IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 20, 2021, accepted May 31, 2021, date of publication June 4, 2021, date of current version June 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3086476

Adversarial Learning Approach to Unsupervised Labeling of Fine Art Paintings

CATHERINE SANDOVAL, (Member, IEEE), ELENA PIROGOVA, AND MARGARET LECH, (Member, IEEE)
School of Engineering, RMIT University, Melbourne, VIC 3000, Australia

Corresponding author: Catherine Sandoval (s3465033@student.rmit.edu.au)

This work was supported in part by the Australian Government Research Training Program, in part by the Defence Science Institute Research Scholarship, and in part by the Defence Science Institute Research Agreement.

ABSTRACT An automatic classification of fine art images is limited by the scarcity of high-quality labels made by art experts. This study aims to provide meaningful automatic labeling of fine art paintings (machine labeling) without the need for human annotation. A new unsupervised Adversarial Clustering System (ACS) is proposed. The ACS is an adversarial learning approach comprising an unsupervised clustering module generating machine labels and a supervised classification module classifying the data based on the machine labels. Both modules are linked through an optimization algorithm iteratively improving the unsupervised clusters. The objective function driving the improvement consists of the within-cluster sum of squares (WCSS) error and the supervised classification accuracy. The proposed method was tested on three different fine-art datasets, including two sets of paintings previously categorized by art experts and one never categorized collection of Australian Aboriginal paintings. The unsupervised clusters were analyzed using standard unsupervised clustering metrics and a reliability measure between machine and human labeling. The ACS showed higher reliability compared to the classical k-means clustering method. The content analysis of unsupervised clusters indicated grouping based on scene composition, type, and shape of the object, edge sharpness and direction, and color palette.

INDEX TERMS Adversarial learning, art classification, data labeling, deep learning, digital humanity, optimization, transfer learning, unsupervised clustering.

I. INTRODUCTION

In the last years, the digital collections of visual artworks such as pictures of paintings, drawings, posters, prints, photographs, or daguerreotypes have dramatically increased. Online exhibitions, virtual tours, auctions, and sales are becoming more common and popular. It creates a growing need for new instruments to automatically perform a rapid large-scale analysis, categorization, recognition, search, and digitized art retrieval.

One of the biggest challenges found in the automatic analysis of art is the semantic gap between objective digital representation of artworks and complex subjective art concepts. The meaning of an image or a style labeling is based on human attributes of personal perception, sensitivity, and ideas. To make it even more challenging, these attributes must be supported by extensive training and expertise. This means that only a very few individuals around the world have ''*the license*'' to define what are the generally agreed styles or trends in art, and thus, only a very few experts can provide ''*ground truth*'' art labels. Such labels have been used to create research datasets to support the development of supervised deep learning techniques for an automatic art classification. These techniques can be highly reliable, but only if the classification system is trained on expert labeling [1]. Limited availability and the high cost of databases labeled by experts reduce potential applications of supervised classification.

The advantage of supervised learning is that computer software learns how to mimic human experts' aesthetic perception. Therefore, the supervised learning quality is assessed by measuring how close are the machine predictions to the human labels. Nevertheless, one can ask what would happen if, instead of relying exclusively on art experts, we would

The associate editor coordinating the review of this manuscript and approving it for publication was Tallha Akram.

like to rely either fully or partially on machines? What if we would like to ask the machines to provide their own labels or categories of art based on objective learning criteria? How to define these criteria? How useful would be such labeling? Would it be close to the human-made labeling? Would such closeness be important? These are some of the frontier research questions that need to be answered to merge machine learning with fine-art and humanities.

In this study, we propose a new unsupervised deep learning paradigm for the automatic labeling of fine-art paintings. We validate the proposed approach on three different datasets of fine-art paintings and demonstrate that it leads to labels that are close to human labeling based on scene composition, type, and shape of the object, edge sharpness and direction, and color palette. This means that the proposed approach can provide a useful tool for unsupervised labeling of fine-art paintings.

The remainder of this paper is organized into five sections: Section II presents a summary of related works. Section III describes the proposed methodology. The datasets used to validate the proposed method are described in Section IV. Experimental results and discussion are presented in Section V, and Section VI concludes the paper.

II. RELATED WORK

While the annotation of natural images is usually based on the content or the objects they represent (e.g., cat, apple, car), the annotation of digitized artwork is based on the meaning or other high-level artistic concepts. Digitized paintings are generally annotated according to their historical period, author, and artistic movement, commonly named ''style'' [2]. However, even for art scholars, art categorization can be a challenging task. It is due to overlapping characteristics of consecutive historical periods, stylistic inconsistencies of the same artists, the presence of unrelated artistic elements that do not belong to a specific period or style, and the ambiguity associated with the assessment of abstract and subjective visual features of art [3], [4].

Computer-based art analysis is an emerging field of research; however, the majority of current studies focus on the supervised classification of art images. The key factor contributing to the popularity of supervised learning is that the learning process is based on human-made expert labels; therefore, the automatic categorization gives outcomes very close to manual classification by humans. Due to centuries of tradition, there is high trust and general social acceptance of these labels. Unfortunately, the manual labeling process is prone to errors, expensive, and time-consuming. In addition, in cases of newly emerging artistic movements or rare cultural collections, the expertise required for the annotation process could not be available at all.

The latest deep learning technologies offer an alternative solution to the art labeling process in the form of unsupervised clustering (or categorization) that can be conducted in a fully automatic way, without the need for manual annotation by experts. We will refer to this process as ''machine-made'' labeling. An unsupervised classification system can organize a set of images into clusters by discovering new patterns or relationships that are not necessarily apparent to human beings. Despite these advantages, the unsupervised labeling of art has not yet been thoroughly explored. At this stage, it is not clear how close this type of categorization is to supervised labeling and whether the unsupervised machine labels can be useful and socially or professionally acceptable in art categorization.

Several earlier studies proposed to address the fine-art categorization from the perspective of supervised learning using techniques ranging from classical feature extraction and machine learning approaches to the implementation of complex deep learning methods. Classical machine learning approaches were, for example, explored in some of the first studies of automatic art analysis [5]–[10]. Arbitrary hand-crafted features were extracted from images of painting and classified using standard classifiers such as Support Vector Machine (SVM), Gaussian Mixture Model (GMM), or k-Nearest Neighbors (k-NN) algorithm. Due to the nature of these techniques, only relatively small datasets of images were needed to train the classifiers. A wide range of image transforms was investigated in a search for optimal image descriptors. It included parameters of the Fourier transform, the scale-invariant feature transform (SIFT), the color scale-invariant feature transform (CSIFT), the opponent-SIFT (OSIFT), local binary patterns (LBP), color LBP, GIST, pyramids of histograms of orientation gradients (PHOG), and the histogram of oriented gradients (HOG).

With the introduction of Deep Learning (DL) techniques and Convolutional Neural Network (CNN) classifiers, it became apparent that the automatic categorization of paintings based on DL outperforms the classical approaches [11], [12]. The initial art classification methods applied CNN models as feature extractors and linear classifiers such as SVMs to classify the features [11]–[17].

One of the most important concepts related to DL is transfer learning [1], [18]–[30]. Popular programming and software development platforms such as *Matlab* or *Python* offer a wide range of pre-trained CNN models of different structures and complexity. These models have been trained on vast datasets (in the order of millions) of images to perform the general task of image object classification. Given such a pre-trained network as a starting point, fine-tuning can be applied on a relatively small training dataset (in the order of thousands) to perform a more specific image classification task [31]. The application of pre-trained CNN models has been instrumental in reducing the time and data requirements of fine-art classification tasks. One of the first studies using transfer learning was reported by Tan *et al.* [22], in which a pre-trained CNN was fine-tuned to perform stylistic classification of digitized paintings. It was shown that the transfer learning approach outperforms methods using CNNs as feature extractors only. Fine-tuned CNNs have also been used to resolve other than classification tasks. In [32], for example,

the use of DL techniques for evaluating the beauty, sentiment, and remembrance of art was explored. In [33], CNN models were applied to detect cracks in paintings.

While a supervised classification of fine art has received plenty of attention from researchers, an unsupervised classification has been relatively less explored. Initial unsupervised studies aimed to determine the best hand-crafted features to perform clustering based on the visual appearance of paintings. In [34], for example, local and global features were investigated for clustering. Low-level features, diverse color statistics, and several features from face detectors were combined to perform an unsupervised classification using the principal component analysis (PCA) for the dimensionality reduction and the k-means algorithm clustering method. Eight clusters were generated, and their correlation with historical art periods was investigated. The results indicated that the clusters were not uniquely aligned with artistic movements. Instead, the grouping followed common colors and contents such as portraits, landscapes, and objects across different art movements. In [35], a nonlinear technique called Unsupervised Feature Learning k-means (UFLK) [36] was applied to extract features from images representing eight artistic movements in an unsupervised fashion. The Spectral Clustering (SC) algorithm grouped the features in an unsupervised way, and the SVM classifier performed a supervised categorization to determine the unsupervised clustering efficiency. The results indicated that the supervised classification accuracy was higher for the clusters of UFLK features than for row image patches clusters. However, the correlation of the unsupervised clusters with the eight artistic movements was not analyzed. A selective clustering approach was proposed in [37] to categorize a collection of fine-art paintings according to the artist. A CNN was implemented as the feature extraction technique, and a robust continuous clustering algorithm [38] complemented by the Bayesian rejection mechanism was used to classify works of six different artists. A deep clustering model was implemented in [39] to categorize a collection of 8,446 paintings from nine artistic periods. This approach adopted the Deep Convolutional Embedding Clustering (DCEC) framework introduced in [40]. The network structure consisted of a convolutional autoencoder with an embedded clustering layer. The outcomes were consistent with [34] in showing that the unsupervised categorization followed the characteristics of a visual similarity rather than an artistic movement.

In this study, we take further the concept of combining an unsupervised clustering with a supervised classification introduced in [39] and propose a new method which (i) employs an unsupervised clustering module trained to cluster the paintings and (ii) a supervised classification network module trained to recognize the categories proposed by the unsupervised module. However, unlike in [39] in our approach, both modules are connected by an optimization algorithm that iteratively improves the unsupervised clustering by increasing the supervised classification accuracy and minimizing the unsupervised clustering error.

III. METHODOLOGY

A. RESEARCH HYPOTHESIS

This research addresses the paradigm of unsupervised labeling of fine-art paintings. The goal was to achieve machine-made labels that could automatically categorize fine art paintings without the need for human annotation. In other words, we wanted to achieve machine-made labels that group art into distinct categories. To establish an objective quality criterion for the machine-made clusters, it was hypothesized that the higher quality of unsupervised clusters should result not only in the lower value of the within-cluster sum of squares (WCSS) error but also in the higher accuracy of the supervised classification based on these clusters.

B. PROPOSED SYSTEM

The proposed ACS combines an unsupervised clustering (or machine labeling) module and a supervised classification module based on machine-made labels. Both modules work as an adversarial Generator-Assessor team. The team is linked via a numerical optimization procedure which iteratively improves the quality of the unsupervised clusters.

The unsupervised clustering module (Generator) generates proposed clusters, and the supervised classifier (Assessor) makes an assessment of their quality. The objective function driving the iterative improvement of the unsupervised clusters consists of two components, the standard clustering error (i.e., the WCSS error) and the accuracy of the supervised classification based on the unsupervised clusters. The clusters are gradually modified in a way that simultaneously minimizes the clustering error and maximizes the classification accuracy.

As shown in Fig. 1, the proposed ACS consists of the following general processing steps:

1) STEP 1: FEATURE EXTRACTION

The unlabeled dataset of images representing fine-art paintings is transformed into features. A pre-trained CNN is employed as a feature extractor generating network embeddings.

2) STEP 2: UNSUPERVISED CLUSTERING (GENERATOR)

The network embeddings are grouped in an unsupervised way using a standard clustering algorithm. At this stage, temporary feature clusters are generated, and the current clustering error is estimated.

3) STEP 3: MACHINE LABELING

Arbitrary machine labels are assigned to each cluster.

4) STEP 4: DATA SPLITTING

Each cluster of embeddings represented by an associated machine label is divided into mutually exclusive training and testing subsets. Combined training subsets from all clusters constitute a machine-labeled training dataset, and the

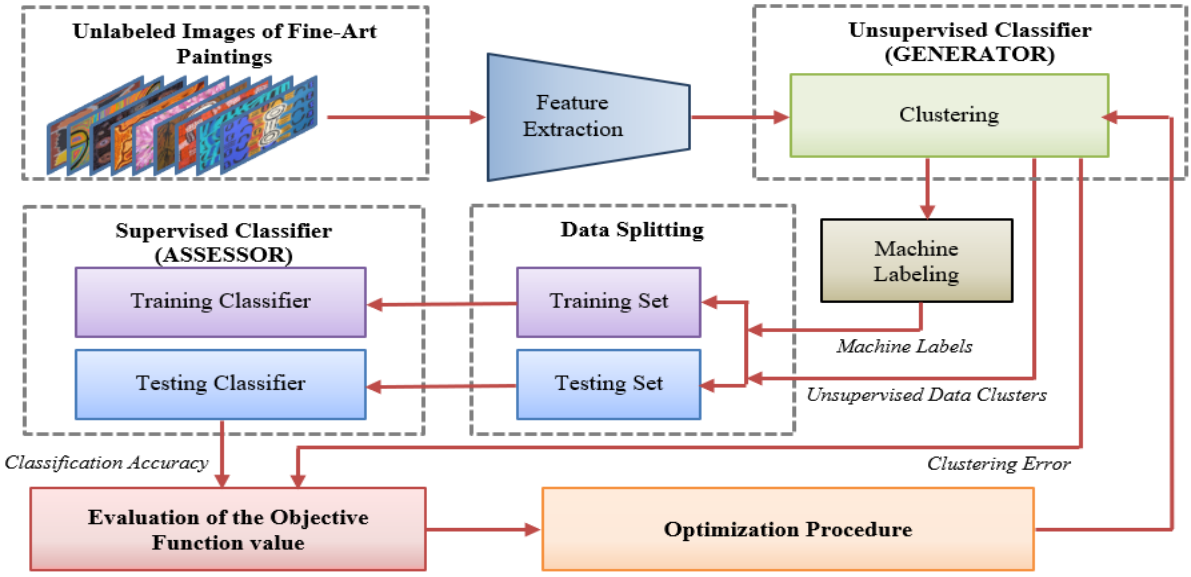


FIGURE 1. Block diagram of the proposed Adversarial Clustering System (ACS) for the unsupervised fine-art categorization.

combined testing subsets constitute a machine-labeled testing dataset.

5) STEP 5: SUPERVISED CLASSIFICATION (ASSESSOR)

Supervised classification of the embeddings into categories defined by machine labels is performed using a standard classifier. The classifier is trained using the training set of features and tested using the testing set. At each iteration, the current average accuracy (across all categories) of the testing procedure is estimated.

6) STEP 6: EVALUATION OF THE OBJECTIVE FUNCTION

The average accuracy of the supervised classification (determined in Step 5) and the clustering error (determined in Step 2) are used to estimate the current value of the objective function.

7) STEP 7: OPTIMIZATION PROCEDURE

A numerical optimization procedure is employed to improve the unsupervised clusters generated in Step 2. The clustering parameters are modified to minimize the objective function value. The modified clustering parameters are passed to Step 2, and Steps 2-7 are repeated until a satisfactory solution defined by an arbitrarily small value of the objective function is reached.

The following sections provide details of the above processing steps.

C. FEATURE EXTRACTION

We used deep network embeddings as inputs to the unsupervised clustering module. Network embeddings extracted from pre-trained CNN models have been shown to provide excellent performance in the unsupervised classification of natural images [41] and speech synthesis [42], outperforming classical image features.

The neural network embedding process represents a categorical (an object label) or discrete variable as a real-valued vector in a continuous multi-dimensional space. The purpose of this kind of mapping is to either cluster the objects in an unsupervised or supervised way or to determine relative distances or relations between objects in the embedding space. To generate an embedding vector, the input object (for example, an image) is put through a neural network pre-trained in a supervised way to perform a specific recognition task. The network parameters resulting from the recognition task form the embedding vector corresponding to the input object. One of the most challenging parts of the embedding process is to decide how to pre-train the network, so the resulting embeddings are meaningful for the intended application. In the case of image clustering, the most often used embeddings are generated by neural networks pre-trained to perform the image object classification task [41]. In our case, the goal was to achieve an unsupervised clustering of art images and to identify the relationship between the unsupervised clusters and the human annotation of artistic styles.

Our previous fine-painting classification study, based on transfer learning from object recognition to style art classification task, has shown that the pre-trained ResNet-50 model [43] achieved good classification results over different pre-trained CNNs and diverse art datasets [1]. Other visual art studies also have reported good classification performance using the ResNet-50 network [18], [27]. Although fine-tuning deeper architectures can exhibit higher classification accuracies, this requires longer training times and higher computational requirements [25], [29]. Therefore, the pre-trained ResNet-50 model offers an outstanding balance between high performance and computational cost.

Consequently, two different types of network embeddings were tested for comparison. The first set of embeddings was given by the ResNet-50 [43] model pre-trained to classify 1000 image object categories [44], and the second set was obtained from the same ResNet-50 model trained on object recognition and additionally fine-tuned on the WikiArt dataset of fine art paintings [45] to recognize 23 artistic styles [1]. Since the first set was extracted from a network that had no pre-requisite knowledge of style, it was expected that these embeddings would lead to object-based rather than style-based unsupervised clusters. On the contrary, the second set of embeddings was extracted from a network that had a pre-requisite knowledge of artistic style. Therefore, it was more likely to provide style-based clusters. In both cases, the features were given as a vector of length 2048 taken from the avg_pool layer of the ResNet-50 model [43]. The first set of features extracted from the CNN network pre-trained on image object classification only is referred to as ''ResNet-50-IO'', and the feature set extracted from the CNN network pre-trained on object classification and fine-tuned on artistic style classification is referred to as ''ResNet-50-AS.''

D. UNSUPERVISED CLUSTERING (GENERATOR)

After testing several potential candidate approaches, a relatively simple k-means algorithm was chosen [46] as an unsupervised clustering method for the proposed system. While more complex techniques may outperform k-means in specific applications, in our case, the simplicity and easy adaptation of clustering parameters offered by k-means was the key factor for making it the best choice. Given a set of N vectors $(x_1, x_2, \ldots, x_N)$, the k-means clustering algorithm groups them into k ($\leq$N) sets (clusters) S = $\{S_1, S_2, \ldots, S_k\}$. The objective is to find an optimal set of clusters S$*$ that minimizes the within-cluster sum of squares (WCSS) given as,

$$WCSS\,(S) = \sum_{i=1}^{k} \sum_{x \in S_i} \|x - m_i\|^2 \tag{1}$$

where m_i is the centroid vector of cluster S_i. Due to the nature of the unsupervised clustering of data, the ground truth information was not available. Therefore, internal clustering metrics had to be used to evaluate the quality of the generated clusters. For this purpose, the Calinski-Harabasz index was adapted. As given in (2), the Calinski-Harabasz index *CHI* was defined as the ratio of the between-cluster variance SS_B and the within-cluster variance SS_W multiplied by a constant factor that depends on the number of clusters k and the total number of data vectors N [47].

$$CHI(k) = \frac{(N-k)}{(k-1)} \times \frac{SS_B}{SS_W} \tag{2}$$

The between-cluster variance SSB was defined as the sum of Euclidean distances between cluster centroids and the mean vector of the whole dataset, which can be denoted as,

$$SS_B = \sum_{i=1}^{k} n_i \|m_i - m\|^2 \tag{3}$$

where n_i is the number of data vectors within cluster i, m_i is the centroid vector of the i-th cluster, m is the mean vector of the whole dataset, and $\|m_i - m\|^2$ is the Euclidean distance between these two vectors.

Similarly, the within-cluster variance SS_W was defined as the sum of Euclidean distances between cluster centroids and all vectors within a given cluster, which can be denoted as,

$$SS_W = \sum_{i=1}^{k} \sum_{x \in c_i} \|x - m_i\|^2 \tag{4}$$

where x is a data vector, S_i is the i-th cluster, m_i is the centroid of cluster i, and $\|x - m_i\|^2$ is the norm, or Euclidean distance between these two vectors.

When clustering the data, it was essential to maximize the Calinski-Harabasz index; This was equivalent to increasing the between-cluster variance SS_B and decreasing the within-cluster variance SS_W. However, apart from well-separated clusters in an objective sense, we also aimed to achieve clusters that group the art images into categories that are ideally close to human-made labels given by art experts. To assess this quality, we used two measures, the unsupervised classification accuracy A_{USup} and the Krippendorff's Alpha-Reliability measure α [48]. To estimate these parameters, we had to create an arbitrary link between the machine labels and human-made labels of artistic styles. As explained in Section V, it was done either by mapping each of the unsupervised clusters to the style represented by the largest number of paintings belonging to this cluster or using the Hungarian mapping algorithm [49]. By doing this, a form of ''ground truth'' was established for each machine-made cluster. The unsupervised classification accuracy A_{USup} was defined as:

$$A_{USup} \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

where TP is the number of true positive assignments, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

The Krippendorff's Alpha-Reliability α is a statistical measure of the agreement achieved between different assessors assigning labels to a set of vectors [48]. In our case, it was applied to determine the agreement between machine labeling and human labeling. The Krippendorff's Alpha-Reliability α was defined as:

$$\alpha = 1 - \frac{D_0}{D_e} \tag{6}$$

where D_0 was the observed disagreement probability in assigning labels to vectors, and D_e is the expected disagreement probability happening by chance. Theoretical and computational details of calculating the Krippendorff's Alpha-Reliability can be found in [48]. The values of α are between 0 and 1. When $\alpha = 1$, there is a perfect agreement between assessors. When $\alpha = 0$, there is no correlation between labels assigned by different assessors, and when $0 < \alpha < 1$, there is a systematic disagreement between assessors exceeding what would be expected by chance. In other words, the closer is the value to 1, the higher is the agreement between assessors.

E. ASSIGNING MACHINE LABELS

Each of the clusters generated by the unsupervised clustering process was assigned an arbitrary label. We call these labels ''machine labels'' as opposed to subjective labels given by human experts during the traditional manual labeling process.

F. TRAINING AND TESTING DATASETS

To achieve a balanced representation of machine-made categories during the supervised training/testing procedure, each of the unsupervised data clusters was split into training subset (80%) and testing subset (20%). The resulting training subsets were grouped together to create a pool of training data, and all testing subsets were grouped to create a pool of testing data. Each data vector came with an associated machine label.

G. SUPERVISED CLASSIFICATION (ASSESSOR)

A classical multiclass Support Vector Machine with linear kernel and Error-Correcting Output Coding (ECOC) algorithm [50] was trained to classify the dataset of network embeddings into machine-made categories. It was trained in a supervised way, with machine labels serving as the ''ground truth'' information. The SVM algorithm was trained to minimize the classification error and thus maximize the supervised classification accuracy A_{Sup} given as,

$$A_{Sup} = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

where TP is the number of true positive, TN is the number of true negative, FP is the number of false positive, and FN is the number of false negative classification outcomes.

H. OPTIMIZATION OF UNSUPERVISED CLUSTERING

A genetic algorithm [51] was implemented as a numerical optimization technique aiming to iteratively improve the unsupervised clustering. Given a fixed number of clusters k, the algorithm conducted a search through the vector space of cluster centroids to find an optimal set of clusters S^* that minimizes the following objective function:

$$f_{obj}(S) = WCSS(S) + A_{Sup}(S)^{-1} \tag{8}$$

The minimization of (8) was equivalent to the simultaneous maximization of the supervised classification accuracy A_{Sup} and minimization of the unsupervised within-cluster sum of squares WCSS. The initial centroids were determined using the k-means++ algorithm [52], and the number of clusters k was set depending on the experimental setup and the tested dataset (see Section V).

IV. DATASETS

A. DATASETS OF FINE ART IMAGES

Three datasets of images depicting fine art paintings were used to evaluate the proposed ACS method.

1) DATASET 1

The first dataset was a collection of 4,105 digitized images representing five stylistic categories: Australian Aboriginal art, Abstract art, Byzantine Iconography, Cubism, and Northern Renaissance. The Australian Aboriginal art collection of images was created by the authors, whereas the remaining four styles were sourced from the publicly available Pandora 18K dataset professionally labeled by art experts [53], [54]. Fig. 2 shows the composition details of this dataset. The style representation was perfectly balanced, with each stylistic category contributing 20% to the total number of paintings.

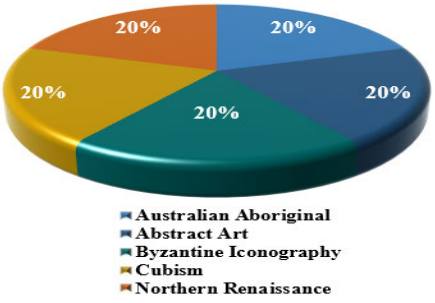


FIGURE 2. Dataset 1 balanced - number of images per style (in percentage).

The Australian Aboriginal art is predominantly characterized by abstract compositions. There are no realistic depictions of objects, people, faces, or sceneries. People and animals are sometimes depicted in a form that resembles x-ray images. Many paintings have structure-like patterns that can be overlaid with symbols. There is no perspective nor dimensionality. Some paintings are almost monochromatic, whereas others use a wide range of colors. A large proportion of Aboriginal paintings show an aerial view depicting abstract elements related to the Australian landscape and the indigenous cosmology, mythology, laws, and belief systems [55].

The remaining four styles in Dataset 1 included artistic movements from ancient and modern historical periods that are only loosely related to each other. The ancient styles are the Byzantine Iconography, which dates between the years 500 and 1400, and the Northern Renaissance movement that began in the year 497 and lasted through to 1550. The two modern styles are the Cubism movement, which started in 1907 and ended in 1920, presenting an overlapping with the second style, Abstract art, which period goes from 1910 to the present [3]. As previously shown in [1], a supervised CNN model was able to categorize the five movements included in Dataset 1 with high accuracy.

2) DATASET 2

The second dataset was the Painting Database for art Movement Recognition Pandora 18K [53], [54]. It included a total of 18,038 images of paintings representing 18 different artistic styles. Fig. 3 illustrates the distribution of images across styles. It shows that the numbers of images were quite evenly distributed, with only a small amount of imbalance. Due to the very rigorous labeling process done by art experts, the Pandora 18K dataset is regarded to be one of

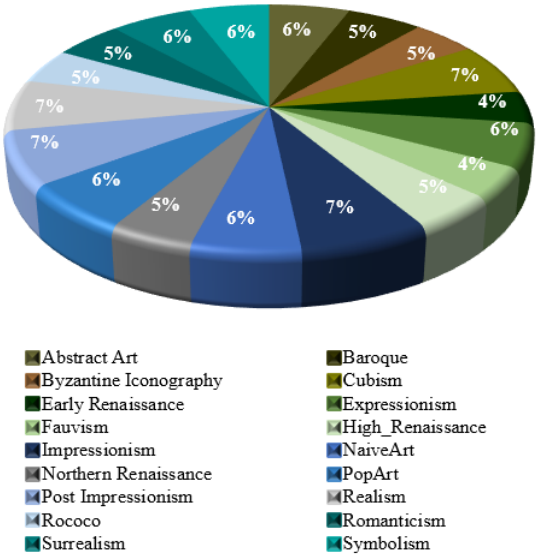


FIGURE 3. Pandora 18K number of images per style (in percentage).

the highest quality research datasets available for the fine art classification.

3) DATASET 3

While in Dataset 1, the Australian Aboriginal style was represented alongside other styles, the Dataset 3 consisted exclusively of the Australian Aboriginal art paintings. It contained a collection of 5,313 images obtained for research purposes from different online galleries [56]–[59] that are members of the Australia Aboriginal Art Association [60]. There were no artistic style labels attributed to these artworks. The authors' informal visual inspection indicated that the collection potentially included several stylistic sub-categories characterized by a specific scene composition and coloring. However, no expert labels identifying these categories were available. The existence of such a vast uncategorized collection of culturally significant artworks was one of the factors validating the purpose of our research. We wanted to find out whether an automatic categorization could provide useful results for the purpose of storage and retrieval of Aboriginal art.

V. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we present the experimental results and discuss the efficacy of the proposed ACS. We compare the ACS method with the standard k-means clustering across metrics given in Section III using two different image features: ResNet-50-IO and ResNet-50-AS. We have a closer look at the stylistic contents of clusters generated by the unsupervised ACS classifier to understand what kind of criteria were used to group the fine art paintings and how close these criteria are to human-made categorization. Since three different datasets of paintings were used in testing, and each dataset had a different size and composition, each dataset's results are analyzed separately.

A. RESULTS FOR DATASET 1

In the case of Dataset 1, we wanted to see how well a mixture of paintings, including the Australian Aboriginal art and four other visually similar styles, can be grouped into separate clusters in an unsupervised way. To evaluate how many distinct machine-made categories we can accurately identify using the unsupervised k-means clustering, we applied the elbow method [61] to cluster the ResNet-50-IO and ResNet-50-AS features extracted from Dataset 1. The variance of the clustering error WCSS given in (1) was calculated for a different number of clusters. The optimal number of clusters was determined by the inflection point (elbow) of the plot of the error variance versus the number of clusters; at the inflection point, the slope of the curve sharply drops. Fig. 4 shows the implementation of the elbow method for Dataset 1 with the two sets of embedding features. It was found that for the ResNet-50-IO features, the optimal number of clusters was four ($k = 4$), and for the ResNet-50-AS, it was five ($k = 5$). The optimal number of clusters given by the ResNet-50-AS was perfectly aligned with the number of human-made categories. It indicated that the ResNet-50-AS features extracted from a network pre-trained on the artistic style recognition task were more likely to follow the style-based distribution than the ResNet50-IO features extracted from a network having no prior knowledge of style. It must be noted that the elbow approach led to an optimal value of k with respect to the k-means using the WCSS error given in (1) as an objective, but not necessarily with respect to the f_{obj} given in (8) and used by the ACS. However, to evaluate the unsupervised clustering accuracy A_{USup} as given in (5) and the Krippendorff's Alpha-Reliability given in (6), we had to establish a hypothetical ground truth for the machine labels. The elbow test results indicated that we could reasonably set the number of clusters k to the number of human-made artistic categories represented by a given dataset. To estimate the performance

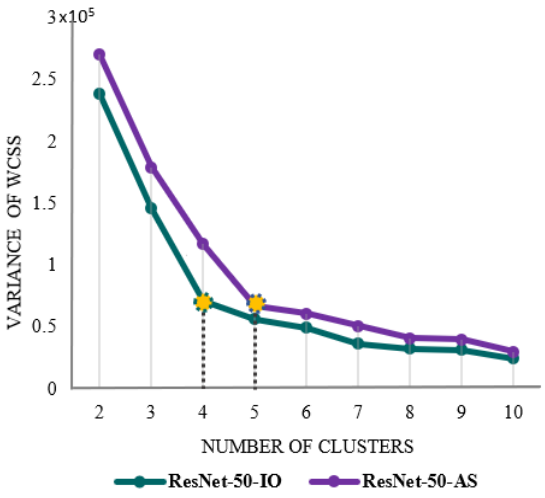


FIGURE 4. Identification of the optimal number of clusters using Elbow method for Dataset 1.

metric presented in Table 1, we had to create a correspondence link (mapping) between the human labels and the machine labels. The mapping process provided the same links when using the Hungarian method and the largest number of images method. Fig. 5 and Fig. 6 show the outcomes of the maximum number of images approach.

TABLE 1. Performance of the Adversarial Clustering System (ACS) and k-means for k = 5 clusters from Dataset 1.

Metric	Features			
	ResNet-50-IO		ResNet-50-AS	
	ACS	k-means	ACS	k-means
WCSS	3840914	379837	3394416	3354143
CHI	154.609	145.382	168.382	157.861
A_{USup}	0.7437	0.6847	0.7690	0.7183
A_{Sup}	0.9502	0.9022	0.9597	0.9101
α	0.6962	0.6276	0.7086	0.6612

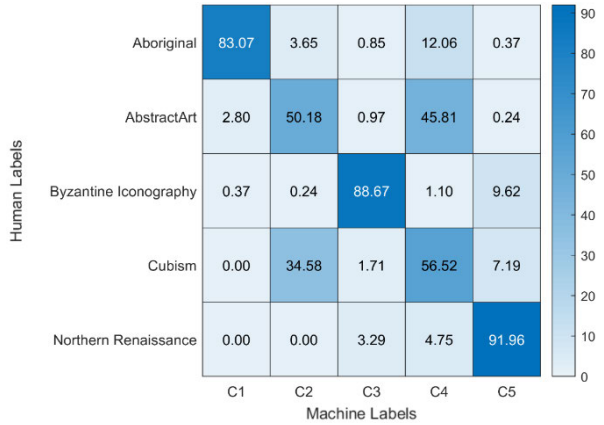


FIGURE 5. Mapping between the human labels and the unsupervised machine labels for ResNet-50-IO features for Dataset 1 using the maximum number of paintings method.

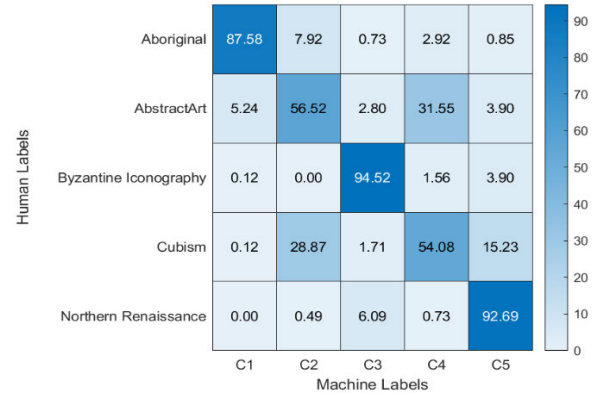


FIGURE 6. Mapping between the human labels and the unsupervised machine labels for ResNet-50-AS features for Dataset 1 using the maximum number of paintings method.

The values shown in Table 1 indicate that in the case of Dataset 1, ResNet-50-AS features outperformed the ResNet-50-IO features on all measures. Specifically, the Krippendorff's Alpha-Reliability α was higher for ResNet-50- AS features, indicating a smaller disagreement between machine-made and human-made labels.

Comparing the metrics obtained with the ACS system against the results from the k-means method for both sets of features; it can be observed that although the WCSS error was slightly smaller for the traditional k-means method, indicating a stronger within-cluster concentration for k-means. However, the CHI parameter and the unsupervised classification accuracy A_{USup} had higher values with the ACS system, pointing to a stronger separation between clusters. The metric of supervised classification (A_{sup}) and Krippendorff's Alpha-Reliability, α, yield higher values with the ACS system. Consequently, the ACS system is in higher agreement with the human annotations than the k-means method.

Table 2 shows the results for different numbers of clusters using both sets of features, ResNet-50-IO and ResNet-50-AS. There are no significant differences between the values of the supervised classification accuracy (A_{Sup}) obtained with ResNet-50-IO and ResNet-50-AS. In both groups of features, the supervised classification accuracy A_{Sup} decreases monotonically as the number of clusters increases achieving about 98% for two clusters and about 90% for 8 clusters. The WCSS error and the Calinski-Harabausz index exhibit similar behavior.

TABLE 2. ACS Clustering results for different number of clusters using Dataset 1 with ResNet-50-IO and ResNet-50-AS features.

	ResNet-50-IO			ResNet-50-AS		
k	WCSS	CHI	A_{Sup}	WCSS	CHI	A_{Sup}
2	4202324	265.01	0.983	3669468	263.16	0.982
3	4024060	222.16	0.979	3524696	228.23	0.975
4	3907662	181.13	0.952	3449428	197.37	0.963
5	3840914	154.62	0.950	3394416	168.38	0.959
6	3780728	137.30	0.942	3346053	149.86	0.941
7	3730566	122.92	0.933	3310627	135.71	0.927
8	3690434	111.87	0.923	3290627	123.93	0.897

The results of mapping between machine-made clusters C1-C5 and the human-made artistic style labels (Aboriginal, Abstract, Byzantine, Cubism, and Northern Renaissance) are presented in Fig. 5 and Fig. 6 for ResNet-50-IO and ResNet-50-AS features, respectively. For both sets of features, cluster C1, cluster C3, and cluster C5 contain a relatively large number of paintings that belong to a single style, namely Aboriginal, Byzantine Iconography, and Northern Renaissance, respectively. However, the ResNet-50-AS clusters attracted significantly more images representing these particular styles compared to the ResNet-50-IO clusters. This can be attributed to the fact that the ResNet-50-AS features were generated by a network pre-trained to differentiate between artistic styles, whereas the ResNet-50-IO came from a network that had no pre-requisite knowledge of artistic styles. Clusters C2 and C4 combined artworks from modern styles (Abstract art and Cubism) for both sets of features. Generally, the Dataset 1 clustering led to more refined differentiation between ancient styles while the modern styles were mixed together.

Examples of the ACS clustering results based on the ResNet-50-AS features are illustrated in Fig. 7. For each

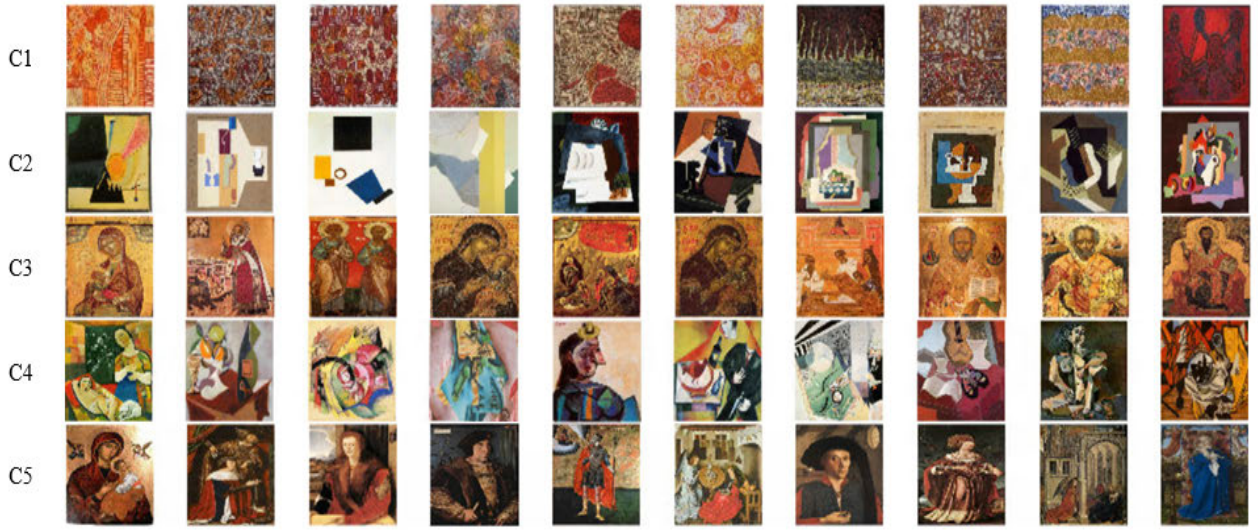


FIGURE 7. Top 10 paintings based on the ascending distance from the cluster centroid – Dataset 1, ResNet50-AS features, ACS clustering with k = 5.

cluster C1 to C5, the images are listed in ascending order based on their distance from the cluster centroid. The artistic movement of these five clusters can be clearly identified. Namely, cluster C1 corresponds to Aboriginal art. Cluster C2 includes paintings with free-form and geometric compositions, which are related to the Abstract style. Cluster C3 shows images with religious and holy figures, corresponding to Byzantine iconography. Cluster C4 shows artworks painted by Pablo Picasso, Franz Marc, and Bela Kadar that belong to Cubism [2]. Finally, intricate portraits of the Northern Renaissance period are grouped in cluster C5.

B. RESULTS FOR DATASET 2

For Dataset 2, the clustering aimed to test unsupervised grouping of a relatively large number of 18 artistic styles with various degrees of similarity into different machine-made categories. In the same way, as for Dataset 1, the elbow method was used to determine the optimal number of clusters for the unsupervised classification. The elbow graph indicated that the number of clusters for the ResNet-50-IO features should be equal to eight ($k = 8$), and for the set of ResNet-50-AS features, it should be equal to ten ($k = 10$). Thus, the number of optimal clusters for both sets of features differed significantly from the number of 18 historical periods (human-made labels) included in Dataset 2. As for Dataset 1, the optimal number of clusters identified for Dataset 2 and the ResNet-50-IO features was smaller than the optimal number of clusters determined for the ResNet-50-AS features.

Since the number of artistic groups (human labels) in Dataset 2 was considerably larger than in Dataset 1, the mapping between machine labels and the human labels by determining which human label (style) was represented by the largest number of images within a given cluster was not straightforward applicable. For this reason, to find the best match between the ACS clusters and the 18 human-made labels, the Hungarian algorithm frequently used in assignment problems was applied [49]. This algorithm implemented a combinatorial optimization procedure to find an optimal distribution matrix mapping the human-made labels (artistic styles) into the machine labels (clusters). It was achieved by minimizing a cost function measuring the amount of style mismatch given by the mapping array.

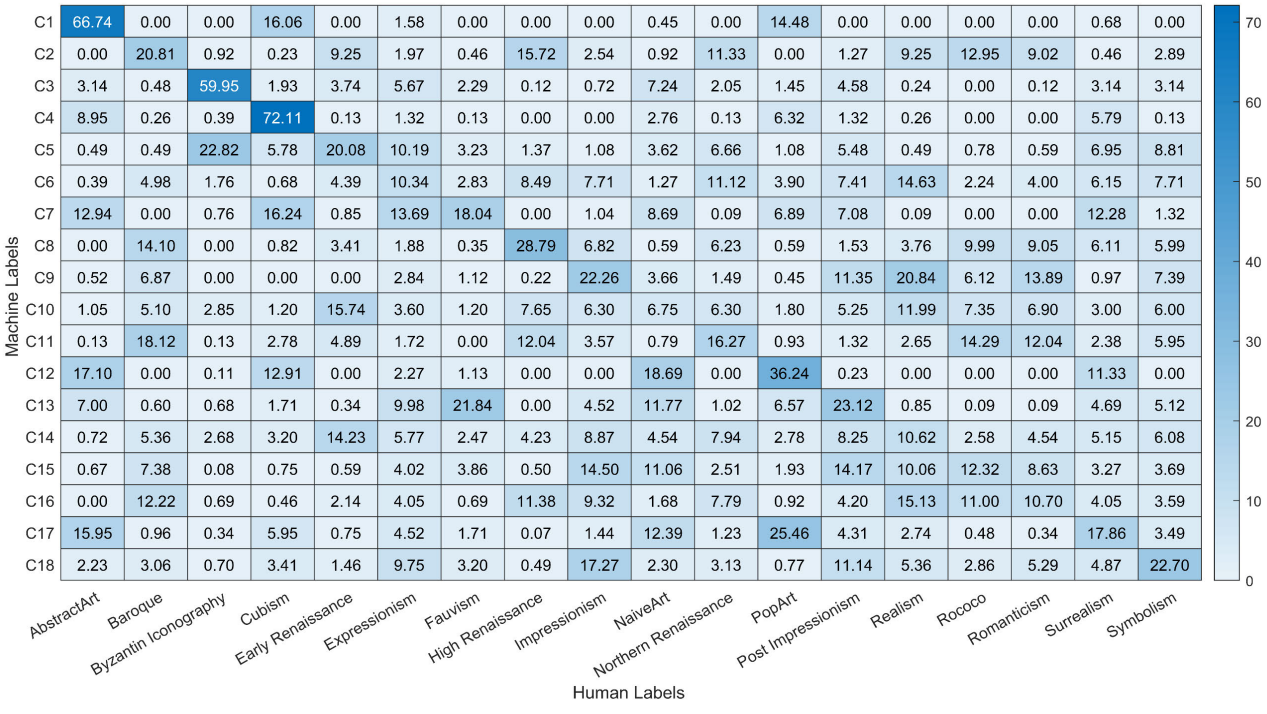
The Hungarian mapping results for 18 clusters are shown in Fig. 8 and Fig. 9 for the ResNet-50-IO and the ResNet-50-AS features, respectively. In both cases, the distribution of the artistic movements between the clusters and the percentage composition of artistic movements within each cluster indicate that all clusters included a mixture of artistic styles; however, some clusters contained a minimal number of closely related styles.

For both sets of features, cluster C3 contains a substantial percentage of images that are members of the Byzantine Iconography whereas, cluster C4 contains the Cubism movement predominantly. The majority of images in cluster C1 represent the Abstract style (66.74% for the ResNet-50-IO C1 cluster and 53.42% for the ResNet-50-AS C1 cluster). However, this was only 27.75% (for ResNet-50-IO C1) and 36.03% (for the ResNet-50-AS) of the total number of Abstract images included in Dataset 2. The remaining clusters include more diverse compositions of images that belonged to different artistic periods.

In general, the number of images within clusters that belonged to a single style was larger when using ResNet-50-AS features than when using the ResNet-50-IO features. This was again consistent with the fact that the ResNet-50-AS features contained artistic style knowledge as they were derived from a CNN trained to classify art images.

When using ResNet-50-AS features (Fig. 9), cluster C3 showed the most uniform style definition with 74% of images representing the Byzantine Iconography; it was 89.5% of the

a)



b)

FIGURE 8. Mapping between the human labels (artistic styles k = 18) and the unsupervised machine labels (ACS clusters) for ResNet-50-IO features for Dataset 2. a) Distribution of the artistic styles (human labels) among the ACS clusters (machine labels). b) Percentage composition of each ACS cluster (machine labels) according to artistic styles (human labels).

total number of all Byzantine Iconography images included in Dataset 2. Cluster C4 and C5 exhibited a significant concentration of paintings belonging to the same artistic period, although with a lower percentage than the one showed in cluster C3. Namely, 70.85% of images in cluster C4 belonged to Cubism, which corresponds to 47.35 % of all Cubism paintings in Dataset 2, and 31.67% of images in cluster C5 represented Northern Renaissance which was 48.20% of all Northern Renaissance paintings in Dataset 2.

Table 3 compares the ACS performance with the k-means method for 18 clusters of Dataset 2 using ResNet-50-IO and ResNet-50-AS features. In comparison with Table 1, Table 3

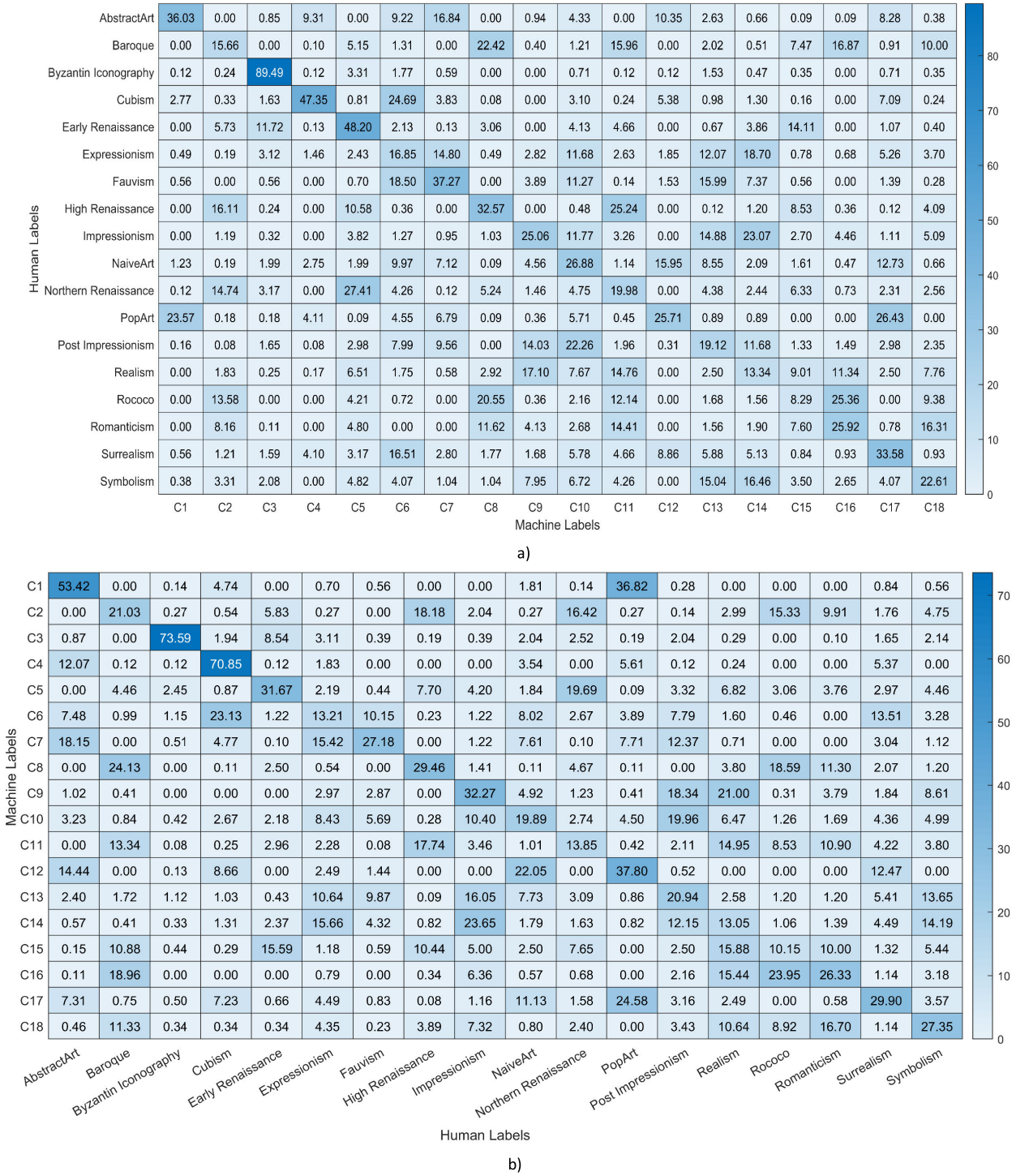


FIGURE 9. **Mapping between the human labels (artistic styles k = 18) and the unsupervised machine labels (ACS clusters) for ResNet-50-AS features for Dataset 2. a) Distribution of the artistic styles (human labels) among the ACS clusters (machine labels). b) Percentage composition of each ACS cluster (machine labels) according to artistic styles (human labels).**

indicates that for the 18 clusters of Dataset 2, the unsupervised classification accuracy A_{USup} and the Krippendorff's Alpha-Reliability α index presented lower values, indicating that the machine-made labels (clusters) were not as strongly related to the human-made labels as in the case of 5 clusters of Dataset 1. Among the Table 3 values, the ResNet-50-AS features presented the best performance in terms of A_{USup} and α, once again confirming that unsupervised clustering of features extracted from a network pre-trained to recognize artistic styles leads to results that are closer to human labeling

TABLE 3. Performance of the Adversarial Clustering System (ACS) and k-means for k = 18 clusters from Dataset 2.

Metric	Features			
	ResNet-50-IO		ResNet-50-AS	
	ACS	k-means	ACS	k-means
WCSS	15027291	14848316	13414572	13272114
CHI	246.56	236.42	260.11	253.42
A_{USup}	0.2643	0.2446	0.3025	0.2816
A_{Sup}	0.811	0.7539	0.817	0.7632
α	0.2132	0.1984	0.2678	0.2397

TABLE 4. ACS Clustering results for different number of clusters using Dataset 2 with ResNet-50-IO and ResNet-50-AS features.

	ResNet-50-IO			ResNet-50-AS		
k	WCSS	CHI	A_{sup}	WCSS	CHI	A_{sup}
8	15838834	436.45	0.876	14157478	463.70	0.884
10	15607485	374.14	0.867	13942176	397.14	0.869
12	15420548	329.66	0.845	13777989	348.30	0.864
14	15263476	296.05	0.835	13632033	312.68	0.842
16	15129280	269.48	0.823	13518310	283.35	0.820
18	15027291	246.56	0.811	13414572	260.11	0.817

TABLE 5. Performance of the Adversarial Clustering System (ACS) and k-means for k = 10 and k = 12 clusters from Dataset 3.

Metric	Features			
	ResNet-50-IO		ResNet-50-AS	
	ACS	k-means	ACS	k-means
	K= 10			
WCSS	5839283	5727923	3982624	3857715
CHI	124.83	111.92	106.95	93.13
A_{Sup}	0.915	0.891	0.904	0.874
	K -12			
WCSS	5721090	5546728	3899992	3779556
CHI	114.16	99.73	99.54	89.76
A_{Sup}	0.910	0.872	0.894	0.863

of art than features extracted from a network train to classify objects in general. Otherwise, Table 3 shows very similar trends to Table 1. Like for Dataset 1, clustering of Dataset 2 shows that the WCSS error values obtained with the k-means method for both sets of features are slightly lower than the values obtained with the ACS. Also, the *CHI*, A_{USup},A_{Sup}, and α values show that the ACS outperforms the k-means method.

The ACS clustering results for both sets of features and the number of clusters ranging from 8 to 18 in steps of 2 are presented in Table 4. The observed trends are consistent with Table 2 for Dataset 1. The supervised accuracy A_{Sup} values do not show significant differences between the two different sets of features. For the ResNet-50-AS clusters, the A_{Sup} is only around 1% better than for the ResNet-50-IO. For both sets of features, the WCSS error, *CHI,* and the A_{Sup} values decrease when the number of clusters increases. For all numbers of clusters, the WCSS error and the *CHI* index are slightly higher for the ResNet-50-AS than for the ResNet-50-IO features.

Fig. 10 shows an example of the unsupervised clustering outcome for the ACS using ResNet-50-AS features and the cluster number set to ten (k = 10). Each row corresponds to a different cluster, and for each cluster (C1-C10), there are the top 10 paintings arranged in the ascending order of their Euclidian distance to the cluster centroid.

It can be observed that in most cases, the grouping into unsupervised clusters was based on the scene composition and types of depicted objects rather than the style classification. A closer inspection of images grouped within each cluster shows that: Cluster C1 combines paintings that depict highly realistic, almost photographic-like landscapes with trees and sharp objects-defining edges as common characteristics; the artworks belong to several styles, including Impressionism, Realism, and Baroque. Cluster C2 concentrates images from Abstract and Cubism, for which semi-geometric squared and linear shapes are common features. Cluster C3 groups artworks depicting semi-realistic landscapes with buildings or people painted with diffused brushstrokes creating very soft edges; these images belong to Fauvism, Post Impressionism, and Expressionism. These three styles in cluster C3 have been previously reported to show high confusion rates when classified with CNN models in a supervised way due to smooth transitions between these artistic movements [1], [4]. Scenes depicting people in dark backgrounds are predominant attributes of cluster C4; this type of scenery is typical of High Renaissance, Rococo, Realism, and Romanticism. Cluster C5 is formed by paintings showing large plain-color areas and a lack of perspective created by the domination of vertical and horizontal lines or object borders. These paintings belong mostly to Byzantine Iconography and the Early Renaissance. Cluster C6 shows a very uniform artistic style representation; it contains Byzantine Iconography works. Cluster C7 groups paintings from Cubism, Surrealism, and Fauvism having an apparent visual similarity in scene composition and colors. The presence of angel figures is a common characteristic of cluster C8, which contains mostly artworks painted by Gustave Moreau, one of the most representative artists of Symbolism [4]. Cluster C9 is comprised of paintings depicting people and having color palettes characteristics to High Renaissance, Baroque, and Rococo; the latter two artistic styles are known to be strongly related [2]–[4]. Cluster C10 comprises portraits from High Renaissance, Baroque, Rococo, and Romanticism.

Although the ACS achieved a better machine-made clustering agreement with the art experts compared to the k-means method, in general, the unsupervised grouping does not fully agree with the labels created by art experts. However, we were able to identify several critical visual criteria used to generate the machine-made clusters. These criteria include scene composition, types of objects, presence or lack of perspective, dominating directions of edges, the sharpness of edges, and the color palette.

C. RESULTS FOR DATASET 3

The unsupervised clustering of Dataset 3 aimed to analyze how well the Australian Aboriginal art can be subdivided into separate categories.

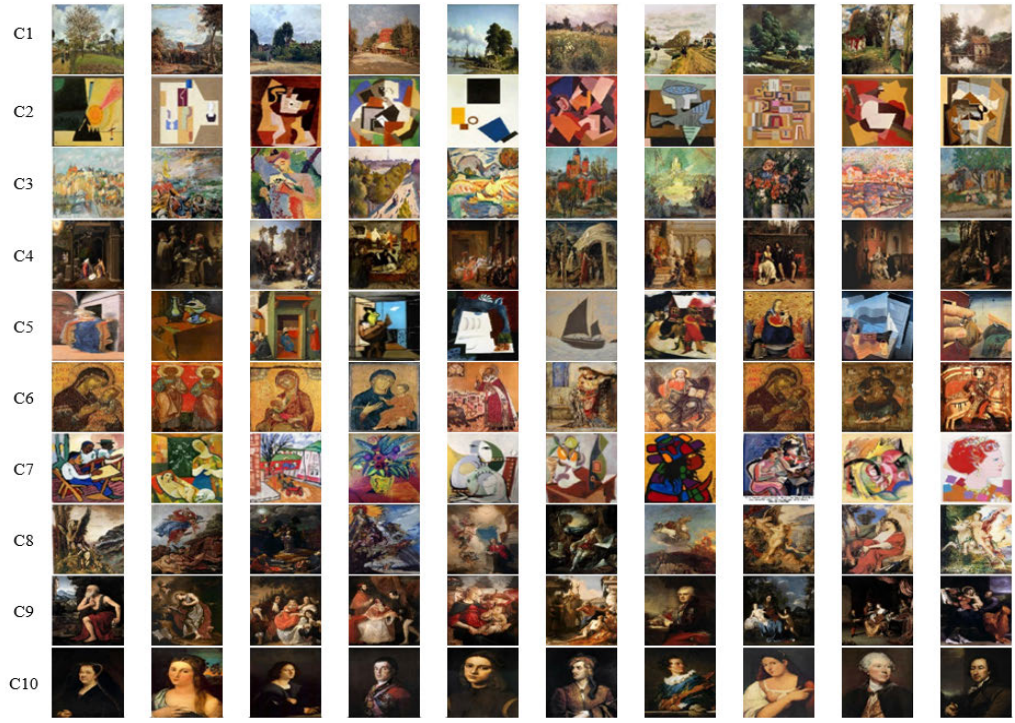


FIGURE 10. Top 10 paintings based on the ascending distance from the cluster centroid – Dataset 2, ResNet50-AS features, ACS clustering with k = 10.

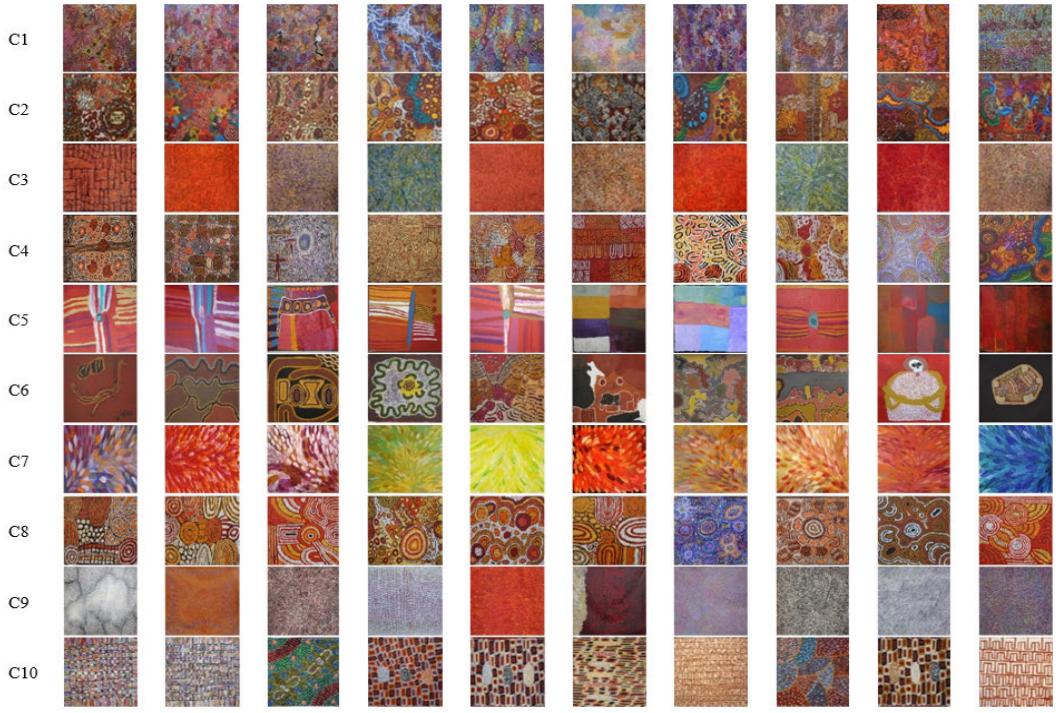


FIGURE 11. Top 10 paintings based on the ascending distance from the cluster centroid – Dataset 3, ResNet50-IO features, ACS clustering with k = 10.

As explained in Section IV, Dataset 3 had no human-made labels annotating artistic styles. Therefore, it was particularly interesting to see what kind of taxonomy will be proposed by the unsupervised clustering procedure and whether this taxonomy can be useful in the absence of ''educated'' labels made by art experts?

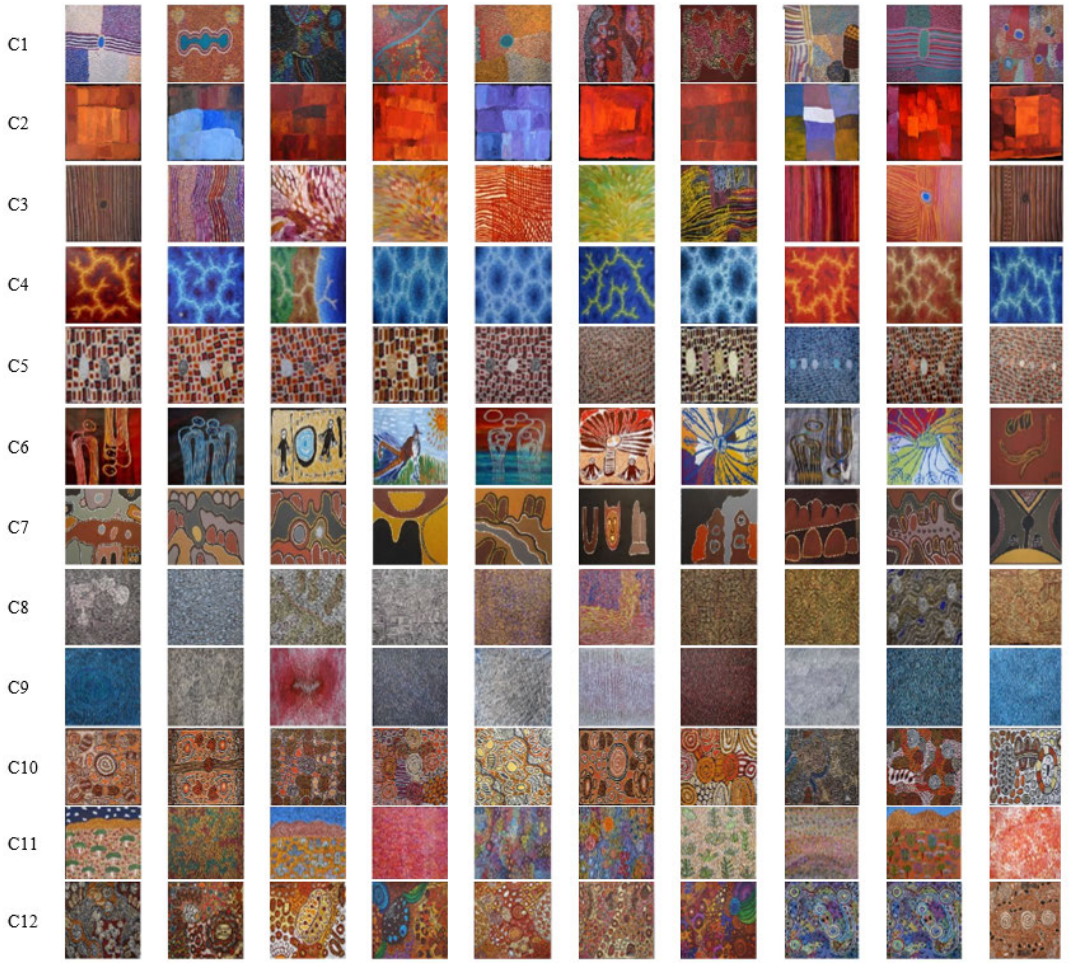


FIGURE 12. Top 10 paintings based on the ascending distance from the cluster centroid – Dataset 3, ResNet50-AS features, ACS clustering with k = 12.

TABLE 6. ACS Clustering results for different number of clusters using Dataset 3 with ResNet-50-IO and ResNet-50-AS features.

	ResNet-50-IO			ResNet-50-AS		
k	WCSS	CHI	A_{Sup}	WCSS	CHI	A_{Sup}
2	6672309	321.62	0.98	4516909	221.76	0.98
4	6343046	204.59	0.94	4303312	165.40	0.95
6	6136290	162.60	0.93	4167927	136.90	0.94
8	5967467	140.83	0.92	4060507	120.38	0.92
10	5839283	124.83	0.91	3982624	106.95	0.90
12	5721090	114.16	0.91	3899992	99.54	0.89
14	5627819	104.92	0.89	3836625	92.31	0.89
16	5533539	98.46	0.88	3773083	87.27	0.87
18	5457815	92.37	0.87	3717664	82.76	0.86

Like in previous cases, the optimal number of clusters for the unsupervised classification of Dataset 3 was determined using the elbow method. It was found that the optimal number of clusters for the ResNet-50-IO features was ten (k = 10), and for the ResNet-50-AS features, it was twelve (k = 12).

The ACS clustering results for different numbers of clusters using both sets of features are presented in Table 6. In all cases, the A_{Sup} shows only a small variation between the two sets of features. Similar to previous experiments with Datasets 1 and 2, A_{Sup} accuracy tends to decrease as the number of clusters increases. However, significant differences between the WCSS values for the ResNet-50-IO and the ResNet-50-AS features suggest that these two types of features contain different information about the paintings. Independent of the number of clusters, the ResNet-50-AS features lead to smaller WCSS values, indicating that the pre-requisite style knowledge inherited by ResNet-50-AS features assists with the formation of objectively better-defined clusters.

Fig. 11 and Fig. 12 show examples of the top ten paintings within each cluster ranked from the lowest to highest based on the distances of its feature vectors to the cluster centroid. Fig. 11 shows the top ten paintings (per cluster) resulting from grouping the ResNet-50-IO features into 10 clusters. Whereas Fig. 12 illustrates the top ten paintings resulting from grouping the ResNet-50-AS features into 12 clusters

When looking at the ResNet-50-IO clusters in Fig. 11, it can be observed that Cluster C2 contains Australian landscapes depicted from the aerial perspective. Various graphical symbols densely embedded into the landscapes tell ancestral Aboriginal stories. This cluster includes several artworks

painted by the famous artists Damien and Yilpi Marks. Cluster C1 also includes landscapes; however, these paintings clearly differ from those in C2 as they do not have the same high density of symbols. In addition, some of the C1 paintings are made using the ''dot technique'', which does not appear in C2. Cluster C4 contains multi-colored backgrounds with subtle patterns and larger, very sparse symbols painted on the top. Cluster C5 contains paintings characterized by large semi-rectangular blocks of solid colors with visible horizontally orientated borders. Cluster C6 contains paintings with curvy objects painted in an ochre color pallet (black, brown, orange, yellow) that evokes the ancestral use of natural pigments such as hard clay and charcoal. Paintings with very well-defined long brushstrokes are grouped in Cluster C7. Most of these artworks were painted by Gloria Petyarre, a renowned Aboriginal artist depicting bush medicine leaves in her paintings. Cluster C3 also contains leaves, but much smaller than in C7, and with stokes evoking the movement of leaves in the wind. Paintings with a distinct presence of concentric circles and U-form symbols representing women's ceremonies [62] are grouped in Cluster C8. C9 groups paintings with monochromatic color palettes and delicate patterns resembling natural stone surfaces. Cluster C10 contains paintings with squares and rectangles forming repetitive print-like patterns that characterize the artworks of Tjapaltjarri brothers and other artists from the Aboriginal linguistic group Pintupi [56].

Fig 12 illustrates what happened when the features were changed to ResNet-50-AS, and the number of clusters was increased from 10 to 12. Although it can be observed that some of the Fig. 12 clusters have similar contents to the Fig. 11 clusters, the indexes of the corresponding clusters are different. It is also apparent that the larger number of clusters led to the discovery of new distinct sub-categories not present in Fig. 11. For example, cluster C4 in Fig. 12 shows a new group of paintings depicting a dark monochromatic background with superimposed glowing irregular mesh in a light color. These artworks are known to refer to climatic events and include many paintings by Tarisse King. Cluster C6 is another example of a new group with human-like figures representing ancestral creation stories. This group captures many artworks of Fiona Omeenyo. Other clusters in Fig. 12 appear to be close equivalents of clusters identified in Fig. 11.

The above analysis allows us to conclude that the proposed unsupervised image clustering system (ACS) can identify Australian Aboriginal art sub-categories based on scene composition, types and shapes of objects, edge orientation, color palette, and brush strokes. Even though the ACS clustering process was not supported by art expertise, it led to the categorization that could be expected from a non-expert person aiming to sort the images according to the above criteria. From this point of view, unsupervised machine learning offers an efficient automatic labeling tool that can be used for the storage and retrieval of art images that have not been yet categorized by experts. These findings are particularly significant because the Australian Aboriginal art does not depict easily identifiable objects, and it has a very abstract and symbolic nature that is not easily understandable to cultural outsiders.

VI. CONCLUSION

The study investigated an unsupervised classification of fine-art paintings. A new unsupervised Adversarial Clustering System (ACS) was proposed and validated using three different databases of fine-art paintings. In contrast to previous studies, the proposed method links the unsupervised clustering with the supervised classification through an optimization algorithm that iteratively improves the clustering process according to given objective criteria. Experimental results revealed that the proposed method leads to efficient automatic labeling of artworks based on scene composition, types, and shapes of objects, presence or lack of perspective, dominating directions of edges, the sharpness of edges, and color palette.

A comparison with the standard k-means clustering method showed that the ACS method leads to a stronger separation between clusters and gives higher reliability values between the machine-made and human-made labels. A comparison between different types of image features indicated that network embeddings, obtained from networks pre-trained to recognize fine art, provide more efficient clustering than features obtained from networks pre-trained on a general object recognition task. Clustering of Australian Aboriginal art paintings, which were never-before labeled by art experts, led to the discovery of categorization criteria and art categories that could be useful for storage and retrieval of Australian Aboriginal art collections yet to be analyzed by experts.

In future studies, the proposed ACS method will be evaluated using more complex and diverse cluster techniques, deep network embeddings, and supervised classification methods. In addition, future research will investigate the unsupervised classification of paintings subregions or patches that can provide a better resolution and details of the stylistic composition of the artworks.

REFERENCES

[1] C. Sandoval, E. Pirogova, and M. Lech, ''Two-stage deep learning approach to the classification of fine-art paintings,'' *IEEE Access*, vol. 7, pp. 41770–41781, 2019, doi: 10.1109/ACCESS.2019.2907986.

[2] *Introduction to Art Concepts*. Accessed: Jun. 2, 2020. [Online]. Available: https://courses.lumenlearning.com/atd-sac-artappreciation/

[3] *The Art Story: Modern Art Movement Timeline*. Accessed: May 20, 2020. [Online]. Available: http://www.theartstory.org/section_movements_timeline.htm

[4] D. J. DeWitt, R. M. Larmann, and M. K. Shields, *Gateways to Art: Understanding the Visual Arts*, 2nd ed. London, U.K.: Thames & Hudson, 2015.

[5] M. M. AlyanNezhadi, H. Dabbaghan, S. Moghani, and M. Forghani, ''A painting artist recognition system based on image processing and hierarchical SVM,'' in *Proc. 5th Conf. Knowl. Based Eng. Innov. (KBEI)*, Feb. 2019, pp. 537–541, doi: 10.1109/KBEI.2019.8734911.

[6] Z. Falomir, L. Museros, I. Sanz, and L. Gonzalez-Abril, ''Categorizing paintings in art styles based on qualitative color descriptors, quantitative global features and machine learning (QArt-learn),'' *Expert Syst. Appl.*, vol. 97, pp. 83–94, May 2018, doi: 10.1016/j.eswa.2017.11.056.

[7] S. Agarwal, H. Karnick, N. Pant, and U. Patel, "Genre and style based painting classification," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Jan. 2015, pp. 588–594, doi: 10.1109/WACV.2015.84.
[8] F. S. Khan, S. Beigpour, J. van de Weijer, and M. Felsberg, "Painting-91: A large scale database for computational painting categorization," *Mach. Vis. Appl.*, vol. 25, no. 6, pp. 1385–1397, Aug. 2014, doi: 10.1007/s00138-014-0621-6.
[9] R. S. Arora and A. Elgammal, "Towards automated classification of fine-art painting style: A comparative study," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 3541–3544.
[10] L. Shamir, T. Macura, N. Orlov, D. M. Eckley, and I. G. Goldberg, "Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art," *ACM Trans. Appl. Perception*, vol. 7, no. 2, pp. 1–17, Feb. 2010, doi: 10.1145/1670671.1670672.
[11] Y. Bar, N. Levy, and L. Wolf, "Classification of artistic styles using binarized features derived from a deep neural network," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2015, pp. 71–84, doi: 10.1007/978-3-319-16178-5_5.
[12] S. Karayev, A. Hertzmann, M. Trentacoste, H. Han, H. Winnemoeller, A. Agarwala, and T. Darrell, "Recognizing image style," in *Proc. Brit. Mach. Vis. Conf.*, 2014, pp. 1–20, doi: 10.5244/C.28.122.
[13] F. Bianconi and R. Bello-Cerezo, "Evaluation of visual descriptors for painting categorisation," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 364, Jun. 2018, Art. no. 012037, doi: 10.1088/1757-899X/364/1/012037.
[14] W.-T. Chu and Y.-L. Wu, "Image style classification based on learnt deep correlation features," *IEEE Trans. Multimedia*, vol. 20, no. 9, pp. 2491–2502, Sep. 2018, doi: 10.1109/TMM.2018.2801718.
[15] B. Saleh and A. Elgammal, "Large-scale classification of fine-art paintings: Learning the right metric on the right feature," *Int. J. Digit. Art Hist.*, vol. 1, no. 2, pp. 71–93, Oct. 2016, doi: 10.11588/dah.2016.2.23376.
[16] N. van Noord, E. Hendriks, and E. Postma, "Toward discovery of the artist's style: Learning to recognize artists by their artworks," *IEEE Signal Process. Mag.*, vol. 32, no. 4, pp. 46–54, Jul. 2015, doi: 10.1109/MSP.2015.2406955.
[17] G. Folego, O. Gomes, and A. Rocha, "From impressionism to expressionism: Automatically identifying van Gogh's paintings," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 141–145, doi: 10.1109/ICIP.2016.7532335.
[18] A. Elgammal, M. Mazzone, B. Liu, D. Kim, and M. Elhoseiny, "The shape of art history in the eyes of the machine," in *Proc. 32nd AAAI Conf. Artif. Intell.*, New Orleans, LA, USA, Feb. 2018, pp. 1–9. [Online]. Available: https://arxiv.org/abs/1801.07729v2
[19] E. Cetinic, T. Lipic, and S. Grgic, "Fine-tuning convolutional neural networks for fine art classification," *Expert Syst. Appl.*, vol. 114, pp. 107–118, Dec. 2018, doi: 10.1016/j.eswa.2018.07.026.
[20] T. Sun, Y. Wang, J. Yang, and X. Hu, "Convolution neural networks with two pathways for image style recognition," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4102–4113, Sep. 2017, doi: 10.1109/TIP.2017.2710631.
[21] A. Lecoutre, B. Négrevergne, and F. Yger, "Recognizing art style automatically in painting with deep learning," in *Proc. ACML*, 2017, pp. 327–342.
[22] W. R. Tan, C. S. Chan, H. E. Aguirre, and K. Tanaka, "Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3703–3707, doi: 10.1109/ICIP.2016.7533051.
[23] K.-L. Hua, T.-T. Ho, K.-A. Jangtjik, Y.-J. Chen, and M.-C. Yeh, "Artist-based painting classification using Markov random fields with convolution neural network," *Multimedia Tools Appl.*, vol. 79, nos. 17–18, pp. 12635–12658, Jan. 2020, doi: 10.1007/s11042-019-08547-4.
[24] S. Bianco, D. Mazzini, P. Napoletano, and R. Schettini, "Multitask painting categorization by deep multibranch neural network," *Expert Syst. Appl.*, vol. 135, pp. 90–101, Nov. 2019, doi: 10.1016/j.eswa.2019.05.036.
[25] S.-H. Zhong, X. Huang, and Z. Xiao, "Fine-art painting classification via two-channel dual path networks," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 1, pp. 137–152, Jan. 2020, doi: 10.1007/s13042-019-00963-0.
[26] E. Cetinic, T. Lipic, and S. Grgic, "Learning the principles of art history with convolutional neural networks," *Pattern Recognit. Lett.*, vol. 129, pp. 56–62, Jan. 2020, doi: 10.1016/j.patrec.2019.11.008.
[27] A. Belhi, A. Bouras, A. K. Al-Ali, and S. J. Foufou, "A machine learning framework for enhancing digital experiences in cultural heritage," *J. Enterprise Inf. Manage.*, Jun. 2020, doi: 10.1108/JEIM-02-2020-0059.
[28] H. Yang and K. Min, "Classification of basic artistic media based on a deep convolutional approach," *Vis. Comput.*, vol. 36, no. 3, pp. 559–578, Mar. 2020, doi: 10.1007/s00371-019-01641-6.
[29] M. O. Kelek, N. Calik, and T. Yildirim, "Painter classification over the novel art painting data set via the latest deep neural networks," *Procedia Comput. Sci.*, vol. 154, pp. 369–376, Jan. 2019, doi: 10.1016/j.procs.2019.06.053.
[30] N. Garcia, B. Renoust, and Y. Nakashima, "ContextNet: Representation and exploration for painting classification and retrieval in context," *Int. J. Multimedia Inf. Retr.*, vol. 9, no. 1, pp. 17–30, Mar. 2020, doi: 10.1007/s13735-019-00189-4.
[31] Q. Yang, Y. Zhang, W. Dai, and S. J. Pan, *Transfer Learning*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2020.
[32] E. Cetinic, T. Lipic, and S. Grgic, "A deep learning perspective on beauty, sentiment, and remembrance of art," *IEEE Access*, vol. 7, pp. 73694–73710, 2019.
[33] R. Sizyakin, B. Cornelis, L. Meeus, H. Dubois, M. Martens, V. Voronin, and A. Pizurica, "Crack detection in paintings using convolutional neural networks," *IEEE Access*, vol. 8, pp. 74535–74552, 2020, doi: 10.1109/ACCESS.2020.2988856.
[34] M. Spehr, C. Wallraven, and R. W. Fleming, "Image statistics for clustering paintings according to their visual appearance," in *Proc. Comput. Aesthetics, Eurograph. Workshop Comput. Aesthetics Graph., Vis. Imag., Eurograph.*, 2009, pp. 57–64.
[35] E. Gultepe, T. Edward. Conturo, and M. Makrehchi, "Predicting and grouping digitized paintings by style using unsupervised feature learning," *J. Cultural Heritage*, vol. 31, pp. 13–23, May 2018, doi: 10.1016/j.culher.2017.11.008.
[36] A. Coates, A. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 215–223.
[37] Y. Deng, F. Tang, W. Dong, F. Wu, O. Deussen, and C. Xu, "Selective clustering for representative paintings selection," *Multimedia Tools Appl.*, vol. 78, no. 14, pp. 19305–19323, Jul. 2019, doi: 10.1007/s11042-019-7271-7.
[38] S. A. Shah and V. Koltun, "Robust continuous clustering," *Proc. Nat. Acad. Sci. USA*, vol. 114, no. 37, pp. 9814–9819, Sep. 2017, doi: 10.1073/pnas.1700770114.
[39] G. Castellano and G. Vessio, "Deep convolutional embedding for digitized painting clustering," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 2708–2715.
[40] X. Guo, X. Liu, E. Zhu, and J. Yin, "Deep clustering with convolutional autoencoders," in *Neural Information Processing*, vol. 10635. Cham, Switzerland: Springer, 2017, pp. 373–382, doi: 10.1007/978-3-319-70096-0_39.
[41] J. Guérin, O. Gibaru, S. Thiery, and E. Nyiri, "CNN features are also great at unsupervised classification," 2017, *arXiv:1707.01700*. [Online]. Available: http://arxiv.org/abs/1707.01700, doi: 10.5121/CSIT.2018.80308.
[42] D. Shitov, E. Pirogova, T. A. Wysocki, and M. Lech, "Learning acoustic word embeddings with dynamic time warping triplet networks," *IEEE Access*, vol. 8, pp. 103327–103338, 2020, doi: 10.1109/ACCESS.2020.2999055.
[43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[44] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015, doi: 10.1007/s11263-015-0816-y.
[45] *Visual Art Encyclopedia*. (2018). [Online]. Available: https://www.wikiart.org
[46] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.
[47] T. Calinski and J. Harabasz, "A Dendrite method for cluster analysis," *Commun. Statist., Theory Methods*, vol. 3, no. 1, pp. 1–27, 1974, doi: 10.1080/03610927408827101.
[48] K. Krippendorff, "Computing Krippendorff's alpha-reliability," Annenberg School Commun., Dept. Papers, Philadelphia, PA, USA, Tech. Rep. 43, 2011. [Online]. Available: https://repository.upenn.edu/asc_papers/43
[49] H. W. Kuhn, "The Hungarian method for the assignment problem," *Nav. Res. Logistics Quart.*, vol. 2, nos. 1–2, pp. 83–97, Mar. 1955.
[50] E. L. Allwein, R. E. Schapire, and Y. Singer, "Reducing multiclass to binary: A unifying approach for margin classifiers," *J. Mach. Learn. Res.*, vol. 1, pp. 113–141, Dec. 2000.

[51] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA, USA: Addison-Wesley, 1989.
[52] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seedin," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms (SODA)*, 2007, pp. 1027–1035.
[53] C. Florea, C. Toca, and F. Gieseke. *Paintings Dataset for Recognizing the Art Movement, Pandora18K*. Accessed: Jun. 2, 2020. [Online]. Available: http://imagpub.ro/pandora/pandora_download.html
[54] C. Florea, C. Toca, and F. Gieseke, "Artistic movement recognition by boosted fusion of color structure and topographic description," presented at the IEEE Winter Conf. Appl. Comput. Vis. (WACV), Mar. 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7926652/
[55] W. Caruana, L. G. Corrin, S. Gilchrist, P. McClusky, and S. A. Museum, *Ancestral Modern: Australian Aboriginal Art London: Seattle Art Museum*. New Haven, CT, USA: Yale Univ. Press, 2012, p. 173.
[56] J. A. A. Gallery. *Japingka Aboriginal Art*. Accessed: 2018. [Online]. Available: https://japingkaaboriginalart.com/
[57] T. Artery. *The Artery Contemporary Aboriginal Art*. Accessed: 2020. [Online]. Available: https://www.artery.com.au/artworks
[58] *Artlandish Aboriginal Art Gallery*. Accessed: 2020. [Online]. Available: https://www.aboriginal-art-australia.com/
[59] K. O. Gallery. *Kate Owen Gallery Contemporary Aboriginal Art*. Accessed: 2020. [Online]. Available: https://www.kateowengallery.com/
[60] Aboriginal Art Association of Australia. *Code of Ethics & Business Practice*. Accessed: Jun. 2020. [Online]. Available: https://www.aboriginalart.org.au/
[61] D. J. Ketchen and C. L. Shook, "The application of cluster analysis in strategic management research: An analysis and critique," *Strategic Manage. J.*, vol. 17, no. 6, pp. 441–458, Jun. 1996.
[62] D. Wroth. *Australian Aboriginal Art Symbols and Their Meanings, Japingka Aboriginal Art Gallery*. Accessed: Dec. 2020. [Online]. Available: https://japingkaaboriginalart.com/articles/aboriginal-art-symbols/

CATHERINE SANDOVAL (Member, IEEE) received the B.E. degree in electronic engineering from Pontificia Javeriana University, Colombia, in 2001, and the M.S. degree in electronic engineering from RMIT University, Australia, in 2015, where she is currently pursuing the Ph.D. degree in electrical and electronic engineering. Her research interests include artificial intelligence, machine learning, art analysis, deep learning, and image processing.

ELENA PIROGOVA received the B.Eng. degree (Hons.) in chemical engineering from the National Technical University of Ukraine, in 1991, and the Ph.D. degree in biomedical engineering from Monash University, Australia, in 2002. She is currently a Professor of biomedical engineering with the School of Engineering. RMIT University, Australia. Her research interests include biomedical electronics and instrumentation, bioelectromagnetics, and protein modeling.

MARGARET LECH (Member, IEEE) received the M.S. degree in physics from Maria Curie-Skłodowska University, Poland, the M.S. degree in biomedical engineering from the Warsaw University of Technology, Poland, and the Ph.D. degree in electrical engineering from The University of Melbourne, Australia. She is currently a Professor of signal processing and artificial intelligence with the School of Engineering, RMIT University, Australia. Her research interests include machine learning applications in speech and image processing, system modeling, and optimization.