Semantic Description of Digital Images of Printmaking Art Based on Image Decomposition

Jing Liu[1*]

[1]Jinzhong College of Information, Jinzhong, 030800, China
lianyiowen@163.com

Corresponding author: Jing Liu, lianyiowen@163.com

Abstract. In this paper, the semantic description of digital images of printmaking art is deeply studied and analyzed by the method of image decomposition. On the semantic significant region extraction of digital images of printmaking art, an image significant region extraction algorithm based on the theory of low-rank matrix decomposition is proposed. The principle of the algorithm is to divide the semantic content of the image into significant and non-significant parts from the perspective of low-rank matrix decomposition theory. The non-significant part of the matrix can theoretically correspond to a low-rank structure due to the high redundancy of its content, while the significant target has a high variance of one or more features and thus can correspond to a sparse component. In this way, the salient regions in the image are extracted and an effective representation model is provided to satisfy further image semantic annotation. The results of the saliency map based on the proposed algorithm are compared with seven other algorithms. Experimental results of the proposed algorithm are given on two current eye-movement databases, MIT, and Bruce, and the MSRA database. The algorithm performs better in low-fuel images, the proposed algorithm significantly outperforms the other methods, and the proposed algorithm is more consistent with the human visual attention process. A multi-task joint sparse representation of printmaking art digital image classification algorithm is proposed. Printmaking art figures have richer structural information. Accordingly, a structured analysis algorithm for printmaking art digital images is proposed. Firstly, the printmaking art digital image is decomposed into four parts: painting body, inscription, white space, and seal, then a series of unique color and texture features are extracted according to the visual and creative characteristics of each part, and finally, a multi-task joint sparse representation model is introduced to effectively fuse the features of the four parts and classify them. Experiments on a large set of digital images of printmaking art show that the proposed structure analysis algorithm can effectively decompose the structure of the painting images, and the performance of the classification strategy based on the multi-task joint sparse representation outperforms that of the global-based classification method.

Keywords: image decomposition; printmaking art; digital images; semantic description
DOI: https://doi.org/10.14733/cadaps.2024.S2.69-83

1 INTRODUCTION

Printmaking is a kind of visual art with "indirectness" and "pluralism", through the plate making and printing plate to transfer the image to the carrier. Printmaking has a long history at home and abroad, and can be divided into four major types of prints: letterpress, intaglio, lithography, and perforation, depending on the creation process [20]. The unique plate material and creation process also brings a unique artistic language for printmaking, such as the vicissitudes and brightness of woodblock prints, the solidity of copper block prints, the graininess of lithographs. The production characteristics of screen-prints break through the limitations of the material or shape of the substrate, such as curved and uneven surfaces, as well as fabric, metal, glass, etc. can be used as the carrier of screen-prints. Due to the characteristics of "marks" and "plural", printmaking is a type of painting that relies heavily on technological change and has more diversified forms of expression and space for development than other types of painting [14]. At present, the rapid development of digital technology has changed people's habits and aesthetic concepts and has brought about radical changes in the creation of visual art, and digital art has also emerged. As an emerging type of printmaking, digital printmaking has brought a new vitality to traditional printmaking art - greatly expanding the artistic language of printmaking, bringing artists more creative inspiration, and enriching the content and form of printmaking. In contrast to traditional printmaking, digital printmaking is a form of printmaking that uses digital equipment and technology to create, make plates, and print onto substrates [5]. Therefore, it is necessary to explore the artistic language of digital printmaking, the relationship between digital printmaking and the art of painting, and the way of using the digital printmaking medium and the visual effects it is suitable for creating.

The accuracy and speed of image recognition and image understanding technologies are still far from meeting the needs of the people. There is also a great demand and reliance on image analysis and image understanding for the public security system, where the case processing process needs to draw on various comprehensive information of the tracked object. How to quickly search and locate the tracking target from a large number of video images according to the existing text or picture clues, and identify the activity scene of the tracking object as well as the emotional semantics, and synthesize the obtained large amount of image information to finally give accurate and comprehensive information about the tracking target, the results of these image understanding and processing can provide strong and useful auxiliary information for the public security department to deal with the case [13]. Similarly, some human image recognition judgment analysis ability is very strong, can capture the subtle changes in the image, and timely find the connection between things, but when facing the processing of massive data, manual labeling is difficult to cope with. If the process of human processing and analysis of images can be implemented in the automatic image recognition and processing system, it can effectively improve the accuracy of the system to process images automatically. Image fusion is a multi-level, multi-level, multi-directional comprehensive processing of two or more images with complementary information and redundant characteristics, reducing the redundancy and ambiguity of images, making the fused image information richer, more accurate, and more reliable, and providing a technical basis for subsequent image detection, image recognition, image classification, image understanding, and other related research and applications [2]. The blowout growth of digital images has put forward higher demands for image understanding. First, the intelligence of image understanding requires that the semantic gap between image content representation and image semantic understanding be narrowed as much as possible. Secondly, the "spurt" of massive images also puts higher demands on the temporal performance of image processing. Therefore, automatic semantic annotation of images efficiently and accurately is urgent.

With the development of information technology, digital images have become an important carrier for information-bearing and expression, and are deeply integrated into daily life, so people's demand for editing digital images has increased. In the live video industry, image segmentation is one of the important technologies for editing images, and replacing the live background or editing the foreground area for live video can improve user experience and attract more users. The editing of live video is online editing, and due to the problems of content distribution and network transmission, online editing can cause delays on the viewing side, so real-time semantic segmentation methods are especially important for the live video industry. Real-time semantic segmentation has the advantage of low computational effort, so deploying real-time semantic segmentation models on cheap or low-power hardware devices can further improve the application of image segmentation and reduce the cost of devices. In addition, research on real-time semantic segmentation can also inspire other areas of real-time development, such for instance segmentation techniques that integrate semantic segmentation and target detection, and real-time semantic segmentation will undoubtedly improve the overall computational speed of instance segmentation; the work on combining semantic segmentation with target tracking also demonstrates the importance of its real-time segmentation.

2 CURRENT STATUS OF RESEARCH

From the summary of existing work related to image retrieval and image classification, it is known that the extraction of image visual feature descriptors is the basis for achieving efficient image classification performance. Since digital images of printmaking art are richer and more abstract in semantic information compared to usual natural images, research work on the automatic classification of digital images of printmaking art is still in its infancy [16]. This section briefly describes the existing work on low-level visual feature extraction and description for digital images of printmaking art [7]. Color features are an important visual information attribute, and since color information plays a pivotal role in expressing art semantics and emotions, has simple and stable characteristics compared to low-level features such as texture and shape, and color descriptors (e.g., color histogram, etc.) have better invariance for translation, rotation, scaling, etc., color features are the most effective for describing digital images of digital works of printmaking art. Most of the existing work investigating the automatic classification of digital images of printmaking art extracts global color descriptors of images to characterize the semantic content of printmaking art figures [12]. Since hue, saturation, and luminance are most commonly used by artists to create, thus existing algorithms usually extract global color features such as color histogram features, color moment features, and color correlation map features in color spaces such as HSL and CLE Luv [9]. The literature proposes a general classification framework for the digital classification problem of printmaking art, where the algorithm extracts the coefficients of the multiscale wavelet transform to characterize the semantic information of the painting images and builds a semantic model for each painting author's image by a learner. The algorithm experimentally collects the paintings of bit painters and the test results show that its classification accuracy is good. Although the algorithm achieves a finer granularity of painting classification based on the author, the algorithm is less scalable as the model requires separate semantic modeling for each author [11]. The literature proposes an automatic painting image semantic classification algorithm that combines image analysis and domain ontology techniques, opening a new research direction.

The algorithms construct visual ontologies for art images and propose the concept of non-realistic semantics of images for art images [4]. Since traditional low-level visual features cannot adequately represent the rich semantic information of printmaking art digital images, ontologies are introduced to semantic classification of printmaking art digital images to achieve high-level semantic retrieval and classification goals that match user perception [18]. For the characteristics of digital images of printmaking art, two semantic descriptions are proposed in the literature, which is specific semantics understood in a general sense and so-called non-realistic semantics. Where specific semantics includes general or specific scenes, objects, or actions. Intent recognition is a much-researched area in the field of spoken language comprehension and is an important evaluation task in human-computer dialogue technology [17]. When a user is engaged in a human-computer dialogue, intention recognition can be used to understand the user's behavioral intention that flows from the user's discourse, and by combining this behavioral intention with the content of the discourse, a more satisfactory response can be provided to the user. In information retrieval, since users often just use concise query language for information retrieval, it is difficult for search engines to identify the query intent of users with different backgrounds, resulting in search results that are often difficult to meet the personalized needs of users and match their search interests [6]. Therefore, user intent, as an important factor affecting the performance of information retrieval, is both a key issue in the field of information retrieval and a higher requirement for information retrieval systems.

With the development of Internet technology, search engines can model users' intent and interests by referring to their query history in the network or system, and provide users with personalized search results that meet their needs through the model, which can effectively improve the accuracy of search engines. In this case, the information sources used for user intent and interest modeling include users' historical search documents, users' Internet browsing records, users' download data, and relevant information in users' social networks; at the same time, the interaction information in social networks can also provide effective references for building more accurate user intent and interest models. To further bridge the semantic gap, many researchers have researched advanced semantic features from the perspective of semantic concepts. Research on concept detection sub aspects was conducted, and his study on news videos showed that a small number of noun features can improve the performance of text-based retrieval.

3 SEMANTIC DESCRIPTION ANALYSIS OF DIGITAL IMAGES OF PRINTMAKING ART FOR IMAGE DECOMPOSITION

3.1 Image Decomposition Semantic Description Design

To achieve dimensionality reduction as well as to improve the efficiency of operations, subspace feature extraction methods based on global statistical feature transformations are often used to extract the features of an image [15]. The form of data often used to represent images is a matrix, and the advantage of using matrices to organize and manage image data is that the data can be analyzed with algebraic features, giving full play to the capabilities of matrix theory in extracting features, using a two-dimensional matrix to organize information in grayscale images and a three-dimensional matrix to organize information in color images. Common analysis methods in algebraic theory are principal element analysis as a typical representative, independent element analysis to achieve blind source separation, current research hot method of non-negative matrix decomposition, and linear discriminant analysis. Principal element analysis (PCA) is an orthogonal linear transformation, i.e., K-L transformation, based on statistical features, and its role is like that of SVD. Principal element analysis is a commonly applied analysis method in dimensionality reduction of high-dimensional data and is also a typical subspace technique. The principal element analysis method uses the second-order statistics of the signal to aid in the analysis of the input vector and output data, resulting in a linear mapping relationship between the two matrices. The process of principal element analysis is to first obtain n orthogonal eigenvectors independent of each other and the vector space into which the vectors branch, the axes of the higher dimensional space being the orthogonal eigenvectors; the sample data is then represented as a linear combination of the eigenvectors.

$$C = \frac{1}{N}\sum_{k=1}^{N}\left(x_k + x^2\right)\left(x_k - x^2\right)^T \tag{1}$$

The individual eigenvectors obtained after the transformation together form the coordinate axes of the vector space and using the eigenvectors as unit orthogonal bases, the samples can represent linear combinations. The eigenvectors with large eigenvalues gather the main information energy, so the practice of discarding the eigenvectors with small eigenvalues does not affect the quality of the image.

$$V = \left[V_1^2, V_2^2, \ldots, V_n^2\right] \tag{2}$$

As an internal feature of an image, texture describes the distribution pattern of the gray values of pixel points in an image [8]. The texture feature of an image is an important visual feature of an image and is often adopted by many image recognition or image segmentation algorithms. But visually, the texture of an image is not as visually impactful as features such as shape and color. Even in some cases, texture information is not so easily recognized by the human eye. The texture can represent the distribution of grayscale of each pixel point mainly with the help of texture primitives, and extracting texture primitives is the core of solving texture problems. Methods that are often used to extract image texture features are spectral analysis methods, structural analysis methods, model analysis methods, and statistical analysis methods. Spectral analysis methods describe texture features by analyzing the frequency characteristics of the image such as time domain and frequency domain analysis implementing some kind of transformation (on a multi-scale basis) and then extracting the eigenvalues whose frequency characteristics are relatively stable during the transformation process, and this eigenvalue can be used to represent the magnitude of differences between different regions and the consistency within the same local region [10]. The Gabor transform and wavelet transform are two typical methods of spectral analysis in existing studies. Structural analysis methods consider that complex texture information can be made by the repetitive arrangement of simple texture primitives, i.e., texture primitives are regular. To use the structural analysis method to describe texture features, the type and number of texture primitives are required to be satisfied. There are many structural analysis methods for describing textures, including tree grammar structural texture representation, using graph structure definition to describe texture models, and Voronoi polygon-based texture segmentation. SIFT descriptors are usually used to extract local feature vectors of an image, as shown in Figure 1. The model analysis method sets the texture features to be distributed in the image according to the model, so the model analysis method is used to estimate the parameters of the model, in which the parameters can be used to reflect the properties of the texture primitives. The specifics of the texture features can be estimated by the model parameters, which are the core of the model analysis method.

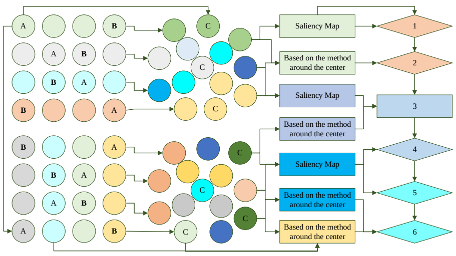

Figure 1: Two significance measures for extracting significance maps.

To study and carry out the computational model of visual attention mechanism, eye movement data were recorded by oculomotor for the distribution of subjects' gaze points in each interval. Based on the distribution of Fixation in each area of the image, a saliency map of human vision is formed as an evaluation of the performance of the computable model predicted to get the saliency map. Probabilistic and information-theoretic-based models require density estimation of features as a measure of feature saliency, and the performance of density estimation in high-dimensional spaces leads to limitations in the saliency analysis of existing methods due to the high dimensionality and complexity of the distribution of image feature spaces. Although the current saliency metric based on global contrast analysis has better improvement than local contrast, the effect is still far from meeting the perceptual characteristics of human vision for images of complex scenes. Therefore, how to construct a reasonable saliency metric for the characteristics of multiple features of images is a key issue in the visual attention mechanism.

$$s.t.X_{ij} = M_{ij}\left(i, j\right) \tag{3}$$

$$\left\|E_0\right\|^2 = \sum_i \sqrt{\sum_j \left(E_0^2\left(j.i\right)_2^4\right)} \tag{4}$$

$$S\left(P_i\right) = \sum_i \sqrt{\sum_j \left(E_j^2\left(j.i\right)_j^3\right)} \tag{5}$$

Since video consists of many image frames, when the temporal correlation between video frames is not considered, the extraction of video static image features can theoretically be accomplished using conventional image feature extraction methods. Frame-based static features usually contain color features, texture features, shape features, etc. Color features can be extracted either from the whole image or from the divided image blocks, and the computational complexity of color features is low. Texture features are inherent visual features of the object surface, containing key information about the way the object surface is organized and related to the surrounding environment; for videos with more obvious texture information, texture features can provide an effective description of the characterized video content. Shape features can be extracted from the contours or regions of an object and are usually implemented as a histogram of the edge distribution after edge detection. The static features of the video can also be obtained after obtaining the feature representation of the image with the help of a deep learning framework without considering the temporal correlation between the video frames.

$$UI = \left\{UI_1, UI_2, ..., UI_m\right\} \tag{6}$$

$$L = VA_{va} \ln P\left(SIV_i^k, SIAV_i^k a\right) \tag{7}$$

Under the Gaussian difference scale space, the scale extreme value points can be extracted by SIFT transform. The SIFT operator is insensitive to noise, rotation, scaling, and other deformation processing. The Gaussian differential function is used to identify potential points of interest according to the scale of the points of interest and the selection of invariant properties to detect the image locations on all scales and provide alternative locations for determining key points in the next step, as shown in Figure 2.

The semantic similarity algorithm defines a formula for calculating the similarity between semantics based on an information-theoretic perspective, which achieves a better semantic similarity calculation. The existing image fusion frameworks are difficult to produce accurate and effective fusion results according to the actual needs of users, and there is a semantic gap between the user's needs and the image fusion process and the resulting fused images. In this paper, a concept-based semantic annotation model is used to address the semantic similarity matching of image descriptions. The concept model can find semantically similar phrases based on the user's ideas and the intrinsic meaning of the query terms to achieve sanitization of the image fusion mechanism [1]. The model features matching based on the user's request for the image, even if the words in the user's request do not appear in the description of the image, but if both parties express the same concept and similar semantics. For example, although "Nanjing" and "Jiangsu Province" are not the same, there is a semantic affiliation between the two words. After the method can meet the semantic requirements of the user as much as possible, and improve the search rate and accuracy.

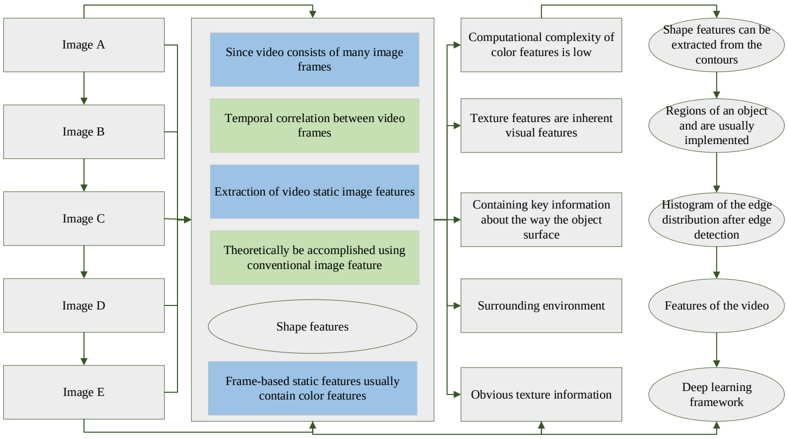

Figure 2: Hierarchy of image semantic fusion.

The core idea of building a corresponding scene semantic tree for a scene is to use a tree structure to represent the semantic structure in that scene so that its root node contains all the training images of that scene, and the semantically annotated words at the root node also represent the semantic concepts that appear most frequently in that scene. The more unique semantic concepts in the scene appear at the leaf nodes, and accordingly, the images containing the semantics of these features are also clustered at the leaf nodes, and the scene semantic tree represents a continuous progressive semantic relationship from the root node to the leaf nodes. The construction of the scene semantic tree is based on the visual features of the images; therefore, the scene semantic tree contains both the connections between semantic concepts and the relationships between visual features. When image annotation is performed, the image to be annotated is first classified to a specific scene based on its visual features, and then a leaf node is accessed from the root of the scene semantic tree corresponding to that scene by iterative matching to obtain the annotation word corresponding to that path.

3.2 Semantic Description Analysis of Digital Images of Printmaking Art

The training learning process of the model is based on making full use of many training images to extract the visual features and semantic annotation information of the images, and then using the visual features of the images, the clustering of the images is divided to form a scene semantic tree, and the structure of the scene semantic tree represents the association of the visual features between the images. Each node in the scene semantic tree aggregates some images according to the visual features, and the annotated words of these images represent the semantics represented by the concept node, and the similarity between two concepts can be measured by the length of the path between the concept nodes. The shorter the path between two nodes, the closer the semantics are represented by the nodes, and the greater the similarity takes value. Conversely, the longer the path between two nodes, the greater the difference in semantics represented by the two nodes, and the smaller the similarity takes on a value.

Although the forms of contemporary printmaking art are complex and varied, there is a constant factor behind it, namely its spiritual direction. In other words, art must keep up with the times and always point to an avant-garde spirit and attitude [19]. Artists need to adhere to the spirit of skepticism and criticism, make clear and accurate judgments about current phenomena in life, and reflect the contemporary spirit by shaping typical examples and images drawn from life through their artworks. Therefore, the value of printmaking works stems from the artist's expression of the times in which he or she lives. The development of printmaking art in contemporary times is divided into two main routes. One route is the exploration of printmaking art techniques and expressions, which in contemporary times is mainly reflected in the integration with other art forms. The other route is to serve the concept, the idea, which may be political, historical, or the artist's own. Contemporary printmaking is precisely the development of the linkage of the two routes, where the artist uses a highly personal artistic language and reflects the reality of social life and the themes of today's times through a multifaceted blend of creative techniques, as shown in Figure 3.

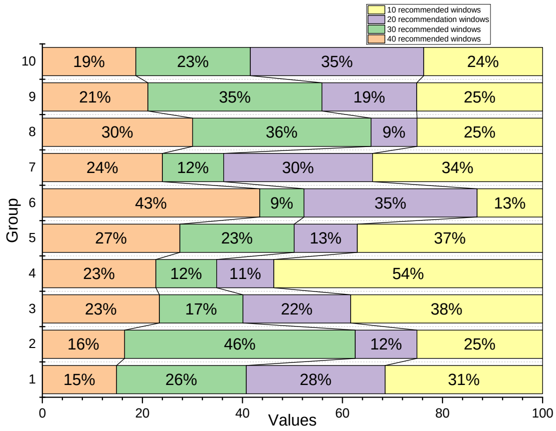

Figure 3: Relationship between the number and size of recommendation windows.

The main reason for the intent gap is that people's demand for images and videos in the retrieval process is not only limited to the content level but also involves many complex factors, such as personal emotions, emotions expressed by images and videos, etc. The influence of these factors is not enough just by measuring images and videos semantically. As social networks become increasingly widespread, people are sharing increasingly happy and sad emotions, parental details, and information about clothing, food, housing, and transportation on social network platforms, and constantly interacting and updating this information on the network. For example, in the circle of friends, when users browse the status, pictures, and information posted by friends, the act of liking and commenting can directly reflect the personal information of users' interests; this personal information provides an effective way to analyze users' interests and habits and identify their intentions. At the same time, the dissemination of information in social networks follows a strong relational characteristic, and the user's interaction behavior in social networks can directly reflect the user's social relationship and other information; this information provides a reference for analyzing the user's interests and personal information through the user's social relationship. These references help the retrieval system to reorder the images and videos obtained from the search to better match the user's retrieval intention. Therefore, how to use social information to identify users' interests, habits, etc., to better understand their retrieval intent is a very important research problem in the field of retrieval and recommendation.

The rendered color images, digital keying target real values, and foreground object maps extracted based on the color images and digital keying target real values. The entire rendering dataset consists of 100 color images rendered from 100 3D foreground objects of various morphologies. Since the exact spatial coordinates of the 3D objects and the associated transparency properties are known during the rendering process, the alpha values at each pixel point can be easily obtained. To augment the data, the extracted foreground objects can also be used with various natural backgrounds and panned, rotated, and mirrored to form more new picture scenes. In this experiment, one of the 24 classes is selected as an unknown new class, and the other 23 classes are used as known classes. t is set to 60%, and 15% of the sub videos in the selected unknown class are used as input videos to test whether they can be recognized as unknown classes. As can be seen from Figure 4, almost all the classes selected as unknown classes can be identified, proving the ability of the method to identify unknown classes. The only exception in Figure 4 is when category 3 is selected as an unknown category, it is identified as category 15, even though they are visually distinct. However, when the 15th category is selected as the unknown category, it can be correctly identified as the unknown category. This means that the 15th classifier can be used to distinguish the 3rd class from the other classes, while the 3rd classifier cannot be used to distinguish the 15th class from the other classes.

In the CCWEB video dataset, many videos do not belong to any of these 24 categories, called category-free data. To demonstrate the capability of the proposed parallel 3D CNN in identifying whether the input videos belong to the existing categories, this method is tested using the category-free videos as input. As shown in Figure 4, the recognition rate of the test videos selected from the 24 categories reaches 93%, demonstrating the good performance of the model in video classification. Due to the complexity and diversity of the category-free videos, their recognition rate is not as high as that of the category videos, but it can be improved when more sub-videos of the same video are used. When T is 50%, the recognition rate is about 55%, and when T is 70%, the recognition rate reaches about 67%. Given that traditional classifiers cannot recognize unknown categories, this method has achieved relatively good performance in recognizing unknown categories of videos.

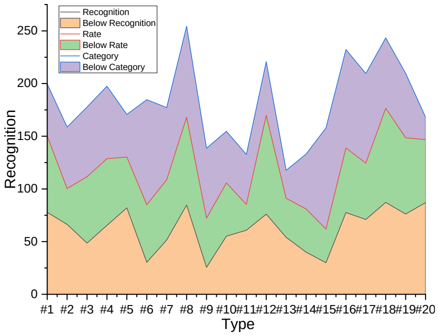

Figure 4: Identification of unknown categories.

The output process is the printing of the processed images from the computer onto the substrate through digital media. In this process, the choice of output equipment directly affects the final output effect [3]. At present, digital printmaking output is generally achieved by using an airbrush as the output device, and the ink used is generally water-based, because water-based ink has the characteristics of bright and stable ink color, strong adhesion, and low pollution. After many practices, it can be found that the type of substrate has a slight difference in the expressiveness of the picture, and the smoother the substrate, the brighter the color. However, the digital printmaking output belongs to the creation of artworks, which is different from the general printed matter. The artist conveys the artist's thoughts through the form and language of art, and uses the corresponding means and artwork presentation to complete a work of art, if this means can convey the artist's thoughts and is presented in the form and language of art, it should be a work of art. Therefore, out of consideration for the artistic value and artistic concept of digital prints, we must improve the quality of digital prints with the standard of painting art in the output technology of digital prints. Since the material involves art collection, we should try our best to avoid lowering the value of digital prints, so it is generally not recommended to use copperplate paper or photo paper as the substrate of digital prints, but water color paper and professional printmaking paper as the substrate of digital prints, to achieve the standard of art collection.

4 ANALYSIS OF RESULTS

4.1 Results of Semantic Description of Image Decomposition

The important parameter that the algorithm in this paper needs to focus on is the scale of the division of the spatial pyramid. The impact of different division scales on the experiments should not be neglected. However, it also increases the time and space cost of the algorithm, and the division scale will not improve the classification accuracy but affect the accuracy of the algorithm after reaching a certain threshold. In the experiment, we examine the change of image checking rate when the scale level=1, level=2, and level=3 respectively, as in Figure 5. According to the experimental results, within a certain range of values, as the scale increases, it can indeed effectively improve the effect of image segmentation. But the segmentation accuracy of level=3 is closer to that of level=2, and the accuracy at both scales tends to be similar with the increase of image types.

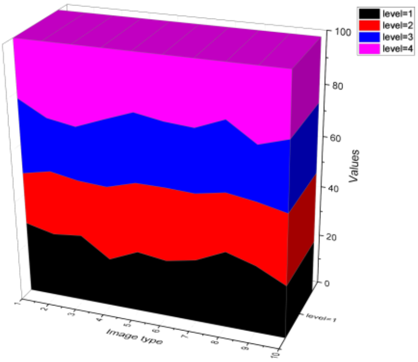

Figure 5: Comparison of image annotation accuracy.

In this paper, we propose an image segmentation algorithm based on fuzzy clustering and spatial information, which combines the color histogram and spatial pyramid of an image to achieve the extraction of image color histogram information at different scales, prompting a more flexible classification and segmentation of image sub-blocks, and because the spatial pyramid itself contains the location information of image sub-blocks at each level, it greatly improves the accuracy of image segmentation. First, the method extracts the global color features of the image. Second, the image at each pyramid scale is divided into sub-blocks equally, and feature information is extracted from each sub-block, and the obtained feature vectors are interconnected. Next, based on the difference between the image sub-blocks and the surrounding adjacent image sub-blocks, the visually significant regions in the image are calculated and the significance of each sub-block is expressed numerically as the visual information weight of that sub-block in the joint feature vector; finally, the joint feature vector obtained by weighting reflects the information of the image sub-blocks is subjected to the clustering operation to obtain relatively good classification results, as shown in Figure 6.

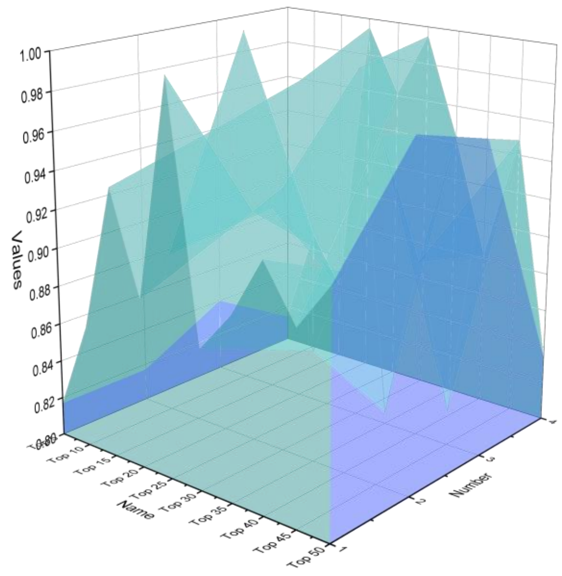

Figure 6: Accuracy comparison chart with other algorithms.

Image semantic fusion algorithm based on semantic similarity and multi-feature fusion. In this paper, we propose an image semantic fusion algorithm based on semantic similarity and multi-feature fusion, which fuses the semantic information of two images to obtain more comprehensive and accurate image information. Firstly, the global visual features of the two images are combined to classify them into corresponding scenes respectively. Secondly, determine whether the two images belong to the same scene. If they belong to different scenes, there is no semantic correlation between the two images, and the conclusion is directly given that the two images are not related; if they belong to the same scene, the semantically annotated words of each of the two images are further obtained and the semantic similarity between the two images is then obtained based on the annotated words. Finally, the result is given according to the similarity magnitude of the two images, combined with the queue value. If the similarity is less than the threshold value, the two images are judged to be disjoint; if the similarity is greater than the queue value, the two images are judged to be related and continue to select the top positions of their respective annotated words for clustering, and the clustering result is used as the fusion information of the two images. First, give me the method of calculating the similarity between semantic concepts to measure the difference between different semantics. Second, different features are assigned different weights according to their role in image differentiation, and their corresponding semantic annotated words are given the same weight. Third, in the same scene, there is partly common semantic information between the annotated words of different images and different semantics that distinguish each other, especially the different semantic information weights should be higher, so that the obtained image semantic fusion results will be more comprehensive and complete. Fourth, FCM fuzzy clustering is used in the process of forming the final semantic annotated words, and many annotated word groups with redundancy are clustered by fuzzy clustering to retain the core annotated words among them, which improves the accuracy and completeness of the image annotation.

4.2 Results of Semantic Description of Digital Images of Printmaking Art

To make the experimental results more general, a variety of existing digital keying methods will be selected as the digital keying solver as a posterior. Together with the three-part segmentation maps drawn by volunteers or the cumulative label maps generated by the recurrent neural network-based interactive keying method proposed in this chapter, the digital keying prediction results are fed into the digital keying solver and the root means the square error is obtained with the target true value, as shown in Figure 7.

In Figure 7, Tramp refers to the three-part segmentation map provided by the volunteers, Ideal refers to a locally optimal sequence i.e., the benchmark method mentioned above B1, and Proposed refers to a recommendation model based on recurrent neural networks utilized as proposed in this chapter. The plus sign is followed by a particular digital keying method. This experiment also compares the proposed recurrent neural network-based regional recommendation model with the locally optimal sequence, the user-drawn three-part segmentation map on the entire online standard digital keying dataset, and the average value on the last column of the dataset in Figure 7. The quantitative comparison in Figure 7 leads to the conclusion that: Despite the higher difficulty, challenging picture scenario for the digital keying task, the recommended region strategy based on recurrent neural networks proposed in this chapter still performs comparably to the manually drawn three-part segmented map and the local optimal sequence.

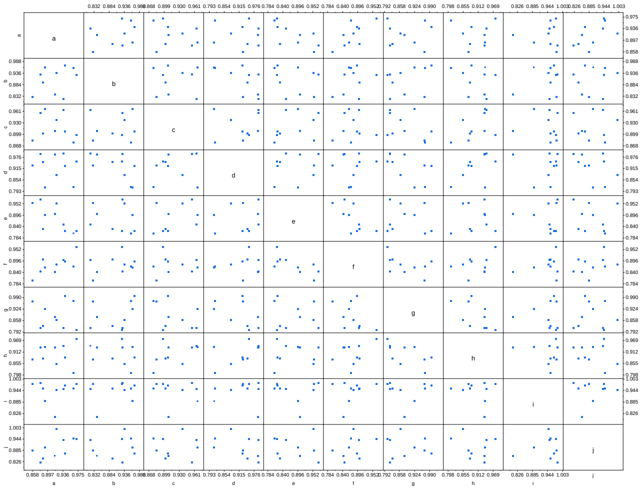

Figure 7: Prediction results obtained from the three-part segmentation map and the proposed model with different digital keying methods.

Also considering that drawing a higher quality three-part segmentation map is extremely time-consuming (at least 3 minutes) and obtaining the local optimal sequence requires repeated traversal of all possible regions (more than 1 hour) and reference to the true value of the digital keying target, the region recommendation method model based on recurrent neural networks proposed in this research work takes only about 1 second to recommend a window, completes 20 region recommendations. The total time required to complete 20 region recommendations and complete the digital keying task is about 1 minute. In summary, the model of the interactive keying method based on the recurrent neural network proposed in this chapter has better performance overall, as shown in Figure 8.

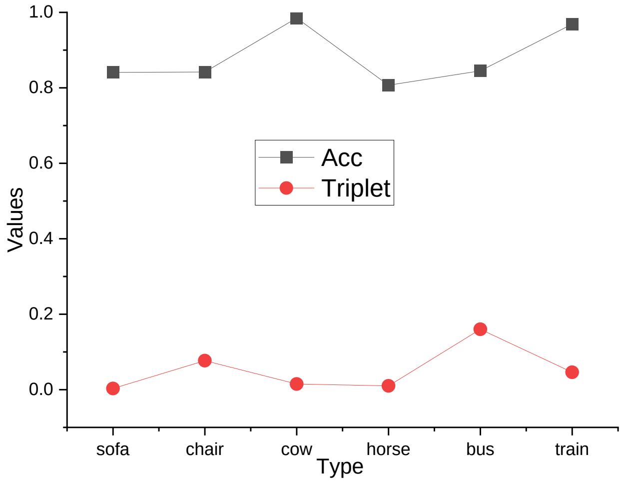

Figure 8: Loss function performance comparison.

Figure 8 shows the Triplet Loss comparison experimental data. This experimental condition is using void pyramid pooling + feature pyramid network + sample importance weights, and the experimental data is Pascal VOC2012, and the experimental results are shown by the average accuracy and 6-category segmentation accuracy in the table. It is hinted that the distance between classes of these 6 categories is small. From Figure 8, Triplet Loss improves the average precision and for all other 6 categories, which verifies the effectiveness of Triplet Loss as a loss function. In addition, Triplet Loss improves the model segmentation accuracy using both correlation and absolute value measures of sample feature distance, but as far as Figure 8 is concerned, the experimental results of Triplet Loss using correlation are slightly better than absolute value.

5 CONCLUSION

For the bottom-up mechanism, we propose a simple and effective computational model of visual attention mechanism to analyze the saliency information of images; visual words extracted on image sub-regions with higher saliency have higher weights; conversely, visual words extracted on regions with lower saliency have lower weights. For the top-down mechanism, the algorithm incorporates a supervised learning strategy based on the semantic category labels of the digital printmaking images, and the visual words are semantically weighted by counting the frequency of the visual words in each semantic category and then constructing a histogram of the frequency of category-related visual words. Accordingly, a digital image of printmaking art can be represented as a semantic visual word occurrence frequency histogram feature based on visual saliency weighting and category correlation. Finally, a support vector machine classifier is used to implement a semantic visual word packet model based on a digital image classification algorithm for printmaking art digital images on the constructed printmaking art digital image database. A digital image of printmaking art is decomposed into four parts: painting subject, inscription, white space, and seal. Then a series of unique color and texture features are extracted based on the visual and creative characteristics of each part. Finally, a multi-task joint sparse representation model is introduced to effectively fuse the features of the four parts and classify them. Experiments on a large set of digital images of printmaking art show that the proposed structural analysis algorithm in this paper can effectively perform structural decomposition of the painting images, and the performance of the classification strategy based on multi-task joint sparse representation outperforms that of the global-based classification method.

Jing Liu, https://orcid.org/0009-0004-1787-465X

ACKNOWLEDGEMENT

The research is supported by: Design professional practice teaching reform from the perspective of "integration of literature and engineering" (2022XJ016).

REFERENCES

[1] Andrikos, C.; Rassias, G.; Tsanakas, P.: et al. An enhanced device-transparent real-time teleconsultation environment for radiologists, IEEE Journal of Biomedical and Health Informatics, 23(1), 2018, 374-386. https://doi.org/10.1109/JBHI.2018.2824312

[2] Apollonio, F. I.; Ballabeni, M.; Bertacchi, S.: et al. Digital documentation and restoration tools reusing existing imagery, A multipurpose model of the Neptune's Fountain in Bologna, Applied Geomatics, 10(4), 2018, 295-316. https://doi.org/10.1007/s12518-018-0210-x

[3] Fu, Y.; Yu, H.; Yeh, C.K.:High relief from brush painting. IEEE transactions on visualization and computer graphics, 25(9), 2018, 2763-2776. https://doi.org/10.1109/TVCG.2018.2860004

[4] Isaac, A.M.C.: Digital images: Content and compositionality, Journal of the American Philosophical Association, 3(1), 2017, 106-126. https://doi.org/10.1017/apa.2017.13

[5] Kang, X.; Zhao, F.; Lin, G.: et al. A novel hybrid of DCT and SVD in DWT domain for robust and invisible blind image watermarking with optimal embedding strength, Multimedia Tools and Applications, 77(11), 2018, 13197-13224. https://doi.org/10.1007/s11042-017-4941-1

[6] Koyuncu, H.; Dixit, M.; Koyuncu, B.: An analysis of content-based image retrieval, International Advanced Researches and Engineering Journal, 5(1), 2021, 123-141. https://doi.org/10.35860/iarej.811927

[7] Kumar, M. P. P.; Poornima, B.; Nagendraswamy, H.S.; et al. A comprehensive survey on non-photorealistic rendering and benchmark developments for image abstraction and stylization, Iran Journal of Computer Science, 2(3), 2019,131-165. https://doi.org/10.1007/s42044-019-00034-1

[8] Li, Y.; Wang, D.; Tang, L.: Robust and secure image fingerprinting learned by neural network, IEEE Transactions on Circuits and Systems for Video Technology, 30(2), 2019, 362-375. https://doi.org/10.1109/TCSVT.2019.2890966

[9] Liu, A.; Zeng, X.; Yuan, Y.; et al. Joint artifact correction and super-resolution of image slicing and mapping system via a convolutional neural network, Optics Express, 29(5), 2021,7247-7260. https://doi.org/10.1364/OE.413076

[10] Ng, W.L; Chan, A.: Ong, Y.S.: et al. Deep learning for fabrication and maturation of 3D bioprinted tissues and organs, Virtual and Physical Prototyping, 15(3), 2020, 340-358. https://doi.org/10.1080/17452759.2020.1771741

[11] Nie, X.; Li, X.; Chai, Y.: et al. Robust image fingerprinting based on feature point relationship mining, IEEE Transactions on Information Forensics and Security, 13(6), 2018, 1509-1523. https://doi.org/10.1109/TIFS.2018.2790953

[12] Oldman, D.; Tanase, D.; Santschi, S.: The problem of distance in digital art history: A Research Space case study on sequencing Hokusai print impressions to form a human curated network of knowledge, International Journal for Digital Art History, (4), 2019,5.29-5.45.

[13] Panotopoulou, A.; Paris, S.; Whiting, E.: Watercolor woodblock printing with image analysis, Computer Graphics Forum, 37(2), 2018, 275-286. https://doi.org/10.1111/cgf.13360

[14] Schelkens, P.; Ebrahimi, T.; Gilles, A.: et al. JPEG Pleno: Providing representation interoperability for holographic applications and devices, ETRI Journal, 41(1), 2019, 93-108. https://doi.org/10.4218/etrij.2018-0509

[15] Shechtman, A.: Command of Media's Metaphors, Critical Inquiry, 47(4), 2021, 644-674. https://doi.org/10.1086/714512

[16] Tan, L.; Hu, K.; Zhou, X.; et al. Print-scan invariant text image watermarking for hardcopy document authentication, Multimedia Tools and Applications, 78(10), 2019, 13189-13211. https://doi.org/10.1007/s11042-018-5771-5

[17] Turco, M.L.; Giovannini, E.C.; Tomalini, A.: The Cultural Value of the Copy in the Museum Domain, Img Journal, (4), 2021, 156-179.

[18] Uccheddu, F.; Carfagni, M.; Governi, L.: et al. 3D printing of cardiac structures from medical images: an overview of methods and interactive tools, International Journal on Interactive Design and Manufacturing 12(2), 2018, 597-609. https://doi.org/10.1007/s12008-017-0415-y

[19] Walia, S.; Kumar, K.: Digital image forgery detection: a systematic scrutiny, Australian Journal of Forensic Sciences, 51(5): 2019, 488-526. https://doi.org/10.1080/00450618.2018.1424241

[20] Xu, C.; Han, Y.; Baciu, G.; et al.: Fabric image re-colorization based on intrinsic image decomposition, Textile Research Journal, 89(17), 2019, 3617-3631. https://doi.org/10.1177/0040517518817051