Intelligent enhancement of ancient Chinese murals based on multi-scale parallel structure

Yuxin Ge*

Shanghai Foreign Language School Affiliated to SISU, Shanghai, China

Abstract. Ancient mural artwork preserves the historical background and cultural customs of that time through intricate details and bright colors. However, after the natural environment and man-made damage, these works of art are damaged in color, texture and content and lose their quality. In order to identify and enhance murals with large areas of color damage, we propose a multi-scale parallel GAN and parallel Unet structure, which can extract features from multiple scales or images to adapt to the changing scale of the target and provide a more diverse set of features. This structure can reduce the risk of overfitting the training data by learning more general features. The verification results of indicators such as PSNR on the ancient mural data set show that the method has a certain performance improvement effect.

1 Introduction

Murals are a valuable part of world cultural heritage and have great historical and research value. They are the spiritual home of modern people and a symbol of world civilization, reflecting the development of society, politics, economy, religion, culture and art in various countries in the world. However, due to long-term natural influence, changes in temperature and humidity, ancient mural artworks generally suffer from different degrees of degradation. Most of the degradation is manifested as cracks, peeling, peeling, etc., which greatly reduces the aesthetic value and appreciation of artworks. On the other hand, artificial intelligence (AI) technology is an activity dedicated to making machines intelligent, and it is developing rapidly in all fields of human beings. Virtual restoration based on AI technology can make a significant contribution to the field of cultural heritage protection.

Jianfang et al. [1] proposed a novel method for digital restoration of ancient murals. The main innovation lies in optimizing the network model by combining local and global discriminative networks, and this combination strengthens the consistency between the global and local representations of mural images output by the generative network. A mask of the original mural image is first selected, simulating the area to be repaired, and then added to the original mural image. The restored image is fed into a convolutional layer for mural image feature extraction. Finally, the mural image features are fed into the

* Corresponding author: geyuxin0806@126.com

deconvolution layer to restore the dimensions of the new image to those of the original image.

The paper [2] proposes a method of applying a generator-discriminator network to repair damaged murals. The whole repair process includes passing the damaged mural image through the generator network to obtain the repaired image. The corrupted images are then stitched together and fed into the discriminator network, which determines whether the input image was generated for the model or captured as a real image. The generator network is based on a modified version of the U-Net model and consists of an encoder and a decoder. The encoder and decoder are directly connected through a residual network. The encoder consists of eight encoding units.

The research in [3] presented a deep learning-based virtual repair technique for the China's Forbidden City's worn beams. This paper restored the three sections of ancient Chinese paintings separately using different technologies than a typical one to restore paintings. First, U-Net MobileNet was applied to the background of paintings to turn the color restoration problem into a semantic segmentation one, decreasing the repetitious work of restorers by teaching the computer the mapping relationship between actual colors and the oxidized colors. Second, traditional image processing technology such as Canny Edge Detection, morphological operations, and Otsu's thresholding method was used to gain the golden edges from the color maps produced by the semantic segmentation. In order to create a realistic dragon pattern based on hand-drawn dragon skeletons, the image translation method Pix2pix was used after drawing the skeletons in accordance with the dragon patterns. These three steps are superimposed to achieve the restoration of the photos of the weathered beams taken in the Forbidden City.

However, ancient murals usually have a long history, and have been affected by the natural environment and human activities for a long time, resulting in varying degrees of degradation. Although image restoration methods can help, it is still difficult to have a significant restoration effect on ancient murals that have suffered severe smoke or extensive color damage. In addition, how to identify the information under the pigment and mine the information that is difficult to identify requires further research on more complex image noise reduction and strong information extraction techniques.

We propose a multi-scale parallel GAN and Unet structure. First, multi-scale is introduced to extract features from images of multiple scales or resolutions to help the model better understand the content of the image, especially when the size of the object of interest changes; parallel Unet is used to increase the diversity of extracted features, Each U-Net in a parallel setting may learn to extract different types of features, thus providing a more diverse feature set for downstream tasks. Parallel GAN is more robust against overfitting. By training multiple U-Nets in parallel, the model can learn more general features, reducing the risk of overfitting the training data.

The remainder of this paper is structured as follows: the second section delves into related work; the third section introduces the multi-scale parallel GAN and U-Net; the fourth section presents our experiments and their results; and finally, we conclude with a summary.

2 Related works

To address the issue of thangka murals being damaged, the paper [4] suggest a damage sensitive and original restoration driven (DSORD) Thanka inpainting technique that can learn both mask features and image features. First of all, a method is provided for creating masks that is responsive to damage form, simulating mask shapes as actual damage to the Thanka paintings. Our masks imitate actual Thanka damage better than arbitrary masks, thus increasing the model's sensitivity to actual damages. Secondly, an original restoration-driven, two-phase learning process is proposed, in which varied loss functions are used for various training phases correspondingly, in contrast to existing learning-based approaches, which typically use fixed loss functions to train the network. The first phase's loss function is intended to direct the model toward reconstructing pixel information, while the second phase's loss function is intended to focus on high-level characteristics. Partial convolution-based Unet encoder-decoder architecture is applied to prevent the drawbacks of standard convolutions of CNN-based methods, as well as the failure of GAN based methods in image training due to its poor performance on a small dataset.

Deep learning technology has several challenges to overcome in order to be applied to mural painting [5]. First of all, gaining data sets might be hard. Furthermore, the supervised-learning-based network cannot be promoted since it cannot be used on real mural damages. Lastly, the unmasked area is not changed, because the unmasked region in the label image and the corresponding masked area in the produced image are combined in the deep neural network's output. To overcome these difficulties, the authors applied weak supervised learning to the inpainting of murals in order to fill in gaps and enhance the murals' overall aesthetic. A mural inpainting model based on the translation method with three domains is proposed, including a dense spatial attention with mask block embedded in the mapping net, which greatly improves the ability of the network to record the long-distance relationship between deep feature mapping and learns the transformation between the two deep spaces by paired data, and the SVD, which efficiently removes the high-frequency information while maintaining the majority of the structural and detail information, increasing the overlap between image features.

Cao et al. introduced an algorithm for super-resolution reconstruction of art mural images [6]. The structure of the algorithm is divided into two parts: generation network and discriminant network. The goal of the generative network is to output high-resolution images after super-resolution reconstruction. The purpose of the discriminative network is to judge the authenticity of the output image of the generative network and the real mural image. The design architecture of the generative network follows the encoder-decoder structure, which is mainly divided into feature extraction and image reconstruction. The algorithm also introduces a new loss function that combines the mean squared error (MSE) and the adversarial loss of the discriminative network. This combination helps to optimize the network model until the MSE tends to be relatively stable.

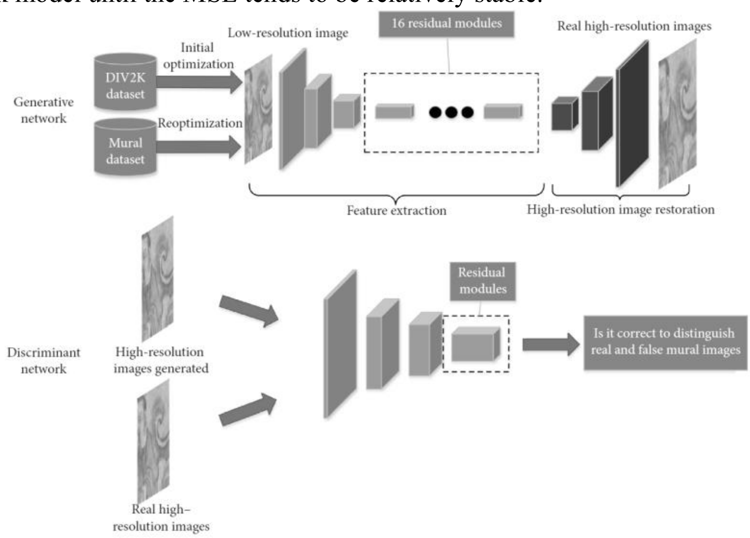

Fig. 1. The super resolution reconstruction of artistic mural image model [6].

3 Multi-scale parallel structure

Wang et al. [7] proposed a gradient guided dual-branch generative adversarial networks (GANs) for high-quality relic sketch generation, consisting of the sketch generation branch (SGB) and the auxiliary gradient-image generation branch (GGB). The SGB uses true sketches as the goal outputs and painted cultural relic images as the inputs, capturing elements from the original image including color and texture. The GGB focuses on the strong gradient guidance in sketches by using the gradient of relic photos as inputs and gradient images of true sketches as goal outputs. The combination of these two GANs effectively diminishes problems like noises, broken lines, recognition disabilities of complex scenes, detail loss and poor style similarity between generated and real sketches caused by methods simply based on gradient, learning or CNN. Furthermore, a feature transmission module and fusion block are applied for the transference of characteristics between two branches. The experiment results proved that GGD-GAN improves the accuracy, completeness and coherence of sketches of painted cultural relics.

The paper [8] proposed a digital restoration technique for Indian mural paintings in terms of structure and texture, which is much more efficient than manual repainting made by traditional artisans. It disregards the random and irregular characteristics of murals, using cGAN based image translation to develop and update masks automatically after convolution mechanism and to identify the damaged or missing parts. Additionally, it should have a strong inpainting model that can deal with irregularly shaped holes and provide predictions that blend seamlessly with the rest of the image without the need for any additional post-processing or merging. Peak Signal to Noise Ratio, Mean Squared Error, and Structural Similarity Index were among the performance parameters where the suggested technique produced the best reconstruction results. A Pconv-inpainting framework based on deep-learning makes use of various loss functions. Compared to the most recent state-of-the-art approaches, this technology offers more accurate and truthful findings.

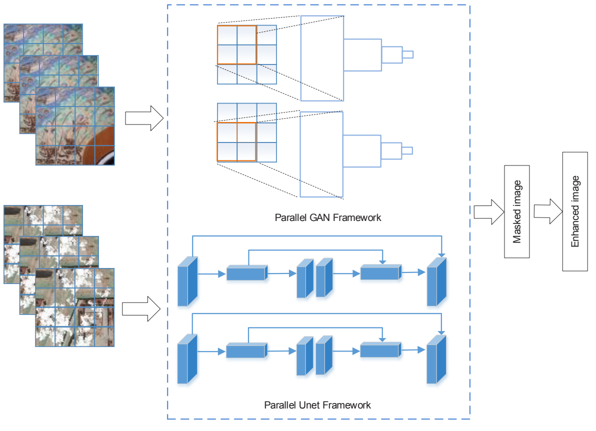

Fig. 2. The proposed multi-scale parallel GAN and Unet structure.

Wang et al. built a multi-scale dense matching repair network based on fine adversarial generative network model for image restoration to solve the problem of painting loss or fading caused by both environmental and manmade events [9]. In order to enhance the repair impact of intricate textures, a dense combination of dilated convolutions is first applied. A multi-scale dense fusion network was introduced in the middle of the convolutional layer in the model's generative network section. Four convolutional layers with varying expansion rates are given the features derived by the convolutional layer, and after that, dilated convolution with different expansion rates is used to extract the multi-scale features. Multiple sparse features are joined through accumulation in order to generate dense Multiscale features. The loss function used to direct the training of the generative network includes mean absolute Error, (Visual Geometry Group, VGG) feature matching, auto-guided regression, and geometric alignment. A self-guided regression constraint loss was developed to compute the feature difference map obtained by the convolutional layer in addition to the frequently used feature matching loss, thus estimating and correcting the semantic layer information. A geometric alignment constraint loss was also created to make up for and forecast the separation between high-level features and actual features. The discriminant network is trained using a discriminator with local and global branches in order that the restored picture appears in both local and global content. By applying this generative adversarial network to restoration, the restored image is both natural in color and cohesive in detail, making it easier to appreciate and display the mural.

We present a multi-scale parallel GAN combined with a U-Net architecture, as shown in Figure 2.. The introduction of multi-scale processing aids in the extraction of features from images across various scales or resolutions, enhancing the model's comprehension of the image content. This approach is particularly effective when the size of the subject matter varies. The integration of parallel U-Nets enriches the diversity of the extracted features, with each U-Net in the parallel configuration potentially learning to extract different feature types, thereby offering a more varied feature set for subsequent tasks. The parallel GAN enhances the model's robustness against overfitting. By training numerous U-Nets in a parallel fashion, the model is better equipped to learn more generalized features, thereby minimizing the likelihood of overfitting to the training data.

4 Experiments and results

To verify the capability of the proposed method, we conduct model training and testing on the following datasets: Dunhuang mural image dataset contains a total of 1926 mural images. All images in this dataset are from the electronic resource album Complete works of Dunhuang murals in China. The mural images are divided into categories of different dynasties including Northern Wei, Northern Zhou, Sui, Tang, Five, and West Wei dynasties. Wutaishan mural image dataset contains 800 mural images with good photographic quality from the temples of Wutaishan, Shanxi Province. After image augmentation and expansion, 12,000 mural images were obtained. 10,000 of these were used as the training set, and the remaining 2000 images were used as the test set. The Dunhuang Grotto Painting dataset is based on the digitization of the northern and southern murals in Cave 7 of Mogao Grottoes, and according to the principle of image content integrity, 600 images of different murals were selected, with a resolution of 500-800 pixels [10-13]. Each image was downscaled to a resolution of 512 x 512 pixels.

We use PSNR (peak signal-to-noise ratio), SSIM (structural similarity), and NIQE (natural image quality evaluator) to evaluate the performance level of various algorithms in mural enhancement [1-4]. PSNR evaluates the quality of an image by comparing the difference between corresponding pixels of two images. The higher the PSNR, the smaller the distortion and the better the super-resolution reconstruction. The PSNR calculation formula is as follows:

$$\mathrm{PSNR} = 10\log_{10}\frac{255^2 \times W \times H}{\sum_{i=1}^{W}\sum_{j=1}^{H}[X(i,j)-Y(i,j)]^2} \tag{1}$$

where W is the width of the image, H is the height of the image, X and Y are the pixel values.

The value of SSIM is between [0, 1], which is an index for evaluating image similarity in terms of brightness, contrast, and structure. The calculated equation is as follows:

$$\mathrm{SSIM}(x,y)=\frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \tag{2}$$

where x and y denote the reconstructed image and the original image respectively, μ is the mean, σ is the variance, σxy is the covariance, and c is the constant.

NIQE analyzes patches of the same size in the image to calculate the same NSS features, fits them with the MVG model, and then compares its MVG fit to the natural MVG model. The quality of the distorted image is expressed as the distance between the quality-aware NSS feature model and the MVG fitted to the features extracted from the distorted image:

$$D(\nu_1,\nu_2,\Sigma_1,\Sigma_2)=\sqrt{\left((\nu_1-\nu_2)^T\left(\frac{\Sigma_1+\Sigma_2}{2}\right)^{-1}(\nu_1-\nu_2)\right)} \tag{3}$$

where ν and Σ are the mean vector and covariance matrix of the natural MVG model and the distorted image MVG model.

We compare RNN, GMCNN, MR-CNN, cGAN with our proposed method, which is designated as parallel GAN, multi-scale GAN, and proposed multi-scale parallel GAN and Unet structure. Area_1, Area_2, and Area_3 respectively represent the restoration of different areas of the mural. Shown in Fig.3 are the PSNR scores of these algorithms. As can be seen from the figure, MR-CNN and cGAN have better PSNR scores in each region, parallel GAN, multi-scale GAN results are close among our three methods, and the final comprehensive method has the best performance.

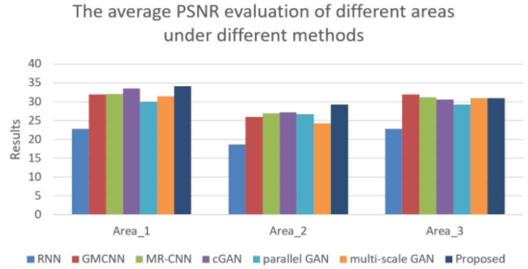

Fig. 3. The average PSNR evaluation of different areas under different methods.

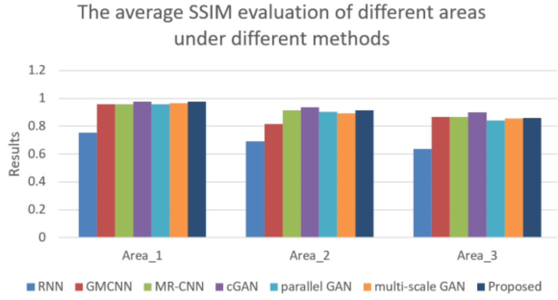

Fig. 4. The average SSIM evaluation of different areas under different methods.

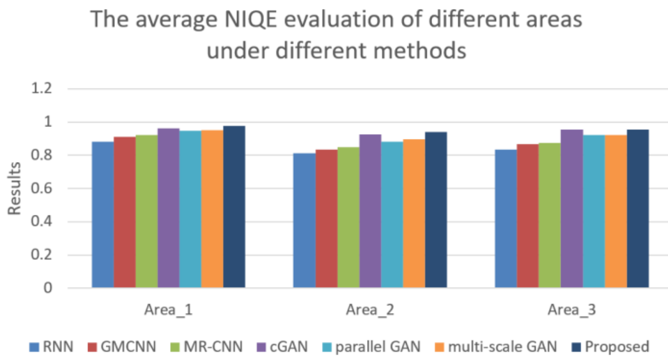

Fig. 5. The average NIQE evaluation of different areas under different methods.

Fig.4 depicts the average SSIM evaluation of different areas under different methods on the dataset. It can be seen from this that in addition to multi-scale parallel GAN and Unet structure, the results of parallel GAN and multi-scale GAN are also better, especially on area1. Fig.5 presents the average NIQE evaluation of different areas under different methods. Again, the results of the proposed structures are still relatively consistent, with the latter three yielding slightly better results.

5 Conclusions

Ancient Chinese murals serve as historical documents that reflect the cultural, social and religious aspects of the period in which they were created. Restoration of these murals provides a better understanding of the historical context, artistic techniques and evolution of cultural practices in the wider region. Our proposed multi-scale parallel GAN and parallel Unet structure can extract features from multiple scales or images, adapt to the changing scale of the target, and provide a more diverse feature set for downstream tasks. The advantage of this structure is that it reduces the risk of overfitting the training data by learning more general features. The verification results on the ancient mural data set preliminarily show that the method has a certain objective index performance improvement.

References

1. Cao J, Zhang Z, Zhao A, et al. Ancient mural restoration based on a modified generative adversarial network[J]. Heritage Science, 2020, 8(1): 1-14.
2. Li J, Wang H, Deng Z, et al. Restoration of non-structural damaged murals in Shenzhen Bao'an based on a generator–discriminator network[J]. Heritage Science, 2021, 9(1): 1-14.
3. Zou Z, Zhao P, Zhao X. Virtual restoration of the colored paintings on weathered beams in the Forbidden City using multiple deep learning algorithms[J]. Advanced Engineering Informatics, 2021, 50: 101421.
4. Wang N, Wang W, Hu W, et al. Damage Sensitive and Original Restoration Driven Thanka Mural Inpainting[C]//Pattern Recognition and Computer Vision: Third Chinese Conference, PRCV 2020, Nanjing, China, October 16–18, 2020, Proceedings, Part I 3. Springer International Publishing, 2020: 142-154.
5. Zhu X, Yu Y, Deng X, et al. Bring Ancient Murals Back to Life[C]//International Conference on Neural Information Processing. Cham: Springer International Publishing, 2022: 231-242.

6. Cao J, Zhang Z, Zhao A. Application of a modified generative adversarial network in the superresolution reconstruction of ancient murals[J]. Computational Intelligence and Neuroscience, 2020, 2020.
7. Wang J, Zhang E, Cui S, et al. GGD-GAN: Gradient-Guided dual-Branch adversarial networks for relic sketch generation[J]. Pattern Recognition, 2023, 141: 109586.
8. Rakhimol V, Maheswari P U. Restoration of ancient temple murals using cGAN and PConv networks[J]. Computers & Graphics, 2022, 109: 100-110.
9. Wang Q, Hou M, Lyu S. Virtual Restoration of Missing Paint Loss of Mural Based on Generative Adversarial Network[J]. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2021, 46: 807-811.
10. Mittal A, Soundararajan R, Bovik A C. Making a "completely blind" image quality analyzer[J]. IEEE Signal processing letters, 2012, 20(3): 209-212.
11. Cao J, Yan M, Chen H, et al. Dynasty recognition algorithm of an adaptive enhancement capsule network for ancient mural images[J]. Heritage Science, 2021, 9: 1-15.
12. Cao J, Zhang Z, Zhao A, et al. Ancient mural restoration based on a modified generative adversarial network[J]. Heritage Science, 2020, 8(1): 1-14.
13. Yu T, Zhang S, Lin C, et al. Dunhuang grottoes painting dataset and benchmark[J]. arXiv preprint arXiv:1907.04589, 2019.