Image Inpainting Based on Multi-frequency Probabilistic Inference Model

Jin Wang *1 Chen Wang *1 Qingming Huang [2] Yunhui Shi [1] Jian-Feng Cai [3] Qing Zhu [1] Baocai Yin [1]

[1]Beijing University of Technology [2]University of Chinese Academy of Sciences [3]Hong Kong University of Science and Technology

Motivation

Image inpainting methods usually fail to reconstruct reasonable structure and fine-grained texture simultaneously. This paper handles this problem from a novel perspective of predicting low-frequency semantic structural contents and high-frequency detailed textures respectively, and proposes a multi-frequency probabilistic inference model(MPI model) to predict the multi-frequency information of missing regions by estimating the parametric distribution of multi-frequency features over the corresponding latent spaces.

Methodology

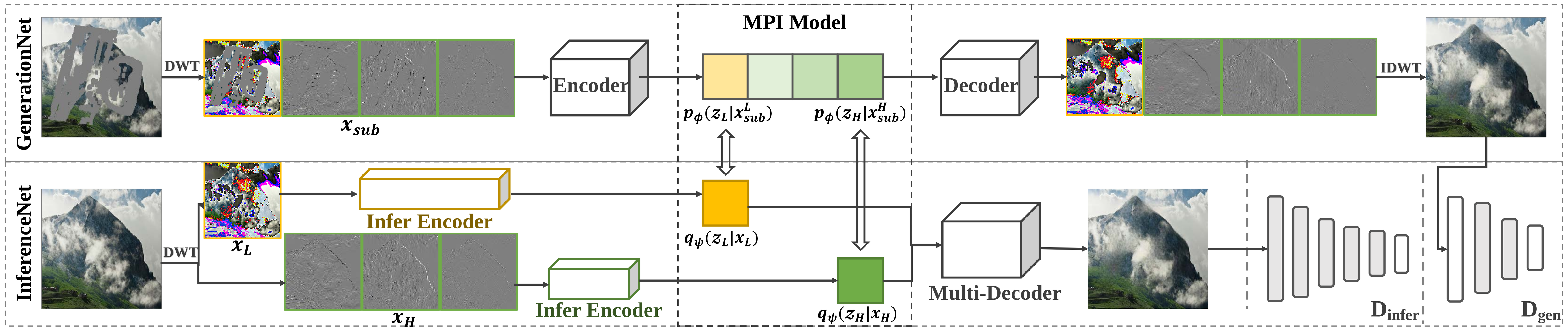

Figure 1. Overall architecture of our method based on the MPI Model.

Network Architecture

The wavelet transform is firstly utilized to decompose the input image into low-frequency and high-frequency subbands. Furthermore, an MPI model is designed to estimate the underlying multi-frequency distribution of input images. Finally, based on the MPI model, the InferenceNet predicts the potential multi-frequency distribution of ground truth, and the GenerationNet generates the final visually realistic images sampling from the distribution of the low-frequency features and high-frequency features estimated by the InferenceNet.

Multi-frequency Probabilistic Inference Model

Firstly, the variational lower bound of the conditional log-likelihood of the observed training instances can be written as follows:

$$\log p(x_L|x^L_{sub}) \geq -KL(q_\psi(z_L|x^L_{sub}, x_L) \,\|\, p_\phi(z_L|x^L_{sub})) + \mathbb{E}_{q_\psi(z_L|x^L_{sub}, x_L)}[\log p_\theta(x_L|z_L, x^L_{sub})] \quad (1)$$

$$\log p(x_H|x^H_{sub}) \geq -KL(q_\psi(z_H|x^H_{sub}, x_H) \,\|\, p_\phi(z_H|x^H_{sub})) + \mathbb{E}_{q_\psi(z_H|x^H_{sub}, x_H)}[\log p_\theta(x_H|z_H, x^H_{sub})] \quad (2)$$

Next, an alternative CVAE variant assumes that conditional prior is independent of the I_m and fixed, which means

$$\log p(x_L) \geq -KL(q_\psi(z_L|x_L) \,\|\, p(z_L)) + \mathbb{E}_{q_\psi(z_L|x_L)}[\log p_\theta(x_L|z_L)] \quad (3)$$

$$\log p(x_H) \geq -KL(q_\psi(z_H|x_H) \,\|\, p(z_H)) + \mathbb{E}_{q_\psi(z_H|x_H)}[\log p_\theta(x_H|z_H)] \quad (4)$$

Finally, our framework defines all unconditional multi-frequency subbands of ground truth I_{gt} as sharing a common latent space with adaptive priors of the corresponding frequency in corrupted images I_m.

$$\log p(x_L|x^L_{sub}) \geq -KL(q_\psi(z_L|x_L) \,\|\, p_\phi(z_L|x^L_{sub})) + \mathbb{E}_{q_\psi(z_L|x_L)}[\log p_\theta(x_L|z_L, x^L_{sub})] \quad (5)$$

$$\log p(x_H|x^H_{sub}) \geq -KL(q_\psi(z_H|x_H) \,\|\, p_\phi(z_H|x^H_{sub})) + \mathbb{E}_{q_\psi(z_H|x_H)}[\log p_\theta(x_H|z_H, x^H_{sub})] \quad (6)$$

Results

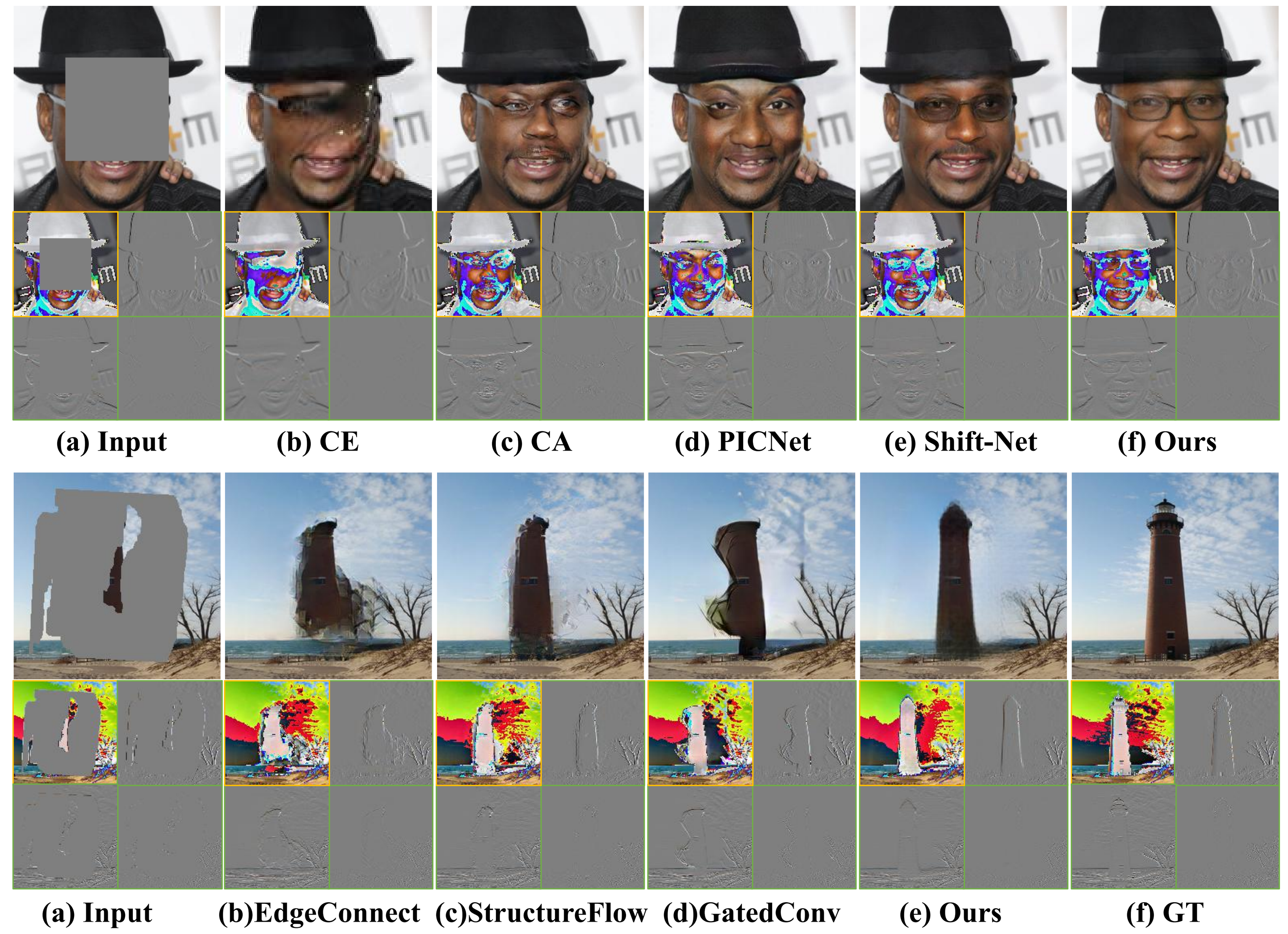

Figure 2. Qualitative comparisons with both centering and irregular holes on CelebAMask-HQ and Places2. Even rows are the visualization of multi-frequency subbands.

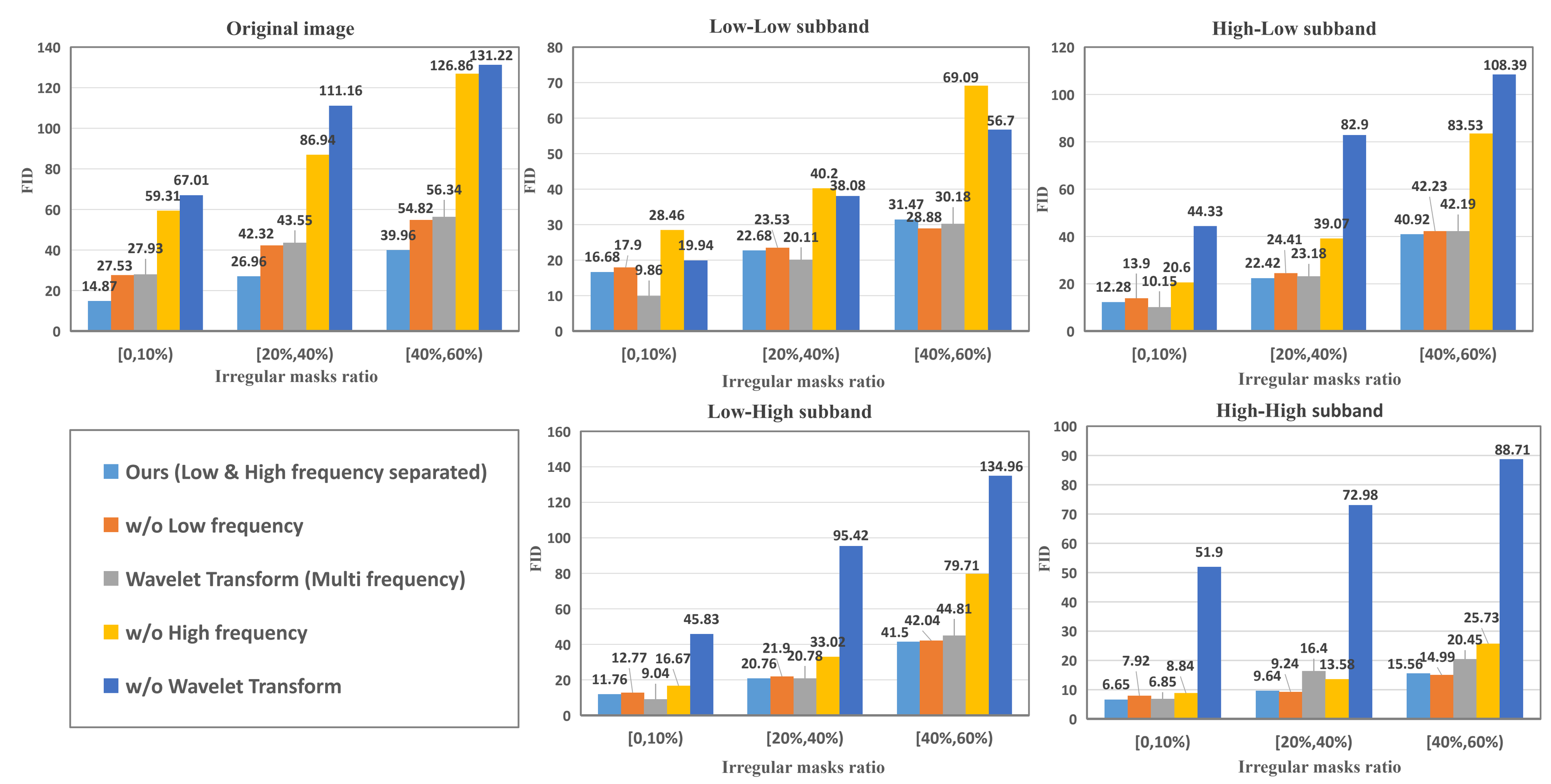

Figure 3. The evaluation result of ablation studies, which are based on irregular masks over CelebA using quantitative measures of FID.

References

[1] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016.

[2] Yurui Ren, Xiaoming Yu, Ruonan Zhang, Thomas H Li, Shan Liu, and Ge Li. Structureflow: Image inpainting via structure-aware appearance flow. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 181–190, 2019.

[3] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5505–5514, 2018.

[4] Chuanxia Zheng, Tat-Jen Cham, and Jianfei Cai. Pluralistic image completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1438–1447, 2019.