International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 3225-3235

https://doi.org/10.70517/ijhsa463271

Using data visualization techniques to analyze the similarities and differences between Giacometti's style and that of other artists.

Jing Ma[1,*]
[1] College of Child Education and Development, Hangzhou Campus, Zhejiang Normal University, Hangzhou, Zhejiang, 321004, China
Corresponding authors: (e-mail: MJZJHZ197765@126.com).

Abstract As a representative of post-war surrealist sculptor, Alberto Giacometti's painting and sculpture style has experienced from realism to surrealism and finally back to the deeper study of human beings, forming his own unique style of work. In this paper, text clustering method is used for data visualization, TF-IDF text keyword extraction method and K-means++ clustering algorithm are used to mine Giacometti's painting art style, and two-pair and two-independent t-test methods are used to analyze the differences existing in different painting art styles. The study categorized Giacometti's painting art styles into 8 clusters and the clustering module value = 0.9146 with Q value far <0.3, clustering is significant. There are significant differences in different artistic painting styles, Surrealist vs. Impressionist, Impressionist vs. Fauvist, Romantic vs. Expressionist styles are significantly different. There is no significant difference between the styles of Surrealists and Romantic Poets and Expressionists.

Index Terms TF-IDF, keyword extraction, k-means++, artistic style

I. Introduction

Alberto Giacometti, born in Stampa, Switzerland, is a master of existentialist sculpture [1]-[3] as well as a unique visual experience in the field of contemporary art that opens up new perspectives in figurative painting. He has been influenced by Freud, interested in Sartre's existentialism, studied ancient art, and has been immersed in Cubism and Surrealism [4], [5]. After the trauma of World War II, he eventually turned to realism, and his bean-sprouted, battered figures, which always seemed to be searching for something, revealed the trauma of war and brought him a lifetime of fame [6]-[8].

Giacometti's works are a kind of exploration, not only about human beings, but also about commonplace objects. His philosophical reflections into visual language, and his association with other artists or genres not only highlights the diversity of interpretations of the “real” in 20th century art, but also reflects the differences in artistic expression in cross-cultural contexts [9]-[12]. His sculpture rejects the action of appearances and discovers the hidden places of the heart. Giacometti is good at eliminating what interferes with vision and liberating people or objects from practical pretenses [13], [14]. This liberation intensifies his melancholy and gives him a certain strength to succeed in his quest [15]. Standing in front of Giacometti's sculptures, all that exists seems to tear at one's wounds, reeling one back into solitude [16], [17]. Giacometti's life also kept going back and forth in loneliness, eventually hiding into the elegant, slim, and tall figures he portrayed. His work travels through time and space, connecting life and death, thus reaching the realm of art [18]-[20].

In this paper, different genres of art styles are analyzed from different perspectives such as mood, subject matter, theme, words, etc., and the research text is first preprocessed by using word division and removal of stop words. From the relevant review texts, K-means++ clustering algorithm was used to analyze the co-occurrence and cluster analysis of Giacometti's painting art style, and a cluster diagram of Giacometti's painting art style was drawn. The artistic characteristics and stylistic differences of Surrealists, Impressionists, Impressionists and Fauvists, Romantics and Expressionists were analyzed using two-pair and two-independent t-tests.

II. Study design

The stylistic characteristics of the artist can be analyzed from different perspectives such as mood, subject matter, theme, and words, and the article focuses on the stylistic characteristics of the artist through words. The research is divided into the following 4 steps in design:

Step1.Select four or five less controversial representative artists from Surrealist, Impressionist, Fauvist, Romantic, and Expressionist schools respectively, and use their artworks as research objects.

Step2.Segment the art works by jieba segmentation and calculate the TF-IDF values of different keywords.

Step3. Use the improved cluster analysis model to cluster the collected keywords, cluster the keywords with the same characteristics together, and parse the similarities between the styles of Giacometti and other artists.

Step4.Analyze the difference between any two different genres using paired-sample t-test to parse the differences between Giacometti's style and other artists' styles.

II. A.Subjects of study

II. A. 1) Surrealists

Alberto Giacometti is a Swiss artist who early on studied traditional sculpture and medieval Italian painting. His artistic style breaks through the traditional sculptor's way of looking at things, not being bothered by the details on the surface of things, and visually creates a distanced image of the work, completely integrating the subject into the object, thus displaying an irrepressible sense of loneliness.

II. A. 2) Impressionism

Claude Monet was a central figure in Impressionism, known for his bold expressions of color and changing light. Focusing on outdoor sketching to capture the effects of instantaneous light and shadow, his works have a unique color and brushwork.

II. A. 3) The Beastie Boys

Henri Matisse was born in Le Cateau, France, and graduated from the Académie Julian, the founder and main representative of Fauvism, a French painter, sculptor and printmaker. His works are known as the master of color, full of sharp colors and strong contrasts. Good use of pure color, through the color - collision and fusion, to create both simple and expressive picture. Under his pen, color is not only a visual enjoyment, but also a conveyance of emotion.

II. A. 4) The Romantic School

Eugène Delacroix is a famous French painter and a typical representative of the Romantic school of painting. His abandoned the neoclassical style, with warm colors, pioneering has been more free, more romantic style of painting, the late rise of the impressionist painters and van Gogh's style of painting has a great influence. His major works include Freedom Leads the People and Dante and Virgil.

II. A. 5) Expressionism

Edvard Munch is an expressionist painter, the pioneer of modern expressionist painting, his works have a strong subjectivity and sad and depressing mood, mostly to life, death, love, terror and loneliness as the subject matter, with contrasting lines, color blocks, simple and generalized exaggerated modeling to express feelings and emotions, representative works include The Scream, The Dance of Life, The Night on Cal John Street and other works. Night" and other works.

II. B.Research methodology

II. B. 1) Natural Language Processing

1) Data Preprocessing

The text will be read from the database, using jieba.lcut() for word splitting, while using the Baidu deactivation table to filter deactivated words, etc., to facilitate subsequent processing.

(1) Segmentation. Use the word separation tool (such as jieba) on the given Chinese text for word separation, the text is converted into a sequence of words. Use the default exact mode words=jiaba.luct(sentence).

(2) De-deactivate words. Before the text keyword extraction, we need to do the deactivation process, mainly to remove some meaningless high-frequency words, such as "the, is, had, and, and" and so on. The frequency of these words is very high, but they do not have much semantic information, which is not very helpful for keyword extraction. At the same time, removing these useless words can also reduce the time and computation of text processing. Deactivation word processing is usually done by creating a deactivation word list containing the useless words that need to be removed. During text processing, for each word, it needs to be compared with the words in the deactivation list, and if the word belongs to the deactivation list, it will be removed, otherwise it will be retained. This can remove some useless high-frequency words and improve the accuracy and efficiency of keyword extraction. In this paper, after comparing three commonly used Chinese deactivation word lists, we decided to use Baidu deactivation word list to filter deactivated words.

2) Calculate TF-IDF Score

TF-IDF, which is called “Word Frequency - Inverse Document Frequency” in Chinese, is a statistical method used to evaluate the importance of a word in a document. It consists of two parts: word frequency and inverse document frequency, and its core idea is that the more times a word appears in a document, and at the same time, the less times it appears in other documents, then the more representative it is of that document.

TF (word frequency) refers to the frequency of a word in a document, the higher the TF, that is, the word appears in the document more often, the more important.

IDF (Inverse Document Frequency) is the inverse of how often a word appears in all documents. If a word occurs frequently in all documents, its IDF will be low, indicating that the word is not very useful in distinguishing documents. On the contrary, if a word occurs only in a few documents, then its IDF will be high, indicating that the word is of great use in distinguishing documents.

Considering TF and IDF together, the TF-IDF value of a word can be calculated, and the higher it is the more important the word is in the document. The calculation formula is as follows:

$$TF-IDF(w)=TF(w)\times IDF(w) \tag{1}$$

$TF(w)$ denotes the number of occurrences of a given word w in the text, calculated as follows:

$$TF(w)=\frac{cout(w)}{|D_i|} \tag{2}$$

where $cout(w)$ denotes the number of occurrences of the word w in the text, and $|D_i|$ denotes the number of all words in the text D_i.

$IDF(w)$ denotes the inverse document frequency of the word w, which is calculated as follows:

$$IDF(w)=\lg\frac{N}{DF(w)} \tag{3}$$

where N denotes the total number of texts and $DF(w)$ denotes the number of texts containing the word w.

II. B. 2) K-means++ cluster analysis

1) K-means algorithm

K-means algorithm is an iterative clustering algorithm based on partitioning, which aims to divide n data points into k clusters, and assign each data point to the cluster corresponding to its nearest center of mass according to the similarity degree, so as to stabilize the squared error criterion function between the data points within the clusters at the minimum value [21]. The Euclidean distance is usually used as an evaluation index of the degree of sample similarity to divide the samples, and the center of mass of each cluster is updated to be the average of all data points in the cluster through iterative calculations until the center of mass position no longer changes or a predetermined number of iterations is reached.

The squared error criterion function, i.e., the sum of squared error function (SSE), is given by Eq:

$$\min J=\sum_{i=1}^{k}\sum_{m=1}^{N}d_{mi}\,\|x_m-c_i\|^2 \tag{4}$$

$$d_{mi}=\begin{cases}1, x_m\in R_i\\ 0, x_m\notin R_i\end{cases} \tag{5}$$

where: J is the error criterion function; k is the number of clusters; R_i is the i th class; c_i is the center of mass of R_i; m is the sample number; N is the number of samples; x_m is the m th sample, i.e., a vector of correlations of the purposes to be clustered; d_{mi} indicates whether the m th sample belongs to R_i.

2) K-means++ algorithm

In general, the initial center of mass of K-means clustering is randomly selected, the results of the eye algorithm are sensitive to the initial value, the choice of initialization center of mass has a large impact on the final clustering results and running time, completely random selection may lead to the algorithm convergence is very slow or fall into the local optimum. The K-means++ algorithm can improve the randomness of the initial center of mass.

The K-means++ algorithm uses a heuristic random seeding method to find the seeds of the centers of mass for K-means clustering. The basic principle and modus operandi of the K-means++ algorithm is to assign a different probability to each sample, making it more likely that points farther away from the existing centers of mass will be selected as the initial centers of mass, in order to make sure that the distances between the initial centers of mass are as far as possible.

Assuming the number of clusters is k, the K-means++ algorithm selects the center of mass in the following way:

(1) Initialize the first center of mass: randomly and uniformly select a data sample as the first center of mass.

(2) Select other centers of mass: for the m th sample x_m in the dataset, compute the shortest Euclidean distance $d(x_m, c_i)$ from the selected center of mass c_i for the i th class.

$$d(x_m, c_i) = \sqrt{\| x_m - c_i \|^2} \tag{6}$$

Assign each sample to its nearest center of mass by distance. The probability x_m of a sample P_m being chosen as the next center of mass is:

$$P_m = \frac{d^2(x_m, c_i)}{\sum_{|h; x_h \in R_i|} d^2(x_h, c_i)} \tag{7}$$

where x_h is the h th sample in R_i.

That is, when selecting each subsequent center, the probability of each sample being selected is proportional to its distance to the nearest center that has been selected, i.e., a data point that is farther away is more likely to be the next center of mass to ensure that the distance between clusters is as far as possible.

(3) Step (2) is repeated until a predetermined k centers of mass are selected.

This initialization method allows for a wider distribution of initial centers of mass, which helps to avoid the problem of the K-means algorithm falling into a local optimal solution, thus improving the stability and accuracy of the clustering results. Literature has demonstrated through simulation studies on several cluster orientations that K-means++ consistently outperforms K-means when calculating the sum of squares of distances from points to the center of mass within a cluster, and that the K-means++ converges to a lower sum faster and has a faster running time compared to the K-means algorithm. Preliminary experiments were conducted on several real datasets and it was observed that K-means++ greatly outperforms standard K-means in terms of both speed and accuracy.

II. B. 3) Repeated Measurement Analysis of Variance (ANOVA)

Univariate analysis of variance for repeated measures data requires that the data satisfy the following three conditions: normality, equal variance, and sphericity of the covariance array consisting of each time point (principal diagonal elements (variances) are equal and non-principal diagonal elements (covariances) are zero). The sphericity of the covariance array can be tested using the Mauchly method, and when sphericity is not satisfied, the degrees of freedom can be corrected using Greenhouse-Geisse, Huynh-Feldt, or lower-bound. The SAS program in this section will test for sphericity and correct for degrees of freedom before performing the ANOVA.

Subjects are viewed as random effects and time and group are viewed as fixed effects. The repeated measures ANOVA model is where:

$$y_{ijk} = \mu + \tau_i + \varphi_k + (\tau\varphi)_{ik} + \pi_{j(i)} + \varepsilon_{ijk} \tag{8}$$

where, y_{ijk} is the measurements for the i (1 or 2) group, the $j(1,2,...,n)$ subjects, the measurements at the $k(1,2,3)$ th time point, μ is the overall mean of the measurements, τ_i is the group effect for group i, φ_k is the group effect at time point k, $(\tau\varphi)_{ik}$ is the interaction effect for the i th group and the k th time point, $\pi_{j(i)}$ is the effect for the j th subject in the i th group, ε_{ijk} is the error at time i group, j subject, and k, obeying an independent normal distribution.

II. B. 4) Paired samples t-test

The t-test for two independent samples is used to test whether two independent samples come from totals with the same mean. That is, it tests whether two independent normal totals have equal means.

1) Formulation of the null hypothesis

The two independent samples t-test involves testing whether the means of two totals are significantly different. The null hypothesis is $H_0 : \mu_1 - \mu_2 = 0$, where μ_i and μ_2 are the means of the two aggregates, respectively.

2) Selection of test statistic

The two independent samples mean test presupposes that the two independent overall distributions obey the normal distribution $N = (\mu_1, 1^2)$ and $N = (\mu_2, 2^2)$ where 12 and 22 are the variances of the two overalls, respectively.

The t-statistic is used for the test of the mean of two independent samples under the condition that the null hypothesis holds. The t-statistic for constructing two independent samples is selected and analyzed in two cases.

(1) When the two overall variances are unknown but equal, i.e., $1^2 = 2^2$, the t-test statistic constructed is:

$$t = \frac{X_1 - X_2 - (\mu_1 - \mu_2)}{S_v \sqrt{\frac{1}{n_2} + \frac{1}{n_2}}} \tag{9}$$

where n_1 and n_2 are the two-sample sizes, S_1 and S_2 are the two-sample standard deviations, and $S_v^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$, and the statistic obeys a t-distribution with degrees of freedom $n_1 + n_2 - 2$.

(2) When the two overall variances are unknown and do not want to be equal, i.e., $1^2 \neq 2^2$, the constructed t-test statistic is:

$$t = \frac{X_1 - X_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \tag{10}$$

This statistic obeys a t-distribution with modified degrees of freedom:

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2}} \tag{11}$$

In statistical analysis, if the variances of two totals are equal, it is called satisfying variance chi-square. Determining the chi-squaredness of the variances of two independent samples is the key to constructing and selecting the two independent samples t-test statistic. The LeveneF variance chi-square test can be utilized to test whether the variances of the two totals are significantly different.

To conduct LeveneF variance alignment test, the null hypothesis is first formulated $H_0 : 1^2 = 2^2$ When executing the test procedure, if the probability ρ value is less than the given level of significance (usually 0.05), the null hypothesis H_0 is rejected as unequal variances of the two aggregates; otherwise, the variances of the two aggregates are considered to be non-significantly different.

The formula for calculating the value of F statistic in F test is:

$$F = \frac{\max(S_1^2, S_2^2)}{\min(S_1^2, S_2^2)} \sim F(n_1 - 1, n_2 - 1) \tag{12}$$

where $n_i - 1$ is the degree of freedom of $\max(S_1^2, S_2^2)$ and $n_2 - 1$ is the degree of freedom of $\min(S_1^2, S_2^2)$.

3) Calculate the observations of the test statistic and their probability of occurrence

Given the null hypothesis, the test value 0 is brought into the $\mu_1 - \mu_2$ part of the t-statistic to obtain the observed value of the test statistic as well as the probability ρ value computed from the distribution function of the t-distribution.

4) Given the significance level, statistical inference results

When the probability ρ value of the test statistic is less than the significance level, the null hypothesis is rejected and the overall mean is considered to be significantly different from μ_0; conversely, the null hypothesis is accepted and the two overall means are considered to be not significantly different.

III. Findings

III. A. Analysis of Giacometti's Artistic Style

III. A. 1) Keyword co-occurrence analysis

Generally speaking, keywords with high frequency represent the common concerns of all research groups in a period of time, that is, research hotspots. In addition, the centerline reflects the importance of a node, which reflects the importance of a node in the network. Keyword frequency and centrality indicators together reflect the importance of nodes. Through the collection of review texts, it can be seen that "spiritual embodiment" and "artistic emotion"

have the highest frequency and the largest node, followed by "humanism" and "realism", but because this paper chooses "realism" and "artistic emotion" as search terms in the subject search. Table 1 shows a list of high-frequency co-occurrence words sorted according to word frequency, and it can be seen from the table that the frequency of node "loneliness" is the highest 38 times, indicating that the characters in Giacometti's paintings are always at a distance from the audience, and the viewer can feel that the characters in the paintings exude a strong sense of curvature through these images. In addition, the high-frequency keywords of Giacometti's research include "humanism" (33), "visual senses" (27), and "thematic embodiment" (23), which are consistent with the social development of the time and the real social situation in which the author lived. To sum up, the high-frequency keywords for the study of Giacometti's artistic style can be divided into three main aspects: one is the distance of loneliness, the second is the visual perception, and the third is the artistic line.

Table 1: High-frequency keywords

Serial number	Frequency	Centrality	Key words
1	38	0.09	Lonely emotion
2	33	0.18	Humanism
3	27	0.49	Visual perception
4	23	0.06	Theme
5	21	0.03	Picture reality
6	21	0.02	Art line
7	19	0	Atmosphere flow
8	19	0.05	Realism
9	17	0.06	Art essence
10	16	0.29	Picture space
11	13	0.05	Life meaning
12	11	0.31	Artistic innovation
13	11	0.06	Picture
14	11	0.35	Sense of order

Table 2 shows the list of co-occurrence words of keywords sorted by centrality. From Table 2, it can be seen that "loneliness", "vision" and "space" are the main characteristics of the art style, and the centrality of the three is 0.51, 0.47 and 0.41, respectively. The centrality of "realism", "three-dimensional picture", "spatial atmosphere", "realism" and "sense of order in the painting" is not less than 0.25, which lays the foundation for the stability of the text network. In addition, the centrality of "visual distance", "spatial depth" and "spiritual atmosphere" is not less than 0.1, which also plays a key role in the text network.

As can be seen through Tables 1 and 2, some of the keywords appear in both the high-frequency keyword co-occurrence frequency table and the high centrality keyword co-occurrence frequency table, and they are loneliness emotion (frequency 38, centrality 0.09), visual perception (frequency 27, centrality 0.49), thematic embodiment (frequency 23, centrality 0.06), realism (frequency 19, centrality 0.05), artistic spirit (Frequency 17, Centrality 0.06), Picture Space (Frequency 16, Centrality 0.29), Meaning of Life (Frequency 13, Centrality 0.05), Artistic Innovation (Frequency 11, Centrality 0.31), Picture Stereoscopy (Frequency 11, Centrality 0.06), and Sense of Order (Frequency 11, Centrality 0.35), which constitute the main hot spots of research.

III. A. 2) Keyword clustering analysis

Keyword clustering is the formation of interconnected network clusters of keywords with similar research topics in a research field, and the connotation of each cluster is identified by the title words used with high frequency in the respective containing articles. In this study, the keyword network is cropped by Pathfinder algorithm and clustering is carried out by log-likelihood ratio algorithm, and finally the name of the feature word with the highest value taken by LLR algorithm in the class is used as the name of the cluster, and the clustering results are obtained as shown in Figure 1. As can be seen from the figure, a total of eight clusters are obtained, which are Realism, Loneliness Emotion, Theme, Visual Perception, Social Reality, Spatial Atmosphere, Sense of Order, and Sense of Authenticity; the average outline value of the clusters, Silhouette (S-value), is equal to 0.992, with the S-value much greater than 0.7, and the clustering is plausible; and the clustering modularity value, Modularity Q (Q-value), is equal to 0.9146, with the Q-value much greater than 0.3, clustering is significant.

Table 2: High school cardiology keywords

Serial number	Frequency	Centrality	Key words
1	27	0.51	Solitude
2	2	0.47	Vision
3	6	0.41	Space
4	12	0.35	Real sense
5	9	0.33	picture
6	4	0.27	Space atmosphere
7	12	0.27	Realism
8	16	0.25	The sense of order in the painting
9	33	0.15	Art line
10	2	0.11	Visual distance
11	10	0.10	Spatial depth
12	37	0.07	Spiritual atmosphere
13	6	0.05	Image construction
14	9	0.05	Artistic charm
15	4	0.04	Aesthetic structure
16	4	0.03	subjectivity
17	9	0.03	Color element
18	12	0.03	Innovate
19	17	0.03	Social reality
20	22	0.03	Personal emotion
21	2	0.02	Artistic characteristics
22	2	0.02	Pass on
23	5	0.02	Cultural protection
24	6	0.02	Mind
25	6	0.02	Theme
26	7	0.02	Experience oneself
27	11	0.02	Worldliness
28	14	0.02	Social development
29	19	0.02	Picture design

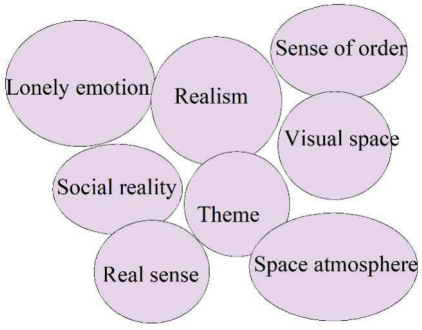

Figure 1: Art style keyword cluster map

III. B. Processing analysis of NLP in different artist styles

Using the jieba module and through Python programming, the poetry ensembles of each genre are subdivided, and the TF-HDF values of specific keywords in a particular genre are calculated after removing deactivated words and single-word words, and a total of 90,202 keywords in each genre are obtained as TF-IDF values. Partial results are shown in Table 3.

$$TF-IDF_i = k_i \times \left[\ln\left(\frac{1+N}{1+n_i} \right) + 1 \right] \tag{13}$$

Table 3: Partial results of TF-IDF values

Key words	Surrealism	Impressionism	Fauvism	Romanticism	Expressionism
Society	83.000	87.000	24.000	145.000	61.000
Reality	62.000	60.000	18.000	58.000	56.000
Affections	51.000	30.000	36.000	60.000	51.000
…	…	…	…	…	…
Colour	6.386	0.000	0.000	0.000	0.000
Theme	6.386	0.000	0.000	0.000	0.000
Vision	7.000	34.000	4.000	33.000	30.000
Realism	0.000	0.000	0.000	0.000	3.015
Space	0.000	0.000	0.000	0.000	3.015
Abstraction	0.000	0.000	0.000	0.000	3.015
Spirit	0.000	0.000	0.000	0.000	3.015

III. C. Statistical analysis

III. C. 1) K-means++ cluster analysis

Taking keywords as samples and factions as indicators, the keywords are clustered using the k-means++ cluster analysis model to cluster the keywords with the same characteristics together, setting the number of clusters k=50, and the operation is realized by SPSS software. Table 4 shows part of the results of clustering, from the table it can be seen that the same points of the artistic style of painting genres are emotion, reality, three-dimensionality, atmosphere and color.

Table 4: Clustering results of TF-IDF

Key words	Surrealism	Impressionism	Fauvism	Romanticism	Expressionism	Categories
Affections	6.000	35.000	1.000	22.000	17.00	1
Reality	10.000	31.000	2.000	13.000	7.000	1
Solid	8.000	25.000	2.000	16.000	19.000	1
…	…	…	…	…	…	…
Space	13.000	12.000	2.000	39.000	27.000	25
Develop	30.000	23.000	25.000	61.000	45.000	27
Culture	45.000	58.000	21.000	56.000	45.000	28
…	…	…	…	…	…	…
Atmosphere	3.379	9.158	0.000	1.278	6.015	50
Colour	0.000	8.239		1.278	3.692	50
Vision	1.278	10.491	0.000	2.573	1.278	50

The number of keywords in each cluster is shown in Figure 2. Based on Figure 2, the number of keywords in most of the categories is <1000, and only the 9th, 10th, 22nd, and 40th categories have >3000, of which the 22nd category is as high as 77,282, indicating that there is also a strong spatial clustering between the TF-IDF values of keywords of different genres. The TF-IDF values of different genres for each category are averaged to form the TF-IDF matrix of categories and genres, and some results are shown in Table 5.

III. C. 2) ANOVA and t-tests

Repeated measures ANOVA was used to test whether there were significant differences between the TF-IDF values of the keywords for different painting genres, and paired samples t-test was used to test the size of the difference between any 2 different genres, and the operation was realized through SPSS software. The results of the test are shown in Table 6. From the table, the value of Mauchly's sphericity test < 0.05 indicates that there is dissatisfaction with soccer symmetry between the TF-IDF values of the keywords of each painting genre, and it is necessary to use the correction coefficient to correct the degrees of freedom.

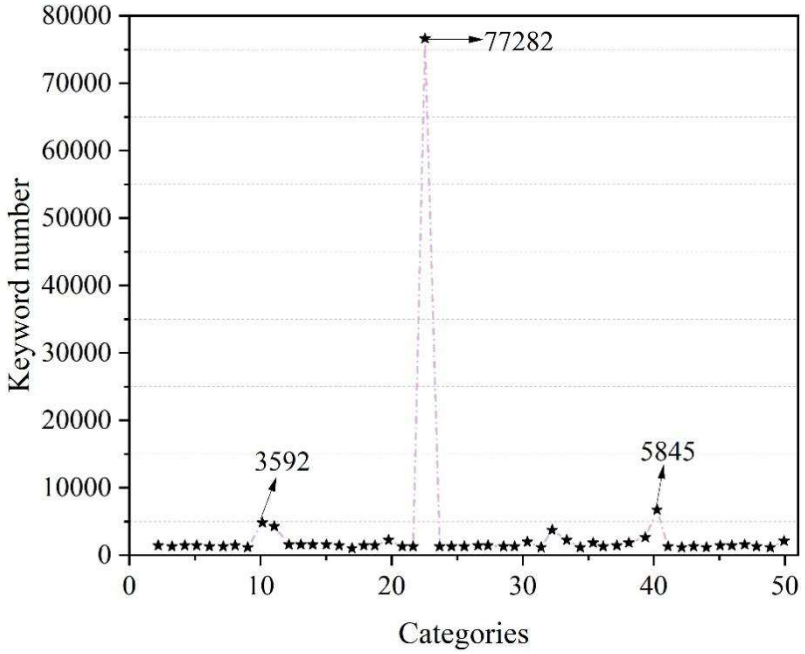

Figure 2: The number of keywords per category

Table 5: The mean value of TF-IDF for each category keyword

Categories	Surrealism	Impressionism	Fauvism	Romanticism	Expressionism
1	9.780	34.287	6.379	14.553	13.719
2	0.000	44.733	1.294	30.631	5.854
3	24.000	77.000	16.000	11.000	34.000
…	…	…	…	…	…
23	18.894	23.754	8.279	57.973	30.436
24	7.375	17.518	2.963	32.174	24.127
25	30.000	22.000	25.000	61.000	45.000
…	…	…	…	…	…
48	30.758	36.375	13.783	67.752	30.000
49	27.000	61.000	26.000	57.000	57.000
50	0.973	8.054	0.551	1.847	1.715

Table 6: Mauchly sphericity test results

Internal effect	Moche W	Approximate card	Freedom	Significance	Epsilon		
					Greenhouse-geisler	Sinfield	Lower limit
Period	0.408	47.384	9	0.000	0.773	0.829	0.254

The results of the corrected tests are shown in Table 7. As can be seen from the table, at the given significance level α=0.01, the corrected p-values are all $<\alpha$, i.e., the painting styles between different painting genres are considered to be significantly different at the 99% confidence level.

Table 7: Multivariate test results

Effect		T test value	F	Assumed freedom	Error degree of freedom	Significance
Period	Bielet trajectory	0.593	16.783b	5	49	0.000
	Wilk lambda	0.417	16.783b	5	49	0.000
	Hotlin trajectory	1.446	16.783b	5	49	0.000
	Roy's biggest root	1.446	16.783b	5	49	0.000

A total of C_5^2 = 10 line tests are required to test any 2 different genres using the paired samples t-test, and Figure 3 shows the collated test results. Given the significance level α =0.05, the styles of the Surrealists and the Impressionists, the Impressionists and the Fauvists, and the Romantics and the Expressionists are all considered to be significantly different at the 95% confidence level. There is no significant difference between the styles of Surrealists and Romantic Poets and Expressionists.

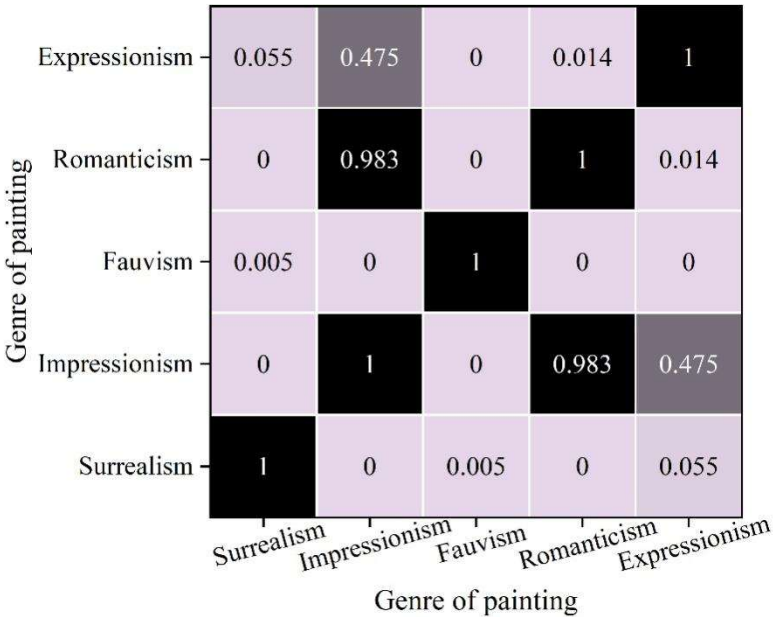

Figure 3: The p-value of the paired sample t-test

To sum up, Surrealist Giacometti shares the same artistic style with the writers of paintings of the Impressionist, Fauvist, and Expressionist schools in terms of emotion, color, reality, culture, and society, whereas he differs from the Romantic authors in terms of space, atmosphere, and romance.

IV. Conclusion

This paper analyzes the characteristics of Giacometti's art style by combining the keywords, and based on the clustering results, it is concluded that Giacometti's art style has realism, lonely emotion, theme, visual perception, social reality, spatial atmosphere, and sense of order. Through TF-HDF, K-means++ clustering method to prove the existence of differences in the painting art style of surrealists, impressionists, fauvists, romantics and expressionists were analyzed, and the results of the study showed that the corrected p-values were <0.01, and there were significant differences in the painting styles between different painting genres, and the similarities of the painting art styles were emotion, color, reality, culture The similarities of painting art styles are emotion, color, reality, culture, and society, and the differences are space, atmosphere, and romance.

References

[1] Jedlińska, E. (2021). A discourse between and about two artists: Alberto Giacometti and Francis Bacon. Art Inquiry, (23), 215-238.
[2] Wenjun, Z. H. U. (2020). Bodies in Samuel Beckett's Theatre from the Perspective of Alberto Giacometti's Sculptures. Journal for Literary and Intermedial Crossings, 5, e1.
[3] Rodin, A. (2020). Rodin-Giacometti. Fondation Pierre Gianadda, Martigny, Suisse, du, 27.
[4] Costello, L. (2015). Portraiture and the Ethics of Alterity: Giacometti vis-à-vis Levinas. October, (151), 62-77.
[5] Eidhan, A. F., & ZBOON, A. J. (2024). Self-awareness in nihilistic philosophy to interpret aesthetic experience (Sculptor Alberto Giacometti as a model). Al-Academy, (113).
[6] Elias, A. J. (2024). Sixteen." The Giacometti Effect": Reconsidering Beauford Delaney's 1966 Portrait Bust of James Baldwin. In Speculative Light: The Arts of Beauford Delaney and James Baldwin (pp. 240-248). Duke University Press.
[7] Frank, C. (2018). Alberto Giacometti's Caress/Despite the Hands: Developing and vanished life as a reversible figure—nucleus of an adequate expression of the struggle for the acknowledgment of space and time?. In The Female Body (pp. 65-89). Routledge.
[8] Fiduccia, J. (2022). Scale of the Nation: Alberto Giacometti's Miniature Monument. Art History, 45(1), 126-156.
[9] Bush, E. (2019). Two Brothers, Four Hands: The Artists Alberto and Diego Giacometti by Jan Greenberg. Bulletin of the Center for Children's Books, 72(8), 345-346.
[10] Stenson, E. (2022). Alberto Giacometti's artistry reenters the limelight in" Towards the Ultimate Figure". UWIRE Text, 1-1.
[11] Taylor, M. R. (2022). Fathers and Sons: Dalí, Giacometti, and the Legend of William Tell. Journal of Surrealism and the Americas, 13(1).
[12] Wilson, L. (2022). Central Organizing Fantasies and Their Relationship to the Signature Style and Masterworks of Two Artists. American

Imago, 79(2), 335-363.
[13] Rae, J. (2016). A Giacometti portrait, postmodern emergence, and creativity. Methodological Innovations, 9, 2059799116672876.
[14] Hall, J. (2020). Freud and sculpture. The Burlington Magazine, 162(1404), 272.
[15] Shaw, J. (2016). Critically discuss the development of Giacometti's process of portrait painting. Word count: 3296.
[16] Perrey, B. (2017). Borges's Blindness and Giacometti's Eyes. In In (ter) discipline (pp. 32-40). Routledge.
[17] Labrusse, S. (2022). Alberto Giacometti ou l'art critique. ALKEMIE. Revue semestrielle de littérature et philosophie, 2(30), 171-178.
[18] McKendrick, J. (2017). Discomfort ye: Alberto Giacometti's stations on the route to perception. TLS. Times Literary Supplement, (5964), 20-21.
[19] Di Monte, M. G. (2015). Persons and Things. The Temptation of Contemporary Art in Giacometti and Bacon. In Gesicht und Identität/Face and Identity (pp. 63-77). Brill Fink.
[20] Monument, A. G. S. M. (2022). Scale of the Nation. Art History, 45(1), 126-156.
[21] Liu Meidan. (2025). A Digital Classification Method of Rural Public Affairs Governance Based on K-means Clustering Algorithm. International Journal of High Speed Electronics and Systems,(prepublish).