Arts, Communication & Digital Media

journal homepage: https://j.starnexus.press/index.php/acdm/en

Research on Deep Learning-based Digital Restoration Method for Images of Intangible Cultural Heritage Thangka

Jianling Li[1], Qiong Sun[2*]

Management College, Beijing Union University, Beijing, China

ARTICLE INFO

Keywords:
Thangka image
generative adversarial network
U-Net
image restoration

ABSTRACT

Thangka is part of China's intangible cultural heritage, but many Thangka images have certain defects due to problems such as age and improper preservation. This paper addressed the problem of digital restoration of Thangka images. Based on the generative adversarial network (GAN) in deep learning, a method was designed with U-Net as the generative model and PatchGAN as the discriminative model. The loss function of the model was composed of reconstruction loss, perceptual loss, and adversarial loss. Thangka images were collected to establish a dataset to verify the proposed approach. The results showed that at a mask ratio of 1%-10%, the peak signal-to-noise ratio, structural similarity index, and learned perceptual image patch similarity of the proposed approach were 33.21 dB, 0.9825, and 0.0187, and its performance was superior to that of methods such as EdgeConnect and progressive reconstruction of visual structure network at different mask ratios. The findings demonstrate the reliability of the proposed approach for digital restoration of images. It can be applied to the digital restoration of actual Thangka images.

1. Introduction

Intangible cultural heritage (ICH), such as traditional fine arts, craftsmanship, dance, and music, is an important manifestation of the world's cultural diversity [1]. Digital technology has also been increasingly widely applied in ICH protection [2]. The Internet, as a communication and dissemination platform, provides a guarantee for the digital protection of ICH [3]. Technologies such as virtual reality, artificial intelligence (AI), and augmented reality are increasingly common in the digital protection of heritage [4]. Gu [5] studied how South Korea used virtual reality, 3D scanning, and geographic information systems to protect tangible and intangible cultural assets. Song et al. [6] explored the digital modeling methods of urban cultural spaces based on geographic information systems and 3D laser scanning technology, and conducted case analysis to propose an integration path for digital modeling and cultural heritage protection. Shen et al. [7] studied the application of computer technology in the digital protection of Bashu's ICH and established a digital protection model, providing useful references for global ICH protection. Wu et al. [8] applied 3D digital programming and virtual reality to the protection and inheritance of Dongyang bamboo weaving craftsmanship, providing references for the digital protection and inheritance of ICH. Thangka, as an embodiment of Tibetan culture, is a typical representative of ICH [9]. The images of Thangka embody the ideas, culture, and religious sentiments of the Tibetan people, with rich variations in color and texture, diverse content, and high ornamental value. However, many Thangka images have been damaged in the process of being passed down due to age and improper preservation, and need to be restored. Traditional manual restoration methods have very high demands on the technical skills of restorers and are less efficient. Image restoration methods that rely on digital technology are now widely used in fields such as medicine and culture [10]. Dalal et al. [11] put forward a convolutional neural network (CNN)-Di strategy bobcat optimization algorithm. It achieved a PSNR of 56.76 dB On the Set12 and Berkeley segmentation dataset 68, demonstrating efficient image restoration. Sirikongtham et al. [12] introduced a generative adversarial network (GAN)-based model enhanced with spectral normalization to restore QR codes. The model achieved an average peak signal-to-noise ratio (PSNR) of 28.5 dB and a structural similarity index (SSIM) of 0.91, providing a viable approach for QR code restoration. Abtahi et al. [13] compared the restoration effects of the real-enhanced super-resolution GAN and the Swin Transformer-based image restoration network (SwinIR) for tooth bite wing X-ray films. They found through experiments on 4,004 bitewing X-ray films that SwinIR performed better, with a PSNR of 35.56 dB and a SSIM of 0.9287. Han et al. [14] proposed a dynamic hybrid Transformer-CNN model to restore remote sensing images: DHTCformer. A large number of experimental outcomes showed that this method outperformed current leading techniques. This paper designed a GAN-based restoration model for the digital restoration of Thangka images and verified its usability on a Thangka image dataset, providing a reference method for the digital protection of Thangka and some references for the image restoration of similar cultural relics.

2. Deep Learning Methods for Digital Restoration

In deep learning, digital restoration methods based on a GAN have been widely applied in image restoration. The principle of GAN is the adversarial game between the generative model (G) and the discriminative model (D). The objective function can be written as:

$$\min_G \max_D V(D,G) = E_{x\sim P_{data}(x)}[\log D(x)] + E_{z\sim P_z(z)}[\log(1-D(x))],$$

where x stands for a real sample, $P_{data}(x)$ stands for true sample distribution, z is the random noise, and $P_z(z)$ is the distribution of noise vector data.

Received September 10, 2026; accepted October 10, 2026
* Corresponding author.
E-mail address: sunqiongbuu@163.com

https://doi.org/10.68791/acdm2026002
Print ISSN: 3135-8381 | Online ISSN: 3135-839X

In the training process, G should produce data as similar as possible to the real sample, and D should strive to enhance its ability to distinguish between real and fake data. Through mutual game, D is difficult to distinguish between the output of G and real samples ultimately. However, in the actual training process, GANs have problems such as vanishing gradients and pattern crashes, and as a result, innovative methods such as conditional generative adversarial networks (CGAN) and BigGAN [15] have emerged.

For the digital restoration of Thangka images, this paper makes improvements based on the GAN architecture to enhance the image restoration effect. The model architecture is shown in Figure 1.

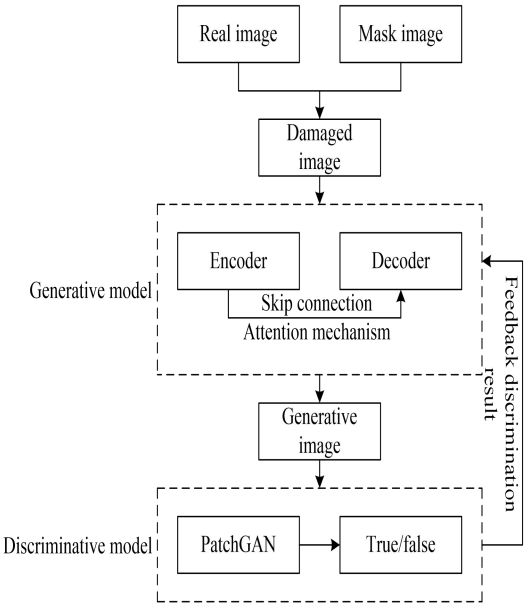


Figure 1. Thangka image restoration model

According to Figure 1, in the generative model part, the U-Net structure is used, and skip connections and attention mechanisms are introduced to extract the context and semantic information of the image more deeply. In the discriminative model part, the PatchGAN structure [16] is used, and spectral normalized convolution is used instead of traditional convolution to enhance the texture quality of the image. The design is as follows.

(1) Generative model

The advantage of U-Net is its ability to capture deep semantic information, which is beneficial for generating fine details for Thangka images and enhancing the consistency between the restored and original images. Based on the original U-Net structure, four layers of dilated convolution are added to expand the range of the perceptual field. Dilation achieves the expansion of the convolution kernel by inserting spaces between the elements of the convolution kernel, thereby increasing the receptive field exponentially. For a standard 3×3 convolution, the kernel size is 3×3, the step size is 1, and the receptive field is 3. The core parameter of dilation is the expansion rate (d). When it takes 1, it is normal convolution; if $d>1$, for example, $d=2$, at this time, the convolution kernel no longer acts on adjacent pixels but takes a point every one pixel for operation, then the effective receptive field becomes: $k+(k-1)\times(d-1)=3+(3-1)\times(2-1)=5$, the covering range is 5×5=25 pixel regions, but the parameter quantity remains to be 9.

After extracting deep features through dilated convolution, the self-attention mechanism is introduced. For an input feature map $\boldsymbol{x}\in\mathbb{R}^{C\times H\times W}$, where C is the number of channels, the query matrix (Q), key matrix (K), and value matrix (V) are generated through convolution. The attention score is calculated based on the similarity between features:

$$\text{Attention}(Q,K,V)=\text{Softmax}(\frac{Q^{\mathrm{T}}K}{\sqrt{d}})V,$$

where d is the feature dimension. The feature map finally output is:

$$\boldsymbol{x}_{\text{out}}=\boldsymbol{x}+\gamma\cdot\ \text{Attention}(Q,K,V),$$

where γ is the learnable parameter.

For the selection of activation functions, Conv1-Conv5 uses rectified linear unit (ReLU) to reduce computational complexity, Conv6-Conv8 uses Leaky-ReLU to avoid gradient loss, and Conv9-Conv10 uses Tanh to accelerate convergence. The formulas for each loss function are as follows.

① ReLU

$$\text{ReLU}(x)=max(0,x)$$

② Leaky-ReLU

$$\text{LeakyReLU}(x)=\begin{cases}x,x>0\\ \alpha x,x\le 0\end{cases}$$

where α takes a value of 0.2 in case the gradient disappears.

③ Tanh

$$\text{Tanh}(x)=\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$$

Table 1 shows the structure of the generative model.

Table 1. Generative model structure

Layer	Kernel	Step	Output
Conv1	5 × 5	1 × 1	64
Conv2	3 × 3	2 × 2	128
Conv3	3 × 3	1 × 1	128
Conv4	3 × 3	2 × 2	256
Conv5	3 × 3	2 × 2	256
Dilated1	3 × 3	1 × 1	256
Dilated2	3 × 3	1 × 1	256
Dilated3	3 × 3	1 × 1	256
Dilated4	3 × 3	1 × 1	256
Conv6	3 × 3	1 × 1	256
Conv7	3 × 3	2 × 2	256
Deconv1	2 × 2	2 × 2	128
Conv8	3 × 3	1 × 1	128
Deconv2	2 × 2	2 × 2	64
Conv9	3 × 3	1 × 1	32
Conv10	3 × 3	1 × 1	3

(2) Discriminative model

The principle of PatchGAN is to divide an image into N blocks of the same size, determine the authenticity of each block individually, and then take the average as the score of the entire image. Compared with global discrimination, PatchGAN pays

more attention to local details and reduces computational overhead. Based on PatchGAN, spectral normalization convolutional layers are used. Spectral normalization constrains the Lipschitz constant of the discriminative network to 1 to enhance the training stability of the GAN.

The definition of the Lipschitz constraint [17] is as follows. For a function $f:\mathbb{R}^m \to \mathbb{R}^n$, if there exists a constant $K \geq 0$ such that for any two inputs x_1 and x_2 within the domain both satisfy $\|f(x_1)-f(x_2)\| \leq K \cdot \|x_1-x_2\|$, then f is said to satisfy the Lipschitz constraint and K is the Lipschitz constant.

(3) Loss function

The loss function optimizes the model by quantifying the deviation between the predicted and true values of the model. The loss functions used in this paper are as follows.

① Reconstruction loss: It measures the difference between the image generated by the model and the real image. The formula is:

$$L_{mse}=\frac{1}{M}\sum_{m=1}^{M}(P_{data}(Y_m)-p_z(y_m))^2,$$

where M is the training image set, $P_{data}(Y_m)$ is the input image, and $p_z(y_m)$ is the output image.

② Perceptual loss: Through a fixed network, such as VGG16 [18], real images and model-generated images are used as inputs respectively, and then the loss is constructed using the corresponding output features:

$$L_{perc}=E\left[\sum_n \left\|\phi_n(I_{out})-\phi_n(I_{gt})\right\|\right],$$

where ϕ_n denotes the activation map obtained by applying the input image to the n-th pooling layer of VGG16.

③ Adversarial loss:

$$L_{Adv}^{SR}=\sum_{n=1}^{N}-\log D_{\theta_D}\left(G_{\theta_G}(I^{LR})\right),$$

where D_{θ_D} is the probability that the generated image is true and $G_{\theta_G}(I^{LR})$ is the generated image.

3. Results and Analysis

3.1 Experimental Configuration

The programming environment was Ubuntu and Python 3.6, and the model was trained and tested on the PyTorch framework. The hardware configurations were an NVIDIA RTX 3090 graphics processor, an Intel Core i9-10900K central processing unit (CPU), and 32GB memory. The model was optimized using the Adam optimizer. The batch size was set to 8. In the initial stage of training, the learning rate of the generative model was set to 10-4, and the learning rate of the discriminative model was set to 10-5. The learning rate of the generative model was reduced to 10-5 after it stabilized. The training consisted of 100 epochs.

3.2 Dataset

The Thangka images were collected through the online platforms, including publicly available digital cultural relic resources from the Tibet Museum (https://www.tibetmuseum.com.cn/) and digital display resources of Thangka published by China ICH Network (https://www.ihchina.cn/) and the Beijing Palace Museum (https://www.dpm.org.cn/). Totally 2,600 original images were collected, and 2,000 valid samples were retained after excluding images that met any of the following criteria: pixel dimensions below 512×512, blurriness, severely fading, heavy watermarking, or duplication. The image processing library PIL of Python was used to unify the size of the collected images to 512×512. Then, data augmentation was performed: random horizontal flip: The flipping probability was 0.5; random scaling: the scaling ratio range was 0.8-1.2, and after scaling, it was re-cropped to 512 × 512; random rotation: the rotation angle range was -5° to 5°. Finally, 10,000 images were finally obtained as the experimental dataset. It was segmented into a training set, a test set, and a validation set in an 8:1:1 ratio. The mask images used the irregular mask dataset made public by Liu et al. [19]. Some of the images are shown in Figures 2 and 3.



Figure 2. Examples of Thangka images



Figure 3. Mask images

3.3 Evaluation Indicators

(1) PSNR

It represents the error of the restored image $\hat{I}$ and the original image I_{gt} at the pixel level (unit: dB). The formula is:

$$MSE=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}(I_{gt}(i,j)-\hat{I}(i,j))^2,$$

$$PSNR=10\log_{10}\left(\frac{MAX^2}{MSE}\right),$$

where $H\times W$ is the image size and MAX is the maximum pixel value of the image.

(2) SSIM

SSIM can evaluate the similarity of an image in terms of brightness, contrast, and structural information:

$$\mathrm{SSIM}(I_{\mathrm{gt}},\hat{I})=\frac{(2\mu_{I_{\mathrm{gt}}}\mu_{\hat{I}}+C_1)(2\sigma_{I_{\mathrm{gt}}\hat{I}}+C_2)}{(\mu_{I_{\mathrm{gt}}}^2+\mu_{\hat{I}}^2+C_1)(\sigma_{I_{\mathrm{gt}}}^2+\sigma_{\hat{I}}^2+C_2)},$$

where $\mu_{I_{\mathrm{gt}}}$ and $\mu_{\hat{I}}$ are the mean pixel of an image, $\sigma_{I_{\mathrm{gt}}}$ and $\sigma_{\hat{I}}$ are the standard deviations of an image, $\sigma_{I_{\mathrm{gt}}\hat{I}}$ is the covariance of an image, and C_1 and C_2 are constants that avoid the denominator to be zero.

(3) Learned perceptual image patch similarity (LPIPS)

LPIPS compares the depth features extracted by the pre-trained network (VGG16), and its formula is:

$$\mathrm{LPIPS}=\sum_{n=1}^{N}\frac{1}{H_nW_n}\sum_{h,w}w_n\left|\phi_n(I_{\mathrm{gt}})h,w-\phi_n(\hat{I})h,w\right|_2^2,$$

where ϕ_n is the normalized feature of the output of the n-th layer of the network and W_n is the weight parameter obtained through network training.

3.4 Result Analysis

According to Table 2, the performance of Thangka image restoration decreased as the mask ratio increased. Specifically, when the mask ratio was 1-10%, the PSNR obtained by this method was 33.21 dB, the SSIM was 0.9825, and the LPIPS was 0.0187. At this point, for minor defects in the image, this method can achieve a relatively high-quality restoration. When the mask ratio was 50%-60%, the PSNR was 18.57 dB, the SSIM was 0.6426, and the LPIPS was 0.2662. At this point, the quality of the restored image showed a significant decline. The results indicate that as the defect area expands, the difficulty of restoring Thangka images increases and the effect of restoration declines.

Table 2. Restoration performance at different mask ratios

	PSNR/dB↑	SSIM↑	LPIPS↓
1%-10%	33.21	0.9825	0.0187
10%-20%	29.14	0.9567	0.0456
20%-30%	25.46	0.9034	0.0835
30%-40%	22.78	0.8485	0.1374
40%-50%	20.46	0.7451	0.1933
50%-60%	18.57	0.6426	0.2662

Note: ↑ indicates that the larger the value, the better; ↓ indicates that the smaller the value, the better.

To verify the advantages of the proposed method in Thangka image restoration, it was compared with some current methods.

Table 3. Comparison results with other methods

		EdgeConnect (EC) [20]	Progressive reconstruction of visual structure (PRVS) network [21]	End-to-end cascaded refinement with mask awareness [22]	Aggregated contextual-transformation GAN (AOT-GAN) [23]	The proposed model
PSNR/dB↑	1% - 10%	25.57	31.12	31.85	32.57	33.21
	10% - 20%	21.62	27.09	27.83	28.51	29.14
	20% - 30%	18.15	23.41	24.12	24.83	25.46
	30% - 40%	15.82	20.72	21.43	22.15	22.78
	40% - 50%	13.79	18.39	19.12	19.84	20.46
	50% - 60%	12.11	16.51	17.23	17.92	18.57
SSIM↑	1% - 10%	0.8253	0.9621	0.9694	0.9763	0.9825
	10% - 20%	0.7992	0.9328	0.9415	0.9492	0.9567
	20% - 30%	0.7528	0.8773	0.8861	0.8954	0.9034
	30% - 40%	0.7005	0.8192	0.8295	0.8391	0.8485
	40% - 50%	0.6102	0.7128	0.7241	0.7356	0.7451
	50% - 60%	0.5227	0.6079	0.6203	0.6318	0.6426
LPIPS↓	1% - 10%	0.0873	0.0332	0.0281	0.0234	0.0187
	10% - 20%	0.1921	0.0665	0.0593	0.0521	0.0456
	20% - 30%	0.3266	0.1104	0.1012	0.0923	0.0835
	30% - 40%	0.5062	0.1705	0.1593	0.1482	0.1374
	40% - 50%	0.6834	0.2351	0.2214	0.2076	0.1933
	50% - 60%	0.8884	0.3147	0.2985	0.2821	0.2662

According to Table 3, first of all, in the comparison of PSNR, taking the mask ratio of 1%-10% as an example, the proposed method (33.21 dB) was 0.64 dB higher than the AOT-GAN method (32.57 dB), which performed the best, and 7.64 dB higher than the EC method (25.57 dB), indicating that even in the case of a relatively small degree of damage, the proposed method also showed certain advantages. In terms of SSIM comparison, the proposed method maintained an SSIM of more than 0.90 under 1%-30% mask ratio. In the face of large area defects in Thangka images (50%-60% mask ratio), the SSIM (0.6426) of the proposed model was 0.0108 higher than that of the AOT-GAN method (0.6318) and 0.1199 higher than the EC method (0.5227), demonstrating the superiority of this method in maintaining structural consistency. In terms of LPIPS comparison, the LPIPS of the proposed model was 0.2662 when the mask ratio was 50% -60%, while the LPIPS of the EC method was 0.8884, indicating poor perceptual quality. This result indicates that in the face of large areas of image defect, the proposed model can make relatively reasonable inferences about the occluded semantic information and obtain higher-quality restored images.

To verify the contributions of different components in the model, an ablation experiment was conducted, and the results are presented in Table 4.

Table 4. Ablation experiment

	PSNR/dB↑	SSIM↑	LPIPS↓
Baseline (U-Net+PatchGAN+MSE loss)	30.12	0.9437	0.0426
Baseline+dilated convolution	31.58	0.9612	0.0315
Baseline+dilated convolution+self-attention mechanism	32.27	0.9726	0.0258
Baseline+dilated convolution+self-attention mechanism+spectral normalization	32.79	0.9774	0.0213
The complete model	33.21	0.9825	0.0187

According to Table 4, compared to the baseline, the dilated convolution technique resulted in a 1.46B improvement in PSNR, effectively enhancing the overall structural restoration by expanding the receptive field; the self-attention mechanism contributed greatly to the improvement in SSIM, enabling the model to focus more on key semantic regions in Thangka images, such as faces and decorative patterns; spectral normalization effectively reduced LPIPS, decreasing artifacts and blurring and making the generated textures more natural; finally, the loss function used further reduced LPIPS, resulting in stronger visual consistency between the restored regions and the original regions. These results indicate that each component has a positive effect on the digital restoration of Thangka images.

4. Discussion

Thangka, as one of the forms of ICH art, has certain constraints on the AI restoration due to its artistic characteristics. The color scheme of Thangka follows a fixed pattern, and the colors used in AI restoration must conform to the color logic of Thangka. Thangka images contain a large number of clothing patterns, auspicious clouds, and flames, thereby having high requirements for detail generation. Additionally, handprints and ritual objects in Thangka may be related to religious constraints, and incorrect restoration may lose their cultural connotations. Therefore, the AI restoration of Thangka images has relatively high requirements. This paper developed a restoration model based on GAN. One of its main improvement was the introduction of dilated convolution, enabling the model to expand the receptive field without increasing the parameter quantity and match the large-scale color blocks and overall composition of Thangka. Moreover, it combined the self-attention mechanism to strengthen the feature association with key semantic regions. For the processing of texture details, a spectral-normalized PatchGAN discriminator was used to enhance the model's attention to fine textures. Finally, a multi-dimensional loss function was used to consider both objective quality and subjective visual quality, improving the image restoration effect.

It can be seen from the experimental results that the GAN-based model achieved high-quality repair (PSNR = 33.21 dB) when the defect degree was relatively mild. However, as the mask ratio increased, the repair performance decreased. When the mask ratio reached 50%-60%, the PSNR was 18.57 dB. However, it can be found from Table 3 that under the condition of a large defect area, the repair effect of the GAN-based model was better than that of methods such as EC and PRVS network. Finally, the ablation experiment results showed that each component in the GAN-based model contributed to the improvement of the repair effect, jointly achieving the enhancement of model performance.

However, this study also has some limitations. The dataset mainly consisted of Thangkas with mainstream themes of Tibetan Buddhism, and lacks validation of the restoration performance on lesser-known themes such as folk Thangkas. In the evaluation process, only objective image evaluation indicators were used, without introducing qualitative assessments by Thangka restoration experts or relevant experts. The cultural legitimacy and artistic value of the restoration results are still unclear. Therefore, in future work, methods such as few-sample learning can be considered to strengthen the model's capability to restore rare themes, and the restoration algorithms for high-resolution (4K and above) Thangka images can further explored. Combined with the opinions of cultural restoration experts, the integration of technology and cultural experience can be achieved.

5. Conclusion

This paper designed a GAN-based restoration model for the digital restoration of Thangka images and tested its restoration performance on the Thangka image dataset. The results showed that this model had a PSNR of 33.21 dB, an SSIM of 0.9825, and an LPIPS of 0.0187 when the mask ratio was 1%-10%. Even in the case of large area defects (with a mask ratio of 50%-60%), it still showed better restoration performance than the other methods, with better pixel accuracy, structural consistency, and perceptual quality. This paper can support the digital protection and inheritance of Thangka images.

Acknowledgment

The authors declare that there are no acknowledgments to report.

Conflict of Interest

The author declares no conflict of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author's contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Li Jianling and Sun Qiong. The first draft of the manuscript was written by Li Jianling and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

1. Yang, S. Safeguarding and inheriting intangible cultural heritage music in the Chongqing Three Gorges Reservoir area: A case study on Lore. Curator, 2025, 68(1), 125-140.
2. Ding, F.; Ding, B.; Bai, Y.; Zhang, J. Emotional Drain in Digital Design for Inheritance of Intangible Cultural Heritage. E3S Web of Conferences, 2021, 236, 05029.
3. Liu, X.; Wang, T.; Taylor, M. Intangible Cultural Heritage Digital Protection System Based on Parallel Processing of Hardware and Software. The International Conference on Cyber Security Intelligence and Analytics, 2022.
4. Wen, W.; Wang, X.; Ye Z. Protection and immersive experience of cultural heritage in the digital age: Technological empowerment, spatial reproduction, and future risks. Arts & Communication, 2025, 3(4), 025040005.
5. Gu, B. K. Digital Solutions for Cultural Heritage Protection: Korea's ODA Contributions to the BIMP-EAGA. SUVANNABHUMI, 2025, 17(1), 33-52.
6. Song, C.; Liu, L.; Wang, Z. Digital modeling of urban cultural space and research on cultural heritage protection method. Journal of

Computational Methods in Sciences and Engineering, 2025, 25(6).
7. Shen, Z.; Xu, Q.; Wang, M.; Xue Y. Application and Exploration of Computer Technology in Digital Protection of Bashu Intangible Cultural Heritage. 2024 5th International Conference on Information Science, Parallel and Distributed Systems (ISPDS), 2024, 492-497.
8. Wu, J.; Guo, L.; Jiang, J.; Sun, Y. The Digital Protection and Practice of Intangible Cultural Heritage Crafts in the context of New Technology. E3S Web of Conferences, 2021, 236, 05024.
9. Zhong, P.; Hu, W.; Zhao, Y.; Zhang, F. Geo-TCAM: a Thangka captioning method integrating topic modeling with geometry-guided spatial attention. npj Heritage Science, 2026, 14(1), 1-18.
10. Shang, X. P. Research on Lingnan Culture Image Restoration Methods Based on Multi-Scale Non-Local Self-Similar Learning. IEEE Access, 2025, 13, 101639-101646.
11. Dalal, H. N.; Patel, H. K.; Prajapati, N. K.; Shyal, N. M. Image Restoration Using Convolutional Neural Network-Di Strategy Bobcat Optimization Algorithm. International Journal of Intelligent Engineering & Systems, 2025, 18(5), 690.
12. Sirikongtham, P.; Nimkoompai, A. Enhancing the Scanability of Damaged QR Codes Through Image Restoration Using GANs Combined with the Spectral Normalization Technique. International Journal of Advanced Computer Science & Applications, 2025, 16(10), 638.
13. Abtahi, M.; Majidinia, S.; Mansourzadeh, M.; Zarch, S.H.H.; Asadi, G.; Noghani, Y.T.; Shakiba, M.R.; Saeedi, A.; Yousefi, S. Enhancing diagnostic quality in dental bitewings using transformer and GAN-Based image restoration. Oral Radiology, 2026, 42(3), 625-634.
14. Han, X.; Yang, H.; Xie, C. DHTCformer: A dynamic hybrid transformer-CNN for weather-degraded remote sensing image restoration. Signal, Image and Video Processing, 2026, 20(8), 475.
15. Pandey, A.; Trehan, J.; Ram, P. K.; Singh, S. Comparative Analysis of Loss Functions in BigGAN for Medical Image Synthesis: A Case Study on Brain MRI Scans. 2025 International Conference on Cognitive Computing in Engineering, Communications, Sciences and Biomedical Health Informatics (IC3ECSBHI), 2025.
16. Li, H.; Wang, N.; Fang, J.; Jia, Y.; Ji, L.; Chen, X. Anime Sketch Coloring Based onSelf-attention Gate andProgressive PatchGAN. 2023 Chinese Conference on Pattern Recognition and Computer Vision (PRCV), 2024, 237-249.
17. Zhang, Y.; Nie, B.; Gao, Y. Robust Locomotion Policy With Adaptive Lipschitz Constraint for Legged Robots. IEEE Robotics and Automation Letters, 2025, 10(1), 272-279.
18. Paul, H.; Udayangani, H.; Umesha, K.; Lankasena, N.; Liyanage, C.; Thambugala, K. Maize leaf disease detection using convolutional neural network: a mobile application based on pre-trained VGG16 architecture. New Zealand Journal of Crop and Horticultural Science, 2024, 53(2), 367-383.
19. Liu, G.; Reda, F. A.; Shih, K. J.; Wang, T. C.; Tao, A.; Catanzaro, B. Image Inpainting for Irregular Holes Using Partial Convolutions. 15th European Conference on Computer Vision, 2018, 89-105.
20. Nazeri, K.; Ng, E.; Joseph, T.; Qureshi, F.; Ebrahimi, M. EdgeConnect: Structure Guided Image Inpainting using Edge Prediction. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 2019, 3265-3274.
21. Li, J.; He, F.; Zhang, L.; Du, B.; Tao, D. Progressive Reconstruction of Visual Structure for Image Inpainting. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, 5961-5970.
22. Zhu, M.; He, D.; Li, X.; Li, C.; Li, F.; Liu, X. Image Inpainting by End-to-End Cascaded Refinement With Mask Awareness. IEEE Transactions on Image Processing, 2021, 30, 4855-4866.
23. Zeng, Y.; Fu, J.; Chao, H.; Guo, B. Aggregated Contextual Transformations for High-Resolution Image Inpainting. IEEE Transactions on Visualization and Computer Graphics, 2023, 29(7), 3266-3280.



Sun Qiong (1984 –), female, native of Beijing; holds a Ph.D.; she is a Professor and Doctoral Supervisor at the School of Management, Beijing Union University, and serves as the Publishing Executive Editor of Arts, Communication & Digital Media. Her research focuses on tourism economy, digital cultural tourism, and big data applications. She has published over 70 papers in core journals such as China Soft Science and Journal of Beijing Union University (Humanities and Social Sciences).



Li Jianling (1972 –), female, native of Hengshui, Hebei Province; holds a Ph.D. in Engineering and completed postdoctoral research in Management Science and Engineering; she is a Professor and Master's Supervisor at the Business College, Beijing Union University. Her research focuses on enterprise management, ecological strategy, and digital-intelligent innovation. She has published over 40 papers in core journals such as Hebei Academic Journal and Academic Exchange, including 16 CSSCI-indexed papers.