Supplementary Material
360PanT: Training-Free Text-Driven 360-Degree Panorama-to-Panorama Translation

Hai Wang* and Jing-Hao Xue
University College London
{hai.wang.22, jinghao.xue}@ucl.ac.uk

1. Introduction

This supplementary material begins by providing an intuitive explanation for the choice of α. Subsequently, we detail the process of producing target prompts for both real-world and synthesized datasets. Further visual results obtained under different control conditions are then presented. Finally, we showcase additional translated results from using different methods on real-world and synthesized 360-degree panoramic images.

Explanation for the choice of α. To intuitively explain the choice of the split constant α, Figure 1 visually depicts the cropping process in 360PanT at denoising step t (where $t \in \{T, T-1, \cdots, 1\}$) for three distinct α values. The top row displays the input 360-degree panorama I_{in} and a diagram of the cropping operations based on the sliding window mechanism employed in the seamless tiling translation with spatial control. Each cropped patch, including the special cropped patch (stitch patch), then undergoes independent denoising guided by a target prompt. Subsequent rows highlight the cropped patches matching I_{in} during the sliding window process, indicated by red or yellow dashed boxes. Observe that when $\alpha = W$ or $\alpha = \frac{W}{2}$, two cropped patches matching I_{in} but in different locations are denoised at each step t. Conversely, when $\alpha = \frac{3W}{4}$, only a single cropped patch matching I_{in} undergoes denoising at each step. Crucially, the continuity of boundaries of these highlighted patches are not considered during denoising. Consequently, at each denoising step t, the fewer cropped patches matching I_{in} are denoised, the better the boundary continuity of the final translated 360-degree panorama. Therefore, we set α to $\frac{3W}{4}$ in this paper, which results in a final translated 360-degree panorama with seamlessly connected boundaries, effectively avoiding local visible cracks.

Generation process of target prompts. Figure 2 illustrates the target prompt generation process for each real-world 360-degree panorama within the *360PanoI-Pan2Pan* dataset. Utilizing a consistent template, "a photo of {image name}", an original prompt is constructed for each 360-degree panoramic image. Subsequently, a target prompt is formulated by combining a randomly selected translation type with the original prompt. Figure 3 depicts the analogous process for the *360syn-Pan2Pan* dataset comprising synthesized 360-degree panoramas. Initially, 120 synthesized 360-degree panoramas are generated using a text-to-360-degree panorama model [6] guided by 120 original prompts. Similar to the real-world dataset, each target prompt consists of a randomly chosen translation type and its corresponding original prompt.

Translation using other control conditions. Diverse control conditions are extracted from corresponding 360-degree panoramic images using the methods described in FreeControl [3]. If a control condition lacks continuous boundaries, the translated result by our 360PanT (F) will exhibit noticeable content inconsistency at the boundaries. For instance, Figure 4 illustrates how using an extracted depth map I_{in} with discontinuous boundaries as input leads to visible cracks in the extended input map $\hat{I}_{in}$. Consequently, the translated image by 360PanT (F) shows content inconsistency in the stitched area. In contrast, we observe that extracted Canny edge maps and segmentation masks effectively maintain continuous boundaries. As shown in Figure 5, when using them as control conditions, FreeControl fails to preserve boundary continuity, but our 360PanT (F) consistently produces translated 360-degree panoramas with continuous boundaries, regardless of the input conditions.

Visual results of various methods. To further demonstrate the efficacy of 360PanT for 360-degree panorama translation, we present additional visual comparisons with SDEdit [2], Pix2Pix-zero [4], P2P [1], PnP [5] and FreeControl [3] on both real-world and synthesized 360-degree panoramas. As illustrated in Figures 6, 7, 8, 9, 10, and 11, 360PanT outperforms these methods in maintaining visual continuity at the boundaries while also adhering to the structure and

*Corresponding author.

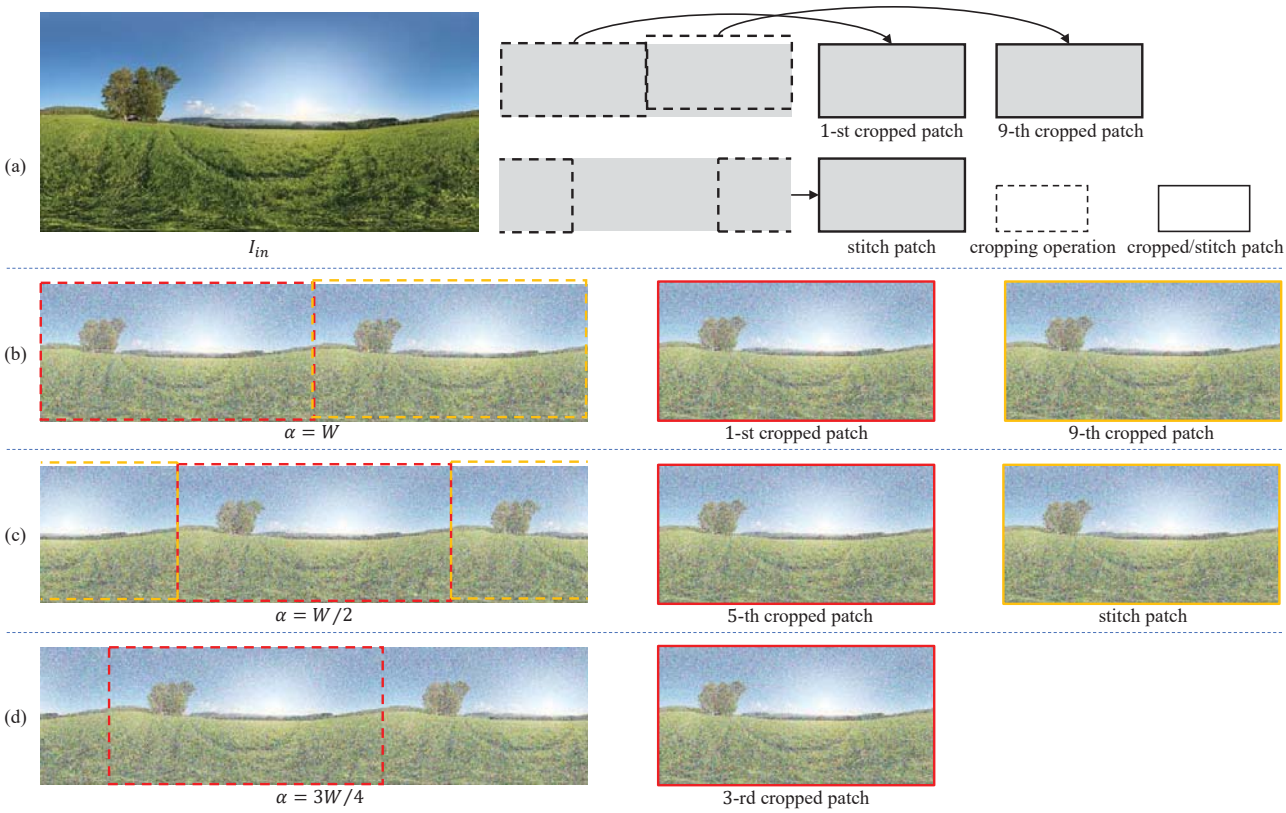

Figure 1. **Intuitive explanation for the choice of split constant α.** The cropped patches matching I_{in} during the sliding window process are highlighted by red or yellow dashed boxes. Note that the stitch patch is a special cropped patch. At each denoising step t, when $\alpha = W$ in (b) or $\alpha = \frac{W}{2}$ in (c), two cropped patches matching I_{in} but in different locations are denoised. Conversely, when α is set to $\frac{3W}{4}$ in (d), only one cropped patch matching I_{in} undergoes denoising. To ensure better boundary continuity in the final translated result, we choose to set α to $\frac{3W}{4}$.

semantic layout of the input 360-degree panoramic images.

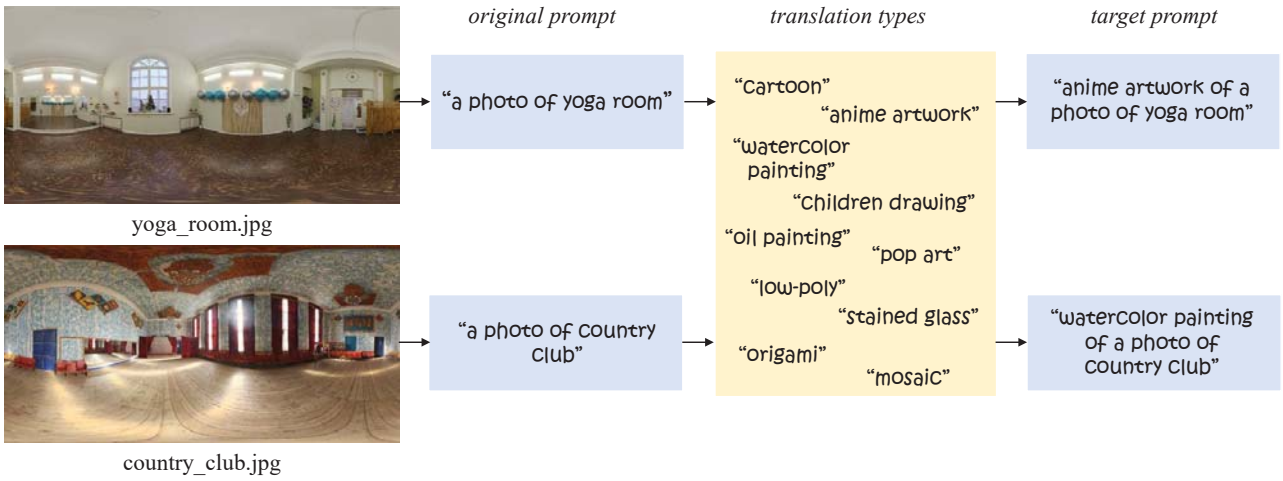

Figure 2. **Target prompt generation for real-world 360-degree panoramas within the *360PanoI-Pan2Pan* dataset.** Our 10 translation types are presented. A target prompt is formulated by combining a randomly selected translation type with the original prompt.

Figure 3. **Target prompt generation for synthesized 360-degree panoramas in the *360syn-Pan2Pan* dataset.** Each target prompt consists of a randomly chosen translation type and its corresponding original prompt.

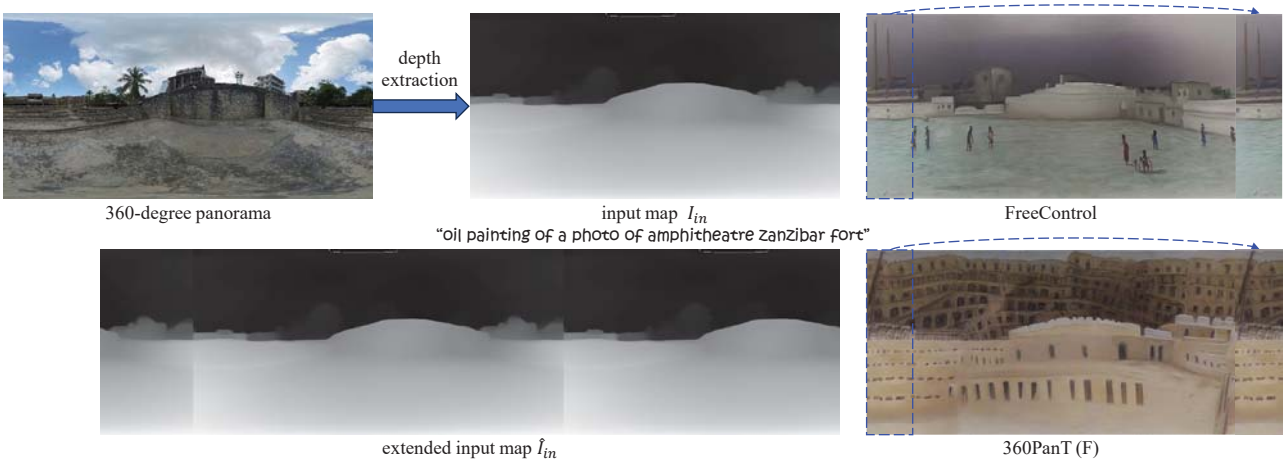

Figure 4. **Depth map with discontinuous boundaries as the control condition.** The boundaries of depth map I_{in} extracted from the 360-degree panorama are not continuous, resulting in visible cracks in the extended input map $\hat{I}_{in}$. In this situation, the translated panorama by our 360PanT (F) exhibits content inconsistency in the stitched area.

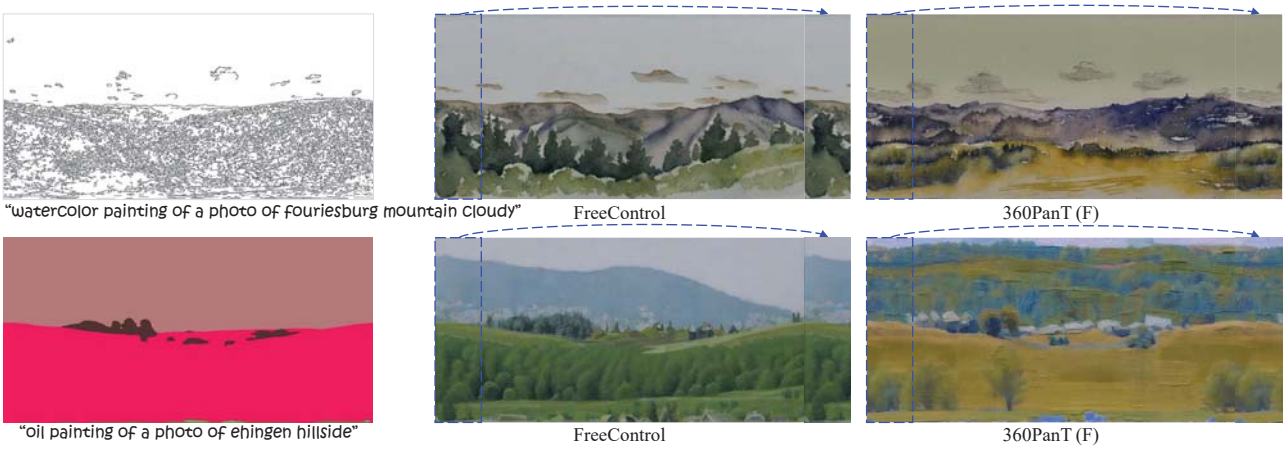

Figure 5. **Visual results using other control conditions.** The extracted Canny edge map and segmentation mask can both effectively maintain continuous boundaries. When using them as control conditions, respectively, FreeControl is unable to guarantee the boundary continuity of the translated panoramas. In contrast, our 360PanT (F) enables the translated 360-degree panoramas with continuous boundaries regardless of the input conditions.

Figure 6. **Visual results on real-world 360-degree panorama.** To easily identify visual continuity or discontinuity at the boundaries, we copy the left area of the panorama indicated by the blue dashed box and paste it onto the rightmost side of the image.

Figure 7. **Visual results on real-world 360-degree panorama.**

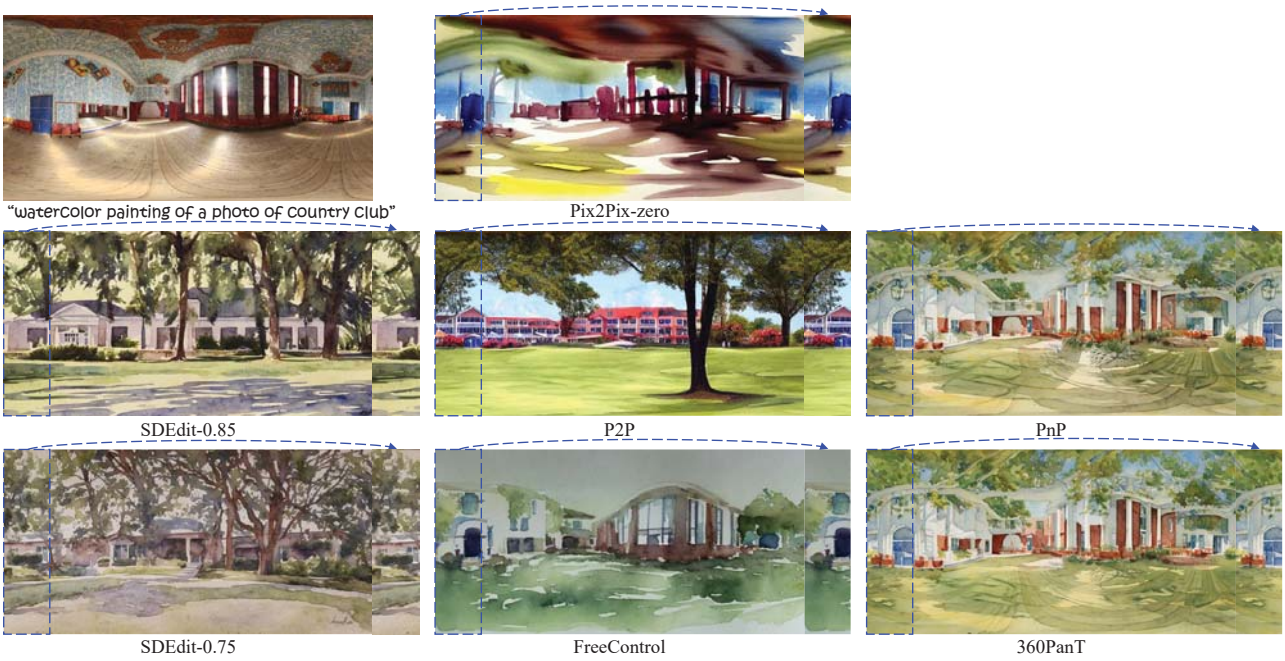

Figure 8. **Visual results on real-world 360-degree panorama.**

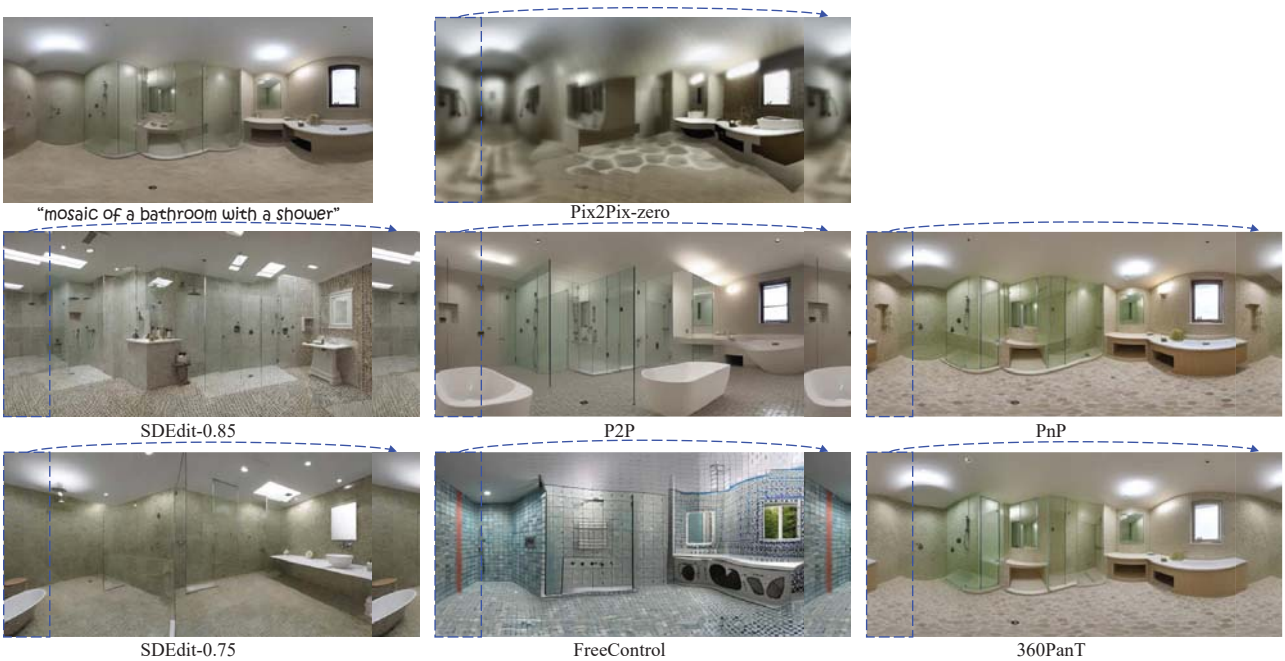

Figure 9. **Visual results on synthesized 360-degree panorama.** To easily identify visual continuity or discontinuity at the boundaries, we copy the left area of the panorama indicated by the blue dashed box and paste it onto the rightmost side of the image.

Figure 10. **Visual results on synthesized 360-degree panorama.**

Figure 11. **Visual results on synthesized 360-degree panorama.**

References

[1] Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Prompt-to-prompt image editing with cross attention control. *arXiv preprint arXiv:2208.01626*, 2022. 1

[2] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. SDEdit: Guided image synthesis and editing with stochastic differential equations. In *International Conference on Learning Representations*, 2022. 1

[3] Sicheng Mo, Fangzhou Mu, Kuan Heng Lin, Yanli Liu, Bochen Guan, Yin Li, and Bolei Zhou. Freecontrol: Training-free spatial control of any text-to-image diffusion model with any condition. *arXiv preprint arXiv:2312.07536*, 2023. 1

[4] Gaurav Parmar, Krishna Kumar Singh, Richard Zhang, Yijun Li, Jingwan Lu, and Jun-Yan Zhu. Zero-shot image-to-image translation. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–11, 2023. 1

[5] Narek Tumanyan, Michal Geyer, Shai Bagon, and Tali Dekel. Plug-and-play diffusion features for text-driven image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1921–1930, 2023. 1

[6] Hai Wang, Xiaoyu Xiang, Yuchen Fan, and Jing-Hao Xue. Customizing 360-degree panoramas through text-to-image diffusion models. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 4933–4943, 2024. 1