

GTA-HDR: A Large-Scale Synthetic Dataset for HDR Image Reconstruction

Hrishav Bakul Barua
Monash University & TCS Research
hrishav.barua@monash.edu

Kalin Stefanov
Monash University
kalin.stefanov@monash.edu

KokSheik Wong
Monash University
wong.koksheik@monash.edu

Abhinav Dhall
Monash University
abhinav.dhall@monash.edu

Ganesh Krishnasamy
Monash University
ganesh.krishnasamy@monash.edu

Abstract

High Dynamic Range (HDR) content (i.e., images and videos) has a broad range of applications. However, capturing HDR content from real-world scenes is expensive and time-consuming. Therefore, the challenging task of reconstructing visually accurate HDR images from their Low Dynamic Range (LDR) counterparts is gaining attention in the vision research community. A major challenge is the lack of datasets, which capture diverse scene conditions (e.g., lighting, weather, locations) and various image features (e.g., color, contrast, saturation). To address this gap, we introduce GTA-HDR, a large-scale synthetic dataset of photo-realistic HDR images sampled from the GTA-V video game. We perform thorough evaluation of the proposed dataset, which enables significant qualitative and quantitative improvements of the state-of-the-art HDR image reconstruction methods. Furthermore, we demonstrate the effectiveness of the proposed dataset and its impact on additional computer vision tasks including 3D human pose estimation, human body part segmentation, and holistic scene segmentation. The dataset, data collection pipeline, and evaluation code are available at: <https://github.com/HrishavBakulBarua/GTA-HDR>.

1. Introduction

High Dynamic Range (HDR) [3] content (i.e., images and videos) has been adopted widely in various domains including entertainment [23], gaming and augmented/virtual reality [57], medical imaging [26], computational photography [48], and robotics/robot vision [70]. However, capturing HDR content from real-world scenes is costly and time-consuming. Therefore, HDR image reconstruction from Low Dynamic Range (LDR) counterparts has been an active area of research in the last several years [29, 33, 62, 63, 65, 66]. The literature proposes a multitude of methods for

HDR image reconstruction that are gradually shifting from traditional non-learning techniques [27, 35, 41, 44] towards data-driven learning-based, such as Generative Adversarial Networks (GAN) [17] and Diffusion Models [11].

Given that the performance of any data-driven learning-based method for HDR image reconstruction largely depends on the size and diversity of the datasets used for development, there is a significant gap in the publicly available datasets required to advance this research direction. Specifically, the existing datasets are either: 1) Not sufficiently large [13, 14, 28, 30, 38, 47, 50]; 2) Not having satisfactory resolution [15, 73]; 3) Having limited scene diversity [28, 30, 50]; 4) Having limited image variations [8, 28, 32, 40]; or 5) Absence of ground truth HDR images [12, 58, 64]. Furthermore, currently, there are no available datasets that adequately address the problem of no-reference HDR image quality assessment, which demands vast collections of ground truth HDR and distorted HDR pairs [2, 4, 5]. In summary, there is a substantial research gap pertaining to benchmark datasets needed to advance the research on HDR image reconstruction, hence motivating the creation of an appropriate large-scale dataset.

Video games have been used for creation and annotation of various large-scale datasets in diverse computer vision tasks [74] including 3D human pose and motion reconstruction [9, 72], semantic segmentation [1], 3D scene layout and visual odometry [54], pedestrian detection and tracking [16], object detection and 3D mesh recovery [25], optical flow and depth estimation [36]. Drawing inspiration from the success of various data-driven learning-based methods developed with video game data, in this paper, we propose GTA-HDR, a large-scale synthetic dataset for HDR image reconstruction, sampled from the photo-realistic (i.e., HDR-10 enabled) game Grand Theft Auto V (GTA-V) by Rockstar Games. Previous work has also used other video games including Hitman [55], Witcher 3 [52], and Far Cry Primal [71] which contain highly realistic and detailed

Table 1. **Publicly available datasets for HDR image reconstruction.** *GT*: Ground truth; *Dis*: Distorted; *: Minimum image resolution. Cf., Section 3.2 for description of *In-the-wild*, *Scene diversity* and *Image diversity*

Dataset	Year	Type	#HDR _{GT}	Resolution	In-the-wild	HDR _{Dis}	Scene diversity	Image diversity
HDR-Eye [47]	2015	Synthetic	46	512 × 512	✗	✗	✗	✗
City Scene [15, 73]	2017	Mixed	41222	128 × 64	✗	✗	✓	✗
Kalantari <i>et al.</i> [30]	2017	Real	89	1500 × 1000	✗	✗	✗	✗
Endo <i>et al.</i> [14]	2017	Synthetic	1043	512 × 512	✗	✗	✗	✗
Eilertsen <i>et al.</i> [13]	2017	Synthetic	96	1024 × 768	✗	✗	✗	✗
Lee <i>et al.</i> [38]	2018	Synthetic	96	512 × 512	✗	✗	✗	✗
Cai <i>et al.</i> [8]	2018	Synthetic	4413	3072 × 1620*	✗	✗	✗	✗
Prabhakar <i>et al.</i> [50]	2019	Real	582	1200 × 900*	✗	✗	✗	✗
LDR-HDR Pair [28]	2020	Real	176	1024 × 1024	✗	✗	✗	✗
HDR-Synth & HDR-Real [40]	2020	Mixed	20537	512 × 512	✗	✗	✗	✓
SI-HDR [20, 21]	2022	Real	181	1920 × 1280	✗	✗	✓	✗
GTA-HDR (ours)	2024	Synthetic	40000	512 × 512 1024 × 1024	✓	✓	✓	✓

worlds with high fidelity. However, those games lack the diversity of scenes, which is the main trait of GTA-V.

The thorough evaluation of the proposed dataset demonstrates important advantages it brings to the state-of-the-art in HDR image reconstruction: 1) The GTA-HDR dataset in combination with other real and synthetic datasets enables significant improvements in the quality of the reconstructed HDR images; and 2) The GTA-HDR dataset fills a gap not covered by any of the publicly available real and synthetic datasets and as such, contributes towards better generalization capabilities for HDR image reconstruction.

- We propose GTA-HDR, a large-scale synthetic dataset to complement existing real and synthetic HDR image reconstruction datasets (cf., Sec. 3).
- We perform thorough validation to highlight the contribution of GTA-HDR to the quality of HDR image reconstruction (cf., Sec. 5 and *Supplementary*).
- We demonstrate the impact of GTA-HDR on the state-of-the-art in other tasks including 3D human pose estimation, human body part segmentation, and holistic scene segmentation (cf., *Supplementary*).

2. Related Work

2.1. Inverse Tone Mapping

Tone mapping [19] is the process of mapping the colors of HDR images with a wide range of illumination levels to LDR images appropriate for standard displays with limited dynamic range. Inverse tone mapping [66] is the reverse process accomplished with either traditional non-learning methods or data-driven approaches. Given the sensor irradiance E and exposure time Δt , the function $f_{crf}(E\Delta t)$ represents the tone mapping process, which outputs I_{LDR} given I_{HDR} images captured by the camera. The main goal of any HDR image reconstruction technique is to reverse the tone mapping process using another function $f_{crf}^{-1}(I_{LDR})/\Delta t$, which outputs reconstructed I_{HDR}

given I_{LDR} images. The main challenge is that the steps in $f_{crf}(E\Delta t)$ are generally not reversible [37].

Khan *et al.* [31] proposed a feedback mechanism based on Convolutional Neural Network (CNN) to generate HDR images from single-exposed LDR. Barua *et al.* [7] utilized multi-exposed features and perceptual losses along with low- and mid-level feature guidance in generating visually accurate HDR images. Le *et al.* [37] leveraged a Neural Network (NN) architecture for camera response inversion to generate pixel radiance and details for various exposures. Liu *et al.* [40] proposed an architecture consisting of three CNN that approximates the three sub-tasks in the tone mapping process but in the reverse order. Li *et al.* [39] presented a combination of an attention mechanism and CNN that can recover over/underexposed regions of LDR images. Santos *et al.* [56] leveraged a feature masking mechanism that helps in reconstructing saturated pixels of LDR images resulting in better visual and perceptual quality of the reconstructed HDR images. Eilertsen *et al.* [13] proposed CNN for accurate prediction of HDR pixels from the under/overexposed counterparts in LDR images. Luzardo *et al.* [42] addressed the low peak brightness issues in the reconstructed HDR images to enhance the artistic intent. Cao *et al.* [10] proposed a method that combines the outputs of preliminary HDR results from a channel-decoupled kernel and pixel-wise output from another architecture resulting in high-quality HDR images. Jang *et al.* [28] explored the concept of histogram and color differences between HDR and multi-exposed LDR pairs. Neural Radiance Fields (NeRF) [46] have been used to learn implicit color and radiance fields and perform HDR view synthesis [26]. Some methods address dynamic scenes [30], while others address multiple tasks, e.g., denoising and deblurring and fuse them into an HDR image reconstruction pipeline.

2.2. Datasets

There has been a shift in the research on HDR image reconstruction from traditional non-learning methods to data-

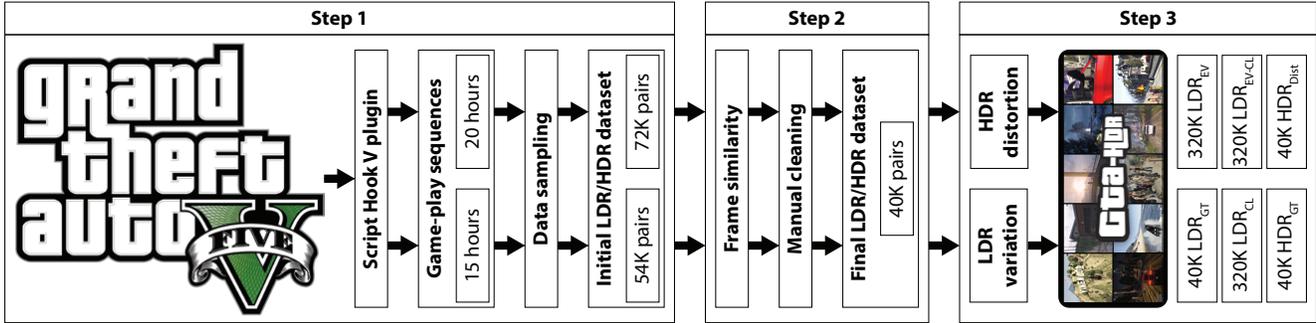


Figure 1. **GTA-HDR dataset collection pipeline.** *GT*: Ground truth; *Dis*: Distorted; *EV*: Exposure value; *CL*: Contrast level. *Note*: The GTA-V logo is retrieved from Google Images. *Cf.*, Section 3.1 for detailed description of the dataset collection pipeline.

driven approaches based on GAN [51], CNN [59], Diffusion Models [11], and NeRF [26]. These data-driven techniques require a significant amount of training data drawing attention to the limitations of the current publicly available datasets. Table 1 summarizes existing public datasets.

Real datasets [12,20,21,28,30,50,58,64] include images with sufficient resolution, however, their main limitation is their size (*i.e.*, low number of images). It is difficult, time-consuming, and costly to collect real-world data using an HDR camera that covers a variety of scenes (*e.g.*, indoor, outdoor, in-the-wild), lightning conditions, and image characteristics (*e.g.*, different levels of contrast, radiance, saturation). The RAISE dataset [12] consists of real images and is of moderate size but it lacks ground truth HDR images and therefore is applicable for evaluation purposes only.

Synthetic datasets [8,13,14,38,47], on the other hand, provide ground truth HDR images, however, only a few consist of large number of images. In addition, these datasets generally lack images with appropriate resolution and diversity. The synthetic video dataset proposed in [32], originally designed for the development of super-resolution video generation methods, could potentially support the development of data-driven HDR image reconstruction methods after appropriate data pre-processing.

Several datasets [15,40,73] include images from both real and synthetic scenes. These datasets provide a sufficient amount of ground truth HDR images, however, they lack appropriate image resolution and diversity.

3. GTA-HDR Dataset

The GTA-HDR dataset addresses some of the limitations of the existing datasets for HDR image reconstruction. The main characteristics of GTA-HDR are the diversity of scenes and variety of images included in the dataset (*e.g.*, forests, mountains, coasts, cities). GTA-HDR includes scenes from different times (*e.g.*, morning, evening, daytime, night) and different weather conditions (*e.g.*, rainy, snowy, sunny, misty). This variety of scenes is an expensive and effort-demanding task to collect in a real-world con-

text. To our knowledge, this work is the first to use video game data to collect and curate synthetic {LDR,HDR} image pairs to support the development of inverse tone mapping data-driven methods.

3.1. Dataset Collection

We performed a thorough data collection and curation, adopting a similar approach as described in [74]. We used 2 full game-play sequences (*i.e.*, playing the story from the beginning until the end) from the GTA-V game to extract {LDR,HDR} image pairs at multiple resolutions (*i.e.*, 512×512 and 1024×1024). GTA-V has built-in HDR-10 support for displaying video sequences on HDR displays. Fig. 1 depicts the entire data collection pipeline for GTA-HDR.

Similar to [74], we used Script Hook V plugin to capture HDR images from GTA-V game-play sequences. Other tools for GTA-V game data extraction include RenderDoc Debugger and customized RenderDoc for the Game Data platform. The normal duration of a GTA-V game-play is approximately 31.5 hours (*i.e.*, going through all basic aspects of the story), and it can extend up to about 82 hours (*i.e.*, visit all aspects of the story thoroughly). We collected data from 2 game-play sequences, one of around 15 hours and another of approximately 20 hours. We sampled 1 frame per second, resulting in approximately 54K and 72K {LDR,HDR} image pairs, respectively.

The steps in the data collection process are as follows: 1) We used Script Hook V to extract {LDR,HDR} image pairs with 1Hz frequency; 2) We removed frames that are similar to the previous or next frames in the sequence. The similarity between two consecutive frames in the sequence is based on [34]; we discarded frames that have a similarity score higher than 0.8. We also did a manual cleaning of the collected data to remove unwanted scenes (*e.g.*, images containing violence and other objectionable actions or items). Finally, we ensured that the collected data has an even distribution of scenes from indoor, outdoor, and in-the-wild environments, resulting in a total of 40K {LDR,HDR} pairs; and 3) We performed transformations on the original

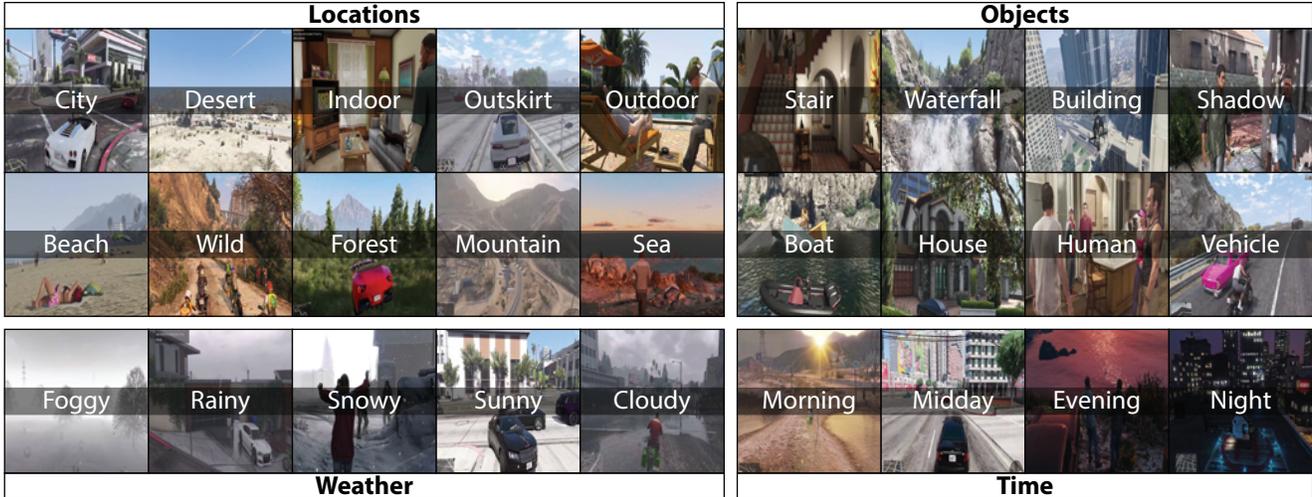


Figure 2. **GTA-HDR dataset scene diversity.** Samples from the dataset with multiple variations in location, weather, objects and time.

LDR images to generate multi-exposed LDR images (*i.e.*, exposure values EV 0, ± 1 , ± 2 , ± 3 , and ± 4) [37] and different contrast levels [8]. This step results in $40K \times 25 = 1M$ LDR images. Apart from the 40K original HDR images, we also generated 40K distorted HDR images by utilizing the following state-of-the-art methods: 20K images were generated using [31], 10K with [37], and [40, 48] were used to produce 5K images each.

3.2. Dataset Characteristics

One of the limitations of existing datasets is the low diversity of scenes and images. To address this limitation, the GTA-HDR dataset includes a wide variety of scenes (*e.g.*, indoor, outdoor, in-the-wild, multiple locations, weather conditions, lighting conditions, and time-of-day) and images (*e.g.*, LDR images with 9 different exposure values EV 0, ± 1 , ± 2 , ± 3 , and ± 4 and contrast levels).

Scene Diversity. Real-life scenes can have a wide range of variety in terms of locations, landscapes, objects, humans, animals, buildings, weather, and lighting conditions. Fig. 2 depicts samples from the GTA-HDR dataset with multiple variations in location, weather, objects and time. The diverse set of locations ensures a thorough coverage of pixel colors, brightness, and luminance. The weather conditions contribute to the rich gamut of brightness levels, *e.g.*, sunny weather scenes will have a larger number of bright pixels than cloudy or rainy scenes. Scenes at different time (*i.e.*, morning, midday, evening, and night) also contribute to different lighting conditions. The diversity of objects captures different color hues, contrast and saturation levels.

Image Diversity. Images can have a diverse range of color hues, saturation, exposure, and contrast levels. For any image-to-image translation dataset, it is important to include a sufficient amount of samples from these categories. Therefore, we introduced different exposure, brightness,

and contrast levels in the GTA-HDR dataset. Considering all variations, the dataset includes 24 versions of the original LDR images. The final set of images amounts to a total of $40K \times 25 = 1M$ LDR, 40K HDR, and 40K distorted HDR images. Fig. 3 provides samples from the GTA-HDR dataset with 9 exposure levels (*i.e.*, exposure values EV 0, ± 1 , ± 2 , ± 3 , and ± 4) and 9 contrast levels of the LDR images. The first two rows show the LDR images with varying contrast and EV levels. The LDR images with normal contrast level are shown in the middle, while LDR images with increasing EV and contrast are shown in sequential order toward the right, and vice versa. On the extreme right, the corresponding HDR and a sample of distorted HDR (*i.e.*, saturation altered HDR) images are illustrated. Here, saturation alteration, contrast alteration, color hue alteration, and noise addition are applied to the HDR to produce the respective distorted HDR images. Similarly, the second and third rows show the LDR images with only varied EV while the contrast is kept constant (*i.e.*, keeping the contrast level of the original LDR image). Finally, in the fifth and sixth rows, the EV is kept constant (*i.e.*, keeping the EV 0 of the original LDR image) and the contrast levels are varied.

4. Experiments

4.1. Experimental Setup

Methods. To demonstrate the effectiveness of GTA-HDR for HDR image reconstruction, we evaluated state-of-the-art methods including FHDR [31], SingleHDR [37, 40], HDRCNN [13], HDR-GAN [49], DrTMO [14], ArtHDR-Net [7], and HistoHDR-Net [6]. Most of these methods are designed for single-exposed LDR image inputs, while HDR-GAN is designed for three multi-exposed LDR image inputs, SingleHDR [37] for two or more multi-exposed LDR image inputs without HDR supervision, and

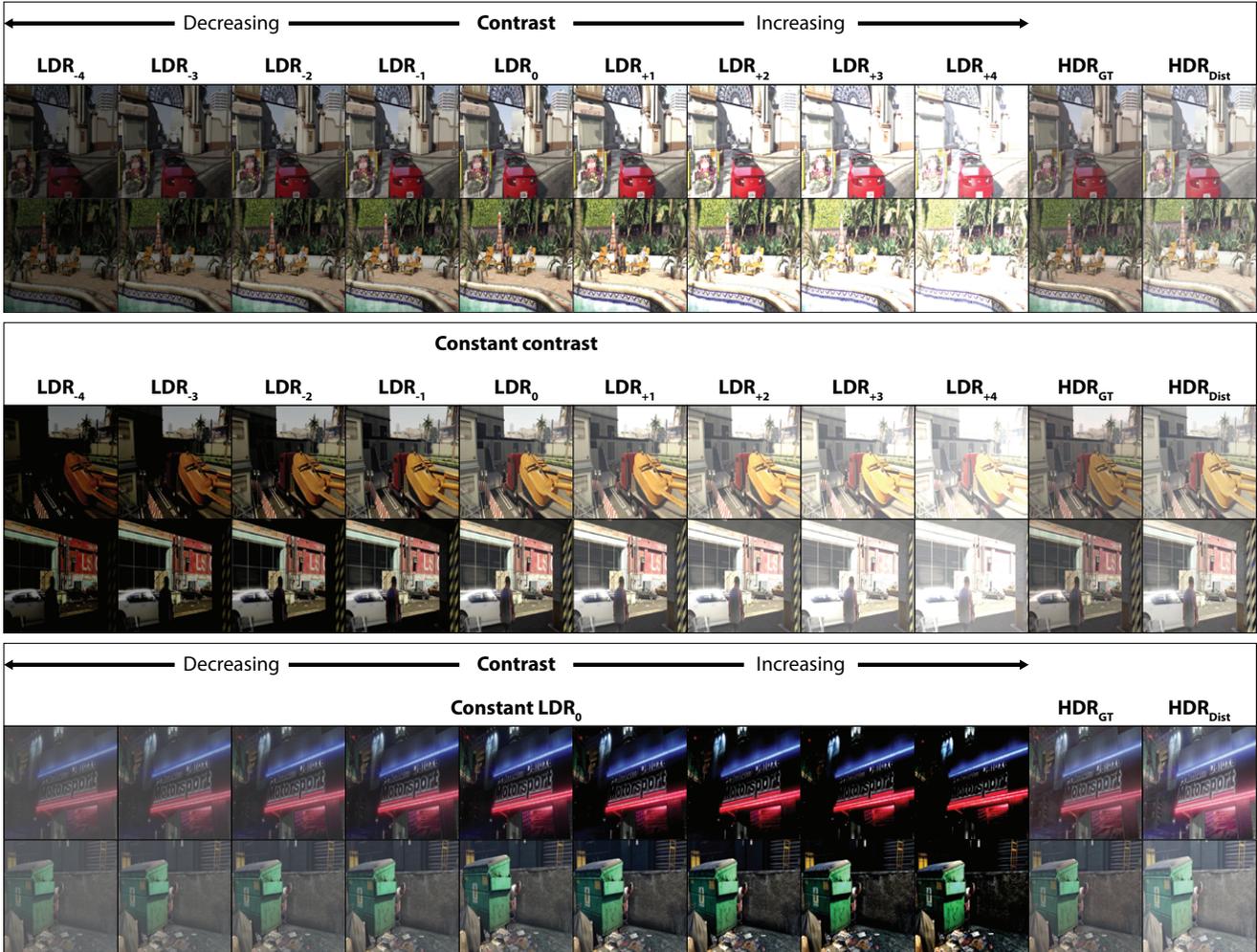


Figure 3. **GTA-HDR dataset image diversity.** Samples from dataset with multiple exposure values, contrast levels and their combinations.

ArtHDR-Net attempts to reconstruct perceptually realistic HDR images using features from multi-exposed LDR images. HistoHDR-Net uses histogram-equalized LDR along with original LDR images to facilitate the recovery of color, contrast, saturation, and hue in over/underexposed regions.

We used the official implementations and training strategies for these methods. Since SingleHDR [37] generates multi-exposed LDR images as output, we used the state-of-the-art tool Photomatix [22] to merge the LDR stack and obtain an HDR image. For single-exposed LDR image input methods, we consider all available LDR images in the datasets. For multi-exposed LDR image inputs: 1) For methods with three inputs, we consider an overexposed, a normally exposed, and an underexposed LDR image; and 2) For methods with two inputs, we consider an overexposed and an underexposed LDR image. For datasets with single-exposed LDR images, we generated the missing exposure versions. The diversity of the considered methods enables a thorough evaluation of the proposed GTA-HDR dataset.

Datasets. We considered most publicly available datasets for HDR image reconstruction in our experiments [15, 28, 30, 40, 50, 73]. We split the data into a train set consisting of the datasets proposed in [15, 28, 30, 50, 73] and test set including the dataset proposed in [40]. We considered different datasets for training and testing to demonstrate that GTA-HDR both enables significant improvements in the quality of the reconstructed HDR images and contributes towards better generalization capabilities of the considered state-of-the-art HDR image reconstruction methods. Identical protocol for training and testing was used in all experiments to ensure a fair comparison. All images were resized to 512×512 resolution. All HDR images displayed in this text have been tone-mapped using the method in [53].

4.2. Evaluation Metrics

We used three metrics to report the quantitative results. High Dynamic Range Visual Differences Predictor (HDR-VDP-2) [43] or Q-Score (Mean Opinion Score Index) is

Table 2. **Impact of the GTA-HDR dataset on the performance of the state-of-the-art in HDR image reconstruction.** *R*: Real data combines the datasets proposed in [28, 30, 50] and real images from the datasets proposed in [15, 73]; *R* \oplus *S*: Real and synthetic data combines all five datasets [15, 28, 30, 50, 73]; *GTA-HDR*: Proposed synthetic dataset; *E2E*: End-to-end training; *FT*: Finetuning of the original pre-trained models. *Note*: The performance of all methods is evaluated on a separate dataset proposed in [40].

Method	Configuration	Datasets	PSNR \uparrow	SSIM \uparrow	Q-score \uparrow
HDRCNN [13]	E2E	R	19.1	0.67	59.2
DrTMO [14]	E2E	R	19.2	0.68	60.3
FHDR [31]	E2E	R	24.4	0.80	65.1
SingleHDR [40]	E2E	R	29.1	0.81	66.2
HDR-GAN [49]	E2E	R	36.9	0.92	65.3
SingleHDR [37]	E2E	R	34.7	0.91	66.9
ArtHDR-Net [7]	E2E	R	35.1	0.91	67.2
HistoHDR-Net [6]	E2E	R	35.2	0.92	67.4
<hr/>					
HDRCNN [13]	E2E	R \oplus S	20.1 (+1.0)	0.69 (+0.02)	60.8 (+1.6)
DrTMO [14]	E2E	R \oplus S	20.3 (+1.1)	0.68 (+0.00)	61.5 (+1.2)
FHDR [31]	E2E	R \oplus S	26.7 (+2.3)	0.81 (+0.01)	65.3 (+0.2)
SingleHDR [40]	E2E	R \oplus S	30.4 (+1.3)	0.82 (+0.01)	66.1 (-0.1)
HDR-GAN [49]	E2E	R \oplus S	37.8 (+0.9)	0.94 (+0.02)	66.7 (+1.4)
SingleHDR [37]	E2E	R \oplus S	35.2 (+0.5)	0.92 (+0.01)	67.1 (+0.2)
ArtHDR-Net [7]	E2E	R \oplus S	35.3 (+0.2)	0.93 (+0.02)	67.4 (+0.2)
HistoHDR-Net [6]	E2E	R \oplus S	35.3 (+0.1)	0.94 (+0.02)	67.5 (+0.1)
<hr/>					
HDRCNN [13]	E2E / FT	GTA-HDR	22.4 (+2.3) / 22.1 (+2.0)	0.72 (+0.03) / 0.71 (+0.02)	61.3 (+0.5) / 61.4 (+0.6)
DrTMO [14]	E2E / FT	GTA-HDR	23.5 (+3.2) / 23.4 (+3.1)	0.71 (+0.03) / 0.71 (+0.03)	64.3 (+2.8) / 64.5 (+3.0)
FHDR [31]	E2E / FT	GTA-HDR	27.7 (+1.0) / 27.6 (+0.9)	0.84 (+0.03) / 0.84 (+0.03)	68.0 (+2.7) / 68.1 (+2.8)
SingleHDR [40]	E2E / FT	GTA-HDR	32.3 (+1.9) / 32.1 (+1.7)	0.86 (+0.04) / 0.85 (+0.03)	68.8 (+2.7) / 69.0 (+2.9)
HDR-GAN [49]	E2E / FT	GTA-HDR	38.7 (+0.9) / 38.5 (+0.7)	0.94 (+0.00) / 0.93 (-0.01)	69.5 (+2.8) / 69.7 (+3.0)
SingleHDR [37]	E2E / FT	GTA-HDR	41.2 (+6.0) / 41.5 (+6.3)	0.96 (+0.04) / 0.96 (+0.04)	70.2 (+3.1) / 70.0 (+2.9)
ArtHDR-Net [7]	E2E / FT	GTA-HDR	41.6 (+6.3) / 41.5 (+6.2)	0.97 (+0.04) / 0.97 (+0.04)	70.4 (+3.0) / 70.2 (+2.8)
HistoHDR-Net [6]	E2E / FT	GTA-HDR	41.7 (+6.4) / 41.5 (+6.2)	0.98 (+0.04) / 0.98 (+0.04)	70.5 (+3.0) / 70.4 (+2.9)
<hr/>					
HDRCNN [13]	E2E / FT	R \oplus S \oplus GTA-HDR	22.6 (+0.2) / 22.3 (+0.2)	0.70 (-0.02) / 0.69 (-0.02)	61.6 (+0.3) / 62.0 (+0.6)
DrTMO [14]	E2E / FT	R \oplus S \oplus GTA-HDR	23.6 (+0.1) / 23.5 (+0.1)	0.71 (+0.00) / 0.72 (+0.01)	64.6 (+0.3) / 64.8 (+0.3)
FHDR [31]	E2E / FT	R \oplus S \oplus GTA-HDR	27.9 (+0.2) / 27.4 (-0.2)	0.83 (-0.01) / 0.83 (-0.01)	67.5 (-0.5) / 68.1 (+0.0)
SingleHDR [40]	E2E / FT	R \oplus S \oplus GTA-HDR	32.5 (+0.2) / 31.6 (-0.5)	0.85 (-0.01) / 0.84 (-0.01)	68.7 (-0.1) / 68.8 (-0.2)
HDR-GAN [49]	E2E / FT	R \oplus S \oplus GTA-HDR	40.1 (+1.4) / 39.4 (+0.9)	0.95 (+0.01) / 0.97 (+0.04)	69.2 (-0.3) / 69.5 (-0.2)
SingleHDR [37]	E2E / FT	R \oplus S \oplus GTA-HDR	41.5 (+0.3) / 41.9 (+0.4)	0.97 (+0.01) / 0.98 (+0.02)	70.3 (+0.1) / 70.0 (+0.0)
ArtHDR-Net [7]	E2E / FT	R \oplus S \oplus GTA-HDR	41.6 (+0.0) / 42.1 (+0.6)	0.98 (+0.01) / 0.98 (+0.01)	71.2 (+0.8) / 70.9 (+0.7)
HistoHDR-Net [6]	E2E / FT	R \oplus S \oplus GTA-HDR	41.8 (+0.1) / 42.3 (+0.8)	0.99 (+0.01) / 0.99 (+0.01)	71.5 (+1.0) / 71.4 (+1.0)

used for evaluation based on the human visual system. For structural similarity, luminance, and contrast evaluation, the Structural Similarity Index Measure (SSIM) [67–69] is used. For pixel-to-pixel evaluation, the Peak Signal-to-Noise Ratio (PSNR) [18] is applied.

5. Results

5.1. HDR Reconstruction

This section presents the results from state-of-the-art HDR image reconstruction methods trained with different configurations of data, including real data, mixed data, and synthetic data. The results are based on two training strategies chosen to evaluate the contributions of the proposed GTA-HDR dataset: 1) End-to-end training (*i.e.*, the models are fully trained with different combinations of data) and 2) Finetuning (*i.e.*, only the final layers of the pre-trained original models are trained with different combinations of data). The results of this experiment are summarized in Table 2.

The results show a consistent improvement in PSNR, SSIM, and HDR-VDP-2 (Q-score) for all methods after including the GTA-HDR dataset in the training process. Furthermore, the results also demonstrate that all considered state-of-the-art methods trained with GTA-HDR data alone

achieved better performance than when they are trained with existing real and synthetic datasets (third sub-table). Moreover, Table 2 shows how the performance of the state-of-the-art methods improve consistently when we add more variations to the training data (first sub-table and second sub-table). Further improvements are achieved when mixing the existing datasets with GTA-HDR in both end-to-end and fine-tuning strategies. It is noteworthy that these consistent improvements are based on evaluation on a separate dataset, thus both demonstrating the contribution of the proposed GTA-HDR dataset towards increased quality of the reconstructed HDR images and better generalization capabilities of the state-of-the-art methods.

5.2. Scene Diversity

We studied the feature space coverage of different datasets as reported by common feature extraction backbones including MobileNet [24], InceptionV3 [61], ResNet50 [59], and VGG19 [60]. The results reveal a gap in the feature space, *i.e.*, certain regions are not covered by existing datasets. These regions are filled, to a certain extent, by the proposed GTA-HDR dataset. Fig. 4 illustrates the feature plots for different datasets and backbones.

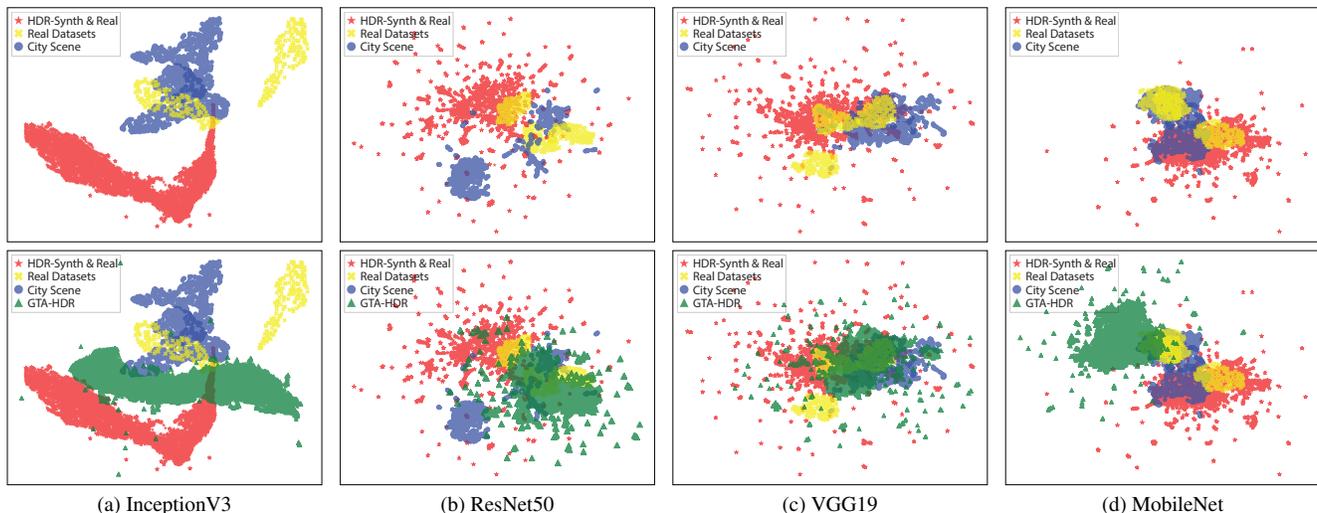


Figure 4. **Feature space covered by different HDR image reconstruction datasets.** We used UMAP [45] to visualize the features extracted from common pre-trained feature extraction backbones. *Real Datasets*: Real datasets proposed in [28,30,50]; *City Scene*: Mixed datasets proposed in [15,73]; *HDR-Synth & HDR-Real*: Mixed dataset proposed in [40]; *GTA-HDR*: Proposed synthetic dataset.

Features extracted with different backbones can be significantly different based on the underlying architecture, which affects the performance on downstream tasks *e.g.*, HDR image reconstruction. The main goal of backbones in LDR to HDR image conversion is the recognition of bright and dark regions and detection of the light source [59] to ensure that underexposed and overexposed regions are treated separately. The first column in Fig. 4 shows the output of InceptionV3 on existing datasets (top) and the GTA-HDR dataset included (bottom). There is a significant gap in the feature space between the HDR-Synth & Real dataset (red) and City Scene dataset (blue) as well as real datasets combined (yellow), which is filled, to some extent, by GTA-HDR (green). In the second column (ResNet50), GTA-HDR fills the gap in the lower right corner of the feature space. Here, we also observed a significant overlap of GTA-HDR with other datasets. In the third column (VGG19), GTA-HDR alone covers a significant area of the feature space which is covered by all other datasets combined. Finally, in the last column (MobileNet), GTA-HDR extends the feature space by covering a considerable area of the upper left corner along with some overlap with existing datasets.

To further investigate the contribution of the GTA-HDR dataset, we replaced the feature extraction block of one of the most recent state-of-the-art methods, SingleHDR [37] (which originally utilizes VGG19), with the described feature extraction backbones. Table 3 summarizes the quantitative results of this experiment in terms of PSNR and SSIM. When the model is trained with existing real and mixed datasets in an end-to-end fashion, the observed improvements are proportional to the size of the backbones (*i.e.*, the number of parameters). However, when using only GTA-HDR data for training, there is a significant im-

Table 3. **Performance of SingleHDR [37].** Different versions of the state-of-the-art method utilizing different feature extraction backbones, trained with and without GTA-HDR data in an end-to-end fashion. $R \oplus S$: Real and synthetic data combines all five datasets [15,28,30,50,73]; *GTA-HDR*: Proposed synthetic dataset. *Note*: Evaluated on a separate dataset proposed in [40].

Method	#Param	Datasets (training)	PSNR \uparrow	SSIM \uparrow
+MobileNet	13M	$R \oplus S$	32.3	0.89
+InceptionV3	24M	$R \oplus S$	32.8	0.89
+ResNet50	25.6M	$R \oplus S$	33.2	0.90
+VGG19	144M	$R \oplus S$	35.2	0.92
+MobileNet	13M	GTA-HDR	38.4 (+6.1)	0.94 (+0.05)
+InceptionV3	24M	GTA-HDR	38.8 (+6.0)	0.95 (+0.06)
+ResNet50	25.6M	GTA-HDR	39.5 (+6.3)	0.95 (+0.05)
+VGG19	144M	GTA-HDR	41.2 (+6.0)	0.96 (+0.04)
+MobileNet	13M	$R \oplus S \oplus$ GTA-HDR	38.6 (+0.2)	0.95 (+0.01)
+InceptionV3	24M	$R \oplus S \oplus$ GTA-HDR	39.5 (+0.7)	0.95 (+0.00)
+ResNet50	25.6M	$R \oplus S \oplus$ GTA-HDR	40.1 (+0.6)	0.96 (+0.01)
+VGG19	144M	$R \oplus S \oplus$ GTA-HDR	41.5 (+0.3)	0.97 (+0.01)

provement for all the backbones. It is worth noting that regardless of the size of the backbone feature extractor, there is an improvement in both PSNR and SSIM when using GTA-HDR. Interestingly, when including the GTA-HDR dataset, the performance of the considerably smaller (*e.g.*, MobileNet) backbones is better than the large ones (*e.g.*, VGG19) trained without using GTA-HDR.

5.3. Image Diversity

The existing datasets include either single-exposed or multi-exposed LDR images and the corresponding HDR images. GTA-HDR also introduces LDR images with multi-contrast levels for each HDR image. In this section, we report the results of an experiment that aims to establish the contribution of multi-exposed and multi-

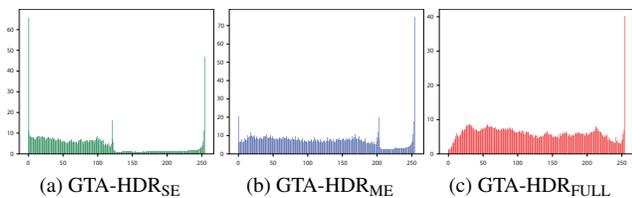


Figure 5. **Histograms of different versions of the GTA-HDR dataset.** (a) GTA-HDR_{SE}: Only single-exposed LDR images; (b) GTA-HDR_{ME}: Only multi-exposed LDR images; and (c) GTA-HDR_{FULL}: All LDR images.

contrast LDR images on the HDR image reconstruction performance. The experiment includes three versions of the GTA-HDR dataset: 1) GTA-HDR_{SE} consists of only single-exposed LDR images; 2) GTA-HDR_{ME} comprises only multi-exposed LDR images; and 3) GTA-HDR_{FULL} includes all LDR images (*i.e.*, multi-exposed and multi-contrast). Fig. 5 illustrates the histograms of the three versions of the dataset (we treat RGB channels as intensity values). The histogram for GTA-HDR_{SE} reveals that there is a miss-balance in the dataset, where most image pixel intensity values are on the left-hand side of the histogram (*i.e.*, underexposed pixels [0, 150]). The histogram for GTA-HDR_{ME} demonstrates that a portion of the miss-balance has been addressed (with a small gap in the interval [200, 250]). The histogram for GTA-HDR_{FULL} confirms that the intensity values are more evenly distributed.

Table 4 summarizes the quantitative results of this analysis. In this experiment, we consider the recent state-of-the-art method ArtHDR-Net [7]. The first three rows report the performance of ArtHDR-Net trained with each version of the GTA-HDR dataset and tested on reconstructing HDR images from underexposed LDR images. The performance of the model trained with GTA-HDR_{SE} dataset is surprisingly good in the case of underexposed LDR images, supporting the qualitative results from the left histogram in Fig. 5. The middle three rows report the performance of ArtHDR-Net trained with each version of the GTA-HDR dataset and tested on reconstructing HDR from overexposed LDR images. In this case, the performance steadily increases with the addition of exposure and contrast levels in the training set. A similar trend can be observed on the combined over/underexposed LDR images, reported in the bottom three rows. Images of real-life scenes have diverse exposure and contrast levels suggesting a data distribution similar to the proposed GTA-HDR_{FULL} dataset. In addition, multi-exposed and multi-contrast LDR images help mitigate model bias towards certain classes of images.

6. Conclusions

This paper describes GTA-HDR, a large-scale synthetic dataset and data collection pipeline to complement existing

Table 4. **Performance of ArtHDR-Net [7].** Assessment of the state-of-the-art method utilizing different training and testing datasets. *SE*: GTA-HDR without exposure and contrast variations; *ME*: GTA-HDR with different exposure levels; *FULL*: Proposed synthetic dataset; *U*: Underexposed LDR images; *O*: Overexposed LDR images; *N*: Normally exposed LDR images. *Note*: Evaluated on a separate dataset proposed in [40].

Datasets (training)	Datasets (testing)	PSNR \uparrow	SSIM \uparrow	Q-score \uparrow
GTA-HDR _{SE}	U	42.9	0.99	73.3
GTA-HDR _{ME}	U	40.7	0.97	71.2
GTA-HDR _{FULL}	U	39.1	0.95	69.4
GTA-HDR _{SE}	O	32.8	0.89	64.9
GTA-HDR _{ME}	O	33.9	0.92	65.7
GTA-HDR _{FULL}	O	34.7	0.93	66.4
GTA-HDR _{SE}	U + O + N	37.2	0.93	66.2
GTA-HDR _{ME}	U + O + N	39.7	0.95	68.5
GTA-HDR _{FULL}	U + O + N	41.6	0.97	70.4

real and synthetic datasets for HDR image reconstruction. The thorough experimental validation using existing real and synthetic datasets and state-of-the-art methods highlights the contribution of the proposed dataset, specifically to the quality of HDR image reconstruction and the recovery of image details with high fidelity (*cf.*, *Supplementary*). In the paper *Supplementary*, we further demonstrate the impact of GTA-HDR on the state-of-the-art in other computer vision tasks, including 3D human pose estimation, human body part segmentation, and holistic scene segmentation. The proposed dataset represents an important contribution that will enable the development of advanced techniques for visually accurate HDR image reconstruction. Preliminary discussion has been presented for developing no-reference quality assessment methods utilising the GTA-HDR dataset. This is possible direction for future research in addition to creating a video-based HDR reconstruction dataset.

Acknowledgements

This research is supported by the Global Excellence and Mobility Scholarship^{1,2}, Monash University. This research is supported, in part, by the E-Science fund under the project: *Innovative High Dynamic Range Imaging - From Information Hiding to Its Applications* (Grant No. 01-02-10-SF0327).

Copyright

Publisher’s policy allows the use of GTA-V game-play data provided it is for non-commercial purposes.

¹<https://www.monash.edu.my/research/support-and-scholarships/gems-scholarship>

²<https://www.thestar.com.my/news/education/2022/04/29/gems-a-rewarding-global-mobility-programme>

References

- [1] Matt Angus, Mohamed ElBalkini, Samin Khan, Ali Harakeh, Oles Andrienko, Cody Reading, Steven Waslander, and Krzysztof Czarnecki. Unlimited Road-scene Synthetic Annotation (URSA) Dataset. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 985–992. IEEE, 2018. **1**
- [2] Alessandro Artusi, Francesco Banterle, Fabio Carra, and Alejandro Moreno. Efficient Evaluation of Image Quality via Deep-Learning Approximation of Perceptual Metrics. *IEEE Transactions on Image Processing*, 29:1843–1855, 2019. **1**
- [3] Alessandro Artusi, Rafał K Mantiuk, Thomas Richter, Philippe Hanhart, Pavel Korshunov, Massimiliano Agostinelli, Arkady Ten, and Touradj Ebrahimi. Overview and evaluation of the JPEG XT HDR image compression standard. *Journal of Real-Time Image Processing*, 16:413–428, 2019. **1**
- [4] Francesco Banterle, Alessandro Artusi, Alejandro Moreo, and Fabio Carrara. Nor-Vdpnet: A No-Reference High Dynamic Range Quality Metric Trained On Hdr-Vdp 2. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 126–130. IEEE, 2020. **1**
- [5] Francesco Banterle, Alessandro Artusi, Alejandro Moreo, Fabio Carrara, and Paolo Cignoni. NoR-VDPNet++: Real-Time No-Reference Image Quality Metrics. *IEEE Access*, 11:34544–34553, 2023. **1**
- [6] Hrishav Bakul Barua, Ganesh Krishnasamy, KokSheik Wong, Abhinav Dhall, and Kalin Stefanov. HistoHDR-Net: Histogram Equalization for Single LDR to HDR Image Translation. *arXiv preprint arXiv:2402.06692*, 2024. **4, 6**
- [7] Hrishav Bakul Barua, Ganesh Krishnasamy, KokSheik Wong, Kalin Stefanov, and Abhinav Dhall. ArtHDR-Net: Perceptually Realistic and Accurate HDR Content Creation. In *2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 806–812. IEEE, 2023. **2, 4, 6, 8**
- [8] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a Deep Single Image Contrast Enhancer from Multi-Exposure Images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018. **1, 2, 3, 4**
- [9] Zhongang Cai, Mingyuan Zhang, Jiawei Ren, Chen Wei, Daxuan Ren, Jiatong Li, Zhengyu Lin, Haiyu Zhao, Shuai Yi, Lei Yang, et al. Playing for 3D Human Recovery. *arXiv preprint arXiv:2110.07588*, 2021. **1**
- [10] Gaofeng Cao, Fei Zhou, Kanglin Liu, Anjie Wang, and Leidong Fan. A Decoupled Kernel Prediction Network Guided by Soft Mask for Single Image HDR Reconstruction. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(2s):1–23, 2023. **2**
- [11] Dwip Dalal, Gautam Vashishtha, Prajwal Singh, and Shanmuganathan Raman. Single Image LDR to HDR Conversion Using Conditional Diffusion. In *2023 IEEE International Conference on Image Processing (ICIP)*, pages 3533–3537. IEEE, 2023. **1, 3**
- [12] Duc-Tien Dang-Nguyen, Cecilia Pasquini, Valentina Conotter, and Giulia Boato. RAISE: A Raw Images Dataset for Digital Image Forensics. In *Proceedings of the 6th ACM multimedia systems conference*, pages 219–224, 2015. **1, 3**
- [13] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K Mantiuk, and Jonas Unger. HDR image reconstruction from a single exposure using deep CNNs. *ACM transactions on graphics (TOG)*, 36(6):1–15, 2017. **1, 2, 3, 4, 6**
- [14] Yuki Endo, Yoshihiro Kanamori, and Jun Mitani. Deep reverse tone mapping. *ACM Trans. Graph.*, 36(6):177–1, 2017. **1, 2, 3, 4, 6**
- [15] Lalonde et al. The Laval HDR sky database. <http://hdrdb.com/>, 2016. [Online; accessed 3-July-2023]. **1, 2, 3, 5, 6, 7**
- [16] Matteo Fabbri, Guillem Brasó, Gianluca Maugeri, Orcun Cetintas, Riccardo Gasparini, Aljoša Ošep, Simone Calderara, Laura Leal-Taixé, and Rita Cucchiara. MOTSynth: How Can Synthetic Data Help Pedestrian Detection and Tracking? In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10849–10859, 2021. **1**
- [17] B-C Guo and C-H Lin. Single-Image HDR Reconstruction Based on Two-Stage GAN Structure. In *2023 IEEE International Conference on Image Processing (ICIP)*, pages 91–95. IEEE, 2023. **1**
- [18] Prateek Gupta, Priyanka Srivastava, Satyam Bhardwaj, and Vikrant Bhateja. A modified PSNR metric based on HVS for quality assessment of color images. In *2011 International Conference on Communication and Industrial Application*, pages 1–4. IEEE, 2011. **6**
- [19] Xueyu Han, Ishtiaq Rasool Khan, and Susanto Rahardja. High Dynamic Range Image Tone Mapping: Literature review and performance benchmark. *Digital Signal Processing*, page 104015, 2023. **2**
- [20] Param Hanji, Rafał Mantiuk, Gabriel Eilertsen, Saghi Hajisharif, and Jonas Unger. Comparison of single image HDR reconstruction methods—the caveats of quality assessment. In *ACM SIGGRAPH 2022 conference proceedings*, pages 1–8, 2022. **2, 3**
- [21] Param Hanji, Rafał Mantiuk, Gabriel Eilertsen, Saghi Hajisharif, and Jonas Unger. SI-HDR-dataset for comparison of single-image high dynamic range reconstruction methods. *University of Cambridge*, 2022. **2, 3**
- [22] HDRsoft. Photomatix. <https://www.hdrsoft.com/>. [Online; accessed 3-Nov-2023]. **5**
- [23] Gang He, Kepeng Xu, Li Xu, Chang Wu, Ming Sun, Xing Wen, and Yu-Wing Tai. SDRTV-to-HDRTV via Hierarchical Dynamic Context Feature Mapping. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 2890–2898, 2022. **1**
- [24] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. **6**
- [25] Yuan-Ting Hu, Jiahong Wang, Raymond A Yeh, and Alexander G Schwing. SAIL-VOS 3D: A Synthetic Dataset and Baselines for Object Detection and 3D Mesh Reconstruction from Video Data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1418–1428, 2021. **1**

- [26] Xin Huang, Qi Zhang, Ying Feng, Hongdong Li, Xuan Wang, and Qing Wang. HDR-NeRF: High Dynamic Range Neural Radiance Fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18398–18408, 2022. [1](#), [2](#), [3](#)
- [27] Yongqing Huo, Fan Yang, Le Dong, and Vincent Brost. Physiological inverse tone mapping based on retina response. *The Visual Computer*, 30:507–517, 2014. [1](#)
- [28] Hanbyol Jang, Kihun Bang, Jinseong Jang, and Dosik Hwang. Dynamic Range Expansion Using Cumulative Histogram Learning for High Dynamic Range Image Generation. *IEEE Access*, 8:38554–38567, 2020. [1](#), [2](#), [3](#), [5](#), [6](#), [7](#)
- [29] Anoop K Johnson. High Dynamic Range Imaging - A Review. *Int. J. Image Process.(IJIP)*, 9:198, 2015. [1](#)
- [30] Nima Khademi Kalantari, Ravi Ramamoorthi, et al. Deep High Dynamic Range Imaging of Dynamic Scenes. *ACM Trans. Graph.*, 36(4):144–1, 2017. [1](#), [2](#), [3](#), [5](#), [6](#), [7](#)
- [31] Zeeshan Khan, Mukul Khanna, and Shanmuganathan Raman. FHDR: HDR Image Reconstruction from a Single LDR Image using Feedback Network. In *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1–5. IEEE, 2019. [2](#), [4](#), [6](#)
- [32] Soo Ye Kim, Jihyong Oh, and Munchurl Kim. Deep SR-ITM: Joint Learning of Super-Resolution and Inverse Tone-Mapping for 4K UHD HDR Applications. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3116–3125, 2019. [1](#), [3](#)
- [33] Yuma Kinoshita and Hitoshi Kiya. Scene segmentation-based luminance adjustment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 28(8):4101–4116, 2019. [1](#)
- [34] Giorgos Kordopatis-Zilos, Symeon Papadopoulos, Ioannis Patras, and Ioannis Kompatsiaris. ViSiL: Fine-grained Spatio-Temporal Video Similarity Learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019. [3](#)
- [35] Rafael P Kovaleski and Manuel M Oliveira. High-Quality Reverse Tone Mapping for a Wide Range of Exposures. In *2014 27th SIBGRAPI Conference on Graphics, Patterns and Images*, pages 49–56. IEEE, 2014. [1](#)
- [36] Philipp Krähenbühl. Free Supervision From Video Games. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2955–2964, 2018. [1](#)
- [37] Phuoc-Hieu Le, Quynh Le, Rang Nguyen, and Binh-Son Hua. Single-Image HDR Reconstruction by Multi-Exposure Generation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 4063–4072, 2023. [2](#), [4](#), [5](#), [6](#), [7](#)
- [38] Siyeong Lee, Gwon Hwan An, and Suk-Ju Kang. Deep Chain HDRI: Reconstructing a High Dynamic Range Image from a Single Low Dynamic Range Image. *IEEE Access*, 6:49913–49924, 2018. [1](#), [2](#), [3](#)
- [39] Jinghui Li and Peiyu Fang. HDRNET: Single-Image-based HDR Reconstruction Using Channel Attention CNN. In *Proceedings of the 2019 4th International Conference on Multimedia Systems and Signal Processing*, pages 119–124, 2019. [2](#)
- [40] 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. [1](#), [2](#), [3](#), [4](#), [5](#), [6](#), [7](#), [8](#)
- [41] Gonzalo Luzardo, Jan Aelterman, Hiep Luong, Wilfried Philips, Daniel Ochoa, and Sven Rousseaux. Fully-Automatic Inverse Tone Mapping Preserving the Content Creator’s Artistic Intentions. In *2018 Picture Coding Symposium (PCS)*, pages 199–203. IEEE, 2018. [1](#)
- [42] Gonzalo Luzardo, Jan Aelterman, Hiep Luong, Sven Rousseaux, Daniel Ochoa, and Wilfried Philips. Fully-automatic inverse tone mapping algorithm based on dynamic mid-level tone mapping. *APSIPA Transactions on Signal and Information Processing*, 9:e7, 2020. [2](#)
- [43] Rafał Mantiuk, Kil Joong Kim, Allan G Rempel, and Wolfgang Heidrich. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on graphics (TOG)*, 30(4):1–14, 2011. [5](#)
- [44] Belen Masia, Ana Serrano, and Diego Gutierrez. Dynamic range expansion based on image statistics. *Multimedia Tools and Applications*, 76:631–648, 2017. [1](#)
- [45] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018. [7](#)
- [46] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *Communications of the ACM*, 65(1):99–106, 2021. [2](#)
- [47] Hiromi Nemoto, Pavel Korshunov, Philippe Hanhart, and Touradj Ebrahimi. Visual attention in LDR and HDR images. In *9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, number CONF, 2015. [1](#), [2](#), [3](#)
- [48] Hue Nguyen, Diep Tran, Khoi Nguyen, and Rang Nguyen. PSENet: Progressive Self-Enhancement Network for Unsupervised Extreme-Light Image Enhancement. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1756–1765, 2023. [1](#), [4](#)
- [49] Yuzhen Niu, Jianbin Wu, Wenxi Liu, Wenzhong Guo, and Rynson WH Lau. HDR-GAN: HDR Image Reconstruction from Multi-Exposed LDR Images with Large Motions. *IEEE Transactions on Image Processing*, 30:3885–3896, 2021. [4](#), [6](#)
- [50] K Ram Prabhakar, Rajat Arora, Adhitya Swaminathan, Kunal Pratap Singh, and R Venkatesh Babu. A Fast, Scalable, and Reliable Deghosting Method for Extreme Exposure Fusion. In *2019 IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2019. [1](#), [2](#), [3](#), [5](#), [6](#), [7](#)
- [51] Prarabdh Raipurkar, Rohil Pal, and Shanmuganathan Raman. HDR-cGAN: Single LDR to HDR Image Translation using Conditional GAN. In *Proceedings of the Twelfth Indian Conference on Computer Vision, Graphics and Image Processing*, pages 1–9, 2021. [3](#)

- [52] CD Project RED. The Witcher 3: Wild Hunt. <https://www.thewitcher.com/in/en/witcher3>, 2016. [Online; accessed 3-Dec-2023]. 1
- [53] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. Photographic tone reproduction for digital images. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 661–670. 2023. 5
- [54] Stephan R Richter, Zeeshan Hayder, and Vladlen Koltun. Playing for Benchmarks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2213–2222, 2017. 1
- [55] Stephan R Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for Data: Ground Truth from Computer Games. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14*, pages 102–118. Springer, 2016. 1
- [56] Marcel Santana Santos, Tsang Ing Ren, and Nima Khademi Kalantari. Single Image HDR Reconstruction Using a CNN with Masked Features and Perceptual Loss. *ACM Transactions on Graphics (TOG)*, 39(4):80–1, 2020. 2
- [57] Pinar Satilmis and Thomas Bashford-Rogers. Deep Dynamic Cloud Lighting. *arXiv preprint arXiv:2304.09317*, 2023. 1
- [58] Pradeep Sen, Nima Khademi Kalantari, Maziar Yaesoubi, Soheil Darabi, Dan B Goldman, and Eli Shechtman. Robust Patch-Based HDR Reconstruction of Dynamic Scenes. *ACM Trans. Graph.*, 31(6):203–1, 2012. 1, 3
- [59] Seungjun Shin, Kyeongbo Kong, and Woo-Jin Song. CNN-based LDR-to-HDR conversion system. In *2018 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–2. IEEE, 2018. 3, 6, 7
- [60] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 6
- [61] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015. 6
- [62] Gaurav Tiwari and Pushpi Rani. A Review On High-Dynamic-Range Imaging With Its Technique. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 8(9):93–100, 2015. 1
- [63] Okan Tarhan Tursun, Ahmet Oğuz Akyüz, Aykut Erdem, and Erkut Erdem. The State of the Art in HDR Deghosting: A Survey and Evaluation. In *Computer Graphics Forum*, volume 34, pages 683–707. Wiley Online Library, 2015. 1
- [64] Okan Tarhan Tursun, Ahmet Oğuz Akyüz, Aykut Erdem, and Erkut Erdem. An Objective Deghosting Quality Metric for HDR Images. In *Computer Graphics Forum*, volume 35, pages 139–152. Wiley Online Library, 2016. 1, 3
- [65] Hu Wang, Mao Ye, Xiatian Zhu, Shuai Li, Ce Zhu, and Xue Li. KUNet: Imaging Knowledge-Inspired Single HDR Image Reconstruction. In *The 31st International Joint Conference On Artificial Intelligence (IJCAI/ECAI 22)*, 2022. 1
- [66] Lin Wang and Kuk-Jin Yoon. Deep Learning for HDR Imaging: State-of-the-Art and Future Trends. *IEEE transactions on pattern analysis and machine intelligence*, 44(12):8874–8895, 2021. 1, 2
- [67] Zhou Wang and Alan C Bovik. Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. *IEEE signal processing magazine*, 26(1):98–117, 2009. 6
- [68] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 6
- [69] Zhou Wang, Ligang Lu, and Alan C Bovik. Video quality assessment based on structural distortion measurement. *Signal processing: Image communication*, 19(2):121–132, 2004. 6
- [70] Xuesong Wu, Hong Zhang, Xiaoping Hu, Moein Shakeri, Chen Fan, and Juiwen Ting. HDR reconstruction based on the polarization camera. *IEEE Robotics and Automation Letters*, 5(4):5113–5119, 2020. 1
- [71] Huazhe Xu, Yang Gao, Fisher Yu, and Trevor Darrell. End-to-end Learning of Driving Models from Large-scale Video Datasets. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2174–2182, 2017. 1
- [72] Zhitao Yang, Zhongang Cai, Haiyi Mei, Shuai Liu, Zhaoxi Chen, Weiye Xiao, Yukun Wei, Zhongfei Qing, Chen Wei, Bo Dai, et al. SynBody: Synthetic Dataset with Layered Human Models for 3D Human Perception and Modeling. *arXiv preprint arXiv:2303.17368*, 2023. 1
- [73] Jinsong Zhang and Jean-François Lalonde. Learning High Dynamic Range from Outdoor Panoramas. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4519–4528, 2017. 1, 2, 3, 5, 6, 7
- [74] Lin Zhang, Anqi Zhu, Shiyu Zhao, and Yicong Zhou. Simulation of Atmospheric Visibility Impairment. *IEEE Transactions on Image Processing*, 30:8713–8726, 2021. 1, 3