Semantic Aware Diffusion Inverse Tone Mapping: Supplementary Material

1 Introduction

The supplementary material includes additional figures and illustrations to support the work. Figure Figure 1 shows how varying (α, β) parameters influence the inpainted details. Figures 2 to 4 were used for the subjective experiment in Section 4.2 in the main paper. These show a single representative exposure of the methods used and the results from [1] (EIL), [2] (SAN) and [3] (MAR) which were chosen for the experiment due to their performance from Table 1 in the main paper. Figure 5 shows further results for DITMO using both [4], and [5]. The input image is shown first, followed by three exposures. Our method is good at generating missing details in various areas, such as clouds, sky, and interior details in these scenes using both inpainting methods. Figure 6 shows results comparing our method to GlowGAN and illustrates that our method inpaints more details in clipped regions. We also show the average dynamic range for the reconstructions in Table 1, and show some more failure cases in Figures 7 and 8.

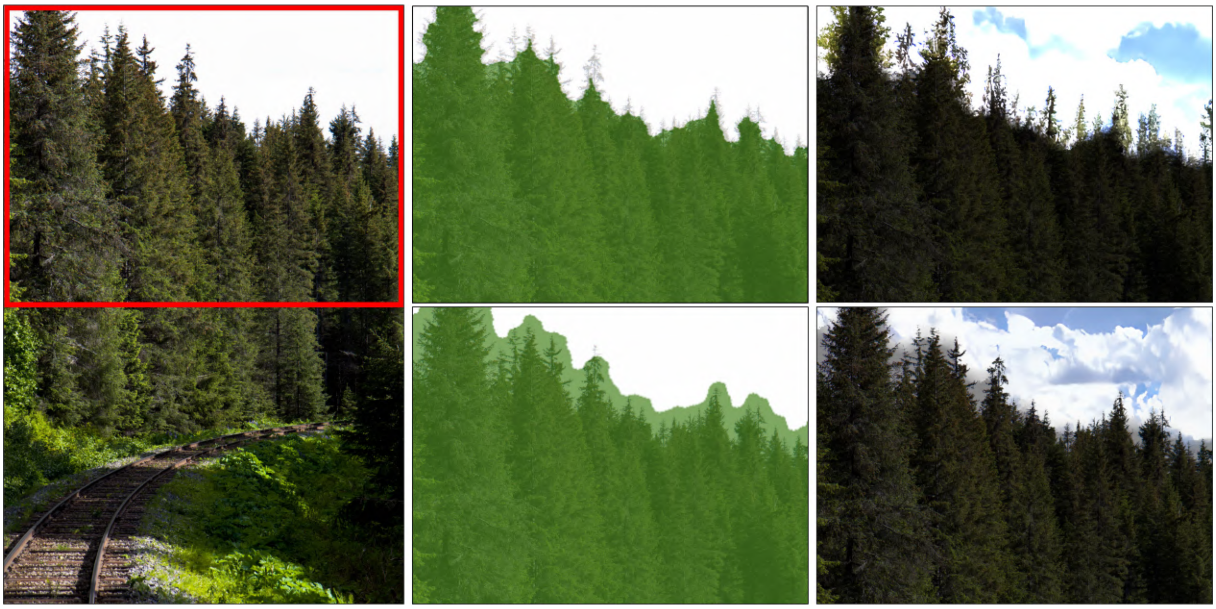

Figure 1: *Regulating $(\alpha,\ \beta)$ parameters in morphological opening.* The dilation parameters ensure that information outside the inpainting mask (middle) does not influence the reinvented details (right).

As described in Section 3.4 (Parameter Selection) in the paper (α, β) parameters play a vital role in generating plausible details. Figure 1 shows how varying parameters influence the inpainted details. The masks generated using $(2, 5)$ (top) show how parts of the neighboring *vegetation* information are included (in green mask) and it creates improper details in the inpainted results. The mask generated using $(10, 5)$ (bottom) avoids this by carefully focusing the information going into the stable-diffusion module.

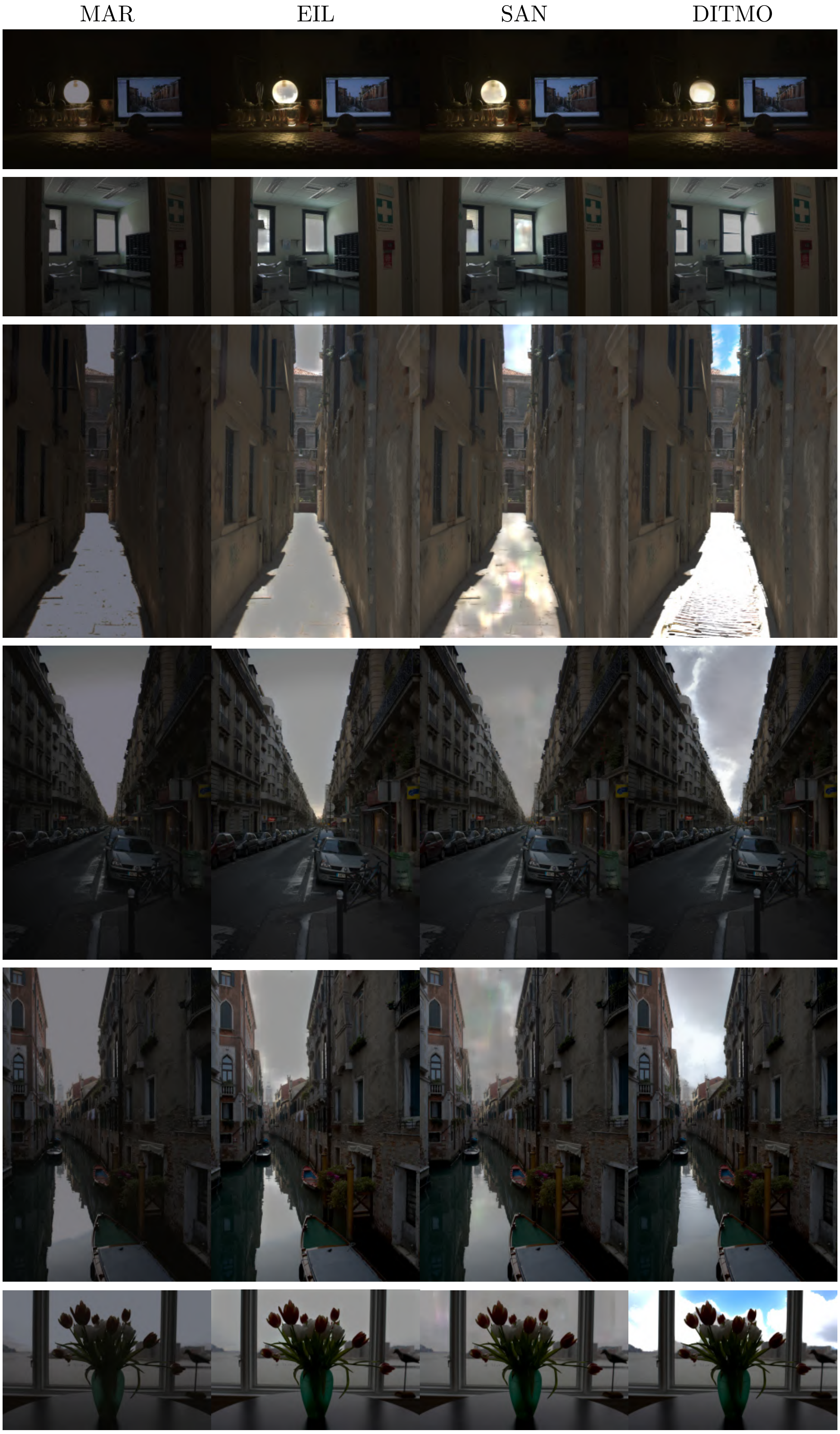

Figure 2: Images used in the experiment (Section 4.2 in the main paper) showing the same exposure (-2 stops) of the generated HDR image from all the methods.

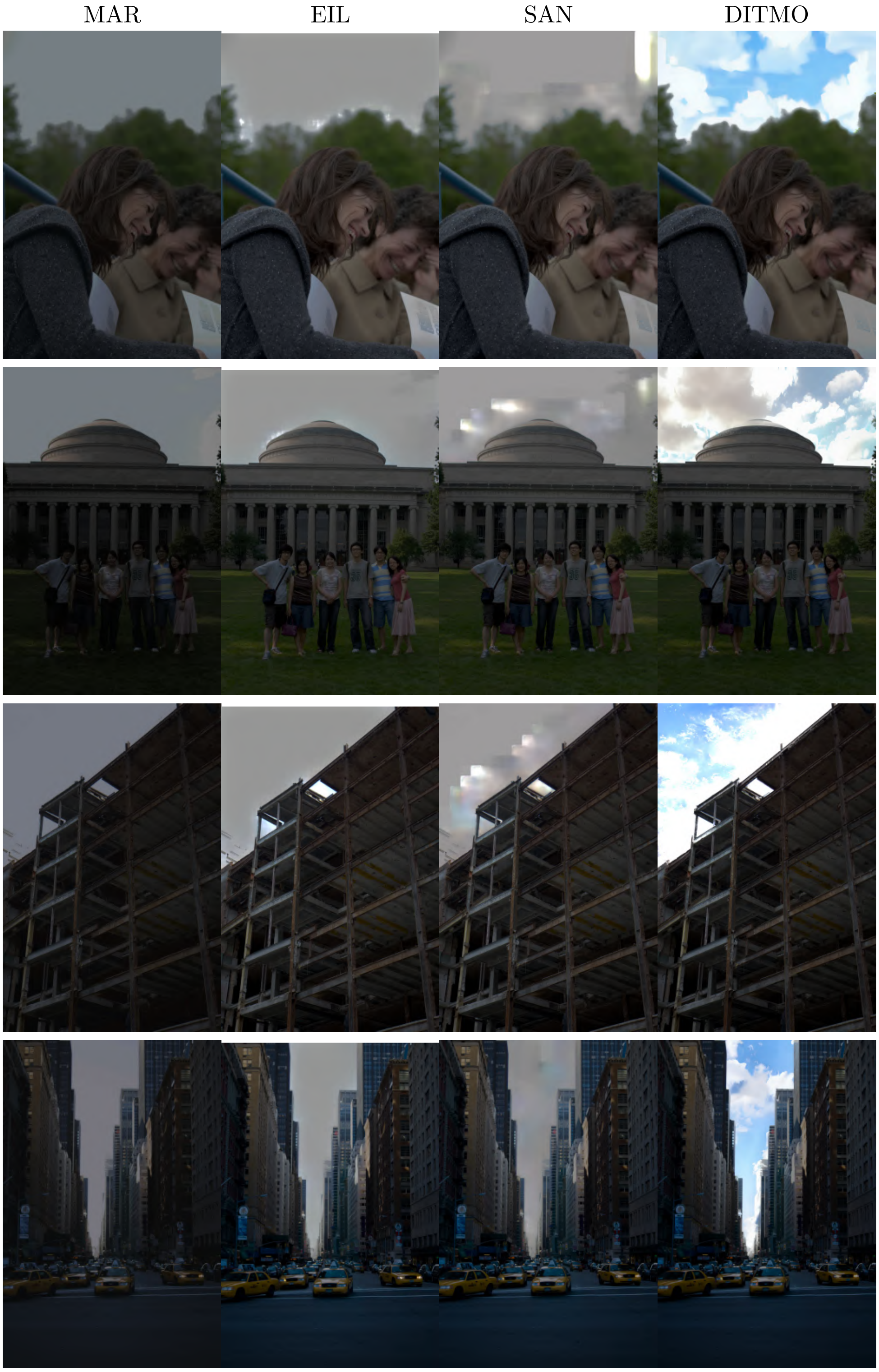

Figure 3: Images used in the experiment (Section 4.2 in the main paper) showing the same exposure (-2 stops) of the generated HDR image from all the methods.

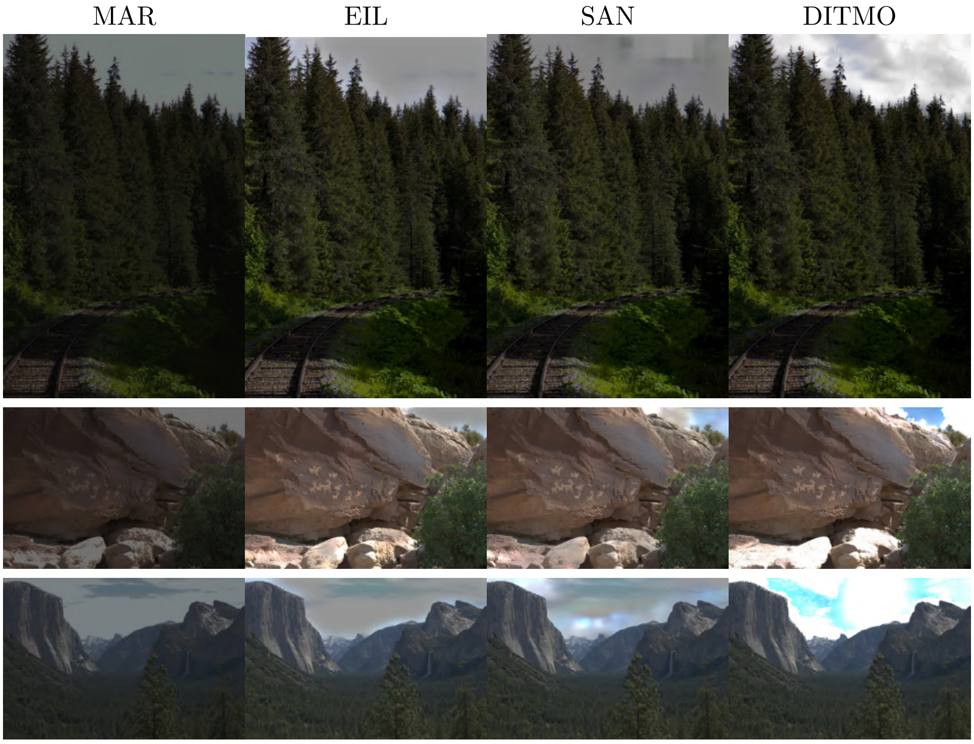

Figure 4: Images used in the experiment (Section 4.2 in the main paper) showing the same exposure (-2 stops) of the generated HDR image from all the methods.

	SAN	YU	MAR	EIL	LIU	DITMO
DR	18.492	14.560	9.955	18.289	14.493	17.710

Table 1: Average dynamic range for iTMOs

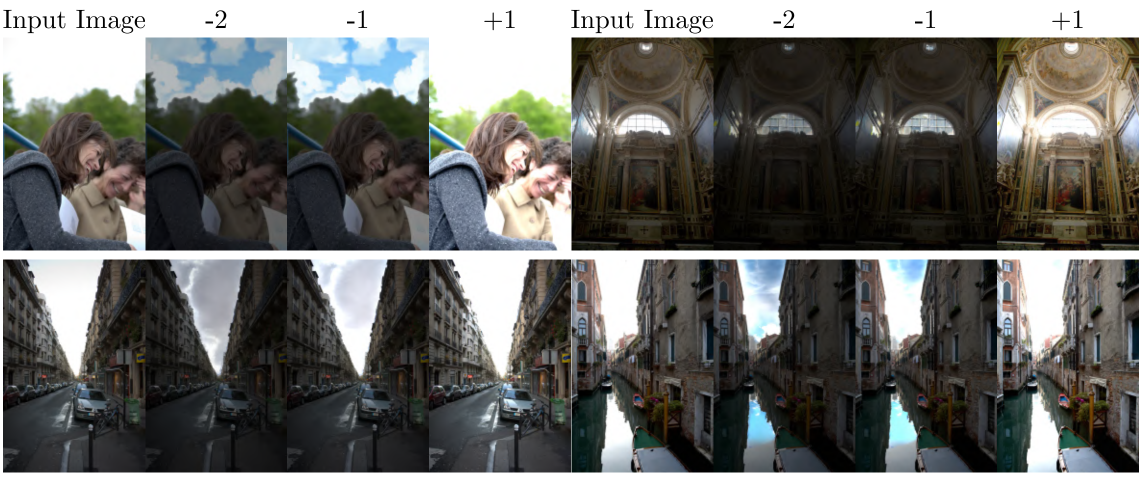

Figure 5: Results using DITMO with inpainting from the SD model [4] (top row) and CN model [5] (bottom row) over a range of images in multiple exposures. The input image is shown first, followed by three exposures. Our method is good at generating missing details in various areas, such as clouds, sky, and interior details in these scenes using both inpainting methods. We use these two different models to illustrate the generalizability of DITMO and choose these methods as they are recent prompt driven diffusion models.

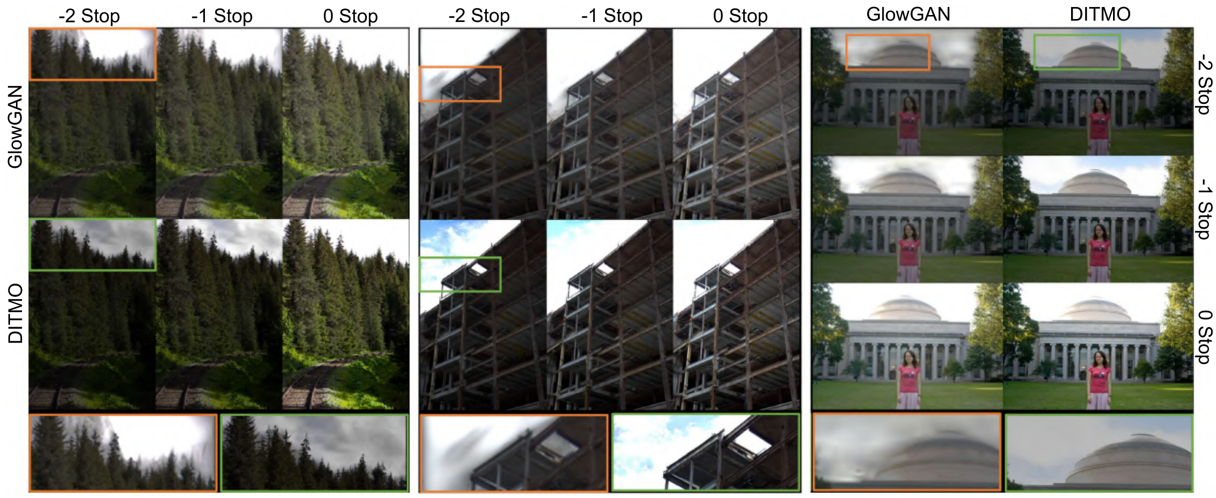

Figure 6: GlowGAN Comparisons. We present 3 different scenes expanded by GlowGAN and DITMO and their multiple exposures (-2, -1 and 0 stops). Insets highlight the qualitative differences between GlowGAN (in green) and DITMO (in orange). GlowGAN suffers from artifacts and poor detail generation.

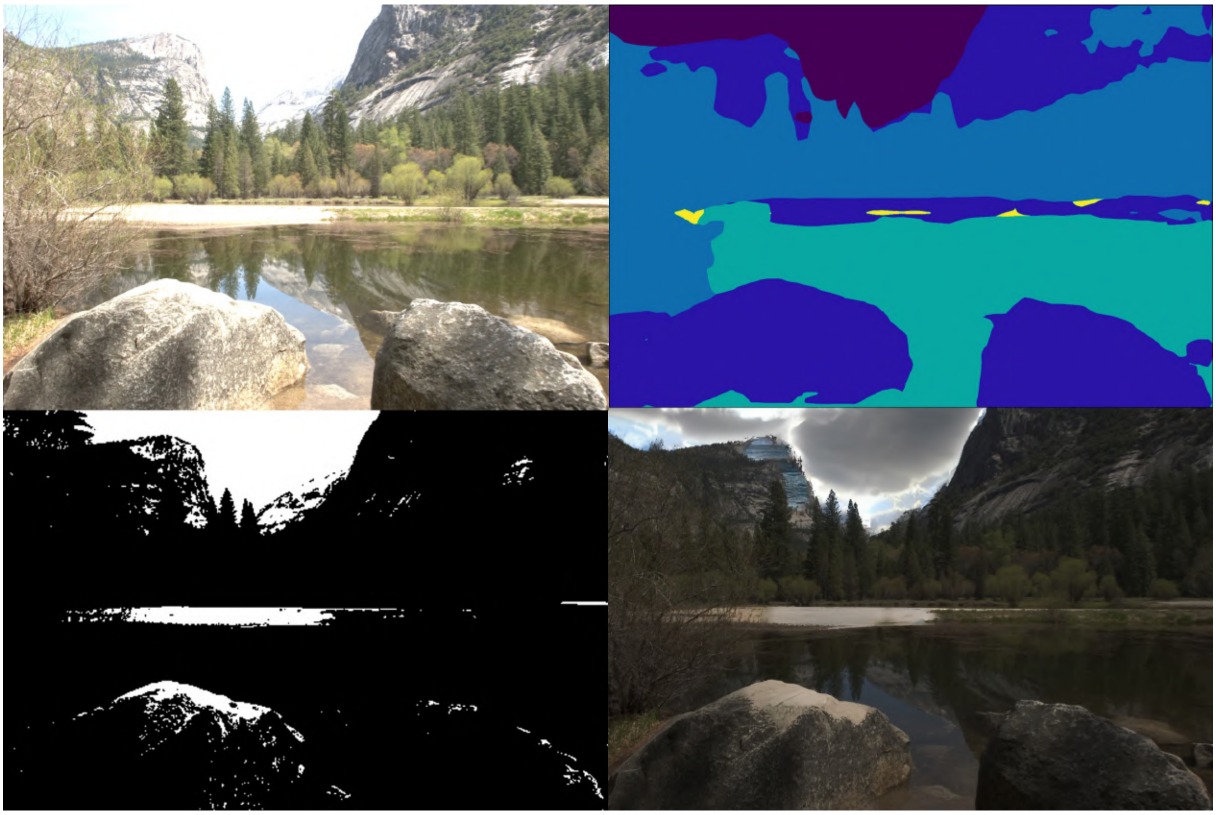

Figure 7: Failure Case 1: Improper semantic segmentation can lead to inconsistent results. The semantic segmentation mask (top-right) of the input image (top-left) shows the sky and parts of the mountain incorrectly labeled as *Sky*. As a result, the generated inpainting (bottom-right, -2 stops) using the saturation mask (bottom-left) is inconsistent. The quality of inpainting is not problematic as can be observed from the reconstruction of foreground clipped regions (boulder and sand bank) but the consistency of the background is not as desired.

Figure 8: Failure Case 2: More consistent prompts. The saturated clouds in the sky (input image - left) are inpainted using DITMO. The sky specific prompts fail to generate a good result as they fail to inpaint 'localised cloud textures' and instead inpaints sky features inside the clipped clouds (generated inpainting - right, -1 stop).

References

[1] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger, "HDR image reconstruction from a single exposure using deep cnns," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 178:1–178:15, 2017.

[2] M. S. Santos, T. I. Ren, and N. K. Kalantari, "Single Image HDR Reconstruction Using a CNN with Masked Features and Perceptual Loss," *ACM Trans. Graph.*, vol. 39, no. 4, 2020.

[3] D. Marnerides, T. Bashford-Rogers, J. Hatchett, and K. Debattista, "ExpandNet: A Deep Convolutional Neural Network for High Dynamic Range Expansion from Low Dynamic Range Content," *Comput. Graph. Forum*, vol. 37, no. 2, pp. 37–49, 2018.

[4] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 684–10 695.

[5] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 3836–3847.