Learning Continuous Exposure Value Representations for Single-Image HDR Reconstruction

Su-Kai Chen[1,2] Hung-Lin Yen[1] Yu-Lun Liu[1] Min-Hung Chen[3]
Hou-Ning Hu[2] Wen-Hsiao Peng[1] Yen-Yu Lin[1]
[1]National Yang Ming Chiao Tung University [2]MediaTek Inc. [3]NVIDIA
https://skchen1993.github.io/CEVR_web/

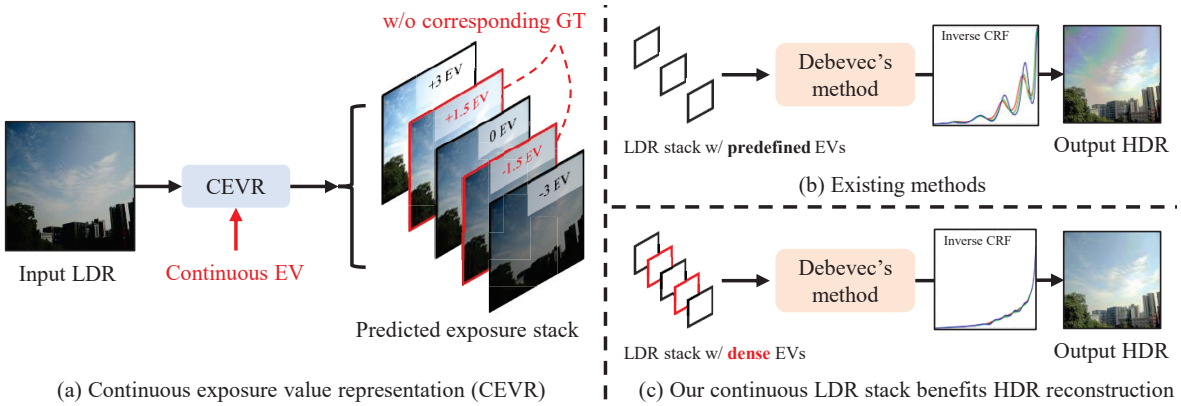


Figure 1: **Single-image HDR reconstruction from continuous LDR stack.** (a) Continuous Exposure Value Representation (CEVR) generates LDR images with continuous exposure values (EV) without corresponding ground truth during training. (b) Existing methods build LDR stacks only with EVs covered by training data, which brings less visible details for Debevec's method [11] to estimate an accurate inverse camera response function (CRF), resulting in artifacts on HDR results. (c) Our CEVR model enriches the LDR stack by including additional LDR images with continuous and dense EVs (**red** frames), allowing Debevec's method to predict a more precise inverse CRF and reconstruct more visually pleasing HDR images.

Abstract

Deep learning is commonly used to reconstruct HDR images from LDR images. LDR stack-based methods are used for single-image HDR reconstruction, generating an HDR image from a deep learning-generated LDR stack. However, current methods generate the stack with predetermined exposure values (EVs), which may limit the quality of HDR reconstruction. To address this, we propose the continuous exposure value representation (CEVR), which uses an implicit function to generate LDR images with arbitrary EVs, including those unseen during training. Our approach generates a continuous stack with more images containing diverse EVs, significantly improving HDR reconstruction. We use a cycle training strategy to supervise the model in generating continuous EV LDR images without corresponding ground truths. Our CEVR model outperforms existing methods, as demonstrated by experimental results.

1. Introduction

High dynamic range (HDR) images can capture detailed appearances in regions with extreme lighting conditions, like sun and shadow. As conventional cameras only capture a limited dynamic range in real-world scenes, one approach to address this issue is to blend multiple LDR images with different exposures into a single HDR image. However, this method is limited to static scenes and may result in ghosting or blurring artifacts in dynamic scenes. Additionally, this method is not applicable when multiple images of the same scene are unavailable, such as an image on the internet.

Another branch of methods, e.g., [12, 13, 23, 24, 26, 32, 43], takes a single LDR image as input to generate the HDR counterpart without suffering from misalignment, which is referred to as *single-image HDR reconstruction*. These approaches, e.g., [23,24], are trained on particular datasets and build an LDR stack with a single LDR image to generate an HDR image using Debevec's method [11].

EV stack of *real images*	PSNR RH's TMO		PSNR KK's TMO		HDR-VDP-2	
	m	σ	m	σ	m	σ
[-2, 0, 2]	24.96	3.18	23.52	2.75	45.12	5.74
[-2, -1.3, 0, 1.3, 2]	25.15	2.98	23.71	3.12	45.26	5.12
[-2, -1.3, -0.7, 0, 0.7, 1.3, 2]	25.32	3.01	23.92	3.04	45.51	4.86

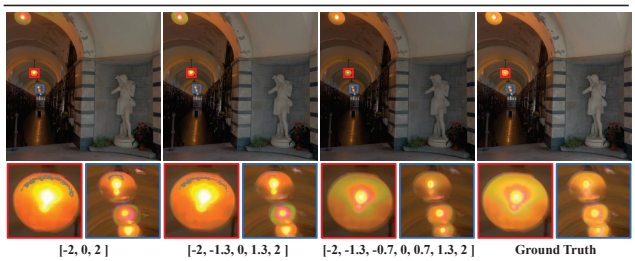


Figure 2: **Motivation.** We observe that an LDR stack with dense EVs improves HDR reconstruction even with the same exposure range (from -2EV to +2EV). A list "[-2,0,2]" means the stack contains three LDR images with -2, 0, and 2 EVs. An example of visual comparison is given.

Using more LDR images with richer EVs improves HDR image quality, as demonstrated in Fig. 2 with different EV stack settings using Debevec's method on the *real* LDR images of the HDREye dataset [36]. We compare tone-mapping operators RH [39] and KK [20] and use HDR-VDP-2 to evaluate HDR quality. However, accessible datasets have predefined and quantized EVs and may not cover optimal values for HDR reconstruction, causing information loss.

Previous studies [8, 35, 46] show the effectiveness of implicit neural representations in modeling continuous relationships, motivating our research. Inspired by the observation in Fig. 2, we address the issue of predefined, quantized EVs by leveraging an implicit neural function to model relationships between image appearance and continuous EVs. It turns out our method can generate LDR images with arbitrary EVs even if the corresponding ground truth is unavailable. More importantly, LDR stacks enriched by images with these continuous and dense EVs can reconstruct HDR images of better quality.

Specifically, the proposed approach, *continuous exposure value representation* (CEVR), exploits an implicit neural function to generate LDR images with continuous exposure values, as shown in Fig. 1(a). Based on the flexibility of our CEVR model, we further develop two strategies, cycle training and continuous stack, to improve the quality of the LDR stack and the final HDR result.

Cycle training utilizes CEVR to generate continuous EV images without relying on direct supervision from corresponding ground truths. We train the model using two continuous EVs that sum up to a predefined EV, with the proportion of these two continuous EVs randomly sampled. This strategy enforces the cycle consistency constraint, improving the model's ability to synthesize images with varying EVs and enhancing the quality of the LDR stack. We then use the enriched LDR stack containing seen and unseen EVs in training data for Debevec's method to produce more accurate inverse camera response functions (CRFs) and visually appealing tone-mapped images (Fig. 1(c)) compared to previous methods [23, 24] (Fig. 1(b)).

Extensive evaluations demonstrate the effectiveness of our proposed continuous stack and cycle training on the VDS [23] and HDREye [36] datasets. Both quantitative and qualitative evaluations show that CEVR significantly outperforms existing methods. The following summarizes our three primary contributions:

- We propose the CEVR approach, which can generate LDR images with continuous exposure values by modeling relationships between image appearances and exposure values.
- With the flexibility of the CEVR model, we design a training strategy, cycle training, to explore continuous EV information and enhance the quality of the estimated LDR stack.
- We propose the continuous stack, which consists of LDR images with continuous and dense exposure values and can improve the quality of final HDR images.

2. Related Work

Multi-image HDR reconstruction. Modern cameras typically have limited dynamic ranges and cannot well capture all visible details of a scene with a wide range of illumination. To address this issue, one practical solution is to take multiple LDR images at different exposure levels and blend them into an HDR image. To this end, conventional methods such as [11, 30] are developed to estimate the CRF [15], upon which multiple LDR images are converted into the radiance field of the scene and transformed into an HDR image. Recent methods, e.g., [17, 51], use CNNs for directly fusing LDR images and reconstructing their HDR counterpart. However, both conventional and CNN-based methods require multiple differently exposed images of a static scene. Furthermore, for working on dynamic scenes, additional mechanisms are needed to alleviate the misalignment problem and avoid blurring or ghosting artifacts [18, 29, 47]. However, misalignment itself is a complicated issue to resolve.

Single-image HDR reconstruction. This task aims to reconstruct the HDR image using just one LDR input, also called inverse tone mapping [3–6], and can bypass the misalignment problem. Existing methods need to enlarge the dynamic range [1, 33, 40, 44] and restore the lost details. Generation techniques [14, 25, 55, 56, 58] for image synthesis are essential to methods of this category. Due to the

superior mapping power of CNNs [16, 45] and GAN [14], deep neural networks are widely adopted for HDR reconstruction.

One branch of research efforts [12,32,43,53,57] focuses on learning the mapping from the input LDR image to the HDR image. For example, Marnerides et al. [32] use CNNs to generate the HDR image based on an LDR input. To further improve the performance, Santos et al. [43] filter out the saturated regions in the LDR input and pretrain the deep network for an inpainting task. However, learning the LDR-to-HDR mapping is ill-posed since different LDR images can be mapped to the same HDR image [13].

Another branch of methods, e.g., [13, 19, 23, 24], aims to synthesize a stack of differently exposed LDR counterparts given an LDR image as input. Then, the conventional multi-image methods can be applied to the synthesized LDR stack to complete HDR reconstruction. For example, Endo et al. [13] use 3D convolutions, with exposure variation being one dimension, to learn the relationship between the LDR input and its counterparts with different exposure values. Their approach can generate the LDR stack directly. The LDR stack can be synthesized in a recursive manner [19,23,24]. For example, Lee et al. [24] use GAN to generate an image with relative exposure value change. The LDR stack is constructed by recursively using their model.

Nevertheless, existing stack-based methods can only generate LDR images with predefined exposure values present in the training data. Inspired by the fact that the real-world captured images can have any EV value depending on different shutter settings instead of predefined ones, we present a method that can synthesize LDR images with continuous exposure values that are even unseen in the training data. Our method can generate an enriched and denser stack with which significantly better HDR results are achieved.

Implicit neural representations. An implicit function space is a shared function space that contains the neural representation of different objects or images learned by a shared implicit function. It is commonly a latent space where a latent code is mapped to an image using an encoder-decoder structure [9,34,41,42,52]. This approach has been widely used in image super-resolution [8, 22], 3D shape, surface modeling [2,9,46], and view synthesis of 3D structures [35,37]. Methods using implicit functions have shown that the learned latent space can be continuous [8,10,35,38], allowing for exploring continuous relationships of exposure differences between images.

More and more radiance field reconstruction research aims to generalize the trained model across scenes unseen in training data. The methods in [7, 48, 54] propose advanced model architectures and training strategies, making the learned implicit function space achieve the generalization on unseen views. Our method, similar to [7,48,54], can generalize well to all images without fine-tuning.

3. Approach

In this section, we present our proposed Continuous Exposure Value Representation (CEVR) which generates LDR images with continuous EV. We provide an overview of our method in Section 3.1, followed by the architectural design in Section 3.2, which includes the implicit module and intensity transformation. Additionally, we propose two strategies, cycle training, and continuous stack, to further enhance the flexibility of CEVR, which are discussed in detail in Section 3.3 and 3.4, respectively.

3.1. Overview

Based on the observation in Fig. 2, we propose the CEVR model to generate an enriched and denser LDR stack for high-quality HDR reconstruction. Our model, shown in Fig. 3, utilizes a hierarchical U-Net structure (Fig. 3(a)) and incorporates the implicit neural representation into the design to predict LDR images with continuous EVs (Fig. 3(c)). To maintain accurate color and image structure while adjusting brightness, we introduce intensity transformation (Fig. 3(d)), which generates an adjustment map from each scale of the feature map. As the ground-truth LDR images with unseen exposure values are lacking, we train the model using unsupervised cycle training (Fig. 4), enabling our method to learn images with varying EVs and enhance the quality of the predicted LDR stack.

3.2. Continuous Exposure Value Representation

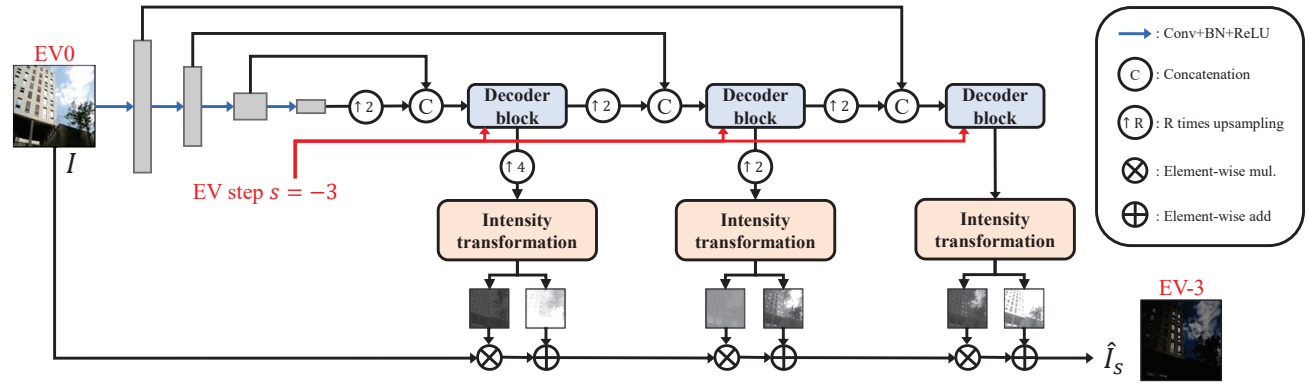
We show our CEVR model in Fig. 3(a). Our CEVR model employs the hierarchical U-Net structure, where the encoder is a pre-trained VGG-Net and the decoder is a cascade of decoder blocks (Fig. 3(b)), each of which comprises an implicit module that compiles the feature map with an input EV step s. Each decoder block is followed by an intensity transformation to adjust the intensity of input image at that scale. Specifically, the CEVR model F takes an LDR image I and the specified EV step s as input and generates another LDR image $\hat{I}_s$, a counterpart of I with the relative exposure value change s, via

$$\hat{I}_s = F(I, s). \tag{1}$$

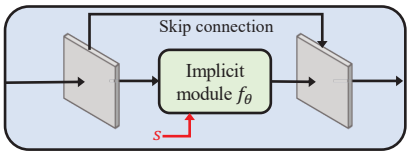
Take the widely used VDS dataset [23] as an example. An LDR image with EV0 in this dataset can serve as I. An LDR stack can be generated by applying our CEVR F to I and every EV step in $\{s \in \mathbb{Z} | -3 \leq s \leq 3\}$.

Implicit module. To synthesize an LDR image conditioned on a continuous EV step s even unseen in the training data, each decoder block in Fig. 3(b) has an associated, learnable implicit module f_θ, which is built by MLPs and shown in Fig. 3(c). The implicit module f_θ parameterized by θ takes the form:

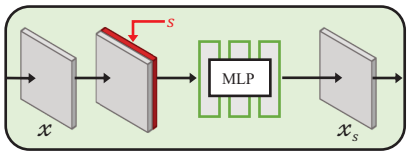
$$x_s(p, q) = f_\theta([x(p, q), s]), \tag{2}$$



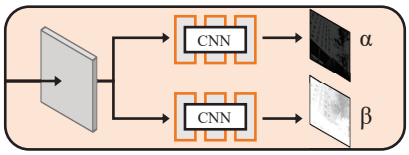
(a) CEVR network



(b) Decoder block



(c) Implicit module



(d) Intensity transformation

Figure 3: **Proposed network architecture.** (a) The proposed CEVR model takes an image I and an EV step s as input, and produces an LDR image $\hat{I}_s$ with a relative exposure value change s. It adopts the U-Net structure, where the encoder is a pre-trained VGG-Net, and the decoder is a cascade of decoder blocks. (b) Each decoder block comprises an implicit module to enable continuous EV representation learning, as shown in (c). (d) Following each decoder block, an intensity transformation module is learned to produce the α and β maps for image brightness transformation.

where $x \in \mathbb{R}^{H\times W\times C}$ is the input feature map, $x(p,q) \in \mathbb{R}^C$ is the feature vector at location (p,q), and $[x(p,q), s] \in \mathbb{R}^{C+1}$ refers to the concatenation of $x(p,q)$ and s. The output feature map x_s is generated by repeatedly applying the implicit module f_θ to all $H \times W$ locations of x with the desired relative exposure value change s.

Intensity transformation. In Fig. 3(a), our CEVR leverages U-Net to perform multi-scale synthesis to generate a better LDR image with a different EV. The input and output images, I and $\hat{I}_s$, cover the same scene under different exposures. Thus, their content should not undergo significant changes. To preserve the image structure and allow the model to focus on the brightness changes for detail reconstruction at each scale, the proposed intensity transformation module in Fig. 3(d) takes the resized feature map from the decoder block as input and produces the α and β maps. As shown in Fig. 3(a), the α and β maps carry out affine brightness transformation at each scale. The output $\hat{I}_s$ is synthesized through multi-scale transformations.

Reconstruction Loss. Suppose that we are given a training set of N images $\{I_n\}_{n=1}^N$ with a set of M EV steps $\{s_m\}_{m=1}^M$. For each training image I_n, its ground-truth LDR stack $\{I_n^*(s_m)\}_{m=1}^M$ is provided, where $I_n^*(s_m)$ is the counterpart of I_n with the relative exposure value change s_m. We train the CEVR model F in Eq. (1) by minimizing the L_1 reconstruction loss:

$$\mathcal{L}_{\text{rec}} = \sum_{n=1}^{N}\sum_{m=1}^{M} \|I_n^*(s_m) - F(I_n, s_m)\|_1. \tag{3}$$

3.3. Cycle Training Strategy

Existing training datasets, such as the VDS dataset [23], provide the ground truth for a sparse set of predefined EV steps, e.g., $[-3, -2, ..., 3]$. Inspired by the success of cycle consistency training in video frame interpolation [27] and to make our CEVR work well for synthesizing images with arbitrary EVs, we introduce the cycle training strategy to train the model with *continuous* EV steps without the corresponding ground-truth images. For each training image I_n and each EV step s_m covered by the training set, the cycle training strategy shown in Fig. 4, derives the CEVR model with two branches. The first branch takes I_n and s_m as input. Since the ground-truth image $I_n^*(s_m)$ is available, the reconstruction loss $\mathcal{L}_{\text{rec}}$ is used to supervise this branch.

The second branch implements a two-step process. We randomly sample a real value $a \in [0, 1]$ for each image I_n at each training iteration, and decompose the EV step s_m into two sub-steps: $u = as_m$ and $v = (1-a)s_m$, with $u + v = s_m$. Our CEVR model is applied *twice* with the two EV sub-steps, respectively. Although the ground truth for

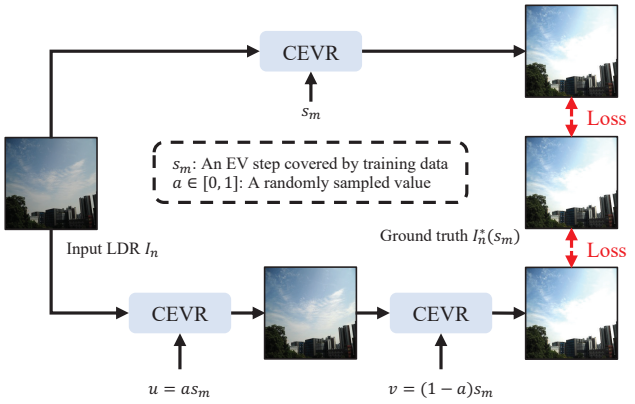


Figure 4: **Cycle training.** We derive the CEVR model in an unsupervised cycle training strategy without using the corresponding ground truth. In this way, our model exploits the cycle consistency constraint and learns more continuous information by varying the EV sub-step u.

the randomly sampled sub-step u is unavailable, we expect that the output of taking the two sub-steps should be similar to the ground truth $I_n^*(s_m)$ because of $u+v=s_m$. Thereby, we enforce the proposed cycle loss:

$$\mathcal{L}_{\text{cyc}} = \sum_{n=1}^{N} \sum_{m=1}^{M} \|I_n^*(s_m) - F(F(I_n, u), v)\|_1. \quad (4)$$

The sub-step u in Eq. (4) is randomly sampled for each training image with each covered EV step at each training iteration. It is used to simulate arbitrary EV step input to our CEVR model. To compensate for the lack of the ground truth of the intermediate output $F(I_n, u)$, the cycle loss $\mathcal{L}_{\text{cyc}}$ in Eq. (4) offers indirect supervision, ensuring the continuity of our CEVR model with continuous EV steps.

The objective function used to derive the proposed CEVR is defined by

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \lambda \mathcal{L}_{\text{cyc}}, \quad (5)$$

where we empirically set λ to 0.1 in our experiments.

3.4. Continuous Stack

In the inference phase, with the implicit module and the cycle training strategy, our CEVR model can generate high-quality LDR images with continuous EVs. The LDR stack containing more LDR images with various EVs can help Debevec's method [11] estimate a more accurate inverse CRF, as shown in Fig. 1(c), and improve the HDR image reconstruction, as shown in Fig. 2. Inspired by this observation, we proposed the continuous stack, which predicts additional LDR images with continuous EVs from our CEVR model. The predicted continuous and dense LDR stack further benefits the stack fusion process and enhances the final HDR quality, as shown in Fig. 1(c).

Table 1: **Quantitative comparison of the predicted LDR stacks on the VDS dataset [23].** CEVR outperforms existing approaches in estimating LDR stacks for all EVs. With cycle training, our method can generate high-quality LDR images even with large EV changes.

EV	Method	PSNR		SSIM		MS-SSIM	
		m	σ	m	σ	m	σ
+3	Deep chain HDRI [23]	28.18	2.77	0.953	0.065	0.983	0.015
	Deep recursive HDRI [24]	28.97	2.92	0.944	0.044	0.981	0.014
	CEVR (Ours)	**34.34**	3.46	**0.973**	0.021	**0.989**	0.007
+2	Deep chain HDRI [23]	29.65	3.06	0.959	0.065	0.986	0.016
	Deep recursive HDRI [24]	29.43	2.85	0.952	0.039	0.986	0.010
	CEVR (Ours)	**35.30**	3.08	**0.981**	0.016	**0.993**	0.004
+1	Deep chain HDRI [23]	31.90	3.43	0.969	0.039	0.992	0.008
	Deep recursive HDRI [24]	32.02	2.85	0.969	0.026	0.992	0.006
	CEVR (Ours)	**37.64**	2.96	**0.989**	0.009	**0.996**	0.004
-1	Deep chain HDRI [23]	29.01	3.83	0.935	0.056	0.980	0.017
	Deep recursive HDRI [24]	31.22	3.69	0.951	0.031	0.986	0.090
	CEVR (Ours)	**34.62**	3.47	**0.980**	0.011	**0.992**	0.005
-2	Deep chain HDRI [23]	26.72	4.54	0.952	0.029	0.974	0.021
	Deep recursive HDRI [24]	31.08	3.07	0.948	0.041	0.986	0.014
	CEVR (Ours)	**33.89**	4.34	**0.978**	0.017	**0.988**	0.010
-3	Deep chain HDRI [23]	24.33	4.57	0.919	0.036	0.948	0.037
	Deep recursive HDRI [24]	29.15	4.75	0.910	0.061	0.966	0.025
	CEVR (Ours)	**30.58**	5.32	**0.954**	0.046	**0.972**	0.032

4. Experiments

4.1. Experimental Setup

Datasets. We train our model using the training set of the VDS dataset [23], which contains image stacks of 48 scenes. The testing sets of the VDS and HDREye datasets [36], which contain 48 and 42 scenes respectively, serve as the testing sets for evaluations. The auto-bracketing feature of the camera produces seven photos with predefined exposure values for each scene in the VDS dataset (EV-3 to EV+3). We follow the common evaluation protocol [23, 24] and select the image with the zero exposure value, which is expected to have the most evenly distributed histogram, as the input to the model.

Training details. For training, we consider each training scene n from the VDS dataset [23] and take the corresponding EV0 LDR image as input I_n. We also take each EV step $s_m \in \{-3, -2, -1, 0, 1, 2, 3\}$ into account. We feed I_n and s_m to the CEVR model and estimate the LDR image with EV s_m for model training. Since the inverse CRF is usually asymmetrical, we train two different models with the same architecture to handle the increasing and decreasing exposure changes, respectively. For upsampling, we use bicubic upsampling, followed by a 3×3 2D convolution with stride 1 and padding 1. The model is trained for 1,250 epochs with Adam optimizer [21] and cosine annealing warmup with restarts as the scheduler. We use random rotation and flip to augment the data.

Evaluation metrics. We employ PSNR, SSIM [49], and MS-SSIM [50] as the metrics for evaluating the qualities of

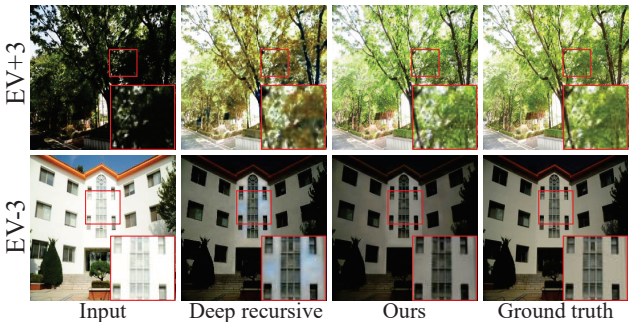


Figure 5: **Qualitative comparison of LDR image predictions in the VDS dataset [23].** Our approach recovers more details compared to Deep recursive HDRI [24] in the EV-3 example. In the EV+3 example, our approach generates LDR images with a color tone similar to the ground truth.

the predicted LDR stacks and HDR tone-mapped images. We also utilize HDR-VDP-2 [31], a metric based on the human visual system, to evaluate the quality of the reconstructed HDR images. We follow the setting of [23, 24], which sets a 24-inch monitor with a viewing distance of 0.5 meters, a peak contrast of 0.0025, and a gamma of 2.2 for measuring the HDR-VDP-2 metric.

HDR reconstruction and tone-mapping operators. Our approach uses Debevec's approach [11] to reconstruct HDR images with the predicted LDR stack and utilizes Reinhard's method [39] or Kim and Kautz's method [20] to tone-map the HDR images.

4.2. Comparison of LDR Stacks Prediction

Quantitative comparisons. The quantitative comparisons of the estimated LDR exposure stacks from the VDS dataset are shown in Tab. 1. The table shows that the proposed method performs favorably against existing methods at every exposure value. The output LDR image quality decreases as the exposure value gap increases because more extensive over- and under-exposed regions reconstruction are required, which makes the task more difficult. However, with our cycle training, our model can still generate high-quality LDR images in the cases of EV+3 and EV-3 by incorporating continuous and dense EV information into the training process. The continuous EV generation during training helps the model learn how to explicitly infer LDR images with arbitrary exposure values.

Qualitative comparisons. With the cycle training, our method can generate a high-quality LDR image even with large EV changes. A detailed qualitative comparison is presented in Fig. 5. The first row of this figure shows that Deep recursive HDRI [24] often gets less accurate color tone in estimating the LDR images, which may further degrade the quality of the HDR images fused by the LDR stack. On the contrary, our method can better estimate the LDR images in all EVs with more accurate color tones. In addition, the second row demonstrates that our method can estimate the LDR images without severe artifacts. More visual comparisons can be found in the supplementary material.

Table 2: **Quantitative comparison of HDR and TMO images.** Two tone-mapping approaches, Reinhard's approach [39], and Kim and Kautz's approach [20], are denoted as RH's and KK's TMO.

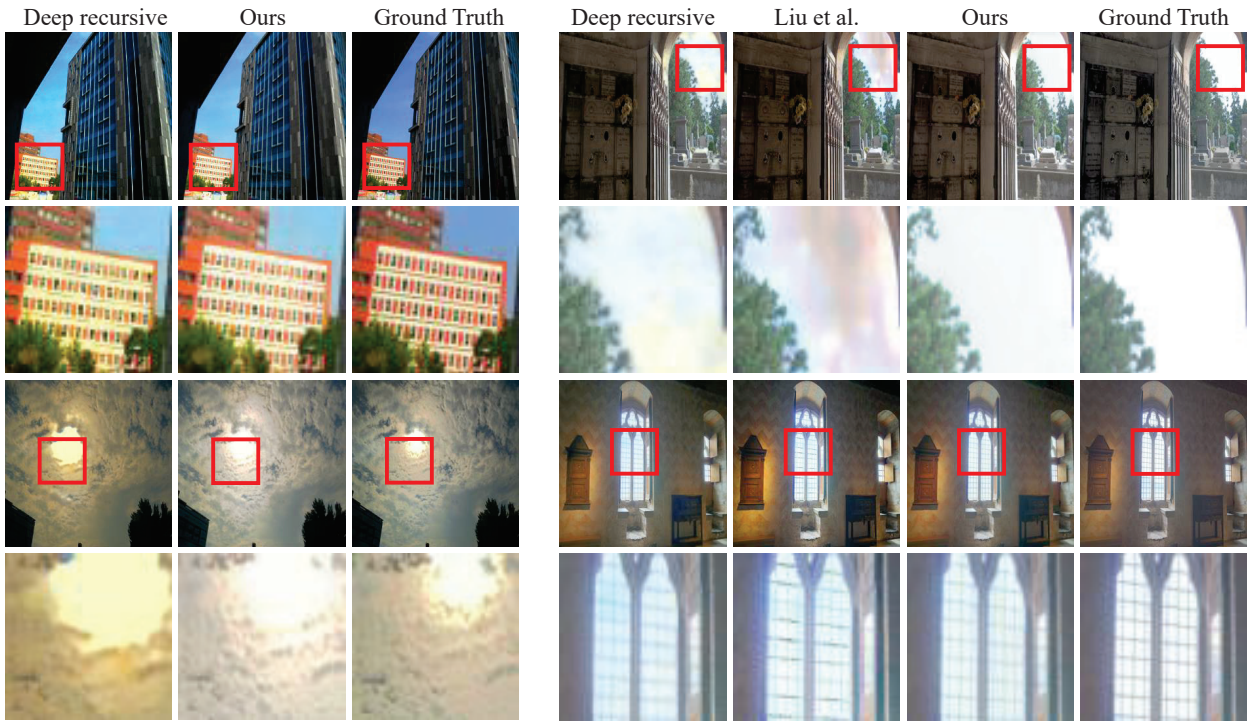
Dataset	Method	PSNR RH's TMO		PSNR KK's TMO		HDR-VDP-2	
		m	σ	m	σ	m	σ
VDS	DrTMO [13]	25.49	4.28	21.36	4.50	54.33	6.27
	Deep chain HDRI [23]	30.86	3.36	24.54	4.01	56.36	4.41
	Deep recursive HDRI [24]	32.99	2.81	28.02	3.50	57.15	4.35
	Santos et al. [43]	22.56	2.68	18.23	3.53	53.51	4.76
	Liu et al. [26]	30.89	3.27	28.00	4.11	56.97	6.15
	CEVR (Ours)	**34.67**	3.50	**30.04**	4.45	**59.00**	5.78
HDREye	DrTMO [13]	23.68	3.27	19.97	4.11	46.67	5.81
	Deep chain HDRI [23]	25.77	2.44	22.62	3.39	49.80	5.97
	Deep recursive HDRI [24]	26.28	2.70	24.26	2.90	52.63	4.84
	Santos et al. [43]	19.89	2.46	19.00	3.06	49.97	5.44
	Liu et al. [26]	26.25	3.08	24.67	3.54	50.33	6.67
	CEVR (Ours)	**26.54**	3.10	**24.81**	2.91	**53.15**	4.91

4.3. Comparison of HDR Image Prediction

We compare our method with five recent single-image HDR reconstruction methods, including Santos et al. [43], DrTMO [13], Deep chain HDRI [23], Deep recursive HDRI [24], and Liu et al. [26]. For Santos et al. [43], Deep recursive HDRI [24] and Liu et al. [26], we use their official implementations along with the released pre-trained model weight to generate all the quantitative and qualitative results on the VDS [23] and HDREye [36] datasets. For DrTMO [13] and Deep chain HDRI [23], we compare our results to the numbers reported in their papers. For HDR image prediction, our approach adopts the continuous stack strategy where the EV steps are enriched from $\{-3, -2, ..., +3\}$ to $\{-3, -2.5, ..., +3\}$, and the images with the extra EV steps are also synthesized by using the proposed CEVR model.

Quantitative evaluations. As shown in Tab. 2, our method performs favorably against the competing methods on the VDS dataset [23]. The HDREye dataset [36] serves as a blind test bed, our HDR prediction still achieves better qualities using the same tone-mapping operators. Our proposed cycle training makes the model explicitly learn the continuity as EV steps change, leading to better generalization on the unseen dataset, HDREye. With the continuous stack, more LDR images with various EVs are involved in the fusion process, which helps Debevec's approach [11] estimate a more accurate inverse CRF and generate HDR images with better qualities.

Qualitative comparisons. To generate the tone-mapped images for visual comparisons, we first reconstruct HDR images by fusing the LDR stacks with Debevec's approach [11]. Then we use Reinhard's method [39] to gen-



(a) VDS dataset (b) HDREye dataset

Figure 6: **Qualitative comparison of tone-mapped HDR images.** We adopt Reinhard's method [39] to generate HDR images shown in the figure. Deep recursive HDRI [24] often suffers from erroneous color tones and artifacts. As shown in the figure, our method can better recover details when compared to Liu et al.'s method [26]. In most regions, our method reconstructs the tone-mapped images with more accurate high-frequency details, leading to visually pleasing results.

erate HDR TMO images for all comepting methods. As Liu et al. [26] generate an HDR image directly, we apply Reinhard's tone-mapping operator [39] for tone-mapping the HDR images. Note that Liu et al. [26] do not train their method on the VDS dataset; hence we only compare the qualitative results with their method on the HDREye dataset, as shown in Fig. 6.

In Fig. 6, it can be observed that Deep recursive HDRI [24] often suffers from inaccurate color tone: the color of the building is inaccurate in the tone-mapped images, and artifacts are present in severely exposed regions. Liu et al. [26] directly estimate and reverse the whole camera pipeline to generate HDR images. It sometimes struggles with generating detailed textures and produces artifacts in severely exposed regions, e.g., the over-exposed window frame and sky in the daylight. With the intensity transformation, our model can preserve the image structure and generate the tone-mapped images with similar tones to the ground truth. It also produces fewer artifacts. More visual comparisons can be found in the supplementary material.

4.4. Ablation Studies

In the following, we validate three design contributions to improving the quality of the LDR stack and HDR images.

Table 3: **Ablation studies on the predicted LDR stack.** Intensity transformation balances two distinct tasks: brightness adjustment and precise color tone generation while preserving the image structure. Cycle training provides the model with extra information about changing EVs. Both designs improve LDR stack quality.

Intensity transformation	-		✓		✓	
Cycle training	-		-		✓	
	PSNR		PSNR		PSNR	
	m	σ	m	σ	m	σ
EV+3	31.95	4.20	33.90	3.57	**34.34**	3.46
EV+2	33.19	3.16	34.89	3.12	**35.30**	3.08
EV+1	35.09	2.56	37.49	3.07	**37.64**	2.96
EV-1	33.67	2.06	34.43	2.55	**34.62**	3.47
EV-2	32.53	3.37	32.91	4.41	**33.89**	4.34
EV-3	30.23	5.47	30.35	5.86	**30.58**	5.32

Intensity transformation. Learning to adjust image brightness while maintaining color tone accuracy and image structures can be challenging. The CEVR model, which directly outputs results from the U-net structure without using intensity transformation, can struggle to adjust brightness or produce inaccurate LDR images with artifacts, as shown in Fig. 7. The intensity transformation module can restrict

Table 4: **Ablation studies on the reconstructed HDR images on the VDS dataset [23].** Intensity transformation and cycle training enhance the quality of LDR stacks, and continuous stack benefits the stack fusion process.

Intensity transformation	-	✓	✓	✓
Continuous stack	-	-	✓	✓
Cycle training	-	-	-	✓
PSNR	32.52	34.20	34.47	**34.67**

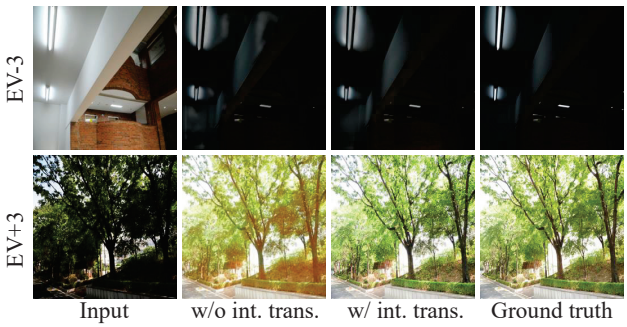


Figure 7: **Ablation on intensity transformation.** With intensity transformation, CEVR can adjust LDR image intensity while preserving the image structure and color tone.

the model's capacity, producing LDR images with more accurate brightness, color tones, and structures. This design improves the quality of the LDR stack and HDR results, as demonstrated in Tab. 3 and Tab. 4.

Cycle training. With cycle training, the model can be supervised on continuous EV steps without using the corresponding ground truth. It can learn how to change the exposure value continuously, which improves the quality and reduces the artifacts of the estimated LDR images, which also leads to better HDR quality, as shown in Fig. 8 ,Tab. 3 and Tab. 4. To further demonstrate the effectiveness of cycle training, we conducted the *hold-out* experiment, excluding EV-1 and +1 LDR images during training. Then, we used the model to estimate EV-1 and +1 LDR images for each scene and evaluated the PSNR. The table shows that our model can generate better LDR images with unseen EVs when the cycle training strategy is adopted.

Table 5: **Hold-out experiment.** Hold-out experiment excludes EV-1 and +1 LDR images during training. With cycle training, the model generate better LDR images with unseen EVs (EV-1, +1).

Cycle training	×	✓
EV-1	27.75	**33.77**
EV+1	33.37	**36.96**

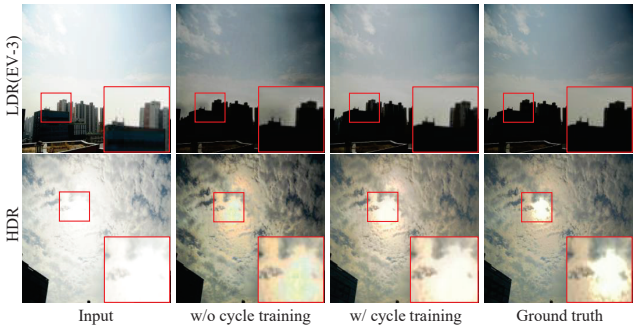


Figure 8: **Ablation on cycle training for LDR and HDR images generation.** With the cycle training, the model captures the finer granularity of "EV changing" and generates more accurate and visually pleasing LDR and HDR images.

Continuous stack. Debevec's method [11] uses the LDR stack to recover response curves and reconstruct HDR images. A denser and continuous EV LDR stack helps produce an accurate inverse CRF, enhancing HDR quality. We compare two stack settings: "predefined stack" and "continuous stack." The CEVR model estimates seven LDR images (EVs: -3, -2, -1, ..., +3) for the predefined stack, which is the setting used in existing methods, while the continuous stack has 13 LDR images with various EVs (-3, -2.5, -2, ..., +3). Tab. 4 and Fig. 1(b)(c) show that the tone-mapped image from the continuous stack has superior quality and is more visually pleasing. We can further validate the effectiveness of the continuous stack by visualizing the CRF of both the predefined stack and the continuous stack. As shown in Fig. 9, the denser EV setting can help generate a smoother CRF compared to the predefined EV setting. Additional analysis of the inverse CRF can be found in the supplementary material.

4.5. Failure cases

Although the proposed method performs favorably against other existing methods in quantitative and qualitative results, we do not explicitly design the module to address the over-exposed issue, which may make the CEVR model fail to generate reasonable content in large saturated regions, as shown in Fig. 10. It is a promising direction to take the emerging generative model designs, e.g. [28, 59], into account to address this issue.

5. Conclusion

We introduce CEVR, a learning-based method that produces LDR EV stacks from continuous EV input. Our approach combines U-Net with implicit functions, and allows the network to generate LDR images with continuous EVs. We propose two strategies, including (1) cycle training for learning on continuous EV changes unseen in the training dataset and (2) continuous stack for improving LDR stack

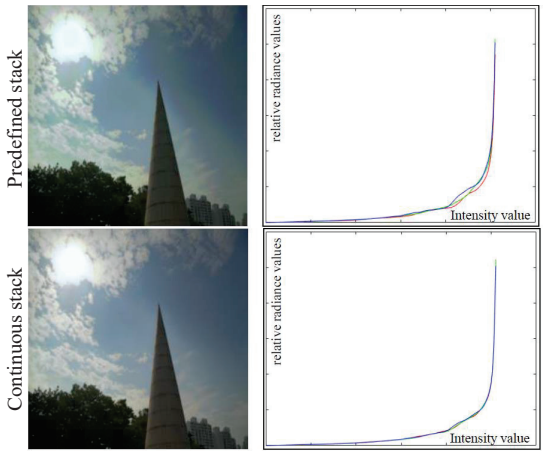


Figure 9: **Analysis of the estimated inverse CRF.** Continuous stack, adding additional LDR images with dense and continuous EVs, can help Debevec's method [11] generate more accurate inverse CRF and generate HDR images with better quality.

Figure 10: **Failure case.** Existing methods and our proposed method cannot generate reasonable content as the sky region is severely over-exposed.

fusion using additional images with dense and continuous EVs. Our approach with the two strategies greatly enhances LDR stack quality and improves HDR image results, as demonstrated through extensive quantitative and qualitative evaluations on two benchmark datasets.

Acknowledgments. This work was supported in part by National Science and Technology Council (NSTC) under grants 111-2628-E-A49-025-MY3, 112-2221-E-A49-090-MY3, 111-2634-F-002-023, 111-2634-F-006-012, 110-2221-E-A49-065-MY3 and 111-2634-F-A49-010. This work was funded in part by MediaTek.

References

[1] Ahmet Oğuz Akyüz, Roland Fleming, Bernhard E Riecke, Erik Reinhard, and Heinrich H Bülthoff. Do HDR displays support LDR content?: A psychophysical evaluation. *ACM Transactions on Graphics*, 2007. 2

[2] Matan Atzmon and Yaron Lipman. SAL: Sign agnostic learning of shapes from raw data. In *CVPR*, 2020. 3

[3] Francesco Banterle, Kurt Debattista, Alessandro Artusi, Sumanta Pattanaik, Karol Myszkowski, Patrick Ledda, and Alan Chalmers. High dynamic range imaging and low dynamic range expansion for generating HDR content. *Computer Graphics Forum*, 2009. 2

[4] Francesco Banterle, Patrick Ledda, Kurt Debattista, and Alan Chalmers. Inverse tone mapping. In *International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia*, 2006. 2

[5] Francesco Banterle, Patrick Ledda, Kurt Debattista, and Alan Chalmers. Expanding low dynamic range videos for high dynamic range applications. In *Spring Conference on Computer Graphics*, 2008. 2

[6] Francesco Banterle, Patrick Ledda, Kurt Debattista, Alan Chalmers, and Marina Bloj. A framework for inverse tone mapping. *The Visual Computer*, 2007. 2

[7] Anpei Chen, Zexiang Xu, Fuqiang Zhao, Xiaoshuai Zhang, Fanbo Xiang, Jingyi Yu, and Hao Su. Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo. In *ICCV*, 2021. 3

[8] Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *CVPR*, 2021. 2, 3

[9] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *CVPR*, 2019. 3

[10] Julian Chibane, Thiemo Alldieck, and Gerard Pons-Moll. Implicit functions in feature space for 3d shape reconstruction and completion. In *CVPR*, 2020. 3

[11] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. *ACM Transactions on Graphics*, 1997. 1, 2, 5, 6, 8, 9

[12] 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*, 2017. 1, 3

[13] Yuki Endo, Yoshihiro Kanamori, and Jun Mitani. Deep reverse tone mapping. *ACM Transactions on Graphics*, 2017. 1, 3, 6

[14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NeurIPS*, 2014. 2, 3

[15] Michael D. Grossberg and Shree K. Nayar. What is the space of camera response functions? In *CVPR*, 2003. 2

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 3

[17] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM Transactions on Graphics*, 2017. 2

[18] Erum Arif Khan, Ahmet Oguz Akyuz, and Erik Reinhard. Ghost removal in high dynamic range images. In *ICIP*, 2006. 2

[19] Junghee Kim, Siyeong Lee, and Suk-Ju Kang. End-to-end differentiable learning to hdr image synthesis for multi-exposure images. In *AAAI*, 2021. 3

[20] Min H Kim, Jan Kautz, et al. Consistent tone reproduction. In *Proceedings of the Tenth IASTED International Conference on Computer Graphics and Imaging*, 2008. 2, 6

[21] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2014. 5

[22] Jaewon Lee and Kyong Hwan Jin. Local texture estimator for implicit representation function. In *CVPR*, 2022. 3

[23] 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*, 2018. 1, 2, 3, 4, 5, 6, 8

[24] Siyeong Lee, Gwon Hwan An, and Suk-Ju Kang. Deep Recursive HDRI: Inverse tone mapping using generative adversarial networks. In *ECCV*, 2018. 1, 2, 3, 5, 6, 7

[25] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In *ECCV*, 2018. 2

[26] 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 *CVPR*, 2020. 1, 6, 7

[27] Yu-Lun Liu, Yi-Tung Liao, Yen-Yu Lin, and Yung-Yu Chuang. Deep video frame interpolation using cyclic frame generation. In *AAAI*, 2019. 4

[28] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *CVPR*, 2022. 8

[29] Stephen Mangiat and Jerry Gibson. High dynamic range video with ghost removal. In *Applications of Digital Image Processing XXXIII*, 2010. 2

[30] Steve Mann and Rosalind W. Picard. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proceedings of IS&T*, 1995. 2

[31] Rafat 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*, 2011. 6

[32] Demetris Marnerides, Thomas Bashford-Rogers, Jonathan Hatchett, and Kurt Debattista. ExpandNet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. In *EG*, 2018. 1, 3

[33] Belen Masia, Ana Serrano, and Diego Gutierrez. Dynamic range expansion based on image statistics. *Multimedia Tools and Applications*, 2017. 2

[34] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *CVPR*, 2019. 3

[35] 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*, 2021. 2, 3

[36] 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)*, 2015. 2, 5, 6

[37] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *CVPR*, 2020. 3

[38] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. DeepSDF: Learning continuous signed distance functions for shape representation. In *CVPR*, 2019. 3

[39] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. Photographic tone reproduction for digital images. *ACM Transactions on Graphics*, 2002. 2, 6, 7

[40] Allan G Rempel, Matthew Trentacoste, Helge Seetzen, H David Young, Wolfgang Heidrich, Lorne Whitehead, and Greg Ward. Ldr2Hdr: On-the-fly reverse tone mapping of legacy video and photographs. *ACM Transactions on Graphics*, 2007. 2

[41] Shunsuke Saito, Zeng Huang, Ryota Natsume, Shigeo Morishima, Angjoo Kanazawa, and Hao Li. Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In *ICCV*, 2019. 3

[42] Shunsuke Saito, Tomas Simon, Jason Saragih, and Hanbyul Joo. Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. In *CVPR*, 2020. 3

[43] 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*, 2020. 1, 3, 6

[44] Ana Serrano, Felix Heide, Diego Gutierrez, Gordon Wetzstein, and Belen Masia. Convolutional sparse coding for high dynamic range imaging. *Computer Graphics Forum*, 2016. 2

[45] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015. 3

[46] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *NeurIPS*, 2020. 2, 3

[47] Abhilash Srikantha and Désiré Sidibé. Ghost detection and removal for high dynamic range images: Recent advances. *Signal Processing: Image Communication*, 2012. 2

[48] Qianqian Wang, Zhicheng Wang, Kyle Genova, Pratul P Srinivasan, Howard Zhou, Jonathan T Barron, Ricardo Martin-Brualla, Noah Snavely, and Thomas Funkhouser. Ibrnet: Learning multi-view image-based rendering. In *CVPR*, 2021. 3

[49] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *TIP*, 2004. 5

[50] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003. 5

[51] Shangzhe Wu, Jiarui Xu, Yu-Wing Tai, and Chi-Keung Tang. Deep high dynamic range imaging with large foreground motions. In *ECCV*, 2018. 2

[52] Qiangeng Xu, Weiyue Wang, Duygu Ceylan, Radomir Mech, and Ulrich Neumann. DISN: Deep implicit surface network for high-quality single-view 3d reconstruction. *NeurIPS*, 2019. 3

[53] Xin Yang, Ke Xu, Yibing Song, Qiang Zhang, Xiaopeng Wei, and Lau Rynson. Image correction via deep reciprocating hdr transformation. In *CVPR*, 2018. 3

[54] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelnerf: Neural radiance fields from one or few images. In *CVPR*, 2021. 3

[55] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Generative image inpainting with contextual attention. In *CVPR*, 2018. 2

[56] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. *ICCV*, 2019. 2

[57] Jinsong Zhang and Jean-Francois Lalonde. Learning high dynamic range from outdoor panoramas. In *ICCV*, 2017. 3

[58] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 2

[59] Haitian Zheng, Zhe Lin, Jingwan Lu, Scott Cohen, Eli Shechtman, Connelly Barnes, Jianming Zhang, Ning Xu, Sohrab Amirghodsi, and Jiebo Luo. Image inpainting with cascaded modulation gan and object-aware training. In *ECCV*, 2022. 8