

Intrinsic Single-Image HDR Reconstruction

Sebastian Dille* , Chris Careaga* , and Yağız Aksoy 

Simon Fraser University
 Burnaby, BC V5A 1S6, Canada
sdille@sfu.ca



Fig. 1: In this work, we propose a physically-inspired HDR reconstruction pipeline by conducting reconstruction individually on intrinsic components. Our final reconstruction, as a result, is effective in recovering the high-luminance details and the lost colors.

Abstract. The low dynamic range (LDR) of common cameras fails to capture the rich contrast in natural scenes, resulting in loss of color and details in saturated pixels. Reconstructing the high dynamic range (HDR) of luminance present in the scene from single LDR photographs is an important task with many applications in computational photography and realistic display of images. The HDR reconstruction task aims to infer the lost details using the context present in the scene, requiring neural networks to understand high-level geometric and illumination cues. This makes it challenging for data-driven algorithms to generate accurate and high-resolution results. In this work, we introduce a physically-inspired remodeling of the HDR reconstruction problem in the intrinsic domain. The intrinsic model allows us to train separate networks to extend the dynamic range in the shading domain and to recover lost color details in the albedo domain. We show that dividing the problem into two simpler sub-tasks improves performance in a wide variety of photographs.

Keywords: HDR Reconstruction · Inverse Tonemapping · Computational Photography

(*) denotes equal contribution.

1 Introduction

Real-world scenes contain a wider range of luminance than most consumer-grade cameras can record or store. As a result, everyday photographs can only represent a low dynamic range (LDR) version of the scene that lacks details and color in very dark or bright regions. High dynamic range (HDR) imaging aims to capture and store the full contrast present in the scene. HDR images allow realistic display of photographs using high-contrast displays as well as many computational photography applications such as tone mapping and illumination-aware image editing. However, HDR imaging typically requires multi-exposure image stacks which limits their applications to static scenes.

To extend the HDR-enabled applications to everyday photographs, there has been a growing interest in methods to estimate HDR images from only LDR photographs called single-image HDR reconstruction, or LDR-to-HDR conversion. Single-image HDR reconstruction aims to extend the dynamic range of LDR photographs by recovering the details lost during capture. The missing information in LDR images is typically characterized by saturated pixels in bright regions, lacking textural details and color. This makes HDR reconstruction a high-level problem, requiring networks to utilize geometric and context-dependent cues in the scene to infer the lost information. Formulating this reconstruction problem using standard color spaces such as RGB or Luv requires the networks to statistically model the complex relationship between luminance, surface colors, and geometry, which limits their performance.

In this work, we present a physically-inspired remodeling of the HDR reconstruction problem in the intrinsic domain. Intrinsic image decomposition is a mid-level vision problem that separates the effect of the illumination in the scene from the reflectance of surfaces [3], representing the input as a combination of a *shading* and an *albedo* layer. The shading of an image is highly correlated with the scene geometry and illumination, while the albedo contains the color information and the high-frequency textures. The saturated pixels we aim to recover create different degradations in the shading and albedo layers. The clipping of high values is observed in the shading, while the loss of color information causes desaturation in the albedo. We propose an HDR reconstruction pipeline that individually addresses the loss of information in the LDR image in the two intrinsic components. Our HDR shading reconstruction network extends the dynamic range, taking advantage of the high correlation between shading and the scene geometry. Our LDR albedo reconstruction network, on the other hand, recovers the lost color information in saturated regions. A final refinement stage combines the two components to generate our full HDR reconstruction. Figure 1 shows different stages of our pipeline.

In our formulation, we take advantage of the intrinsic image formation model by dividing the recovery of lost details and colors into two separate tasks which are more convenient for networks to model. We show that our physically inspired approach can generate high-fidelity details with rich colors in complex in-the-wild scenes. We evaluate our method qualitatively and quantitatively to demonstrate the state-of-the-art performance of our formulation in a wide variety of scenes.

2 Related work

Common imaging sensors for photography are inherently limited in the portion of the scene’s dynamic range they can capture in a single shot. This results in saturated pixels in bright regions, representing a significant challenge in the capture and realistic display of real-world scenes.

Multi-exposure fusion methods can reconstruct an HDR image by combining a stack of LDR images taken at different exposures [9, 26, 40]. Their formulation successfully addresses contrast range extension and reversal of the quantization and nonlinearity of LDR images. The multi-exposure capture, however, assumes a static camera and scene, requiring a challenging alignment process and limiting their application scenario. To enable HDR-related applications in everyday photography, single-image HDR reconstruction approaches aim to recover the lost information directly from an LDR image. Earlier approaches apply specially designed filters [18, 34] or median-cut [1, 8] to fill in overexposed regions, using different heuristics to scale the intensity of a selected area. Similarly, Bayesian [41] and graph-based formulations [39] have been introduced for chroma restoration. Due to the high-level nature of the reconstruction problem, these low-level approaches struggle to generate realistic details in real-world scenarios.

In their concurrent publications, Eilertsen *et al.* [10] and Endo *et al.* [11] present the first data-driven approaches to the single-image HDR reconstruction problem. They train their networks for direct conversion to HDR from the LDR input, addressing quantization, nonlinearity, and saturation in the LDR image in a single step. Many works in recent literature adapt this unified inverse model [24] and develop tailored solutions for underexposed [6, 14] or overexposed [36] areas or losses for color recovery [30]. The direct estimation of HDR outputs, however, creates the challenge of an unconstrained color distribution that hinders training efficiency. Several methods [5, 11, 17, 19, 20, 40] instead adopt the concept of exposure stacks and predict a set of over- and underexposed images from the input LDR image. With each prediction still in the LDR domain, the HDR image can later be created with multi-exposure fusion [9]. Other methods [38, 42] employ generative approaches, either generating separate LDR detail layers [42] or full HDR images directly [38], with image-specific HDR reconstruction via GAN inversion. Liu *et al.* [24] approach HDR reconstruction by inverting the in-camera image formation process, modeling the dequantization, nonlinearity, and saturated pixel reconstruction in three separate steps. They show that this decomposition of the task simplifies the modeling of each individual sub-problem.

In this work, we adopt the modular formulation of Liu *et al.* [24] and focus on recovering the lost details and color information in the saturated pixels. Our method takes inspiration from prior works that formulate computational photography problems in the intrinsic image domain [4, 29]. Rather than defining a direct mapping from LDR to HDR, we divide this task into two sub-problems: shading reconstruction that extends the dynamic range, and reflectance reconstruction that recovers the color information. Our physical deconstruction of the problem significantly improves the recovery of details and color information.

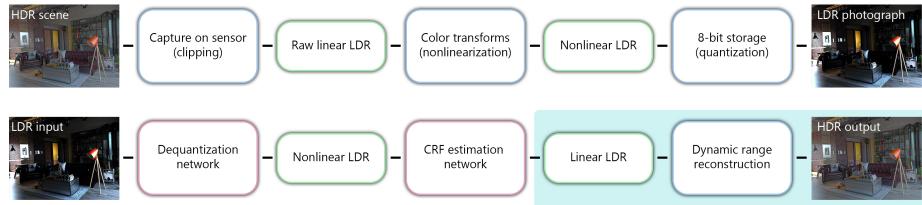


Fig. 2: The in-camera pipeline modeling LDR image formation is shown at the top. Liu *et al.* [24] propose the inverse pipeline shown on the bottom, addressing each type of information loss individually in three steps. In this work, we use the pre-trained networks of Liu *et al.* [24] for the dequantization and linearization tasks (shown in pink) and focus on developing an intrinsic model for the dynamic range reconstruction problem highlighted in teal.

3 Background and Motivation

The focus of our work is the formulation of the single-image LDR-to-HDR reconstruction problem in the intrinsic domain. We begin our discussion by first introducing the LDR imaging pipeline and the intrinsic model of image formation. We then draw connections between the two models and motivate our intrinsic HDR formulation.

3.1 LDR image formation and HDR reconstruction

The acquisition of standard RGB photographs involves a series of operations in the camera-internal image processing pipeline, or ISP, that result in different forms of information loss. The initial acquisition of the image is in a raw format, where the luminance received by the sensor from a real-world scene is scaled and stored in 16- or 32-bit data structures. The scale of the captured raw values is determined by the analog camera settings such as the shutter speed, the aperture size, and the ISO. Given the natural high dynamic range of real-world scenes, the raw values are clipped at a maximum due to saturation at the sensor level. As shown in Figure 2, the remainder of the ISP applies nonlinear color and gamma corrections before finally decreasing the bit rate to 8 bits per channel for storage.

The LDR image formation process causes three different types of corruption of the HDR information: quantization, non-linearization, and clipping. Single-image HDR reconstruction methods, as a result, have to address all three of these challenges. While many methods in the literature address all three simultaneously within a single architecture, Liu *et al.* [24] design their pipeline to address each type of information loss with a separate dedicated network. They propose a dequantization network to first increase the bit depth of the input LDR image and then estimate a CRF curve for linearization. Both tasks are convenient for a network to learn, taking advantage of the ease of dataset generation for dequantization and a simple parametric curve estimation for linearization.

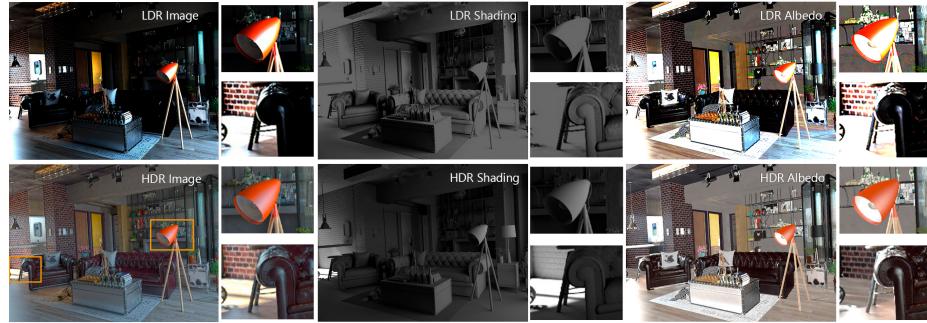


Fig. 3: We show the different effects of clipping on the albedo and the shading for a tonemapped HDR image and its LDR counterpart. When clipping the image, the LDR shading loses information compared to HDR. We scale the intensity of both shadings to show the effect. At the same time, the LDR albedo becomes saturated.

They then reconstruct the values lost to clipping with a third network in what they call the hallucination stage. This last stage of the inverse pipeline addresses the most significant challenge in single-image HDR reconstruction, namely the recovery of the highlights and the color information in clipped regions. Diverging from Liu *et al.*’s terminology, we refer to this stage as dynamic range extension. We adopt Liu *et al.*’s inverse pipeline and use their pre-trained networks for dequantization and linearization. In the following, we focus on the dynamic range extension problem and present our solution using an intrinsic model of image formation.

3.2 Intrinsic image decomposition

Intrinsic image decomposition is a fundamental mid-level computer vision problem that represents a core step in inverse rendering pipelines [2, 12]. It aims to separate the image into illumination and reflectance components following the physical image formation model. Adopting the commonly used Lambertian shading assumption with a single-channel shading representation [3], the intrinsic equation can be written as:

$$I_{rgb} = A_{rgb} \cdot S, \quad (1)$$

where I represents the linear image reflecting relative luminance, A represents albedo, i.e. the surface reflectance containing the color information, and S represents shading reflecting the effect of illumination in the scene.

Shading and albedo represent two orthogonal characteristics of a scene, one containing the effect of light and its interaction with the scene geometry and the other representing material properties. As a result, the two intrinsic components have very different characteristics as demonstrated in Figure 3. Shading is highly correlated with surface orientations, appearing smooth over continuous surfaces

and having a high gradient in depth discontinuities and shadow boundaries. Albedo, on the other hand, contains the high-frequency textures present on the materials in the scene while having a sparse set of values per image as the variation in brightness due to geometry is isolated in the shading layer. In the single-channel shading representation, albedo also contains any variation caused by colorful illumination [3].

3.3 Intrinsics in LDR and HDR

The bounded representation of LDR images limits the range of luminance values they can contain. HDR reconstruction aims to extend this range of luminance by exploiting statistical properties of real-world photographs derived from high-level scene properties such as geometry and appearance.

A brief look at the intrinsic components of HDR images, as shown in Figure 3, reveals the different characteristics of the shading and albedo layers. The distribution of reflectance values in real-world scenes typically has a narrow distribution, making the albedo a variable that can be represented linearly between $[0, 1]$. The high dynamic range of the HDR image, on the other hand, is reflected in the shading layer that similarly has an unbounded representation.

As a result, the limited dynamic range of LDR images has different effects on the shading and albedo layers. As Figure 3 demonstrates, the clipping of HDR values results in a similar clipping in shading, showing the relationship between the dynamic range and illumination that has been previously discussed in the literature [10]. The loss of colors induced by clipping, on the other hand, appears in the albedo layer as the dullification of colors in over-exposed regions.

Based on this observation, we formulate the HDR reconstruction problem separately in shading and albedo. In our setup, the network tasked with converting LDR shading into HDR shading models the high dynamic range of the illumination in the scene, taking advantage of the high correlation between shading and high-level scene properties such as geometry. We task another network with converting the LDR albedo into HDR albedo, training it to recover the lost color information in clipped regions. We detail our full pipeline in Section 4.

Our approach effectively separates the two challenges in dynamic range reconstruction, reconstruction of high luminance regions and recovery of the color information, into two separate tasks. Formulating the reconstruction of high luminance in the shading domain results in a more convenient way for the network to statistically model the problem. This can be attributed to the high correlation between shading and geometry, which is often violated in luminance due to textured surfaces as highlighted in Figure 5. Formulating the color recovery problem in the albedo domain simplifies the task similarly, allowing the exploitation of the sparsity of reflectance values in real-world scenes. Our albedo LDR-to-HDR network, as a result, acts comparable to a simple inpainting model, filling in the lost color details based on the scene content.

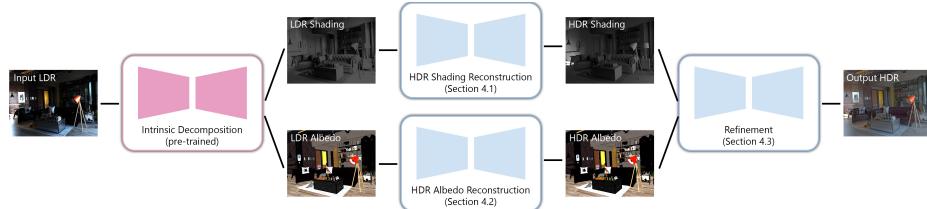


Fig. 4: Our HDR reconstruction pipeline starts with the intrinsic decomposition of the image using an off-the-shelf method [3]. The LDR reconstruction is then done individually on shading and albedo layers, which are then combined in a final refinement stage. The input LDR image is provided to each network shown in blue.

4 Dynamic Range Extension via Intrinsics

In this work, we propose a physically inspired system for dynamic range extension in the single-image HDR reconstruction pipeline. The input to our pipeline is the dequantized and linearized LDR photograph I_L . We first compute the intrinsic components of I_L , the LDR shading S_L , and albedo A_L , using the in-the-wild intrinsic decomposition method by Careaga and Aksoy [3]. We then estimate an HDR shading and an HDR albedo from the corresponding LDR components using dedicated networks. The estimated HDR intrinsic components are then combined and refined by a subsequent HDR reconstruction network yielding our final result. We present an overview of our pipeline in Figure 4. In this section, we present our formulation and training details for each step.

4.1 HDR shading reconstruction

The shading layer in the intrinsic model contains the effect of illumination in the scene. The high dynamic range of illumination in the scene, shadows, and specular objects cause a long-tailed distribution of values in the shading map. An LDR input with clipped values similarly produces a clipped LDR shading layer after intrinsic decomposition. We train our HDR shading reconstruction network to extend the range of shading values beyond this clipping.

Several works in the literature define the dynamic range extension task in the luminance domain, targeting a similar goal of isolating the effect of illumination. In their work, Eilertsen *et al.* [10] motivate this choice from a perceptual standpoint, citing earlier work that explores the effect of illumination versus surface colors in human perception [13]. The luminance, however, is computed as a linear combination of RGB channels and, as a result, still contains information on the reflectance. Shading, on the other hand, directly reflects the illumination and is accordingly highly correlated with the scene geometry as Figure 5 demonstrates. Defining the range extension task in shading instead of luminance benefits the network to model the clipped shading values by easily utilizing the geometry-related cues in the input.



Fig. 5: We show the difference between the shading representation and the L-channel from LUV. The shading correlates well with the geometry, while the L-channel includes texture information as shown in the green insets.

Following Careaga and Aksoy [3], we represent shading in the inverse domain that fits the long-tailed shading in a representation bounded in $[0, 1]$:

$$D_L = \frac{1}{S_L + 1}. \quad (2)$$

The motivation behind this inverse representation is similar to the use of the log space in HDR reconstruction: It allows the definition of balanced loss functions on variables that could take very large or small values otherwise. The advantage of the inverse domain over the log space lies in its bounded nature that allows using the variable as input without clipping or normalization as well as the use of bounded activation functions such as sigmoid.

Our HDR shading reconstruction network receives the LDR image I_L and the estimated LDR inverse shading D_L as a concatenated $h \times w \times 4$ input. Its output is defined as the HDR inverse shading D_H . We use multiple loss functions to train our network. The first one is the standard mean-squared error loss defined in the inverse space:

$$\mathcal{L}_{mse}^D = MSE(D_H^*, D_H), \quad (3)$$

where D_H^* and D_H denote the ground-truth and estimated HDR inverse shading, respectively. We also use the multi-scale gradient loss [21] as an edge-aware smoothness loss:

$$\mathcal{L}_{msg}^D = MSG(D_H^*, D_H) = \frac{1}{N} \sum_m SAD(\nabla D_H^{*m}, \nabla D_H^m), \quad (4)$$

where ∇ denotes the spatial gradient and D_H^m denotes the estimated inverse HDR shading at scale m .

Although the output of our HDR shading reconstruction network is defined only as the HDR inverse shading, we can use the ground-truth HDR image I_H^* and the intrinsic model in Eq. 1 to compute an *implied* HDR albedo \hat{A}_H [3]:

$$\hat{A}_H = \frac{I_H^*}{S_H}, \quad S_H = \frac{1 - D_H}{D_H}. \quad (5)$$

We define the MSE and multi-scale gradient loss also for the implied albedo:

$$\mathcal{L}_{mse}^{\hat{A}} = MSE(A_H^*, \hat{A}_H), \quad \mathcal{L}_{msg}^{\hat{A}} = MSG(A_H^*, \hat{A}_H), \quad (6)$$

and backpropagate these losses to our network through Eq. 5.

Our HDR shading reconstruction network only estimates the HDR shading and the HDR albedo component will be estimated with a separate network defined in the next section. Estimating the HDR intrinsic components separately results in possible reconstruction errors when these components are combined to produce an estimation for the HDR image. While we will refine the HDR reconstruction in our final stage with a dedicated network, we define these losses on the implied albedo to promote a better fit within the intrinsic model.

Our final loss for the HDR shading reconstruction network is defined as:

$$\mathcal{L}^D = \mathcal{L}_{mse}^D + \gamma \cdot \mathcal{L}_{msg}^D + \mathcal{L}_{mse}^{\hat{A}} + \gamma \cdot \mathcal{L}_{msg}^{\hat{A}}, \quad \gamma = 0.1. \quad (7)$$

4.2 HDR albedo reconstruction

The clipping in the LDR image formation process results in the loss of color information in the saturated pixels of the LDR albedo map. Our HDR albedo reconstruction network is tasked with recovering the corrupted colors.

While the LDR shading map has saturated values, signaling the regions to be reconstructed to our HDR shading reconstruction networks, the LDR albedo map does not carry the same signal as albedo maps are not saturated. To denote the regions that require color reconstruction, we provide the network with a soft guidance mask α defined as:

$$\alpha = \frac{\max(0, I_L - \lambda)}{1 - \lambda}, \quad \lambda = 0.8, \quad (8)$$

similar to the soft mask defined by Liu *et al.* [24]. It should be noted that, unlike Liu *et al.*, we use this mask as an input to our network rather than a blending mask.

Our HDR albedo reconstruction network receives the LDR image I_L , the estimated LDR albedo A_L , and the soft attention mask α , as a concatenated $h \times w \times 7$ input. Similar to our loss formulation for shading, we define the MSE and multi-scale gradient losses:

$$\mathcal{L}_{mse}^A = MSE(A_H^*, A_H), \quad \mathcal{L}_{msg}^A = MSG(A_H^*, A_H), \quad (9)$$

as well as losses on implied inverse shading $\hat{D}_H = A_H/(I_H^* + A_H)$ using the ground-truth HDR image I_H^* and the intrinsic model in Eq. 1:

$$\mathcal{L}_{mse}^{\hat{D}} = MSE(D_H^*, \hat{D}_H), \quad \mathcal{L}_{msg}^{\hat{D}} = MSG(D_H^*, \hat{D}_H). \quad (10)$$

Our final loss for the HDR albedo reconstruction network is defined as:

$$\mathcal{L}^A = \mathcal{L}_{mse}^A + \gamma \cdot \mathcal{L}_{msg}^A + \mathcal{L}_{mse}^{\hat{D}} + \gamma \cdot \mathcal{L}_{msg}^{\hat{D}}, \quad \gamma = 0.1. \quad (11)$$

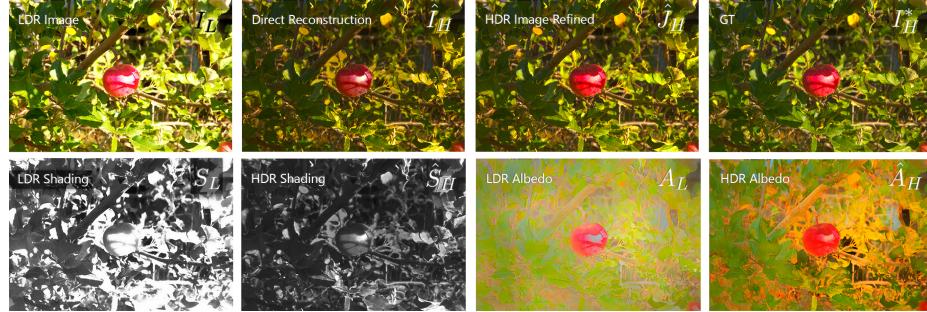


Fig. 6: We show the intermediate results of our pipeline from an LDR input compared with its ground truth. The direct HDR reconstruction from the estimated HDR intrinsics can lead to small artifacts which are reduced by our refinement network.

4.3 Refinement

The outputs of our HDR reconstruction networks for shading and albedo can be combined using the intrinsic model to give us an initial estimation for the HDR image:

$$\hat{I}_H = A_H * S_H, \quad S_H = \frac{1 - D_H}{D_H}. \quad (12)$$

However, as the two intrinsic components are estimated separately, the perfect reconstruction of the HDR image is not guaranteed as shown in Figure 6. In the final step of our pipeline, we train a refinement network that takes the intrinsic HDR reconstruction results as input and generates our final HDR reconstruction result.

This network has a similar goal to the hallucination network used in Liu *et al.* [24]: Reconstructing the final HDR image by extending dynamic range. However, as it receives the intrinsic reconstruction of the HDR image readily as input, its task is significantly easier. Unlike a network trained for direct LDR-to-HDR conversion, it does not need to reason about the high-level cues in the image which is performed by our HDR shading reconstruction network. Similarly, it does not need to recover the corrupted colors in saturated regions, which is performed by our HDR albedo reconstruction network. It is simply tasked with combining the rich input information it receives to reconstruct the final HDR image, which allows it to generate more accurate and sharper results.

Following the inverse representation we adopt for shading, we use the inverse representation for the HDR images as a bounded representation that fits the long-tailed distribution in HDR images between $[0, 1]$:

$$J_H = \frac{1}{I_H + 1}, \quad \hat{J}_H = \frac{1}{\hat{I}_H + 1}, \quad J_H^* = \frac{1}{I_H^* + 1}, \quad (13)$$

where J_H , and J_H^* denote the estimated and ground-truth inverse HDR images and \hat{J}_H is the inferred inverse HDR image following Eq. 12.

Our refinement network receives the LDR image I_L , the inferred inverse HDR image \hat{J}_H , the estimated HDR inverse shading D_L , and the estimated HDR albedo A_L as a concatenated $h \times w \times 10$ input. We use the MSE and multi-scale gradient losses to train our refinement network:

$$\mathcal{L}_{mse}^J = MSE(J_H^*, J_H), \quad \mathcal{L}_{msg}^J = MSG(J_H^*, J_H), \quad (14)$$

with the final loss defined as $\mathcal{L}^J = \mathcal{L}_{mse}^J + \gamma \cdot \mathcal{L}_{msg}^J$, $\gamma = 0.1$.

4.4 Network architecture and training

We utilize the same architecture for our 3 networks. We use EfficientNet [37] as the backbone with the decoder from [33], replacing the output activation with a **sigmoid**. We use RAdam [23] as the optimizer with a learning rate of $1e^{-4}$ and CosineAnnealing [25] scheduling.

To train our intrinsic formulation, we use the provided synthetic intrinsic ground truth from Hypersim [35] and construct ground truth intrinsics from the raw images of the MultiIllum dataset [31] as described in [3]. We follow the preprocessing steps from Liu *et al.* [24] to create clipped, linearized LDR images with randomly sampled exposure values $t \in [-3..3]$ from the ground truth data which we then decompose into our LDR input intrinsics using [3].

Our refinement network in contrast does not require supervision in intrinsic space. We can thus extend our training set by real images from MultiRAW [22], LSMI [16], and the HDR-Real training subset from Liu *et al.* [24]. For this stage, we follow the ISP model in [24] to convert the ground truth images into JPEG files, use their pre-trained networks to retrieve linearized LDR images, and again retrieve the intrinsic decomposition via [3]. We provide a more detailed description of our training procedure in the supplementary material.

5 Experiments

In this section, we present a thorough evaluation of our proposed method using multiple standard benchmarks. First, we discuss the datasets, metrics and prior work used in our experimental setup, then we present qualitative and quantitative comparisons to state-of-the-art approaches. We present a more detailed evaluation and our ablation studies in the supplementary material.

5.1 Quantitative evaluation

To quantitatively evaluate our approach against existing methods we follow the evaluation protocol proposed by Hanji *et al.* [15]. For a given ground truth HDR image, we follow the preprocessing of Liu *et al.* [24] and map the values into the range [1 – 1000]. We determine the best scale to align the input reconstruction with the ground truth with least-squares, using only values between the 10th and 90th percentiles to avoid outliers. We then perform CRF correction to match the reconstruction to the ground truth by fitting a per-channel third-degree polynomial in log-RGB space.

Table 1: Quantitative results against state-of-the-art using the HDR-VDP3 metric. Our method achieves state-of-the-art performance on the difficult high-definition SI-HDR benchmark dataset. We are state-of-the-art or competitive on all other datasets.

Method	SIHDR	HDR-EYE	HDR-REAL	HDR-SYNTH	RAISE
Ours	8.96	8.96	7.57	8.21	8.89
DrTMO [11]	8.27	8.72	7.57	<u>8.27</u>	8.64
HDR-CNN [10]	8.39	8.61	6.70	8.03	8.47
ExpandNet [30]	8.67	8.49	6.97	7.69	8.27
Single-HDR [24]	8.79	8.96	<u>7.47</u>	8.51	8.95
Mask-HDR [36]	8.25	8.60	7.10	8.14	8.44
HDRUNet [6]	<u>8.82</u>	8.46	5.71	7.61	8.04
Multi-Exp Gen. [19]	8.64	<u>8.87</u>	7.34	8.16	8.66
Lightweight [14]	8.26	8.20	7.03	7.31	7.68

Metrics We follow the guidance of Hanji *et al.* [15] when choosing metrics to evaluate our approach. In their study, they find that certain metrics better correlate with results from qualitative experiments. We focus our evaluation on the popular HDR-VDP3 [28] metric as it is purpose-built for HDR evaluation and models the subjective evaluation of a human observer. We specifically use the "quality" task of HDR-VDP3. Furthermore, we evaluate multiple recommended SDR metrics using the PU21 [27] encoding and provide those results in the supplementary material.

Datasets We evaluate our method on five publicly available datasets. We use the recent high-resolution SI-HDR Benchmark [15] as well as commonly used real-world datasets HDR-Eye [32], HDR-Real [24], HDR-Synth [24], and RAISE [7].

We compare our method against a collection of baselines that share their implementation publicly. Table 1 shows the results of our quantitative evaluation. Each value represents the average HDR-VDP3 score in Just-Objectionable-Differences (JOD) units. We show a significant improvement in quality in the SI-HDR Benchmark [15] while achieving state-of-the-art performance on 3 of the other 4 datasets. As our qualitative evaluation in the next section and our results presented in the supplementary material demonstrate, the favorable performance of our method comes from our physically-motivated intrinsic formulation, allowing us to reliably reconstruct the high luminance details and recover lost colors. We provide additional metrics and analysis in the supplementary material.

5.2 Qualitative evaluation

We present qualitative comparisons in Figures 7 and 8, showing the differences in both detail and color reconstruction. On SI-HDR [15], prior works struggle with predicting accurate details around edges where over- and under-exposed pixels meet. This becomes notable around the leaves in the [roof] scene with visible halos and color shifts. On the floor in the [book] scene, most competitors fail to

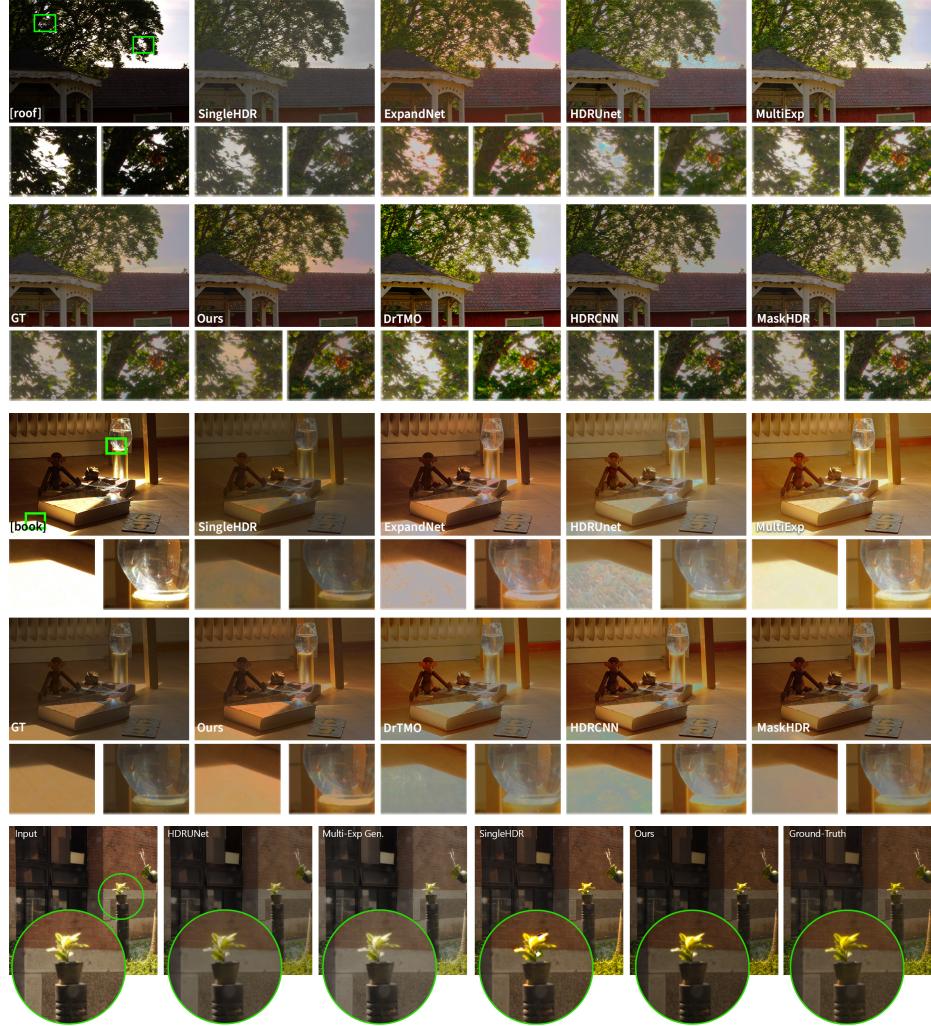


Fig. 7: We show results from our method on SI-HDR [15] to the other baselines. We refer to Sec. 5.2 and the supplementary material for an in-depth discussion.

reconstruct the homogenous surface and hallucinate patterns and color blobs. In both cases, our method seamlessly recovers missing details in over-exposed areas without artifacts. For the woman in Fig. 8, HDRUnet [6] and SingleHDR [24] both exhibit strong artifacts as bright spots in the hand and the face, similar to the faces of the [bust] and the [kids]. Our method in contrast can recover the correct color information due to our modeling of the albedo. The results showcase the advantage of the intrinsic formulation for creating visually pleasing results. We provide an extensive qualitative comparison in the supplementary material.

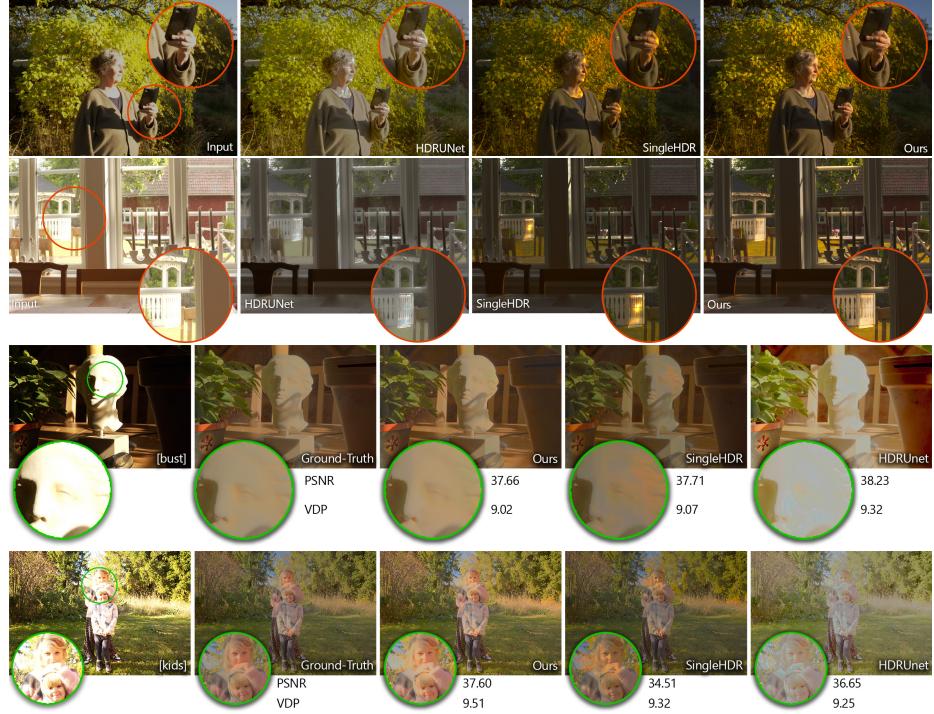


Fig. 8: We compare results from our method with a focus on color reconstruction against HDRUNet [6] and SingleHDR [24]. We refer to Sec. 5.2 and the supplementary material for an in-depth discussion.

6 Conclusion

In this work, we remodel the HDR reconstruction problem in the intrinsic domain and propose a multi-stage pipeline to recover lost details. The luminance range extension can be conveniently modeled in the shading space which allows the network to better model the correlation between illumination and geometry. Similarly, neural networks can easily take advantage of albedo sparsity to recover lost color information with high fidelity. Through qualitative and quantitative evaluations, we show that our physically-inspired modeling of the HDR reconstruction problem achieves state-of-the-art performance, improving upon prior work in terms of both detail recovery and color reproduction in a wide variety of in-the-wild photographs.

Acknowledgements

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [RGPIN-2020-05375].

References

1. Banterle, F., Ledda, P., Debattista, K., Chalmers, A.: Inverse tone mapping. In: Proc. GRAPHITE (2006)
2. Barrow, H., Tenenbaum, J., Hanson, A., Riseman, E.: Recovering intrinsic scene characteristics. *Comput. vis. syst.* **2**(3-26), 2 (1978)
3. Careaga, C., Aksoy, Y.: Intrinsic image decomposition via ordinal shading. *ACM Trans. Graph.* **43**(1) (2023)
4. Careaga, C., Miangoleh, S.M.H., Aksoy, Y.: Intrinsic harmonization for illumination-aware compositing. In: Proc. SIGGRAPH Asia (2023)
5. Chen, S.K., Yen, H.L., Liu, Y.L., Chen, M.H., Hu, H.N., Peng, W.H., Lin, Y.Y.: CEVR: Learning continuous exposure value representations for single-image HDR reconstruction. In: Proc. ICCV (2023)
6. Chen, X., Liu, Y., Zhang, Z., Qiao, Y., Dong, C.: HDRUnet: Single image HDR reconstruction with denoising and dequantization. In: Proc. CVPR (2021)
7. Dang-Nguyen, D.T., Pasquini, C., Conotter, V., Boato, G.: RAISE: a raw images dataset for digital image forensics. In: Proc. MMSys (2015)
8. Debevec, P.: A median cut algorithm for light probe sampling. In: ACM SIGGRAPH 2005 Posters. ACM (2005)
9. Debevec, P.E., Malik, J.: Recovering high dynamic range radiance maps from photographs. In: ACM Trans. Graph. (1997)
10. Eilertsen, G., Kronander, J., Denes, G., Mantiuk, R.K., Unger, J.: HDR image reconstruction from a single exposure using deep CNNs. *ACM Trans. Graph.* **36**(6) (2017)
11. Endo, Y., Kanamori, Y., Mitani, J.: Deep reverse tone mapping. *ACM Trans. Graph.* **36**(6) (2017)
12. Garces, E., Rodriguez-Pardo, C., Casas, D., Lopez-Moreno, J.: A survey on intrinsic images: Delving deep into lambert and beyond. *Int. J. Comput. Vision* (2022)
13. Gilchrist, A., Jacobsen, A.: Perception of lightness and illumination in a world of one reflectance. *Perception* **13**, 5–19 (02 1984)
14. Guo, C., Xiuhua, J.: LHDR: HDR reconstruction for legacy content using a lightweight DNN. In: Proc. ACCV (2022)
15. Hanji, P., Mantiuk, R., Eilertsen, G., Hajisharif, S., Unger, J.: Comparison of single image HDR reconstruction methods — the caveats of quality assessment. In: ACM Trans. Graph. (2022)
16. Kim, D., Kim, J., Nam, S., Lee, D., Lee, Y., Kang, N., Lee, H.E., Yoo, B., Han, J.J., Kim, S.J.: Large scale multi-illuminant (LSMI) dataset for developing white balance algorithm under mixed illumination. In: Proc. CVPR (2021)
17. Kim, J., Lee, S., Kang, S.J.: End-to-end differentiable learning to HDR image synthesis for multi-exposure images. In: Proc. AAAI (2021)
18. Kovaleski, R.P., Oliveira, M.M.: High-quality reverse tone mapping for a wide range of exposures. In: Proc. SIBGRAPI (2014)
19. Le, P.H., Le, Q., Nguyen, R., Hua, B.S.: Single-image HDR reconstruction by multi-exposure generation. In: Proc. WACV (2023)
20. Lee, S., An, G.H., Kang, S.J.: Deep recursive HDRI: Inverse tone mapping using generative adversarial networks. In: Proc. ECCV (2018)
21. Li, Z., Snavely, N.: Learning intrinsic image decomposition from watching the world. In: Proc. CVPR (2018)
22. Li, Z., Lu, M., Zhang, X., Feng, X., Asif, M.S., Ma, Z.: Efficient visual computing with camera raw snapshots. *IEEE Trans. Pattern Anal. Mach. Intell.* (2024)

23. Liu, L., Jiang, H., He, P., Chen, W., Liu, X., Gao, J., Han, J.: On the variance of the adaptive learning rate and beyond. In: Proc. ICLR (2020)
24. Liu, Y.L., Lai, W.S., Chen, Y.S., Kao, Y.L., Yang, M.H., Chuang, Y.Y., Huang, J.B.: Single-image HDR reconstruction by learning to reverse the camera pipeline. In: Proc. CVPR (2020)
25. Loshchilov, I., Hutter, F.: SGDR: Stochastic gradient descent with warm restarts. In: Proc. ICLR (2017)
26. Mann, S., Picard, R.: On being "undigital" with digital cameras: Extending dynamic range by combining differently exposed pictures. In: Proc. IS&T Annual Conference (1995)
27. Mantiuk, R.K., Azimi, M.: Pu21: A novel perceptually uniform encoding for adapting existing quality metrics for hdr. In: Proc. PCS (2021)
28. Mantiuk, R.K., Hammou, D., Hanji, P.: HDR-VDP-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)
29. Maralan, S.S., Careaga, C., Aksoy, Y.: Computational flash photography through intrinsics. In: Proc. CVPR (2023)
30. Marnerides, D., Bashford-Rogers, T., Hatchett, J., Debattista, K.: Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. Comput. Graph. Forum **37**(2) (2018)
31. Murmann, L., Gharbi, M., Aittala, M., Durand, F.: A multi-illumination dataset of indoor object appearance. In: Proc. ICCV (2019)
32. Nemoto, H., Korshunov, P., Hanhart, P., Ebrahimi, T.: Visual attention in LDR and HDR images. In: Proc. VPQM (2015)
33. Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., Koltun, V.: Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. IEEE Trans. Pattern Anal. Mach. Intell. **44**(3) (2022)
34. Rempel, A.G., Trentacoste, M., Seetzen, H., Young, H.D., Heidrich, W., Whitehead, L., Ward, G.: LDR2HDR: on-the-fly reverse tone mapping of legacy video and photographs. ACM Trans. Graph. **26**(3) (2007)
35. Roberts, M., Ramapuram, J., Ranjan, A., Kumar, A., Bautista, M.A., Paczan, N., Webb, R., Susskind, J.M.: Hypersim: A photorealistic synthetic dataset for holistic indoor scene understanding. In: Proc. ICCV (2021)
36. Santos, M.S., Ren, T.I., Kalantari, N.K.: Single image HDR reconstruction using a CNN with masked features and perceptual loss. ACM Trans. Graph. **39**(4) (2020)
37. Tan, M., Le, Q.: EfficientNet: Rethinking model scaling for convolutional neural networks. In: Proc. ICML (2019)
38. Wang, C., Serrano, A., Pan, X., Chen, B., Myszkowski, K., Seidel, H.P., Theobalt, C., Leimkühler, T.: Glowgan: Unsupervised learning of HDR images from LDR images in the wild. In: Proc. ICCV (2023)
39. Xu, D., Doutre, C., Nasiopoulos, P.: Correction of clipped pixels in color images. IEEE Trans. Vis. Comput. Graph. **17**(3) (2010)
40. Zhang, N., Ye, Y., Zhao, Y., Wang, R.: Revisiting the stack-based inverse tone mapping. In: Proc. CVPR (2023)
41. Zhang, X., Brainard, D.H.: Estimation of saturated pixel values in digital color imaging. J. Optical Soc. Am. A **21**(12) (2004)
42. Zhang, Y., Aydin, T.: Deep HDR estimation with generative detail reconstruction. Comput. Graph. Gorum **40**(2), 179–190 (2021)