Deep Polarimetric HDR Reconstruction

Juiwen Ting, Moein Shakeri, Hong Zhang

Abstract—This paper proposes a novel learning based high-dynamic-range (HDR) reconstruction method using a polarization camera. We utilize a previous observation that polarization filters with different orientations can attenuate natural light differently, and we treat the multiple images acquired by the polarization camera as a set acquired under different exposure times, to introduce the development of solutions for the HDR reconstruction problem. We propose a deep HDR reconstruction framework with a feature masking mechanism that uses polarimetric cues available from the polarization camera, called Deep Polarimetric HDR Reconstruction (DPHR). The proposed DPHR obtains polarimetric information to propagate valid features through the network more effectively to regress the missing pixels. We demonstrate through both qualitative and quantitative evaluations that the proposed DPHR performs favorably than state-of-the-art HDR reconstruction algorithms.

Index Terms—High dynamic range imaging, polarization, convolutional neural network, feature masking

I. INTRODUCTION

In real-world scenes, the luminance varies over several orders of magnitude and is presented as high-dynamic-range (HDR). Unfortunately, standard digital cameras can only capture a limited fraction of this range due to sensor constraints. As a result, the images captured by such cameras are low-dynamic-range (LDR) with over- or under-exposed regions that do not reflect the human visual system’s ability to simultaneously identify details in both bright and dark regions of a scene. In LDR images, information is lost in the over- or under-exposed regions, and thus presents challenges to computer vision related robotic tasks [1]–[3]. In HDR reconstruction, the method seeks to restore the lost information, and aims to output an HDR image that prevails a wider range of illumination than is available directly from the camera sensor.

Classical HDR reconstruction methods propose to generate an HDR image from a sequence of LDR images of the scene taken at different exposures [4], [5]. However, these approaches require multiple exposure images as input, which may not always be available and practical. Single-shot HDR techniques resolves the above issue by requiring only a single capture. As a result, they have attracted considerable research efforts in recent years.

A plethora of single-shot HDR works have been proposed [6]–[11]. However, almost all existing methods have some significant drawbacks including ignoring reconstruction for under-exposed regions, neglecting dynamic range expansion for properly exposed pixels, or generating artifacts (i.e., color distortions, halo artifacts). An interesting recent development of imaging technology is a polarization camera that uses multi-directional on-chip linear polarizers to capture multi-polarization images in single-shot. Polarization is a property of light, which can convey rich illumination information about the environment to help with HDR reconstruction. Recent studies show that the polarization camera can handle the aforementioned drawbacks due to the acquisition of the different polarization filtered images. These polarization filtered images can capture different intensity values, where a region that is poorly exposed in one image may be properly exposed in another image. An existing polarimetric method by Wu et al. [12] is a traditional model based approach, and is vulnerable to deal with a) poorly exposed pixels in polarized images, and b) pixels with low degree of polarization. Another existing polarimetric method by Ting et al. [13] is a deep-learning-based approach, whose input consists of the four images from the polarization camera. However, where the degree of polarization is high, the method proposed by Ting et al. [13] is unable to take advantage of the known physics of the polarization images as compared to the method by Wu et al. [12]. Specifically, Ting et al. [13] does not utilize degree of polarization to construct HDR images.

Motivated by these shortcomings, we present a deep-learning-based solution to the problem of HDR reconstruction using a polarization camera, called Deep Polarimetric HDR Reconstruction (DPHR). We design our approach based on two main observations. First, pixels in an LDR image can be a) under-exposed, b) properly exposed, or c) over-exposed.
In HDR imaging, the goal is to not only recover the lost information in poorly exposed regions, but to also increase the bit-depth of the properly exposed pixels beyond 8-bits. Therefore, we use the polarimetric cues to design a novel reconstruction process to a) identify reconstruction regions based on our feature masking mechanism for the convolutional neural network (CNN) to focus its learning upon, and b) propagate the properly exposed pixels that contains the valid features through the network. Second, the method by Wu et al. [12] is able to expand the dynamic range in areas that have a strong measure of polarized light. Therefore, we combine the HDR images generated by a traditional model based approach [12] and our deep-learning-based DPHR model to obtain the final HDR result. We evaluate the effectiveness of our method on outdoor scenes captured by a polarization camera. Fig. 1 shows a sample result of the proposed DPHR method. The main contributions of our work are as follows:

- We propose a feature masking mechanism to solve the HDR reconstruction problem using a CNN and a polarization camera. This masking approach is based on the domain knowledge of polarization, and effectively identifies the invalid content that requires HDR recovery. Thus, the quality of the final results expands the dynamic range by synthesizing realistic textures and has reduced artifacts.
- We introduce an HDR reconstruction formulation that combines the complementary results from the traditional model-based and deep-learning-based approaches to generate the final output.

The remainder of this paper is organized as follows. Related works on HDR reconstruction and the background physics of light polarization are summarized in Section 2. Section 3 presents the details of our proposed DPHR method based on the polarization camera. Experimental results and discussions are described in Section 4, and concluding remarks in Section 5.

II. RELATED WORKS

In this section, we first review the relevant works to the problem of single-shot and multiple-shot HDR reconstruction, and then explain the background physics of light polarization that is the basis of our proposed method.

A. Multiple-shot HDR Reconstruction

Multiple-shot HDR reconstruction is a well-known area of research with a series of works that focuses on removing ghosting artifacts caused by moving objects or misalignment in the images shot from multiple acquisitions. Several traditional works perform alignment and HDR reconstruction in a unified optimization system. For example, Sen et al. [14] and Hu et al. [15] proposed patch-based optimization systems to fill in the missing details due to the saturated regions in the reference image using other images within the stack. Other works by Oh et al. [16], Lee et al. [17] and Li et al. [18] use rank minimization where image misalignments are considered as sparse outliers, to align and reconstruct an HDR image. However, the above techniques are not data-driven and produce unsatisfactory results in challenging cases where the reference has pronounced saturated regions.

To alleviate the aforementioned issues, a group of works explores learning based approaches to directly align and reconstruct pixels in all saturated regions. Some works proposed a pre-processing step to align the multi-exposure images prior to feeding them to a CNN. For example, pre-processed steps based on optical flow and homography transformation are proposed, respectively by Kalantari et al. [19] and Wu et al. [20]. Other works implemented an end-to-end HDR framework. For example, Yan et al. [4]. Liu et al. [21] and Pu et al. [22] presented attention-merging networks to generate ghost-free HDR imaging. Ma et al. [23] proposed an attention-merging network with guided image filtering to optimize the fusion weight maps of the multi-exposure images. Niu et al. [24] implemented a generative adversarial network (GAN) based model to fuse the multi-exposure images into an HDR image. In a similar vein, Kalantari et al. [25] uses two connected networks to fuse multiple LDR frames, and generate an HDR video. All of these methods need multiple shots, though some methods have accounted for motions between the input images [4], [21], [22]. suppressing severe motions remain a challenge thus is not practical in applications with moving cameras. On the other hand, perfectly aligned polarization images can be captured in just a single-shot, which is suitable for applications with both stationary and moving cameras.

B. Single-shot HDR Reconstruction

Single-shot HDR reconstruction is a well-studied area of research with a plethora of solutions developed for various applications in computer vision and robotics. A classical and important group of these solutions is expansion operators, where different heuristic operators are implemented to transform a single LDR image to an HDR image [26]–[32]. However, these techniques make assumptions about the scene and use heuristics that result in deviation from the true luminance value, and thus tend to induce pixel artifacts.

To address these limitations, the second group of methods take advantage of the power of CNNs and directly expands the dynamic range to recover pixels in poorly exposed regions. Several novel architectures have been proposed to utilize contextual information extracted by convolutional layers to improve HDR reconstruction results [6]–[11]. For example, Marnerides et al. [6] proposed a three-branch CNN to extract global, semi-local, and local features that exploit feature recovery at various scales, for reconstruction in all poorly exposed regions. However, their method outputs overly bright and smooth images as it over-enhances the extracted illumination features. Liu et al. [11] presented an HDR framework that incorporates the domain knowledge of the LDR image formation pipeline. They trained three specialized networks to reverse the image formation steps of dynamic range clipping, non-linear mapping from a CRF, and quantization to reconstruct an HDR image. Chen et al. [33] proposed a U-Net like network to solve the joint tasks of HDR reconstruction, denoising, and dequantization. Zhang et al. [34] designed a three-stage cascade network to decompose the LDR image to learn the HDR
content. Eilersten et al. \cite{7} proposed a U-Net like architecture to regress details in over-exposed regions with a fixed mask, and later combined the prediction with the input image for the final result. Santos et al. \cite{8} developed an autoencoder structure with a feature masking mechanism to also regress details in over-exposed regions, and then reconstruct the HDR image through combining with the input image for the end result. Namely, the feature masking mechanism by Santos et al. \cite{8} is based on a soft mask computed from the pixel intensity of the input image. Thus, dark pixels in the soft mask correspond to over-exposed pixels where the CNN focuses its learning upon, and the soft mask is continuously updated by the feature weights of each convolutional layer. However, both methods by Eilersten et al. \cite{7} and Santos et al. \cite{8} neglect HDR reconstruction for under-exposed regions, and is unable to improve HDR estimation for properly exposed regions, as it applies a mask that focuses the reconstruction on over-exposed regions. Furthermore, the methods by Eilersten et al. \cite{7} and Santos et al. \cite{8} has unreliable reconstruction in all poorly exposed regions due to the lack of prior to enforce consistency.

The third group of methods rely on specialized cameras or filters, such as neuromorphic camera \cite{35}, saturation self-reset camera \cite{36}, modulo camera \cite{37}, snapshot camera \cite{38}–\cite{40}, polarization camera \cite{12} or cross-screen filter \cite{41} to recover HDR images. Alternatively, some works presented end-to-end joint learning of optics and HDR reconstruction. For example, Metzler et al. \cite{42} and Sun et al. \cite{43} jointly trained an optical-based element and a CNN to hallucinate the HDR content from a single LDR image. As another example, Martel et al. \cite{44}, Alghamid et al. \cite{45}, \cite{46}, Potiadou et al. \cite{47} and Serrano et al. \cite{48} proposed joint designs to optimize a snapshot sensor and a deep learning algorithm to reconstruct HDR images. A snapshot camera is an imaging device capable of acquiring multiple images in a single capture. This is beneficial for HDR imaging as more information (i.e., luminance) about the scene can be captured in a single-shot, to help recover the scene’s dynamic range. For example, Nayar et al. \cite{39} proposed a learning-based approach to handle the images captured by a spatially varying exposure camera (i.e., multiple exposure images), and generate an HDR image. Recently, traditional model based \cite{12} and deep-learning-based \cite{13} approaches using a polarization camera have been proposed for HDR reconstruction, and showed promising results. However, the method by Wu et al. \cite{12} is vulnerable to challenging cases with significant poorly exposed and low degree of polarization pixels. Though the method by Ting et al. \cite{13} can reconstruct better details than Wu et al. \cite{12}, Ting et al. \cite{13} still outputs limited details. In contrast to the method by Ting et al. \cite{13} that uses only the polarization images to achieve HDR reconstruction, we propose an enhanced deep-learning-based HDR reconstruction method called DPHR, that uses both the polarization images and polarimetric cues captured by the polarization camera to reconstruct HDR images as will be explained in Section \ref{iii}

\section{Background Physics of Light Polarization}

The proposed DPHR method is based on a polarization camera, which implements pixel-level polarizer filters that has made it possible to acquire real-time and high resolution measurement of incident polarization information. Each calculation unit has four pixels with four on-chip directional polarizer filters, at 0°, 45°, 90°, and 135°, to capture four spatially aligned polarization images. Each polarization image is a filtered version of the image irradiance \( I \) captured by the polarization camera, and thus the effect of a polarizer on image irradiance can be written as \cite{49}:

\[
I_0, I_{90} = 0.5 \times I(1 \pm \rho \cos 2\theta) \\
I_{45}, I_{135} = 0.5 \times I(1 \pm \rho \sin 2\theta)\tag{1}
\]

where \( \rho \in [0, 1] \) denotes the degree of linear polarization (DoLP), and measures the portion of light that is polarized for a given pixel. \( \theta \in [0^\circ, 180^\circ] \) denotes the angle of linear polarization (AoLP), and measures the direction of light in which the polarized light oscillates at a given pixel. The DoLP can be defined as follows:

\[
\rho = \frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{max}} + I_{\text{min}}}\tag{2}
\]

where \( I_{\text{max}} \) and \( I_{\text{min}} \) are the maximum and the minimum measured radiance at every pixel, respectively. These two parameters along with AoLP \( \theta \) can be determined by three or more polarization images with different but known polarizer filter angles. As our camera can simultaneously capture four images, the DoLP and AoLP can be readily computed, and their closed form solutions can be found in the work by Huynh et al. \cite{50}.

In general, the relation between image irradiance \( I_i \) and pixel value \( L_i \) at exposure time \( t_0 \) can be written as \cite{51}:

\[
L_i = f(I_i t_0)\tag{3}
\]

where \( f \) is the camera response function. By substituting \cite{1} into \cite{3}, and applying the reciprocity relation in the work byDebevec and Malik \cite{51}, we obtain:

\[
t_1, t_3 = 0.5 \times t_0 (1 \pm \rho \cos 2\theta) \\
t_2, t_4 = 0.5 \times t_0 (1 \pm \rho \sin 2\theta)\tag{4}
\]

The polarization image formation model in the case of HDR reconstruction is provided by \cite{4}. The key observation is that when the incoming light is not entirely unpolarized (\( \rho \neq 0 \)), the four polarizer orientation images experience different exposure times, effectively creating the condition for multiple exposures. In other words, each polarizer orientation image covers a different luminance range, and thus can be combined to recover a wider dynamic range by synthesizing realistic textures.

\section{Proposed Method}

Fig. 2 shows an overview of the deep neural network used in constructing \( H_d \) in our proposed DPHR framework. The DPHR network architecture is summarized in Table \ref{table1}. Our goal is to reconstruct HDR values for all properly exposed and poorly exposed pixels. We achieve this using a CNN that takes as input four LDR polarization images captured by the polarization camera, and estimates the missing HDR information in all properly exposed and poorly exposed regions with the help from the feature masking mechanism. We compute
CNN to estimate HDR for all properly exposed and poorly exposed pixels extracted at each convolutional layer, and guides the CNN to estimate HDR for all properly exposed and poorly exposed regions. On the other hand, where DoLP is high, it indicates large pixel variations among the four polarization images, where existing rich and reliable polarimetric cues are available for HDR recovery. Equipped with this domain knowledge about polarization, it allows us to use this polarimetric cue as a prior in the form of soft mask for our feature masking mechanism, to reconstruct HDR using polarization images.

In the proposed DPHR method, the first step is to stack the four LDR polarization images \( L_i, i = 1,2,3,4 \) (512x512x12), to account for the contributions from filters oriented at \( 0^\circ, 45^\circ, 90^\circ, \) and \( 135^\circ \), respectively for input to the CNN. The values in \( L_i \) are in the range of \([0,1]\). Then we use \( L_i \) to compute the feature mask \( M_{l,i} \) where the subscripts \( l \) and \( i \) are the layer in the CNN and the index of each polarizer orientation, respectively. Specifically, we compute the feature mask \( M_{l,i} \) for input to the feature masking mechanism. The feature mask is constructed where if the pixel value of \( L_i \) is properly exposed and DoLP is low, then the reconstruction is dominated by the predicted image \( H_{d} \). However, if \( L_i \) is poorly exposed and DoLP is high, then the reconstruction is dominated by the traditional model based method \( H_t \). Then for \( L_i \) and DoLP values that lie somewhere in between, the reconstruction is shared by the \( H_d \) and \( H_t \) images. The feature mask \( M_{l,i} \) is defined as follows:

\[
M_{l,i} = \frac{\rho + K_i}{\max(\rho + K_i)}
\]

where the subscript 1 indicates the first layer in the CNN (\( l = 1 \)). The values in \( K_i \) are in the range of \([0,1]\), and indicate the proper exposedness of each input pixel (of \( L_i \)) based on the pixel intensity, as shown in Fig. 3. The value 0 in \( K_i \) indicates the input pixel is completely over-exposed. Therefore, \( M_{l,i} \) computes the well exposedness of each input pixel based on the combination of polarization and intensity information. The values in \( M_{l,i} \) are in the range of \([0,1]\). The value 0 in \( M_{l,i} \) indicates the feature is computed from low polarization and poorly exposed pixels, and thus represents invalid content that require reconstruction by the CNN. Then the feature mask is updated at each convolutional layer \( l \) to obtain \( M_{l,i} \). As a result, the reconstruction is achieved by using the mask \( M_{l,i} \)

![Input Mask](image)

**Fig. 2:** Overview of the deep neural network used in constructing \( H_d \) in our proposed DPHR framework. The LDR input images \( L_i, i = 1,2,3,4 \) are propagated through the network, while updated by the corresponding mask \( M_{l,i} \) before going through the convolutional layers. The mask at each layer is obtained by updating the mask from the previous layer.

**TABLE I:** Overview of DPHR network architecture

| Layer | Stage | # Filters | Filter size | Conv. stride | Activation |
|-------|-------|-----------|-------------|--------------|------------|
| 1     | conv+act | 64        | (3,3)       | (2,2)        | ReLU       |
| 2     | conv+bn+act | 128       | (3,3)       | (2,2)        | LeakyReLU  |
| 3     | conv+bn+act | 256       | (3,3)       | (2,2)        | LeakyReLU  |
| 4     | conv+bn+act | 512       | (3,3)       | (2,2)        | LeakyReLU  |
| 5     | conv+bn+act | 512       | (3,3)       | (2,2)        | LeakyReLU  |
| 6     | conv+bn+act | 512       | (3,3)       | (2,2)        | LeakyReLU  |
| 7     | conv+bn+act | 512       | (3,3)       | (2,2)        | LeakyReLU  |
| 8     | conv+bn+act | 512       | (3,3)       | (2,2)        | LeakyReLU  |
| 9     | conv+bn+act | 512       | (3,3)       | (2,2)        | LeakyReLU  |
| 10    | conv+bn+act | 512       | (3,3)       | (2,2)        | LeakyReLU  |
| 11    | conv+bn+act | 512       | (3,3)       | (2,2)        | LeakyReLU  |
| 12    | conv+bn+act | 256       | (3,3)       | (1,1)        | LeakyReLU  |
| 13    | conv+bn+act | 128       | (3,3)       | (1,1)        | LeakyReLU  |
| 14    | conv+bn+act | 64        | (3,3)       | (1,1)        | LeakyReLU  |
| 15    | conv+act   | 3         | (3,3)       | (1,1)        | LeakyReLU  |

An input mask image for the feature masking mechanism. Then the input mask updates the features at each convolutional layer using the obtained polarimetric information, in order to reduce the magnitude of the features generated by the invalid areas, and thus better propagate the valid features. Finally, we combine the predicted HDR image with the HDR image generated from the traditional model based approach [12] to obtain the final HDR result.

**A. Deep Polarimetric HDR Reconstruction**

As discussed in Section 11-C, the polarization state of light can be quantified by DoLP, which determines the portion of light that is polarized for a given pixel. Where DoLP is low, it indicates a low measure of polarization that can be attributed to similar pixel values or pixel saturation among the four polarization images. Such regions correspond to a higher weight in the CNN, and thus are regions that the CNN will direct its learning upon. In other words, the polarimetric information provided by DoLP weighs the contribution of the features extracted at each convolutional layer, and guides the CNN to estimate HDR for all properly exposed and poorly exposed regions.
where the perceptual loss has been demonstrated useful to improve visual quality \cite{8, 43}, and its key idea is to enhance the similarity in feature space between the ground truth and predicted image. Our perceptual loss is a combination of the VGG loss $\mathcal{L}_v$ and style loss $\mathcal{L}_s$ as follows:

$$\mathcal{L}_p = \lambda_3 \mathcal{L}_v + \lambda_4 \mathcal{L}_s$$

(10)

where $\lambda_3 = 1.0$ and $\lambda_4 = 120.0$ in our implementation.

The VGG loss is defined as follows:

$$\mathcal{L}_v = \sum_i \| \phi_l(H_g) - \phi_l(H_d) \|_1$$

(11)

where $\phi_l$ is the feature map from the $l$-th layer of the pretrained VGG-19. Then to recover more vivid textures, we add the style loss to (10), and the style loss is defined as follows:

$$\mathcal{L}_s = \sum_l \| G_l(H_g) - G_l(H_d) \|_1$$

(12)

where $G_l$ is the Gram matrix \cite{52} applied on the feature map at the $l$-th layer of the pretrained VGG-19. For the perceptual loss, we extract feature maps from pool1, pool2 and pool3 layers of the VGG-19 network.

Fig. 3 shows the mask $M_{l,i=1}$ at different layers of the network. We can observe from the input image that the building and the fence regions are over-exposed. We can also observe from the input mask that these regions indicate a lower polarization and a poorer exposed. The input mask informs the network where the invalid features are for the CNN to direct its learning. Therefore, the building and the fence regions consistently indicate a lower polarization and a poorer exposed pixel (i.e., darker pixel) compared to other regions in the masks at each convolutional layer. Also observed is that the masks become blurrier and more uniform in the deeper layers, as the receptive fields of the convolution filters become larger.

B. HDR Formulation

With the CNN predicted image $H_d$, we formulate the final HDR image $H$ through HDR integration as follows:

$$H = \alpha \cdot H_t + (1 - \alpha) \cdot H_d$$

(13)

where $\alpha$ is as follows:

$$\alpha = \frac{\rho}{\rho + (1 - M_t)}$$

(14)

In (13), $H_t$ is computed as follows \cite{12}:

$$H_t = \frac{\sum_{i=1}^{2} W(L_i + L_{i+2}) \left( g(L_i) + g(L_{i+2}) \right)}{\sum_{i=1}^{2} W(L_i + L_{i+2}) t_0}$$

(15)
where $W$ is the Gaussian weighted function ($\sigma = 0.2$ in our study), and $g$ is the inverse camera response function.

The HDR formulation by (13) uses a combination of $H_t$ which is a traditional model based method, and $H_d$ which is a deep-learning-based method, to estimate HDR in all areas. $H_t$ can estimate HDR well for all areas with high DoLP [12]. Then we let $H_d$ to estimate HDR for all areas with low DoLP, using the predicted pixels from the CNN. Additionally, normalization is performed where necessary.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we first present the experimental setups, and then evaluate the experimental results of the proposed DPHR method on our collected outdoor polarization images. Specifically, we evaluate the effectiveness of the different components in our HDR formulation. To show the capability of the proposed DPHR method, we compare the reconstructed results with state-of-the-art methods.

A. Experiment Setups

**Dataset.** Since there is a lack of polarization images for training our network for HDR reconstruction, we have built our own dataset. We collected an outdoor dataset, under sunlight conditions, named EdPolCommunity. The dataset is collected by the IMX250MYR polarization camera [53]. For each scene, four high-resolution 1024x1224 colored polarization images is captured at 17 exposure times ($t_0 = 0.03, 0.045, 0.068, 0.101, 0.152, 0.228, 0.342, 0.513, 0.769, 1.153, 1.73, 2.595, 3.592, 5.839, 8.758, 13.137, 19.705$ ms). Then four 512x512 patches are cropped from each image for improved readability. In total, the EdPolCommunity dataset has 736 pairs of ground truth HDR and LDR images available to train and test the network. We follow a 70/20/10 practice to split our dataset into train/val/test sets. The images used for training are not used in the evaluation. Fig. 5 shows examples of input LDR images $L_1$ at $t_0 = 0.769$ ms from the EdPolCommunity dataset.

To gain further insights on the EdPolCommunityOutdoor dataset, a mixture model and the expectation maximization algorithm [54] is used to fit the DoLP distribution. The mixture model is given by:

$$f = w \cdot \gamma + (1 - w) \cdot U$$

(16)

where $w$ is the weight of the distributions, $\gamma$ is the Gamma distribution, and $U$ is the uniform distribution. For the EdPolCommunityOutdoor dataset: $w = 0.934$, $\gamma = 6.264$, $\gamma_{\beta} = 0.023$, $U_{\text{start}} = 0.0$ and $U_{\text{end}} = 1.0$. We note that (16) is a general case that considers the presence of shot noise which is prevalent in poorly-lit scenes. For the EdPolCommunity- Outdoor dataset, $f$ in (16) is predominantly weighted by the Gamma distribution since its weight $w$ is 0.934, and thus the data can be explained well with only the Gamma distribution. The negative likelihood calculated for the mixture model is -548.895. Additionally, the acquisition settings are summarized in Table II.

**TABLE II: Polarized capture settings**

| Aperture | Focal length | FOV | Pixel format | Pixel range | ADC bit depth | Frame rate | # Exposures |
|----------|--------------|-----|--------------|-------------|--------------|------------|-------------|
| 16       | 8            | 58’x49’x73” | Polarized 8  | 0-255       | 12           | 73         | 17          |

**Implementation Details.** The network adopts a U-Net architecture [55]. All the encoder layers use the leaky ReLU activation ($\alpha = 0.2$), while we use ReLU in all the decoder layers, except for the last one, which has a linear activation. Also the feature masking strategy is used in all the convolutional layers. The framework is implemented using PyTorch and trained on NVIDIA GeForce GTX 1080 Ti. We initialized the network using the Xavier approach. Then trained using an Adam optimizer [56] with the default parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and mini-batch size of 4.

**Evaluation Metrics.** To evaluate the quality of the HDR images, we used the popular HDR-VDP2 [57] metric, and adopted the means squared error (MSE) metric as used in other works [8]. We normalize the predicted and reference ground truth HDR images [6]. Then to account for the human visual system’s sensitivity to lighting at different luminance values, we apply perceptual uniform (PU) encoding to calculate PU-PSNR and PU-SSIM metrics [58], to further evaluate the accuracy of the HDR images.

B. Evaluation of the HDR Formulation

In this experiment, we validate the model design choice for the HDR formulation, which is performed by comparing the following variants of the HDR formulation:

(a) **DPHR (i.e., $H = \alpha \cdot H_t + (1 - \alpha) \cdot H_d$).** The full HDR formulation, which is the combination of traditional model based and deep-learning-based HDR reconstruction results.

(b) **DPHR w/o $H_t$ (i.e., $H = H_d$).** We remove the traditional model based result in this variant, and obtain the final HDR image $H$ using our deep-learning-based result to show the effect of polarimetric information in comparison with the regular images in traditional deep-learning methods.

The HDR formulation with (a) is effective for estimating HDR values. As shown in Fig. 6, compared with the input $L_i$, (b) can restore more details as our CNN uses the polarimetric
Fig. 6: Qualitative results on variants of DPHR. The DPHR model formulates HDR using a combination of the traditional model based method $H_t$ and the deep-learning-based method $H_d$, to estimate HDR in all areas. In contrast to $L_i$ and DPHR w/o $H_t$, DPHR obtains results with better details that are more closely matched to the ground truth. The input mask $M_{1,i}$ identifies dark pixels as invalid regions that require CNN to direct its learning. The Photomatix Enhanced tone-mapping operator is used.

information as a prior. In particular, the CNN itself can outperform the traditional deep-learning approach by Santos et al. [8], where the (b) variant provides better quantitative results than [8], as shown in Table V (seventh row). However, (b) suffers color distortions, and is an overall darker image with limited recovery of details in the poorly exposed regions. On the other hand, (a) can recover better details in both under- and over-exposed regions, and alleviate color distortions and visible artifacts. An explanation for the removal of color distortions in (a) is that DoLP is a mask shared among the colored channels, and thus can enforce consistency between images which helps to remove the colorization artifacts for a better HDR reconstruction. The quantitative results are shown in Table III and is aligned with the observation that (a) achieves a better performance. The key difference between these two variants is the effect of DoLP. In the (a) variant, DoLP is
TABLE III: Quantitative results on variants of DPHR (the higher the better, except for MSE)

|          | PU-PSNR | PU-SSIM | HDR-VDP2 | MSE    |
|----------|---------|---------|----------|--------|
| DPHR w/o $H_t$ | 20.34±1.83 | 0.83±0.044 | 51.13±2.94 | 0.0357±0.0029 |
| DPHR     | 30.59±3.19 | 0.94±0.068 | 56.15±5.28 | 0.0081±0.0006 |

(a) $L_r$ (b) $L_p$ (c) $L_r + L_p$ (d) ground truth

Fig. 7: Qualitative results on different loss functions, where our choice of $L_r + L_p$ loss presents color consistency and preserves details that are better matched with the ground truth. The Photomatix Enhanced tone-mapping operator is used.

TABLE IV: Quantitative results on variants of loss function (the higher the better, except for MSE)

|          | PU-PSNR | PU-SSIM | HDR-VDP2 | MSE    |
|----------|---------|---------|----------|--------|
| $L_r$    | 29.89±2.89 | 0.90±0.034 | 54.15±3.96 | 0.0080±0.0081 |
| $L_p$    | 29.94±2.87 | 0.93±0.023 | 55.35±4.80 | 0.0088±0.0061 |
| $L_r + L_p$ | 30.59±3.19 | 0.94±0.068 | 56.15±5.28 | 0.0081±0.0006 |

used as a strong accurate prior to recover HDR, where for regions with high DoLP the traditional model based method recovers HDR details well compared to the deep-learning-based method. Therefore, we achieve HDR formulation using the (a) variant.

In addition, the results from both (a) and (b) variants can outperform a recent deep-learning-based HDR reconstruction approach developed using the polarization camera [13]. Unlike [13] that uses only the polarization images to achieve HDR reconstruction, and thus neglects the polarimetric cue available from the polarization camera, the (b) variant integrates the polarimetric cue as a prior in the form of soft mask for our feature masking mechanism, to reconstruct better HDR images than [13] as shown in Table IV (fifth row). Then the (a) variant further enhances the reconstruction results of [13] in Table IV by integrating a prior that the traditional model based method $H_t$ is able to recover well for regions with high DoLP to generate the final HDR image.

C. Evaluation of the Training Loss Function

In this experiment, we analyze the effect of each component of our training loss function. The results in Table IV and Fig. 7 shows that $L_r + L_p$ loss is more powerful for preserving details as discussed in [8]. We thus train our model using $L_r + L_p$ loss.

D. Comparison with State-of-the-art Methods

We use the test dataset to evaluate the proposed DPHR method and compare it with state-of-the-art methods, including seven deep-learning-based methods developed for the conventional cameras [6]–[9], [11], [20], [24], one traditional model based method developed for the polarization camera [12], and one deep-learning-based method developed also for the polarization camera [13]. For all methods, we used the codes and pretrained weights provided by the authors. Fig. 8 illustrates the quantitative results of the comparative study. We can observe that images reconstructed with the DPHR method can recover more accurate details in all properly exposed and poorly exposed regions, and alleviate color distortions and visible artifacts. The HDR-GAN [24] and DeepHDR [20] results are relatively bright, and have difficulty restoring details in the over-exposed areas. The ENet [6] result tends to be overly bright and smooth, as it over-enhances the extracted illumination features. Additionally, it’s unable to improve HDR estimation for properly exposed regions. The DrTMO [9] method suffers from blocking artifacts, and has difficulty recovering details in all properly exposed and poorly exposed regions. The DRCP [11] method can restore details in the poorly exposed regions, but exhibits color distortions. Furthermore, the result is generally darker, thus contents lost in the under-exposed regions can not be restored. The PHDR [12] result is overall bright where pixels remain lost in the poorly exposed regions. The DPHR [13] result reveals more details, but remains unable to reconstruct sufficient details in both under- and over-exposed regions.

The HDRCNN [7] method can recover some contents in the over-exposed sky region, but the result is overall dim and presents visual artifacts. The HDRCNN-Mask [8] method follows the approach by HDRCNN to estimate details in over-exposed regions, and then reconstruct the final HDR image by combining with the input. However, unlike HDRCNN, HDRCNN-Mask proposes a feature masking mechanism to propagate valid features for properly exposed pixels. The output of HDRCNN-Mask is overly smooth with color artifacts, and reconstructs limited details of the cloud in the over-exposed sky area. Also, the content remains lost in the under-exposed areas. On the other hand, the proposed DPHR method presents color consistency, free of visible artifacts, and is overall able to recover richer textures in both under- and over-exposed regions. This is because the proposed DPHR method is designed to handle polarization images, as we have integrated the polarization information (i.e., DoLP) into the network during training and inference to formulate the final HDR image. In addition to visual evaluation, the quantitative results are summarized in Table IV Since the proposed DPHR method uses DoLP as a strong accurate prior to help HDR recovery for polarization images, it performs favorably compared to state-of-the-art methods under various evaluation metrics, including both HDR (HDR-VDP2, MSE) and HDR tone-mapped LDR evaluation metrics (PSNR, SSIM, FSIM).

V. LIMITATIONS AND FUTURE WORK

Despite that our method can reconstruct better details compared to existing works, its expressive power is still limited. Fig. 10 shows two examples of difficult scenes where our approach fails to reconstruct. In Fig. 10a our approach is
Fig. 8: Qualitative comparative results with state-of-the-art methods. The proposed DPHR method presents color consistency, free of visible artifacts, and overall able to recover richer textures in all poorly exposed regions. This is because DPHR is designed to handle polarization images, as it integrated the polarimetric cues into the framework. The Photomatix Enhanced TMO is used.

Unable to reconstruct the details and incorrectly reconstructs the color in the underexposed image captured from an indoor poorly-lit setting. This is mainly because our method is unable to handle noisy pixels. In particular, shot noise is prevalent...
when insufficient light reaches the sensor due to an exposure that is too brief for the lighting condition, and it results in random pixel variations which we observe as noisy pixels. This issue can be alleviated by exploring techniques to reduce
### TABLE V: Quantitative comparative results with state-of-the-art methods (the higher the better except for MSE)

| Input sources | Methods       | PU-PSNR | PU-SSIM | HDR-VDP2 | MSE      |
|---------------|---------------|---------|---------|----------|----------|
| Multiple-shot | HDR-GAN [24]  | 15.62 ± 1.92 | 0.48 ± 0.018 | 50.54 ± 4.34 | 0.0599 ± 0.0037 |
|               | DeepHDR [20]  | 18.69 ± 2.04 | 0.60 ± 0.037 | 49.55 ± 5.08 | 0.0293 ± 0.0043 |
| Single-shot   | ENet [6]      | 16.86 ± 1.45 | 0.88 ± 0.049 | 49.45 ± 5.08 | 0.0293 ± 0.0042 |
|               | DrTMO [17]    | 21.51 ± 1.74 | 0.84 ± 0.043 | 53.28 ± 3.28 | 0.0591 ± 0.0027 |
|               | HDRCNN [7]    | 16.48 ± 1.41 | 0.88 ± 0.063 | 49.25 ± 4.64 | 0.0784 ± 0.0062 |
|               | HDRCNN-Mask [8] | 21.82 ± 1.90 | 0.90 ± 0.066 | 53.66 ± 4.77 | 0.0391 ± 0.0037 |
|               | PHDR [12]     | 17.87 ± 1.39 | 0.83 ± 0.059 | 47.72 ± 3.22 | 0.0470 ± 0.0025 |
|               | DPSHDR [13]   | 16.43 ± 1.91 | 0.73 ± 0.067 | 45.92 ± 4.45 | 0.0639 ± 0.0048 |
|               | DPHR          | 20.22 ± 2.13 | 0.83 ± 0.041 | 50.22 ± 4.36 | 0.0787 ± 0.0054 |
|               | PHDR          | 30.59 ± 3.19 | 0.94 ± 0.068 | 56.15 ± 5.28 | 0.0081 ± 0.0006 |

Fig. 10: Limitation analysis. In the top row, the proposed DPHR method fails to reconstruct the indoor image captured from a poorly-lit setting. In the bottom row, the proposed DPHR method also fails to reconstruct the high dynamic-range image captured from an outdoor scene, where the details of the clouds in the sky are not recovered. The Photomatix Enhanced TMO is used.

To follow up on this study, we will enrich our training dataset to improve the performance, we can collect a greater variety of scenes to include other outdoor scenes, such as buildings and people. Additionally, synthesizing HDR images using 3D computer graphics represents a promising avenue we would like to explore. Another interesting avenue of future work is to evaluate the benefits of the polarization camera with other image capturing devices in reconstructing HDR images. We also plan to expand our method to other camera response functions, investigate demosaicing artifacts, and explore the use of the AoLP image, as an additional polarimetric cue, to further the image dynamic range and HDR detail recovery.

### VI. Conclusion

In this paper, we have proposed a novel learning based HDR reconstruction method using a polarization camera. The proposed DPHR method explores polarimetric information obtained by a polarization camera to enhance the reconstruction in all properly exposed and poorly exposed regions, which is a long-standing challenge in computer vision and robotics applications. In the proposed DPHR method, we first compute the input and mask image from the polarization images captured by the polarization camera, and use those cues in the feature masking mechanism to better propagate the valid features through the CNN. Then we combined the deep-learning predicted HDR image with a traditional model based HDR image to formulate the final HDR result. Our experimental results on the polarization image dataset show that the DPHR method demonstrated better qualitative and quantitative performances over state-of-the-art methods.

### References

[1] M. Shakeri and H. Zhang, “Illumination invariant representation of natural images for visual place recognition,” pp. 466–472, 2016.
[2] ——, “Moving object detection in time-lapse or motion trigger image sequences using low-rank and invariant sparse decomposition,” Oct 2017.
[3] ——, “Moving object detection under discontinuous change in illumination using tensor low-rank and invariant sparse decomposition,” June 2019.
[4] Q. Yan, D. Gong, Q. Shi, A. van den Hengel, C. Shen, I. Reid, and Y. Zhang, “Attention-guided network for ghost-free high dynamic range

the pixel noise in the pre-processing step. In Fig. [10b], our approach fails to recover the details of the clouds in the overexposed sky regions from an outdoor high dynamic-range scene. This is mainly due to images with broad regions of overexposed sunshine are scarce in our training dataset. This issue can be mitigated by augmenting our training dataset to incorporate such extreme cases to improve performance.

To follow up on this study, we will enrich our training dataset to improve the performance, we can collect a greater variety of scenes to include other outdoor scenes, such as buildings and people. Additionally, synthesizing HDR images using 3D computer graphics represents a promising avenue we would like to explore. Another interesting avenue of future work is to evaluate the benefits of the polarization camera with other image capturing devices in reconstructing HDR images. We also plan to expand our method to other camera response functions, investigate demosaicing artifacts, and explore the use of the AoLP image, as an additional polarimetric cue, to further the image dynamic range and HDR detail recovery.
[53] “Imx250 cmos sensor,” https://www.sony-semicon.co.jp/e/products/IS/industry/technology/polarization.html, 2021.
[54] T. K. Moon, “The expectation-maximization algorithm,” IEEE Signal Processing Magazine, vol. 13, no. 6, pp. 47–60, 1996.
[55] R. Olaf, F. Philipp, and B. Thomas, “U-net: Convolutional networks for biomedical image segmentation,” in Medical Image Computing and Computer-Assisted Intervention (MICCAI), ser. LNCS, vol. 9351. Springer, 2015, pp. 234–241.
[56] P. K. Diederik and B. Jimmy, “Adam: A method for stochastic optimization,” International Conference on Learning Representations (ICLR), 2015.
[57] 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, vol. 30, no. 4, Jul. 2011.
[58] R. Mantiuk and M. Azimi, “Pu21: A novel perceptually uniform encoding for adapting existing quality metrics for hdr,” 06 2021, pp. 1–5.