Affective and Pleasurable Design, Vol. 71, 2023, 241–251
https://doi.org/10.54941/ahfe1003264

Painting Style Alignment: Restoration of Ancient Chinese Landscape Paintings Driven by Aesthetic Cognition and Aesthetic Computation

Rong Chang[1] and Jingran Wang[2]

[1]Beijing Institute of Graphic Communication, Beijing, 102600, China
[2]Beijing International Studies University, Beijing, 100024, China

ABSTRACT

Painting style alignment is critical for ancient painting restoration. To reduce the aesthetic cognition conflicts between human and artificial intelligence in Chinese landscape painting restoration, we propose a new research paradigm, joint aesthetic cognition, which aligns painting style in three stages: descriptive painting style based on human aesthetic cognition, predictive painting style with AI aesthetic computation, and prescriptive restoration test via human-AI collaboration. To keep interaction of these stages continuous, a hybrid research method based on both data design and self-supervised learning is further proposed to adaptively construct the joint cognition of specific painting style. Preliminary tests on two restorative scenarios, analogized generation and cognitive recompositing, indicate that our joint aesthetic cognition is effective and feasible for painting style alignment, and can thus facilitate human-AI collaboration in ancient painting restoration.

Keywords: Painting style alignment, Ancient painting restoration, Chinese landscape painting, Human-AI collaboration, Joint aesthetic cognition

INTRODUCTION

AI assisted ancient painting restoration requires accurate painting style alignment. During the constant interaction of human and AI, instructions expressed in language may generate cognition conflicts, resulting in unstable output. Therefore, there is a very real need for a research framework, that enables accurate and stable output by integrating the knowledge of restoration experts and the computing power of artificial intelligence.

We propose a joint aesthetic cognition paradigm to interpret, learn, predict, and prescribe painting style under certain circumstances in given an ancient Chinese landscaping painting. The joint aesthetic cognition research framework is composed of two peer systems- a aesthetic cognition system and an aesthetic computation system (see Fig. 1). In the aesthetic cognition system, human experts extract style descriptions from the specific painting. In the aesthetic computation system, the AI computationally predicts the painting style. These two systems interact with each other by integrating expert's

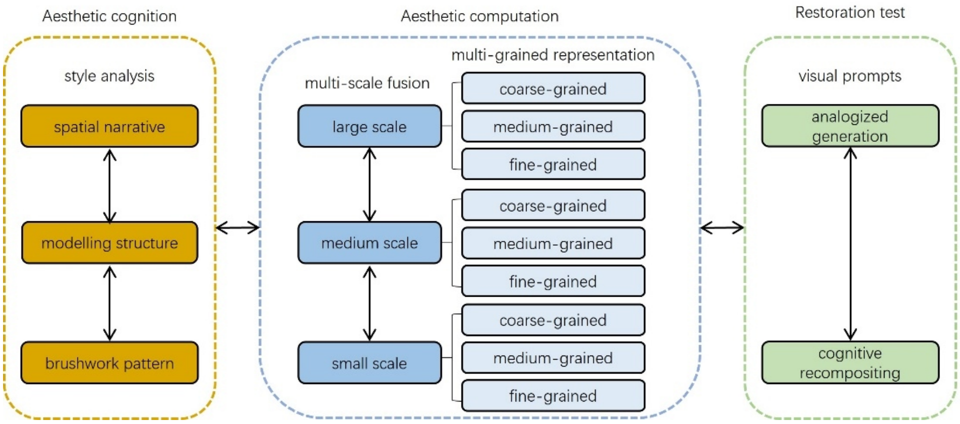


Figure 1: The joint aesthetic cognition framework bidirectionally driven by aesthetic cognition and aesthetic computation.

domain knowledge into data design, and then AI models learn how to colorize as a painter from a data flow to train and adapt its model. By recombining the multi-scale and multi-grained cognition, the model will predict painting style in computational experiments. Finally, guided by the expert's visual prompts, the model will provide restoration solutions in the expected painting style.

This research empowers experts in the field of ancient painting restoration from the following three aspects: (1) a series data design methods is proposed to help the experts generate high-quality datasets from rare ancient paintings easily; (2) the impacts of data granularity, scale, and mask are discussed qualitatively and quantitatively, that can be adopted by experts as suggestions to improve style alignment by selecting and updating sub-datasets; and (3) a restoration model is trained, which may assist experts to adjust and evaluate restoration solutions intuitively.

RELATED WORK

In the field of ancient Chinese painting restoration, artificial intelligence technology has not been widely used. On the one hand, the restoration of ancient paintings requires profound professional knowledge; on the other hand, most restoration experts lack the resources and knowledge to modify the AI model for a specific painting style. A more targeted approach is to empower restoration experts with AI through the construction of joint cognition paradigm. This section introduces related research focusing on data design, self-supervised learning, and human-AI interaction.

The Generation of Chinese Painting Style

Inspired by the deep learning, researchers have begun to use convolutional models to simulate Chinese painting style, such as transfer photos to Chinese paintings based on CNN (Sheng et al., 2019) or generated Chinese landscape paintings using GAN or its variants (He et al., 2018; Xue, 2021). The diffusion model (Nichol, 2021) and its related applications (Rombach et al., 2022) have boosted the popularity of AI painting. However, the current painting models are not ideal for stable style output. This may be due to the limitations of language-supervised learning. After all, it is difficult for people to express irregular shapes, movements, complex structures and subtle strokers in words (Van Den Oord, Kalchbrenner and Kavukcuoglu, 2016). Despite the learning mechanism, one fundamental issue is that painting models often trained on the data collected from the Internet (Kreutzer et al., 2022), which is insufficient in quality and mixed in style. The previous study has revealed that the models trained on diverse datasets are not always better than those trained on specialized datasets (Cole et al., 2022). Therefore, we need to construct the hybrid method based on both data design and self-supervised learning.

Date Design

To simulate the general drawing steps of human, we adopted the data form of contour-stroker pairs (Isola et al., 2017). In addition to the data form, we also considered the granularity and scale. Human cognition is a process from large-scale and coarse-granularity to small-scale and fine-granularity; AI computing is a process from small-scale and fine-granularity to large-scale and coarse-granularity. In order to achieve the interaction between the two, we designed the datasets as a structure of multi-granularity (Liu et al., 2022), multi-scale (Ronneberger, Fischer and Brox, 2015), and performed mask augmentation (He et al., 2022). The existing studies have emphasized the impact of sub-dataset on model performance (Ghorbani and Zou, 2019; Hohman et al., 2020). Such a data design is also convenient for the experts to select and update sub-datasets.

Human-AI Interaction

Some researchers suggested that end users should be widely involved in all aspects of artificial intelligence research and development, such as data, frameworks, models, and applications (Lee et al., 2019). The human-in-the-loop approach can integrate tacit knowledge into computing. As a result, the problem that aesthetic cognition is difficult to be encoded in language would be alleviated to a certain extent. Although the study on human-AI interaction is still in its infancy (Cooper et al., 2014; Dove et al., 2017; Kurdziolek, 2022), it has been confirmed that, comparing to relying solely on data, the feedback of human can improve learning efficiency (Ouyang et al., 2020) and enhance the value alignment of human and AI (Gabriel, 2020).

Based on the previously mentioned research and technical support, we built a joint cognition paradigm consisting of three stages: descriptive painting style based on human aesthetic cognition, predictive painting style with AI aesthetic computation, and prescriptive restoration test via human-AI collaboration. We will elaborate on these three stages in the subsequent sections.

DESCRIPTIVE PAINTING STYLE AND DATA DESIGN

The first step in building the joint aesthetic cognition is to build a dataset that characterizes the basic attributes of the target style. This section will analyse the painting style of “Dwelling in the Fuchun Mountains” based on human aesthetic cognition and introduce the data design method.

Multi-Scale Fusion

During the Yuan Dynasty (1271-1368 A.D.), the style of Chinese landscape painting changed from depiction to super-expression. “Dwelling in the Fuchun Mountains”, an authentic work of Huang Gongwang, is representative of this new style. The spatial narrative of “Dwelling in the Fuchun Mountains” is a mixture of lofty-distance, flat-distance, and far-distance, which challenges AI by making it face both macro and micro at the same time. As we know, the AI network constructed by up-sampling and down-sampling highlights the dominant objects at the cost of ignoring small objects. In order to enable small objects to be clearly captured and fairly calculated, data design needs to support multi-scale fusion.

Based on the high-definition digital photo of “Dwelling in the Fuchun Mountains”, we preformed random sampling at three scales. We adjusted the vertical axis of the digital photo to 512 pixels, 1024 pixels and 2048 pixels, then we randomly segmented them with a window of 512*512 pixels (see Fig. 2). In this way, we obtained large-scale samples covering 100% of the vertical axis, medium-scale samples covering 50% of the vertical axis, and small-scale samples covering 25% of the vertical axis.

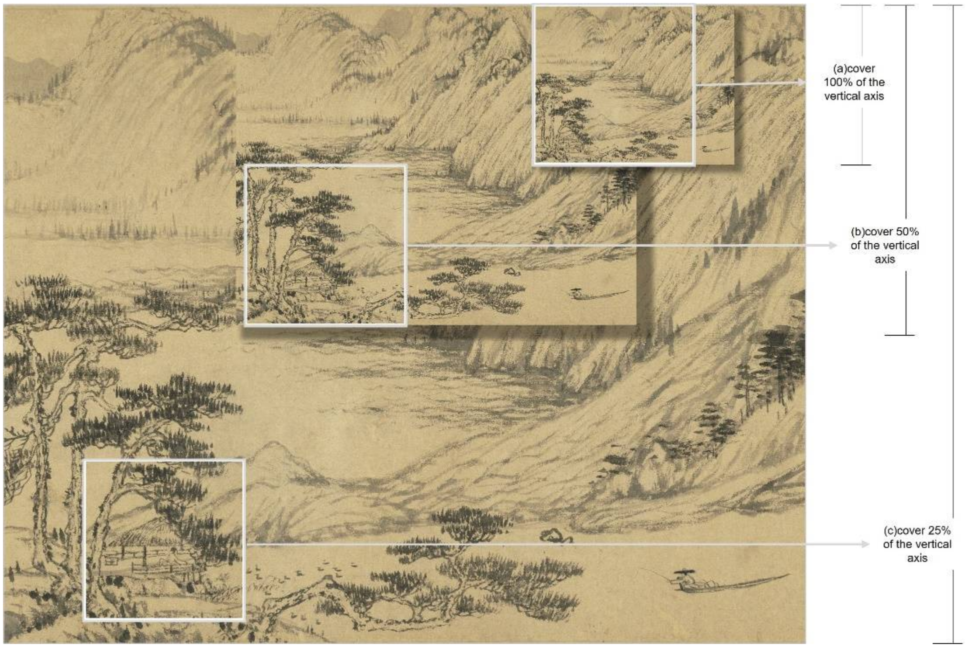


Figure 2: The segmentation at multi-scale: (a) large-scale sample, (b) medium-scale sample, and (c) small-scale sample. (The HD digital photo of “Dwelling in the Fuchun Mountains” is adapted from Taipei Palace Museum.)

Multi-Grained Representation

Huang Gongwang has developed a unique modelling structure in “Dwelling in the Fuchun Mountains” that can be called “lines in lines”. The abstract lines in the painting are roughly parallel, which make it difficult for AI to distinguish the external contours from the internal strokes. In addition, Huang’s brushwork pattern is vivid and of no apparent regularity. Our data design should serve not only the capture of main contours, but also the preservation of details. As shown in Fig. 3, we formed a multi-grained representation in the data design: coarse-grained representation of the main contours (factor $= 1$); medium-grained representation of intermediate contours (factor $= 0.5$); fine-grained representation of brushwork details (factor $= 0.1$).

Data Augmentation

When watching “Dwelling in the Fuchun Mountains”, the viewer’s eyes tend to focus on the detailed parts. We added irregularly distributed masks on the aesthetic focus areas (see Fig. 4). These masked data can drive AI to learn better in context and apply learned style to new compositions by analogy.

Based on the above methods, we formed a painting style dataset containing 44,132 contour-stroker pairs (see Table 1).

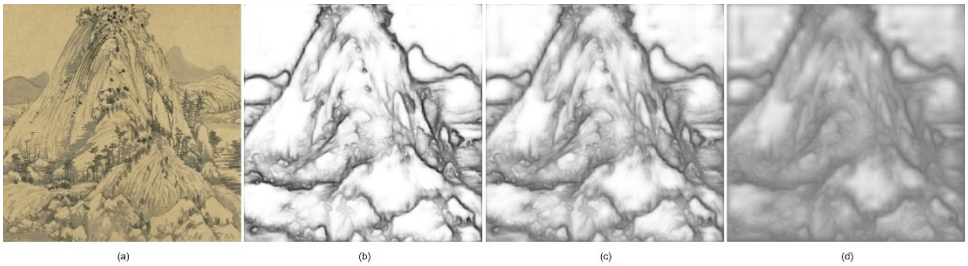


Figure 3: The multi-granularity representation: (a) original image, (b) coarse-grained representation, (c) medium-grained representation, (d) fine-grained representation. (The HD digital photo of “Dwelling in the Fuchun Mountains” is adapted from Taipei Palace Museum.)

Figure 4: Examples of manual masked data. (The HD digital photo of “Dwelling in the Fuchun Mountains” is adapted from Taipei Palace Museum.)

Table 1. The composition of painting style datasets.

	Coarse grained	Medium grained	Fine grained
Large scale	6600 (Including 2000 masked samples)	4000	4000
Medium scale	4754	4454	4454
Small scale	5370	5250	5250

PREDICTIVE PAINTING STYLE AND COMPUTATION

The second step in building a joint aesthetic cognition is to train the model to predict painting style through computational experiments.

Computational Framework

The parts of our computational framework are shown in Fig. 5. The framework is composed of a discriminator and a generator (Isola et al., 2017). It has the architecture to fuse multi-scale features through layer-skip connections (Ronneberger, Fischer, and Brox, 2015). To reduce the hardware requirements, we minimized the dimensions and the hidden layers of the architecture.

Experiment Operation

We designed a series of experiments according to different alignment goals. As listed in Table 2, the updates of sub-datasets took place in every experiment. Sub-dataset selection and update can be viewed as an interaction problem. The optimal interaction pattern will determine the signals that human sent to AI model.

The training parameters were consistent from E1 to E5. The parameters of E6 were optimized: the buffer_size was expanded from 800 to 1600, which increased the randomness of the data used in each iteration; the generator optimizer rate was in-creased from 2e-1 to 4e-1 to further expand the opportunities for analogized generation.

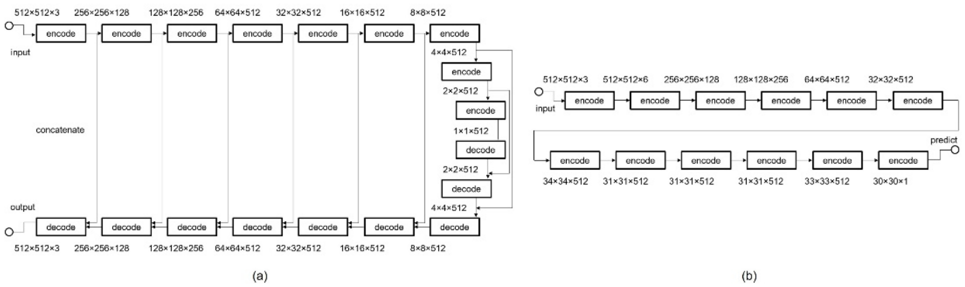


Figure 5: Architecture of the computational framework: (a) discriminator and (b) generator.

Table 2. Sub-datasets updates in different experiment.

Experiment	Goal	Large-scale	Med-scale	Small-scale	Coarse-grained	Med-grained	Fine-grained	Masked
E1	Outline	✓			✓			
E2	Strokers	✓			✓	✓	✓	
E3	Spatial narrative	✓	✓	✓	✓			
E4	Modelling structure	✓	✓	✓	✓	✓	✓	
E5	Analogy	✓	✓	✓	✓	✓	✓	✓
E6	Analogy	✓	✓	✓	✓	✓	✓	✓

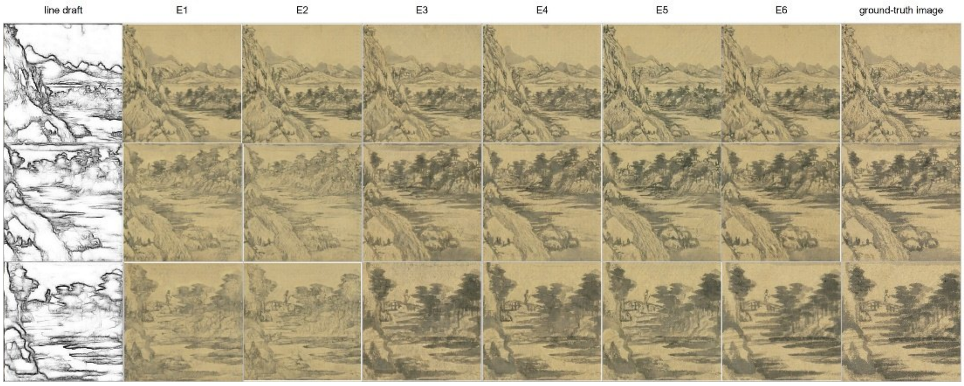


Figure 6: Comparison of experimental results.

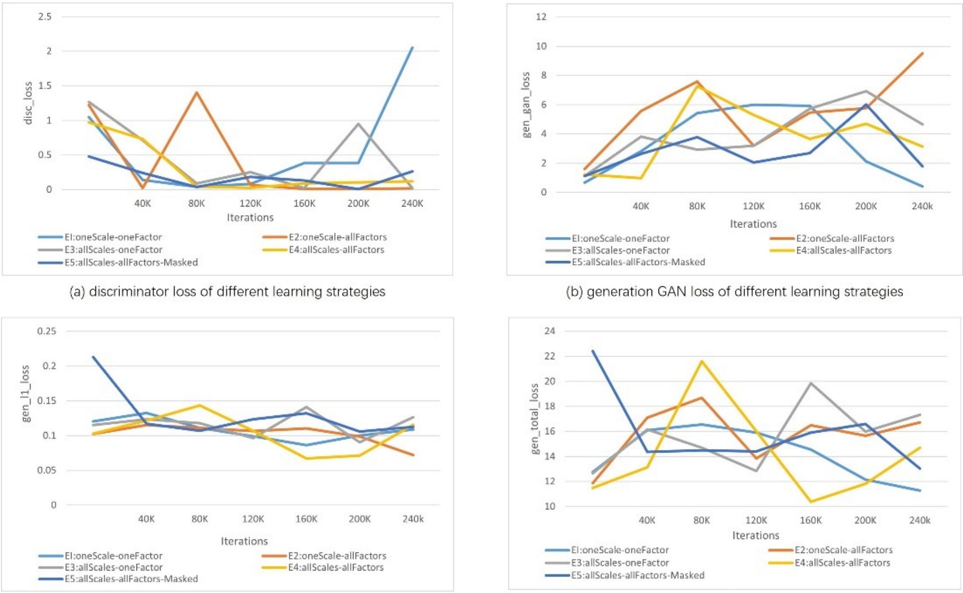
Experimental Results

The given style was constantly learned during the training. Fig. 6 demonstrates the comparison of the experimental results. From E1 to E6, the lofty-distance scenes were clearly reproduced first, followed by the flat-distance scenes, and far-distance scenes. As the computation progressed, the modelling structure and brushwork details were represented more and more completely. All kinds of objects were captured in a relatively balanced manner, even the point-view figure as small as pod. Note that part of the visual content generated by E5 and E6 is not in the input sample data, which indicates that the model has a certain generation ability based on analogy.

The Impact of Data Selection

In order to gain a deeper understanding of the effect of data selection, we analysed the dynamically changing of the loss (see Fig. 7).

Trough analysing the loss, we have the following findings. Firstly, compared to E1, E2 reduced the discriminator loss and increased the generation loss. Within the generation loss, E2 led to an obvious increase in generation_gan_loss. This shows that the information brought by multi-granularity data significantly helps to improve the alignment of the brushwork pattern. Secondly, compared to E2, E4 increased the discriminator loss while reduced



(a) discriminator loss of different learning strategies

(b) generation GAN loss of different learning strategies

(c) generation L1 loss of different learning strategies

(d) generation total loss of different learning strategies

Figure 7: The dynamic changes of loss: (a) discriminator loss, (b) generation GAN loss, (c) generation L1 loss, and (d) generation total loss.

the generation loss. Within the generation loss, E4 led to a reduction in generation_gan_loss and an increase in generation_L1_loss. This indicates that the spatial cognition integrated in the multi-scale data is more informative for the alignment of modelling structure. Thirdly, compared to E4, E5 led to an increase in both discriminator loss and generation loss. It can be considered that the masked data drives the model to apply the learned style to a new context.

PRESCRIPTIVE RESTORATION TESTS

After the aesthetic computational experiments, the third step is to perform a prescriptive restoration test via human-AI collaboration. Fig. 8 shows some of the results of tests, with each subfigure providing a particular restorative form effect. Our first test scenario was generated by erasing some contours of the painting. As illustrated in column (b), the generated contents inspired by manual erasing include the blank sky, a turbulent river at the foot of the mountains, and the rapids between the valleys. The second test scenario came from adding contours on the painting. As can be seen in column (c), the generated contents guided by the added lines are not only reasonable, but also poetic, blending naturally with the original images.

If an ancient painting has large areas of damage, the restorers typically execute cut and mounting, and then add strokers to the transitional area to make this part look less abrupt. Fig. 9 shows how we simulated this scenario by generating cut-mix drafts containing the parts of draft A and draft B. As can be seen from column (e), the model can basically complete the cognitive recompositing task.

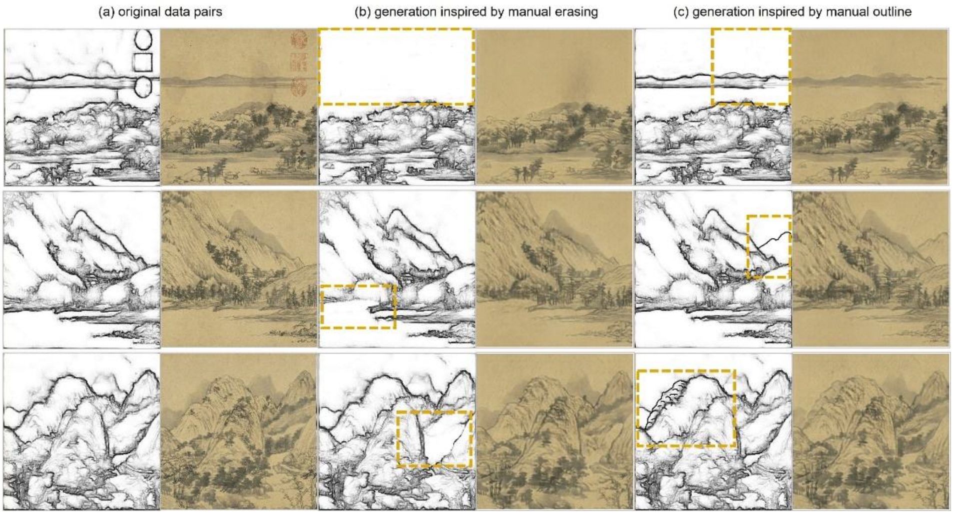


Figure 8: Examples of analogized generation tests: (a) original data pairs, (b) generation inspired by manual erasing, and (c) generation inspired by manual outline.

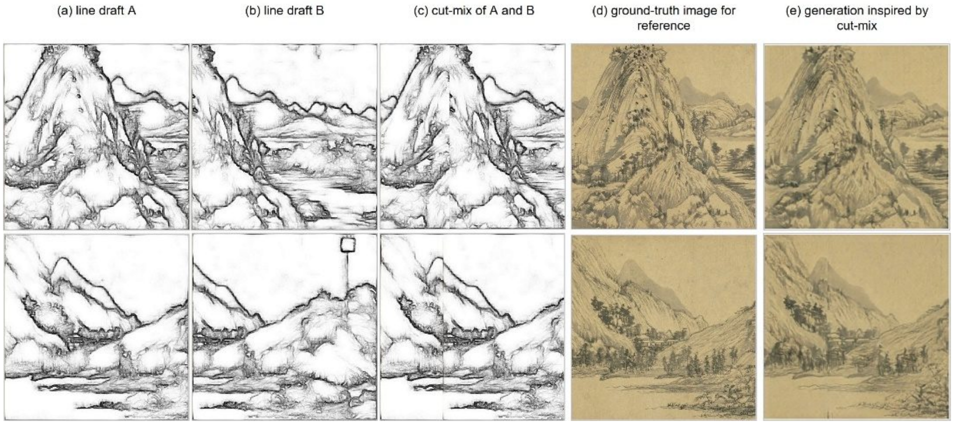


Figure 9: Examples of cognitive recompositing tests: (a) line draft A, (b) line draft B, (c) the cut-mix of A and B, (d) ground-truth image for reference, and (e) generation inspired by cut-mix.

CONCLUSIONS AND DISCUSSIONS

In this paper, we propose a joint cognition paradigm consisting of aesthetic cognition and aesthetic computation. The experiments in this paper are elementary. The methods and conclusions of this study need to be validated in more painting style alignment experiments. In addition, this study has not covered the issue of colour restoration of Chinese landscape painting. All the above will be the directions of our future research. Using painting style alignment is a step in the direction of human-AI collaboration for ancient paintings conservation, and we believe further steps are possible.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support by the Humanities and Social Sciences Research Foundation of the Ministry of Education (No. 21YJA760005).

REFERENCES

Cole, E., Yang, X., Wilber, K., Mac Aodha, O., and Belongie, S. (2022) “When does contrastive visual representation learning work?”, proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Tel Aviv, Israel.

Cooper, A., Reimann, R., Cronin, D., and Noessel, C. (2014). About face: the essentials of interaction design. New York: John Wiley & Sons.

Dove, G., Halskov, K., Forlizzi, J., and Zimmerman, J. (2017) “UX design innovation: Challenges for working with machine learning as a design material”, proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver Colorado, USA.

Gabriel, I. (2020). Artificial intelligence, values, and alignment, MINDS AND MACHINES Volume 30 No. 3.

Ghorbani, A., and Zou, J. (2019) “Data shapely: Equitable valuation of data for machine learning”, proceedings of the 36th International Conference on Machine Learning, California, USA.

He, B., Gao, F., Ma, D., Shi, B., and Duan, L. Y. (2018) “Chipgan: Agenerative adversarial network for Chinese ink wash painting style transfer”, proceedings of the 26th ACM International Conference on Multimedia, Seoul, Republic of Korea.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., and Girshick, R. (2022) “Masked auto encoders are scalable vision learners”, proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA.

Hohman, F., Wongsuphasawat, K., Kery, M. B., and Patel, K. (2020) “Understanding and visualizing data iteration in machine learning”, proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA.

Isola, P., Zhu, J. Y., Zhou, T., and Efros, A. A. (2017) “Image-to-image translation with conditional adversarial networks”, proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA.

Kreutzer, J., Caswell, I., Wang, L., Wahab, A., van Esch, D., Ulzii-Orshikh, N.,... and Adeyemi, M. (2022). Quality at a glance: Anaudit of web-crawled multilingual datasets, TRANSACTIONS OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS Volume 10 No. 5.

Kurdziolek, M. (2022). Explaining the Unexplainable: Explainable AI (XAI) for UX, UXPA MAGAZINE Volume 22 No.3.

Lee, M. K., Kusbit, D., Kahng, A., Kim, J. T., Yuan, X., Chan, A.,... and Procaccia, A. D. (2019) “WeBuildAI: Participatory framework for algorithmic governance”, proceedings of the ACM on Human-Computer Interaction, New York, NY, USA.

Liu, M., Wang, G., Hu, J., and Chen, K. (2022) Multiple heterogeneous network representation learning based on multi-granularity fusion. INTERNATIONAL JOURNAL OF MACHINE LEARNING AND CYBERNETICS Volume 13 No. 10.

Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B.,... and Chen, M. (2021). Glide: Towards photo realistic image generation and editing with text-guided diffusion models. arXiv preprintarXiv:2112.10741.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P.,... and Lowe, R. (2022). Training language models to follow instructions with human feedback. arXiv preprint arXiv:2203.02155.

Ronneberger, O., Fischer, P., and Brox, T. (2015) "U-net: Convolutional networks for biomedical image segmentation", proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention, Munich, Germany.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2022) "High-resolution image synthesis with latent diffusion models", proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA.

Sheng, J., Song, C., Wang, J., and Han, Y. (2019). Convolutional neural network style transfer towards Chinese paintings. IEEE ACCESS Volume 7 No. 11.

Van Den Oord, A., Kalchbrenner, N., and Kavukcuoglu, K. (2016) "Pixel recurrent neural networks", proceedings of the International Conference on Machine Learning, New York City, NY, USA.

Xue, A. (2021) "End-to-end Chinese landscape painting creation using generative adversarial networks", proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, Hawaii, USA.