Research Article

Visual Memory Neural Network for Artistic Graphic Design

Yuchuan Zheng

School of Arts, Shanxi Normal University, Taiyuan, Shanxi 030031, China

Correspondence should be addressed to Yuchuan Zheng; 322092@sxnu.edu.cn

Received 6 June 2022; Revised 11 July 2022; Accepted 15 July 2022; Published 16 August 2022

Copyright © 2022 Yuchuan Zheng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Artistic graphic design is the aesthetic result of the designer’s fusion of various elements, with a high degree of independence. Considering the lack of significant visual design scope and aesthetic indicators of graphic design, our research aims to build an upgraded network model that can categorize different types of artistic graphics with labels and realize the free combination of graphic solutions. We realize the scheme reorganization of artistic graphic design from the perspective of computer vision and propose the artistic graphic design method based on memory neural network. We built a computer vision environment and reconstructed the computer vision network to set up an independent deep camera vision range calculation law. Considering the artistic graphic region segmentation problem, we propose the self-attentive mechanism, which can quantitatively segment different artistic graphic regions according to temporal features, before arranging them in a sequence to obtain the graphic region feature vector. We also add the LSTM structure based on the attention mechanism to match with the self-attention features of the graphical region segmentation module and pass the matched attention feature vector to the LSTM network to extract the labeled text feature information of the graphs. To test the effectiveness of our method, we build a database of artistic graphics and set up an adaptive training process. We also compared deep learning methods of the same type, and the experimental results proved that our method outperformed other deep methods in artistic graphic design by keeping the scheme reorganization accuracy and quantitative evaluation of artistic models above 90%.

1. Introduction

The most critical purpose of art graphic design is to solve the positioning solution and emergency plan of the product in the complete solution. During the establishment of the final solution of the product, different design processes and emergency measures need to be presented in visual communication in real time [1, 2]. Based on the artistic design needs of the client’s feedback, the appearance and approach are reconstructed in terms of appearance, methodological improvement, and style review on different levels such as screen, space, structure, and logic. In the field of visual design, textual language is active in another way in the coordination of solutions. Professional artists and design will communicate textual elements to the audience in visual form according to the client’s perception of the product, and this creative design style is the key point of figuration of artistic graphic design [3]. The representation of artistic graphics is shown in Figure 1.

The elements of art graphic design consist of both static and dynamic, and for all artistic design, graphic design applies to a range of disciplines such as painting, sculpture, and drawing. Graphic design as a basic discipline is widely used in various art industries and is also used in industry. Selecting independent design elements, building perceptual logic models, and constructing visual environments according to different design principles are all part of artistic graphic design. Considering the high professional requirements of artistic graphic design for manual labor, many researchers started to study automatic combination systems for artistic graphic design [4–7]. This research requires the integration of an artistic graphics database, computer vision unit, deep learning algorithm unit, data preprocessing unit, etc. Many researchers have already started their work accordingly.

Artistic graphic design solutions generate many unstable situational factors when it comes to human-environment communication scenarios. When dealing with image key
issues feature extraction, most researchers take the approach of fusion of different algorithms. Behavior recognition algorithms are mainly justified for personal graphic design, and image recognition algorithms are mainly utilized for environmental static buildings [8, 9]. Combining the two algorithms, dynamic features of the person and static features of the environment can be obtained separately, and matching the appropriate classifier can subdivide the person’s action features and map them with static features. Most researchers in artistic graphic designers prefer to use deep learning methods in the experimental phase [10–14]. Deep neural networks can capture different types of graphical features, and in artistic scene construction, different levels of combinations can be generated in the design scheme depending on the database coverage of the graphics. Each combination has independent network training parameters with generalizability. In the early visual communication research, researchers tried to construct 3D scenes using deep cameras, using deep learning algorithms to learn pixel elements in the scenes, and embedding the trained models into the artistic graphic design system, which can automatically match different scene combinations according to the customer’s needs. Customers can choose the corresponding graphic design solutions according to their needs. The application of deep learning accelerates the speed of scenario design and reduces the complexity of scenario design. In convolutional neural network weight screening, the training parameters and weights can be specified within a specified threshold according to the complexity of the original graphic data, and the model with the best graphic design is used in the test results.

The current character recognition algorithms cannot achieve nodal feature connection at the temporal level, resulting in missing longitudinal features between nodes at the time of acquisition. In the process of predicting behavior, artistic graphic design requires the dynamic coexistence of character features and environmental features, but the spatial information error is impossible to compensate for at the temporal level, which causes the problems of time consuming, inefficient, and low accuracy in image design solution generation [15–17]. To meet the special needs of different artistic graphic designs, some researchers have adopted 3D scanners for scene reconstruction and then used motion capture methods to reconstruct characters and environments on demand. Such an approach enables directed scene reconstruction for different graphic design projects, and researchers who adopt this method mostly use RGB images in the selection of behavior recognition data. Although this method has high timeliness and low experimental cost, it is the preferred method for most researchers. However, the method requires a certain experimental environment, and the graphical design scheme can be miscombined in the case of unstable nonstructural factors. To solve this problem, some researchers have used an approximate linear method to optimize the data minimization problem and extracted to generate predefined templates on the scene reconstruction combination problem to prevent combination errors that lead to system whiteout [18, 19].

Considering that the graphic design contains skeletal information of each node in the character model reconstruction, there is a directional problem between the normalization of skeletal information and vector information transfer, which reduces the generalization of the character graphic design. The extraction of temporal level behavioral features from skeletal information becomes exceptionally difficult under the dual effect of nonstructural factors. The depth camera is used as the only channel for depth information extraction, and the correlation between temporal convolution features and behavioral labels is obtained by matching the data video frame rate with the temporal convolution layer. In the matching of skeletal points with skeletal joints, some researchers try to create multiple spatial dimensions to match from different measures as a way to obtain the maximum match. The results of reasonable graphical design tests can be filtered according to the set range of criterion values [21–24]. When setting the graphic windows dynamically, all the dynamic windows are set to a uniform size to avoid the problem of inconsistent output features, and it is the deep neural network that can flexibly identify multiple skeleton sequences when receiving skeletal point feature data to increase the robustness of the model. We analyzed the visual effects and principles of composition of artistic graphic design. Considering the differences and independence of art graphic design, we realize the scheme reorganization of artistic graphic design from the perspective of computer vision and propose a memory neural network-based artistic graphic design method.

The rest of this study is organized as follows. Section 2 presents the history of research and research results on the intelligent design of artistic graphics. Section 3 describes in
detail the principles and implementation process related to visual memory neural network-based art graphics design. Section 4 shows the related experimental setup, the experimental data set, and the analysis of the experimental results. Finally, Section 5 summarizes our research and reveals some further research work.

2. Related Work

In the process of character graphic design, artists and technicians have different requirements for the effect of the dynamic presentation of characters, and there is a gap between the overall retention effect of graphics and the effect achieved by deep neural network models. To balance the requirements of both artistic design and technical processing, some researchers try to use RGB graphic data as the input of the scheme combination, and for the requirement of optical flow information in the scheme, graphic design artists require full traversal of still life appearance and dynamic character features. Some of the node-tracking information is often lost in the implementation process of technicians, and the behavioral linear resolution cannot be completed for dynamic character features. Therefore, some researchers have treated the above features separately, with the study in literature [25] oriented toward still life contour feature acquisition and the study in literature [26] oriented toward dynamic character skeletal point optical flow information feature capture. In addition, researchers in the literature [27] proposed a dense trajectory extraction algorithm for still life feature classification, which is based on the SVM algorithm and classified by feature association, and the efficiency of this classification method is experimentally proven to be excellent. Researchers in the literature [28] optimized the former study based on the addition of RGB cameras to capture optical flow features and optimize the bad trajectory data by matching the trajectory information co-generated by optical flow features with dense trajectories. Traditional RGB algorithms require manual labeling of behavior types and preprocessing management of data labels to prevent behavior label overlap. The biggest drawback of the RGB algorithm is that it relies too much on manual label classification design, and cannot freely combine solutions according to feature types in the graphical design combination scheme.

Traditional graphic design solution combination methods do not perform well at the level of accuracy and speed, and the graphic solution classification and frame rate processing effects cannot achieve real-time results. Some researchers have tried to optimize the visual graphic design solution combination using deep learning methods. Deep neural networks for graphic design can process the visual effects brought by graphic design solutions from the pixel level, and different visual presentation effects are easier to achieve at the pixel level according to the specific needs of the customer. Researchers in the literature [29] proposed a two-layer CNN algorithm when dealing with graphic design optical flow features, and the method can obtain deep pixel features between combinations of graphic schemes. Researchers in the literature [30] optimized based on dual-stream neural networks and proposed an independent frame feature fusion convolution algorithm, which was designed to achieve feature optimization and compensation in the deep convolution of graphic features in the first and second layers, and the experimental final output of the graphic design model was tested with better efficiency.

A convolutional neural network, as the most basic network structure for image recognition, is more relevant for video data with continuous frames. Compared with single-frame data processing, graphic contextual feature information is more easily linked. If a series of natural language processing neural networks are to be adopted for graphic sequences, they will be more closely related to the integration requirements of data sequences. Recurrent neural networks were the initial adaptive training model used in this study to complete the pretraining of primary combinatorial solutions for graphic design. As graphic design became more demanding at the artistic level, the researcher gradually focused his research on the complementarity of the strengths between different neural networks. In the literature [31], a recurrent neural network and a memory unit network were fused in dealing with the graphic design sequence problem, and an ordered graphic sequence integration was accomplished with the assistance of attention mechanisms. The literature [32] utilized computer vision to complete the directed graphical visual scene reconstruction in advance, and the method was also directly cited as a template by subsequent studies, which reduced the time cost and experimental cost and improved the efficiency of visual reconstruction. The literature [33] proposed a two-layer network nesting structure, where the authors selectively arranged some structures of RNN and LSTM algorithms in a spatial network in a flashback manner, and the graphical input was also RGB data. Researchers in the literature [34], on the other hand, proposed a new and improved LSTM approach, where the authors proposed the model for dense graphical target labeling and segmentation, where the input and output times are automatically ordered according to the matched image sequences in the design of the attention mechanism. To address the influence of unstructured factors, some researchers try to represent the graphical design in 3D space with RGB three channels as the width of feature extraction, and the method is experimentally proven to work better in the face of graphical design schemes with depth information. Some researchers extract interference information in the original graphic data and realize label mapping linkage through interference information and unstructured factors, which can directly filter out most unstructured factors during the model training process. The method works better in subsequent experimental demonstrations but is more dependent on manual processing of the raw data, which is more workload.

3. Method

3.1. Computer Vision Reconstruction. In graphic reconstruction design, the mapping relationship between pixel distances and realistic distances of graphic designs is usually determined from a two-dimensional graphic pixel
coordinate system. To ensure the accuracy, most researchers use the camera calibration method, through which the graphic elements under the pixel coordinate system can be converted to the world coordinate system by the camera calibration operation. The system of automatic combination of schemes for artistic graphic design is to realize automatic extraction of artistic graphic features, an automatic combination of schemes, etc. For this purpose, we designed a computer vision reconstruction mathematical equation as follows:

\[
x'_{ji} = f_1(x_{ji}),
\]

\[
y'_{ji} = f_2(y_{ji}),
\]

where \( f \) represents the mapping association between real coordinates and pixel coordinates, and the automatic transformation of the projection matrix can be done given the predetermined pixel coordinates.

According to the art graphic design requirements, the corresponding heights and nodes of different design patterns can be obtained according to the camera calibration. For irregular geometric art graphics, we can zone the graphics and subdivide them into rectangles, squares, triangles, and so on. Rectangles and squares are relatively easy to calculate and can be directly brought into the mathematical equation. For triangular clusters, trigonometric functions need to be added to the original equation to calculate the node distances and feature vector information of the art figure. The computer vision system is built as shown in Figure 2.

For a given graphic design target, dimensional mapping and pixel coordinate positioning can be done within the first line of action. In a real scene, the dynamic movement distance can be accurately measured in the projection and camera angle as long as the range of the graphic design target is within the range limit. The area between the first action line and the second action line belongs to the maximum distance where the depth information of the depth camera acts, and within this area, it can be used for 3D graphic design, and the spatial information of each graphic node can be accurately recorded in the detailed vector information due. Those beyond the second action line belong to the invalid area. According to the calculation equation of the triangle function, we can get the information of camera angle, action line angle, action range angle, and the angle of nodes inside the graphical target. The mathematical equation of the clip angle calculation is shown below.

\[
\alpha = \arccos \left(\frac{d_{ij}^2 + d_{ik}^2 - d_{jk}^2}{2d_{ij} \times d_{ik}}\right) - \arcsin \left(\frac{d_{jd} \times \sin \angle a_{jk}}{d_{ad}}\right)
\]

\[
- \arcsin \left(\frac{d_{jk} \times \sin \angle a_{jk}}{d_{ac}}\right).
\]

3.2. Self-Attentive Based Neural Network for Artistic Graphic Design. We refer to many literature on the construction of neural networks for graph design, and according to our need for graph design solutions with only combinations, we use the partial structure of convolutional neural networks and temporal convolutional networks. In the temporal convolutional network part, we divide the graphics into different regions, all of which will be arranged in a certain order, assuming that the input sequence is \(X = (x_1, x_2, \ldots, x_n)\), \(n\) represents the number of graphic design time regions division, and \(x_i\) represents the spatial dimension of the region in the ith feature data in the feature acquisition stage of the art graphics, which is represented as a projection matrix with dimensions \(T \times M \times D\), where \(D\) represents the temporal label of the region, and \(M\) represents the number of graphic nodes. In the structural design of the temporal convolution layer, we use a \(3 \times 3 \times t\) filter as the initial feature extraction, set the pooling layer to \(3 \times 3 \times 1\), and the rest of the initial convolution layer dimensions are set to \(3 \times 3 \times 3\). The last graphical temporal convolution result will come with the temporal depth information of all the previous convolution layers, and according to the temporal depth range obtained from the test, the best temporal convolution result can be retained for each feature [35].

We segment the art graph into regions and then combine the node vector features of each region into different sequences, each sequence will also be subjected to pooling operation and batch normalization after the initial convolution calculation. Finally, \(n\) feature vectors of different time regions of the art graphics at the time level will be obtained. Each set of sequences represents a combination of art graphics in an independent period. The scale of each graphical feature is \(X \in \mathbb{R}^{N \times M \times K}\), where \(K\) represents the feature dimension of each time region node. On top of this, we add an attention mechanism to weight the weighted features of each art-graphic region. To migrate the self-attentive weighted features to the original art graphics, we reshape the region nodes in the attention network. Assuming that the expression of the region node X after reshaping is \(Y = [y_1, y_2, \ldots, y_m]\), where \(y_i \in \mathbb{R}^{NK}\), by forwarding propagation network we define the self-attention weights calculation \(a_i\) as the following equation:

\[
a_i = \sigma (W^x \times y_i + b_i^x), \quad i = 1, 2, \ldots, m,
\]

where \(W^x\) and \(b_i^x\) denote the self-attentive parameters with matrix bias and \(\sigma(\ast)\) represents the activation function. We denote the mathematical equations of the self-attentive

**Figure 2: Computer vision system construction.**
model as \( s_{\text{att}} = [s_1, s_2, \ldots, s_m] \) in the artistic graph input matching \( X \in \mathbb{R}^{N \times M \times K} \). To ensure the generalization of the input graphics in, we denote the attention graph as \( A \in \mathbb{R}^{N \times M \times K} \), where \( K \) represents the artistic graphics features after \( K \) feature extractions. Finally, we weight all the previously obtained features for conversion, and the mathematical equation is as follows:

\[
\begin{align*}
\mathbf{f}_n &= A \cdot \mathbf{X}, \\
\mathbf{f}_n^{(1)} &= \text{FNN}\left(\text{MultiHead}\left(\mathbf{f}_n^{(0)}, \mathbf{f}_n^{(0)}, \mathbf{f}_n^{(0)}\right)\right),
\end{align*}
\]

where \( g \) denotes the extraction method of LSTM network in artistic graphic region features. \( Q \) represents the combination of temporal features in different dimensional outputs. If the feature outputs in the same dimension are used as local feature sequence inputs, the forgetting gate will not be able to filter useful features due to the existence of blank features between periods. \( \alpha \) denotes the attention of \( n \) features in the predicted output of the memory network unit. \( r \) denotes the weighted sum of attention of all time dimensions. To solve the problem of the inconsistent combination of multiple artistic graphics schemes and uncoordinated input dimensions, we also propose a graphics feature compression method based on variable pooling operations. This method can compress the attention of different dimensions according to uniform dimensions and decompress the uniform features at the graphical design combination network layer, and each network layer has a built-in data preprocessing layer to avoid the problem of confusing data formats. The variable pooling feature compression network is shown in Figure 4.

3.3. Attention-Based LSTM. To facilitate the traversal and association of temporal convolutional features, we used the LSTM algorithm as an association network, and to the LSTM, we added an attention mechanism to associate with the self-attention layer in the temporal convolutional network. Our proposed visual memory neural-based art graphics design network is shown in Figure 3.

In the above figure, \( \gamma \) represents the comprehensive fusion output of the attention feature vector, and \( \rho \) represents the average weight fusion output of the attention feature vector. During the reorganization of the art graphic design scheme, we assigned a different focus of attention mechanism to each graphic area. In the face of different art styles, each art style was evaluated under the artist’s assessment with multiple indicator scores according to the differences in expression. For example, the difference between color art and rule art, color art mainly emphasizes RGB pixel intensity and establishes the assignment point of color art attention mechanism based on the local pixel threshold. The contour intensity is mainly used as the threshold response for rule art, and the attention mechanism assigns weights based on the response threshold. In each period, the improved LSTM network can extract feature vectors of the same dimension from the same attention mechanisms. In this way, different dimensional features are extracted in batches, and finally, the global graphic combination scheme features are obtained in the final class.

Suppose the art graphics data has \( N \) feature sets \( f = [f_1, f_2, \ldots, f_n] \), each of which is obtained in the dual association role of temporal convolutional network and long short-term memory network. When the feature sets are used as input, the inverse network of the temporal convolutional layer receives only one fused feature graph at a time, and the first half of the network can output a high-level feature graph at any time under the action of the LSTM network, and the outputs of the different dimensional attention mechanisms are calculated as shown below.

\[
\begin{align*}
h_i &= g(f_i), i \in (1, n), \\
Q &= [h_1, h_2, \ldots, h_n], \\
\alpha &= \text{softmax}(w^T \tanh(Q)), \\
r &= Q\alpha^T,
\end{align*}
\]

4. Experiment

4.1. Training. Artistic graphic scheme free combination model for scheme requires huge graphic database support. For the creation of the database, the database classification can perceive the differences in art graphics in drawing style, color, and contour, to extract the different features of each type in the deep neural network. We map the depth information in the depth camera to the pixel coordinates, to obtain the representation of the contour nodes in space according to the graphic contour extraction algorithm. The pixel coordinates and depth information will produce information mapping in two different data sequences, and to check whether the graph and combination classification are unified, we set independent region thresholds. It has an excellent scheme combination effect for still life graphics. In the model training process, we plan a reasonable training process according to the demand for artistic graphics combinations, as shown in Figure 5.

4.2. Data set. At the initial stage of the art graphics database creation, we invited professional art aestheticians to evaluate art graphics for aesthetic indicators, object emphasis, color harmony, balance elements, motion blur, and other indicators and developed an independent evaluation system. In the art graphics design data collection, we manually labeled the art graphics that have generated economic benefits, and the labeled information contains graphic size, artistic category, combination direction, etc. For the database sequence classification, we adopted the rule of thirds and experimentally reconstructed a quantitative evaluation model. In the quantitative evaluation model, the art graphics’ ease of use, aesthetics, balance, and contrast remain the performance evaluation benchmarks for the automatic art graphics combination model. In all art graphics data sets, we set 80% of the training set and 20% of the test set. In the experimental implementation of the art graphics data set, we defined only
three categories of artistic graphics at the beginning of the study, namely color category, regular combination category, and multiple arrangement category. The details of the art graphics data set are shown in Table 1.

4.3. Analysis. In the prior experiments, we found in the method validation experiments that machine learning methods have poor accuracy in the visual communication of artistic graphics and cannot achieve real-time requirements. Deep learning methods perform better in the scheme combination of artistic graphics; therefore, in the later experiments, we use deep learning methods as the base reference. In the first stage of experiments, we compare the effect of three methods, CNN [36], RNN [37], and LSTM [38], on the combination of artistic graphics. The first phase experiments are evaluated in terms of precision ($P$) and recall ($R$). The experimental results are shown in Table 2.

According to the experimental data in Table 2, it is clear that the combination of artistic graphic design solutions performs better in the image-based deep learning approach. However, the image-based deep neural network method cannot access the information contained inside the artistic graphic labels and cannot achieve the fusion and generalization of data features at the textual level. The accuracy of CNN-like methods stays around 80% in the three types of artistic graphic scheme combinations. Our method incorporates not only CNN methods but also LSTM methods. The structure of the dual network makes up for the model’s feature capture of textual information of art graphics, and the joint mapping of pixel features and textual features can effectively compensate for the shortcomings of the pure image-like methods. Therefore, the experiment proves that our method can keep the accuracy above 90% in all the art graphics combinations, which is significantly better than other methods.

In the second phase of the experiment, we validated the artistic graphic aesthetic index. Based on the opinions of professional art aestheticians, we chose three important indicators for validation: Balance Element (BE), Color Harmony (CH), and Object Emphasis (OE). The balance element is to verify the visual balance of art graphics after reconstruction. Color Harmony is to verify that the artistic graphic features are fully captured at the pixel level. Object Emphasis is to highlight the art graphics in similar combination schemes. The results of the experiments are shown in Table 3.
The experimental data in Table 3 shows that the best performing type of art graphics among the balanced elements is the multiple arrangement art graphics, which is more advantageous when evaluating the balanced elements because of the multiple arrangement graphics with different combinations of elements. In the color harmony, the best effect is in the color art graphics, and the color harmony index is not good enough because the regular combination and multiple arrangements pay more attention to the outline and space planning. Target emphasis performs well in all
In this study, we analyze the visual effects and principles of composition of artistic graphic design. Considering the differences and independence of artistic graphic design, we realize the scheme reorganization of art graphic design from the perspective of computer vision and propose the method of art graphic design based on memory neural network. Referring to numerous deep learning methods such as CNN, RNN, and LSTM, we experimentally validated each method and finally designed a two-layer network structure based on CNN and LSTM networks. We built a computer vision environment and reconstructed the computer vision network to set up an independent deep camera vision range computation law. Considering the artistic graphic region segmentation problem, we proposed a self-attentive mechanism, which can quantitatively segment different artistic graphic regions based on temporal features, after arranging them in a sequence to obtain the graphic region feature vector. In the last part of the network structure, we propose the LSTM structure based on the attention mechanism to match with the self-attention features of the graphic region segmentation module and pass the matched attention feature vector to the LSTM network to extract the labeled text feature information of the graphics. To test the effectiveness of our method, we build a database of artistic graphics and set up an adaptive training process. We also compare deep learning methods of the same type, and the experimental results demonstrate that our method outperforms other deep methods in artistic graphic design in terms of scheme reorganization accuracy and quantitative evaluation of artistic models.

Intelligent design of art graphics is a very complex study, where visual scene construction and graphic area segmentation are key point technologies. In the research of this paper, we only choose three simple art categories as research, in fact, there are more complex art categories. In future research, we will try more art types and use the two-layer LSTM algorithm to enhance the feature capture range of the model and improve the generalization of the model.

### Data Availability

The data set can be accessed upon request.

### Conflicts of Interest

The author declares that there are no conflicts of interest.

### Acknowledgments

The author thanks the General Project of Philosophy and Social Science Research in Colleges and Universities of Shanxi Provincial Education Department (no. 2013228).

### References

[1] G. Kress and T. Van Leeuwen, *Reading Images: The Grammar of Visual design*, Routledge, vol. 33, no. 6, Oxfordshire, UK, 2020.

[2] R. Hirsch, *Light and Lens: Photography in the Digital age*, Routledge, Oxfordshire, England, UK, 2012.

[3] B. Peterson, *Using Design Basics to Get Creative results*, Adams Media, Avon, MA USA, 1996.

[4] S. Bell, *Elements of Visual Design in the landscape*, Routledge, Oxfordshire, England, UK, 2019.

[5] J. Lin, “Development of scales for the measurement of principles of design,” *International Journal of Human-Computer Studies*, vol. 71, no. 12, pp. 1112–1123, 2013.

[6] W. Rawat and Z. Wang, “Deep convolutional neural networks for image classification: a comprehensive review,” *Neural Computation*, vol. 29, no. 9, pp. 2352–2449, 2017.

[7] A. Carballal, A. Santos, J. Romero, P. Machado, J. Correia, and L. Castro, “Distinguishing paintings from photographs by complexity estimates,” *Neural Computing & Applications*, vol. 30, no. 6, pp. 1957–1969, 2018.

[8] R. G. Condorovici, R. Vranceanu, and C. Vertan, “Saliency map retrieval for artistic paintings inspired from human understanding[[]],” *Proceedings of the SPAMEC*, 2011.

[9] C. H. Bock, J. G. A. Barbedo, and E. Del Ponte, “From visual estimates to fully automated sensor-based measurements of plant disease severity: status and challenges for improving accuracy,” *Phytopathology Research*, vol. 2, no. 1, pp. 1–30, 2020.

[10] P. Machado, J. Romero, M. Nadal, A. Santos, J. Correia, and A. Carballal, “Computerized measures of visual complexity,” *Acta Psychologica*, vol. 160, pp. 43–57, 2015.

### Table 4: Experimental results of different methods of quantitative evaluation model.

|                           | CNN | RNN | LSTM | Ours |
|---------------------------|-----|-----|------|------|
| Graphics design ease of use| 0.7 | 0.7 | 0.8  | 0.9  |
| Graphic combination balance| 0.7 | 0.8 | 0.8  | 0.9  |
| Graphic category aesthetics| 0.8 | 0.7 | 0.8  | 0.9  |


[11] X. Guo, Y. Qian, L. Li, and A. Asano, “Assessment model for perceived visual complexity of painting images,” Knowledge-Based Systems, vol. 159, pp. 110–119, 2018.

[12] E. Cetinic, T. Lipic, and S. Grgic, “Learning the principles of art history with convolutional neural networks,” Pattern Recognition Letters, vol. 129, pp. 56–62, 2020.

[13] L. Liu, E. A. Silva, C. Wu, and H. Wang, “A machine learning-based method for the large-scale evaluation of the qualities of the urban environment,” Computers, Environment and Urban Systems, vol. 65, pp. 113–125, 2017.

[14] Y. Wu, W. AbdAlmageed, and P. Natarajan, “Mantra-net: manipulation tracing network for detection and localization of image forgeries with anomalous features,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9543–9552, 2019.

[15] A. Tomé, M. Kuipers, T. Pinheiro, M. Nunes, and T. Heitor, “Space-use analysis through computer vision,” Automation in Construction, vol. 57, pp. 80–97, 2015.

[16] W. Huang and H. Zheng, “Architectural Drawings Recognition and Generation through Machine learning[1],” in Proceedings of the 38th Annual Conference of the Association for Computer Aided Design in Architecture (ACADIA), ACADIA, Mexico City, 2018.

[17] D. Karimi, H. Dou, and S. K. Warfield, “Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis,” Medical Image Analysis, vol. 65, 2020.

[18] D. Moussazadeh and A. Ayтуг, “The concept of the aesthetic features in architectural structures of the museums[1],” International Journal of Architectural and Environmental Engineering, vol. 12, no. 12, pp. 1142–1150, 2018.

[19] M. Balzani, F. Maietti, and B. M. Mugayar Kühl, “Point cloud analysis for conservation and enhancement of modernist architecture,” The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 1XLIII-2/W3, pp. 71–77, 2017.

[20] C. Belém, L. Santos, and A. Leitão, “On the Impact of Machine Learning: Architecture without Architects? in CAAD Futures, On the Impact of Machine Learning in Architecture, 2019.

[21] S. Stabinger and A. Rodriguez-Sanchez, “Evaluation of deep learning on an abstract image classification dataset,” in Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 2767–2772, IEEE, 2017.

[22] Y. C. Tang, J. I. Huang, M. T. Yao et al., “A review of design intelligence: progress, problems, and challenges,” Frontiers of Information Technology & Electronic Engineering, vol. 20, no. 12, pp. 1595–1617, 2019.

[23] Y. K. Yi, Y. Zhang, and J. Myung, “House style recognition using deep convolutional neural network,” Automation in Construction, vol. 118, Article ID 103307, 2020.

[24] L. Lettry, M. Perdoch, and K. Vanhoey, “Repeated Pattern Detection Using CNN activations,” in Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 47–55, IEEE, Santa Rosa, CA, USA, 2017.

[25] C. Yan, Z. Li, and Y. Zhang, “Depth image denoising using nuclear norm and learning graph model, ACM Transactions on Multimedia Computing,” Communications, and Applications (TOMM), vol. 16, no. 4, pp. 1–17, 2020.

[26] M. Li, J. Lv, and C. Tang, “Aesthetic assessment of paintings based on visual balance,” IET Image Processing, vol. 13, no. 14, pp. 2821–2828, 2019.

[27] H. Wang, A. Kläser, C. Schmid, and C. L. Liu, “Dense trajectories and motion boundary descriptors for action recognition,” International Journal of Computer Vision, vol. 103, no. 1, pp. 60–79, 2013.

[28] F. Perronnin and C. Dance, “Fisher kernels on visual vocabularies for image categorization[1],” in Proceedings of the 2007 IEEE conference on computer vision and pattern recognition, pp. 1–8, IEEE, Minneapolis, MN, USA, 2007.

[29] K. Simonyan and A. Zisserman, “Two-stream convolutional networks for action recognition in videos[1],” Advances in Neural Information Processing Systems, vol. 27, 2014.

[30] K. Thömmes and R. Hübner, “Instagram likes for architectural photos can Be predicted by quantitative balance measures and curvature,” Frontiers in Psychology, vol. 9, Article ID 1050, 2018.

[31] C. C. S. Liem and A. Panichella, “Oracle issues in machine learning and where to find them,” in Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops, pp. 483–488, IEEE, New York, NY, USA, 2020.

[32] N. Murray, L. Marchesotti, and F. Perronnin, “AVA: a large-scale database for aesthetic visual analysis,” in Proceedings of the 2012 IEEE conference on computer vision and pattern recognition, pp. 2408–2415, IEEE, New York, NY, USA, 2012.

[33] A. Lecoutre, B. Negrevergne, and F. Yger, “Recognizing Art Style Automatically in Painting with Deep learning[1],” in Proceedings of the Asian Conference on Machine Learning, pp. 327–342, PMLR, 2017.

[34] J. Llamas, P. M. Lerones, R. Medina, E. Zalama, and J. Gómez-Garcia-Bermejo, “Classification of architectural heritage images using deep learning techniques,” Applied Sciences, vol. 7, no. 10, p. 992, 2017.

[35] K. He, X. Zhang, and S. Ren, “Deep Residual Learning for Image recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, IEEE, Las Vegas, NV, USA, 2016.

[36] J. Gu, Z. Wang, J. Kuen et al., “Recent advances in convolutional neural networks,” Pattern Recognition, vol. 77, pp. 354–377, 2018.

[37] W. Zaremba, I. Sutskever, and O. Vinyals, “Recurrent neural network regularization,” arXiv preprint arXiv, vol. 1409, p. 2329, 2014.

[38] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.