

# Supplementary Material of Diverse Plausible 360-Degree Image Outpainting for Efficient 3DCG Background Creation

## 1. Additional Results

In this section, we show figures of results that could not be included in the main body of the paper due to space constraints. Each result image is a higher-resolution image and is best viewed in color and enlarged on a display. Fig. 12 shows diverse completion results from our method, corresponding to Fig. 6. Furthermore, Fig. 13, 14 and 15 show comparisons with previous methods, corresponding to Fig. 7, 8 and 9.

## 2. Additional Analyses

### 2.1. Generated Textures

Some of the generated results in Fig. 2 and Fig. 12 have sky pixels generated at positions corresponding to the lower poles. This is due to the training image rather than the proposed method. SUN360 padded the pole with sky pixels or marks to hide the camera equipment that took the 360-degree images. A similar appearance can be seen in Fig. 5 of Akimoto *et al.* [2].

In this paper, we reported equirectangular images. However, because of the inherent distortions of the equirectangular projection, it may be challenging to evaluate generated results intuitively. Therefore, we display equirectangular images as perspective images in a supplementary video. Please refer to the video for reprojection results.

### 2.2. Training AdjustmentNet

Fig. 16 illustrates the training pipeline of AdjustmentNet. See Section 3.2 for a detailed explanation.

### 2.3. Different FOV inputs

The “Ours” in Fig. 7 and Fig. 8 are the results of the same trained model but with different FOVs. Furthermore, we show Fig. 17 to discuss the effect of different FOVs on the results. We perform training at  $180^\circ \times 90^\circ$  and inference at multiple FOVs. Fig. 17 shows that the proposed method is robust to differences in the size of the input region. Furthermore, when the input area is small, the color of the generated texture tends to be lighter. This may be

because the input area has less texture and therefore cannot be referenced, resulting in a lack of clarity in the generated texture.

### 2.4. Quantity evaluation for diversity

One of the diversity metrics is the average LPIPS distance, following Wan *et al.* [30]. The score of our proposed method is 0.34 ( $180^\circ \times 90^\circ$  input (mask ratio 75%), 1K inputs, 5 samples per an input, VGG network). We report this score for subsequent works. We cannot evaluate some of the previous methods due to unpublished codes. However, we believe that this quantitative comparison does not change the conclusions of our paper, since 360IC and EnvMapNet are deterministic and SIG-SS is limited only to the type of symmetry.

## 3. Supplementary Video

We publish a video as supplementary material. In this video, we show our diverse completion results and further compare them quantitatively with the results obtained by previous methods through demonstrations. Using the pipeline proposed in Section 5, we demonstrate the camera rotating around virtual objects and also demonstrate a virtual object approaching the camera. Since the original output of EnvMapNet [5] is an HDR image, but the resulting image is not publicly available, we use the same inverse tone mapping method [4] as ours to convert the publicly available LDR images into HDR images for the demonstration. We render each video clip by path tracing. We also perform spherical visualization for some of the comparisons. For the visualization, we use Basic app RICOH THETA [1]. At the end of the video, we show a demo where we use one completion result to produce the background and ground, and a virtual actor runs on it.

## References

- [1] Basic app RICOH THETA, 2021. 1
- [2] Naofumi Akimoto, Seito Kasai, Masaki Hayashi, and Yoshimitsu Aoki. 360-degree image completion by two-stage condi-



Figure 12. Diverse outputs. Our method outputs multiple and diverse results for a single input. Best viewed in color and zoomed in.

- tional gans. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 4704–4708. IEEE, 2019. 1, 3
- [3] Takayuki Hara, Yusuke Mukuta, and Tatsuya Harada. Spherical image generation from a single image by considering scene symmetry. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021. 4
- [4] Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Single-image hdr reconstruction by learning to reverse the camera pipeline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1651–1660, 2020. 1
- [5] Gowri Somanath and Daniel Kurz. Hdr environment map estimation for real-time augmented reality. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11298–11306, 2021. 1, 5



Figure 13. Qualitative comparison with 360IC [2]. Best viewed in color and zoomed in.



Figure 14. Qualitative comparison with SIG-SS [3]. (a) indicates the input image, (b) indicates SIG-SS(rec), (c) indicates SIG-SS(gen), and (d) indicates ours. Best viewed in color and zoomed in.

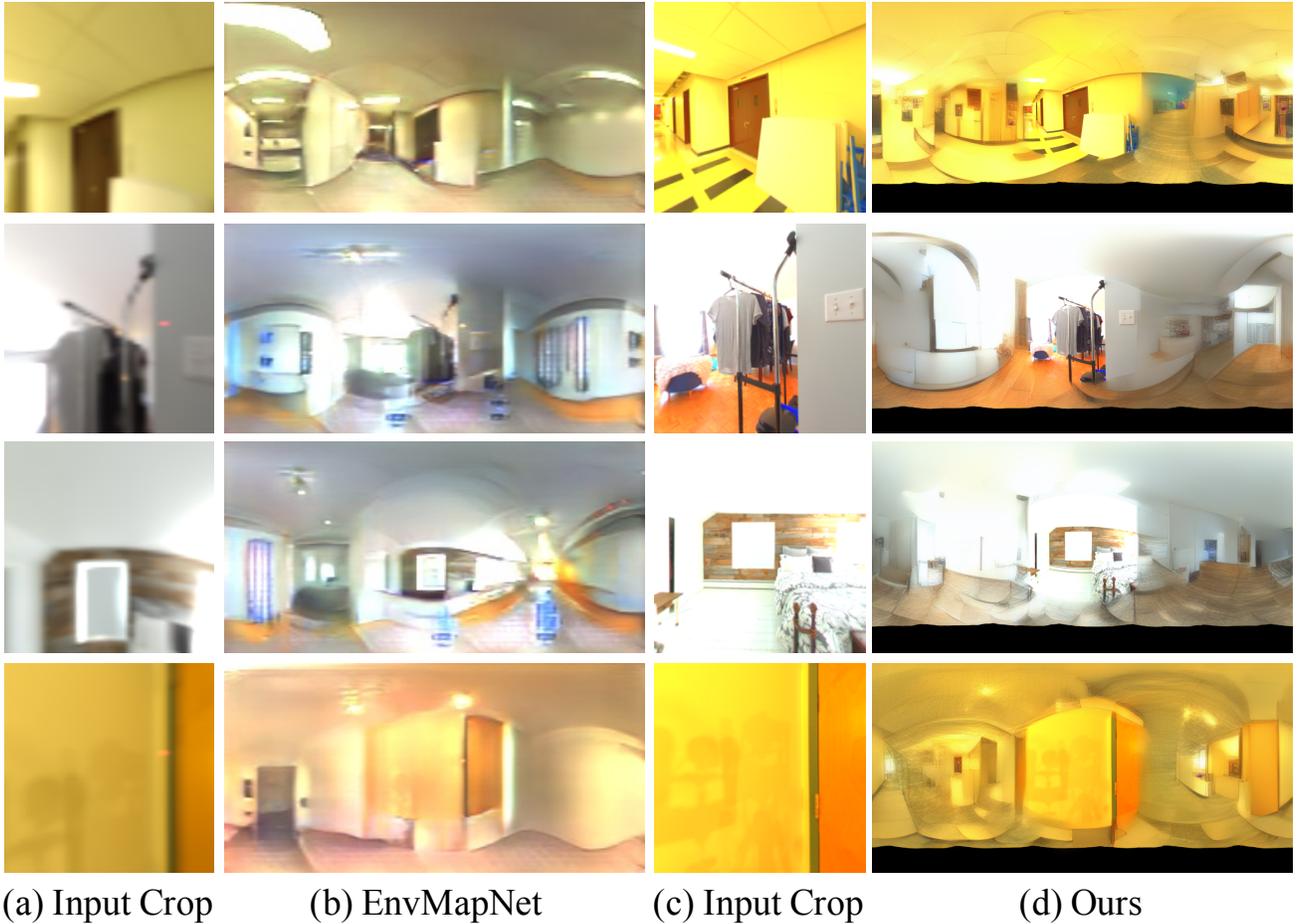


Figure 15. Qualitative comparison with EnvMapNet [5]. We show completion results along with the corresponding input images. Best viewed in color and zoomed in.

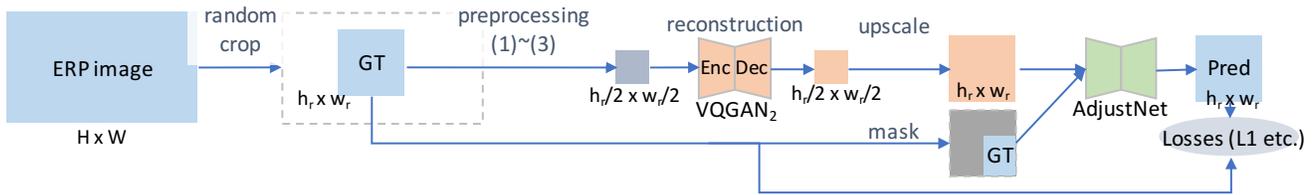


Figure 16. Training pipeline for AdjustNet. Trained VQGAN<sub>2</sub> performs reconstruction, not completion.



Figure 17. Results with different FOV inputs (red boxes). Left to right:  $180^\circ \times 90^\circ$  input,  $90^\circ \times 90^\circ$  input, and  $70^\circ \times 70^\circ$  input.