find similar images, we can then compute a similarity measure between the embeddings.

However, the features vectors are extremely large due to the complexity of the convolutional layers of the CNN. We need a way to reduce the dimension of the embeddings.

Autoencoders

- Autoencoders are a kind of neural network made by two components:
  - *The encoder produces a reduced representation of the given input, keeping its most relevant properties*
  - *The decoder tries to reconstruct the original input starting from the encoded representation computed at the coding layer.*
- Autoencoders allow to compress data to an arbitrary size, spanning the same subspace spanned by PCA.

Example of autoencoder



Description of the work

- This work aims at seeing if the similarities between paintings are kept after reducing the size of their features vectors using an autoencoder.
- A pre-trained CNN was used to compute the features of the paintings made by some artists. The features were extracted by two layers.
- An autoencoder was also developed and trained using the extracted features. After the training, the encoder was used to reduce the dimensionality of the features and see if the similarities between artists are kept in the reduced feature space.

Used technology

- Python 3,9
- Keras 2,4,3
- Local GPU: NVIDIA Geforce MX330
- Google Colab GPU: Tesla K80

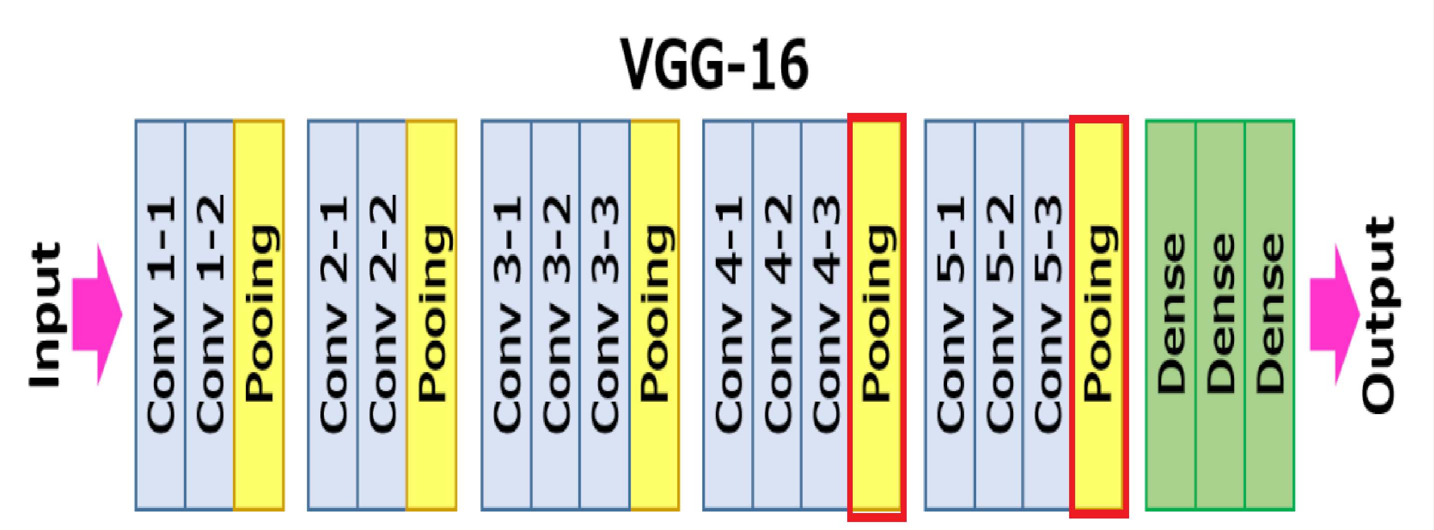
Dataset

- «Best artworks of all time», found on Kaggle
- 7775 paintings made by 49 artists.

Albrecht_Du¦êrer
Alfred_Sisley
Amedeo_Modigliani
Andrei_Rublev
Andy_Warhol
Camille_Pissarro
Caravaggio
Claude_Monet
Diego_Rivera
Diego_Velazquez
Edgar_Degas
Edouard_Manet
Edvard_Munch
El_Greco
Eugene_Delacroix
Francisco_Goya
Frida_Kahlo
Georges_Seurat
Giotto_di_Bondone
Gustav_Klimt
Gustave_Courbet
Henri_de_Toulouse-Lautrec
Henri_Matisse
Henri_Rousseau
Hieronymus_Bosch
Jackson_Pollock
Jan_van_Eyck
Joan_Miro
Kazimir_Malevich
Leonardo_da_Vinci
Marc_Chagall
Michelangelo
Mikhail_Vrubel
Pablo_Picasso
Paul_Cezanne
Paul_Gauguin
Paul_Klee
Peter_Paul_Rubens
Pierre-Auguste_Renoir
Piet_Mondrian
Pieter_Bruegel
Raphael
Rembrandt
Rene_Magritte
Salvador_Dali
Sandro_Botticelli
Titian
Vasiliy_Kandinskiy
Vincent_van_Gogh
William_Turner

Adopted CNN

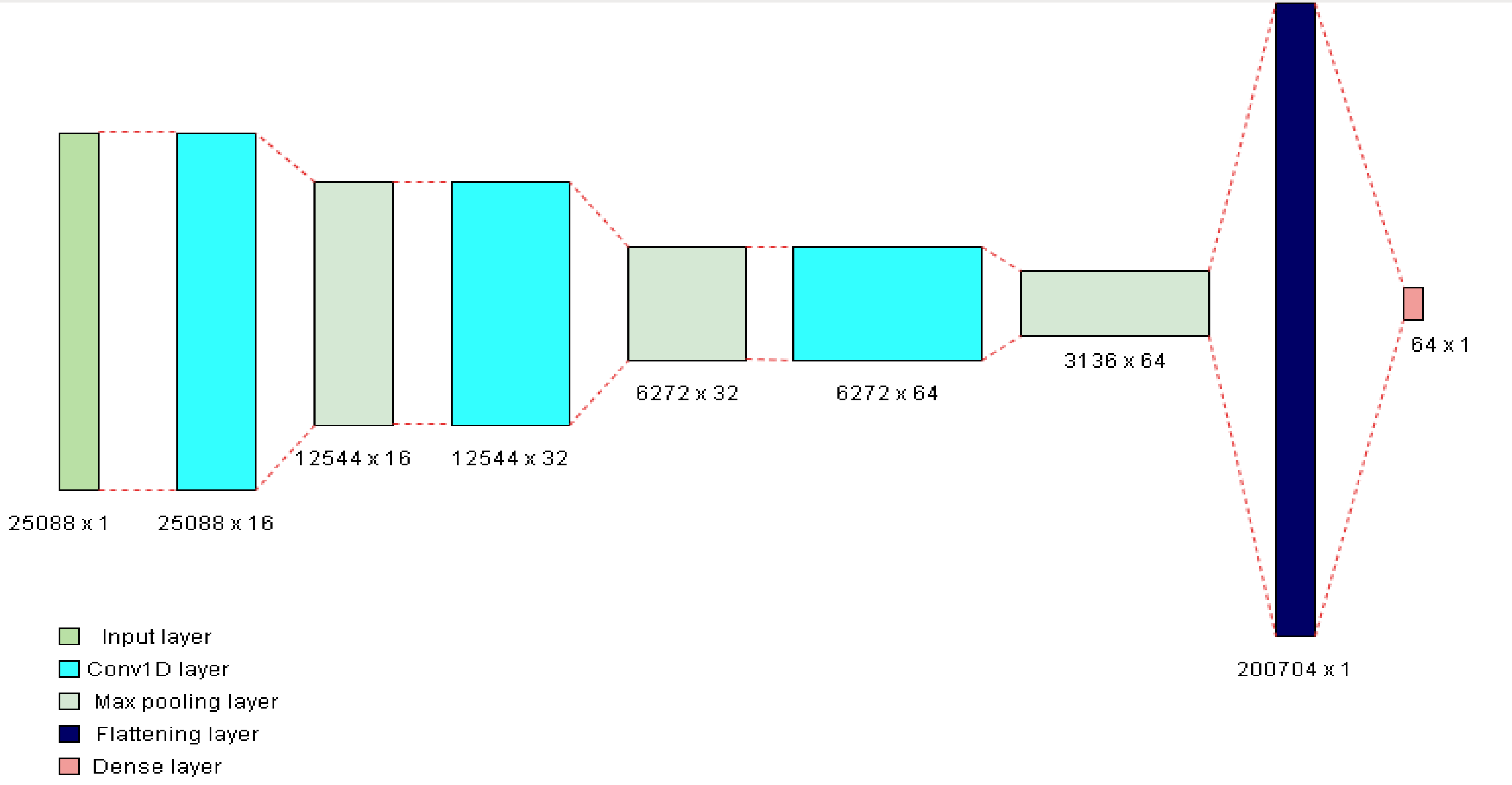
The VGG16 CNN was used for the features extraction. The network had been pre-trained on the Imagenet dataset. The two layers chosen for the feature extraction are the fourth and fifth pooling layers. The dense layers on top are ignored.



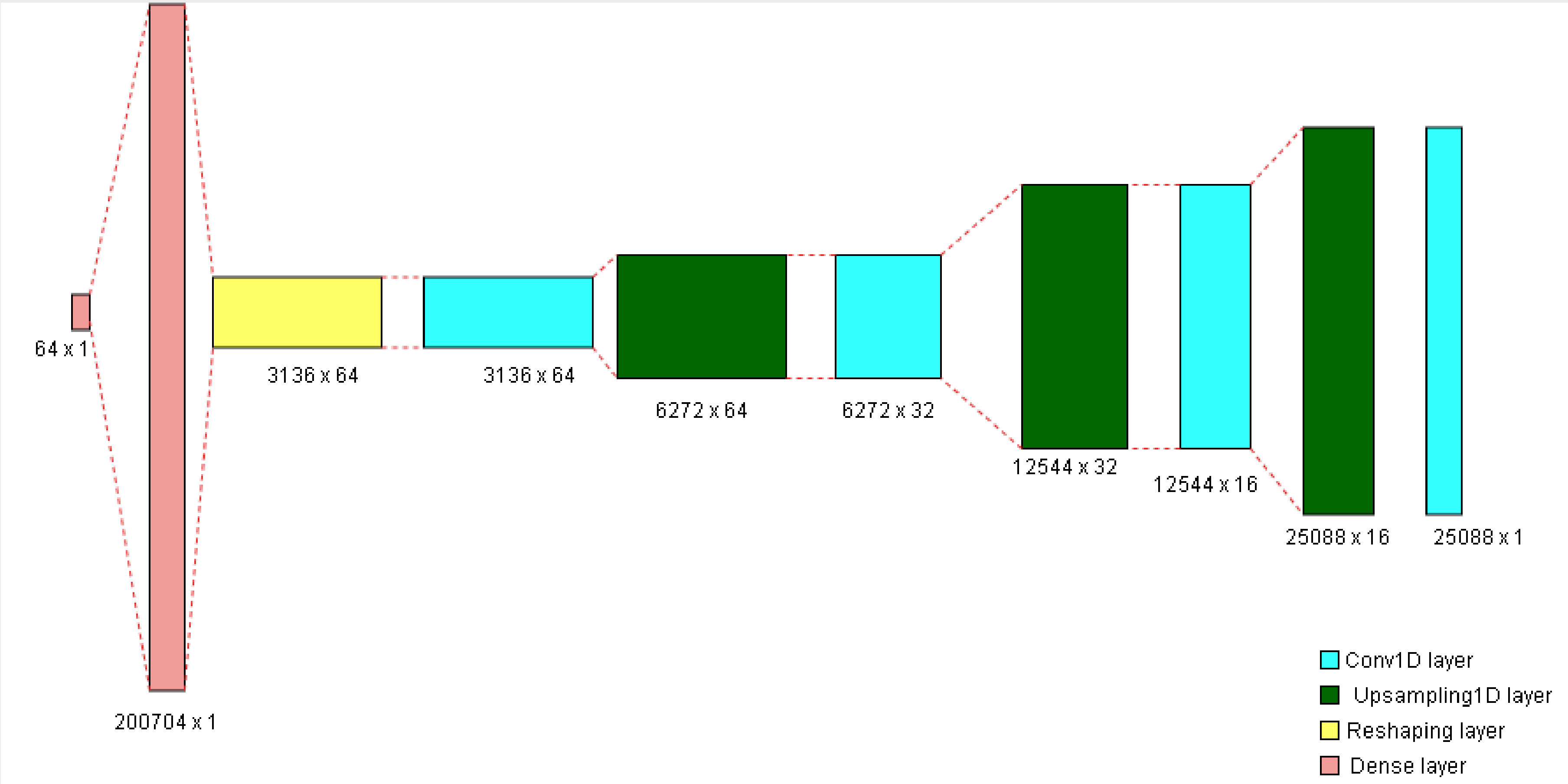
Autoencoder

- For each image, a feature vector containing 100352 elements is extracted from the fourth pooling layer, while the array extracted from the fifth layer contains 25088 features. The arrays extracted from the fourth layer are chunked in four pieces, each 25088 long. In this way, we can use the same architecture to learn five different models.
- Since the input dimension is very large, creating an autoencoder composed by fully connected layers would require an incredibly large amount of memory, which was not available. Hence, to obtain a lighter structure, the first layers of the encoder are 1D convolutional layers. The last part of the encoder contains a fully connected structure.
- At the code layer the initial input size is reduced to 64.

Encoder structure



Decoder structure



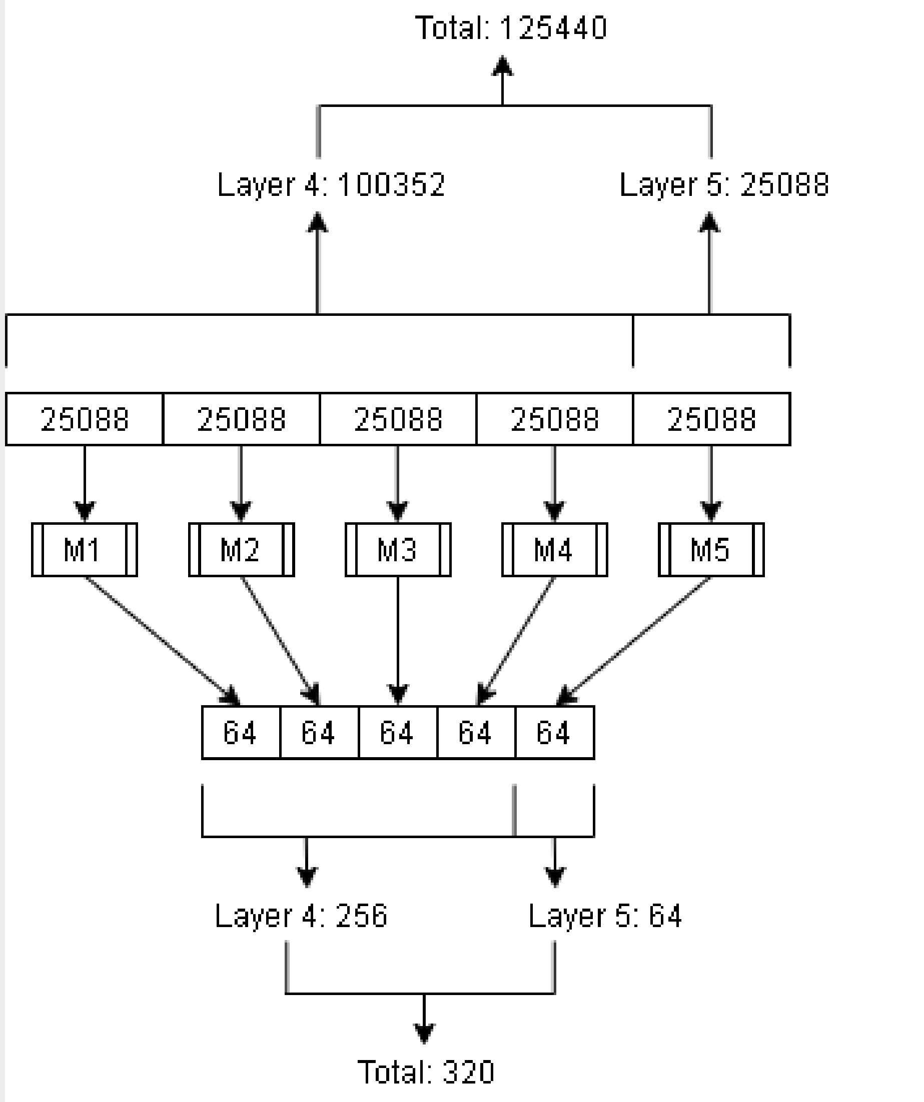


Illustration of the compression procedure

The picture depicts the way in which the features vectors are compressed. As said, the arrays extracted from the fourth layer are chunked and their parts are compressed separately. After this, the compressed parts are concatenated back. Also the compressed arrays extracted from the fifth layer are appended to obtain a final embedding of size 320.

Evaluation

The reduced embeddings of the paintings made by an artist were serialized together in a file.

The paintings from an artist will be part of the same cluster in the new feature space.

Also PCA was used to take the two most relevant components of the embeddings of the paintings and plot them in a scatter plot. By doing this we can have a visual idea of how the clusters are made.

The silhouette score was computed for the clusters representing each pair of artists. Given two clusters c1, c2, the silhouette score is computed as follows:

$$\text{Silhouette\_score (c1, c2)} = \frac{b-a}{\max(a,b)}$$

Where a is the mean distance between a point and the other point of the same cluster, b is the mean distance between a sample and the nearest cluster that the sample is not a part of. A negative silhouette score indicates that some points don’t belong to the right cluster. A score close to zero indicates that the clusters are almost overlapping.

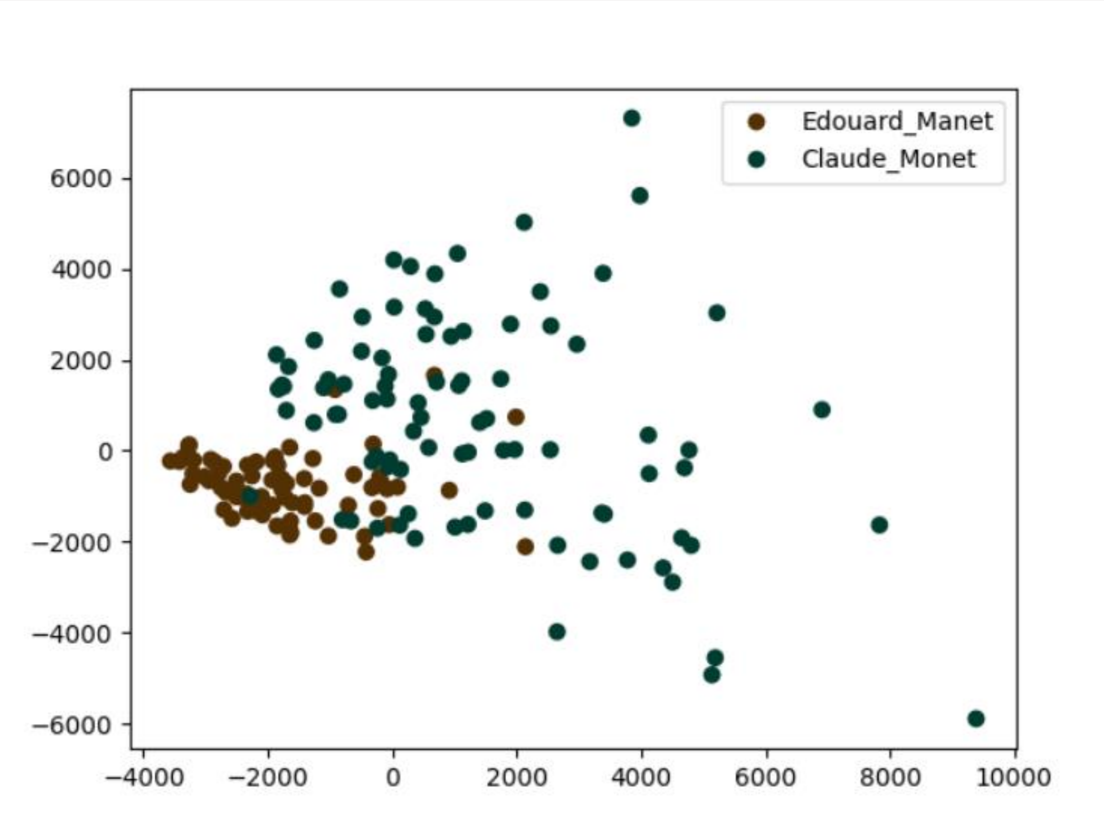
Experiments (1)

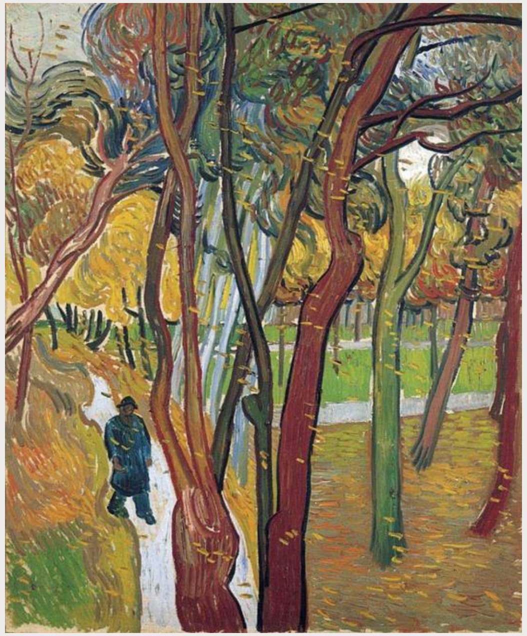
As first experiment, we evaluate the clusters of two similar artists, like Claude Monet and Edouard Manet, both impressionists. We expect their clusters to be very close.

On the right, a painting by Manet (on top) and one by Monet (on bottom).

Experiments (2)

The silhouette score obtained for the pair Monet – Manet is **0.05**. We can observe how close the clusters are in the plot, where we can also notice how the paintings by Manet are more concentrated.





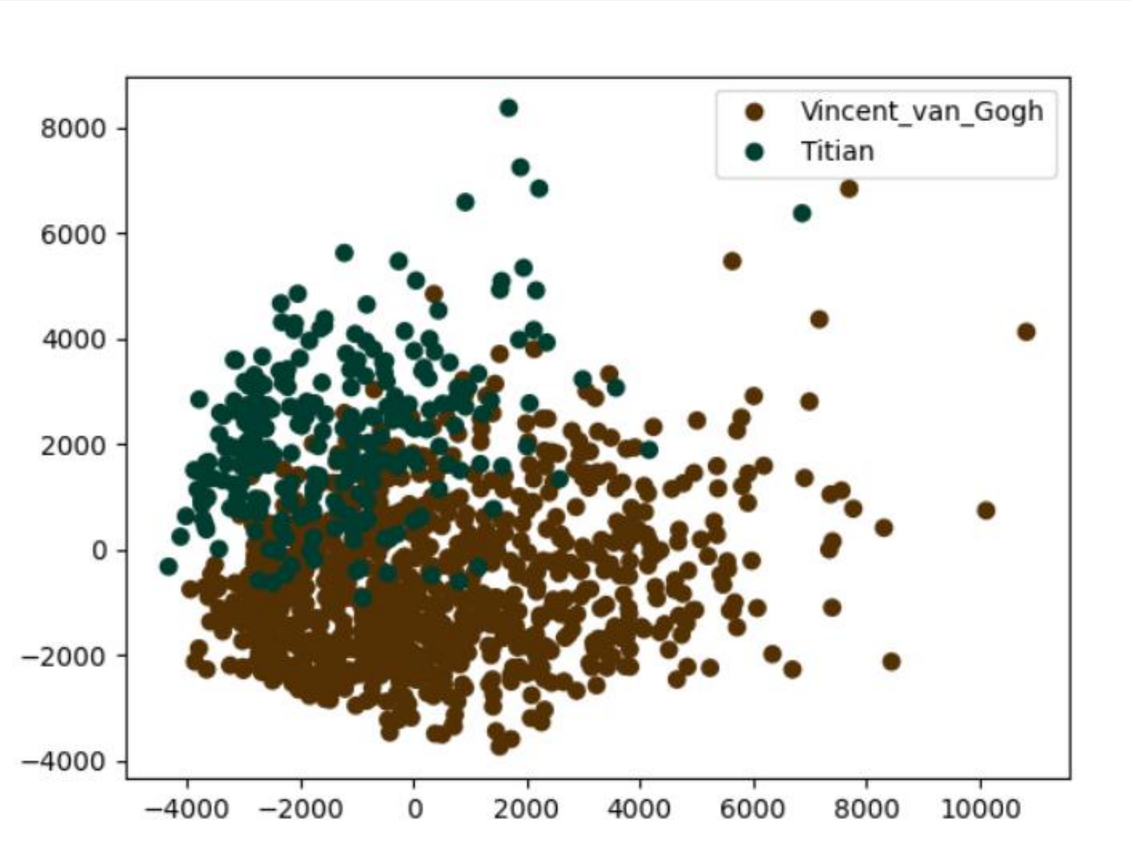
Experiments (3)

We now compare the clusters from two artists who don't have much in common: the Renaissance painter Titian and the post impressionist Vincent Van Gogh. We expect the silhouette score to be high and the clusters to be distant.

On top, a painting made by Titian, on the bottom a painting made by Van Gogh

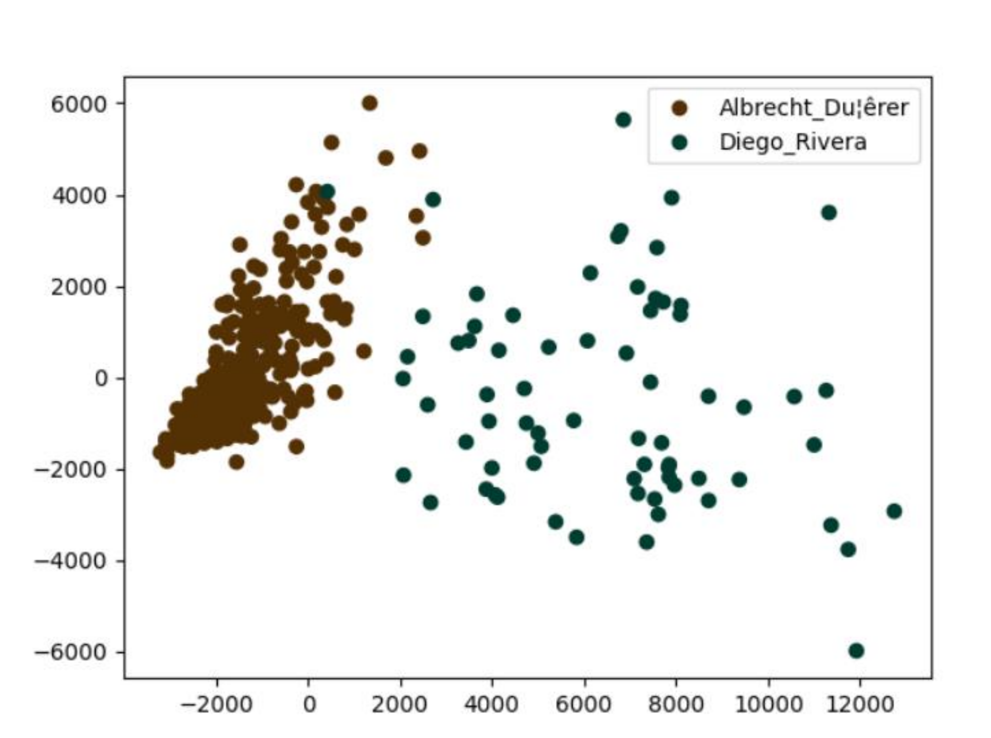
Experiments (4)

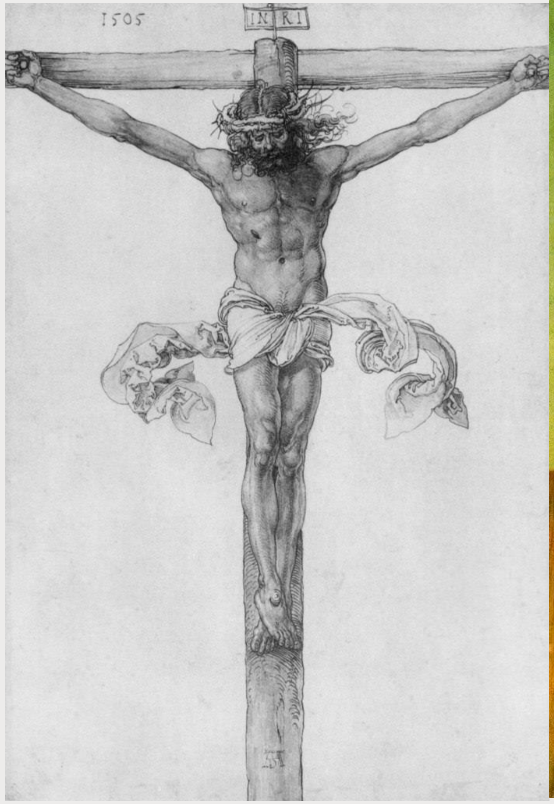
The silhouette score computed for the couple Titian – Van Gogh is **0,05,** much less than expected. The clusters have a large intersection.



Experiments (5)

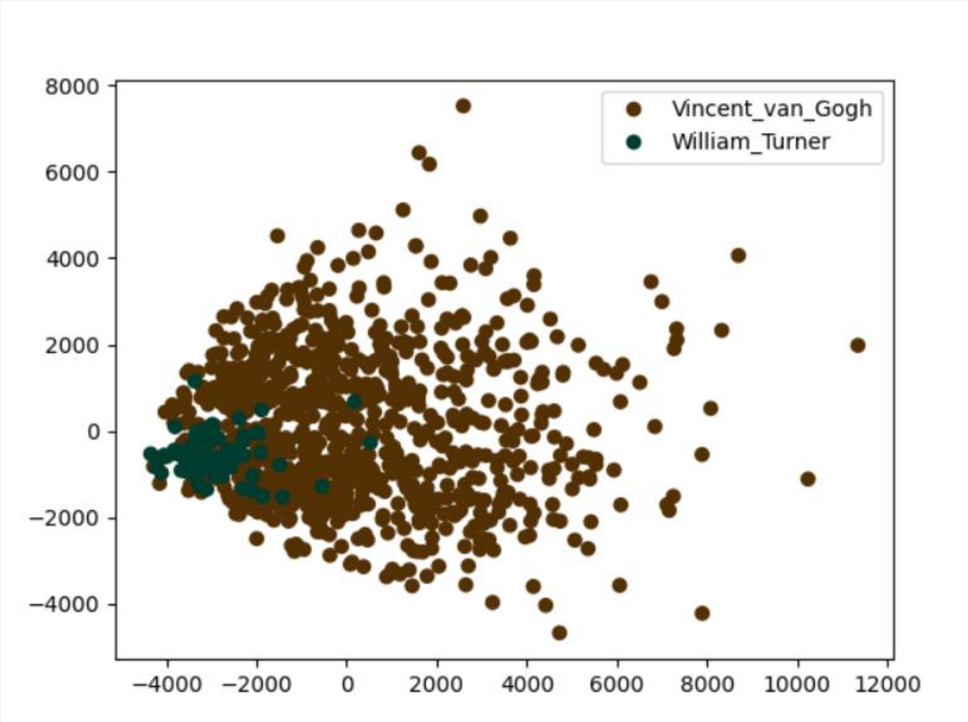
The great majority of the computed silhouette scores are small. The highest obtained value has been equal to 0,45, computed for the couple Albrecht Durer – Diego Rivera, who indeed are artists whose paintings don’t seem to have anything in common.

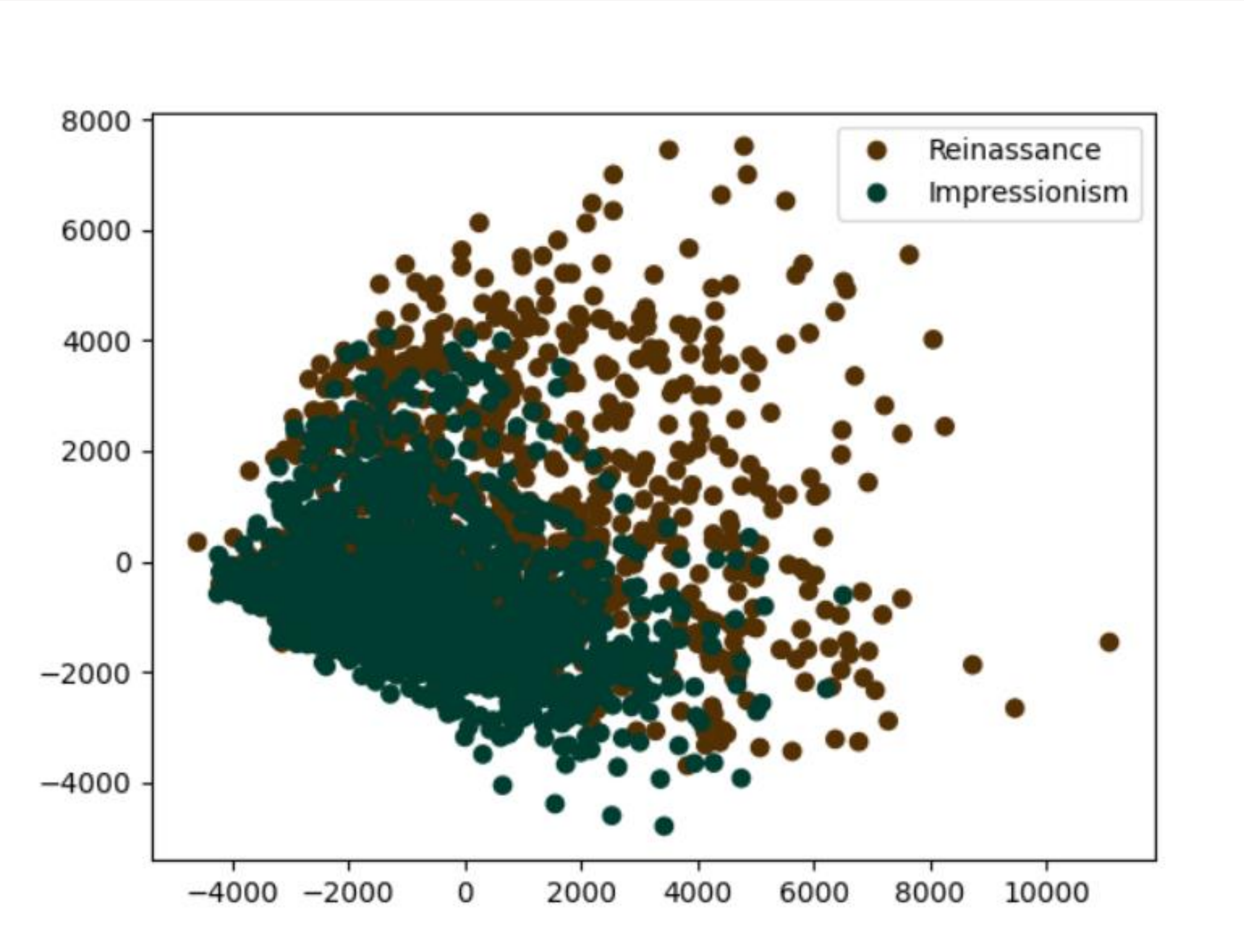




Experiments (6)

Also, several of negative silhouette scores were obtained. The lowest has been computed for the couple William Turner – Vincent Van Gogh, and is equal to **-0,13**.





Experiments (7)

We can also compare clusters of paintings belonging to the same movement. In the example, we compare Impressionism and Renaissance.

We expect the two clusters to be well separated, but, as we can see, this is not the case. The silhouette score is **0,04**.

However we can notice how the Renaissance cluster is wider. In fact this movement lasted 200 years and saw the birth of several sub-movements which had some differences the one from the other.

Conclusions

We can conclude that the new feature space failed in maintaining the distances between clusters. This could have been caused by:

- The excessive reduction of the dimension of the features arrays
- The choice of the layers from which features were extracted. The last layers compute more general features, while the early ones detect the most fine-grained ones. Choosing these layers could have brought better results, but the starting dimension of the features would have been much bigger, making the training much slower.

Thank you for your attention!