Beijing University of Technology, Beijing, China

GENERATIVE IMAGE INPAINTING BASED ON WAVELET TRANSFORM ATTENTION MODEL

Chen Wang, Jin Wang, Qing Zhu, Baocai Yin

S201761817@emails.bjut.edu.cn, {ijinwang,ccgszq,ybc}@bjut.edu.cn

Introduction

Image inpainting is a challenging task in image processing and widely applied in many areas such as photo editing. Traditional patch-based methods are not effective to deal with complex or non-repetitive structures. Recently, deep learning-based approaches have shown promising results for image inpainting. However, they usually generate contents with artificial boundaries, distorted structures or blurry textures. To handle this problem, we propose a novel image inpainting method based on wavelet transform attention model (WTAM). The wavelet transform decomposes features into multi-frequency subbands for extracting and transmitting deep information, and the attention mechanism enhances the ability of wavelet transform to capture significant detailed information in each level's subband images. Extensive experimental results on multiple datasets demonstrate that our method can not only synthesize sharp image structures but also generate fine-detailed textures in missing regions, significantly outperforming the state-of-the-art methods.

Proposed Method

Network Architecture

Considering that long skip connections are limited by the semantic consistency of feature maps, we adopt a redesigned encoder-decoder network of UNet++ as the baseline network, where sub-networks are connected through a series of nested and dense skip pathways. Essentially, it brings the semantic level of the encoder features closer to that in the decoder, which makes optimizing learning tasks easier for recovering the information of missing regions.

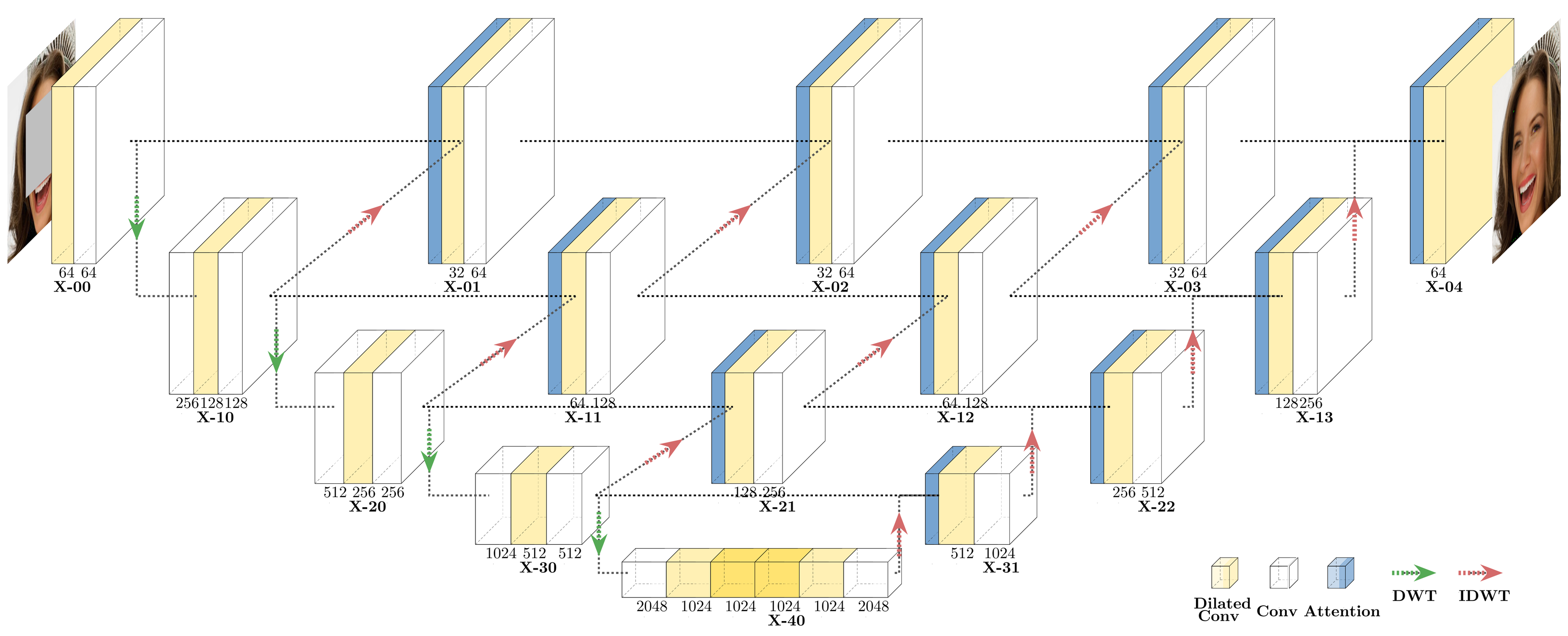

Figura 1: Overview of our architecture with wavelet transform attention model.

Wavelet Transform Attention Model

To extract the deep information from the known regions, we utilize the Wavelet Packet Transform, where multi-frequency information can be represented by a sequence of wavelet coefficients. Moreover, both frequency and location information of feature maps are captured by Discrete Wavelet Transformation (DWT), which is helpful for capturing both global structure information and local texture details when using multi-frequency feature representation. Meanwhile, in order to specifically capture the detailed information of each level's subband images, we design an attention mechanism in wavelet domain, which sequentially generates a wavelet attention map by exploiting the inter relationship of multi-frequency subband. As each subband of the wavelet transform can be considered as a feature detector, attention mechanism focuses on 'what' is meaningful of known regions and highlights salient features.

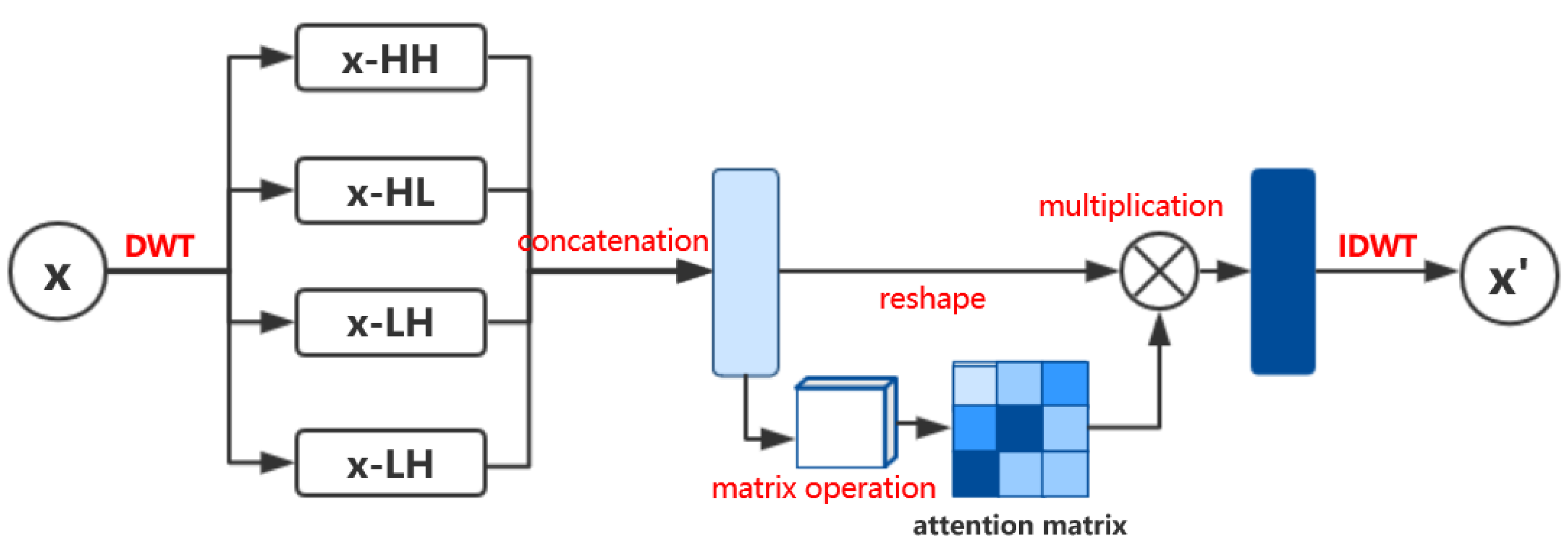

Figura 2: The process of wavelet transform attention model.

Results

Our model can be trained for both center and arbitrary region completion. As shown in Figure 3, our model either based on U-net or UNet++ can achieve better and generate visually more realistic results. Moreover, Figure 4 shows the random mask results of Shift-Net [2] and our model based on UNet++. For textured and smooth regions, both methods perform favorably. While for structural regions, our model is more effective.

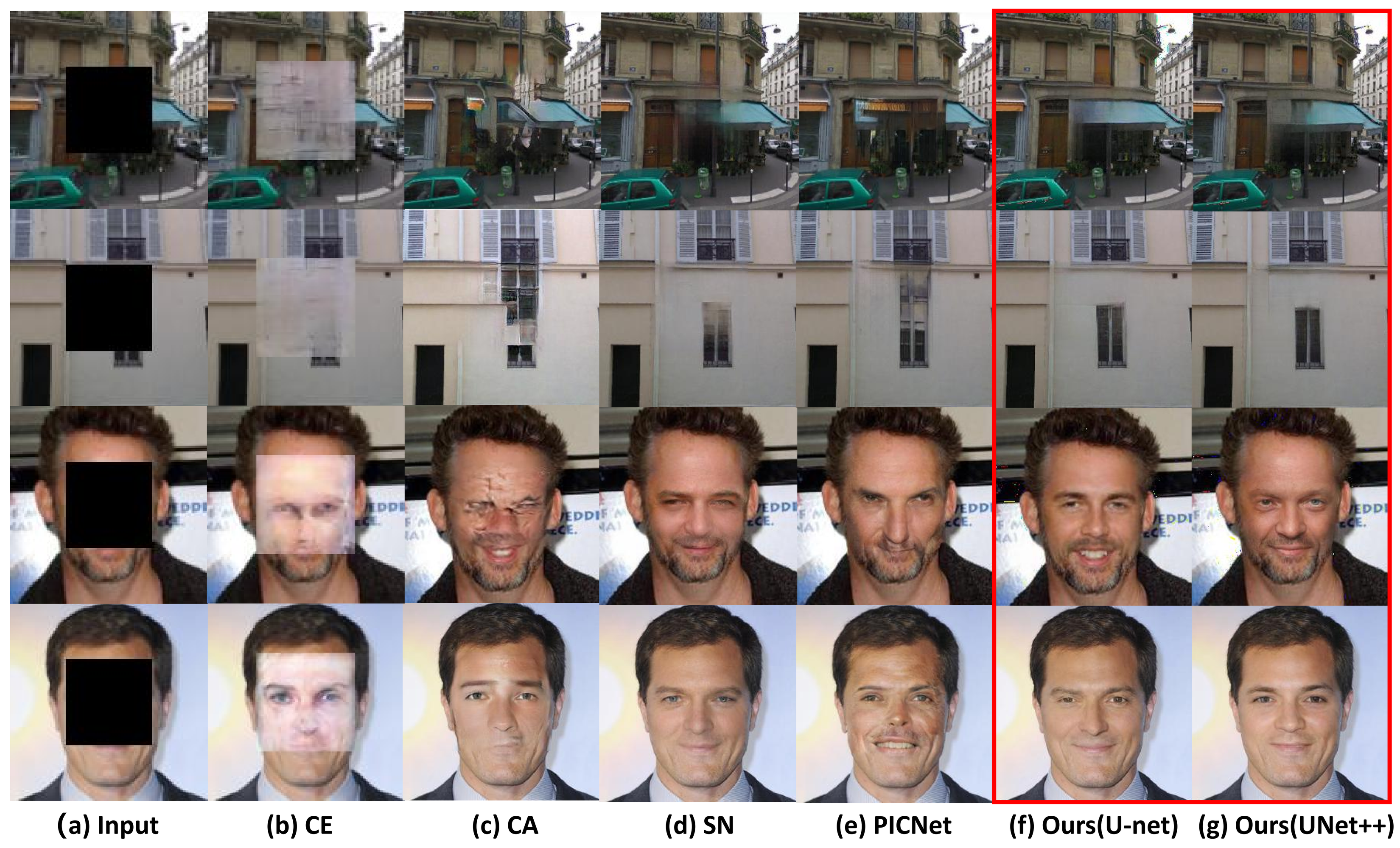

Figura 3: Comparison of visual results with Context Encoder [1], Contextual Attention [3], Shift-Net [2] and PICNet [4] on images taken from Paris Street-View, CelebA and CelebAMask-HQ for center region completion.

Figura 4: Random mask completion results comparison on CelebAMask-HQ and Paris StreetView. From top rows to bottom rows are: input, Shift-Net and Ours.

Conclusion

In this paper, we proposed a novel image inpainting method based on wavelet transform attention model. Compared to existing methods, the proposed method enhances the ability of wavelet transform to capture significant multi-frequency information, thus it obtains images with sharper structures and fine-detailed textures. Extensive experimental results demonstrate its superior performance on the challenging image inpainting tasks.

Referências

[1] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros. Context encoders: Feature learning by inpainting. In *IEEE Conference on Computer Vision Pattern Recognition*, 2016.

[2] Z. Yan, X. Li, L. Mu, W. Zuo, and S. Shan. Shift-net: Image inpainting via deep feature rearrangement. 2018.

[3] J. Yu, L. Zhe, J. Yang, X. Shen, and T. S. Huang. Generative image inpainting with contextual attention. 2018.

[4] C. Zheng, T. J. Cham, and J. Cai. Pluralistic image completion. 2019.