535514 RL Team Project: Reinforcement Learning in High Dynamic Range Reconstruction with Intrinsic Decomposition

Yo-Tin Lin 1* Jin-Ting He 1*
Department of Computer Science
National Yang Ming Chiao Tung University
{ytlin.cs12, jinting.cs12}@nycu.edu.tw

1 Project Overview

1.1 Profile

Track: 3. Application

Abstract and project goal: This project aims to leverage reinforcement learning (RL) techniques for High Dynamic Range (HDR) reconstruction from Low Dynamic Range (LDR) images. HDR reconstruction enhances image quality by increasing dynamic range, which is critical for photography, computer vision, and multimedia applications. The goal is to develop an RL-based framework that optimally fuses multiple image exposures or reconstructs HDR from single LDR images using intrinsic properties like albedo and shading.

TL;DR ("Too Long; Didn't Read"): Develop a reinforcement learning model to generate HDR images by optimizing fusion and intrinsic image reconstruction techniques.

1.2 Motivation

Why is the problem interesting? HDR reconstruction is essential for capturing realistic scenes in various lighting conditions, enabling better visualization and analysis in fields like autonomous driving, medical imaging, and entertainment. Current methods often rely on deterministic algorithms that fail to generalize across diverse environments.

Critical Challenges:

- *Challenge 1:* Designing effective reward functions for RL models that capture the quality of HDR reconstruction.
- *Challenge 2:* Creating a robust state space representation that includes albedo and shading information while accounting for image variability.
- *Challenge 3:* Ensuring computational efficiency and scalability for real-world implementation.

Why is the problem unsolved? Existing methods often lack adaptability to varying lighting conditions or fail to optimize HDR quality across diverse datasets. Prior works such as "Automatic Intermediate Generation With Deep Reinforcement Learning for Robust Two-Exposure Image Fusion" Yin et al. [2022] and "Intrinsic Single-Image HDR Reconstruction" Dille et al. [2024] provide foundational insights but are limited in scope or performance optimization.

State-of-the-Art Methods: CEVR Chen et al. [2023] Since it achieves the highest SSIM and PSNR scores in the single HDR reconstruction task. It is the current SOTA method.

2 Problem Formulation

Environment Model: The problem is modeled as a **Markov Decision Process (MDP)**, defined by the tuple (S, A, P, R, γ), where:

- **State Space (S)**: Each state consists of a pair of images:
  - **Albedo** $\in \mathbb{R}^{H\times W\times 3}$: representing the color and material reflectance.
  - **Shading** $\in \mathbb{R}^{H\times W\times 1}$: representing the illumination component.
- **Action Space (A)**: The action space defines a set of continuous or discrete transformations applied to albedo and shading independently or jointly. These transformations include:
  - Brightness adjustment (for Shading)
  - Contrast adjustment (for Shading)
  - Saturation tuning (for Albedo)

  The agent outputs transformation parameters that modify the current albedo and shading images. These actions are designed to progressively reconstruct high dynamic range (HDR) components from their low dynamic range (LDR) counterparts.

Reward Signal. The reward function $R(s, a)$ is designed to guide the reinforcement learning agent toward generating albedo and shading maps that facilitate the reconstruction of high-quality HDR images. To quantitatively evaluate the quality of the predicted HDR output, the reward integrates two widely-used error metrics: Mean Squared Error (MSE) and Mean Absolute Deviation (MAD). Specifically, the reconstructed HDR image $I_{\text{HDR}}^{\text{pred}}$ is computed as the element-wise product of the predicted HDR albedo and shading components. The reward is then defined as:

$$\text{Reward} = -\left[\lambda_1 \cdot \text{MSE}(I_{\text{HDR}}^{\text{pred}}, I_{\text{HDR}}^{\text{gt}}) + \lambda_2 \cdot \text{MAD}(I_{\text{HDR}}^{\text{pred}}, I_{\text{HDR}}^{\text{gt}})\right],$$

where $I_{\text{HDR}}^{\text{gt}}$ denotes the ground-truth HDR image, and $\lambda_1, \lambda_2 \in \mathbb{R}^+$ are scalar weights that balance the contributions of the MSE and MAD terms. The negative sign ensures that minimizing the error corresponds to maximizing the reward, thereby encouraging the agent to produce HDR outputs that are both numerically accurate and perceptually faithful to the ground truth.

Optimization Problem: Maximize the HDR image quality while minimizing computational overhead. Constraints include maintaining naturalness in reconstructed images and avoiding artifacts.

Assumptions: The input LDR images are preprocessed to ensure compatibility with the RL model, and the training data covers diverse lighting conditions.

3 Empirical Evaluation

3.1 Performance Metrics

- **PSNR:** Measures fidelity between reconstructed HDR and ground truth images.
- **SSIM:** Evaluates structural similarity between images.
- **Perceptual Loss:** Captures human-perceived differences in image quality.

3.2 Baseline Methods

We choose the single HDR reconstruction that is (was) the state-of-the-art:

- **Deep Recursive HDRI** Lee et al. [2018]
  The first framework to create high dynamic range images based on the estimated multi-exposure stack using the conditional generative adversarial network structure.

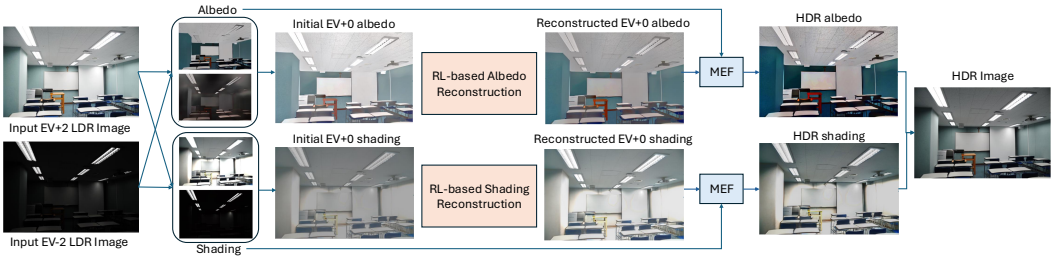

Figure 1: Overall architecture of the proposed HDR reconstruction framework. The system begins by decomposing an LDR image into albedo and shading components. These components are then processed by an RL agent, which learns enhancement strategies to reconstruct high-quality HDR representations.

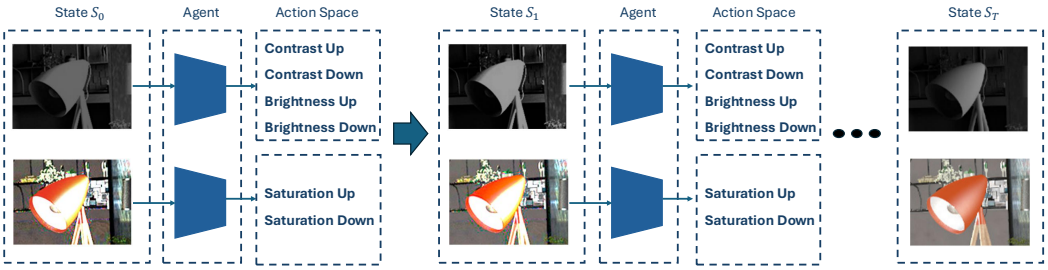

Figure 2: Details of the reinforcement learning module. The RL agent interacts with the environment by applying a set of enhancement actions—such as brightness, contrast, and saturation adjustments—to albedo and shading maps. The agent is trained to optimize these transformations for improved HDR quality.

- **SingleHDR** Liu et al. [2020]
  Decomposes HDR reconstruction into three CNNs reversing dynamic range clipping, nonlinear mapping, and quantization steps.
- **CEVR** Chen et al. [2023]
  Proposes an implicit function to generate LDR stacks with arbitrary exposure values (EVs) for improved HDR reconstruction via cycle training.

3.3 Benchmark Datasets

We will conduct the evaluation on the famous HDR dataset: VDS Dataset Lee et al. [2018]. The dataset consists of a variety of scenarios and are also widely used by other HDR reconstruction works.

4 Methodology

The proposed framework aims to reconstruct a high dynamic range (HDR) image from two low dynamic range (LDR) images taken at different exposure values (EVs), specifically EV–2 (underexposed) and EV+2 (overexposed). As illustrated in Figure 1, the core objective is to synthesize an intermediate EV0 image representing the properly exposed frame by learning to interpolate between the two inputs using a reinforcement learning (RL) model, and subsequently combining all three exposures to generate an HDR image.

4.1 Intrinsic Image Decomposition

We begin by applying an intrinsic image decomposition module to both input images, $I_{\text{EV}-2}$ and $I_{\text{EV}+2}$. This module decomposes each image into its albedo and shading components:

$$I_{\text{EV}} = A_{\text{EV}} \odot S_{\text{EV}}, \tag{1}$$

where $A_{\text{EV}} \in \mathbb{R}^{H\times W\times 3}$ denotes the albedo image, $S_{\text{EV}} \in \mathbb{R}^{H\times W\times 1}$ denotes the shading image, and $\odot$ indicates element-wise multiplication. The decomposition provides an illumination-invariant representation (albedo) and a lighting-dependent structure (shading), which allows for more stable learning and synthesis in varying exposure conditions.

4.2 Initial EV0 Estimation via Interpolation

To estimate an initial approximation of the EV0 scene, we perform a linear interpolation between the decomposed shading and albedo components of $I_{\text{EV}-2}$ and $I_{\text{EV}+2}$:

$$S_{\text{EV0}}^{\text{init}} = \frac{1}{2}(S_{\text{EV}-2} + S_{\text{EV}+2}), \quad A_{\text{EV0}}^{\text{init}} = \frac{1}{2}(A_{\text{EV}-2} + A_{\text{EV}+2}). \tag{2}$$

These interpolated results serve as the starting point for our reinforcement learning policy to refine toward the true EV0 representation.

4.3 Reinforcement Learning-Based Enhancement

The reinforcement learning agent is tasked with refining $S_{\text{EV0}}^{\text{init}}$ and $A_{\text{EV0}}^{\text{init}}$ such that they approximate the true shading and albedo under EV0 illumination. Each state s_t at time step t consists of the current albedo and shading pair:

$$s_t = \left(A_t \in \mathbb{R}^{H\times W\times 3}, \quad S_t \in \mathbb{R}^{H\times W\times 1}\right). \tag{3}$$

The agent applies enhancement actions to this state using parameterized transformations to adjust brightness, contrast, and saturation. Each action is defined by:

$$I_{t+1} = I_t \cdot \alpha + \beta \cdot (1 - \alpha), \tag{4}$$

where α is a scaling factor and β is computed based on the content statistics of I_t, specifically the Y and Cb/Cr channels in the YCbCr color space:

$$\beta = \begin{cases} \overline{I_Y}(t), & \text{if adjusting brightness or contrast} \\ \overline{I_{\text{cbcr}}}(t), & \text{if adjusting saturation} \end{cases} \tag{5}$$

This design ensures that the enhancement is context-aware and adaptive to the input content. The action space comprises six discrete actions, listed in Table 1, enabling both increases and decreases in the aforementioned visual attributes. Empirically, we allow a maximum of 20 sequential actions, after which the process is terminated.

As illustrated in Figure 2, we decouple the enhancement strategy according to the characteristics of intrinsic components. Specifically, the reinforcement learning policy is designed to adjust saturation only for the shading map, while brightness and contrast adjustments are applied exclusively to the albedo image. This decision is based on the observation that shading primarily represents illumination intensity and geometric variations, which are more perceptually impacted by saturation changes, whereas albedo encodes material reflectance properties that are more sensitive to brightness and contrast. Such specialization in action design allows the agent to perform more effective and semantically meaningful corrections in the respective representation spaces.

4.4 HDR Reconstruction from Multiple Exposures

After obtaining the refined EV0 components $(\hat{A}_{\text{EV0}}, \hat{S}_{\text{EV0}})$, we perform HDR reconstruction. First, all three exposures are brought into a unified representation:

$$\text{HDR-Albedo} = \text{MEF}(A_{\text{EV}-2}, \hat{A}_{\text{EV0}}, A_{\text{EV}+2}), \tag{6}$$

$$\text{HDR-Shading} = \text{MEF}(S_{\text{EV}-2}, \hat{S}_{\text{EV0}}, S_{\text{EV}+2}), \tag{7}$$

where `MEF` denotes an HDR fusion algorithm, Multiple Exposure Fusion Yin et al. [2022]. Finally, the full HDR image is reconstructed by element-wise multiplication:

$$I_{\text{HDR}} = \text{HDR-Albedo} \odot \text{HDR-Shading}. \tag{8}$$

This pipeline leverages the physical interpretability of intrinsic decomposition and the adaptive learning capacity of RL to synthesize a high-quality HDR image from sparse LDR observations.

Table 1: Actions for Intermediate Image Generation. We define six actions to adjust brightness, contrast, and saturation. $\overline{I_Y}(t)$ and $\overline{I_{\text{cbcr}}}(t)$ denote the mean value of Y-component and Cb- and Cr-components of $I(t)$, respectively.

Action	α	β	**Action**	α	β
Contrast↑	1.05	$\overline{I_Y}(t)$	Contrast↓	0.95	$\overline{I_Y}(t)$
Brightness↑	1.05	0.00	Brightness↓	0.95	0.00
Saturation↑	1.05	$\overline{I_{\text{cbcr}}}(t)$	Saturation↓	0.95	$\overline{I_{\text{cbcr}}}(t)$

4.5 Deep Reinforcement Agent

To achieve optimal policy learning, we employ the Deep Q-Network (DQN) framework, which approximates the optimal action-value function $Q^*(s, a)$ using a neural network parameterized by θ. This network estimates the expected cumulative reward of taking action a in state s, following the Bellman optimality principle:

$$Q(s_t, a_t; \theta) \approx \mathbb{E}\left[r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-) \,\middle|\, s_t, a_t\right], \tag{9}$$

where θ^- denotes the parameters of the target network, which is periodically updated to improve training stability. The training process minimizes the temporal difference (TD) loss:

$$L(\theta) = \mathbb{E}_{(s_t, a_t, r_t, s_{t+1}) \sim D}\left[\left(y_t - Q(s_t, a_t; \theta)\right)^2\right], \tag{10}$$

with the target y_t defined as

$$y_t = r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-), \tag{11}$$

and D representing the experience replay buffer, which allows the agent to learn from past transitions and break correlations between consecutive samples.

The reward r_t is crafted to reflect HDR reconstruction quality. After each action, the agent reconstructs an HDR image via $I_t^{\text{HDR,pred}} = I_t^{\text{Albedo}} \odot I_t^{\text{Shading}}$, and computes the reward as the negative weighted sum of MSE and MAD errors:

$$r_t = -\left(\lambda_1 \cdot \text{MSE}(I_t^{\text{HDR,pred}}, I^{\text{HDR,gt}}) + \lambda_2 \cdot \text{MAD}(I_t^{\text{HDR,pred}}, I^{\text{HDR,gt}})\right), \tag{12}$$

where $I^{\text{HDR,gt}}$ denotes the ground-truth HDR image, and λ_1, λ_2 are hyperparameters to balance the two metrics.

The policy is optimized using an ε-greedy exploration strategy, allowing the agent to strike a balance between exploration and exploitation—choosing the action with the highest Q-value most of the time while still occasionally exploring alternative actions. Over successive episodes, the agent progressively refines its enhancement policy, discovering effective strategies to improve HDR quality across diverse LDR inputs. As depicted in Figure 2, the complete framework integrates modules for state-action mapping, reward computation, and Q-network training, enabling robust and adaptive HDR reconstruction through reinforcement learning.

By leveraging intrinsic image features as state representations and optimizing the Q-function via DQN, this method provides a principled way to learn enhancement policies that are both robust and interpretable. The combination of discrete, content-aware actions and deep reinforcement learning allows the framework to dynamically adapt to varying lighting conditions and content characteristics, resulting in superior HDR image reconstruction performance without relying on heuristic algorithms.

5 Experiments

We conduct our experiments on the VDS dataset Lee et al. [2018], using the Intrinsic Image Decomposition method via Ordinal Shading Careaga and Aksoy [2023] for input image pre-processing.

5.1 Pre-processing

By leveraging the decomposition method in Careaga and Aksoy [2023], an input image X is decomposed into shading (X_{sha}), albedo (X_{alb}), and residual (X_{res}) components. We then test the

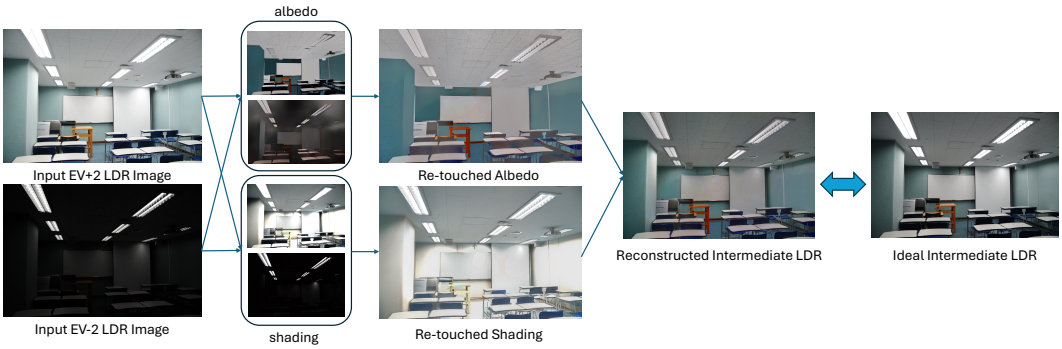

Figure 3: Pre-processing and re-touch architecture of the proposed HDR reconstruction framework. The system first decomposes an LDR image into albedo and shading components, then generates an intermediate LDR image for HDR reconstruction.

Figure 4: Effect of the re-touch stage. The goal is to produce an ideal intermediate LDR image. Once our proposed pre-processing method is applied, the output LDR image shows significant improvement.

robustness of the baseline method Yin et al. [2022] by selecting input images with different exposure values.

5.2 Re-touching Stage

We use the pre-trained weights from the baseline method Yin et al. [2022] to establish a lower-bound performance. Specifically, we choose input images with exposure values of +2 and -2 and attempt to reconstruct a suitable intermediate image for HDR reconstruction. The decomposed albedo and shading components are used as input to the same network, which independently predicts the corresponding intermediate albedo and shading. Figure 3 illustrates the results with and without the proposed pre-processing step.

5.3 Temporary Results

After processing the decomposed shading and albedo images. We can combine the re-touched albedo and shading by $\hat{X} = \hat{X}_{alb} * \hat{X}_{shd} + X_{residual}$. As shown in Figure 4, decomposing images with different exposure values into albedo and shading components leads to a substantial improvement in the quality of the generated intermediate LDR image compared to results without pre-processing.

5.4 Fine-tuning

We found that though the temporary results enhanced the baseline in the color shifting artifact. However, we found that there are still some artifacts around the highlight regions. Therefore, we decided to fine-tune the whole model with the loss function we proposed in (5) for albedo and shading separately. Once we fine-tuned the model, the mentioned artifacts disappeared since the model fits our decomposed environment much better.

Table 2: **Quantitative comparison of the predicted LDR stacks on the VDS dataset**

Method	PSNR ↑	SSIM ↑	HDR-VDP-3 ↑
Liu et al.	30.89	0.8462	7.5610
Deep recursive HDRI	32.99	**0.8731**	9.3884
CEVR	**34.67**	0.8645	**9.2704**
Yin et al.	28.21	0.8092	7.1698
Yin et al. + decomp	28.62	0.8404	7.4320
Yin et al. + decomp + finetune (**Ours**)	**28.72**	**0.8787**	**7.6343**

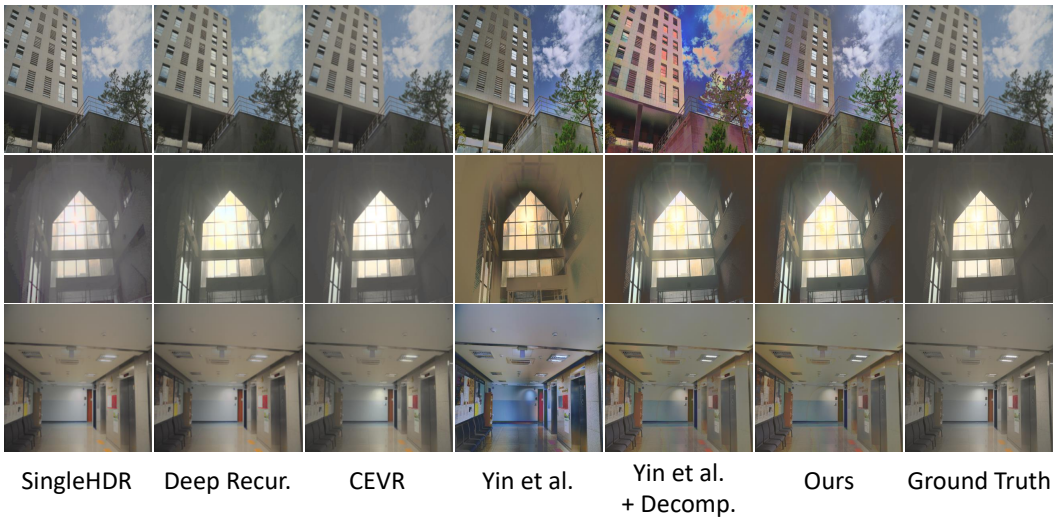

Figure 5: **Qualitative comparisons of HDR reconstruction methods.**

6 Results

In this section, we conduct the comparison with visual comparison and with metrics: PSNR, SSIM, and HDR-VDP-3 on the VDS dataset. We choose the previous state-of-the-art methods: SingleHDR Liu et al. [2020], CEVR Chen et al. [2023], and Deep Recursive HDRI Lee et al. [2018] as our competitors. The quantitative results are shown in Tab. 2, and the qualitative results are shown in Fig. 5.

To show our approach against the baseline method clearly, we also chose a case that demonstrates how our approach affects the results. The original method Yin et al. [2022] might have some color-shifting artifacts. After the decomposition pre-process, the results show a higher fidelity compared to the ground truth, but some artifacts that are not smooth between the highlight regions and the low-light regions. Our approach solves these problems and gains a better visual quality consistently.

7 Conclusion

Our approach demonstrates that reinforcement learning enhances adaptability to diverse lighting conditions, with the learned actions aligning closely with intuitive image adjustments. Additionally, we find that decomposing images into shading and albedo components and applying reinforcement learning to each separately yields superior results compared to direct manipulation of the original image. Experiments on benchmark datasets show consistent improvements across all metrics (PSNR, SSIM, HDR-VDP-3) over both baseline and vanilla methods. These encouraging results indicate that our method has the potential to outperform current state-of-the-art approaches under certain conditions.

Figure 6: **This case demonstrates that our approach has the ability to solve the color-shifting and smoothing artifacts.**

7.1 Future Work

To adapt the original DQN model to decomposed inputs, we plan to retain the existing encoder architecture, which effectively captures low-level image features. However, we propose to modify the action space to better suit the unique characteristics of albedo and shading images. Unlike raw LDR inputs, these decomposed components represent reflectance and illumination independently, and therefore may require more specialized operations. For instance, actions could include independent scaling, tone mapping, or contrast adjustments applied separately or jointly to albedo and shading, rather than global pixel-level transformations.

Furthermore, our current evaluation lacks a comprehensive comparison against state-of-the-art (SOTA) and baseline methods. In future work, we aim to benchmark our approach against leading HDR reconstruction frameworks—both traditional supervised networks and recent decomposition-based methods—under various exposure settings. This will allow us to better assess the effectiveness, robustness, and generalizability of our proposed fine-tuning strategy.

8 Expected Contributions of Each Team Member

Please describe how you would collaborate with each other and specifically what each member would contribute to this project.

- *Yo-Tin Lin:* Research on HDR reconstruction, development of RL models and optimization algorithms, evaluation of performance metrics, and benchmarking against baseline methods.
- *Jin-Ting He:* Research on state space design, development of RL models and optimization algorithms, evaluation of performance metrics, and benchmarking against baseline methods.

References

Jia-Li Yin, Bo-Hao Chen, Yan-Tsung Peng, and Hau Hwang. Automatic intermediate generation with deep reinforcement learning for robust two-exposure image fusion. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12):7853–7862, 2022. doi: 10.1109/TNNLS.2021.3088907.

Sebastian Dille, Chris Careaga, and Yağız Aksoy. Intrinsic single-image hdr reconstruction. In *Proc. ECCV*, 2024.

Su-Kai Chen, Hung-Lin Yen, Yu-Lun Liu, Min-Hung Chen, Hou-Ning Hu, Wen-Hsiao Peng, and Yen-Yu Lin. Learning continuous exposure value representations for single-image hdr reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12990–13000, 2023.

Siyeong Lee, Gwon Hwan An, and Suk-Ju Kang. Deep recursive hdri: Inverse tone mapping using generative adversarial networks. In *proceedings of the European Conference on Computer Vision (ECCV)*, pages 596–611, 2018.

Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Single-image hdr reconstruction by learning to reverse the camera pipeline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1651–1660, 2020.

Chris Careaga and Yağız Aksoy. Intrinsic image decomposition via ordinal shading. *ACM Trans. Graph.*, 43(1), 2023.