Prior Based Human Completion
Supplemental Material

Zibo Zhao[1] Wen Liu[1] Yanyu Xu[2] Xianing Chen[1] Weixin Luo[1] Lei Jin[1]
Bohui Zhu[4] Tong Liu[5] Binqiang Zhao[4] Shenghua Gao[†1,3]

[1]ShanghaiTech University
[2]Institute of High Performance Computing, A*STAR
[3]Shanghai Engineering Research Center of Intelligent Vision and Imaging
[4]Alibaba Group [5] Taobao

{zhaozb,liuwen,chenxn1,luowx,jinlei,gaoshh}@shanghaitech.edu.cn yingmu@taobao.com
xu_yanyu@ihpc.a-star.edu.sg {bhiui.zbh,binqiang.zhao}@alibaba-inc.com

1. Overview

In this supplemental material, we show the detailed **Architecture of Prior Encoding Module** in section 2 and more **Experiment Results** in section 3.

2. Architecture of Prior Encoding Module

We illustrate the architecture of prior encoding module in Figure 1. There are some details to be noted:

1. Prior encoding module takes the concatenation of I_{gt} and S_{gt} as input, where the ground-truth image $I_{gt} \in \mathcal{R}^{3\times256\times256}$, the ground-truth segmentation map $S_{gt} \in \mathcal{R}^{C\times256\times256}$.

2. S_{gt} converts to a one-hot matrix with nine channels ($C = 9$) for the LIP dataset and 12 channels ($C = 12$) for the ChitopiaPlus dataset.

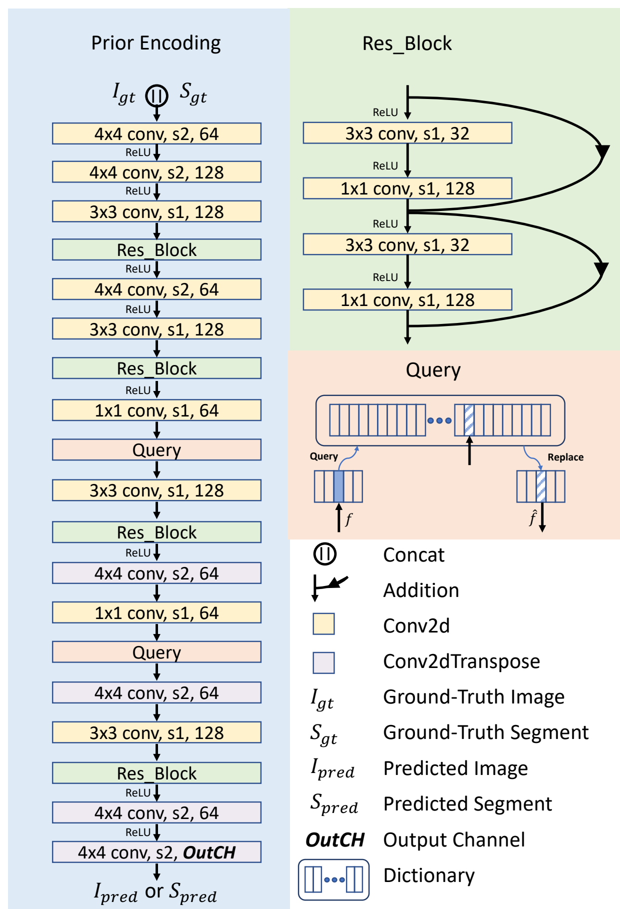

Figure 1. **Network Architecture of Prior Encoding Module**. The value of ***OutCH*** depends on the output. Taking maintaining memory banks on the LIP dataset as an example, the value of ***OutCH*** is three for training texture memory banks and 9 for training structure memory banks.

3. Experiment Results

In this section, we first compare our method with other prior based methods. Then we show more visual results of our method.

3.1. Comparison to other prior based methods.

Baseline. Some colleges have tried to leverage priors to repair corrupted images in recent years. DIP [3] proposes that the structure of a convolutional neural network could capture the textural prior in images, and it can recover the corrupted image by fine-tuning a randomly initialized

† Corresponding author

model. Recently, another prior based method, DGP [2], further claims that are leveraging richer image priors from the large scale dataset, ImageNet [1], to improve the model further. We compare our method with DIP and DGP.

Metric and Dataset. We conduct experiments on the LIP dataset. To evaluate the performance of each method, we use two common metrics: Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM). The PSNR and SSIM evaluate the quality of the generated images. For all metrics, high values mean better performance.

Quantitative comparison. We illustrate some quantitative results in Figure 2. From Figure 2, we can see that although the prior based methods could generate images with a smooth texture, like boy-images produced by both methods and the skier-image produced by DGP, the structure of images are weird. In comparison, our method could generate images with both plausible structure and texture while benefiting from the semantic prior and the structure-texture correlation. It indicates that our method could better encode

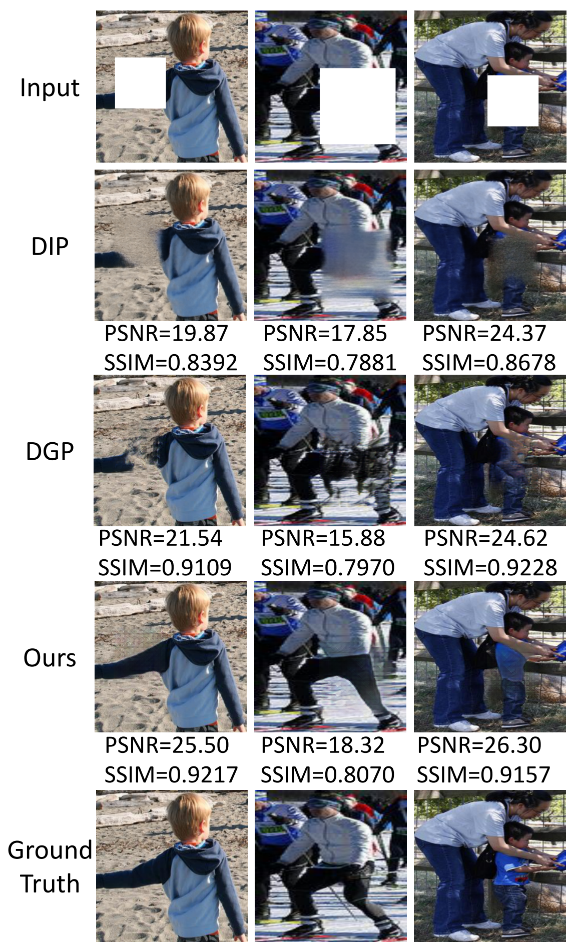

Figure 2. **Qualitative Analysis**. Each column illustrates the input image, outputs of methods, and ground-truth from left to right. The difference between generated images shows that only our method could generate images with smooth texture and reasonable structures, which proves that our model's performance exceeds the other methods.

human body priors.

3.2. More Visual Results

We show more visual results produced by our method in Figure 3 to illustrate its ability better.

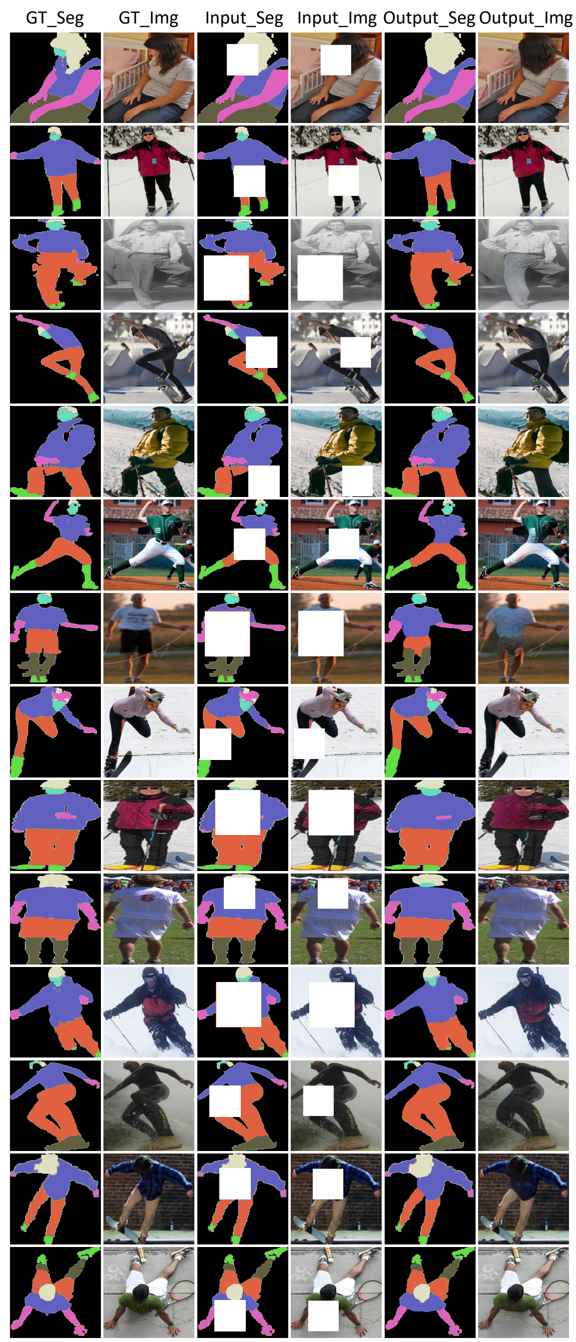

Figure 3. In this figure, we show more visual results produced by our method. Each row illustrates ground-truth segmentation maps, ground-truth images, input segmentation maps, input images, output segmentation maps and output images from left to right. Best viewed with zoom-in.

References

[1] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 2

[2] Xingang Pan, Xiaohang Zhan, Bo Dai, Dahua Lin, Chen Change Loy, and Ping Luo. Exploiting deep generative prior for versatile image restoration and manipulation. *European Conference on Computer Vision (ECCV)*, 2020. 2

[3] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018. 1