
CamEdit: Continuous Camera Parameter Control for Photorealistic Image Editing (Supplementary Material)

Anonymous Author(s)

Affiliation

Address

email

1 This supplementary material provides additional details and results. Section A describes more details
2 about CamEdit50K. Section B includes extra quantitative comparisons, such as shutter speed editing
3 and joint multi-parameter control. Section C shows qualitative comparisons with other methods.
4 Section D presents more visual results to demonstrate the controllability and photorealism of our
5 CamEdit.

6 A More Details of CamEdit50K

7 Our dataset consists of pairs of base images and corresponding captured or synthetic images, with each
8 pair associated with a specific camera parameter. For real captured pairs, most images are collected
9 from the public datasets. Specifically, the focal plane pairs are primarily sourced from [2, 8, 15];
10 aperture-labeled pairs are from [2, 8, 4]; and shutter speed annotations are obtained from [19, 20, 3].
11 For synthetic pairs, we render parameter-specific synthetic images from base images. We randomly
12 sample five parameters per scene to increase diversity. Synthetic images with visual artifacts or poor
13 quality are excluded.

14 A.1 Focal Plane

15 The focal plane refers to the plane corresponding to the focal point, where light entering a lens
16 converges to form a sharp image. Accurately simulating the true focal plane using only EXIF
17 metadata is a challenge. In this paper, we apply the parameter p_f to represent the focal plane. For real-
18 world data, we define the parameter p_f as the median of the normalized depth map $D(x, y) \in [0, 1]$
19 within the sharpness mask \mathcal{M} , as shown in the equation below:

$$p_f = \text{median}\{D(x, y) \mid (x, y) \in \mathcal{M}\}, \quad (1)$$

20 where \mathcal{M} is computed using the method proposed in [5], which identifies regions of peak focus. As
21 illustrated in Figure S1, we show a real captured pair of images with corresponding sharpness masks
22 and depth maps used to compute p_f . The depth map $D(x, y)$ is estimated using Depth Anything
23 V2 [17]. For generating synthetic images, DrBokeh [16] simulates defocus effects by rendering
24 the base image based on p_f . Figure S2 shows a base image, its depth map, and the corresponding
25 synthetic image rendered with a sampled p_f .

26 A.2 Aperture

27 Aperture controls the extent of defocus blur in an image and is typically recorded in the EXIF
28 metadata as an f-number. For real images, aperture values are directly extracted when EXIF metadata
29 is available. Otherwise, we estimate them using edge-based blur kernel analysis [7] combined with
30 focal plane under a thin-lens approximation [13]. The resulting values are quantized to the nearest
31 standard f-number for consistency.

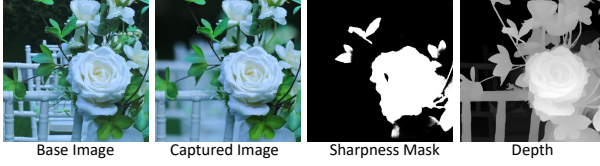


Figure S1: Focal plane estimation from real data. From left to right: base image, captured image, sharpness mask, and depth map.

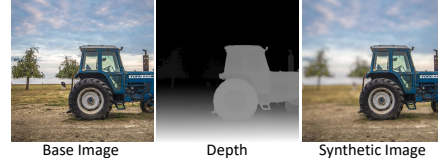


Figure S2: Synthetic rendering with sampled focal plane. From left to right: base image, depth map, and synthetic image.

For synthetic data, we simulate images using BokehMe [14] based on predicted depth maps. The focal plane is fixed on the foreground subject, while background blur is applied according to the specified aperture. To maintain foreground sharpness, we composite the blurred background with the original subject using an alpha matte generated by BRIA . AI [1].

A.3 Shutter Speed

Shutter speed, denoted as p_s , determines the duration of light integration on the camera sensor and directly influences image brightness. Following [18, 9, 10], the image formation process is modeled as:

$$L = ADC \{ \xi \cdot Clip [Poisson(p_s \cdot QE \cdot (H + \mu))] + \mathcal{N}(0, \sigma_{read}^2) \}^{1/\gamma}, \quad (2)$$

where H represents the HDR irradiance, QE is the quantum efficiency, μ is the dark current, ξ is the conversion gain, and σ_{read}^2 denotes the read noise variance. The $Clip$ and ADC operators simulate saturation and digitization respectively. The γ is the gamma correction factor.

For real data, we treat the HDR image as the latent irradiance H in Equation 2, and the corresponding images captured at different exposure levels as observed outputs L . The shutter speed p_s is estimated by inverting Equation 2.

For synthetic data, images are rendered from HDR sources by varying p_s across a predefined range using Equation 2. In addition, each image is treated as input and the corresponding HDR image as the output, with the estimated shutter speed p_s .

B Additional Quantitative Comparisons

Comparison for shutter speed editing. We conduct experiments on test images from our proposed CamEdit50K, which primarily consist of moderately low-light scenes where adjusting the shutter speed is needed to simulate long exposure. As shown in Table S1, our method achieves competitive performance compared to several existing low-light image enhancement approaches. Furthermore, as illustrated in the last two rows of Figure S4, our method effectively enhances image quality in these scenarios.

Table S1: Quantitative comparison on shutter speed editing.

Method	NIQE↓	MUSIQ↑	DINO↑
SCI [11]	5.31	44.30	0.82
CycleR2R [10]	5.08	52.35	0.87
CLODE [6]	4.06	59.98	0.91
Ours	4.12	60.04	0.91

55

Joint control of multiple parameters. We also train a unified model that edits all three camera parameters simultaneously. In this setting, we double the number of learnable parameter heads and add 5,000 synthetic pairs exhibiting multi-parameter variation. As summarized in Table S2, the joint model matches the single-parameter variants in accuracy and perceptual quality, confirming effective multi-parameter control.

Table S2: Performance of CamEdit under single-parameter vs. joint multi-parameter control.

Setting	Aperture			Focal Plane			Shutter Speed		
	NIQE↓	DINO↑	Error↓	NIQE↓	DINO↑	Error↓	NIQE↓	DINO↑	Error↓
CamEdit (Single-Parameter)	3.34	0.83	0.60	4.46	0.82	0.15	4.28	0.93	0.11
CamEdit (Joint-Parameter)	3.42	0.84	0.65	4.70	0.83	0.17	4.12	0.91	0.15

61 C Additional Qualitative Comparisons

62 **Comparison with image editing methods.** Figure S3 shows visual comparisons with state-of-the-
63 art image editing methods. Existing approaches, such as UltraEdit [21] and SuperEdit [12], lack
64 explicit control over camera parameters and fail to accurately reflect parameter variations. Although
65 a retrained version of UltraEdit (UltraEdit*) can simulate aperture changes, it remains insensitive
66 to input values and introduces undesired content. In contrast, CamEdit achieves accurate aperture
67 control while preserving consistency of scene structure and details. Both UltraEdit* and CamEdit
68 are based on the same SD3 backbone, and the performance improvements of CamEdit underscore
69 the effectiveness of our proposed modules: Continuous Parameter Prompting and Parameter-Aware
70 Modulation, as detailed in Sections 4.1 and 4.2 of the main paper, respectively.

71 **Comparison with an Image Generation Method.** We also present visual comparisons with the
72 camera-aware image generation method PhotoGen [18] in Figure S3. Due to limited scene diversity
73 in its training data, PhotoGen has difficulty achieving high content fidelity and accurate control
74 over varying camera parameters. In contrast, CamEdit, guided by both the input image and our
75 CamEdit50K dataset, achieves consistent parameter control and generates more realistic and coherent
76 results.



Figure S3: Visual comparison with diffusion-based methods under varying aperture parameters.

77 D More Visual Results

78 We provide more visual results in Figure S4 to demonstrate the controllability of focal plane, aperture,
 79 and shutter speed. **CamEdit** enables precise adjustment of camera parameters across diverse scenes
 80 and consistently generates photorealistic images with high aesthetic quality.

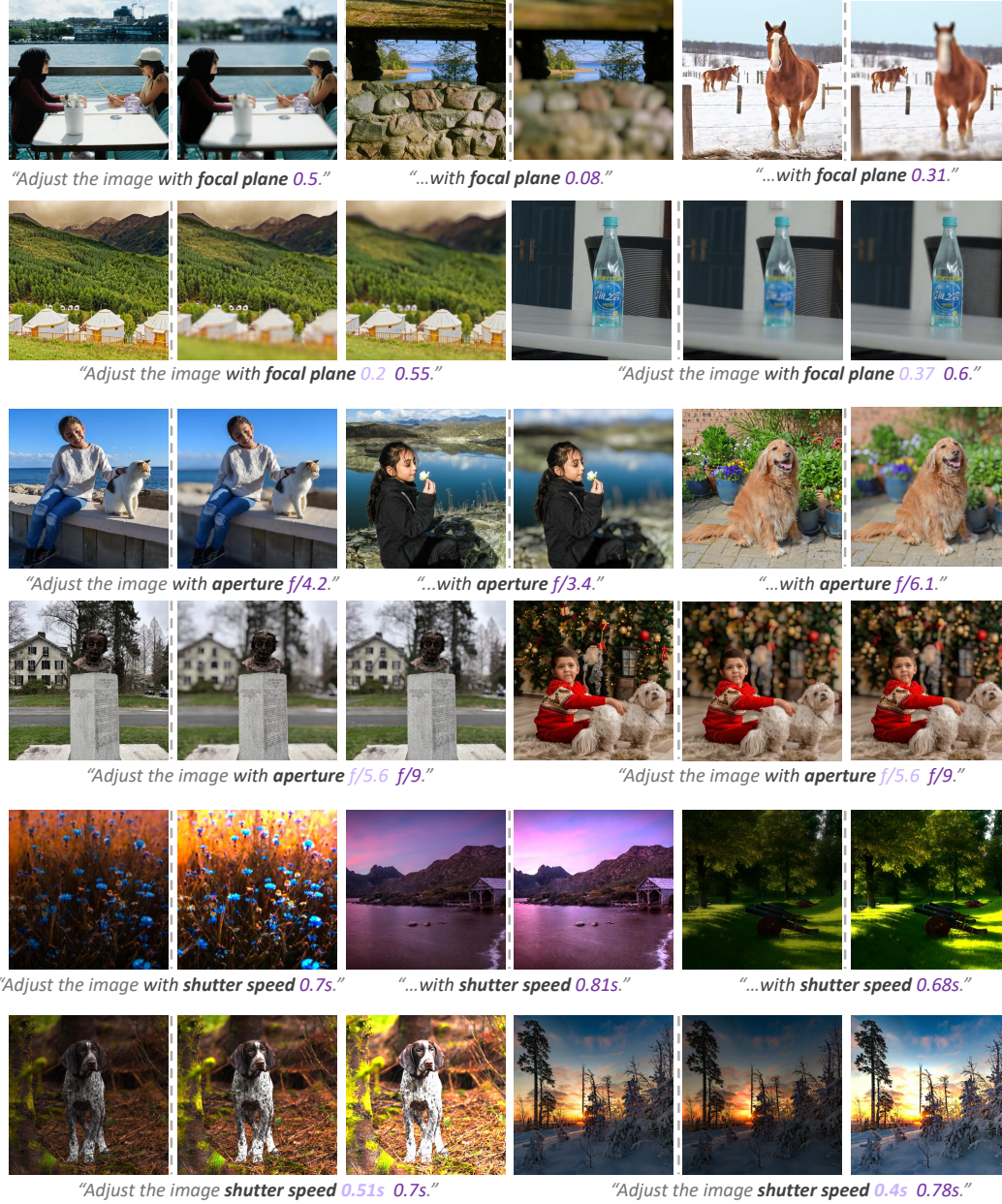


Figure S4: Additional results of CamEdit.

81 **Continuous camera parameter control.** Figure S5 shows that CamEdit supports continuous control
 82 over individual camera parameters. The model allows smooth adjustment of the focal plane from
 83 distant to near focus, gradual changes in aperture values, and progressive variations in shutter speed.
 84 These capabilities extend beyond the physical limitations of real cameras, enabling flexible and
 85 expressive parameter editing. Throughout the adjustment process, both scene structure and fine
 86 details remain consistent, demonstrating strong coherence under continuous parameter changes.

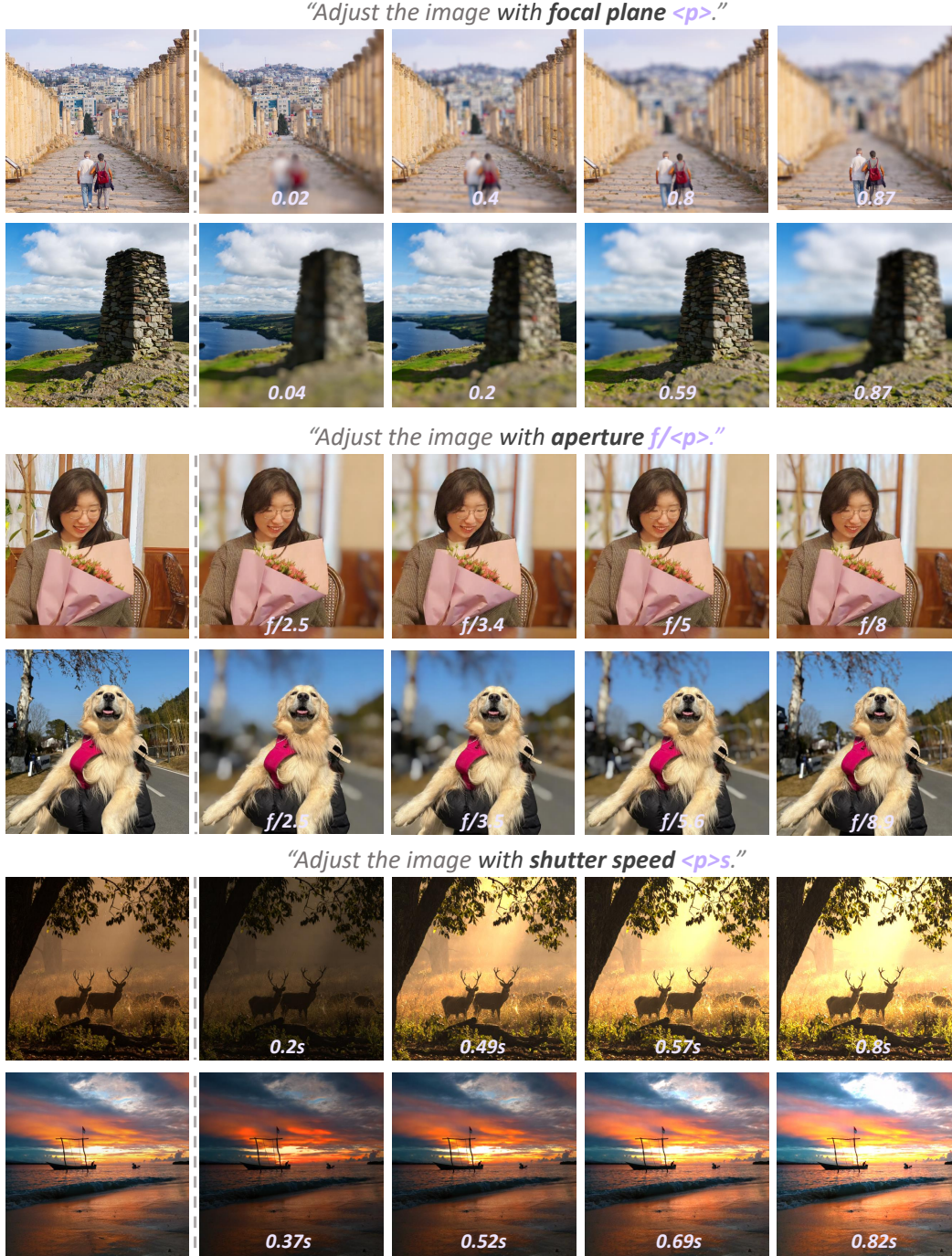


Figure S5: Additional CamEdit results demonstrating continuous control across varying camera parameters.

Progressive camera parameter editing. Figure S6 illustrates a structured editing process of CamEdit involving sequential adjustments of camera parameters. The procedure begins by modifying the focal plane to shift the focus point, followed by adjusting the aperture to vary the depth of field, and ends with shutter speed adjustment to alter the exposure level. The middle blue path in the figure indicates the main editing trajectory, starting from the input image. This step-by-step adjustment simulates real camera operations and enables fine-grained control over different scenes. The resulting images exhibit clear and continuous changes along each camera parameter dimension while preserving the consistency of scene structure and detail.

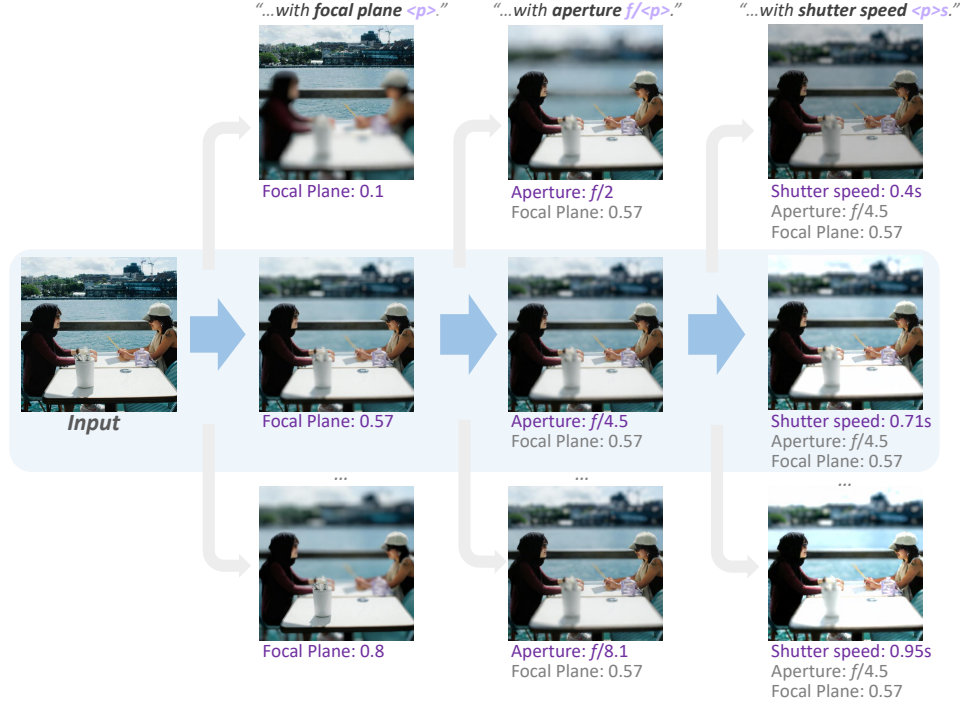


Figure S6: Progressive editing of camera parameters with CameraEdit, including focal plane, aperture, and shutter speed. Blue arrows denote the primary editing trajectory starting from the input image.

References

- [1] Bria-ai. <https://huggingface.co/briaai/RMBG-1.4>, 2024.
- [2] Abdullah Abuolaim and Michael S Brown. Defocus deblurring using dual-pixel data. In *European Conference on Computer Vision*, pages 111–126. Springer, 2020.
- [3] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018.
- [4] Kang Chen, Shijun Yan, Aiwen Jiang, Han Li, and Zhifeng Wang. Variable aperture bokeh rendering via customized focal plane guidance. *arXiv preprint arXiv:2410.14400*, 2024.
- [5] Yuxin Jin, Ming Qian, Jincheng Xiong, Nan Xue, and Gui-Song Xia. Depth and dof cues make a better defocus blur detector. In *2023 IEEE International Conference on Multimedia and Expo (ICME)*, pages 882–887. IEEE, 2023.
- [6] Donggoo Jung, Daehyun Kim, and Tae Hyun Kim. Continuous exposure learning for low-light image enhancement using neural ODEs. *International Conference on Learning Representations*, 2025.
- [7] Ali Karaali and Claudio Rosito Jung. Edge-based defocus blur estimation with adaptive scale selection. *IEEE Transactions on Image Processing*, 27(3):1126–1137, 2017.
- [8] Junyong Lee, Hyeongseok Son, Jaesung Rim, Sunghyun Cho, and Seungyong Lee. Iterative filter adaptive network for single image defocus deblurring. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2034–2042, 2021.
- [9] Zhihao Li, Ming Lu, Xu Zhang, Xin Feng, M. Salman Asif, and Zhan Ma. Efficient visual computing with camera raw snapshots. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(7):4684–4701, 2024.
- [10] Zhihao Li, Ming Lu, Xu Zhang, Xin Feng, M. Salman Asif, and Zhan Ma. Efficient visual computing with camera raw snapshots. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(7):4684–4701, 2024.

- 119 [11] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust
120 low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern
121 recognition*, pages 5637–5646, 2022.
- 122 [12] Fan Chen Xiaoying Xing Longyin Wen Chen Chen Sijie Zhu Ming Li, Xin Gu. Superedit: Rectifying and
123 facilitating supervision for instruction-based image editing. *arXiv preprint arXiv:2505.02370*, 2025.
- 124 [13] Ren Ng. Light field photography with a hand-held plenoptic camera. In *Stanford Tech Report CSTR*, 2005.
- 125 [14] Juewen Peng, Zhiguo Cao, Xianrui Luo, Hao Lu, Ke Xian, and Jianming Zhang. BokehMe: When
126 neural rendering meets classical rendering. In *IEEE/CVF Conference on Computer Vision and Pattern
127 Recognition*, 2022.
- 128 [15] Lingyan Ruan, Bin Chen, Jizhou Li, and Miuling Lam. Learning to deblur using light field generated
129 and real defocus images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern
130 Recognition*, pages 16304–16313, 2022.
- 131 [16] Yichen Sheng, Zixun Yu, Lu Ling, Zhiwen Cao, Xuaner Zhang, Xin Lu, Ke Xian, Haiting Lin, and Bedrich
132 Benes. Dr. Bokeh: Differentiable occlusion-aware bokeh rendering. In *IEEE/CVF Conference on Computer
133 Vision and Pattern Recognition*, pages 4515–4525, June 2024.
- 134 [17] Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao.
135 Depth anything v2. *Advances in Neural Information Processing Systems*, 37:21875–21911, 2024.
- 136 [18] Yu Yuan, Xijun Wang, Yichen Sheng, Prateek Chennuri, Xingguang Zhang, and Stanley Chan. Generative
137 photography: Scene-consistent camera control for realistic text-to-image synthesis. *IEEE/CVF Conference
138 on Computer Vision and Pattern Recognition*, 2025.
- 139 [19] Xingchen Zhang. Benchmarking and comparing multi-exposure image fusion algorithms. *Information
140 Fusion*, 74:111–131, 2021.
- 141 [20] Zhilu Zhang, Haoyu Wang, Shuai Liu, Xiaotao Wang, Lei Lei, and Wangmeng Zuo. Self-supervised high
142 dynamic range imaging with multi-exposure images in dynamic scenes. In *International Conference on
143 Learning Representations*, 2024.
- 144 [21] Haozhe Zhao, Xiaojian Shawn Ma, Liang Chen, Shuzheng Si, Rujie Wu, Kaikai An, Peiyu Yu, Minjia
145 Zhang, Qing Li, and Baobao Chang. Ultraedit: Instruction-based fine-grained image editing at scale.
146 *Advances in Neural Information Processing Systems*, 37:3058–3093, 2024.