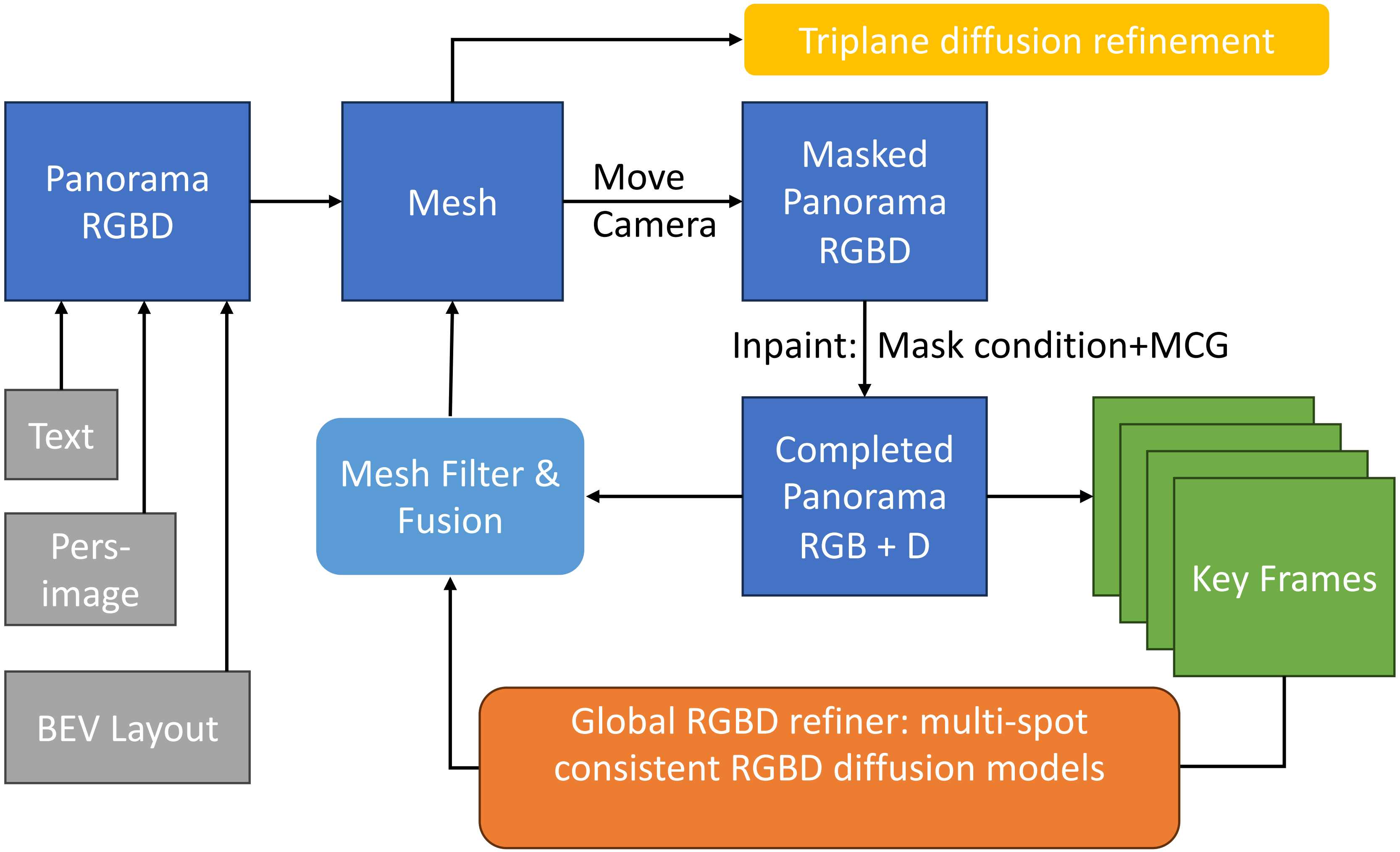
3D Scene Generation using Panoramic RGBD Diffusion Models

Matt Xu

3D Scene Generation using Panoramic RGBD Diffusion Models

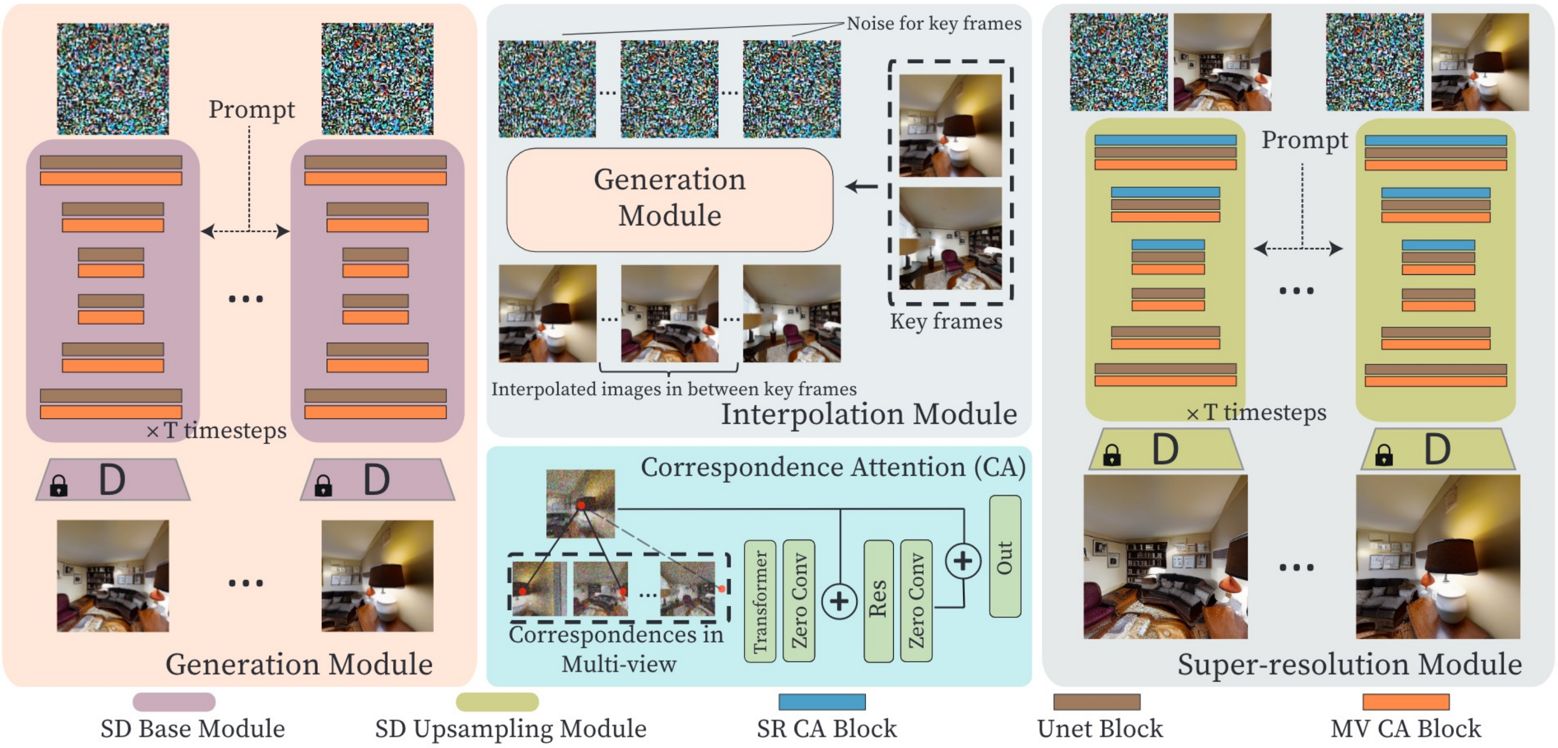


Key novelties:

- A new high-resolution perspective text-2-panorama diffusion model
- A new RGBD diffusion model
- Multi-spot RGBD diffusion models: using projected RGBD alignment error as Unet condition (Better than MVDiffusion and Manifold constraint)

Panorama RGBD generation

- Finetune SDXL using Lora with a loop consistent loss
- Multi-view consistent Img2img refinement in the perspective field
  - Multi-diffusion cannot solve sphere distortion
  - Original MVDiffusion cannot achieve meaningful global structure



For_Honor, The image is a 360-degree panoramic view of a fantasy-themed village, featuring various buildings and structures. The village appears to be located near a body of water, possibly a river or a lake. There are also some trees visible in the background, adding to the overall atmosphere of the scene.

genshin, The image is a 360-degree panoramic view of a cobblestone street with buildings and a windmill in the background. It appears to be a computer-generated or animated scene, possibly from a video game or virtual world.

Ghost_Recon, The image is a 360-degree panoramic view of a mountainous landscape with a lake in the foreground. There are several trees and bushes visible in the scene. The image appears to be a digital or computer-generated representation of the landscape, rather than an actual photograph.

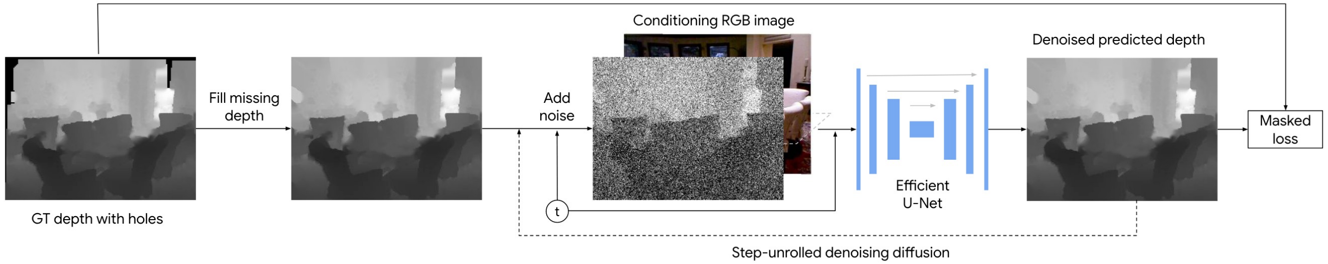
The image is a 360-degree panoramic view of a bedroom, featuring a large bed, a television, and a window with a view of the city. The room is decorated in a modern style, with wooden floors and white walls. There is also a chair and a lamp in the room. The panoramic view provides an immersive experience, allowing the viewer to take in the entire room and its surroundings.

RGBD diffusion

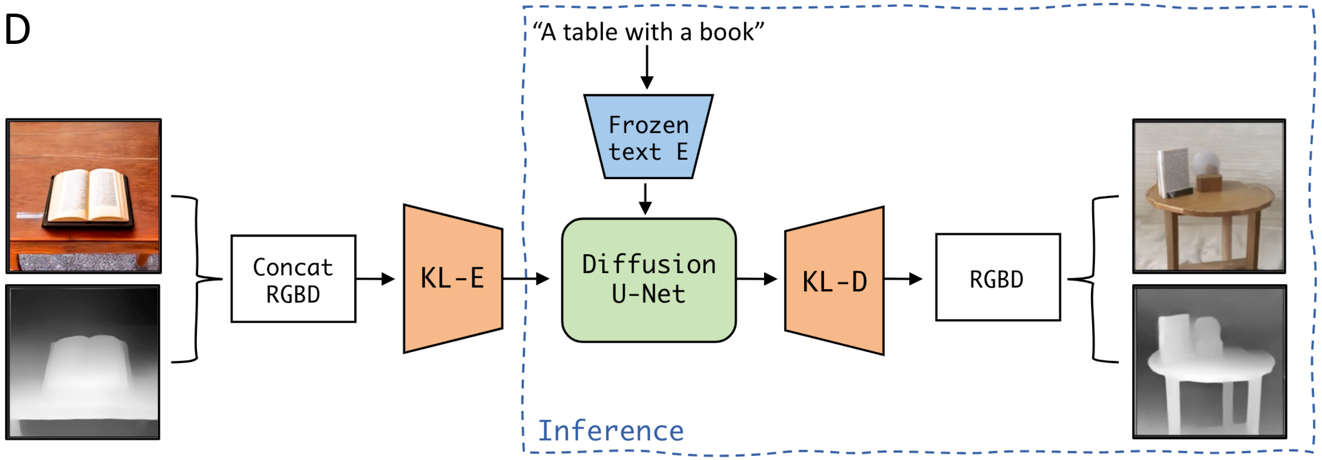
Our depth diffusion:

- Use pretrained RGB diffusion
- Use RGB UNet features as condition to train depth UNet
- Fintune depth Unet with Lora, works well on small scale depth data

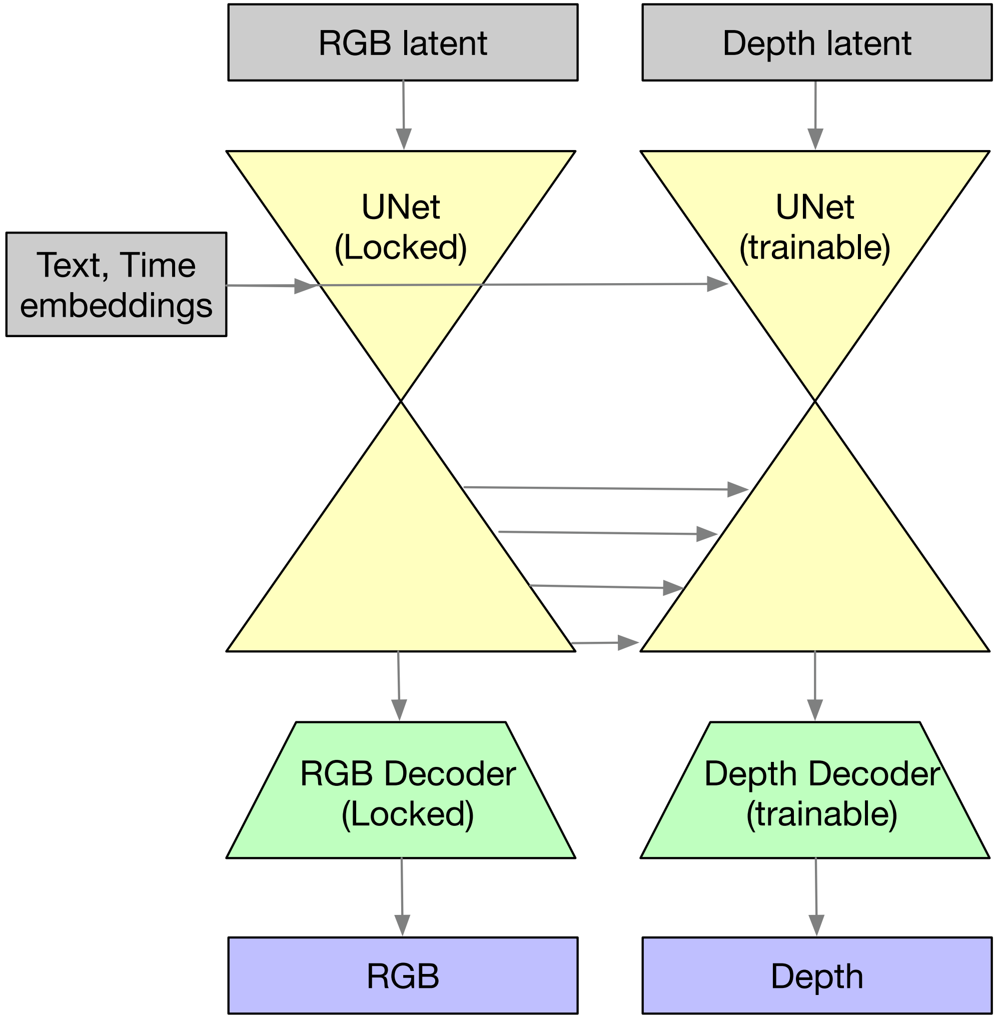
Monocular Depth Estimation using Diffusion Models

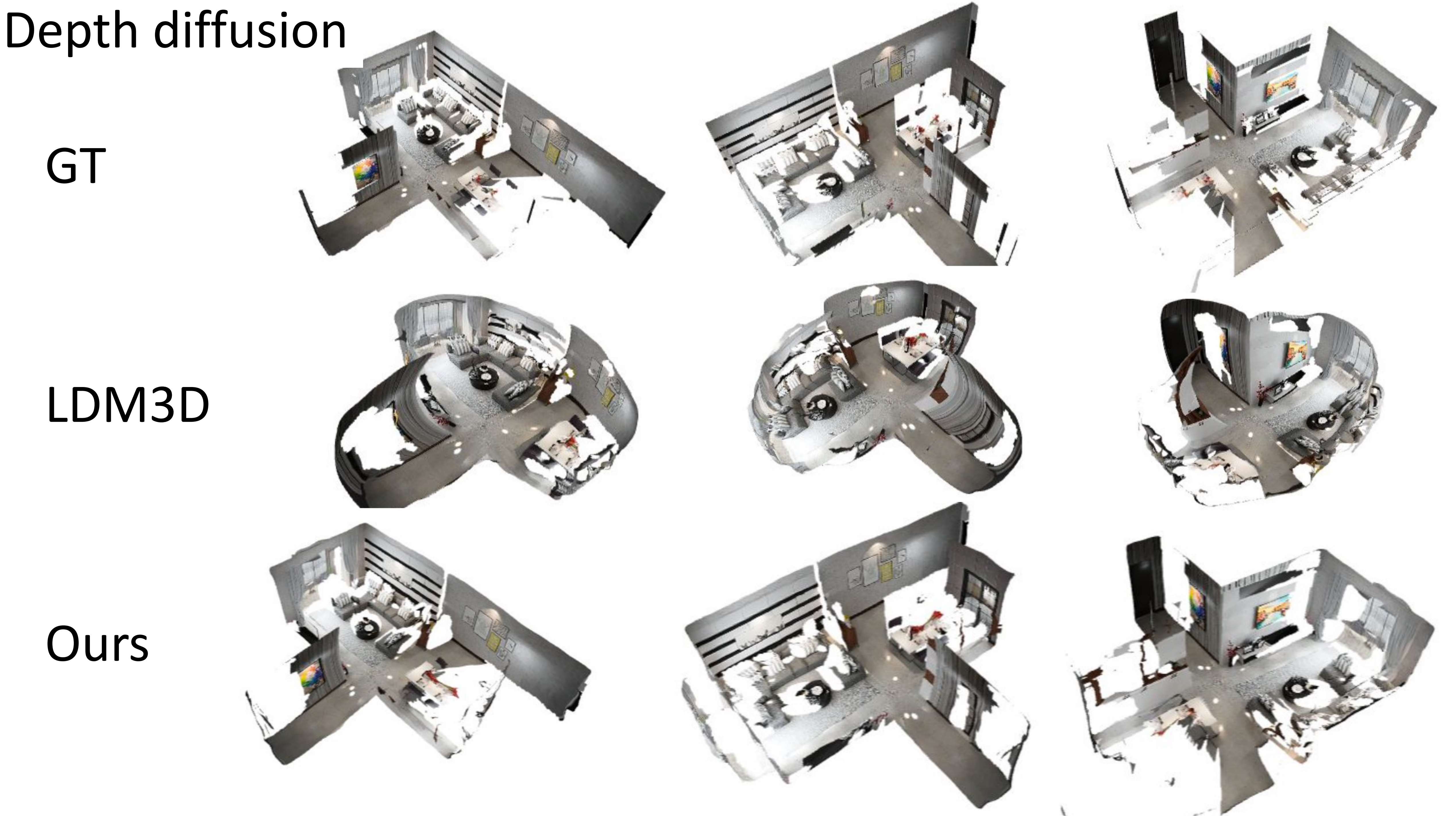


LDM3D



Ours: depth diffusion




Depth diffusion
GT
LDM3D
Ours

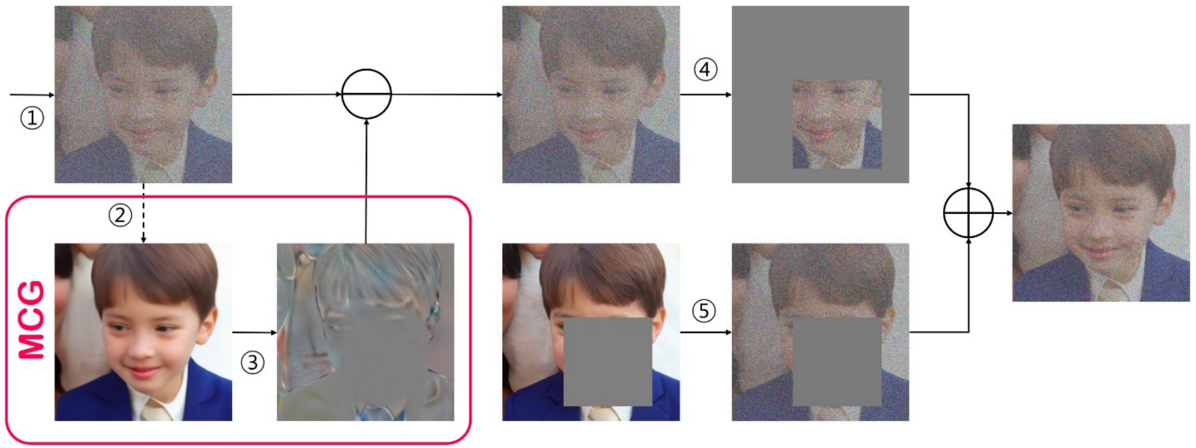
Panorama inpainting: Masked image conditional UNet + Manifold constraint training

Disadvantage of the original manifold constraint generation (MCG):

- Calculate gradient in the inference leading to 6 times slower
- Use MCG + unconditional score functions to solve conditional tasks is challenging, which does not work very well on panoramic inpainting

Proposed method: Manifold constraint training with masked image conditional Unet:

- Fast, no gradient required in inference
- Chamfer distance soft mask
- VAE on around filling masked image, better than gray masked image



$$\boldsymbol{x}'_{i-1} = \boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{s}_\theta) - \alpha \frac{\partial}{\partial \boldsymbol{x}_i} \|\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{H}\hat{\boldsymbol{x}}_0(\boldsymbol{x}_i))\|_2^2 + g(\boldsymbol{x}_i)\boldsymbol{z}, \quad \boldsymbol{z} \sim \mathcal{N}(0, \boldsymbol{I}),$$

$$\boldsymbol{x}_{i-1} = \boldsymbol{A}\boldsymbol{x}'_{i-1} + \boldsymbol{b}.$$

Vanilla FT inpainting: Finetune inpainting models using Lora.

Our MCT inpainting: inpainting model trained with Lora and manifold constraint loss.

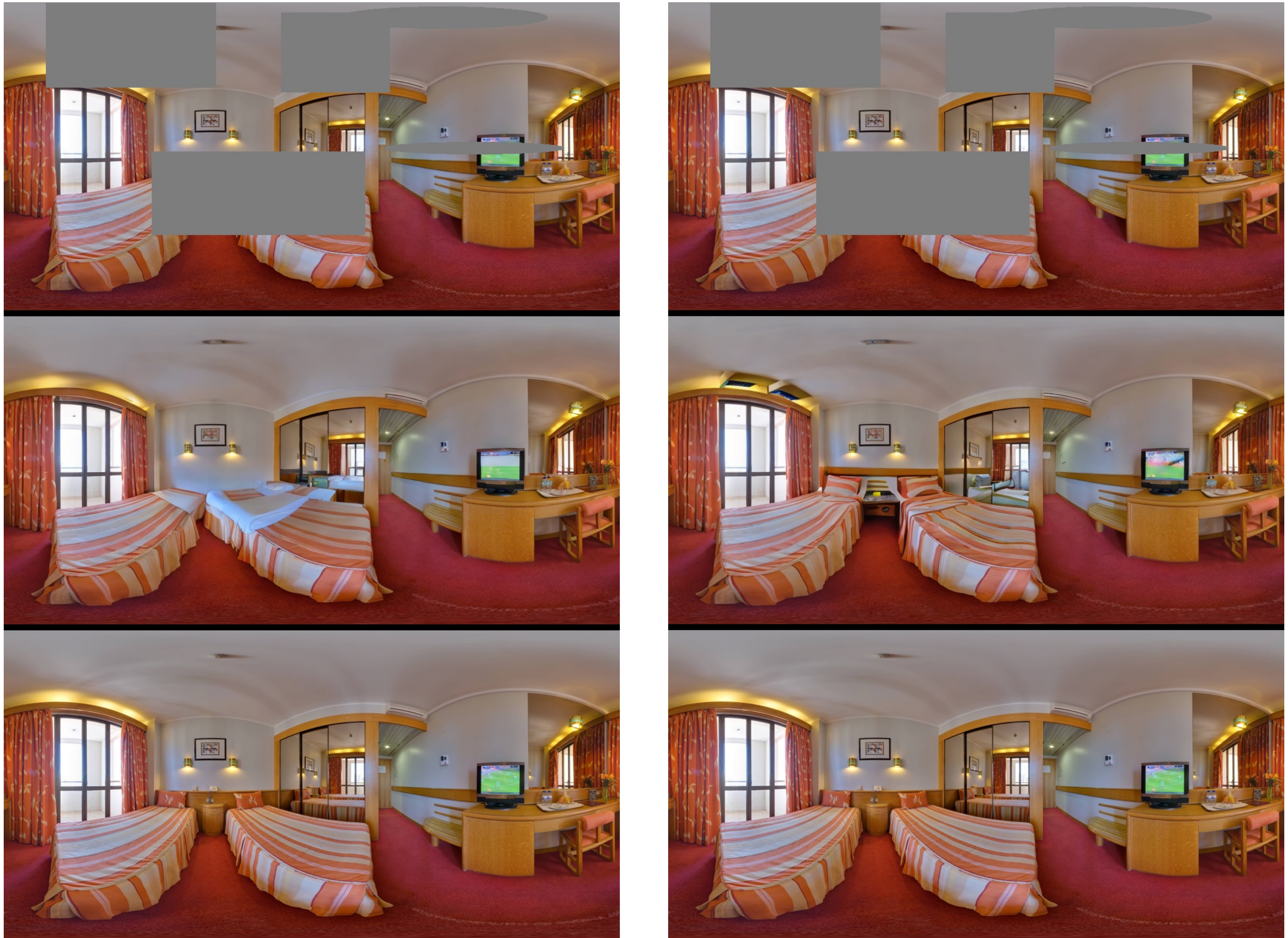
(1) Inpainting diffusion Unet input (5 channels): [masked image latent, mask]
(2) Text2Img Unet input (4 channels): [masked image latent]

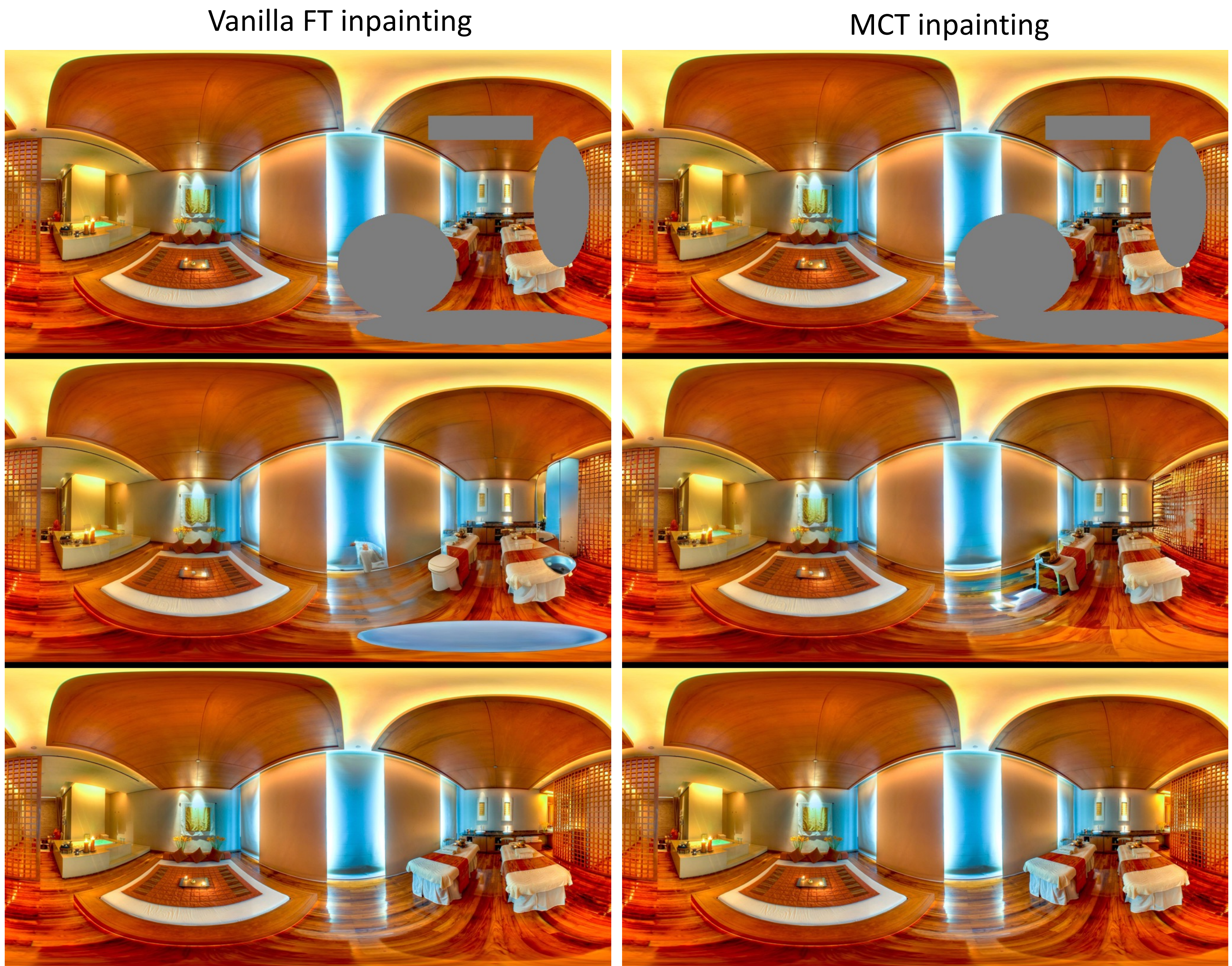
MCG on inference -> on training

MCG on inference: manifold constraint + Text2Img Unet, 6 times slower

MCG on inference: manifold constraint + Inpainting Unet, no gradient in inference

Vanilla FT inpainting | MCT inpainting




Vanilla FT inpainting
MCT inpainting

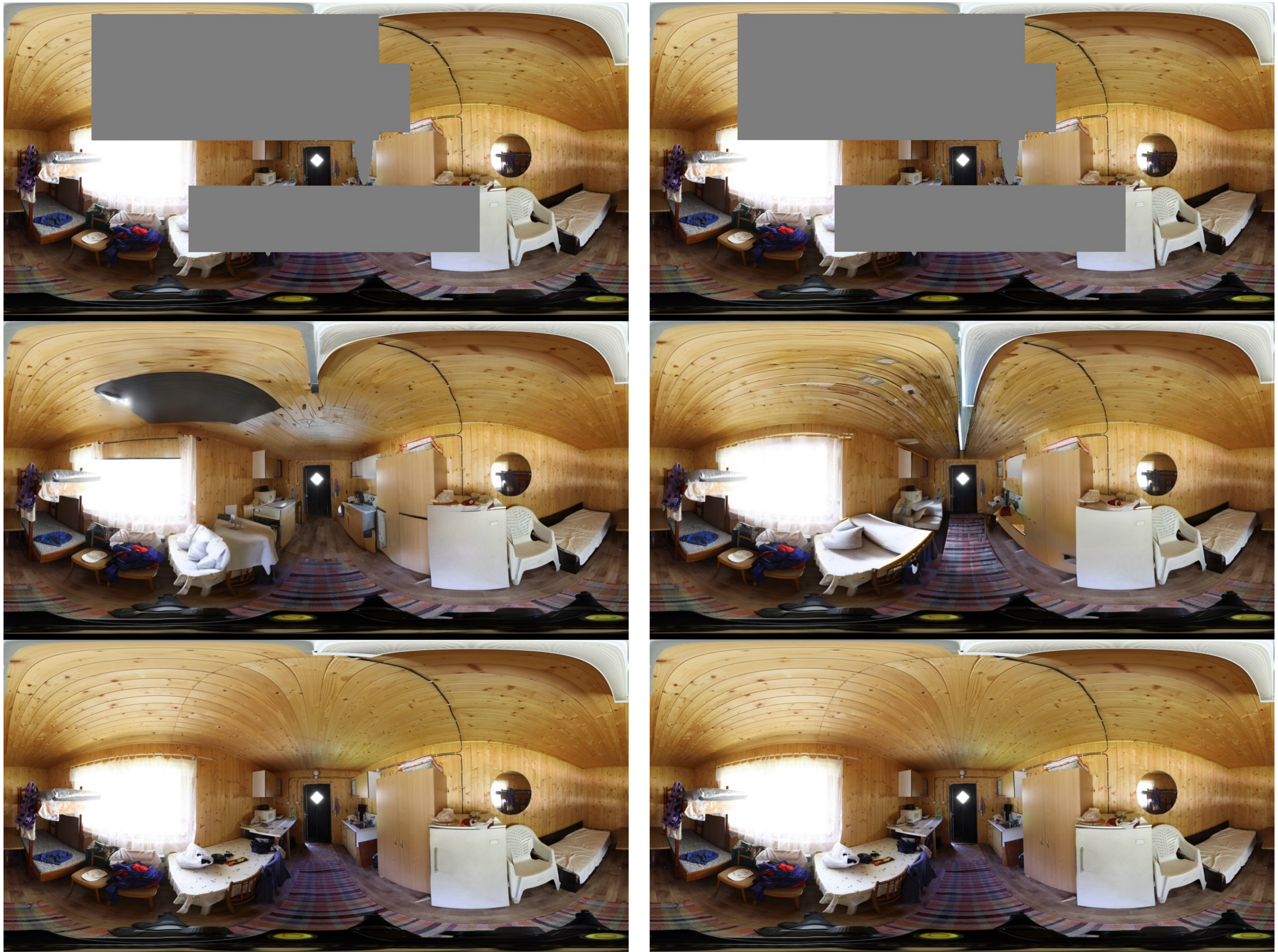
Vanilla FT inpainting

MCT inpainting

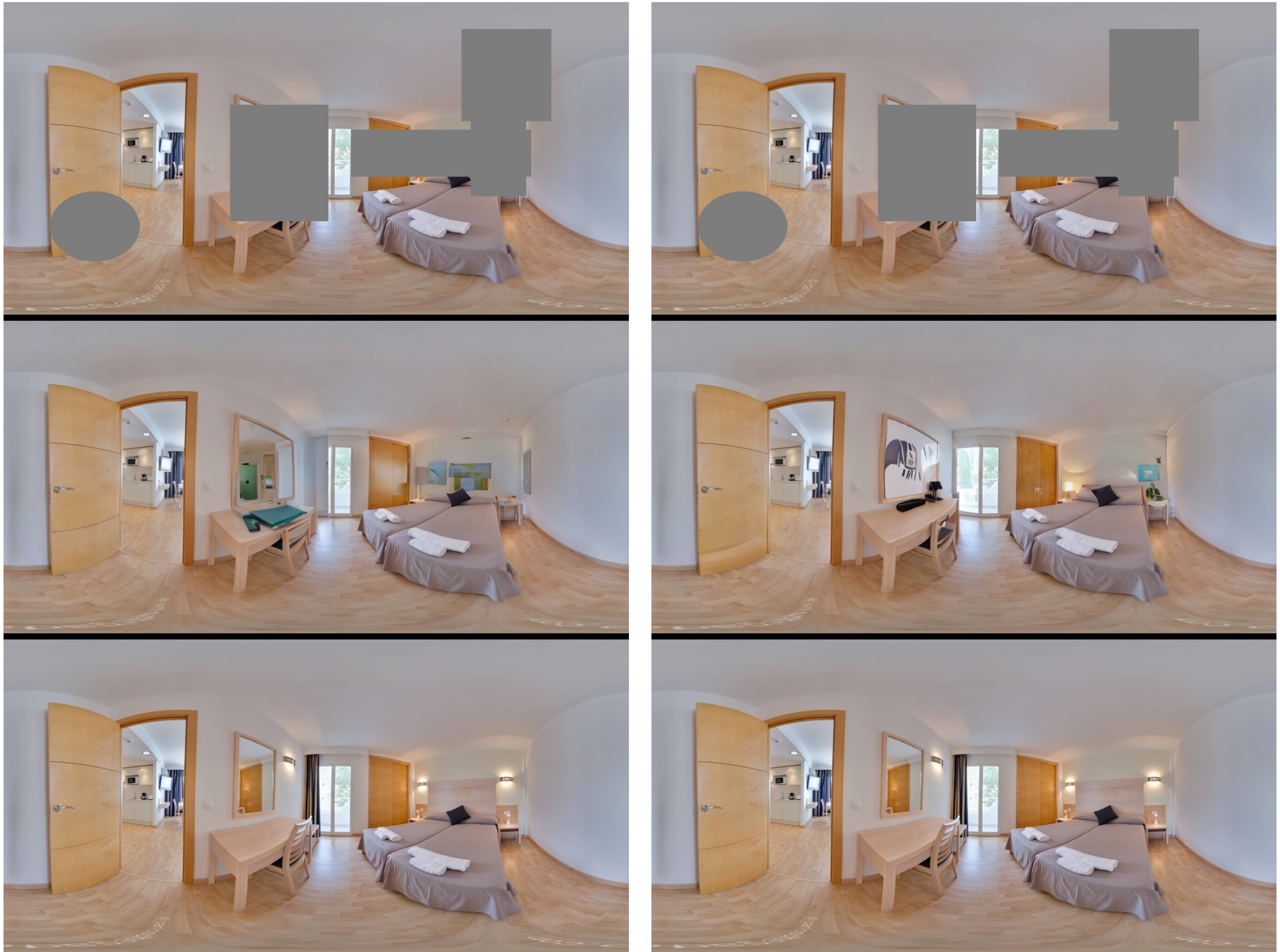
Vanilla FT inpainting

MCT inpainting

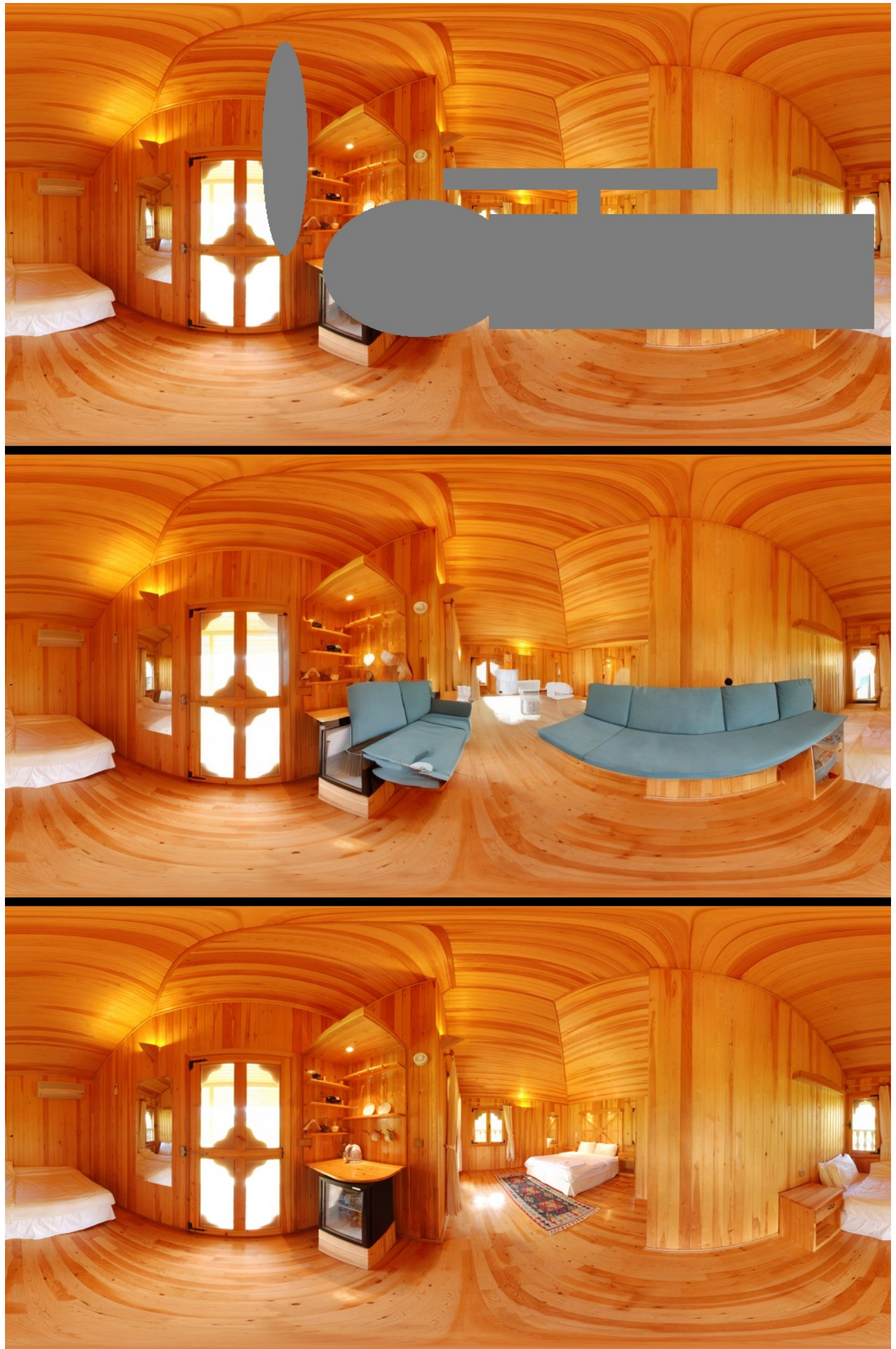
Vanilla FT inpainting | MCT inpainting



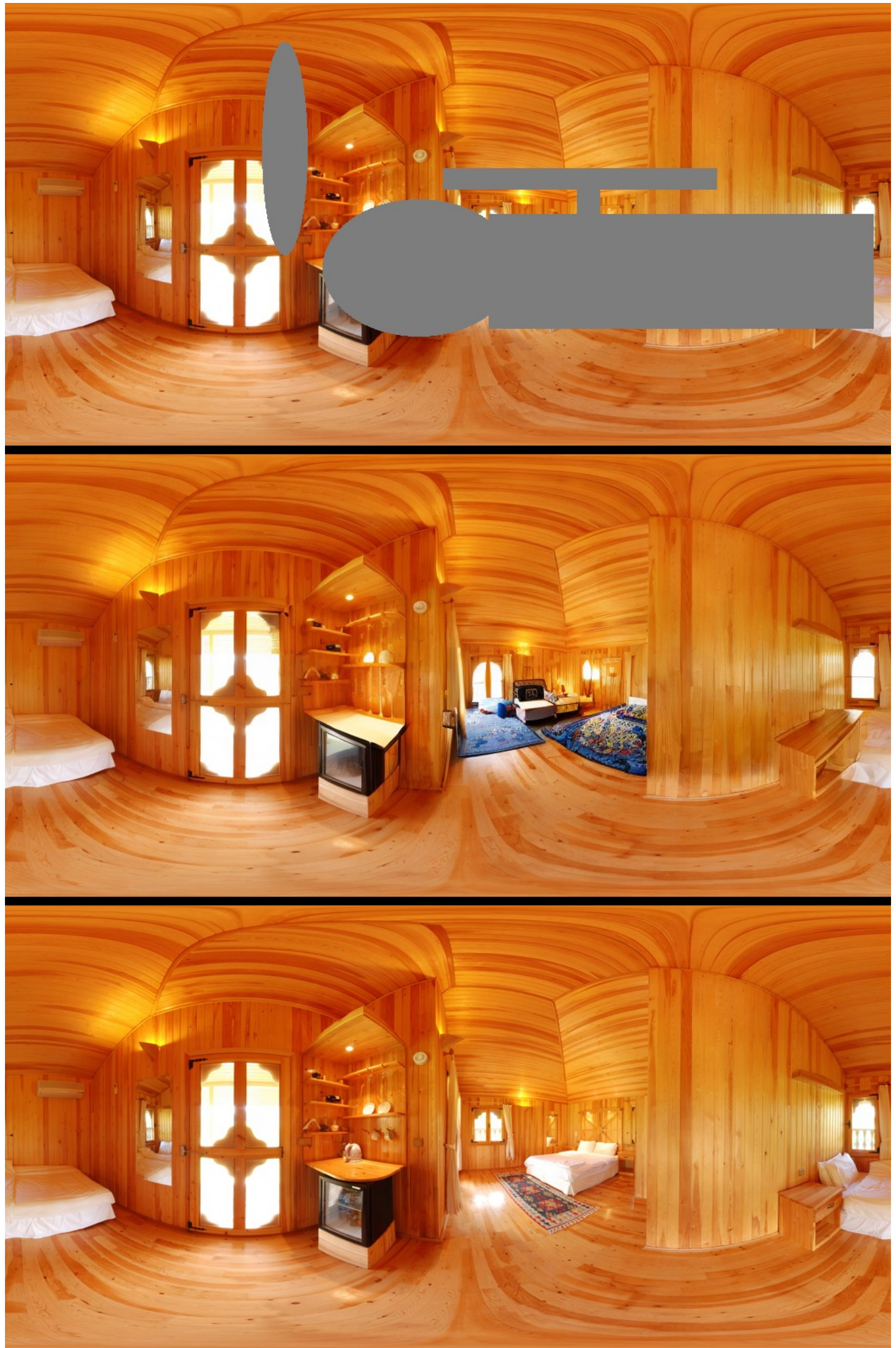
Vanilla FT inpainting MCT inpainting



Vanilla FT inpainting



MCT inpainting



Vanilla FT inpainting

MCT inpainting