A Survey: High Dynamic Range Imaging from Traditional Methods to Deep Learning Methods

Shuhong Zheng
zhengshuhong@pku.edu.cn

Abstract

Since the very first day on which photography was invented, how to obtain high dynamic range (HDR) images from low dynamic range (LDR) images has become an important topic and the goal of researchers. HDR images contain much more details that LDR images don't have. From traditional methods to deep learning methods, the ways to generate HDR images have evolved over time. In this survey, we provide a comprehensive overview of the development of HDR imaging. First, we will give a brief introduction and overview of HDR imaging. Second, we will elaborately analyze a couple of typical works using traditional methods and deep learning methods. Third, we will present a taxonomy and some comparisons focusing on recent deep learning methods. Finally, some future perspectives of this field will be presented from my point of view.

1. Introduction and overview

In the real world, the luminance of the brightest scene is around six to eight orders of magnitude larger than the darkest. However, the dynamic range of digital sensors in cameras is relatively narrow. Moreover, the dynamic range of digital images or prints is even narrower. For example, in the common grayscale system there are only 256 discrete levels to express brightness. Thus, after setting an exposure level for a camera, we can only get the proper mapping of a small range of luminance from natural scenes to digital images, which we call LDR images. As a result, over exposure and under exposure areas frequently occur in LDR images, causing damages to the visual quality. Due to the problem stated above, researchers have two main topics of interest. The first one is how to recover the wide range of luminance from a single LDR image with a certain exposure or a set of LDR images with different exposures, which we call HDR imaging. The second one is how we can display HDR images on our devices or prints which only have a relatively narrow dynamic range, which we call tonemapping.

Back to decades ago, pioneering works such as [5, 68, 57, 42, 50] already had the insight to extend dynamic ranges of digital images to get a better visual quality. The work of [7] by Debevec *et al.* is more or less the starting point of HDR imaging researches in computer graphics. It requires a set of LDR images with different exposures as input. The following works [15, 2] explore deeper into the number of LDR images needed to reconstruct HDR images and the problem of how to choose LDR images for the reconstruction task. As a large proportion of HDR imaging methods obtain the results with a final step of merging all LDR images, there often exist some artifacts due to the misalignment of all LDR images. The works of [31, 1, 54, 61, 75, 23] are devoted to address the issue of ghosting and tearing in the reconstructed HDR images. Since LDR images with different exposures will have different levels of clarity on the same region, a weighting scheme of all LDR images naturally comes to the stage. The works of [34, 17] adopt a weighting scheme when merging the LDR images. There are also some patch-based methods [3, 58, 6, 23] that focus on the reconstruction problem at a different level, a scale between pixel-level and image-level. There are also other innovative HDR imaging methods which use rank constraint [47, 48] and median threshold bitmap (MTB) algorithm [49], achieving satisfactory HDR reconstruction performances.

When the deep learning era finally arrived, researchers began to explore how deep neural networks can be applied to HDR imaging. One popular idea is to design a network that can learn a good way of merging multiple LDR images with different exposures [28, 69, 72, 71, 73]. People believe the capability of deep neural networks to automatically learn latent information behind images can help refine the quality of the reconstructed HDR results. Another idea is to train the neural networks to learn the transformation from LDR images to HDR images [11, 44, 36, 37, 41], thus only requiring a single LDR image to accomplish HDR reconstruction. Using a single image on this task can avoid the alignment step which is indispensable when we have to deal with a bunch of LDR images. Therefore, methods using a single LDR image can successfully escape from the long-standing and intractable problem of ghosting and tearing that we have long struggled against. Researchers also attempt applying techniques that are prevalent in other deep learning tasks, such as attention mechanism [65, 71, 39], generative adversarial networks (GAN) [14, 46, 70] and perceptual loss [27, 56, 41].

Besides, there are some works that focus on how to modify the camera hardware to obtain better performance on HDR imaging. With the specifically designed cameras, we can perform HDR imaging with a single shot. For example, Zhao *et al*. proposed a modulo camera [77] which can achieve unbounded high dynamic range in some sense. A few methods use a beam-splitter to split the light to multiple sensors [63, 45]. Other approaches [19, 20, 16] proposed to perform HDR reconstruction from coded per-pixel exposure. However, these fancy cameras have special and unique optical systems and sensors, which are typically customized and expensive, thus not available to the general public.

Moreover, HDR video generation is also a hot topic in recent years [29, 33, 4]. This task is not simply transforming each frame in the video from LDR images to HDR images. Videos have to be consistent in the time domain. If the HDR reconstruction results in the neighboring frames vary a lot, there will exist some artifacts when videos are played. Also, HDR videos are of great practical use in the commercial market. People will enjoy the wonderful experience of watching high-quality programs in front of an HDR television or other digital screens that can render HDR videos.

Finally, researchers have broad interest in searching for proper metrics to evaluate the quality of HDR reconstruction. Common evaluation metrics for image processing such as peak signal to noise ratio (PSNR) and structural similarity (SSIM) are not perfectly suitable to judge the results of HDR imaging. Kuang *et al*. proposed [35] to give an evaluation metric specially designed for the task of HDR imaging. Karaduzovic-Hadziabdic *et al*. proposed [30] to evaluate HDR reconstruction results in both subjective and objective ways. The work of [43] by Mantiuk *et al*. proposed a calibrated visual metric for HDR images' visibility and quality in all luminance conditions.

The rest of this survey is structured as follows. In Section 2, we will give an exhaustive analysis of several typical methods of HDR imaging, including both traditional methods and deep learning methods. In Section 3, we will make a classification of recent deep learning methods for HDR imaging and compare them with one another. In Section 4, some future prospectives of HDR imaging from my point of view will be proposed. Finally, the conclusion of this survey is presented in Section 5.

2. Analysis on specific works

In this section, first we give our analysis on works [7, 53] which are typical traditional ways of rendering HDR images in Section 2.1. Deep learning methods are analyzed in Section 2.2 and Section 2.3, according to whether they use multi-view LDR images [28, 69] or a single LDR image [36, 37, 41], respectively.

2.1. Traditional methods

In this part, we choose to analyze a classic work that focuses on recovering luminance information of HDR images [7] and a typical work that gives an empirical way to do tonemapping [53].

2.1.1 Recovering HDR radiance maps

Debevec *et al*. proposed a far-reaching way in [7] to recover HDR radiance maps from photographs obtained with conventional imaging equipment. As shown in Figure 1, the image acquisition pipeline has several non-linear mappings which make the task of recovering HDR radiance maps challenging. The algorithm takes a number of photographs taken from the same position with different known exposure durations as input and obtains the estimated HDR radiance maps in two steps. The first step is to attain a non-linear function f as an estimation of the composition of all the non-linear functions in the image acquisition pipeline. If we denote the digital value at pixel i with exposure duration Δt_j as Z_{ij}, the radiance at each pixel i as E_i, then we have the following equation:

$$Z_{ij} = f(E_i \Delta t_j) \tag{1}$$

Eq. 1 holds for any i and j, ranging over all pixels and exposure durations. Then we transfer the equation to the logarithm field and replace the function $\ln f^{-1}$ with notation g for simplicity:

$$f^{-1}(Z_{ij}) = E_i \Delta t_j \tag{2}$$

$$\ln f^{-1}(Z_{ij}) = \ln E_i + \ln \Delta t_j \tag{3}$$

$$g(Z_{ij}) = \ln E_i + \ln \Delta t_j \tag{4}$$

Since we only care about the function value of g on a finite number of positions, which are $\{Z_{ij}\}$ for all pixels and all exposure durations. This task turns into a problem to solve an over-determined system of linear equations, where the goal is to find the best g in a least squared error sense. This mathematical problem can be robustly solved by singular value decomposition (SVD).

However, it is prevalent that certain parts in the LDR images suffer from over exposure or under exposure so there will be only little information or even noisy information in these regions. Thus, we need to adjust their weights in the linear system. An easy and sensible way to assign the weight for a pixel value Z_{ij} is formulated as below:

$$w(Z_{ij}) = \begin{cases} Z_{ij} - Z_{\min} & \text{for} \quad Z_{ij} \leqslant Z_{\text{mid}} \\ Z_{\max} - Z_{ij} & \text{for} \quad Z_{ij} > Z_{\text{mid}} \end{cases} \tag{5}$$

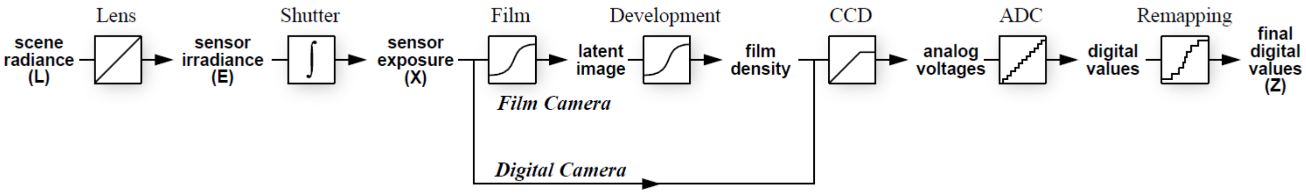

Figure 1. Image acquisition pipeline from [7] which maps the radiance at each position to the corresponding pixel value

where $w(Z_{ij})$ is a weighting function deciding the weights assigned to pixels according to their pixel values. $Z_{\max}$ and $Z_{\min}$ mean the the largest and the smallest pixel value in $\{Z_{ij}\}$ respectively, and Z_{mid} equals to the average of $Z_{\max}$ and $Z_{\min}$.

The objective function with the weighting scheme is formulated as below:

$$\begin{aligned}\mathcal{O} = &\sum_i \sum_j \{w(Z_{ij})[g(Z_{ij}) - \ln E_i - \ln \Delta t_j]\}^2 \\ &+ \lambda \sum_{z=Z_{\min}+1}^{Z_{\max}-1} [w(z)g''(z)]^2\end{aligned} \tag{6}$$

where $g''(z) = g(z-1)+g(z+1)-2g(z)$ serves as a regularization term constraining the smoothness of the function g and λ is a hyperparameter adjusting the importance of the regularization term in the objective function.

Once we obtain the estimation of the response function g, the second step to reconstruct the radiance map becomes much easier. From Eq. 4, we obtain:

$$\ln E_i = g(Z_{ij}) - \ln \Delta t_j \tag{7}$$

We also apply the weighting function in the reconstruction process, so we recover the HDR radiance at each pixel as below:

$$\ln E_i = \frac{\sum_j w(Z_{ij})[g(Z_{ij}) - \ln \Delta t_j]}{\sum_j w(Z_{ij})} \tag{8}$$

Up to now the reconstruction of the HDR radiance map is done. The output of this method is only an estimation of the relative radiance in the logarithm domain, not an absolute one. This is because the value of the objective function $\mathcal{O}$ stays the same when all the function values of g are multiplied by a scalar factor. Afterwards, there are some discussions about the number of images we need for recovering the HDR radiance map and how to deal with color images.

2.1.2 Photographic tone reproduction

Tonemapping is a procedure enabling us to show HDR images on digital devices or prints, which are originally designed to show LDR images. So it can also be regarded as a procedure that maps HDR images to LDR images, meanwhile hoping to preserve as much visual information as possible. Reinhard *et al*. proposed [53] to address this issue utilizing the idea of dodging and burning which are two procedures of classical photography technology centuries ago. The popular trend to display HDR images at that time is related to something called the Zone System, which simply maps the middle brightness region to middle gray and then infers the gray scales of other regions. However, this method only provides users with a small set of subjective controls so the work aims to suggest a new tone reproduction operator to gain more controls.

The first step of this method is to get the key value of a given image. The key value indicates the overall brightness of an image. For example, an image of a white-painted room will have a relatively high key value, while a dim image will have a relatively low key value. Following the works of [67, 64, 21], we get the key value L_k as:

$$L_k = \exp\left[\frac{1}{N}\sum_{x,y} \ln\left(\delta + L_w(x,y)\right)\right] \tag{9}$$

where $L_w(x,y)$ is the real luminance at position (x,y), N is the total number of pixels in the image and δ is a small constant to avoid singularity when doing the logarithm operation. As pointed out by Reinhard in [52], there is a minor mistake in the paper of [53] on the above equation and Eq. 9 here is the correct one.

After obtaining the key value of an image, we can perform a scaling to all pixels in the image:

$$L(x,y) = \frac{a}{L_k} L_w(x,y) \tag{10}$$

where $L(x,y)$ is the scaled luminance value at position (x,y) and a varies according to users' preference of a relatively bright or dark output. After this process, the luminance values of images are adjusted to a similar range no matter what key values they have.

The second step of this method is to perform automatic dodging and burning on the luminance maps obtained from the first step. In classical photography and printing techniques, dodging means withholding a portion of light while burning means adding more light. This will lighten or darken certain regions in the final display, thus creating a better visual quality. The automatic dodging and burning in this method is realized by multi-scale Gaussian filters $\{R_i\}$ defined as below:

$$R_i(x,y,s) = \frac{1}{\pi(\alpha_i s)^2} \exp\left[-\frac{x^2+y^2}{(\alpha_i s)^2}\right] \tag{11}$$

where s indicates the operation scale of the Gaussian filters $\{R_i\}$, $\{\alpha_i\}$ are the hyperparameters defining a series of $\{R_i\}$. In this method we have two Gaussian filters R_1 and R_2, defined by α_1 and α_2, serving as a center detector and a surround detector, respectively.

The filtering procedure can be formulated as:

$$V_i(x,y,s) = L(x,y) \otimes R_i(x,y,s) \tag{12}$$

where the notation "$\otimes$" means the operation of convolution. After we obtain the responses V_1 and V_2, we can define a center-surround function V as follows:

$$V(x,y,s) = \frac{V_1(x,y,s) - V_2(x,y,s)}{\frac{2^\phi a}{s} + V_1(x,y,s)} \tag{13}$$

where center V_1 and surround V_2 responses are derived from Eqs. 11 and 12. The hyperparameter ϕ is a sharpening parameter while a is chosen from Eq. 10.

The center-surround function V is computed for the sole purpose of finding an appropriate Gaussian kernel size s_m, which may be different for each pixel. This procedure of selecting s_m is the key to the success of the proposed automatic dodging and burning technique, as we wish to perform Gaussian filtering in a region where no large contrast changes occur. So we search for the largest s that holds the following inequality:

$$|V(x,y,s) < \varepsilon| \tag{14}$$

where ε is the threshold we choose and the largest s is the appropriate Gaussian kernel size s_m that we are searching for. This is carried out for each pixel so s_m is a function of the pixel position (x,y), so it can be denoted as $s_m(x,y)$. After that, the tone reproduction operator can be formulated as:

$$L_d(x,y) = \frac{L(x,y)}{1 + V_1\left(x,y,s_m(x,y)\right)} \tag{15}$$

where $L_d(x,y)$ is the tonemapping result of an HDR image.

This tone reproduction function constitutes the local dodging and burning scheme, thus producing better visual results. This idea can also be considered as choosing a different parameter a in Eq. 10 for each pixel as the local key value for each pixel varies a lot. Other contemporary works utilized bilateral filtering [9] and gradient fields [12] to perform tonemapping, which also produce impressing results.

2.2. Deep learning methods with multi-view images

In this part, we give our analysis on works [28, 69] which use multiple LDR images with different exposures. These methods have more information of the image but they often suffer from misalignment which will cause ghosting and tearing in the final reconstructed HDR results.

2.2.1 Deep HDR imaging with dynamic scenes

The work of [28] proposed by Kalantari *et al.* is the first one to use deep learning method to reconstruct an HDR image from a set of bracketed exposure LDR images of a dynamic scene. Also, they introduce the first dataset suitable for learning HDR reconstruction, which facilitates future deep learning researches in this domain. Previous methods suffer from the artifacts such as ghosting and tearing in the reconstructed HDR images, due to the inproper alignment of certain pixels with motions. This deep learning method utilizes convolutional neural networks (CNN) to merge LDR images, hoping to mitigate the artifacts caused by misalignment. The overall process can be broken down into two stages: alignment and HDR merge.

The alignment part is accomplished by optical flows [40] and it takes the median exposure image Z_2 of the three input images $\{Z_1, Z_2, Z_3\}$ as the reference image, which means the low exposure image Z_1 and the high exposure image Z_3 are required to be aligned with the structure of Z_2.

In the HDR merge part the authors proposed three different network architectures, each has its pros and cons: (1) the direct CNN, (2) the CNN serving as a weight estimator and (3) the CNN serving as a weight and image estimator. Three CNN structures can all be formally written as:

$$H = g(\mathcal{I}, \mathcal{H}) \tag{16}$$

where g is the mapping conducted by CNN, $\mathcal{I} = \{I_i\}$ is the set of aligned images in LDR domain, $\mathcal{H} = \{H_i\}$ is the set of aligned images in HDR domain by simply performing gamma correction on $\mathcal{I}$, and H is the estimated HDR image. The gamma correction procedure is formulated as:

$$H_i = \frac{I_i^\gamma}{t_i} \tag{17}$$

where I_i and H_i are pixel values of the LDR images and the roughly estimated HDR images, while t_i denotes the exposure time of the image I_i.

The direct CNN, which is the first and simplest architecture, directly outputs the estimated values $\hat{H}$ for HDR images. For the second one, the CNN served as a weight estimator, it outputs the blending weights $\alpha_i(x,y)$ instead. Then the HDR result $\hat{H}$ is produced as below:

$$\hat{H}(x,y) = \frac{\sum_{i=1}^{3} \alpha_i(x,y) H_i(x,y)}{\sum_{i=1}^{3} \alpha_i(x,y)} \tag{18}$$

For the CNN served as a weight and image estimator, the final structure, it outputs not only the weights $\alpha_i(x, y)$ but also the refined aligned images $\{\tilde{I}_i\}$. Thus, the HDR result $\hat{H}$ is calculated by Eq. 18 with $\{H_i\}$ replaced by:

$$\tilde{H}_i = \frac{\tilde{I}_i^{\gamma}}{t_i} \tag{19}$$

The loss function of the network is calculated in the field of tonemapped images. Therefore, the HDR result $\hat{H}$ needs a final tonemapping process. Gamma encoding, defined as $H^{\frac{1}{\gamma}}$ with $\gamma > 1$, is perhaps the simplest way of tonemapping in HDR image processing. However, since it is not differentiable around zero, we use the μ-law compressor which is commonly used in audio processing to compress HDR images to the LDR domain. The transformation is defined as:

$$T = \frac{\log(1 + \mu H)}{\log(1 + \mu)} \tag{20}$$

where μ is the parameter which defines the amount of compression, H is the HDR image in the linear domain, and T is the tonemapped image. After we obtain the estimated HDR image $\hat{T}$ in the tonemapped domain, we can define our loss function as the L2 distance between our estimation $\hat{T}$ and the ground truth T:

$$\mathcal{L} = ||T - \hat{T}||_2^2 \tag{21}$$

Among the above three architectures, the first one is undoubtedly the simplest. Of the rest two architectures, the CNN acting as both a weight and image estimator can produce results with the least numerical errors. However, it will have some overblurred results in dark regions which make its visual quality not as good as the other one.

Furthermore, the authors point out that the estimation step of optical flows can also be designed into the network, as [8, 24] both show that deep neural networks perform fairly well in the task of estimating optical flows. The learning of optical flows can be jointly trained with the merging procedure.

2.2.2 Dealing with large foreground motions

Since previous works that use a set of bracketed exposure LDR images suffer from misalignment, Wu *et al*. proposed [69] to formulate HDR imaging as an image-to-image translation problem [25] without optical flows. As shown in [8] and [79], CNN has the ability to learn misalignment and hallucinate missing details, so it can be a remedy for misalignment and the ghosting problem. The problem setting is the same as Eq. 16. The network takes a set of multiple exposure LDR shots as input, and outputs the estimated HDR result. The HDR set $\mathcal{H}$ in Eq. 16 is obtained the same way as the previous work [28] by Eq. 17.

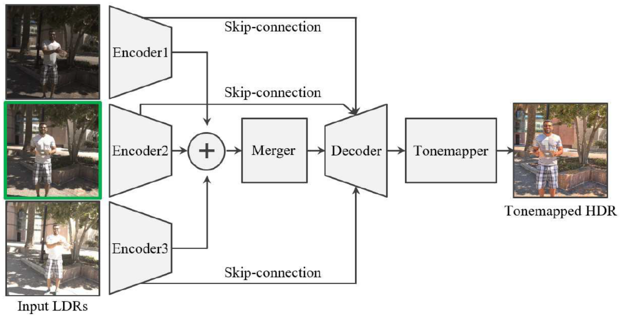

Figure 2. Overall architecture of [69]

The overall architecture of this method is shown in Figure 2, which can be conceptually divided into three components: encoder, merger and decoder. The network structure is implemented by adopting the ideas from both U-Net [55] and ResNet [18]. The U-Net framework has a bottle-neck characteristic that is suitable for image translation problem, as the latent code dimension at the merger stage can be relatively low. Also, its skip-connections can ensure that features extracted at different levels can be effectively fed to the decoder, thus resulting in high-quality image generation. The ResNet framework introduces the technique of residual blocks that can replace a number of middle layers in the U-Net and therefore boost the learning process of the network. Although there are multiple inputs, the encoders for the differently exposed LDR images only vary at the first couple of layers, so they can still share a large portion of parameters.

After we get the outputs from the network, we utilize the μ-law as Eq. 20 to get the tonemapped image and Eq. 21 as the loss function for the network.

The authors of this work point out that their method can handle LDR images with large foreground motions because they choose not to use optical flows to do the alignment. They argue that the distortions and artifacts produced by previous methods are mainly consequences of erroneous alignment performed by optical flows. Also, in cases with large foreground motions sometimes there are serious occlusions or largely saturated regions, the image translation framework can hallucinate plausible details, thus providing us with better visual effects.

2.3. Deep learning methods with a single image

Different from Section 2.2, in this part we will pay our attention to the deep learning methods which use a single LDR image for HDR reconstruction. The works of [11, 10] start this research topic, and they are also representatives of the two main categories of this kind of approaches. The work of [11] by Endo *et al*. represents the category in which we first infer LDR images with different exposures with the single LDR image we have, then we apply merging on the

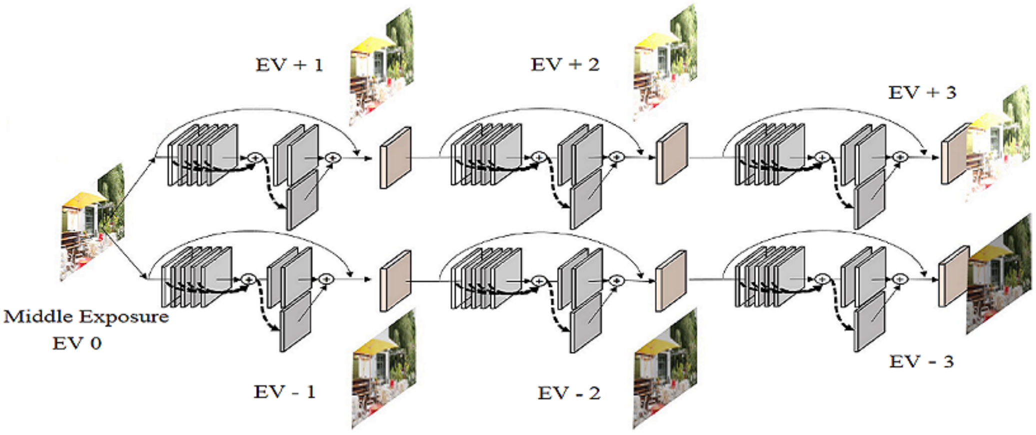

Figure 3. Inference network of [36]

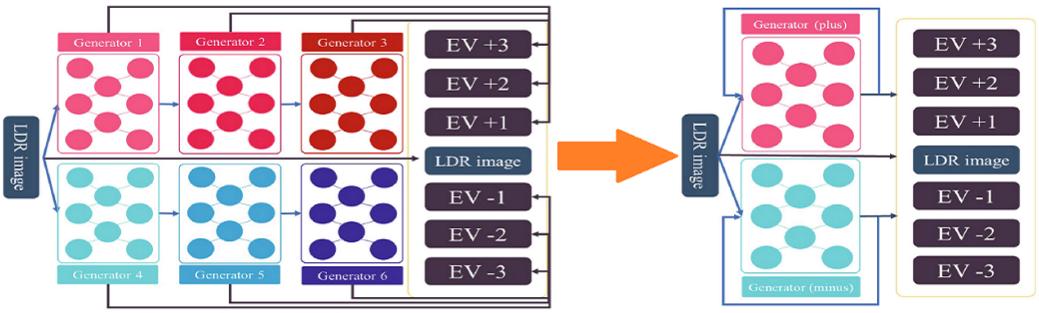

Figure 4. Transformation of the framework from [36] to [37]

LDR images we obtain from the first step. We will show a couple of works of this kind [36, 37] in Section 2.3.1. On the other hand, the work of [10] by Eilertsen *et al*. completes the reconstruction task with the only given LDR image from the beginning to the end, without generating a set of bracketed LDR images. We will analyze an example of this kind [41] in Section 2.3.2.

2.3.1 Deep chain / recursive HDR imaging

Lee *et al*. proposed [36] to use a chain structure to sequentially infer LDR images with various exposures from a single LDR image of the same scene. The inference network is shown in Figure 3. The authors devise such a chain stucture to infer images with various exposures step by step because they find that when the exposure time difference of two images is considerably large, their gap becomes unbridgeable. Therefore, inferring extremely low or high exposure images from a single middle exposure image is way too difficult for neural networks. They validate the superiority of their chain structure by an ablation study which compares their chain structure with a specialized network in the task of generating extremely high exposure images. It turns out that it's hard for the specialized network to output plausible results but the inference results from the chain structure are relatively satisfactory.

Later, they borrow the idea from recurrent neural networks (RNN) and proposed [37] to change the former chain structure to a recursive structure. The change from [36] to [37] is shown in Figure 4. In the recurrent structure, the generator responsible for inferring higher exposure images share the weights and so does the generator responsible for inferring lower exposure images. Therefore, the change from a chain structure to a recurrent structure can make the training process more efficient in both memory and computation.

As for the loss function, in the reconstruction term, they adopt L1 loss rather than L2 loss to enable the robustness of the network. Besides, they add a discriminator to the network to gain authenticity of the final HDR results. However, this discriminator is not the traditional one in common GAN structure, but a Markovian discriminator structure proposed in [38] which generates feature maps that consider the neighboring pixels in an input through convo-

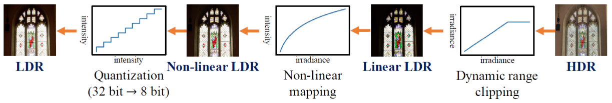

Figure 5. Camera pipeline that performs the transformation

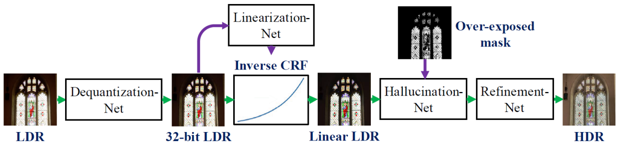

Figure 6. Network structure of [41]

lutional layers. Hence, this network not only focuses on the pixel-wise difference to the ground truth but the patch-wise difference as well.

Due to the unstable training process of GAN, the authors suggest a two-phase training strategy. In the first phase the network only uses L1 loss to get roughly reconstructed HDR images. Then in the second phase the network combines L1 loss and discriminator loss together to refine the results.

2.3.2 Learning to reverse the camera pipeline

Liu *et al.* proposed [41] to step back from aimlessly devising fancy and complicated deep neural networks. They go back to the origin and take a close observation on how the camera pipeline transforms HDR images to LDR images, as shown in Figure 5. According to this, they design three specialized networks to seperately deal with the three steps shown in the pipeline: dynamic range clipping, non-linear mapping and quantization. The three specialized networks are supposed to learn the reverse operation of the pipeline and also arranged in the reverse order of the original pipeline, as shown in Figure 6, which are dequantization, linearization and hallucination.

When LDR images are taken in the whole framework, the first step is to go through the dequantization network, which the authors adopt a U-Net structure. The second step is to transfrom the 32-bit LDR to linear LDR by the linearization network. The primary goal of the linearization framework is to recover the inverse camera response function (CRF). The authors use a ResNet as the backbone to obtain the recovered inverse CRF. The third step is to do the hallucination to recover the regions whose details are lost at the dynamic range clipping stage. The authors follow the work of [10] to design the hallucination network. Finally, the images will go through a refinement network which enhances the quality of the final outputs.

The three networks for dequantization, linearization and hallucination are first pretrained seperately. The loss function for the dequantization network is formulated as:

$$\mathcal{L}_{\text{deq}} = ||\hat{I}_{\text{deq}} - I_n||_2^2 \tag{22}$$

where $\hat{I}_{\text{deq}}$ is the dequantized LDR image and I_n is the corresponding ground truth image.

The loss function for the linearization network is formulated as:

$$\mathcal{L}_{\text{lin}} + \lambda_{\text{crf}}\mathcal{L}_{\text{crf}} \tag{23}$$

where $\mathcal{L}_{\text{lin}} = ||\hat{I}_{\text{lin}} - I_c||_2^2$ is the L2 distance between the linearized image and the corresponding ground truth. $\mathcal{L}_{\text{crf}} = ||\hat{g} - g||_2^2$ is the L2 distance between the reconstructed inverse CRF and ground truth inverse CRF.

The loss function for the hallucination network is formulated as:

$$\mathcal{L}_{\text{hal}} + \lambda_{\text{p}}\mathcal{L}_{\text{p}} + \lambda_{\text{tv}}\mathcal{L}_{\text{tv}} \tag{24}$$

where $\mathcal{L}_{\text{hal}} = ||\log\hat{H} - \log H||_2^2$ is the L2 distance between the reconstructed HDR image and ground truth HDR image in the logarithm domain. $\mathcal{L}_{\text{p}}$ is the perceptual loss [27] calculated at certain layers of the VGG network [62]. $\mathcal{L}_{\text{tv}}$ is the total variation loss calculated on the recovered HDR image, hoping to improve the spatial smoothness of the predicted contents.

After all three networks converge, finally we can jointly fine-tune the entire pipeline by minimizing the combination of all the loss functions mentioned above, which we call $\mathcal{L}_{\text{total}}$:

$$\begin{aligned} \mathcal{L}_{\text{total}} = \lambda_{\text{deq}}\mathcal{L}_{\text{deq}} + \lambda_{\text{lin}}\mathcal{L}_{\text{lin}} + \lambda_{\text{crf}}\mathcal{L}_{\text{crf}} \\ + \lambda_{\text{hal}}\mathcal{L}_{\text{hal}} + \lambda_{\text{p}}\mathcal{L}_{\text{p}} + \lambda_{\text{tv}}\mathcal{L}_{\text{tv}} \end{aligned} \tag{25}$$

The two-phase training strategy reduces error accumulation between separate networks and further improves the HDR reconstruction performance.

3. Taxonomy and comparison

In this section we present our taxonomy and comparison of different deep learning methods. In Section 2 we have already roughly split all deep learning methods into two parts: using a single LDR image and using multiple LDR images. The goal of the methods using multiple LDR images with different exposures is clear, which is to mitigate the effect of ghosting and tearing in the final HDR reconstruction results. These artifacts, as discussed in [69], is mainly caused by the misalignment of multi-view LDR images. Therefore, researchers are devoted in seeking ways to perform better alignment, bringing up methods such as applying attention mechanism to multi-view LDR images [71].

Compared to methods using multiple LDR images, the inter-relationship between those using a single LDR image is much more complex. As implied in Section 2.3, HDR imaging method using a single image can be roughly classified into two categories, with [11, 10] being their pioneers,

Method	Generate a bracketed of LDR images first	Apply end-to-end training	Use adversarial training	Use perceptual loss
EI17 [10]	×	×	×	×
ZH17 [76]	×	✓	×	×
MA18 [44]	×	✓	×	×
YA18 [74]	×	✓	×	×
LI20 [41]	×	×	×	✓
EN17 [11]	✓	✓	×	×
LE18a [36]	✓	✓	×	×
LE18b [37]	✓	×	✓	×
NI18 [46]	✓	✓	✓	×
KH19 [32]	✓	✓	×	✓
XU19 [70]	✓	✓	✓	✓
SA20 [56]	✓	×	×	✓

Table 1. Comparison of 12 selected works that use a single image for HDR reconstruction

respectively. Leading by [11], the methods that first generate a bracketed of LDR images focus on how to make their inference of other LDR images with different exposures more accurate. They have tried adopting adversarial training and perceptual loss which are broadly used in other image generation tasks. For those that don't need to generate a set of LDR images which are led by [10], researchers often take a look back at the camera pipeline. They hope to design a series of networks that can each function as a component in the camera pipeline. Altogether they can do the same as the physical devices so the whole network will learn the mapping between LDR images and HDR images.

In Table 1, we show our comparison of 12 selected works that use a single image for HDR reconstruction. We compare them in four aspects: (1) whether they generate a bracketed of images first, (2) whether they apply end-to-end training, (3) whether they use adversarial training and (4) whether they use perceptual loss.

4. Future prospective

As we can see in the previous sections, more and more works begin to apply techniques like GAN, perceptual loss and attention mechanism which are commonly used in other computer vision tasks. Also, many researchers suggest that we can regard the problem of HDR imaging as an image-to-image translation task, which learns a mapping between the LDR domain and the HDR domain. Thus, the techniques that are often used in image-to-image translation framework can also be applied to the HDR imaging task, like cycle consistency loss in CycleGAN [78].

Inspired by this, we can consider the dynamic range of an image as a kind of semantic, and treat the HDR imaging problem as a semantic editing task. There are a bunch of semantic editing works, including both supervised methods [59, 13, 26] and unsupervised methods [51, 66, 22, 60]. Among them, the work of [26] is especially suitable for the task of HDR imaging. This work offers us a way to perform camera movement and other image manipulation tasks by simply editing latent codes of images in the latent space, as shown in Figure 7. There is a good possibility that the dynamic range of an image can also be controlled in this way by manipulating its latent code. If this idea works well, we can perform the transition between LDR images and HDR images arbitrarily. I believe there are numerous potentials in using semantic editing on HDR imaging for future researches.

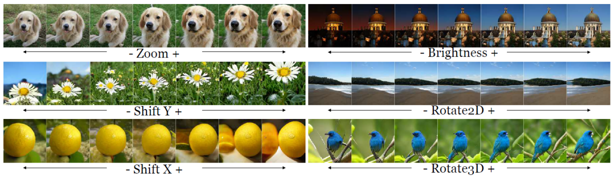

Figure 7. Some semantic editing results in [26]

5. Conclusion

In this survey, we present a comprehensive review of HDR imaging, from traditional methods to recent deep learning ones. We classify them into several categories and summarize the challenges that a group of methods come across. We also elaborate how the methods deal with the challenges and how they have evolved over time. Finally, from current works we come up with some potentials on which no researches have been conducted by now. This survey is definitely not exhaustive, but we hope it can be inspiring for people interested in the field of HDR imaging.

References

[1] A. O. Akyuez and E. Reinhar. Noise reduction in high dynamic range imaging. *Journal of Visual Communication & Image Representation*, 18(5):366–376, 2007. 1

[2] N. Barakat, A. N. Hone, and T. E. Darcie. Minimal bracketing sets for high-dynamic-range image capture. *IEEE Transactions on Image Processing (TIP)*, 2008. 1

[3] C. Barnes, E. Shechtman, B. G. Dan, and A. Finkelstein. The generalized patchmatch correspondence algorithm. *European Conference on Computer Vision (ECCV)*, 2010. 1

[4] G. Chen, C. Chen, S. Guo, Z. Liang, K. Y. K. Wong, and L. Zhang. Hdr video reconstruction: A coarse-to-fine network and a real-world benchmark dataset. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 2

[5] K. Chiu, M. Herf, P. Shirley, S. Swamy, C. Wang, and K. Zimmerman. Spatially nonuniform scaling functions for high contrast images. *Proceedings of Graphics Interface*, pages 245–253, 1993. 1

[6] S. Darabi, E. Shechtman, C. Barnes, D. B. Goldman, and P. Sen. Image melding: Combining inconsistent images using patch-based synthesis. *Proceedings of SIGGRAPH*, 2012. 1

[7] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. *Proceedings of SIGGRAPH*, 1997. 1, 2, 3

[8] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Smagt, D. Cremers, and T. Brox. Flownet: Learning optical flow with convolutional networks. *IEEE International Conference on Computer Vision (ICCV)*, 2015. 5

[9] F. Durand and J. Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. *Proceedings of SIGGRAPH*, 2002. 4

[10] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger. Hdr image reconstruction from a single exposure using deep cnns. *Proceedings of SIGGRAPH Asia*, 2017. 5, 6, 7, 8

[11] Y. Endo, Y. Kanamori, and J. Mitani. Deep reverse tone mapping. *Proceedings of SIGGRAPH Asia*, 2017. 1, 5, 7, 8

[12] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. *Proceedings of SIGGRAPH*, 2002. 4

[13] L. Goetschalckx, A. Andonian, A. Oliva, and P. Isola. Ganalyze: Toward visual definitions of cognitive image properties. *IEEE International Conference on Computer Vision (ICCV)*, 2019. 8

[14] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, X. Bing, and Y. Bengio. Generative adversarial nets. *Conference on Neural Information Processing Systems (NeurIPS)*, 2014. 2

[15] M. D. Grossberg and S. K. Nayar. High dynamic range from multiple images: Which exposures to combine. *Workshop on Color and Photometric Methods in Computer Vision*, 2003. 1

[16] S. Hajisharif, J. Kronander, and J. Unger. Adaptive dualiso hdr reconstruction. *EURASIP Journal on Image and Video Processing*, 2015. 2

[17] S. W. Hasinoff, F. Durand, and W. T. Freeman. Noise-optimal capture for high dynamic range photography. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010. 1

[18] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 5

[19] F. Heide, M. Steinberger, Y. Tsai, M. Rouf, D. Pajkak, D. Reddy, O. Gallo, J. Liu, W. Heidrich, K. Egiazarian, J. Kautz, and K. Pulli. Flexisp: A flexible camera image processing framework. *ACM Transactions on Graphics (TOG)*, 2014. 2

[20] F. Heide, G. Wetzstein, B. Masia, A. Serrano, and D. Gutierrez. Convolutional sparse coding for high dynamic range imaging. *Computer Graphics Forum: Journal of the European Association for Computer Graphics*, 35(2):153–163, 2016. 2

[21] J. Holm. Photographics tone and colour reproduction goals. *CIE Expert Symposium on Colour Standards for Image Technology*, 1996. 3

[22] E. Hrknen, A. Hertzmann, J. Lehtinen, and S. Paris. Ganspace: Discovering interpretable gan controls. *Conference on Neural Information Processing Systems (NeurIPS)*, 2020. 8

[23] J. Hu, O. Gallo, K. Pulli, and X. Sun. Hdr deghosting: How to deal with saturation? *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013. 1

[24] E. Ilg, N. Mayer, T. Saikia, M. Keuper, and T. Brox. Flownet 2.0: Evolution of optical flow estimation with deep networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 5

[25] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 5

[26] A. Jahanian, L. Chai, and P. Isola. On the steerability of generative adversarial networks. *International Conference on Learning Representations (ICLR)*, 2020. 8

[27] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. *European Conference on Computer Vision (ECCV)*, 2016. 2, 7

[28] N. K. Kalantari and R. Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *Proceedings of SIGGRAPH Asia*, 2017. 1, 2, 4, 5

[29] N. K. Kalantari and R. Ramamoorthi. Deep hdr video from sequences with alternating exposures. *Eurographics (EG)*, 2019. 2

[30] K. Karaduzovic-Hadziabdic, J. H. Telalovic, and R. Mantiuk. Subjective and objective evaluation of multi-exposure high dynamic range image deghosting methods. *Eurographics (EG)*, 2016. 2

[31] E. A. Khan, A. O. Akyuz, and E. Reinhard. Ghost removal in high dynamic range images. *International Conference on Image Processing (ICIP)*, 2006. 1

[32] Z. Khan, M. Khanna, and S. Raman. Fhdr: Hdr image reconstruction from a single ldr image using feedback network. *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2019. 8

[33] S. Y. Kim, J. Oh, and M. Kim. Jsi-gan: Gan-based joint super-resolution and inverse tone-mapping with pixel-wise

task-specific filters for uhd hdr video. *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2020. 2

[34] K. Kirk and H. J. Andersen. Noise characterization of weighting schemes for combination of multiple exposures. *British Machine Vision Conference (BMVC)*, 2006. 1

[35] J. Kuang, H. Yamaguchi, C. Liu, G. M. Johnson, and M. D. Fairchild. Evaluating hdr rendering algorithms. *ACM Transactions on Applied Perception*, 4(2), 2007. 2

[36] S. Lee, G. H. An, and S. J. Kang. Deep chain hdri: Reconstructing a high dynamic range image from a single low dynamic range image. *IEEE Access*, 2018. 1, 2, 6, 8

[37] S. Lee, G. H. An, and S. J. Kang. Deep recursive hdri: Inverse tone mapping using generative adversarial networks. *European Conference on Computer Vision (ECCV)*, 2018. 1, 2, 6, 8

[38] C. Li and M. Wand. Precomputed real-time texture synthesis with markovian generative adversarial networks. *European Conference on Computer Vision (ECCV)*, 2016. 6

[39] J. Li and P. Fang. Hdrnet: Single-image-based hdr reconstruction using channel attention cnn. *International Conference on Multimedia Systems and Signal Processing (ICMSSP)*, 2019. 2

[40] C. Liu. Beyond pixels: exploring new representations and applications for motion analysis. *Massachusetts Institute of Technology*, 2009. 4

[41] Y. L. Liu, W. S. Lai, Y. S. Chen, Y. L. Kao, M. H. Yang, Y. Y. Chuang, and J. B. Huang. Single-image hdr reconstruction by learning to reverse the camera pipeline. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1, 2, 6, 7, 8

[42] S. Mann and R. Picard. Being undigital with digital cameras: extending dynamic range by combining differently exposed pictures. *IS&T's Annual Conference*, 1995. 1

[43] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich. Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on Graphics (TOG)*, 30(4):40, 2011. 2

[44] D. Marnerides, T. Bashford-Rogers, J. Hatchett, and K. Debattista. Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. *Eurographics (EG)*, 2018. 1, 8

[45] M. McGuire, W. Matusik, H. Pfister, B. Chen, J. F. Hughes, and S. K. Nayar. Optical splitting trees for high-precision monocular imaging. *IEEE Computer Graphics & Applications*, 2007. 2

[46] S. Ning, H. Xu, L. Song, R. Xie, and W. Zhang. Learning an inverse tone mapping network with a generative adversarial regularizer. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018. 2, 8

[47] T. H. Oh, J. Y. Lee, and I. S. Kweon. High dynamic range imaging by a rank-1 constraint. *International Conference on Image Processing (ICIP)*, 2013. 1

[48] T. H. Oh, J. Y. Lee, Y. W. Tai, and I. S. Kweon. Robust high dynamic range imaging by rank minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 37(6):1219–1232, 2015. 1

[49] F. Pece and J. Kautz. Bitmap movement detection: Hdr for dynamic scenes. *Conference on Visual Media Production (CVMP)*, 2013. 1

[50] Z. U. Rahman, D. J. Jobson, and G. W. Woodell. Multiscale retinex for color rendition and dynamic range compression. *SPIE Proceedings: Applications of Digital Image Processing*, 2847:183–191, 1996. 1

[51] A. Ramesh, Y. Choi, and Y. Lecun. A spectral regularizer for unsupervised disentanglement. *International Conference on Machine Learning (ICML)*, 2019. 8

[52] E. Reinhard. Parameter estimation for photographic tone reproduction. *Journal of Graphics Tools*, 7(1):45–51, 2002. 3

[53] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. Photographic tone reproduction for digital images. *Proceedings of SIGGRAPH*, 2002. 2, 3

[54] R. Revathi and T. Kavitha. Artifact-free high dynamic range imaging. *International Conference on Computational Photography (ICCP)*, 2009. 1

[55] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2015. 5

[56] M. S. Santos, T. I. Ren, and N. K. Kalantari. Single image hdr reconstruction using a cnn with masked features and perceptual loss. *Proceedings of SIGGRAPH*, 2020. 2, 8

[57] C. Schlick. Quantization techniques for visualization of high dynamic range pictures. *Eurographics Workshop on Rendering*, 1994. 1

[58] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman. Robust patch-based hdr reconstruction of dynamic scenes. *Proceedings of SIGGRAPH Asia*, 2012. 1

[59] Y. Shen, J. Gu, X. Tang, and B. Zhou. Interpreting the latent space of gans for semantic face editing. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 8

[60] Y. Shen and B. Zhou. Closed-form factorization of latent semantics in gans. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 8

[61] D. D. Sidibe, W. Puech, and O. Strauss. Ghost detection and removal in high dynamic range images. *European Signal Processing Conference*, 2009. 1

[62] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations (ICLR)*, 2015. 7

[63] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen. A versatile hdr video production system. *ACM Transactions on Graphics (TOG)*, 2011. 2

[64] J. Tumblin. Tone reproduction for computer generated images. *IEEE Computer Graphics and Applications*, 13:42–48, 1993. 3

[65] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *Conference on Neural Information Processing Systems (NeurIPS)*, 2017. 2

[66] A. Voynov and A. Babenko. Unsupervised discovery of interpretable directions in the gan latent space. *International Conference on Machine Learning (ICML)*, 2020. 8

[67] G. J. Ward. A contrast-based scale factor for luminance display. *Graphics Gems IV*, 1994. 3

[68] G. J. Ward. The radiance lighting simulation and rendering system. *Proceedings of SIGGRAPH*, 1994. 1

[69] S. Wu, J. Xu, Y. W. Tai, and C. K. Tang. Deep high dynamic range imaging with large foreground motions. *European Conference on Computer Vision (ECCV)*, 2018. 1, 2, 4, 5, 7

[70] Y. Xu, S. Ning, R. Xie, and L. Song. Gan based multi-exposure inverse tone mapping. *International Conference on Image Processing (ICIP)*, 2019. 2, 8

[71] Q. Yan, D. Gong, Q. Shi, A. Hengel, C. Shen, I. Reid, and Y. Zhang. Attention-guided network for ghost-free high dynamic range imaging. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1, 2, 7

[72] Q. Yan, D. Gong, P. Zhang, Q. Shi, J. Sun, I. Reid, and Y. Zhang. Multi-scale dense networks for deep high dynamic range imaging. *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2019. 1

[73] Q. Yan, L. Zhang, Y. Liu, Y. Zhu, J. Sun, Q. Shi, and Y. Zhang. Deep hdr imaging via a non-local network. *IEEE Transactions on Image Processing (TIP)*, 29:4308–4322, 2020. 1

[74] X. Yang, K. Xu, Y. Song, Q. Zhang, X. Wei, and R. Lau. Image correction via deep reciprocating hdr transformation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 8

[75] S. H. Yong, K. M. Lee, U. L. Sang, Y. Moon, and J. Cha. Ghost-free high dynamic range imaging. *Asian Conference on Computer Vision (ACCV)*, 2010. 1

[76] J. Zhang and J. F. Lalonde. Learning high dynamic range from outdoor panoramas. *IEEE International Conference on Computer Vision (ICCV)*, 2017. 8

[77] H. Zhao, B. Shi, C. Fernandez-Cull, S. K. Yeung, and R. Raskar. Unbounded high dynamic range photography using a modulo camera. *International Conference on Computational Photography (ICCP)*, 2015. 2

[78] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *IEEE International Conference on Computer Vision (ICCV)*, 2017. 8

[79] S. Zhu, S. Liu, C. C. Loy, and X. Tang. Deep cascaded bi-network for face hallucination. *European Conference on Computer Vision (ECCV)*, 2016. 5