ISBN: 978-81-959356-3-5
IIP Proceedings, Volume 2, Book 18, Part 1, Chapter 4

IMAGE CLUSTERING BASED ON GENRES USING AUTOENCODER

Abstract

Image Processing is popular in the 21st century and plays an important role in many industries. This paper mainly focuses on clustering the images using deep learning algorithms. When I start looking for fast and smart analysis. Artificial Intelligence algorithms are major in solving the problem and predicting images by using machine learning and deep learning models. By using algorithms, we can minimize human error in detecting the solution for relevant images through models by learning the automation technique. This can easily detect and predict the clustering of images based on genres of information of images. The main technique used in this paper is Convolution Neural Network encoding and decoding of the image using an auto encoder algorithm. Also performed image reconstruction data augmentation and deep image clustering of performance and analysis of each genre.

Keywords: Convolution Neural Network, K-Means, Cluster, Genres, Deep Image Reconstruction, Data Augmentation

Authors

Abhishek Barandooru Janavejirao
Department of Physics & Informatics
University of Naples Federico II
Naples, Italy
a.barandoorujanavejirao@studenti.unina.it

Alberto Di Maro
Department of Physics & Informatics
University of Naples Federico II
Naples, Italy
albe.dimaro@studenti.unina.it

Pietro Soglia
Department of Physics & Informatics
University of Naples Federico II
Naples, Italy
p.soglia@studenti.unina.it

NOMENCLATURE

Acronyms

ML - Machine Learning

AI - Artificial Intelligence

DL – Deep Learning

UL – Unsupervised Learning

NN – Neural Networks

ANN – Artificial Neural Network

CNN – Convolution Neural Network

I. INTRODUCTION

Science is the way we can more access to save the life of people through the more advanced technology of using Robotics, Artificial Intelligence, Machine Learning, Deep Learning, and many other new computing software and technologies [1]. All these technologies which developed in the deep image field are drastic changes when we compare with the 20th and 21st centuries of past 50 years made it more accessible for changing rapidly. The contribution of ML in painting photo and image science is vast in many fields. Although, ML has developed many algorithms. ML algorithms have succeeded in many oppressive things worldwide. ML algorithms are generally mixed up with mathematics and logic that can easily predict from a given dataset. ML algorithms are very accessible to represent complex datasets by applying ML algorithms. As research was performed on ML algorithms in an image for predicting the optimal solution for the problem [2].

DL NN is a computational processing system that is heavily inspired by the way biological nervous systems (such as the human brain) operate. NNs are mainly comprised of a high number of interconnected computational nodes (referred to as neurons), which work entire in a distributed fashion to collectively learn from the input to optimize its final output [3]. AI, as a machine-based technique with algorithmic power for making predictions, diagnoses, recommendations, and decisions, has grown rapidly. Fostering trust in AI systems is a tremendous obstacle to bringing the most transformative AI technologies into reality, such as the large-scale integration of ML Images.

In image Clustering, it is a grouping of data not like classified. Based on the assumption of data it is grouped [4]. DL is the subset of ML. It mimics the human brain in the NN. We used a multi-layered NN which helps in training the neural network [5]. DL aims to develop a model that matches the level of the human brain in solving complex problems. CNN allows us to encode image-specific features into the architecture, making the network [6].

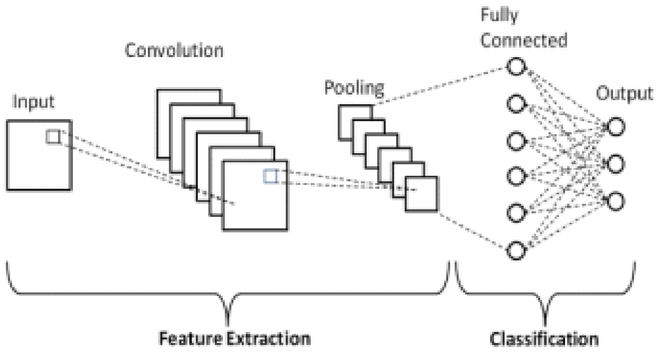


Figure 1: Neural Network

II. RELATED WORK

In this paper image clustering using genres, the main purpose is to make the grouping up of images in clustering form using K-means. When we started working on a research project using different ideas of image processing there are several advantages in clustering in real world. like business, anomaly detection, image search similarities, pattern recognition, marketing research based on previous images, network analysis, search result grouping, medical imaging process, fraud detection. In generally we researched and get feed on all images mixed on many genres not only in particular images.

To get a high level of the process we deeply studied on image processing techniques and also focused on some journal and research paper related to our topic. Some of the papers which is particularly helped in research the final attempt of analysis. Here some research paper extracted and explained to help in further process for better understanding deeply.

CNN using image painting identification. Each image performed various operational techniques such as lens distortion, color deformation, projection, translation, scaling, rotation etc., final evaluation is performed based on Scale-Invariant Feature Transform (SIFT) and CNN based. Test error variates huge between 15.6% and 2% in test error. By this shows which show competent on CNN [7].

In this paper, they proposed very important automatic learning techniques algorithms for learning and clustering the digitalized painting images with embedding for better accuracy reasons. It is not easy when human visual compared with computer vision in programming there is lot of errors in machines. Due to train and testing without suitable and relevant algorithms. starting from the deep convolutional embedding clustering (DCEC) model introduced, they propose DCEC-Paint as a method for grouping digitized paintings in an unsupervised fashion. Clustering is completely on distance measures in the highly multi-dimensional raw pixel space is well-known to be completely ineffective in results [8].

This survey paper demonstrates and carries help to understand deep in datasets and style of genres which follows the tradition of art data which enables performs the recognition and performs with many comparisons of 8 different architecture and in that 3 were never computed before for style [9].

During the analysis of this paper, they proposed deeply in art style recognition in a digital image in various ways. The baseline model is based on a pre-trained CNN, the VGG16 network and an SVM classifier. Trained and evaluated the model on pairs of painting styles using VGG16 and SVM model. Performed the 3 set of datasets which results in good way for better accuracy [10].

III. PROPOSED SYSTEM

1. **Dataset:** Dataset used in this analysis for the model is almost 102K images from the database wiki art and Kaggle. Which consists of 136 styles and a frequency of 10K for each style more or less

2. **Methodology:** Before moving to model of architecture here we analyzed deeply using exploratory data analysis for the image data to understand and filtering of unwanted image is necessarily required to process.

Style analysis

	style
count	102264
unique	136
top	Impressionism
freq	10643

Figure 2: Style Analysis

In this data analysis, we can easily understand the total number of image data is 102264 with 136 unique styles and the topmost data is impressionism.

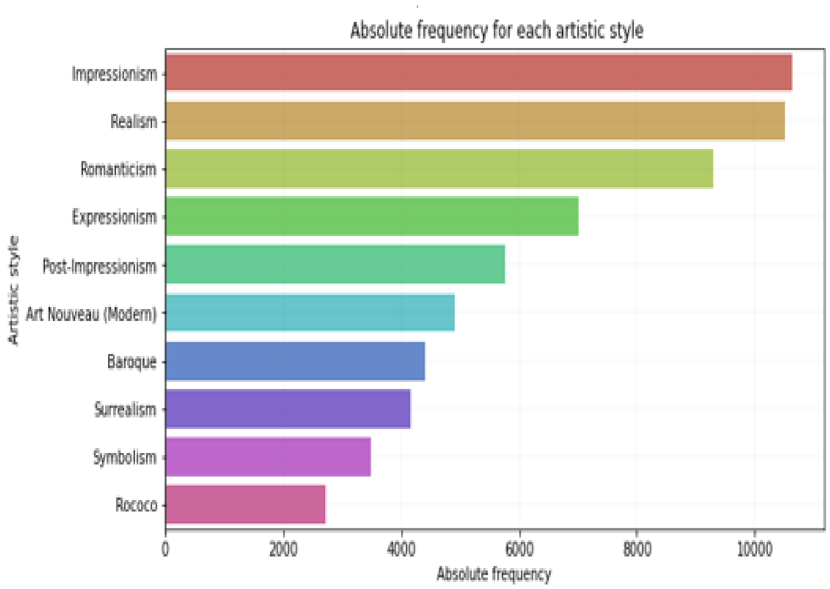


Figure 3: Frequency Analysis of style

136 style is too much to handle the data to computing for cluster analysis. So we chose the top 10 styles of data which is composed of 61% of the top total data

```
The percentage of paintings belonging to the top 10 styles is: 61.0%
The total number of paintings belonging to the top 10 styles is: 62917
```

Date analysis

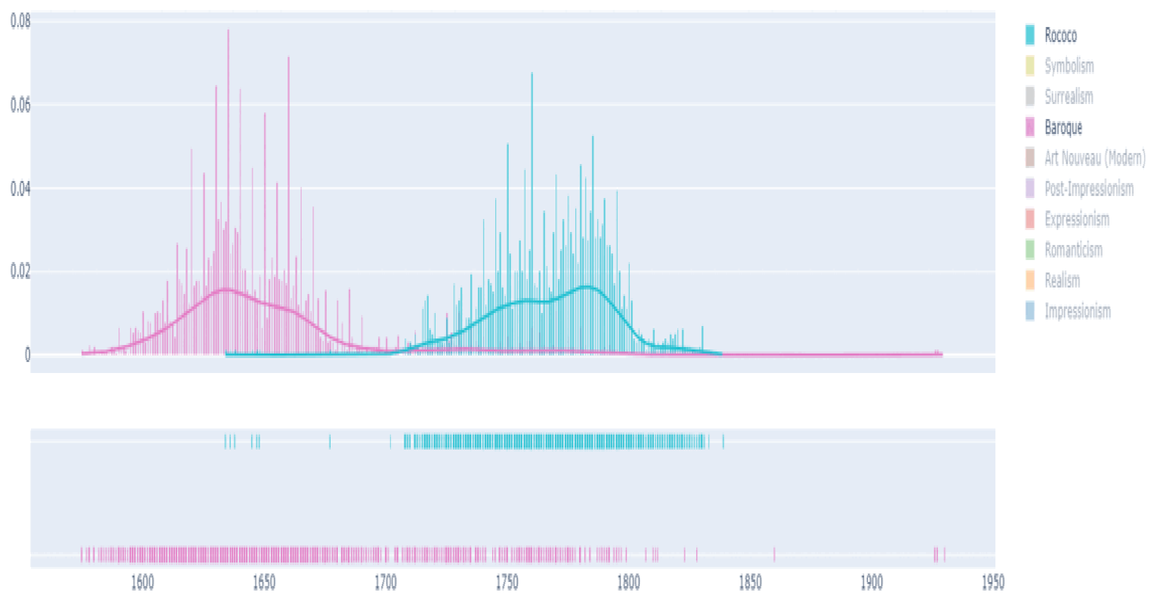


Figure 4: Style vs Year graph Analysis

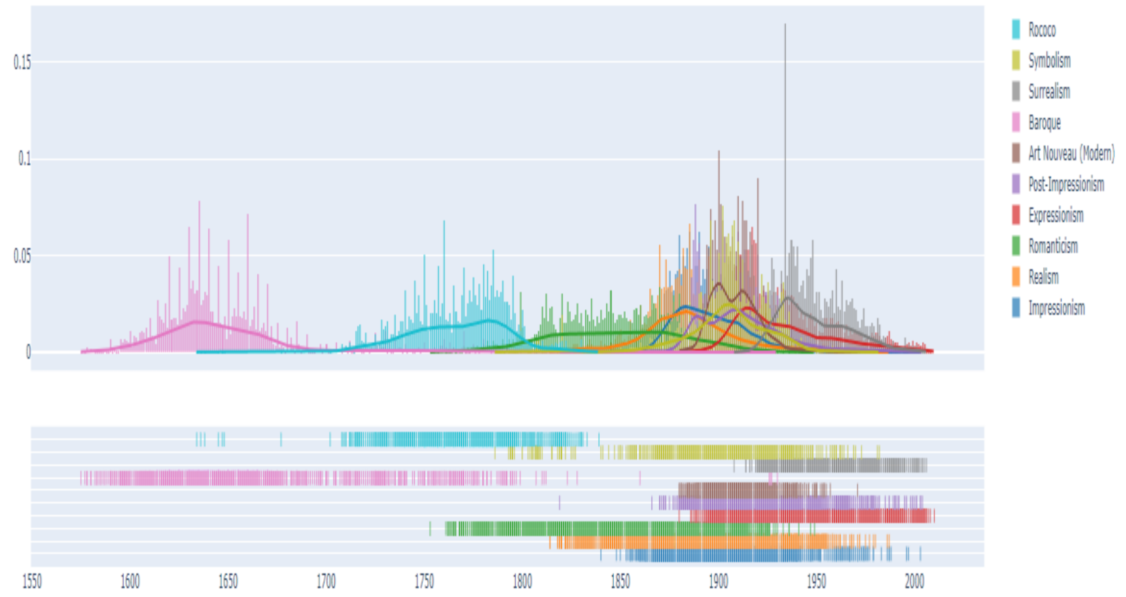
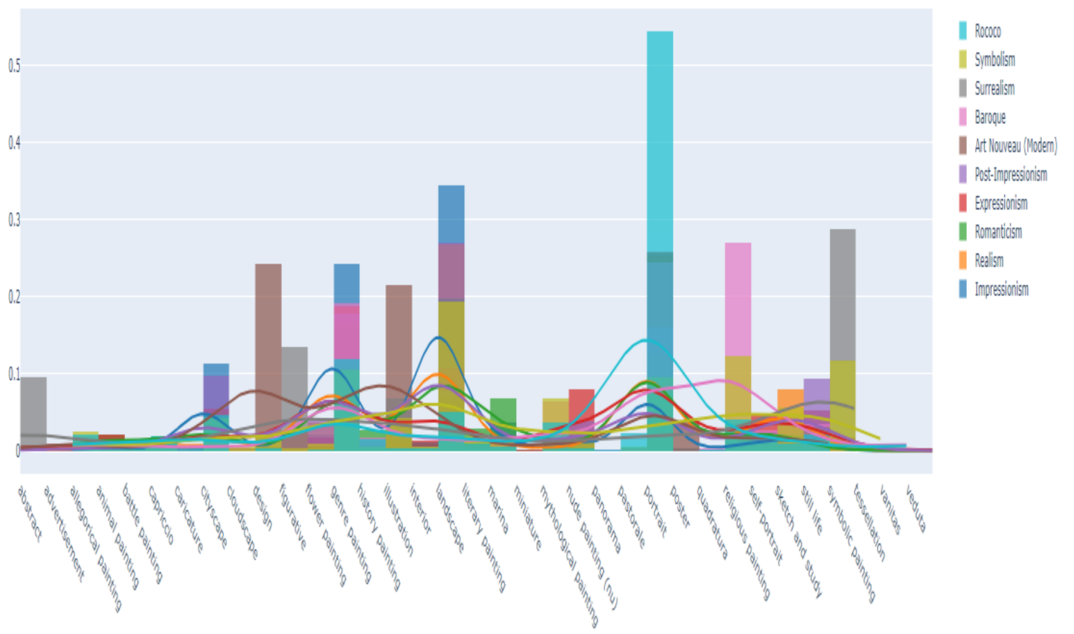


Figure 5: Year comparison of more image data present between 1880 to 1980 of 100-year distribution

Genre analysis



	genre
count	62706
unique	36
top	portrait
freq	12300

Figure 6: Genre Analysis

In this graph we observed carefully the data is not distributed equally it is imbalanced in each genre which contains of total 61% when we compare it with the original data of 100%. The total count of data is 62706, unique data is 36 types, the portrait is the top genre, and the distribution frequency of this data is 12300 which is highly present in portrait.

Pixel distribution analysis: Here in the below graph, the pixel distribution is easily analyzed and we can observe that region in 500 – 3000 in both the x and y axis is densely distributed

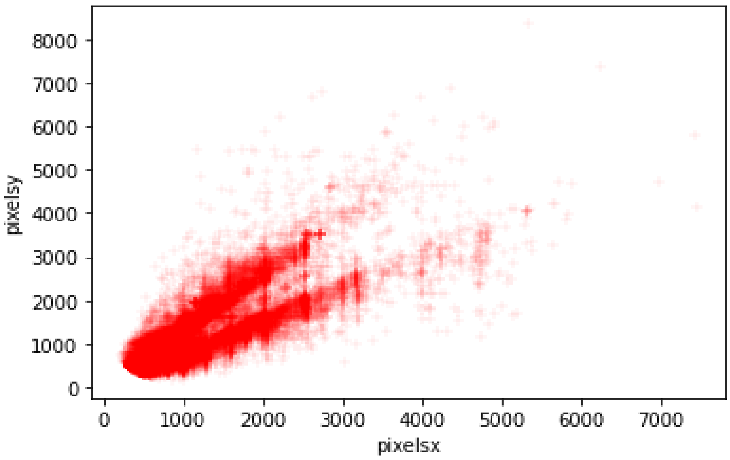


Figure 7: Pixel Distribution

	pixelsx	pixelsy
count	62706.000000	62706.000000
mean	987.697445	994.880474
std	567.324929	595.113339
min	205.000000	170.000000
25%	633.000000	635.000000
50%	838.000000	820.000000
75%	1114.000000	1088.000000
max	7459.000000	8351.000000

Figure 8: Pixels on X-axis and Y-axis

Colors representation and Chi-Square distance: We used chi-square distance as a statistical method to analyze the image of each style. Chi-square distances is finding the distances between 2 arrays in terms of image it will go with the dimensionality of the array for distance calculation. Chi-square is mainly used for similarity of image retrieval, image texture, feature extractions and many other purposes. Here below I mentioned the chi-square formula.

$$X^2 = \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i - y_i)^2}{(x_i + y_i)}$$

All images are represented in in grayscale image format for better computing results. as know that grayscale image which is fat and better computing when we compare with other format of images in image processing

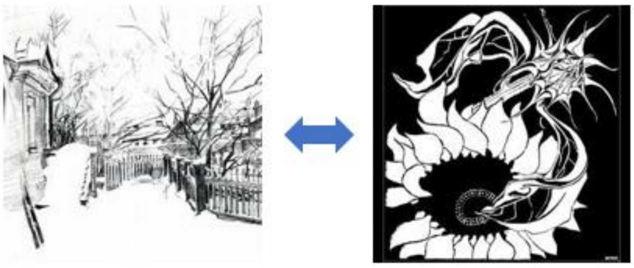


Figure 9: Grayscale Image

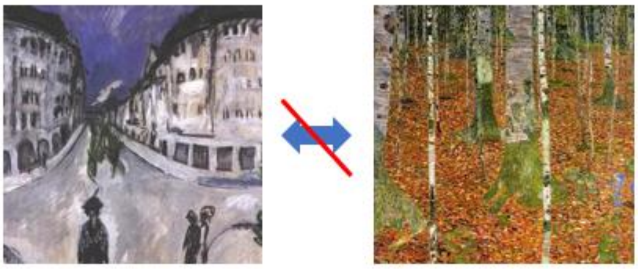


Figure 10: Colour Image

Intra-style distance in terms of color

```
The average ChiSqaure-distance of the style: Art Nouveau (Modern) is: 1017098.95
The average ChiSqaure-distance of the style: Baroque is: 111381.10
The average ChiSqaure-distance of the style: Expressionism is: 7032685.27
The average ChiSqaure-distance of the style: Impressionism is: 706939.05
The average ChiSqaure-distance of the style: Post-Impressionism is: 497990.55
The average ChiSqaure-distance of the style: Realism is: 1525426.65
The average ChiSqaure-distance of the style: Rococo is: 856532.93
The average ChiSqaure-distance of the style: Romanticism is: 934485.68
The average ChiSqaure-distance of the style: Surrealism is: 374415.86
The average ChiSqaure-distance of the style: Symbolism is: 709255.42
```

Here chis-square distance is computed between styles as I mentioned in the above results of each style having different distances based on the style. And we observed carefully in some of the styles are slightly similar because of the similarity of images style. Even though we didn't used grayscale image and we got better results

Pre-processing techniques in image: For any data in analyzing, training, or testing before processing it is the basic steps in all formats of data like an image file or csv it is preliminary to do process. To increase and make it proper design of data to analysis it is most important. Image should be in resizing, orienting, color correctness like gray scale and many other augmentation techniques. But in our preprocessing of image data, we carried 2 preprocessing techniques

For huge images: There is large image of size for that specific image we cropped and in below figure it shows in sign technique for better understanding of how we process the data in large size and we didn't lose the data by cropped into specific part.

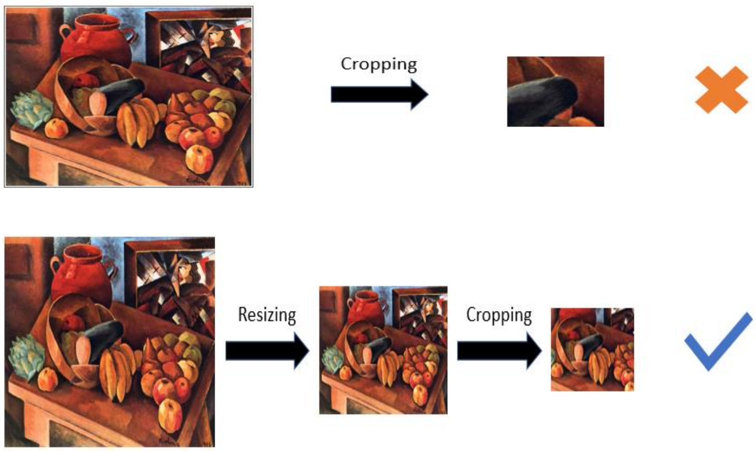


Figure 11: Resizing and Cropping Technique of Image

For portraits: For portraits image data we don't require all the images and for that style image we recognize the face part and crop with descent size for pre-processing the data

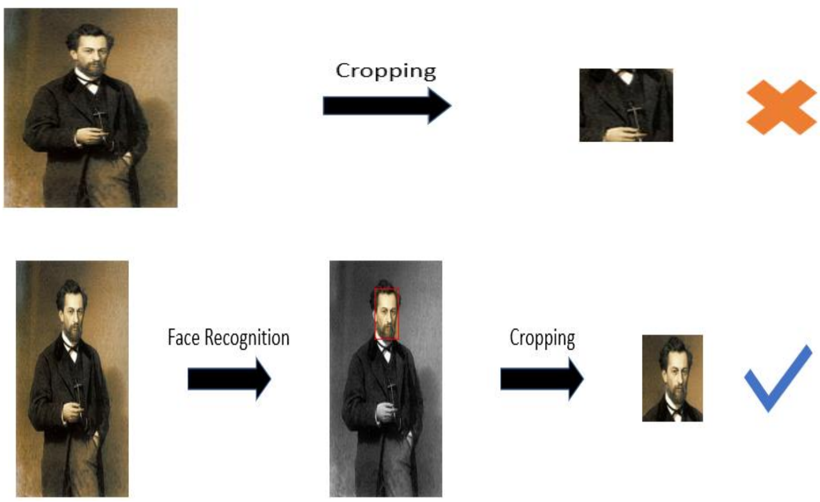


Figure 12: Cropping Technique used for Portraits

Pre-processing techniques in color: It simplifies the algorithms and reduces computational requirements like GPU and less RAM for processing. Due to dimensionality reduction of the color image to the grayscale image, in which RGB contains 3 color and 3 channels and 3 dimension. But in grayscale image are single dimensional, which reduces model complexity. For model evaluation learning in image processing becomes faster and easier for NN and computational power.

Figure 13: Color to a Grayscale Image

3. **Image reconstruction:** Autoencoder techniques use encoding and decoding of input and output data to reduce the noise and dimensionality reduction.

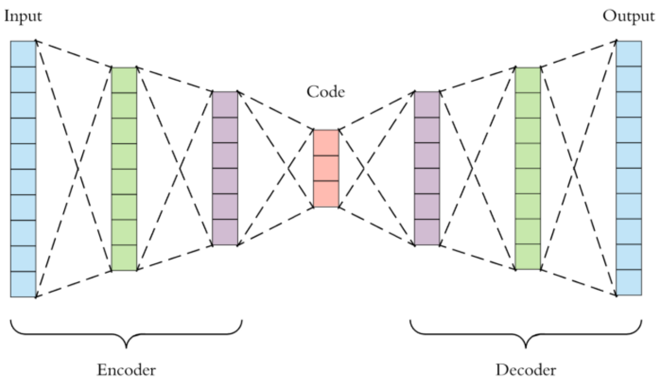


Figure 14: Autoencoder (Encoder & Decoder)

It is an NN of unsupervised data which is mainly compromised of function and nodes and neurons in the network to compute better perform similar to reconstruct the same image in the output of encoding side in autoencoder NN reduction features. An autoencoder compresses the input into a lower-dimensional code, obtained through a data-driven non-linear mapping. Then, it reconstructs the output from this representation [11].

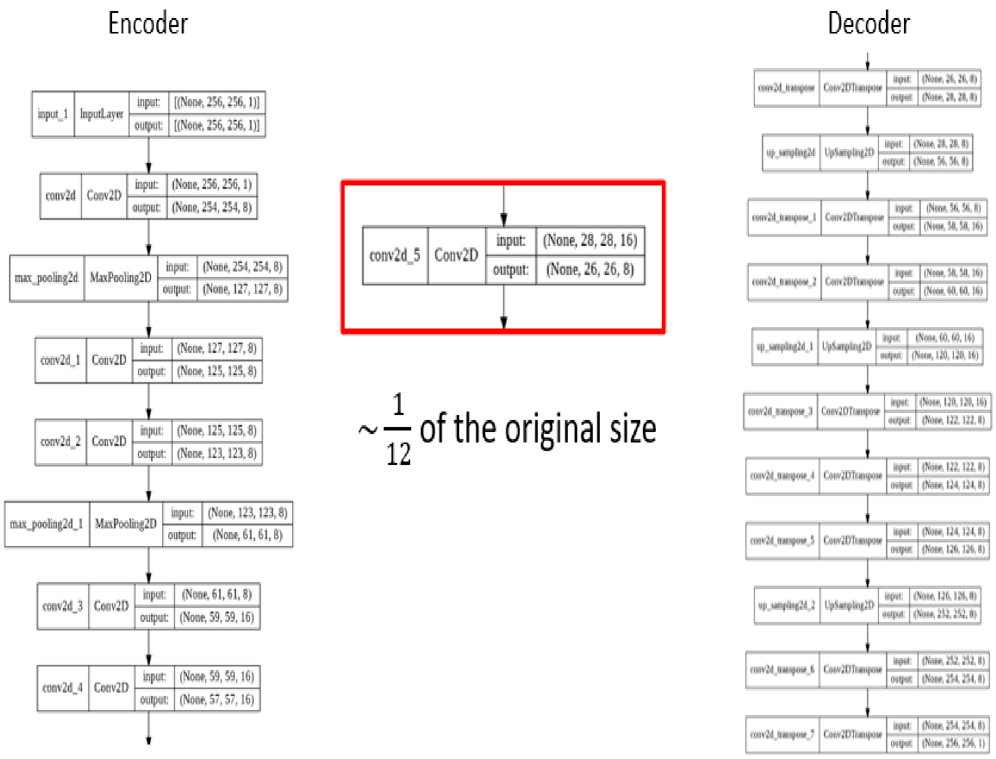


Figure 15: Architecture of Encoder & Decoder with 1/12 of the Original Size Reduced

Training and validation loss per epoch: Training and validation are performed over 140epoches and finally, it values loss of 0.0018 and image reconstruction results in better results more than 99.99%

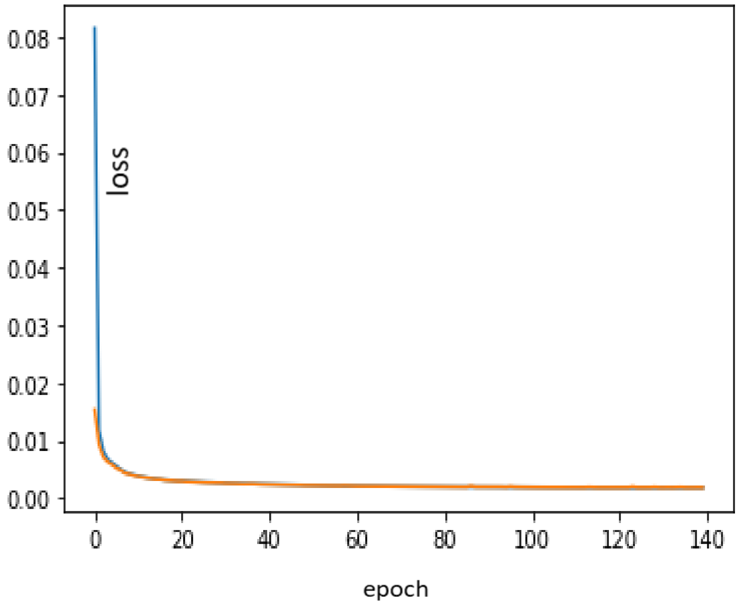


Figure 16: Training & Validation

ISBN: 978-81-959356-3-5
IIP Proceedings, Volume 2, Book 18, Part 1, Chapter 4

Final values:

loss: 0.0018 - val_loss: 0.0018

Results: images reconstruction

Figure 17: Image Reconstruction after Training

Performance evaluation for each genre

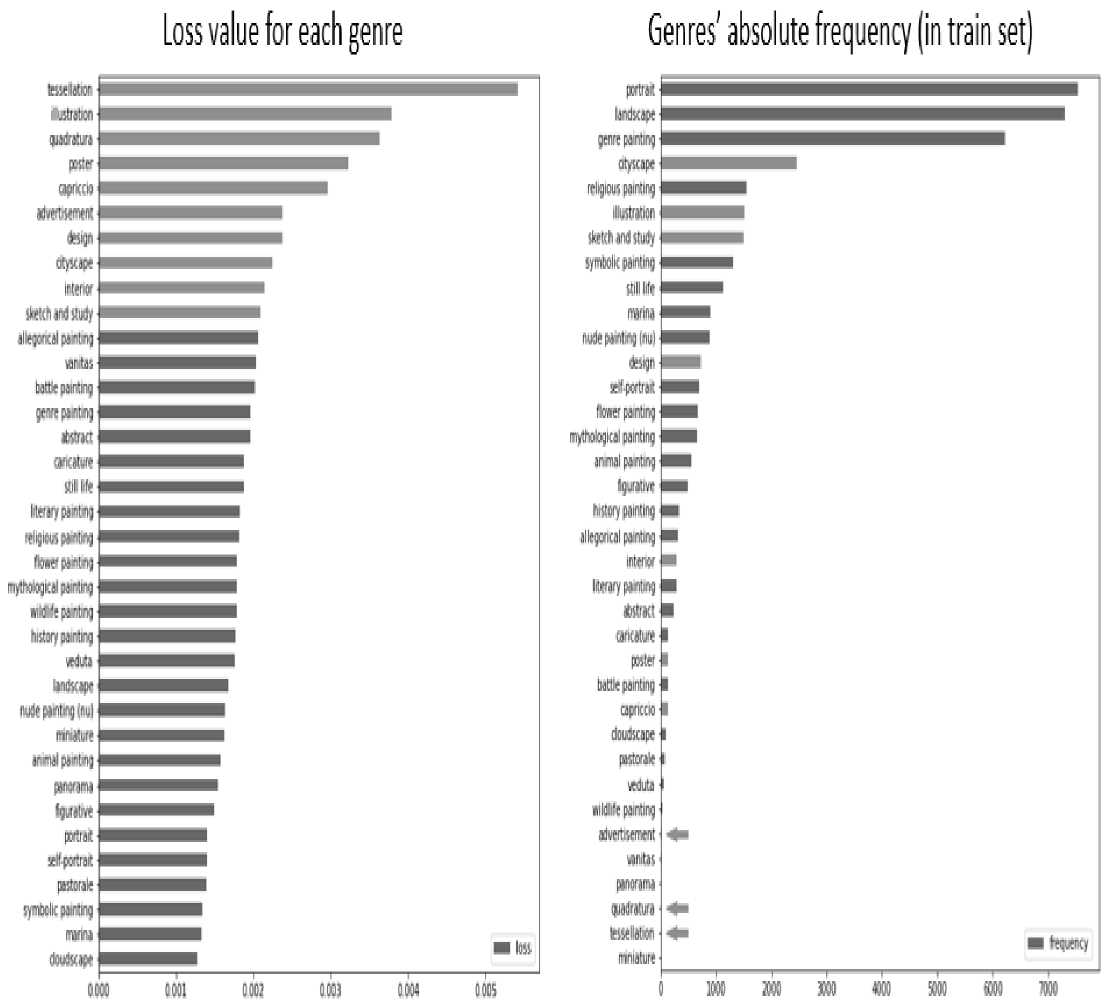


Figure 18: Graph Comparison of Loss Value in Genre & Genre Frequency in the Train Set

The model performs well on several genres like portrait or landscape, whereas it seems to find more difficulties in extracting features from genres like:

Tessellation
Quadratura
Capriccio
Poster
Illustration
Advertisement
Design
Cityscape
Interior
Sketch and Study

Note that the loss for each genre could be explained by looking to the quantity of images the train takes from each genre. Maybe, the images that are not reconstructed so well are the ones with less representatives in the train set.

The imbalance of the dataset could affect the model performances. Some genres showed good performances in terms of loss results, even if with few representatives. It could be due to the fact that low level features, learned on more represented genres, are capable to generalize on them too.

On the other hand, the set of genres identified before (i.e. tessellation, sketch and study, advertisement, quadratura, capriccio, poster, illustration, cityscape, design and interior) seem to be less represented in the train set, as expected.

Thus, an augmentation is used to increment the number of these less represented genres. In particular, the three genres illustration, sketch and study and cityscape will be independently augmented up to 5000 images. With regard to the other five genres, they will be collected in a sub-category (namely, "others") and this entire set will be augmented up to 5000 images. The choice of grouping these 5 genres into a single category is due to the fact that they have really few representatives by their own. Thus, augmenting them independently could lead to overfitting.

Augmentation on Critical Genres

Train and validation loss of critical genres

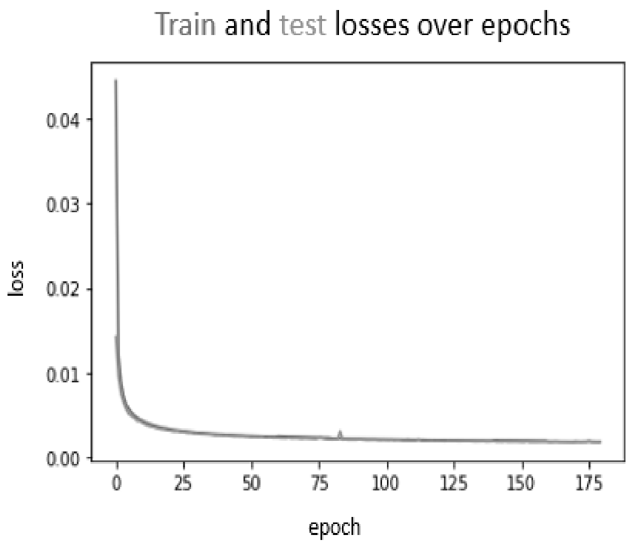


Figure 19: Train & Test for Critical Genre

Finally, let's check out the effects of the augmentation on the specific genres.

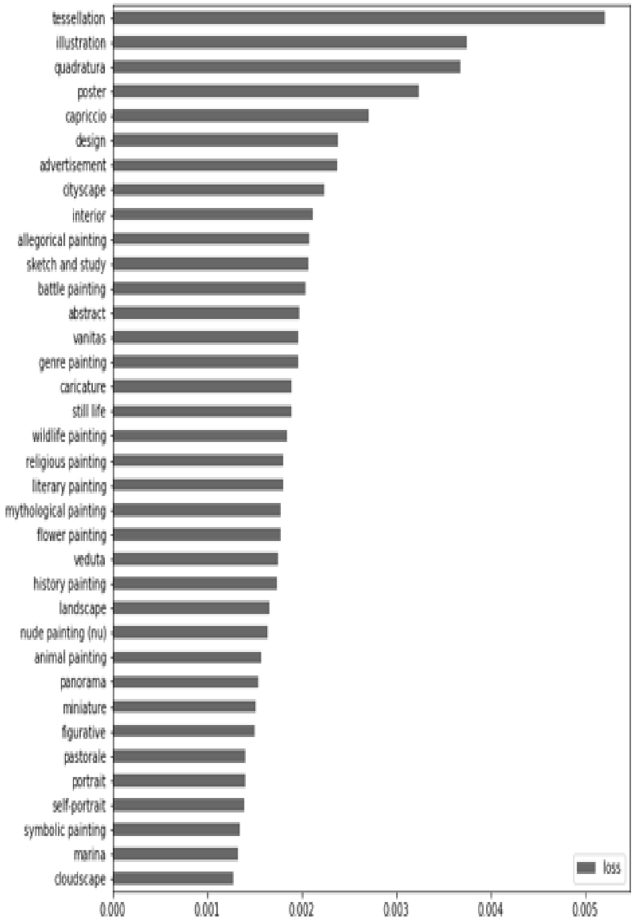


Figure 20: Augmentation for a Specific Genre

The augmentation did not improve the extraction of relevant features in the whole dataset.

IV. RESULTS

Deep clustering

A Deep Embedding Clustering model consists of a Convolutional Autoencoder in which a specific Clustering Layer is attached to the output of the Encoder. The parameters of this neural network are learnt by using a joint optimization. Particularly, the model tries to minimize two different loss functions at the same time, namely, the Squared Mean Error (for reconstruction) and the Kullback–Leibler Divergence (for clustering).

In order to perform the clustering task, the weights of the Clustering Layer need to be initialized as the Centroids obtained by a pre-computed K-Means algorithm executed on the encoded output of the images.

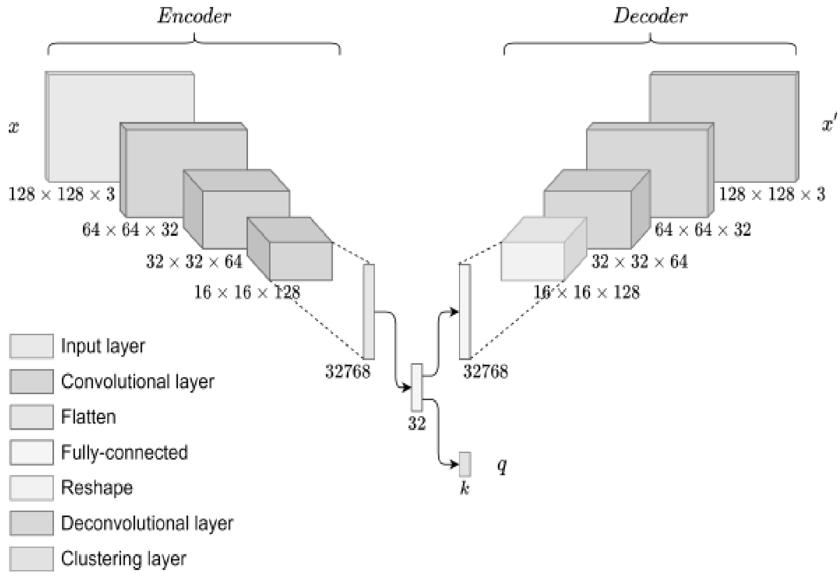


Figure 21: Autoencoder Architecture for Clustering

For the following part of the result, we go back working on the original non-augmented data. Thus, we delete the augmented dataset and its related train and test sets in order to re-upload the original folders on the disk. This model needs to be trained without the iterative slicing performed before, since it would badly affect the performances. Therefore, we need a lighter version of the dataset in which images are of size 128 x 128.

The general workflow will be as follows:

1. Instantiate and train Autoencoder_v2, saving its weights.
2. Extract the Encoder part to create the encoded version of our images.
3. Use these encoded images to train a K-Means algorithm and extract the estimated centroids.
4. Instantiate a Deep Embedded Clustering (DEC) model (Autoencoder + Clustering Layer);
5. Initialize the Autoencoder_v2 weights with those calculated in point 1;
6. Initialize the centroids with those retrieved in point 3;
7. Train the Deep Embedded Cluster model.
8. Show the results.

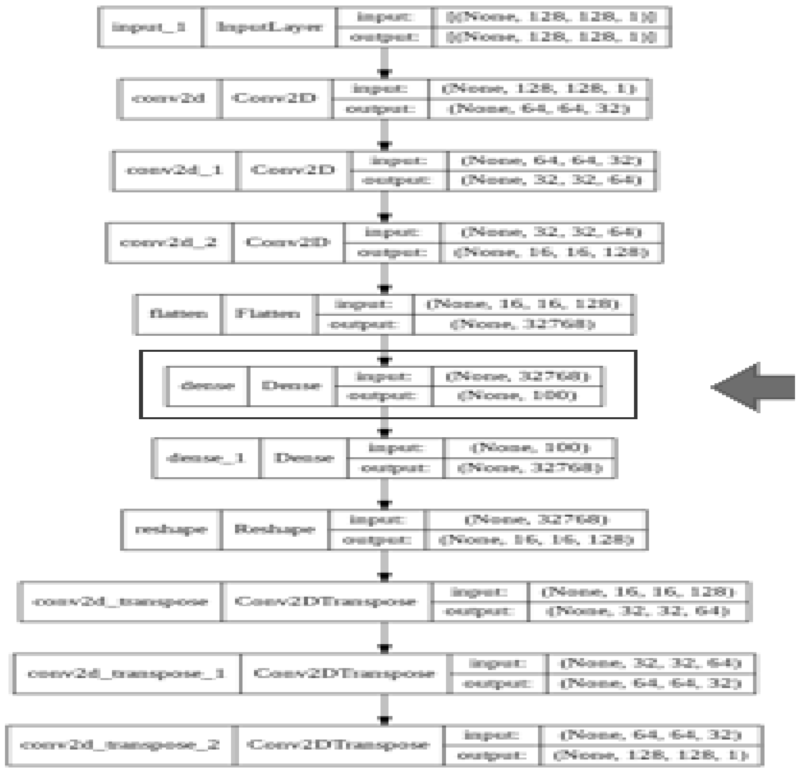


Figure 22: 1/163 of The Original Size

Performance results

```
188/188 [==============================] - 5s 21ms/step - loss: 0.0042
0.004236379638314247
```

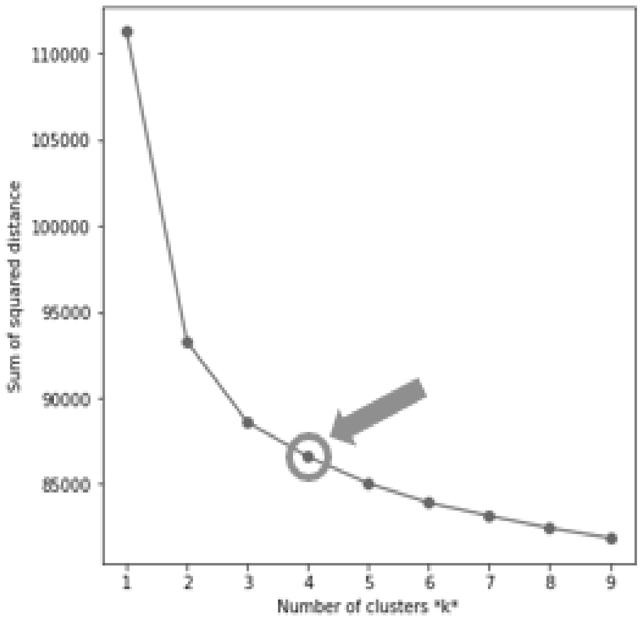


Figure 23: Elbow Method for K = 10 Clusters

We opted for a domain-based approach for the selection of the number of clusters. However, there is a more statistical approach to choosing this number, namely, the Elbow Method.

The Elbow Method gives us an idea on the choice of the number of clusters k, based on the sum of squared distance (SSE) between data points and their assigned clusters' centroids. We pick k at the spot where SSE starts to flatten out and form an elbow.

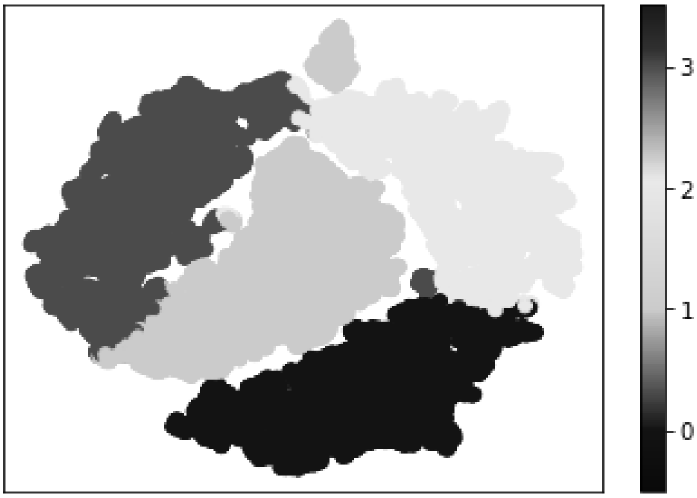


Figure 24: Cluster analysis for K=4

```
For n_clusters = 4, Deep clustering silhouette score is 0.8002052307128906)
```

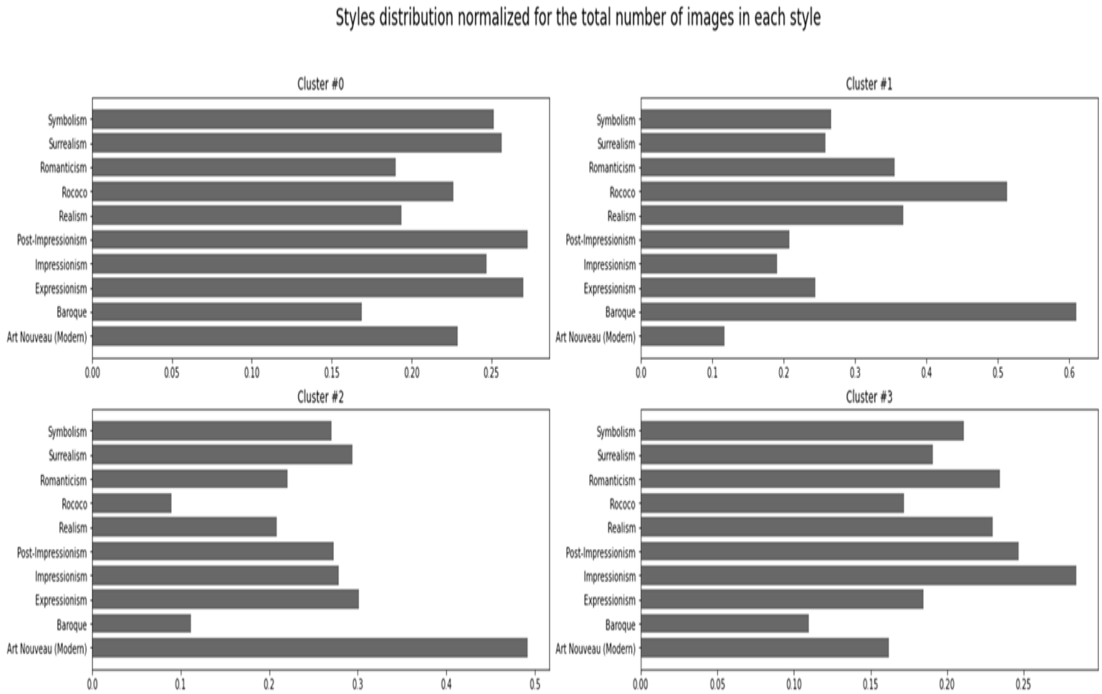


Figure 25: Style Distributed a Total Number of Images Based on Style

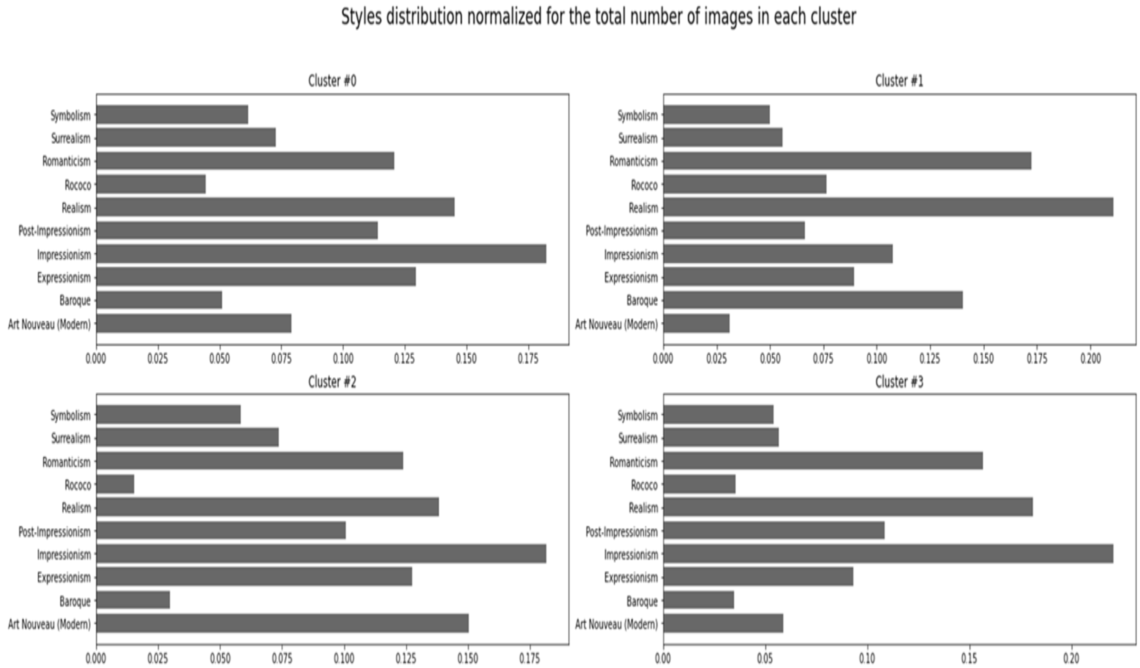
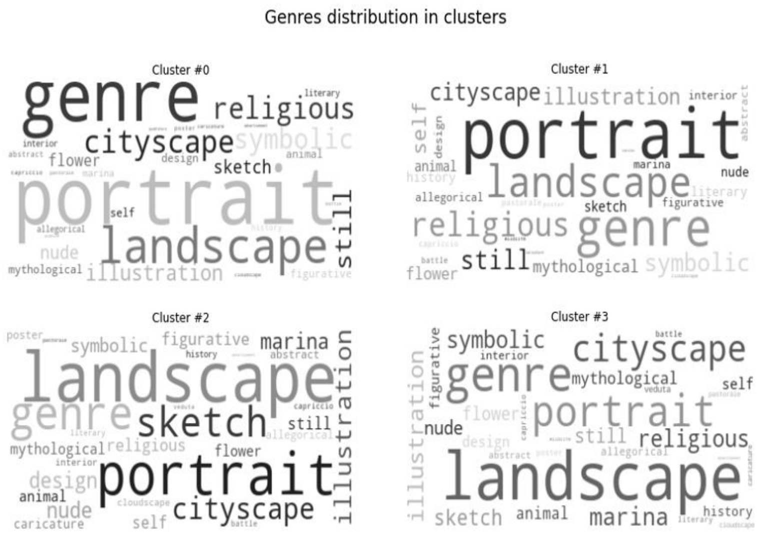
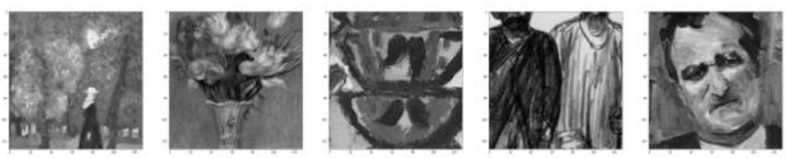


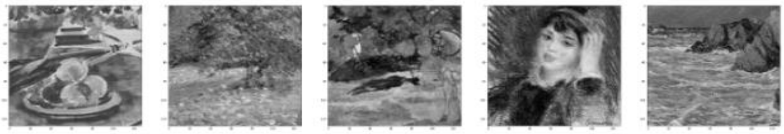
Figure 26: Style Distributed a Total Number of Images on Each Cluster

Genres distribute among the clusters using world cloud



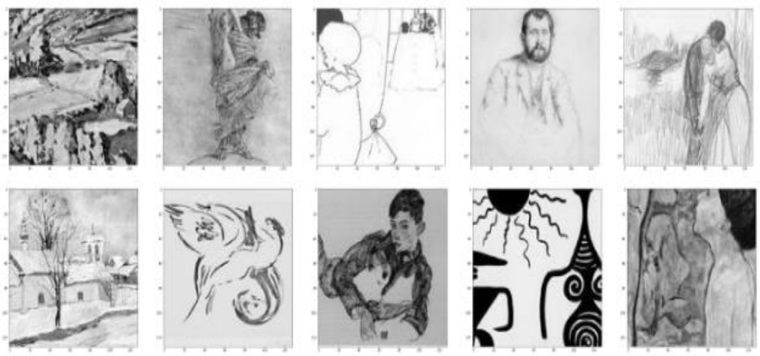
Cluster 0: Particular Brushstrokes



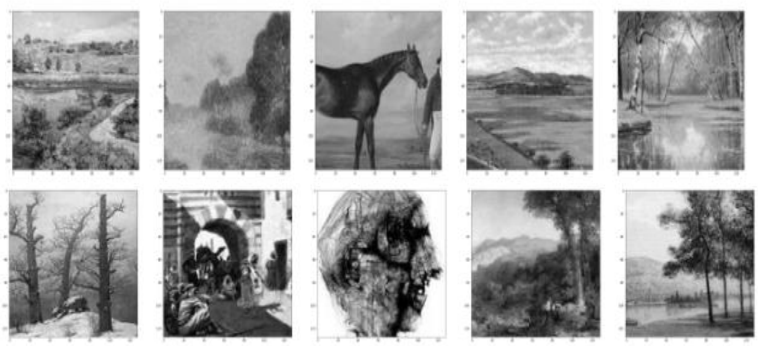


Cluster 1: Faces

Cluster 2: Drawings and Sketches



Cluster 3: Open Spaces



V. CONCLUSION

In this paper, we achieved clustering with an accuracy of 80% and data augmentation techniques which is having less loss in train and testing. The accuracy of the model is not 100 % for all the training and testing. We have to improve our model for more accuracy along with data augmentation accuracy. Machine Learning and Deep Learning play a massive role in clustering.

VI. FUTURE WORK

In this paper, we got good results. But we try to improve more accuracy in the next implementation. And also, we will work on implementing video streaming images to the cluster.

VII. ACKNOWLEDGEMENT

I would like to thank Prof., Giuseppe Longo Data Science Head, President of International IAIA and Department of Physics, and also Prof., Roberta Siciliano University of Naples Federico II for the implementation guidance, discussion and suggestions.

REFERENCES

[1] Munira Ferdous, Jui Debnath and Narayan Ranjan Chakraborty, “Machine Learning Algorithms in Healthcare: A Literature Survey”, *International Conference on Computing, Communication and Networking Technologies (ICCCNT),* IEEE – 49239, 2020
[2] Sarker, I.H. “Machine Learning: Algorithms, Real-World Applications, and Research Directions. *SN COMPUT. SCI”,* vol 2**,** 160 (2021). https://doi.org/10.1007/s42979-021-00592
[3] Jianxin Wu and LAMDA Group, “Introduction to Convolutional Neural Networks”, *“National Key Lab for Novel Software Technology”*, 2017
[4] M. Emre Celebi and Kemal Aydin, “Unsupervised Learning Algorithms”, “Springer Cham Heidelberg New York Dordrecht London”, 2016
[5] Suzana Herculano-Houzel, “The human brain in numbers: a linearly scaled-up primate brain”, “Frontiers in Human Neuroscience”, Volume 3, Article 31, 2009
[6] Keiron O’Shea and Ryan Nash, “An Introduction to Convolutional Neural Networks”, “arXivLabs Neural and Evolutionary Computing”, 2015
[7] Yiyu Hong, Jongweon Kim, “Art Painting Identification using Convolutional Neural Network”, “International Journal of Applied Engineering Research”, h ISSN 0973-4562, Vol. 12, Issue 2, pp. 532-539, 2017
[8] Giovanna Castellano, Gennaro Vessio, “Deep Convolutional Embedding for Digitized Painting Clustering”, “25th International Conference on Pattern Recognition (ICPR)”, ISSN: 1051-4651, Italy, January 2021
[9] Orfeas Menis, Natasa Sofou, Giorgos Stamou, “Deep Ensemble Art Style Recognition”, “International Journal of Applied Engineering Research”, ISSN 978-1-7281-6926-2/20, 2020
[10] Rasmus Hove Johnsen, Andrea Gradečak, “Deep Learning Approaches to Art Style Recognition in Digital Images”, “Department of Computer Science Aalborg University”, Denmark, 2017
[11] Kajaree Das, Rabi Narayan Behera, “A Survey on Machine Learning: Concept, Algorithms and Applications”, “International Journal of Innovative Research in Computer and Communication Engineering”, Vol. 5, Issue 2, February 2017