

Recurrent Dynamic Range Extension

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

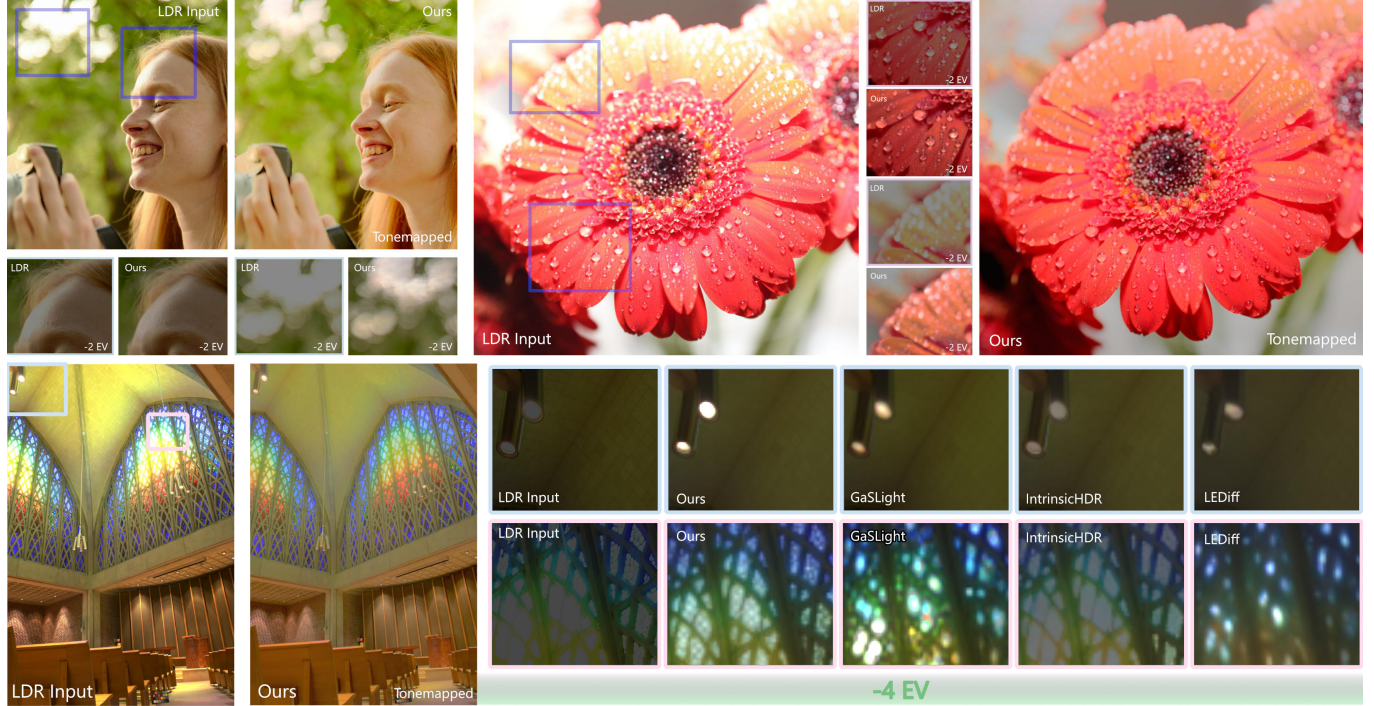


Fig. 1. We present a recurrent highlight reconstruction approach that progressively expands the dynamic range of a single LDR image. By iteratively extending the range, our method recovers missing information in severely clipped regions, including specular highlights, strong light sources, and reflective surfaces. Trained on RAW data in addition to HDR ground-truth, our model generalizes well across diverse scenes and outperforms prior methods in restoring high-intensity details while maintaining consistent structure and color fidelity under large exposure changes. Results are tone-mapped for visualization. Image credits: @Darina Belonogova, Fairchild [2007]

We present an approach to progressively extend the highlights of an image. Instead of reconstructing the full dynamic range of a complex scene directly, we learn a simpler task first: We extend the dynamic range of an input image by a single exposure value. Once this is mastered, we retrieve the full HDR image for the scene by executing our network recurrently, progressively increasing the dynamic range of the input. Our formulation is agnostic to the input dynamic range and targets a bounded output domain. This enables

Authors' Contact Information: 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.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

SA Conference Papers '26, Kuala Lumpur, Malaysia

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2842-6/2026/12

<https://doi.org/10.1145/3829340.3842175>

us to use widely available RAW images for the reconstruction task and adapt adversarial losses to construct realistic images. By incorporating Memory Replay for backpropagation, we can train our network recurrently over multiple inference stages and reduce reconstruction errors. As a consequence, our system reconstructs challenging long-tailed HDR scenes robustly and shows powerful recovery of bright light sources and highlights.

CCS Concepts: • **Computing methodologies** → **Computational photography**.

ACM Reference Format:

Sebastian Dille, Keru Fu, S. Mahdi H. Miangoleh, and Yağiz Aksoy. 2026. Recurrent Dynamic Range Extension. In *SIGGRAPH Asia 2026 Conference Papers (SA Conference Papers '26)*, December 01–04, 2026, Kuala Lumpur, Malaysia. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3829340.3842175>

1 Introduction

Natural scenes often exhibit an extremely wide dynamic range. Radiance values in reflective regions, specular highlights, or visible



Fig. 2. Starting from a single LDR input, we iteratively apply our single-step dynamic range extension network. After each iteration, the predicted image serves as input again to further expand the dynamic range. This recurrent process progressively restores clipped regions and is repeated until the image is free of clipping and the radiance range of the scene is recovered. Results are tone-mapped for visualization. Image credits: SI-HDR benchmark [Hanji et al. 2022]

light sources can be several orders of magnitude larger than the rest of the scene, leading to unbounded long-tailed intensity distributions. Modern imaging sensors are incapable of capturing the full dynamic range of natural scenes in a single photo. In common photographs, when the radiance in a scene exceeds the sensor capacity, high-intensity regions are clipped, producing low-dynamic-range (LDR) images that lack information in bright areas.

Data-driven approaches provide a promising solution for inferring the missing content from casual observations. However, this relies on large and diverse training data to build a strong prior of the radiance in natural scenes. The necessary high dynamic range (HDR) ground truth is typically obtained through multi-exposure bracketing, which requires static scenes, precise alignment, and tripod-mounted cameras. This capturing process is time-consuming and error-prone, significantly limiting the scale and diversity of available real-world HDR data. Thus, existing datasets remain scarce and lack the rich diversity needed to train generalizable models.

At the same time, the long-tailed radiance distribution of HDR scenes poses a challenge for common image-based losses, such as L1 or L2. The high luminance of specular highlights or active light sources can lead to large prediction errors in early training and exploding gradients during backpropagation. The corresponding large weight updates destabilize the training process, since the optimization is likely to oscillate or diverge. Consequently, directly regressing the HDR image requires tone mapping to compress the luminance and stabilize the training. While highlight compression is principally similar to the behavior of the human visual system, it is prone to reducing physical accuracy: The compression hinders local differentiation within the highlight regions, and small inaccuracies in this compressed space translate into large errors in the linear radiance values. For these reasons, recovering the clipped highlights from a single LDR input is a challenging problem.

In this paper, we propose learning a much simpler task first: Given a linear clipped input image, we train a neural network to extend its highlights by a single exposure value (EV). The clipped input $\in [0, 1]$ becomes an extended image $\in [0, 2]$, independent of the nominal exposure value of the input and its initial dynamic range. The network learns to restore the clipped values of saturated pixels, while bright highlights remain clipped if they exceed the new bounds. This single-EV extension is exposure invariant and constrained by design. Since the output is bounded $\in [0, 2]$, we can train our network in *linear space* without risking exploding gradients. It allows us to

omit any compression of the target domain and employ common perceptual losses to enhance realism and preserve visual fidelity. Likewise, extending a clipped linear image to a slightly extended dynamic range can be trained on diverse RAW photo collections. While RAW images do not capture the full dynamic range of a scene, they preserve significantly more radiometric information than standard 8-bit LDR images. We can hence augment our training data by leveraging the exposure-invariant formulation and derive multiple training pairs from a single re-exposed RAW photo. Together, this provides our extension network with a rich and diverse training distribution to enable generalization to in-the-wild photographs.

Crucially, we can move beyond the bounds of our single-EV extension by applying our network recurrently over multiple inference steps. This enables us to progressively recover the full long-tailed distribution of HDR highlights by using the output of one step as input to the next. For this purpose, we implement a recurrent training scheme with supervision from real HDR images. Our network thus expands its learned prior from RAW-based training with the continuous highlight distribution from HDR data. It learns to extend the first and the n -th input of the recurrent execution adaptively, while compensating for prediction errors in previous inference steps.

Our contributions are as follows:

We present a novel exposure-invariant recurrent highlight extension that effectively tackles the main challenges for single-image highlight reconstruction. The unbounded scene luminance is broken down into clearly constrained, separate chunks, which can be trained on RAW images to improve generalization. We further incorporate a recurrent training scheme on HDR images to enable backpropagation for our full ResNet-like architecture over multiple inference steps. With our recurrent training, we can combine the separate steps into a consistent reconstruction of the long-tailed luminance. We show examples in Figure 1. Our method faithfully recovers the skin tone of the woman as well as the bokeh pattern in the background. On the flower, it recovers the color visible through the raindrops while preserving the bright reflections. For the church, it predicts accurate spotlights and a bright sunlit window. Our resulting luminance for the spotlights thereby still exceeds the displayable dynamic range, even with an exposure correction of 4 EV. As we will further demonstrate in Section 4, this design enables accurate and generalizable highlight reconstruction in high resolution for real-world, in-the-wild images. We show an overview of our pipeline in Section 2, and describe the inference and training details in Section 3.

2 Related Work

Inverse tone mapping. Reconstructing clipped areas is a long-standing research problem. Earlier approaches [Abebe et al. 2018; Banterle et al. 2006, 2008; Didyk et al. 2008; Masia et al. 2009; Masood et al. 2009; Rempel et al. 2007] aim to create HDR intensities in the highlights with local-adaptive [Banterle et al. 2006, 2008; Didyk et al. 2008; Kovaleski and Oliveira 2009; Rempel et al. 2007] or global brightness scaling [Masia et al. 2009], heuristically determining the target intensity. Some additionally fill in color via cross-channel correlation [Abebe et al. 2018; Masood et al. 2009]. While enhancing the LDR input, these methods do not reproduce details and struggle to generalize to natural scenes.

LDR-to-HDR regression. With data-driven methods, a straightforward approach to highlight reconstruction is to predict the full dynamic range of the scene directly, including enhanced shadow areas. Such methods operate either in a single step [Chen et al. 2021; Eilertsen et al. 2017; Guo and Xiuhua 2022; Marnerides et al. 2021, 2018; Santos et al. 2020; Yu et al. 2021; Zou et al. 2023] or by breaking down the reconstruction into different sub-tasks [Dille et al. 2024; Liu et al. 2020; Zhang and Aydin 2021]. In both settings, the unbounded range of the highlights poses the main challenge for the optimization. Existing approaches consequently operate in log-space [Eilertsen et al. 2017; Liu et al. 2020; Santos et al. 2020; Yu et al. 2021] or apply tone-mapping [Chen et al. 2021; Dille et al. 2024; Guo and Xiuhua 2022; Marnerides et al. 2021, 2018] to compress the dynamic range. Additionally, direct regression relies exclusively on HDR data for training to learn the typical long-tailed distribution. With the limited availability of HDR datasets and the additional compression during training, the methods do not generalize well to images in the wild and tend to saturate early for specular highlights or visible light sources. Our method focuses specifically on the challenging highlights. It operates within a bounded linear domain, avoiding outliers in the loss function and exploding gradients. Our network faithfully learns a physically accurate HDR distribution without requiring any nonlinearities for supervision. Most importantly, we combine the diversity of RAW image collections and the full radiance information from HDR datasets to train a model that generalizes to everyday photographs through our 2-step training scheme.

Exposure-stack reconstruction. A different direction is the reconstruction of exposure stacks from a single image [Bolduc et al. 2025; Chen et al. 2023; Endo et al. 2017; Jo et al. 2021; Kim et al. 2021; Le et al. 2023; Lee et al. 2018a,b, 2020], which can later be merged into an HDR image using established multi-exposure merging [Debevec and Malik 1997; Mann and Picard 1995]. Closely related to our approach, these methods create differently exposed *LDR* images from a given input. This circumvents the challenge of the unbounded target domain and enables the use of common losses without additional compression. Early approaches [Endo et al. 2017; Lee et al. 2018a] are trained to predict a fixed set of output exposures simultaneously. They are restricted to a defined dynamic range and require architectural changes and retraining on HDR data to adapt to other scenes. Succeeding works estimate the exposures individually, either by conditioning the estimation on a given target exposure [Chen et al. 2023; Kim et al. 2021] or only differentiating between increasing

and decreasing the exposure in a recursive inference setup [Bolduc et al. 2025; Jo et al. 2021; Lee et al. 2018b]. The iterative execution of the same network, however, is prone to accumulated estimation errors and severe reconstruction artifacts.

Critically, these methods do not *extend* the dynamic range but rather create LDR outputs with *shifted* upper and lower bounds. Being supervised on stacked LDR images, the networks learn not only to fill in pixels but also to *remove* valid content from the input to create new under- or over-exposed areas. This behavior is encouraged by cycle-consistent losses [Chen et al. 2023; Lee et al. 2020] but increases the risk of reconstruction errors by removing valuable information. As a result, the methods perform worse on HDR benchmarks such as SI-HDR [Hanji et al. 2022] when compared to direct-regression approaches. In contrast, our approach progressively *extends* the dynamic range, reducing the ambiguity at each step without removing valuable information. Our exposure-invariant formulation enables our network to adapt the dynamic range in a *scene-specific* manner, instead of using ensembles specialized for individual steps [Lee et al. 2018a] or conditioning on the arbitrary range of a target HDR image [Lee et al. 2018b]. Our *recurrent training scheme* allows us to adaptively reconstruct the dynamic range of the highlights without restrictions from the architecture.

Generative approaches. Recently, approaches have been introduced that adapt the multi-exposure generation practice to fine-tune generative diffusion models [Bemana et al. 2025; Bolduc et al. 2025; Wang et al. 2025b]. They benefit from their strong pretrained image prior and apply the constrained exposure-estimation framework to directly generate HDR images. For HDR reconstruction, these approaches fill in clipped areas more effectively, where non-generative methods lack details. At the same time, their prior is learned on *non-linear* data and shows a distribution mismatch to the HDR domain. The models are prone to modifying the image content and creating artifacts. The generative capabilities come at the cost of large networks and a minute-long inference process, while the image resolution is limited. Our recurrent extension is independent of the architecture. Using fully-convolutional networks and training in the *linear* domain, it shows strong generalization to diverse scenarios under challenging settings. We can operate at high resolutions, reconstructing the highlights of a 4k image in half a second, while remaining faithful to the original image.

Recurrent training in computational photography. Recurrent approaches are typically used to model time series in low-dimensional regimes like NLP [Cho et al. 2014; Hochreiter and Schmidhuber 1997; Shiva Prakash et al. 2018]. Their training is often limited due to the high computational burden of backpropagating gradients over multiple inference stages. For computational photography applications, existing approaches remain scarce [Cao et al. 2023; Hidalgo-Carri  et al. 2020; Shi et al. 2020; Teed and Deng 2020; Wang et al. 2017] and rely on low resolutions and shallow recurrent modules to reduce complexity. The same applies to existing HDR reconstruction approaches [Kim et al. 2021], limiting their applicability for professional image editing. In our training, we adapt the Memory Replay strategy [Wu et al. 2022] from NLP literature to the image domain, reducing the compute for recurrent backpropagation significantly. It allows us to train our full ResNet-based network recurrently without sacrificing resolution or network capacity.

3 Recurrent Dynamic Range Extension

We formulate the reconstruction of clipped highlights as an iterative process of multiple bounded extensions of the input dynamic range: We start with a clipped input image $I_L \in [0, 1]$ in linear RGB space. At each inference step, our core extension model is tasked to increase the luminance range of the input by exactly one EV, yielding an extended image $I_E \in [0, 2]$. Instead of directly estimating the extended image, we design our extension network to predict a residual image I_R , and compute the extended result by adding this residual to the input, i.e., $I_L + I_R = I_E$. We visualize this process in Figure 3. Our formulation provides a straightforward mechanism for the network to recover information clipped in I_L , rather than redundantly recreating content that is already present in the input. By only estimating the missing information, the residual formulation naturally preserves other image regions and also keeps the network output bounded within $[0, 1]$.

In our recurrent pipeline, we repeat this process multiple times, dividing the output by two and providing it as input to the same network, until the result is free of clipping and the full luminance range is restored. Lastly, we revert the repeated normalization and use a soft mask M to linearly blend our result over the input image to compensate for any reproduction artifacts in the shadows, with $M = \max(0, I_L - 0.5)/0.5$. This yields the final HDR estimate $\hat{I}_H$ of our method. Note that at inference, the number of network iterations is not fixed and fully depends on the content of the scene. We stop when our network does not leave clipped pixels in its result, i.e., when the maximum value of the output does not reach our empirical threshold of 1.8.

3.1 Learning Single-EV Extension

Our extension mechanism relies on a robust prior over diverse natural images. For this purpose, we generate input and target pairs I_L and I_E that are one EV apart directly from RAW data collections. RAW images represent the linear irradiance captured by the camera without any display-referred compression. With the common 14-bit signal processing of modern sensors, they offer a wider dynamic range than the final LDR photo and yield suitable training targets for our extension after color reproduction. At the same time, they do not require the controlled setup necessary for exposure stacks, but can be readily captured and stored. Publicly available RAW collections [Bychkovsky et al. 2011; Chen et al. 2018; Conenna 2019; Dang-Nguyen et al. 2015; Hasinoff et al. 2016; Liang et al. 2021; Prabhakar et al. 2021] consequently exceed HDR datasets drastically in both size and diversity.

Our constrained extension steps allow us to supervise the reconstruction in linear space. Therefore, we directly follow the adversarial scheme proposed by Huang et al. [2024] for supervision to encourage realistic results. We apply LPIPS [Zhang et al. 2018] as a pixel-wise loss ($\mathcal{L}_{LPIPS}$) and use two distinct discriminators $\mathcal{D}_E$ and $\mathcal{D}_R$ to ensure the realism and plausibility of I_E and I_R , respectively. We also condition $\mathcal{D}_E$ on the input image I_L . Our combined training loss can be summarized as:

$$\mathcal{L} = \mathcal{L}_{LPIPS} + \mathcal{D}_E + \mathcal{D}_R. \quad (1)$$

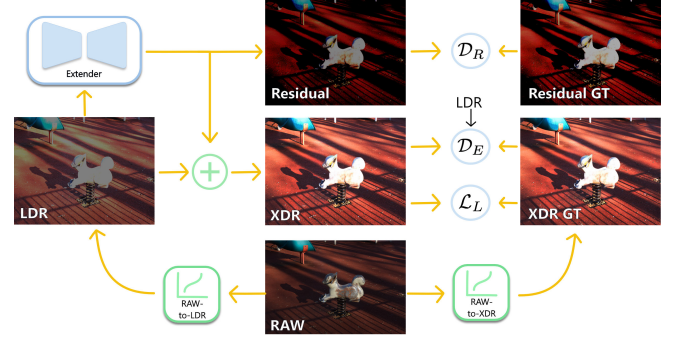


Fig. 3. Training the single-step dynamic range extension. From a given RAW photo, we create an LDR input and an extended target. Our network then predicts a residual image that is added to the LDR input to reproduce the extended image. We supervise both the residual and the extended output using a dual-discriminator adversarial loss to ensure realism.

Image credits: Bychkovsky et al. [2011]

The combination of our two discriminators with their complementary properties is essential to avoid degenerate solutions. Recall that the extended image can still contain clipping from bright highlights exceeding the extended range, as it only increases dynamic range by one EV. Hence, the output characteristics are similar to those of I_L , leaving several options for the extension network to fool the discriminator. Adding I_L as a conditional input to the $\mathcal{D}_E$ serves as an important clue to identify a realistic extension. It prevents the network from producing a trivial or all-zero residual that would result in a simple reproduction of I_L .

On the other hand, cases such as fully white estimates or duplicating I_L in I_R yield a simple reproduction of the input with shifted or scaled brightness. These results are degenerate because the dynamic range remains unchanged, but they resemble real images and appear different from the original I_L . $\mathcal{D}_R$ imposes a constraint on the distribution of the generated residuals and mitigates this effect. At the same time, a single discriminator for the residuals cannot guarantee that the extended image I_E itself looks realistic. It also will not penalize the extension network for modifying pixels not affected by clipping. Using this design, we train our dynamic range extension model with a batch size of 20 for 400 epochs. We provide additional details for our data processing and our losses and an ablation for our dual-discriminator setup in the supplementary material.

3.2 Learning Accurate Highlight Distributions

Our single-exposure extension already provides the network with a prior over diverse RAW images. To successfully reconstruct true HDR images, it additionally has to learn the realistic luminance of highlights and light sources over a wide dynamic range, and to correct its previous estimation errors in later stages. For this purpose, we integrate the recurrent inference as a second stage into its training process.

We show an overview of the recurrent training strategy in Figure 4. During training, we repeatedly apply our network for N inference steps, where the normalized output of each step serves as the

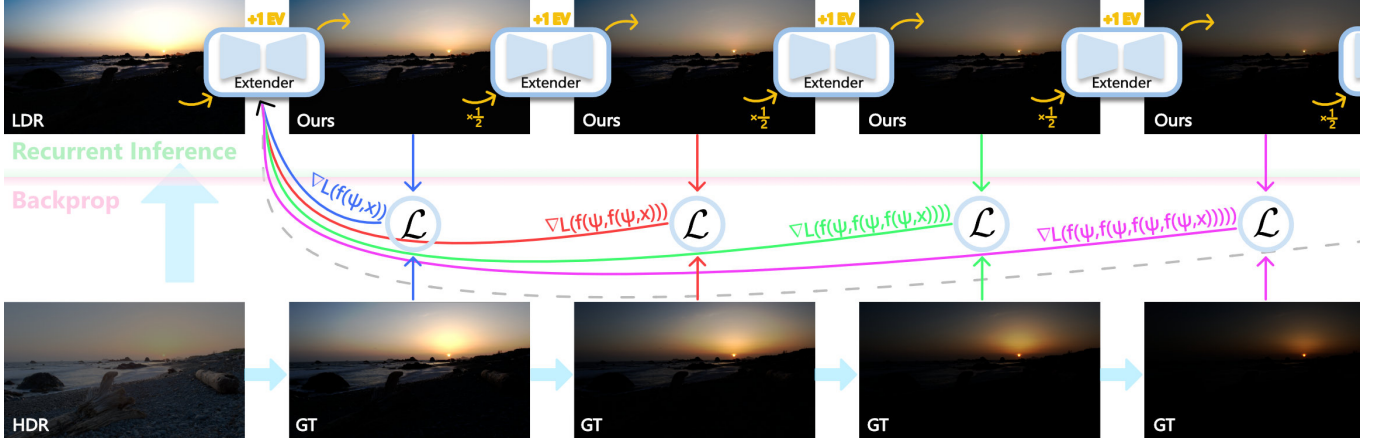


Fig. 4. Our recurrent training pipeline. We create a linear LDR input and targets for N steps from an HDR image. Next, we iteratively perform inference using the prediction of each step as input to the next. We then backpropagate the reconstruction error through the entire inference chain, taking the contribution of all steps into account. Image credits: Fairchild [2007]

input to the next. This produces a sequence of intermediate estimates $\{I_E^n\}_{n=1}^N$ defined as $I_E^n = f(\psi, I_E^{n-1})$ where $f(\cdot)$ denotes the network with parameters ψ . Each intermediate prediction I_E^n is supervised using its corresponding ground truth derived from HDR data. During training, gradients from the loss at each step propagate through the entire chain of iterative network executions. While the result for the first step depends on the network itself (i.e. $f(\psi, I_L)$), the second step becomes $f(\psi, f(\psi, I_L))$, the third $f(\psi, f(\psi, f(\psi, I_L)))$, and so on. By backpropagating through the same network parameters, weight updates also reflect how errors introduced at earlier steps influence subsequent extensions. This allows the model to learn to reduce mistakes made in earlier iterations by propagating corrective gradients back to the beginning of the chain.

This iterative gradient computation leads to a heavy computational burden, which exceeds the memory of most common GPUs. Because of this limitation, previous methods omit backpropagating through recurrent inference, leading to subpar results where prediction errors accumulate over iterations. We efficiently compute the complex recurrent gradients by adapting the Memory Replay strategy from NLP literature [Wu et al. 2022], and successfully integrate the recurrent inference in our training. Different from inference, the number of recurrent steps N in training does not depend on the image content. Starting from our single-EV pretraining with $N = 1$ and applying the same losses, we progressively increase N by 1 and continue training with the fixed step count. We repeat this process three times to a maximum of $N = 4$. After each increment, we train for 100 epochs with the new N to allow our model to adjust. For $N > 4$, the number of valid pixels in our training data is insufficient to yield meaningful supervision signals.

During the recurrent training, it is crucial to provide accurate supervision over the wider dynamic range. We construct our training data this time from HDR images that represent specular highlights and light sources well. Our network thus adapts to the long-tailed distribution of HDR images and simultaneously maintains the prior

from single-stage training. We use HDRT [Peng et al. 2025], S2R-HDR [Wang et al. 2025a], and HDR-Real [Liu et al. 2020] for our recurrent training scheme. We provide the details for our HDR data processing and the Memory Replay in the supplementary material.

3.3 Network Architecture And Training Details

Our dynamic range extension model is built upon the fully convolutional MiDaS 2.1 architecture [Ranftl et al. 2022] with a ResNext101-backbone. It is well-established for high-resolution imaging tasks and has been applied for HDR reconstruction previously [Dille et al. 2024]. For both discriminators, we use the architecture from PatchGAN [Isola et al. 2017] with 3 layers each. We use AdamW [Loshchilov and Hutter 2017] as an optimizer and a Cosine Annealing Scheduler [Loshchilov and Hutter 2016] with $T_{max} = 10$. The learning rate at the start of training is 10^{-4} for the generator and 10^{-5} for both discriminators.

4 Experiments

We evaluate the performance of our method by comparing it against a wide set of existing approaches. HDR-CNN [Eilertsen et al. 2017], ExpandNet [Marnerides et al. 2018], HDRUnet [Chen et al. 2021], and Mask-HDR [Santos et al. 2020] directly estimate an HDR image in a single step without breaking up the dynamic range. Still closely related are Single-HDR [Liu et al. 2020] and IntrinsicHDR [Dille et al. 2024], which use multi-step pipelines but go directly from linearized LDR RGB [Liu et al. 2020] or LDR Shading [Dille et al. 2024] to HDR. They stand in contrast to DrTMO [Endo et al. 2017], Multi-Exp Gen. [Le et al. 2023], CEVR [Chen et al. 2023], and Diff-Syn [Kim et al. 2021], which all create differently exposed LDR images first and combine them second via multi-exposure merging. GaSLight [Bolduc et al. 2025], BracketDiffusion [Bemana et al. 2025], and LEDiff [Wang et al. 2025b] follow a similar strategy, building upon fine-tuned large generative diffusion models. Our method progressively extends the dynamic range of the input to create an HDR image. We provide a numerical evaluation on two representative

Table 1. Quantitative results against state-of-the-art on the SI-HDR [Hanji et al. 2022] and BtP-HDR [Bolduc et al. 2023] benchmarks. We compute PSNR, VSI, and P-H in the perceptual uniform PU21-space [Mantiuk and Azimi 2021]. Grey font color indicates that the BtP-HDR Benchmark was used in training.

Method	SI-HDR benchmark [Hanji et al. 2022]						BtP-HDR benchmark [Bolduc et al. 2023]					
	PSNR↑	VSI↑	VDP3↑	P-H↑	FLIP↓	CVVDP↑	PSNR↑	VSI↑	VDP3↑	P-H↑	FLIP↓	CVVDP↑
IntrinsicHDR [Dille et al. 2024]	36.62	98.27	8.96	32.63	55.92	7.46	34.86	97.86	8.85	31.48	59.50	8.38
HDR-CNN [Eilertsen et al. 2017]	35.91	98.17	8.39	31.87	60.98	6.82	36.18	98.46	<u>9.04</u>	32.83	<u>44.92</u>	8.88
ExpandNet [Mamerides et al. 2018]	36.01	97.65	8.67	32.53	55.69	6.95	<u>36.33</u>	98.15	8.74	<u>33.25</u>	67.16	7.79
Single-HDR [Liu et al. 2020]	35.68	<u>98.30</u>	8.79	31.34	<u>52.40</u>	7.76	33.74	97.85	8.85	30.00	70.04	7.79
Mask-HDR [Santos et al. 2020]	36.72	98.22	8.25	<u>33.28</u>	62.33	6.75	35.96	<u>98.40</u>	9.00	32.88	42.81	8.88
HDRUNet [Chen et al. 2021]	<u>36.92</u>	98.16	8.82	31.58	85.92	5.24	35.71	98.13	8.90	32.23	74.49	6.76
DrTMO [Endo et al. 2017]	33.58	96.73	8.27	28.50	84.58	5.96	34.83	97.94	8.41	31.39	71.09	7.60
Multi-Exp Gen. [Le et al. 2023]	35.36	98.01	8.64	31.23	86.88	5.08	33.72	97.83	8.41	30.41	77.43	6.51
CEVR [Chen et al. 2023]	34.48	97.31	8.46	31.07	64.99	6.59	34.32	97.56	8.46	31.0	63.81	8.01
DiffSyn [Kim et al. 2021]	35.56	97.66	8.60	31.39	63.22	6.47	34.15	97.78	8.33	30.51	65.04	7.84
BracketDiffusion [Bemana et al. 2025]	32.50	95.52	7.54	28.38	64.62	6.06	34.69	97.39	8.46	31.85	53.15	8.29
LEDiff [Wang et al. 2025b]	32.69	96.41	7.89	28.47	81.55	4.63	35.61	98.04	8.64	32.69	58.37	7.62
GaSLight [Bolduc et al. 2023]	34.92	97.97	8.65	29.95	55.90	7.02	37.05	98.25	9.09	33.99	52.41	8.86
OURS	37.94	98.79	9.10	35.09	48.63	<u>7.64</u>	36.37	98.46	9.12	33.42	49.92	<u>8.87</u>

real-world benchmarks in Section 4.1 and a detailed visual analysis of our results in Section 4.2. Further evaluation results and ablations are provided in the supplementary material.

Note that our approach operates in linear RGB space and is independent of the linearization method itself. It can be applied flexibly after dedicated linearization networks such as the one proposed by Liu et al. [2020], or to further extend already linear RAW images. For our numerical evaluation and all visual results, we apply an inverse gamma function [ITU 2011] as a simple baseline to linearize the images. We expect further improvements with dedicated inversion of the camera response function.

4.1 Quantitative Evaluation

We quantitatively evaluate our approach on two HDR benchmarks. For all results, we find the best scale between the reconstructed radiance and the ground-truth based on the valid pixels in the input, and correct the CRF with a third-degree polynomial, following the protocol recommended by Hanji et al. [2022].

Benchmarks. We compare all methods on the established SI-HDR [Hanji et al. 2022] and BtP-HDR [Bolduc et al. 2023] benchmarks. Both sets contain diverse challenging HDR settings such as sunsets, specular reflections, and visible light sources. While SI-HDR focuses strongly on outdoor scenes, BtP-HDR fittingly extends to bright light sources in images derived from the larger Laval Indoor HDR dataset [Gardner et al. 2017]. Both datasets are zero-shot settings for our approach and all baselines, as they have not been used in training. The only exception is GaSLight, which has been trained on BtP-HDR.

Metrics. We evaluate the results using six different metrics. PSNR and VSI are standard metrics for image quality and provide insightful results on HDR images in the perceptual uniform PU21 color space [Mantiuk and Azimi 2021], which aligns them closely with human preference [Hanji et al. 2022]. HDR-VDP3 [Mantiuk et al. 2023], short VDP3, is a popular HDR metric that models human perception. ColorVideoVDP [Mantiuk et al. 2024] (CVVDP) is closely related, with additional focus on color reproduction in HDR.

FLIP [Andersson et al. 2020] computes the reconstruction error over different exposures and is well-suited to evaluate the full dynamic range, including bright highlights. We additionally report results on PSNR-H [Dille et al. 2024], or P-H, which specifically evaluates the highlight sections of the image.

Results. We show our results in Table 1. Our recurrent dynamic range extension outperforms all baselines on SI-HDR with significant improvements on most metrics. The strong results on VDP3 underline the close perceptual alignment with natural scenes, supported also by the competitive results on CVVDP. The large gains on FLIP and P-H indicate that the recurrent extension enables the reconstruction of a wide dynamic range and creates realistic highlights for specular reflections and visible light sources. This is due to our exposure-invariant formulation that allows the network to learn the distribution of highlights well while maintaining robustness for lower brightnesses. With the recurrent training, the network learns meaningful radiance values over a wide dynamic range. It is robust against accumulated errors and clearly outperforms stack-reconstruction methods. This includes BracketDiffusion, LEDiff, and GaSLight, which are built on large generative diffusion networks. Our method is thus the only iterative approach that performs not only on par but better than LDR-to-HDR reconstruction baselines.

We see a similar picture emerging on BtP-HDR with more variance in the numbers. Our method performs competitively with the strongest baseline GaSLight, although the latter has been trained on the dataset. Note that neither BtP-HDR nor the original Laval Indoor HDR dataset is included in the training data for our method. This shows our generalization performance to both outdoor and indoor images and underlines once more the realistic recovery of bright light sources.

Dynamic Range Gain. We evaluate the effect of our second training phase on the resulting dynamic range in Table 2. We show the achieved dynamic range gains on the SI-HDR [Hanji et al. 2022] benchmark. The input is the original LDR with an encoded dynamic range of 256:1, i.e., 8 stops. Matching our formulation, we focus the

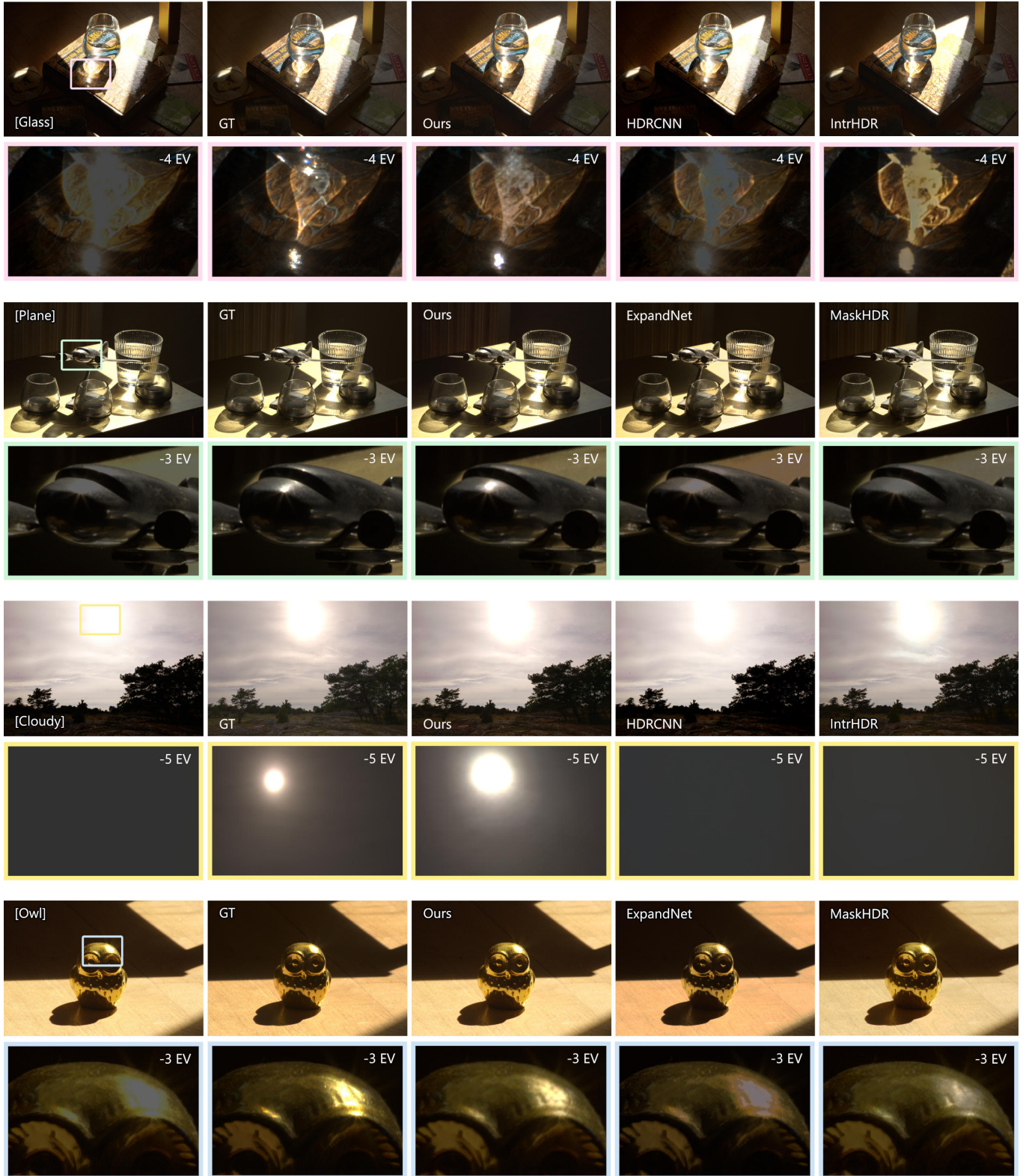


Fig. 5. We show results of our method against the strongest baselines on the highlight-specific P-H metric. Our method recovers plausible highlights on SI-HDR [Hanji et al. 2022], where other methods saturate early or create artifacts.

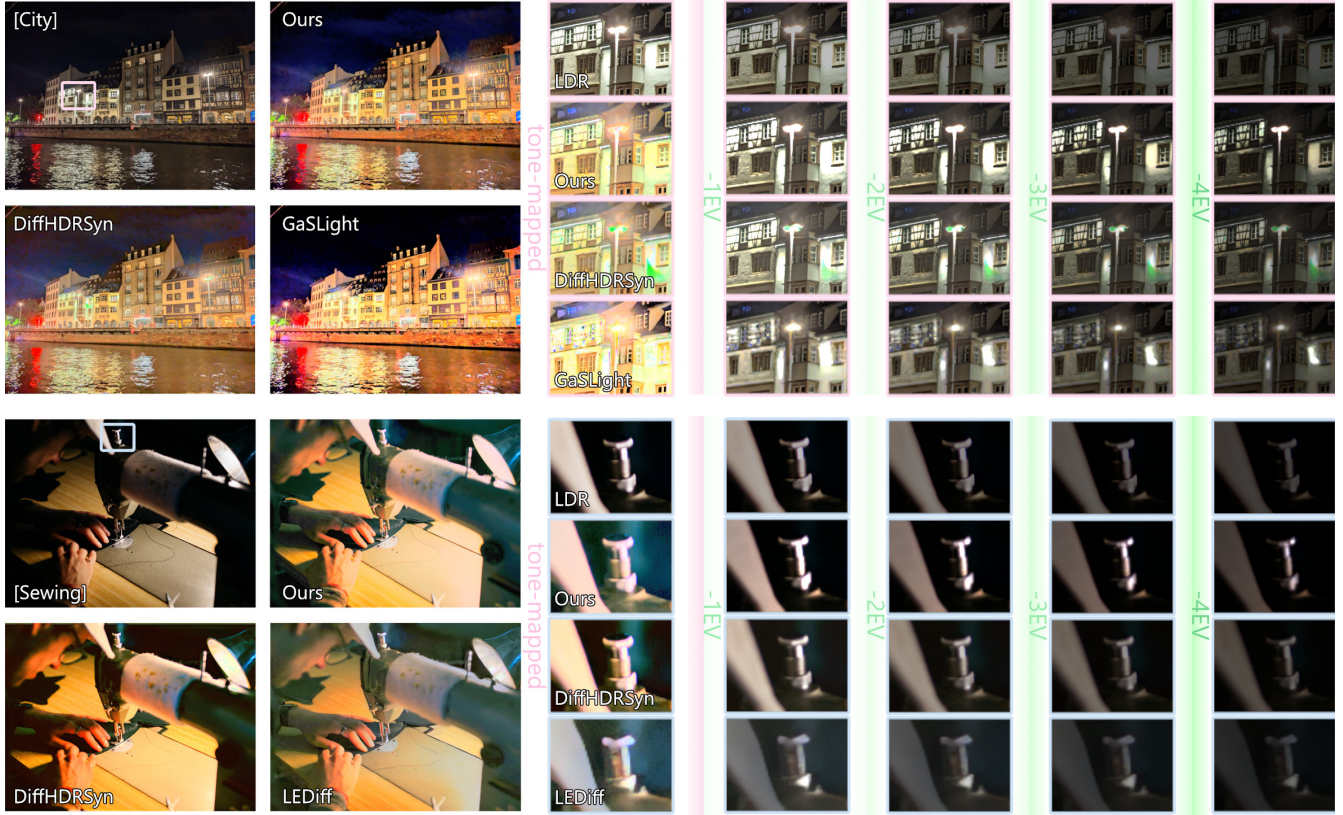


Fig. 6. Compared to multi-exposure generation methods, our approach is powerful in recovering realistic luminance for bright light sources, recovers details in the highlights, and is robust against artifacts. Image credits: Death to the Stock Photo

Table 2. We compare the dynamic range gains on the highlights of the SI-HDR [Hanji et al. 2022] benchmark.

Method	avg. DR	max. DR
GaSLight [Bolduc et al. 2025]	80,000:1	1,049,000:1
Mask-HDR [Santos et al. 2020]	45,000:1	1,921,000:1
HDR-CNN [Eilertsen et al. 2017]	50,000:1	1,078,000:1
OURS - w/o second phase	748:1	2,929:1
OURS	120,000:1	8,778,000:1
GT	529,000:1	11,874,000:1

evaluation on the highlight extension and mask out extended shadows in the baseline methods and the ground truth. The computed dynamic range is thus the ratio of the linearized maximum pixel value to $1/255$ as minimum. We report the average and maximum across ratios.

Without our recurrent training and HDR ground-truth data, the dynamic range extension only slightly increases the dynamic range of the input. The network has only seen RAW images in training. Used to a single extension, it is not able to reproduce the high luminance values of bright highlights. This effect also becomes visible in Figure 3 of the supplementary material, where the single-stage

training saturates early and does not show further improvement for additional network executions.

Once we add the recurrent training with full HDR supervision, the performance changes significantly. Our full approach learns the formation of highlights over multiple consecutive exposures. On average, the network now performs five inference steps and more than doubles the number of stops, outperforming the strongest baselines.

Our second-stage training introduces both recurrent training and exposure to HDR ground-truth to the system, both of which help our method estimate longer-tailed distributions. We further analyze the components of our method in Section D of the supplementary document, where we show the effect of the recurrent training, the supervision with RAW data, and our two discriminators.

4.2 Qualitative Evaluation

We support our numerical findings with visual comparisons of examples from the SI-HDR [Hanji et al. 2022] benchmark and images in the wild in Figures 5, 6, and 7. Our method performs strongly in recovering the highlights even in extreme cases, and simultaneously inpaints clipped areas convincingly in a wide range of scenarios.

Highlight Reconstruction. Recovering the accurate brightness of highlights is one of the main challenges for HDR reconstruction.

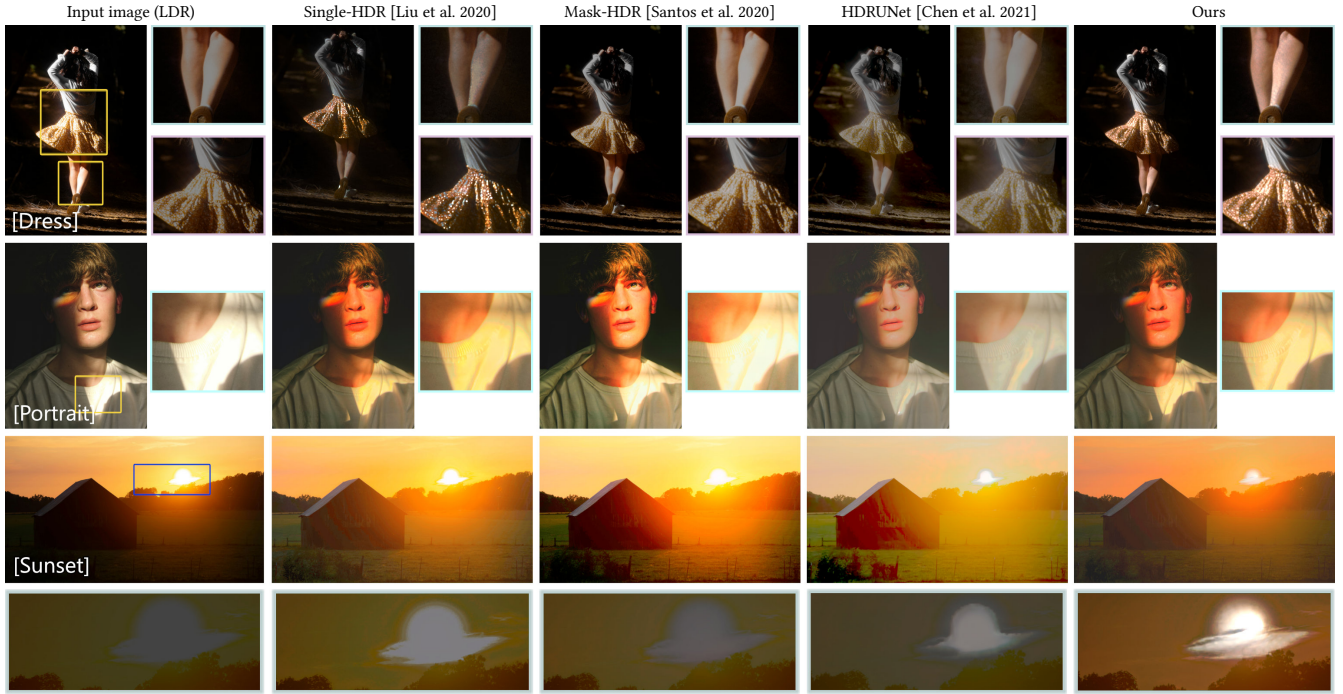


Fig. 7. Our method faithfully recovers clipped areas in images in the wild with colors that match the environment. The full images are tone-mapped.

Image credits: @Matthew Ball, @gryffyn m, @Elmer Cañas

Their extreme values create the long-tailed distribution of the HDR image and force direct LDR-to-HDR reconstruction methods to compress the luminance range during training. We show the results of our method on challenging highlights in the SI-HDR benchmark in Figure 5. For each example, we show the ground-truth and the results from HDRCNN, MaskHDR, IntrinsicHDR, and ExpandNet, all of which demonstrate competitive performance in quantitative baselines (Section 4.1).

In the [Glass] example, the LDR input shows clipping on the book cover where it is hit by the sunlight reflected from the glass. The ground-truth reveals a bright specular spot in the lower part of the clipped area, which remains clipped even when the exposure is lowered by four stops. In the LDR input, the large brightness difference between the specular highlight and the book cover is lost due to clipping. For this reason, both baseline methods equally increase the brightness of the full clipped area and struggle to reproduce the highlight spot. Being trained on a compressed dynamic range, they saturate early, resulting in flat gray regions on the book cover. In contrast, our approach faithfully matches the brightness of the specular highlight. It also inpaints parts of the lost texture of the book cover, while recovering the challenging contrast between the book cover and the specular highlight present in the ground-truth.

We observe a similar behavior for the [Plane] and the [Owl] scenes. For the [Plane], ExpandNet and MaskHDR both leave the clipped highlight unchanged. For the [Owl], MaskHDR faithfully recovers the golden metallic appearance of the owl, but does not reproduce the brightness of the highlight. ExpandNet creates color artifacts and barely increases the dynamic range of the input. In

both cases, our method recovers the strong direct reflection of the nose of the plane and the golden head of the owl and extends the dynamic range beyond four exposure values, in accordance with the ground-truth.

[Cloudy] poses a challenge, as the sun is visible through the cloud cover of the sky. The bright spot of the sun in the ground-truth is in contrast to the overcast sky in the surroundings. As a result, both HDRCNN and IntrinsicHDR reconstruct a cloudy sky with low dynamic range, while our method creates a realistic sun disc with corresponding brightness. Note that our generated dynamic range exceeds the number of recurrent steps used during training.

The examples show the strength of our extension formulation in being adaptive to the content of the scene rather than being limited to the specific number of exposure brackets used during training. Our formulation leaves the luminance uncompressed, and the method learns to recover accurate bright highlights even in these challenging scenarios.

Multi-Exposure Consistency. Multi-exposure reconstruction approaches require multiple inference steps of the same network. This is prone to accumulating prediction errors and creating artifacts in the estimated result. We compare our results with the recent exposure-stack generation methods DiffHDRSyn, LEDiff, and GaSLight over a wide dynamic range in Figure 6. The last two baselines are both based on pretrained generative diffusion models that have been fine-tuned on HDR reconstruction.

For the [City], DiffHDRSyn and GaSLight both fail to preserve the brightness of the light source. DiffHDRSyn additionally creates

strong color artifacts in the clipped areas, while GaSLight exaggerates the brightness of the wall, giving it the appearance of an active light source. Both cases present examples of artifacts caused by iterative inference. In the [Sewing] image, DiffHDRSyn and LEDiff only slightly extend the brightness of the pin, resulting in a more diffuse, non-metallic appearance. Notably, the generative LEDiff incorrectly modifies the shape of the pin, which is a common weakness of diffusion-based methods.

In comparison, our method maintains the strong brightness of the light source in the [City] example and reduces the brightness of the diffuse wall realistically while recovering the color. This shows the importance of our recurrent training, which stabilizes the dynamic range extension over multiple inference steps and reduces accumulated artifacts such as those of the other two methods. For this reason, our method reproduces the metallic appearance of the pin in the [Sewing] example, preserves its shape, and concentrates the intensity of the specular highlight realistically on a single point.

Our method maintains accurate brightness levels for light sources, bright surfaces, and reflections over multiple exposure steps. We show more examples in our supplementary material.

Detail Recovery. We compare the ability to recover clipped details on images in the wild in Figure 7. Generalization to diverse imagery and realistic inpainting of details is another challenge for HDR reconstruction methods, since learning a strong prior on natural images is limited by the scarcity of HDR training data. We compare against LDR-to-HDR reconstruction methods SingleHDR, MaskHDR, and HDRUnet, and show additional results against IntrinsicHDR, DiffHDRSyn, and LEDiff in the supplementary.

The clipped LDR input of the [Dress] requires recovery of both skin color and the pattern on the skirt. Being trained on HDR images alone, SingleHDR produces extremely bright, unnatural spots for the skirt. MaskHDR and HDRUnet, on the other hand, both leave the skin clipped and fail to recover natural color. For the [Portrait], SingleHDR and HDRUnet both fail to accurately inpaint the shirt and produce artifacts. MaskHDR leaves the clipped area completely white, creating a clear boundary. Our method, in contrast, performs well on the [Dress], recovering the color of the skirt and the skin tone. The estimated values match the LDR input without being compressed. We observe similar performance for the [Portrait]: Our method seamlessly fills the clipped areas with matching colors. We attribute the behavior to our training on diverse RAW collections that enables our network to learn a prior over natural images. This prior is maintained through the dynamic range extension, so that our method can add details even in bright highlight regions. This becomes visible for the Church example in our teaser Figure 1, and for the [Sunset] example in Figure 7, where it is the only method reproducing the shape of the cloud in front of the sun.

5 Conclusion And Limitations

We have presented a recurrent dynamic range extension approach that progressively recovers the highlights of an HDR image. Our method generalizes to diverse in-the-wild scenarios by faithfully recovering skin tones, specular reflections, and light sources. The generalization performance comes from our scale-invariant formulation that enables training on rich RAW photo collections. Our



Fig. 8. Limitations of our method. While the network creates local structure and details, it results in a repetitive, unrealistic pattern for the tree. In some cases, our method overestimates the brightness of the clipped area, creating a bright spot on the shirt.

recurrent training setup opens the extension to arbitrary dynamic ranges without restriction during training. We demonstrate the performance in an extensive visual and numerical analysis.

We show limitations of our approach in Figure 8. Stemming from the distribution of RAW image collections, our approach assumes casual photography conditions with reasonably exposed midtones. Under these circumstances, our extension network generates structure matching to the local neighbourhood. For the severe overexposure on the [Tree] trunk with larger clipped areas of the complex structure of the bark, however, its generative capabilities are limited and can lead to repetitive patterns. Similarly, our approach can overestimate the original brightness of affected areas. In the case of the [Shirt], the LDR image contains a very large area of all-white pixels, which does not provide sufficient semantic information for our approach to reason about the context. Our method correctly extends the wrinkles of the cloth, but creates an unrealistic brighter light spot in the center.

Further, our method addresses one part of the larger HDR reconstruction task. Recovering the full radiance of a scene additionally involves the removal of noise and quantization artifacts in the shadows, as well as the inversion of the non-linear CRF. These effects have different characteristics from clipping and pose interesting targets for future work.

Acknowledgments

We thank Alireza Moazeni and Ke Li for our discussions in the early stages of this work. We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [RGPIN-2020-05375].

References

- Mekides Assefa Abebe, Alexandra Booth, Jonathan Kervé, Tania Pouli, and Mohamed-Chaker Larabi. 2018. Towards an automatic correction of over-exposure in photographs: Application to tone-mapping. *Computer Vision and Image Understanding*

- 168 (2018), 3–20.
- Pontus Andersson, Jim Nilsson, Tomas Akenine-Möller, Magnus Oskarsson, Kalle Åström, and Mark D Fairchild. 2020. FLIP: A Difference Evaluator for Alternating Images. *Proc. ACM Comput. Graph. Interact. Tech.* 3, 2 (2020), 15–1.
- Francesco Banterle, Patrick Ledda, Kurt Debattista, and Alan Chalmers. 2006. Inverse tone mapping. In *Proc. GRAPHITE*.
- Francesco Banterle, Patrick Ledda, Kurt Debattista, and Alan Chalmers. 2008. Expanding low dynamic range videos for high dynamic range applications. In *Proc. SCCG*.
- 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, Justine Giroux, Marc Hébert, Claude Demers, and Jean-François Lalonde. 2023. Beyond the pixel: a photometrically calibrated hdr dataset for luminance and color prediction. In *Proc. ICCV*.
- 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*.
- Si-Yuan Cao, Runmin Zhang, Lun Luo, Beinan Yu, Zehua Sheng, Junwei Li, and Hui-Liang Shen. 2023. Recurrent homography estimation using homography-guided image warping and focus transformer. In *Proc. CVPR*.
- Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. 2018. Learning to see in the dark. 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*.
- Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder–decoder approaches. In *Proc. SSS-8*.
- Joey Conenna. 2019. Canon 6D datasets for learning to see in the dark. GitHub Repository. <https://github.com/jconenna/Canon-6D-Datasets-For-Learning-to-See-in-the-Dark>; accessed 03-March-2025.
- Duc-Tien Dang-Nguyen, Cecilia Pasquini, Valentina Conotter, and Giulia Boato. 2015. Raise: A raw images dataset for digital image forensics. In *Proc. MMSys*. 219–224.
- Paul E Debevec and Jitendra Malik. 1997. Recovering high dynamic range radiance maps from photographs. In *Proc. SIGGRAPH*. 369–378.
- Piotr Didyk, Rafal Mantiuk, Matthias Hein, and Hans-Peter Seidel. 2008. Enhancement of bright video features for HDR displays. *Comput. Graph. Forum* 27, 4 (2008).
- Sebastian Dille, Chris Careaga, and Yağız Aksoy. 2024. Intrinsic Single-Image HDR Reconstruction. In *Proc. ECCV*.
- Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafal 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*, Vol. 15. Society of Imaging Science and Technology, 233–238.
- Marc-André Gardner, Kalyan Sunkavalli, Ersin Yumer, Xiaohui Shen, Emiliano Gabbaretto, Christian Gagné, and Jean-François Lalonde. 2017. Learning to predict indoor illumination from a single image. *ACM Trans. Graph.* 36, 6 (2017), 1–14.
- Cheng Guo and Jiang Xiuhua. 2022. LHDR: HDR Reconstruction for Legacy Content using a Lightweight DNN. In *Proc. ACCV*.
- 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.*
- Samuel W. Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T. Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. 2016. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Trans. Graph.* 35, 6 (2016).
- Javier Hidalgo-Carrió, Daniel Gehrig, and Davide Scaramuzza. 2020. Learning monocular dense depth from events. In *Proc. 3DV*.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
- 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).
- Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proc. CVPR*.
- ITU. 2011. Recommendation ITU-R BT.1886: Reference electro-optical transfer function for flat panel displays used in HDTV studio production. Accessed: 2026-04-06. https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.1886-0-201103-I!!PDF-E.pdf
- So Yeon Jo, Siyeong Lee, Namhyun Ahn, and Suk-Ju Kang. 2021. Deep arbitrary HDR: Inverse tone mapping with controllable exposure changes. *IEEE Trans. MM* 24 (2021).
- 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*.
- Rafael Pacheco Kovaleski and Manuel M Oliveira. 2009. High-quality brightness enhancement functions for real-time reverse tone mapping. *The Visual Computer* 25, 5 (2009), 539–547.
- Phuoc-Hieu Le, Quynh Le, Rang Nguyen, and Binh-Son Hua. 2023. Single-image hdr reconstruction by multi-exposure generation. In *Proc. WACV*.
- Siyeong Lee, Gwon Hwan An, and Suk-Ju Kang. 2018a. Deep chain hdi: Reconstructing a high dynamic range image from a single low dynamic range image. *IEEE Access* 6 (2018), 49913–49924.
- Siyeong Lee, Gwon Hwan An, and Suk-Ju Kang. 2018b. Deep Recursive HDR: Inverse Tone Mapping Using Generative Adversarial Networks. In *Proc. ECCV*.
- Siyeong Lee, So Yeon Jo, Gwon Hwan An, and Suk-Ju Kang. 2020. Learning to generate multi-exposure stacks with cycle consistency for high dynamic range imaging. *IEEE Trans. MM* 23 (2020), 2561–2574.
- Jie Liang, Hui Zeng, Miaomiao Cui, Xuansong Xie, and Lei Zhang. 2021. Ppr10k: A large-scale portrait photo retouching dataset with human-region mask and group-level consistency. In *Proc. CVPR*.
- 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*.
- Ilya Loshchilov and Frank Hutter. 2016. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983* (2016).
- Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017).
- S Mann and RW Picard. 1995. On being “undigital” with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proc. IS&T Annual Conference*.
- Rafal K Mantiuk and Maryam Azimi. 2021. PU21: A novel perceptually uniform encoding for adapting existing quality metrics for HDR. In *Proc. PCS*.
- Rafal K Mantiuk, Dounia Hammou, and Param Hanji. 2023. HDR-VPD-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content. *arXiv preprint arXiv:2304.13625* (2023).
- Rafal K Mantiuk, Param Hanji, Maliha Ashraf, Yuta Asano, and Alexandre Chapiro. 2024. ColorVideoVPD: A visual difference predictor for image, video and display distortions. *ACM Trans. Graph.* 43, 4 (2024).
- Demetris Marnerides, Thomas Bashford-Rogers, and Kurt Debattista. 2021. Deep HDR Hallucination for Inverse Tone Mapping. *Sensors (Basel, Switzerland)* 21 (2021).
- 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).
- Belen Masia, Sandra Agustin, Roland W Fleming, Olga Sorkine, and Diego Gutierrez. 2009. Evaluation of reverse tone mapping through varying exposure conditions. In *Proc. ACM SIGGRAPH Asia*.
- Syed Z Masood, Jiejie Zhu, and Marshall F Tappen. 2009. Automatic Correction of Saturated Regions in Photographs using Cross-Channel Correlation. *Comput. Graph. Forum* 28, 7 (2009).
- Jingchao Peng, Thomas Bashford-Rogers, Francesco Banterle, Haitao Zhao, and Kurt Debattista. 2025. HDRT: A large-scale dataset for infrared-guided HDR imaging. *Information Fusion* (2025), 103109.
- K. Prabhakar, Vishal Vinod, N. Sahoo, and Venkatesh Babu Radhakrishnan. 2021. Few-Shot Domain Adaptation for Low Light RAW Image Enhancement. In *Proc. BMVC*.
- René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. 2022. Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 3 (2022).
- Allan G Rempel, Matthew Trentacoste, Helge Seetzen, H David Young, Wolfgang Heidrich, Lorne Whitehead, and Greg Ward. 2007. LDR2HDR: on-the-fly reverse tone mapping of legacy video and photographs. *ACM Trans. Graph.* 26, 3 (2007).
- 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).
- Jieying Shi, Ziheng Zhu, Jianhua Zhang, Ruyu Liu, Zhenhua Wang, Shengyong Chen, and Honghai Liu. 2020. CalibRCNN: Calibrating camera and LiDAR by recurrent convolutional neural network and geometric constraints. In *Proc. IROS*.
- B Shiva Prakash, KV Sanjeev, Ramesh Prakash, and K Chandrasekaran. 2018. A survey on recurrent neural network architectures for sequential learning. In *Soft Computing for Problem Solving: SocProS 2017, Volume 2*. Springer, 57–66.
- Zachary Teed and Jia Deng. 2020. Raft: Recurrent all-pairs field transforms for optical flow. In *Proc. ECCV*.
- Chao Wang, Zhihao Xia, Thomas Leimkühler, Karol Myszkowski, and Xuaner Zhang. 2025b. LEDiff: Latent Exposure Diffusion for HDR Generation. In *Proc. CVPR*.
- Sen Wang, Ronald Clark, Hongkai Wen, and Niki Trigoni. 2017. Deepvo: Towards end-to-end visual odometry with deep recurrent convolutional neural networks. In *Proc. ICRA*.
- Yujin Wang, Jiarui Wu, Yichen Bian, Fan Zhang, and Tianfan Xue. 2025a. S2R-HDR: A Large-Scale Rendered Dataset for HDR Fusion. *arXiv preprint arXiv:2504.07667*

- (2025).
- 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*.
- Hanning Yu, Wentao Liu, Chengjiang Long, Bo Dong, Qin Zou, and Chunxia Xiao. 2021. Luminance attentive networks for HDR image and panorama reconstruction. *Comput. Graph. Forum* 40, 7 (2021).
- Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proc. CVPR*.
- Yang Zhang and TO Aydin. 2021. Deep HDR estimation with generative detail reconstruction. *Comput. Graph. Forum* 40, 2 (2021).
- Yunhao Zou, Chenggang Yan, and Ying Fu. 2023. RawHDR: High Dynamic Range Image Reconstruction from a Single Raw Image. In *Proc. ICCV*.