Recurrent Dynamic Range Extension - Supplementary Material

SEBASTIAN DILLE, Computational Photography Lab., Simon Fraser University, Canada
KERU FU, Computational Photography Lab., Simon Fraser University, Canada
S. MAHDI H. MIANGOLEH, Computational Photography Lab., Simon Fraser University, Canada
YAĞIZ AKSOY, Computational Photography Lab., Simon Fraser University, Canada

ACM Reference Format:
Sebastian Dille, Keru Fu, S. Mahdi H. Miangoleh, and Yağız Aksoy. 2026. Recurrent Dynamic Range Extension - Supplementary Material. In *SIGGRAPH Asia 2026 Conference Papers (SA Conference Papers '26), December 01–04, 2026, Kuala Lumpur, Malaysia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3829340.3842175

We provide further details for our supervision in Section A, implementation details for the Memory Replay strategy in Section B and for our data preprocessing in Section C, as well as ablation studies of the main components and design choices of our method in Section D. We also extend qualitative evaluation of Section 4.2 of the main paper, with additional results and additional baselines in Figures 3 to 8. Further results from all methods for SI-HDR [Hanji et al. 2022] and for the full HDR photographic survey [Fairchild 2007] can be viewed with our provided HTML viewers "sihdr.html" and "itw.html".

A Supervision Details

We train our network with a combination of pixel-wise supervision and the losses and regularization from R3GAN [Huang et al. 2024], including two different discriminator modules D_E and D_R. R3GAN has been successfully used to stabilize the training and mitigate the risk of mode collapse. Following their formulation, we add a gradient penalty in the form of an L2 regularization for both real and fake images according to Equation (1) for our image discriminator D_E:

$$\mathcal{L}_{D_E} = \mathcal{L}_{adv-img} + \frac{\gamma}{2}(\mathcal{R}_{real} + \mathcal{R}_{fake}), \quad \gamma = 0.05, \tag{1}$$

$$\text{where} \quad \mathcal{L}_{adv-img} = f(-(D_E(G(z)|I_L) - D_{img}(I_E|I_L))), \tag{2}$$

$$\mathcal{R}_{fake} = ||\nabla D_E(G(z) + I_L|I_L)||^2, \tag{3}$$

$$\mathcal{R}_{real} = ||\nabla D_E(I_E|I_L)||^2, \tag{4}$$

$$\text{and} \quad f(x) = log(1 + e^x). \tag{5}$$

We use the same set of losses and gradient penalty for the residual discriminator as in Equation (1), but drop the condition:

$$\mathcal{L}_{D_R} = f(-(D_R(G(z)) - D_R(I_R))) + \frac{\gamma}{2}(\mathcal{R}_{real} + \mathcal{R}_{fake}). \tag{6}$$

Authors' Contact Information: Sebastian Dille, Computational Photography Lab., Simon Fraser University, Canada, sdille@sfu.ca; Keru Fu, Computational Photography Lab., Simon Fraser University, Canada, keru_fu@sfu.ca; S. Mahdi H. Miangoleh, Computational Photography Lab., Simon Fraser University, Canada, smh31@sfu.ca; Yağız Aksoy, Computational Photography Lab., Simon Fraser University, Canada, yagiz@sfu.ca.

SA Conference Papers '26, Kuala Lumpur, Malaysia

ACM ISBN 979-8-4007-2842-6/2026/12
https://doi.org/10.1145/3829340.3842175

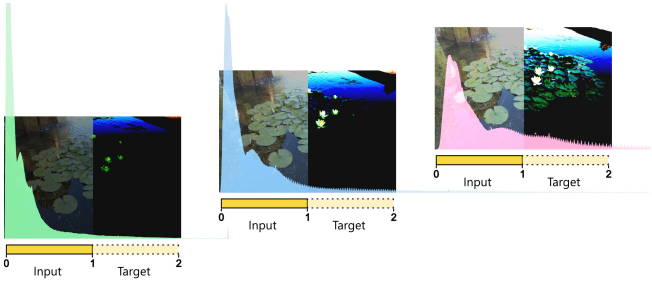

Fig. 1. We show different input images and target residuals for our extension training with their histograms. By leveraging the extended dynamic range of common RAW photos, we re-expose the image to vary the number of pixels above the clipping threshold.

Image credits: Bychkovsky et al. [2011]

The loss for our extension network as generator $\mathcal{L}_G$ is lastly a combination of pixel-wise supervision and adversarial supervision as detailed in Equation (7):

$$\mathcal{L}_G = \mathcal{L}_{adv-img} + \mathcal{L}_{pixel}, \tag{7}$$

$$\text{with} \quad \mathcal{L}_{pixel} = LPIPS(\hat{I}_E, I_E). \tag{8}$$

B Memory Replay

For our recurrent training, we adopt the concept of Memory Replay Back-Propagation [Wu et al. 2022] from neural language processing and adapt it to image inputs. In this setting, the inference chain is executed twice. We first generate estimations for the defined N steps starting from the initial input I_L^0. In this stage, no gradients are computed. We only compute the estimates for further processing and store them in a buffer in RAM to free GPU memory.

In the second run, the chain is executed in reverse order. We start with the estimation of $\hat{I}_E^N$ from the previously computed $\hat{I}_E^{N-1}$, which is now read from the buffer. By utilizing the precomputed inputs, each new estimate is still influenced by all previous steps along the chain, and the training error reflects the relative contribution accordingly. At the same time, we can now treat the steps separately for gradient computation, comparable to a batch of predictions. We accumulate the gradients from all steps and finally backpropagate through the recurrent chain once. By utilizing the buffer and traversing the chain in reverse order, we can lower the memory burden and realize recurrent training while still considering the entire prediction chain.

C Data processing

To generate training pairs for our single-EV extension, we augment the input exposure by randomly scaling the median of the RAW photo before clipping. This effectively varies the number of pixels

Table 1. We ablate the effect of our dual-discriminator setup on the SI-HDR [Hanji et al. 2022] benchmark.

Method	PSNR↑	VSI↑	VDP3↑	P-H↑	FLIP↓
OURS	36.37	**98.46**	**9.12**	33.42	49.92
w/o $\mathcal{D}_E$	did not converge				
w/o $\mathcal{D}_R$	did not converge				

Table 2. We ablate the effect of training on diverse RAW data on the SI-HDR [Hanji et al. 2022] benchmark. We compute PSNR, VSI, and P-H in the perceptual uniform PU21-space [Mantiuk and Azimi 2021].

Method	PSNR↑	VSI↑	VDP3↑	P-H↑	FLIP↓	CVVDP↑
OURS	**37.94**	**98.79**	**9.10**	**35.09**	**48.63**	**7.64**
w/o RAW data	37.71	98.76	9.09	34.63	51.19	7.30

that lie within the valid range and increases the robustness of our network for in-the-wild images or when confronted with varying dynamic ranges during recurrent inference. The [lake] example in Figure 1 shows three different training pairs created this way.

For our recurrent training scheme, we apply a similar scheme as described above, but generate enough targets from a source image to provide ground-truth for all steps in our recurrent chain. The maximum value for step n that the model can predict is 2^n. For a given HDR image I_H, we therefore randomly select a threshold between 2^n and the image maximum, clip at this threshold, and normalize to 2^n. We then create the ground-truth for intermediate steps by iteratively halving the current maximum and clipping again, until the required targets are generated.

Note that clipped pixels can occur in the ground-truth of both training phases. This does not contradict our formulation. We only extend values up to a maximum of 2, and leave brighter areas saturated. For clipped ground-truth, our augmentation reexposes the input such that it is still 1-EV apart from the target, allowing us to perform a full inference step.

For both training stages, we apply random resizing, horizontal and vertical flipping, and lastly crop the images to a resolution of 384. Training our convolutional architecture on varying input resolutions enables inference on high-resolution images, which is only bounded by the memory of the GPU.

D Ablations

We ablate the effect of our dual-discriminator setup in Table 1. As we explain in Section 3.1 of the main paper, the combination of image and residual discriminator is crucial for our method to succeed. Without one or the other, the extension network converges to a local minimum, and the training fails.

In Table 2, we show the effect of our training on diverse RAW data. We compare our results with a network trained exclusively on the HDR datasets listed in Sec. 3.2 of the main paper, and evaluate the results on SI-HDR [Hanji et al. 2022]. For all metrics, the performance improves with the RAW data, with significant gains for PSNR-H, FLIP, and the color-specific CVVDP. With the increased training diversity, our network becomes more robust for images in the wild, including a more faithful reproduction of colors.

Table 3. We ablate the effect of our recurrent training scheme on the SI-HDR [Hanji et al. 2022] benchmark.

Method	VDP3↑	FLIP↓	avg. Range↑
OURS	9.09	51.63	2^{17}
OURS - w/o recurrent training	8.56	61.78	2^{10}

Table 4. Runtime analysis of our method. We report the average run times (in seconds) of our method on the BtP-HDR test set, comparing our original setup with variable inference steps depending on the scene against a fixed number of steps.

Variant	variable steps	4 steps	per step
Time (s)	0.64	0.48	0.12

Table 3 shows the effect of our recurrent training scheme. For this evaluation, we keep training the single-EV extension model without adding the recurrent training and compare the results after 30 additional epochs. After this short training time, the improvement from our recurrent training already becomes visible. The network learns to further extend the dynamic range and produces more realistic HDR images, as shown by the numbers on HDR-VDP3 and Flip. We show the improved extension performance on an example image in Figure 2.

E Runtime

The inference time of our recurrent formulation largely depends on the number of recurrent inference steps, which in turn depend on the content of the scene. We report the average runtime per image on the BtP-HDR benchmark with a resolution of 1200x1200 in Table 4. Our regular framework achieves an average runtime of 0.64 seconds per image. For a more general insight, we adapt our original pipeline to execute four inference steps fixed for every image, independent of the clipping in the scene. In this case, the inference time drops to 0.48 seconds, or 0.12 seconds per step. Even for 4K images, the average time excluding I/O only slightly increases to 0.57 seconds, while occupying roughly 8Gib of GPU memory in total, mostly by the data itself. This makes it well-suited for professional image editing applications.

F Extended Quantitative Evaluation

We extend our quantitative evaluation from Section 4.1 with FID results on SI-HDR [Hanji et al. 2022] in Table 5. While FID has become a popular metric for evaluating image quality, it is not specifically designed for HDR data. We thus follow the evaluation protocol from Wang et al. [2025] and apply tone-mapping [Drago et al. 2003] before randomly creating 60 crops per image, with a crop size of 128. Next to the randomly sampled crops, we create another set of crops sampled specifically around the highlights to match our application scenario. Our method achieves state-of-the-art results for both sets.

Notably, the generative methods perform worse than most other baselines. The tested diffusion-based networks, in comparison, are

Fig. 2. We show the effect of our recurrent training scheme. With single-stage training (top), the model extends the input once but produces no significant difference for further iterations. As we increase the number of training stages, it learns to further extend the dynamic range (middle) and to add meaningful details in the highlights (bottom). We have decreased the image brightness for visualization. Image credits: Fairchild [2007]

solving a harder task for our application as they are always recreating the full image, including well-exposed pixels. This increased the risk of prediction artifacts in the result. For BracketDiffusion, another possible reason is the low inference resolution, which affects the gradient-based FID evaluation. In the case of LEDiff, lower inference resolution can be an explanation as well. Yet, we also notice a difference between the FID results in our evaluation and their reported numbers in their paper, despite following their description and using their public implementation with the provided weights for inference.

G Extended Qualitative Evaluation

We demonstrate additional results of our qualitative evaluation in Section 4.2 of the main paper. We show the performance on highlight reconstruction in Figure 4, complementary to Figure 5 in the main paper, exchanging the baselines for the different examples. The overall result remains the same. None of the baselines manages to recover the specular highlights of the glass or on the plane. None successfully recovers the sun. Different from MaskHDR and our own method, HDRCNN and IntrinsicHDR both create color artifacts for the clipped area on the owl.

We show an extended demonstration of the consistency of our method across multiple exposures from Figure 6 of the main paper in Figure 5. We show the result of LEDiff for the [City] image and the GaSLight estimation for the [Sewing] example. The performance of GaSLight is consistent with the other baselines in failing to reproduce the specular appearance of the metal pin for [Sewing]. LEDiff for the [City] image produces an overall lower contrast. Only the brightness of the street lamp is increased, while the building walls contain artifacts. This becomes visible once more in a third example we show for all methods in the [Vegas] example, where LEDiff and GaSLight both modify the illuminated letters. This reduces not only their luminance incorrectly but also renders them unreadable. In contrast, our method preserves the information in the image for every extension step, enforced by our residual formulation.

Lastly, we extend our examples from Figure 6 of the main paper with results from IntrinsicHDR, DiffHDRSyn, and LEDiff in Figure 3. Most obvious, DiffHDRSyn produces strong artifacts for the skin

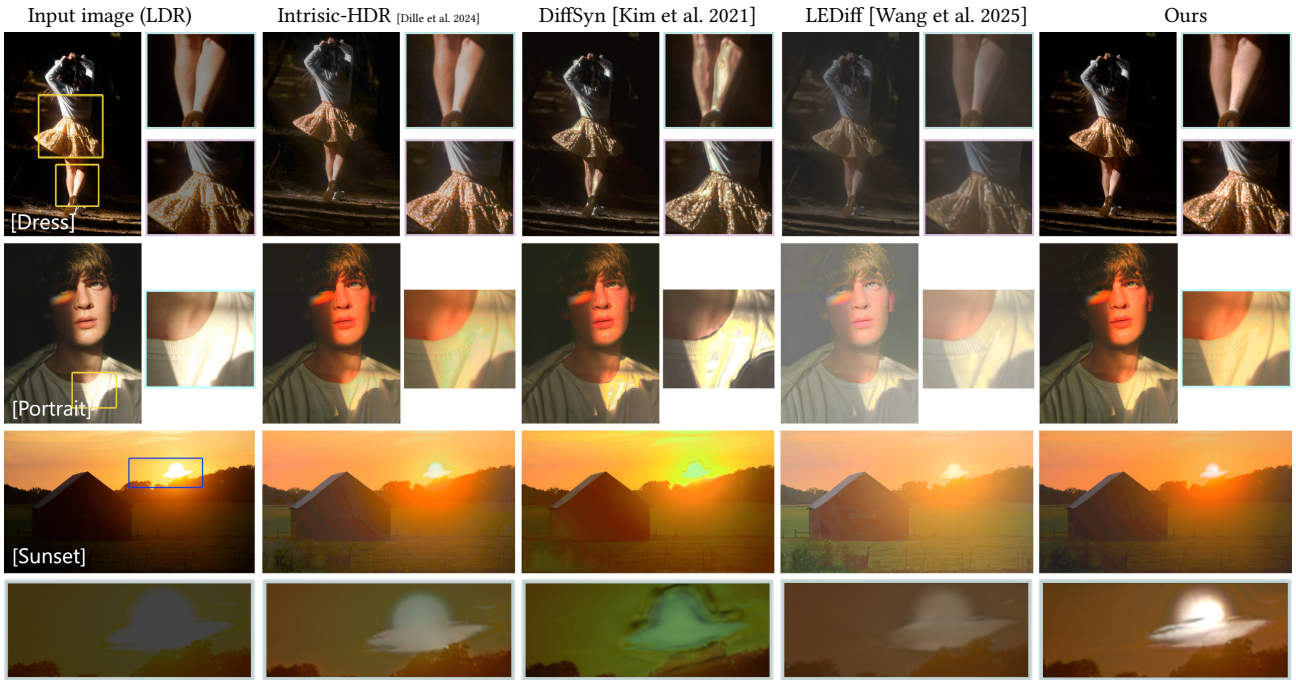

Fig. 3. Our method faithfully recovers clipped areas in images in the wild with colors that match the environment. This serves as an extension to Figure 7 of the main paper with additional baselines. The full images are tone-mapped. Image credits: @Matthew Ball, @gryffyn m, @Elmer Cañas

Table 5. FID↓ results on the clip-95 set from SI-HDR [Hanji et al. 2022]. Following Wang et al. [2025], we apply tone-mapping [Drago et al. 2003] on the results and sample 60 crops per image with a crop size of 128.

Method	random sample	smart sample
IntrinsicHDR [Dille et al. 2024]	20.98	28.20
HDR-CNN [Eilertsen et al. 2017]	24.53	<u>22.45</u>
ExpandNet [Marnerides et al. 2018]	24.30	24.48
Single-HDR [Liu et al. 2020]	<u>19.91</u>	23.02
Mask-HDR [Santos et al. 2020]	27.66	25.85
HDRUnet [Chen et al. 2021]	35.81	33.74
DrTMO [Endo et al. 2017]	36.50	30.11
Multi-Exp Gen. [Le et al. 2023]	44.04	30.29
CEVR [Chen et al. 2023]	29.50	34.25
DiffSyn [Kim et al. 2021]	26.94	32.39
BracketDiffusion [Bemana et al. 2025]	127.41	130.83
LEDiff [Wang et al. 2025]	74.25	74.51
GaSLight [Bolduc et al. 2025]	24.92	39.03
OURS	**18.53**	**21.95**

tones in [Dress], the clipped shirt, and the sun in [Sunset]. Our recurrent training scheme increases the robustness against this type of effect. We provide further qualitative results in Figures 6 to 8 and in the provided zip files. Our method produces consistent results over a wide range of scenarios.

References

Mojtaba Bemana, Thomas Leimkühler, Karol Myszkowski, Hans-Peter Seidel, and Tobias Ritschel. 2025. Bracket diffusion: Hdr image generation by consistent ldr denoising. *Comput. Graph. Forum* 44, 2 (2025).

Christophe Bolduc, Yannick Hold-Geoffroy, and Jean-François Lalonde. 2025. GaSLight: Gaussian Splats for Spatially-Varying Lighting in HDR. In *Proc. CVPR.*

Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. 2011. Learning Photographic Global Tonal Adjustment with a Database of Input / Output Image Pairs. In *Proc. CVPR.*

Su-Kai Chen, Hung-Lin Yen, Yu-Lun Liu, Min-Hung Chen, Hou-Ning Hu, Wen-Hsiao Peng, and Yen-Yu Lin. 2023. CEVR: Learning Continuous Exposure Value Representations for Single-Image HDR Reconstruction. In *Proc. ICCV.*

Xiangyu Chen, Yihao Liu, Zhengwen Zhang, Yu Qiao, and Chao Dong. 2021. HDRUnet: Single image HDR reconstruction with denoising and dequantization. In *Proc. CVPR.*

Sebastian Dille, Chris Careaga, and Yağız Aksoy. 2024. Intrinsic Single-Image HDR Reconstruction. In *Proc. ECCV.*

F. Drago, K. Myszkowski, T. Annen, and N. Chiba. 2003. Adaptive Logarithmic Mapping For Displaying High Contrast Scenes. *Comput. Graph. Forum* 22, 3 (Sept. 2003).

Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K. Mantiuk, and Jonas Unger. 2017. HDR Image Reconstruction from a Single Exposure Using Deep CNNs. *ACM Trans. Graph.* 36, 6 (2017).

Yuki Endo, Yoshihiro Kanamori, and Jun Mitani. 2017. Deep Reverse Tone Mapping. *ACM Trans. Graph.* 36, 6 (2017).

Mark D Fairchild. 2007. The HDR photographic survey. In *Color and imaging conference.* Society of Imaging Science and Technology.

Param Hanji, Rafal Mantiuk, Gabriel Eilertsen, Saghi Hajisharif, and Jonas Unger. 2022. Comparison of Single Image HDR Reconstruction Methods — the Caveats of Quality Assessment. In *ACM Trans. Graph.*

Nick Huang, Aaron Gokaslan, Volodymyr Kuleshov, and James Tompkin. 2024. The gan is dead; long live the gan! a modern gan baseline. *Proc. NeurIPS* (2024).

Junghee Kim, Siyeong Lee, and Suk-Ju Kang. 2021. End-to-end differentiable learning to HDR image synthesis for multi-exposure images. In *Proc. AAAI.*

Phuoc-Hieu Le, Quynh Le, Rang Nguyen, and Binh-Son Hua. 2023. Single-image hdr reconstruction by multi-exposure generation. In *Proc. WACV.*

Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. 2020. Single-image HDR reconstruction by learning to reverse the camera pipeline. In *Proc. CVPR.*

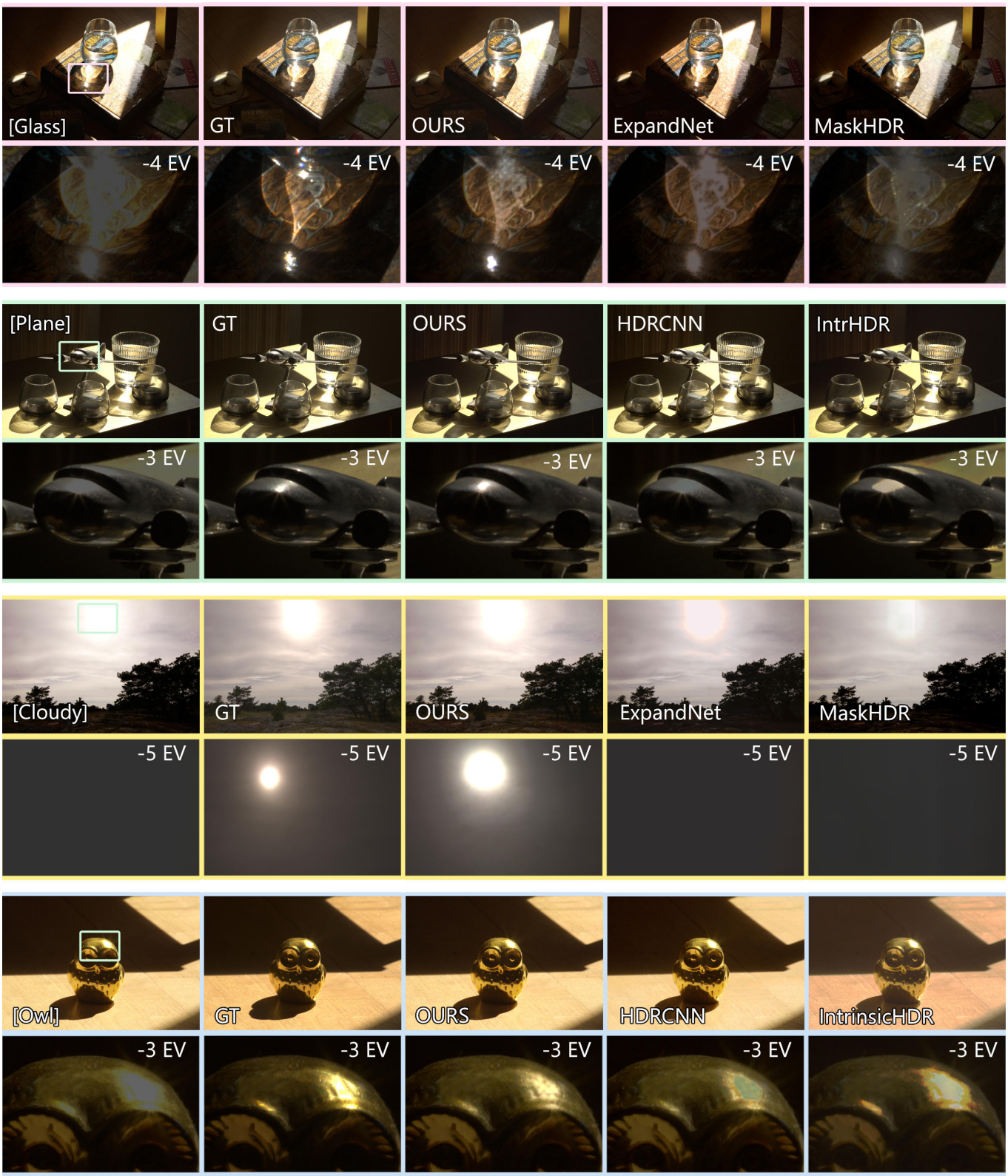

Fig. 4. We show results of our method against the strongest baselines on the highlight-specific P-H metric. Our method recovers plausible highlights on SI-HDR [Hanji et al. 2022], where other methods saturate early or create artifacts.

Rafal K Mantiuk and Maryam Azimi. 2021. PU21: A novel perceptually uniform encoding for adapting existing quality metrics for HDR. In *Proc. PCS*.
Demetris Marnerides, Thomas Bashford-Rogers, Jonathan Hatchett, and Kurt Debattista. 2018. Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. *Comput. Graph. Forum* 37, 2 (2018).
Marcel Santana Santos, Tsang Ing Ren, and Nima Khademi Kalantari. 2020. Single Image HDR Reconstruction Using a CNN with Masked Features and Perceptual Loss. *ACM Trans. Graph.* 39, 4 (2020).
Chao Wang, Zhihao Xia, Thomas Leimkuhler, Karol Myszkowski, and Xuaner Zhang. 2025. LEDiff: Latent Exposure Diffusion for HDR Generation. In *Proc. CVPR*.
Qingyang Wu, Zhenzhong Lan, Kun Qian, Jing Gu, Alborz Geramifard, and Zhou Yu. 2022. Memformer: A Memory-Augmented Transformer for Sequence Modeling. In *Proc. IJCNLP*.

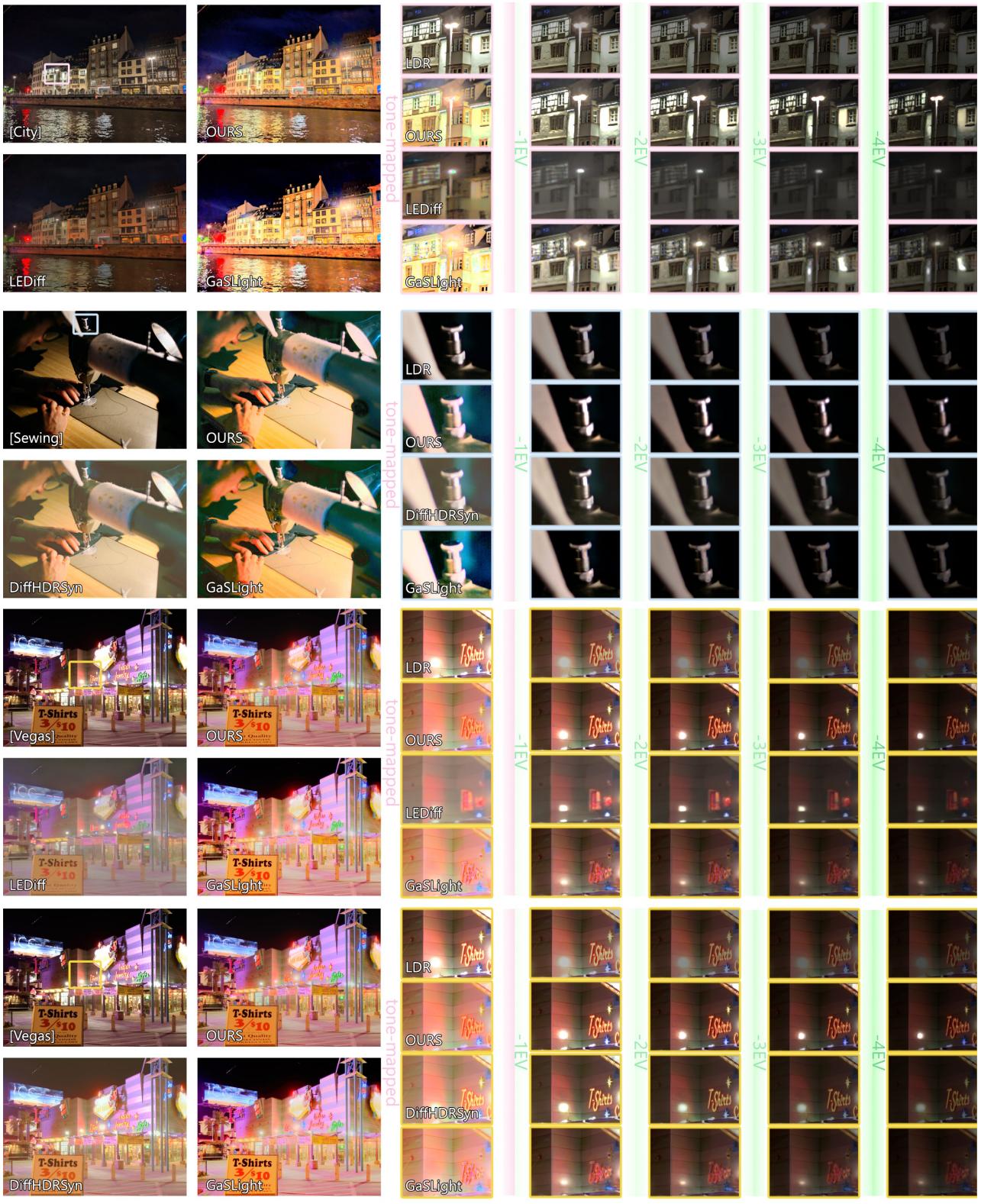

Fig. 5. We compare our results with multi-exposure generation methods as an extension to Figure 6 of the main paper.
Image credits: Death to the Stock Photo, Fairchild [2007]

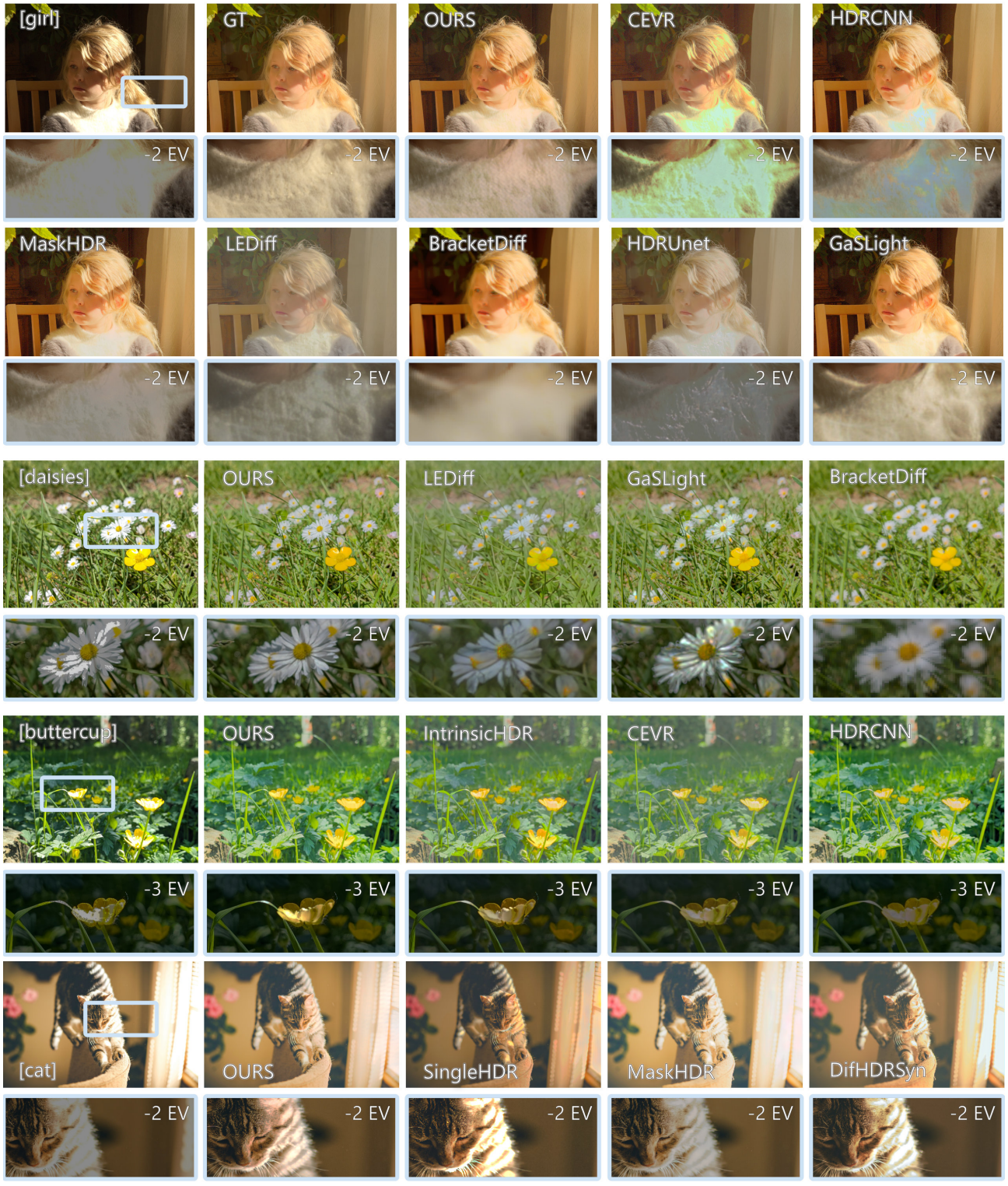

Fig. 6. We show the reconstruction of challenging clipped areas in everyday photographs as an extension to Fig. 7 of the paper. The full images are tone-mapped.
Image credits: Hanji et al. [2022], @Kaiwen Sun

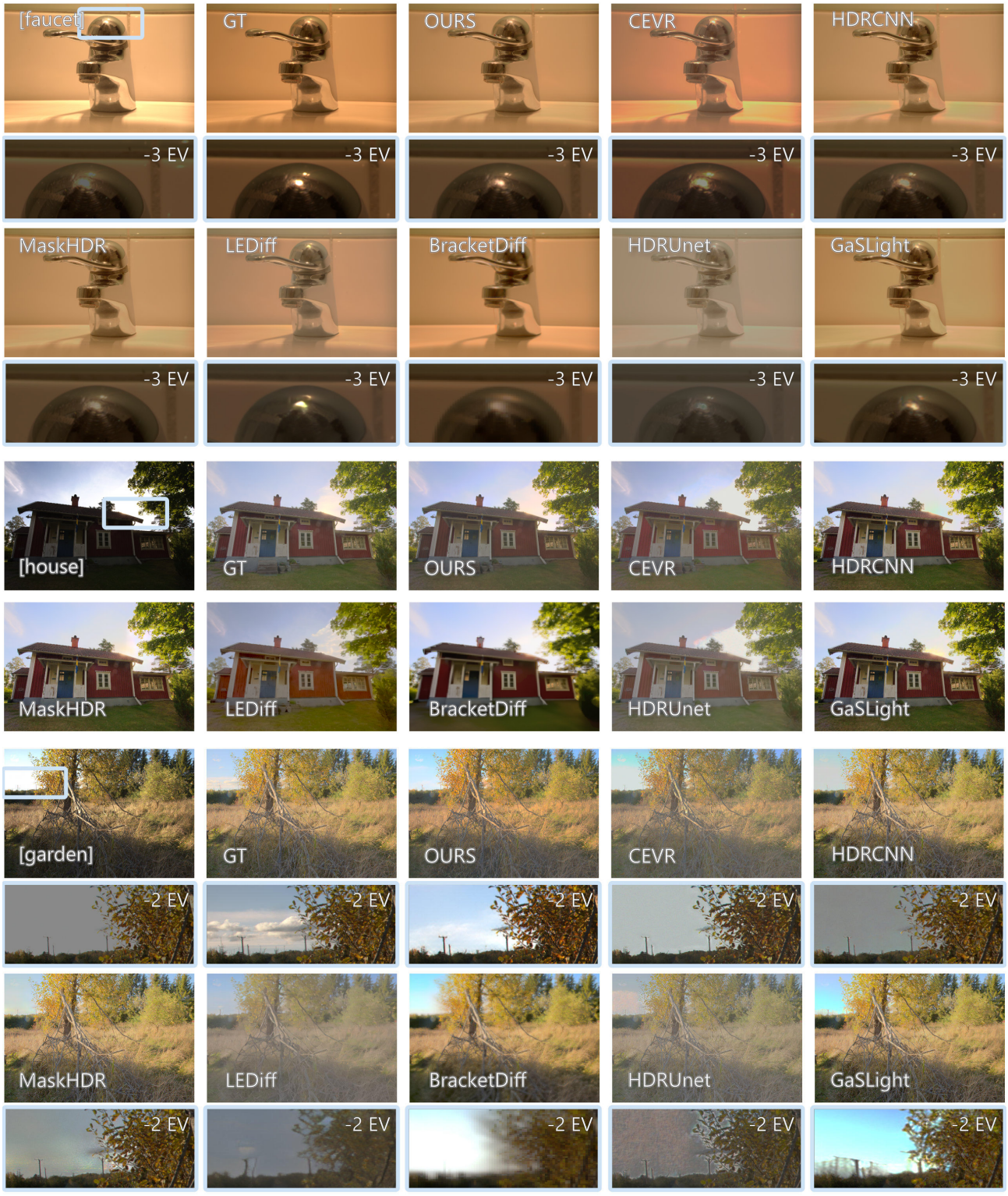

Fig. 7. We show additional examples comparing the highlight reconstruction across all approaches. Images are tone-mapped for visualization.
Image credits: Hanji et al. [2022]

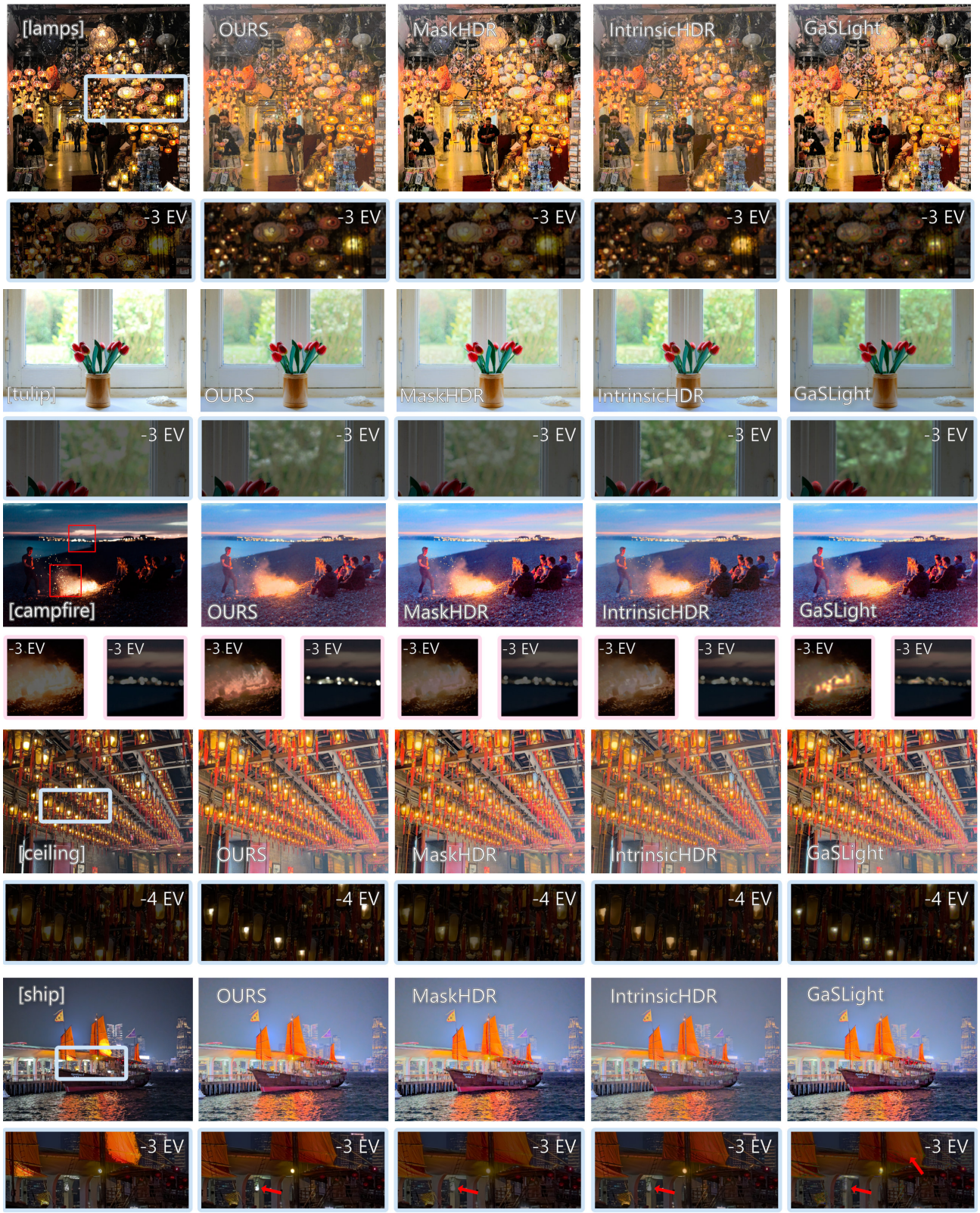

Fig. 8. We show additional examples for highlight reconstruction across all approaches on images in the wild. Images are tone-mapped for visualization. Image credits: @Svetlana Gumerova, @Colin Maynard, Death to Stock Photo