Image Inpainting for Digital Dunhuang Murals Using Partial Convolutions and Sliding Window Method

Ming Chen, Xudong Zhao and Duanqing Xu
Zhejiang University, No.38 Zheda Road, Hangzhou, China
Email: chm@zju.edu.cn
Email: sdu_zxd@163.com
Email: xdq@zju.edu.cn

Abstract. It is a difficult and challenging task to restore the digital murals to a visually pleasant result, and even the result is similar to the original murals without corruption. In this paper, to address the above problem, we propose an image inpainting strategy called PCSW for digital Dunhuang murals using partial convolutions and sliding window method. Specially, a deep neural network based on partial convolutions is used as the underlying model for image inpainting. Because the murals are somewhat damaged or even large areas are missing, in addition, digital murals are large and high resolution, it is unreasonable and impractical to use the original digital murals for training and then restoring the missing areas. Therefore, a data augmentation method based on sliding window technique is applied to increase samples and then improve the model accuracy. Experimental results have shown that the proposed strategy has a certain effect on the restoration of digital Dunhuang murals.

1. Introduction
Dunhuang Grottoes is an important historical and cultural heritage, in which murals are one of the most valuable parts and contains precious Buddhist art. However, the contents of murals have been severely damaged and corrupted caused by the environmental and human factors with the passage of time. Therefore, preserving these precious murals as completely and clearly as possible to allow more people to appreciate and research is an urgent task. Fortunately, with the rapid development of electronic technology, digitizing murals is a viable and effective method to store and preserve these precious materials. After obtaining the high fidelity and resolution digital murals, how to restore the corrupted digital murals in a visually pleasant way even to their original appearance is a difficult and challenging task.

There are two main difficulties in the process of image inpainting for digital Dunhuang murals. On one hand, it is a time-consuming and resource-intensive task to produce a good result whether using traditional image processing methods or deep learning methods. Traditional methods [1-6] are mainly based on the similarity between the missing areas and the known areas. For example, Ghorai et al. [1] have proposed an image inpainting algorithm based on the Markov random field framework, which mainly contains two components, that is patch selection and patch refinement. Especially, the patch selection is performed from groups of similar patches. With respect to deep learning methods [7-13], the accuracy of the model greatly depends on the quality of the training data and the architecture of the network. Liu et al. [10] presented an image inpainting model for irregular holes based on the partial convolutions, which achieves good inpainting results. In our paper, the deep neural network used in the proposed strategy is the same as [10] but used to restore digital murals images. As to image inpainting for digital murals images, in order to make these images clearer, the original images are
large and high resolution, it will consume more time and resources to process each full image in theoretical, however, it is impractical in reality due to the limited resources and computing capabilities. Therefore, it poses great challenge for training the neural network in an efficient way.

On the other hand, there exists small sample size problem when using deep learning techniques. Although there are many murals in the grottoes, most of them are damaged. In other words, part areas of the murals are missing. There are very few complete digital images that can be used directly in the training process when using deep neural networks. Therefore, in order to improve the accuracy of the network, it is necessary to perform reasonable data augmentation operations before the training process when using deep learning based method.

In this paper, an image inpainting strategy called PCSW is proposed to restore the digital Dunhuang murals through using partial convolutions and sliding window method. Specially, a deep neural network based on partial convolutions is applied to restore the digital murals. And then we use sliding window method in the data augmentation process to address the problem of small sample size and improve the model accuracy. Finally, the PCSW strategy based on the above partial convolutions and sliding window method is used for image inpainting. Through using the real digital Dunhuang murals, we have shown the proposed strategy has a certain effect on the image inpainting process.

The main contributions of the paper are as follows:

- An image inpainting strategy called PCSW based on partial convolutions and sliding window method is used for digital Dunhuang mural restoration.
- The partial convolutions based deep neural network is used for image inpainting, and the sliding window based data augmentation process is used to generate enough data to address the small sample size problem and improve the network accuracy.
- By using real digital Dunhuang murals, the effect of the proposed PCSW strategy on image inpainting has been presented.

The remainder of this paper is organized as follows. Section 2 gives the detailed description of the proposed PCSW strategy for image inpainting with respect to digital Dunhuang murals. Section 3 shows the effect on the image inpainting for digital Dunhuang murals of the proposed PCSW strategy. Section 4 is the conclusion of the paper.

2. The Proposed PCSW Strategy

In this section, the proposed PCSW strategy based on partial convolutions and sliding window method is described in detail. We first introduce the deep neural network used in the paper based on the partial convolutions, and then the data augmentation using sliding window method is described. Finally, we present the PCSW strategy for digital murals image inpainting.

2.1. Deep Neural Network Based on Partial Convolutions

Liu et al. [10] firstly proposed the partial convolution and then designed a network similar to UNet-like architecture [14] by stacking partial convolutional layers rather than traditional convolutional layers. Next, the partial convolutional layer and the loss function will be briefly introduced.

The partial convolution operation and the mask update function are two key components of the partial convolutional layer. Assuming that $X$ are the feature values of the digital mural image $I$, $M$ is the binary mask, representing whether there exists missing areas or not. $W$ and $b$ denotes the weights and bias of the convolution filter. Therefore, similar to [15], the partial convolution can be defined as:

$$
x' = \begin{cases} 
W^T (X \ast M) \frac{\text{sum}(1)}{\text{sum}(M)} + b, & \text{if } \text{sum}(M) > 0 \\
0, & \text{otherwise}
\end{cases} \quad (1)
$$

where $\ast$ represents element-wise multiplication, the factor of $\text{sum}(1)/\text{sum}(M)$ is used to adjust the amount of unmasked inputs, it also means that the output of the network depends on the unmasked values. $x'$ is the updated feature values.

After the partial convolution operation, the mask $M$ can be updated as follows:
\[ m' = \begin{cases} 
1, & \text{if } \text{sum}(M) > 0 \\
0, & \text{otherwise}
\end{cases} \] (2)

where \( m' \) is the updated mask.

In our paper, the architecture of the used deep neural network is the same as the Liu et al. [10], in which both the encoder and decoder consists of eight partial convolutional layers. According to [10], the loss function is:

\[
L_{\text{total}} = L_{\text{valid}} + 6L_{\text{hole}} + 0.05L_{\text{perceptual}} + 120(L_{\text{style}} + L_{\text{comp}}) + 0.1L_{\text{tv}}
\] (3)

where \( L_{\text{total}} \) is the final loss weights, \( L_{\text{valid}} \) and \( L_{\text{hole}} \) is related to per-pixel of the image, \( L_{\text{perceptual}} \), \( L_{\text{style}} \) and \( L_{\text{comp}} \) are used to ensure that the missing areas are restored in a natural way. \( L_{\text{tv}} \) is the total variation loss and used for smoothing penalty [16]. Details of these loss functions can be found in the Liu et al. [10].

2.2. Data Augmentation Using Sliding Window Method

As described above, there exists small sample size problem when using the original digital murals images to train the deep neural network. Although the original digital murals are high fidelity and high resolution, most of them have missing areas. In addition, the images obtained only by traditional data augmentation operations such as rotation and scaling are still large and high resolution, it is also unsuitable to directly use these augmented images to train the network. In order to generate enough data in a reasonable way, the sliding window method is used in the data augmentation process. The main idea is that each sliding window is considered as an image patch, and then the stride is less than the size of image patch. In this way, the image content continuity can be ensured while reducing the image size.

Assuming that the size of the original digital mural image \( I \) is \( w \times h \), and the size of sliding window (image patch) is \( p \times q \) (\( p < w, q < h \)). Since the sliding window can slide from both horizontal and vertical directions and in order to guarantee the content continuity of the image, the stride of both directions can be set as \( s \) \((s < p)\) and \( t \) \((t < q)\), respectively. The diagram of the sliding window method used in the original digital mural image is represented as Figure 1.

![Figure 1. The diagram of sliding window method used in the original digital mural image](image-url)
As we can see from the Figure 1, there may exist some residual areas as using sliding window method if the $p, q, s, t$ are set unsuitable. Assuming the residual length in horizontal and vertical directions is $m$ and $n$, respectively. In other words, when $m < s$ or $n < t$, the stride is set to $m$ or $n$ but keep the size of sliding window (image patch) unchanged.

After using the sliding window method, the original large size and high resolution digital mural image can be divided into many small image patches. However, there may be some patches with a wide range of missing areas, in that case, these patches should be selected and removed from the training set, because there is no ground truth about the missing areas of these patches and excessive missing areas can greatly affect the accuracy of the network. In order to obtain more image samples, the traditional data augmentation methods, including rotation, scaling and translation, can be used on suitable image patches. Through using the above data augmentation process, enough data with relatively high quality can be obtained and then used in the training process.

2.3. PCSW Strategy for Image Inpainting
According to the above description, the PCSW strategy for restoring digital mural images based on the partial convolutions and sliding window method can be described as follows:

- Building the neural network based on the partial convolutions according to Liu et al. [10].
- Using ImageNet dataset [17] to pre-train the partial convolutional layers and then use the obtained weights to initialize the parameters of the used network rather than random initialization parameters.
- Applying the sliding window method to handle the original large digital mural images and obtain enough image patches as the augmented dataset to address the small sample size problem according to the Figure 1.
- Using the augmented dataset to train the network, in this training process, the size and shape of the masks are automated generated. In other words, the network is fine-tuned by using the augmented dataset. Because the parameters including weights and biases are initialized by pre-training on the ImageNet dataset, it will greatly decrease the time consuming in this training process.
- After the training process, the network can be used to restore the image patches with irregular or regular missing areas.

Through using the above PCSW strategy on the GPU architecture, the network can be trained in a relatively fast way.

3. Experimental Results
In this section, by using the real digital Dunhuang murals images, the effect of the proposed PCSW strategy is presented through comparing the experimental results with ground truth. The image inpainting results for irregular missing areas are described.

Figure 2 shows the experimental results for image inpainting with the irregular missing areas. In these experiments, four image patches from four different original digital murals are used to compare the experimental results with the ground truth, and each image patch is $500 \times 500$.

Figure 2(a) shows the image with irregular missing areas. Because automatically determining the missing area is not the research point of the paper, in the experiments, the irregular missing areas are automatically generated randomly by programs. In addition, the same shape of the missing area is used in four experiments.

Figure 2(b) and 2(c) shows the inpainting results using the PCSW strategy and the ground truth of the image patches. As we can see from the figure, through comparing with the inpainting results with the ground truth of the corresponding image patch, the PCSW strategy has a certain effect on image inpainting for digital murals with irregular missing areas. In other words, the proposed strategy can restore the missing areas of digital Dunhuang murals in a visually pleasant way. However, in the restored missing areas, the corresponding area is blurred and even the visible traces of the missing areas can be seen. It indicates that the proposed PCSW strategy needs to be further optimized to improve the accuracy of image inpainting.
Figure 2. Experimental results for image inpainting with irregular missing areas.
4. Conclusion

Image inpainting for digital Dunhuang murals is a difficult and challenging task due to the large size of the original image and the small sample size problem. In the paper, we propose an image inpainting strategy called PCSW for restoring the digital Dunhuang murals based on the partial convolutions and sliding window method. Specially, the partial convolutions based deep neural network is used as the underlying model for image inpainting, and the data augmentation process applying sliding window method is used to address the small sample size problem in the training process and then improve the network accuracy. Finally, experimental results on real digital Dunhuang murals have shown the efficiency of the proposed strategy.

5. References

[1] Ghorai M, Mandal S, Chanda B 2018 A Group-Based Image Inpainting Using Patch Refinement in MRF Framework. *IEEE Trans. Image Processing* 27(2): 556-567.

[2] Lu H, Liu Q, Zhang M, Wang Y, Deng X 2018 Gradient-based low rank method and its application in image inpainting. *Multimedia Tools Appl.* 77(5): 5969-5993.

[3] Criminisi A, Pérez P, Toyama K 2004 Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. Image Processing* 13(9): 1200-1212.

[4] He K, Sun J 2014 Image Completion Approaches Using the Statistics of Similar Patches. *IEEE Trans. Pattern Anal. Mach. Intell.* 36(12): 2423-2435.

[5] Barnes C, Shechtman E, Finkelstein A, Goldman D B 2009 PatchMatch: a randomized correspondence algorithm for structural image editing. *ACM Trans. Graph.* 28(3): 24:1-24:11.

[6] Hays J, Efros A A 2007 Scene completion using millions of photographs. *ACM Trans. Graph.* 26(3): 4.

[7] Yang C, Lu X, Lin Z, Shechtman E, Wang O and Li H 2017 High-Resolution Image Inpainting Using Multi-scale Neural Patch Synthesis. *CVPR 2017*: 4076-4084.

[8] Demir U, Ünal G B 2017 Inpainting by deep autoencoders using an advisor network. *SIU 2017*: 1-4.

[9] Yan Z, Li X, Li M, Zuo W and Shan S 2018 Shift-Net: Image Inpainting via Deep Feature Rearrangement. *ECCV (14) 2018*: 3-19.

[10] Liu G, Reda F A, Shih K J, Wang T, Tao A and Catanzaro B 2018 Image Inpainting for Irregular Holes Using Partial Convolutions. *ECCV (11) 2018*: 89-105.

[11] Yeh R A, Chen C, Lim T, Schwing A G, Johnson M H and Do M N 2017 Semantic Image Inpainting with Deep Generative Models. *CVPR 2017*: 6882-6890.

[12] Wang Y, Tao X, Qi X, Shen X and Jia J 2018 Image Inpainting via Generative Multi-column Convolutional Neural Networks. *NeurIPS 2018*: 329-338.

[13] Liu H, Lu G, Bi X, Yan J and Wang W 2018 Image Inpainting Based on Generative Adversarial Networks. *ICNC-FSKD 2018*: 373-378.

[14] Ronneberger O, Fischer P and Brox T 2015 U-Net: Convolutional Networks for Biomedical Image Segmentation. *MICCAI (3) 2015*: 234-241.

[15] Harley A W, Derpanis K G, Kokkinos I 2017 Segmentation-Aware Convolutional Networks Using Local Attention Masks. *ICCV 2017*: 5048-5057.

[16] Johnson J, Alahi A and Li F 2016 Perceptual Losses for Real-Time Style Transfer and Super-Resolution. *ECCV (2) 2016*: 694-711.

[17] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M S, Berg A C and Li F 2015 ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision 115(3)*: 211-252.