Panorama Generation From NFoV Image Done Right
Supplementary Material

Dian Zheng[1,*] Cheng Zhang[2] Xiao-Ming Wu[1]
Cao Li[3,✉] Chengfei Lv[3] Jian-Fang Hu[1] Wei-Shi Zheng[1,4,✉]

[1]Sun Yat-sen University [2]Monash University [3]Alibaba Group
[4]Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China

zhengd35@mail2.sysu.edu.cn wszheng@ieee.org

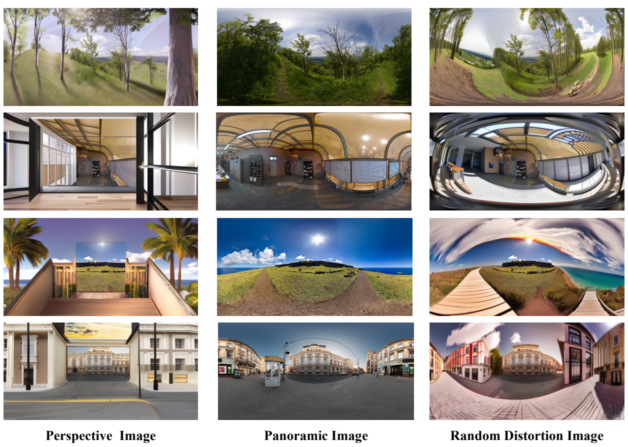


Figure S1. Visual results of panoramic images and corresponding perspective images. Zoom in for best view.

A. Random Distortion Image Generation

Motivation. The random distortion image could accelerate the robustness of our Distort-CLIP to assess distortion. The challenge is how to generate it while sharing the same content with panorama. We observe that the training data of stable diffusion [1] contains panorama image (with the height-width ratio 1:2). However, due to its relatively small proportion in training data, the distortion of the generated panoramic image is quite chaotic, which perfectly meets our requirements.

Method. We apply a mask to the panoramic image, leaving only the central area (*i.e.*, the same mask as shown in the main paper Fig. 3). Then we use the sd-outpainting based on [1] to outpaint the model with an extra prefix "A panorama image of ". In this way, we obtain random distortion images with the same content as panoramas.

Visual results. Here we also show the panoramic, generated perspective and random distortion images in Fig S1. It could be seen that we generate different distortion types images with similar content, ensuring the success of our Distort-CLIP.

B. Condition Registration Mechanism

We show the model architecture of different condition registration mechanisms in Fig S4. We explore the widely used mechanism for condition registration and validate that the all-block registration with addition is the best choice for position-encoding-like conditions.

C. More Applications

Text Editing. We show the text editing results below. Note that we specifically selected this NFoV image (with snow at the bottom), which is extremely challenging for text editing when the text is inconsistent with the snow. However, our method still generates panoramas consistent with the text.

Figure S2. Visual results of text editing.

Random NFoV. We show results below, even the input image is extremely small and the position is tricky, we could generate panoramas with visual appealing.

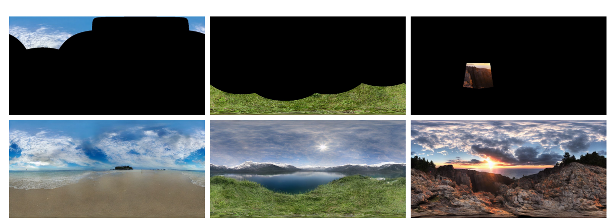


Figure S3. Visual results with random NFoV input.

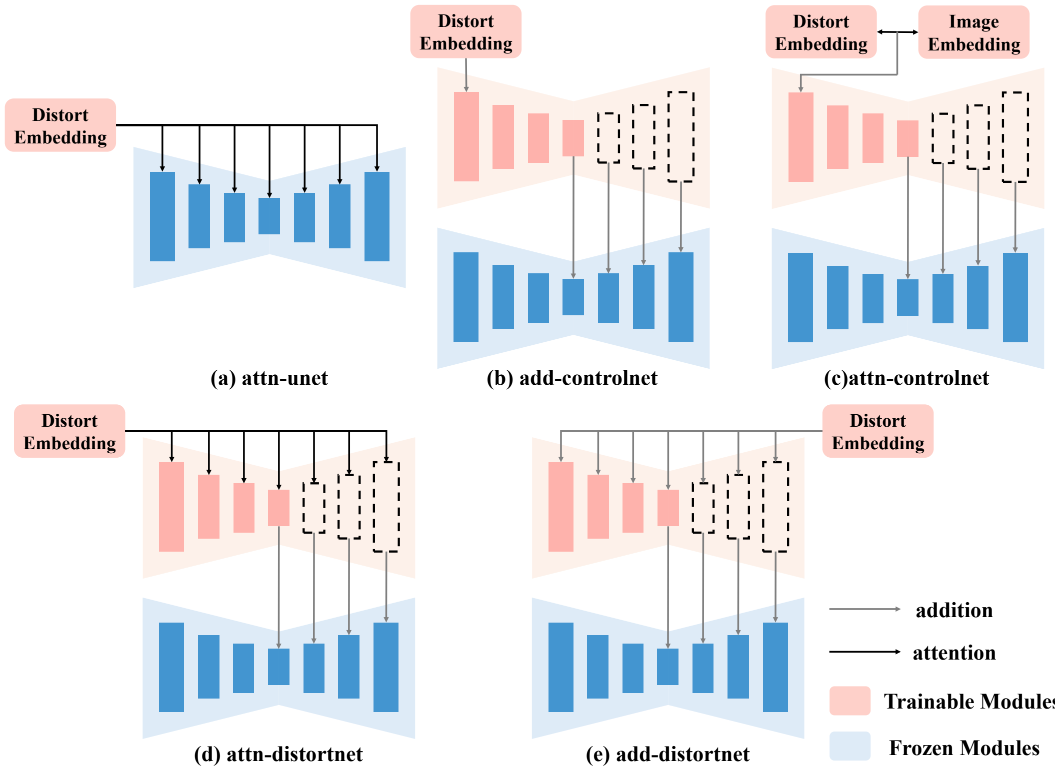


Figure S4. Various condition registration mechanisms. Note that both addition and attention own small learnable parameters (*e.g.*, projection layers and zero convolution for addition; common attention modules for attention), we omit it in figure for simplification. Zoom in for best view.

D. More Visual Results

In Fig S5, we show panoramic and perspective results of generated panoramas by different methods. Our PanoDecouple achieves great image quality and accurate distortion simultaneously. We also show more raw image panorama outpainting results in Fig S6. Enjoy it!

References

[1] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis. *arXiv preprint arXiv:2307.01952*, 2023. 1

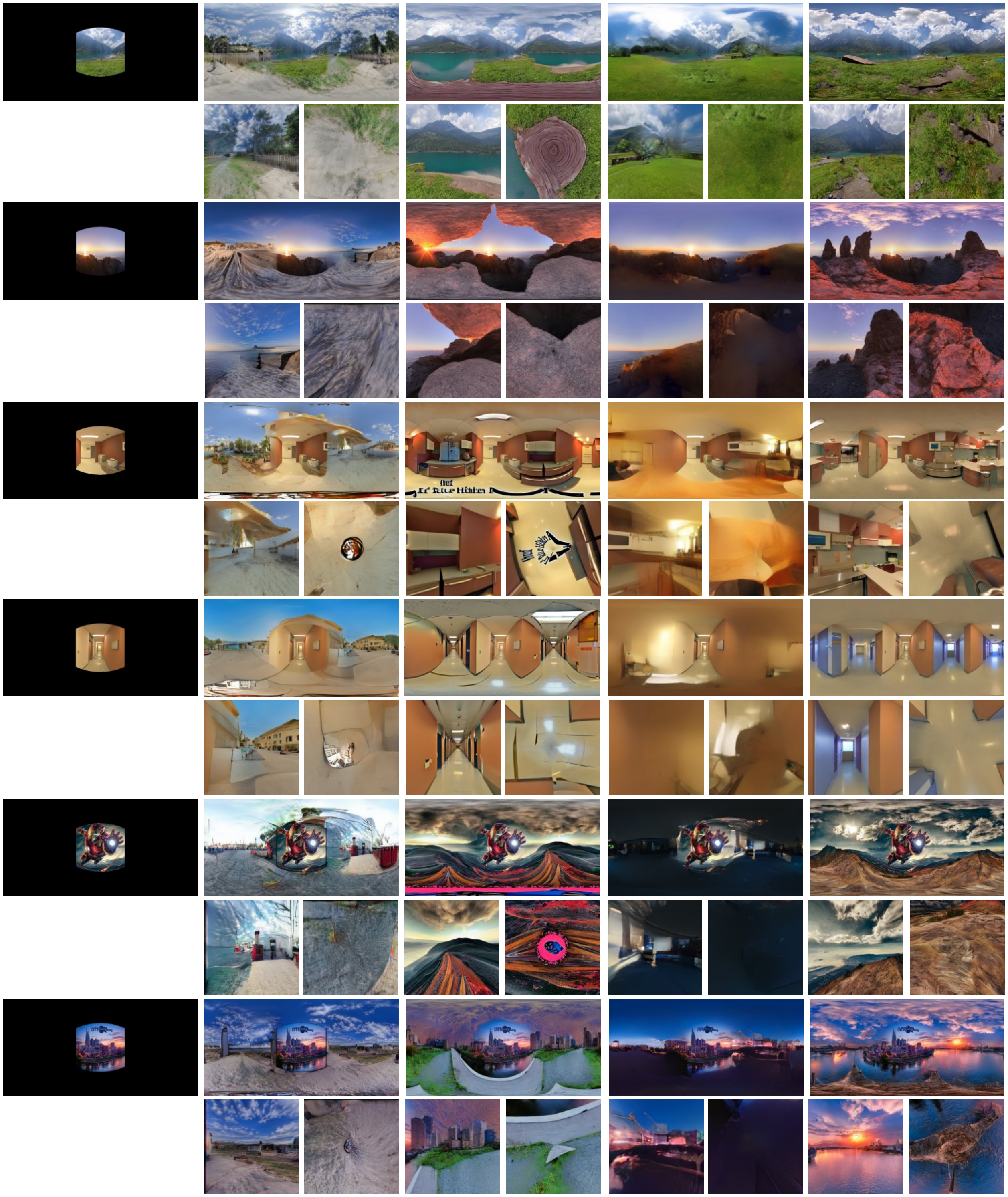
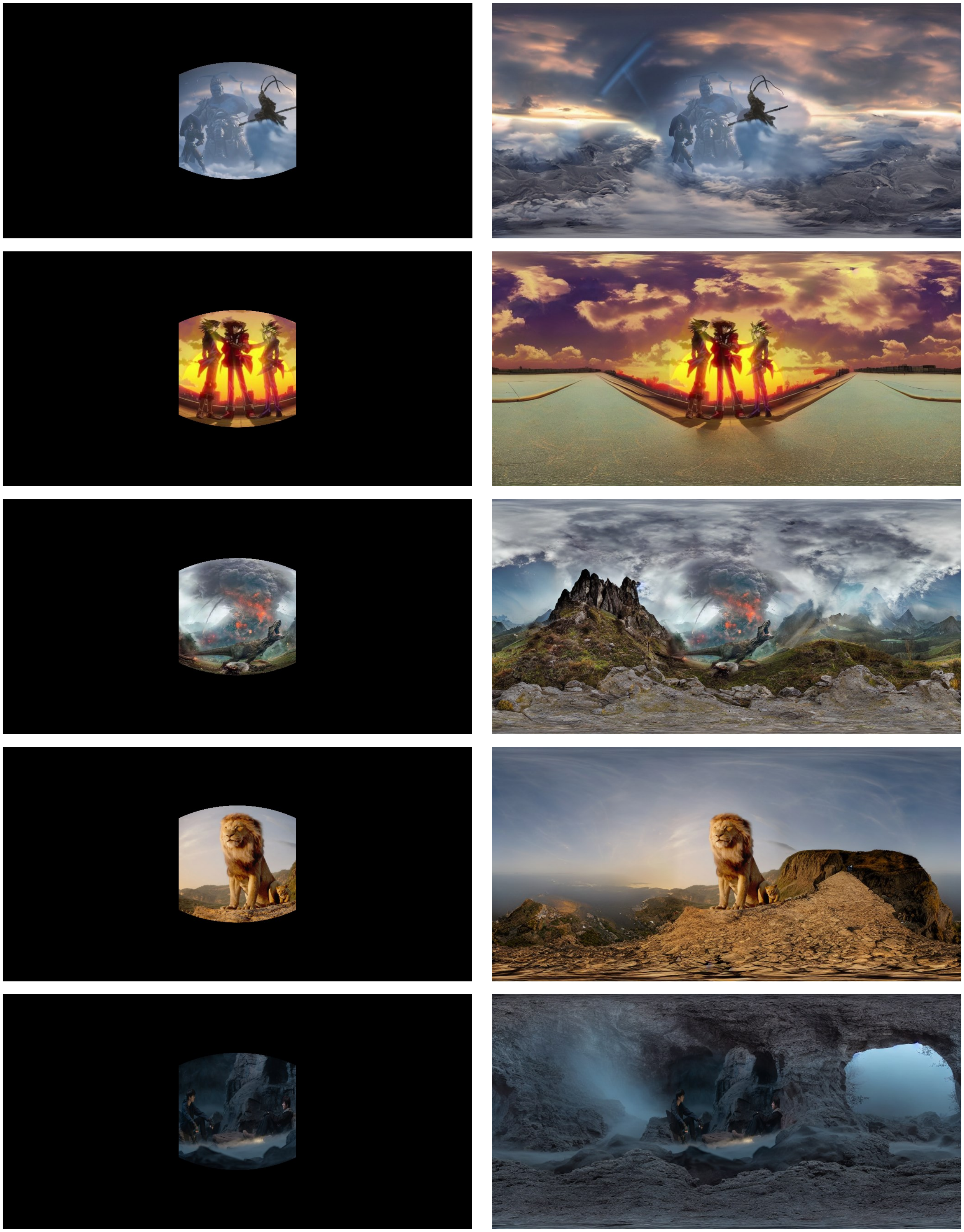


Figure S5. Visual results of panoramic images and corresponding perspective images.



Partial Input | **PanoDecouple (Ours)**

Figure S6. Visual results with raw image input. **Note that the images we use are for academic purposes only. If any copyright infringement occurs, we will promptly remove them**.