

HISTOHDR-NET: HISTOGRAM EQUALIZATION FOR SINGLE LDR TO HDR IMAGE TRANSLATION

Hrishav Bakul Barua^{*®§}, Ganesh Krishnasamy^{*}, KokSheik Wong^{*‡}, Abhinav Dhall[‡], Kalin Stefanov[†]

^{*}School of Information Technology, Monash University Malaysia, Malaysia

[®]Robotics and Autonomous Systems Lab, TCS Research, Kolkata, India

[‡]College of Science and Engineering, Flinders University, Australia

[†]Faculty of Information Technology, Monash University, Australia

{hrishav.barua,ganesh.krishnasamy,wong.koksheik,abhinav.dhall,kalin.stefanov}@monash.edu

ABSTRACT

High Dynamic Range (HDR) imaging aims to replicate the high visual quality and clarity of real-world scenes. Due to the high costs associated with HDR imaging, the literature offers various data-driven methods for HDR image reconstruction from Low Dynamic Range (LDR) counterparts. A common limitation of these approaches is missing details in regions of the reconstructed HDR images, which are over- or under-exposed in the input LDR images. To this end, we propose a simple and effective method, HistoHDR-Net, to recover the fine details (e.g., color, contrast, saturation, and brightness) of HDR images via a fusion-based approach utilizing histogram-equalized LDR images along with self-attention guidance. Our experiments demonstrate the efficacy of the proposed approach over the state-of-art methods.

Index Terms— Data fusion, Self-attention, High dynamic range imaging, Deep learning, Histogram equalization

1. INTRODUCTION

High Dynamic Range (HDR) imaging [1–3] has been a topic of much focus in the vision community for the last several years. Many applications [1, 4] require images as detailed as real-world scenes, which cannot be captured by standard Low Dynamic Range (LDR) cameras. The general camera pipeline [5] captures real-world scenes with a high dynamic range of intensity values and maps them to a low dynamic range by performing quantization, non-linear mapping, and dynamic range clipping. Therefore, the focus in the community has been to reconstruct HDR images from their LDR counterparts using sophisticated data-driven methods that aim to approximate the reverse of the camera pipeline process.

[§]This research is supported by the Global Excellence and Mobility Scholarship (GEMS), Monash University (Australia & Malaysia).

[‡]This research is supported, in part, by the E-Science fund under the project: *Innovative High Dynamic Range Imaging - From Information Hiding to Its Applications* (Grant No. 01-02-10-SF0327).

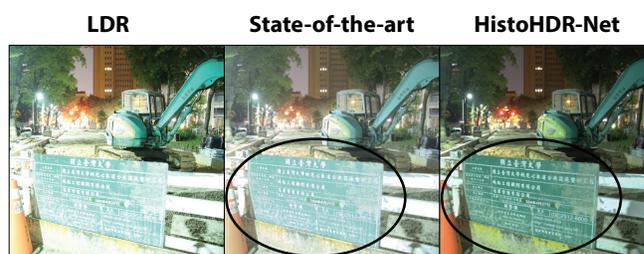


Fig. 1: Our method HistoHDR-Net (right) can recover the text on the notice board better than the state-of-the-art [6] (middle) given an extremely over-exposed LDR image (left) as input.

Most methods proposed in the literature utilize Convolutional Neural Networks and Generative Adversarial Networks [5, 7–13], where a single LDR image is used as input to reconstruct the HDR image. Methods based on single-exposed LDR images often produce unwanted color, contrast, saturation, hue, luminance, brightness, and radiance levels in the resulting HDR images [5, 7–13]. Some methods [14–17] utilize multi-exposed LDR images (generally 2 or 3) as input but often produce many artifacts in the resultant HDR images [14–17]. In contrast, ArtHDR-Net [6] uses features from multi-exposed LDR images to reconstruct the final HDR images without artifacts, but it still lacks naturalness in contrast and saturation. Recent methods employ Neural Radiance Fields [18, 19] drawing inspiration from physics/optics to replicate implicit color and radiance fields involved in the general camera pipeline and enable photo-realistic HDR view synthesis [19, 20]. Dalal *et al.* [21] used Diffusion Models and exposure loss to replicate the reverse of the camera pipeline process through a conditional diffusion architecture with a noise induction process without requiring any classifier. These advanced methods can reconstruct high-quality HDR images with increased clarity of both foreground and background. However, further improvements are still required especially for images with highly contrasting and saturated regions.

There is limited study on the impact of histograms [22] in HDR image reconstruction. Jang *et al.* [23] proposed a method using histogram and color difference [24] between LDR/HDR image pairs to reconstruct HDR images. Other methods explored image fusion techniques using multi-resolution [25] and/or multi-exposed [11] images to enhance contrast levels and remove artifacts in both the HDR and tone-mapped LDR images. Attention-based methods have also been explored [3, 9, 26–28]. Li and Fang [9] proposed a way to tackle the problem of color quantization using a multi-scale Convolutional Neural Network with an attention mechanism. The method is capable of reducing the loss of information in over- and under-exposed regions. In this work, we propose a technique that combines the capabilities of histogram equalization and data fusion to reconstruct high-quality HDR images with better saturation and contrast levels. Fig. 1 depicts the clarity of the images generated by the proposed method compared to the recent state-of-the-art [6]. The contributions of our work, HistoHDR-Net, are:

- We design a simple ResNet50 [29] based pipeline to extract features from ground truth LDR and histogram-equalized LDR images using two parallel ResNet50 blocks and perform fusion on them. We further perform self-attention on the fused feature maps before employing them for HDR image reconstruction (see Sections 2.1 and 2.2).
- We design a novel loss function based on Weber’s law [30], Multi-Scale Structural Similarity Index Measure [31–33], and Color [34, 35] to guide the network for better and reliable HDR image reconstruction (see Section 2.3).
- We perform a thorough analysis of the proposed loss function, different combinations of fusion and attention mechanisms, and feature extraction backbones, to establish the contribution of each of the components (see Section 4).

2. METHOD

The proposed method employs two ResNet50 [29] backbones for feature extraction (encoders), a feature fusion module, a self-attention block [27] for weighing feature relevance, and a fully connected convolutions block (decoder) [6, 36] for HDR image reconstruction. Fig. 2 illustrates the proposed architecture. We also propose a novel loss function that employs Weber’s law [30] based Peak Signal-to-Noise Ratio (PSNR), Color information using the ΔE^* score [34, 35], and Multi-Scale Structural Similarity Index Measure (MS-SSIM) [31–33] to supervise the model in reconstructing reliable, artifact-free, and realistic HDR images with appropriate contrast and saturation levels. To the best of our knowledge Weber’s law is not used as a loss in HDR reconstruction tasks so far.

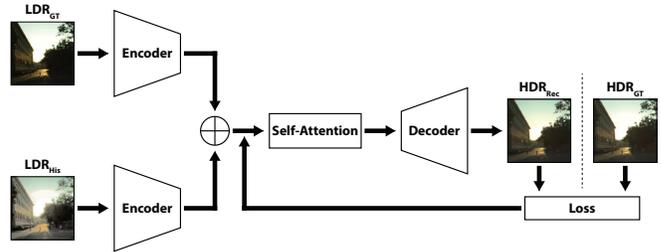


Fig. 2: Architecture of the proposed HistoHDR-Net method.

2.1. Fusion Module

Histogram equalization is a process to improve the contrast and saturation levels. The frequent intensity values of pixels are spread out across 256 bins (*i.e.*, 0 – 255) which leads to intensity range equalization. In the case of images with heavily under- and over-exposed regions, equalization may lead to revealing hidden or non-perceivable areas of the images due to extreme lighting conditions or shadows and darkness. We use OpenCV [37] for performing histogram equalization of LDR images (see Fig. 3 for examples). We denote the original LDR images by LDR_{GT} and the histogram-equalized LDR images by LDR_{His} . The feature maps extracted by the two ResNet50 blocks from LDR_{GT} and LDR_{His} are denoted as f_{LDR}^{GT} and f_{LDR}^{His} , respectively. Then, the output of the fusion module *i.e.*, $f_{fuse}(\cdot)$ can be represented as:

$$f_{fuse}(f_{LDR}^{GT}, f_{LDR}^{His}) = f_{LDR}^{GT} \oplus f_{LDR}^{His}, \quad (1)$$

where \oplus represents the concatenation of feature maps.

2.2. Self-Attention Module

The fused feature maps are passed through a simple self-attention mechanism [27] with the intention to further improve the efficacy of the method. This process increases the receptive field by allowing the new value of a pixel to be influenced by all other pixels in the image. However, the main characteristic of the self-attention module is to add weights to the different features in the extracted feature maps. Henceforth the feature importance is considered for adjusting the amount of influence it will have on the HDR image reconstruction. We represent the output of the self-attention module as $f_{att}(\cdot)$. Hence, the output of this module can be represented as $f_{att}(f_{fuse}(f_{LDR}^{GT}, f_{LDR}^{His}))$. The output of the self-attention module is added to the output of the fusion module to produce the self-attention feature map. This information is used by the decoder block to reconstruct the HDR image from a single-exposed LDR image.

2.3. Loss Functions

All loss functions use tone-mapped versions of the reconstructed HDR (HDR_{Rec}^{TM}) and ground truth HDR (HDR_{GT}^{TM}) im-

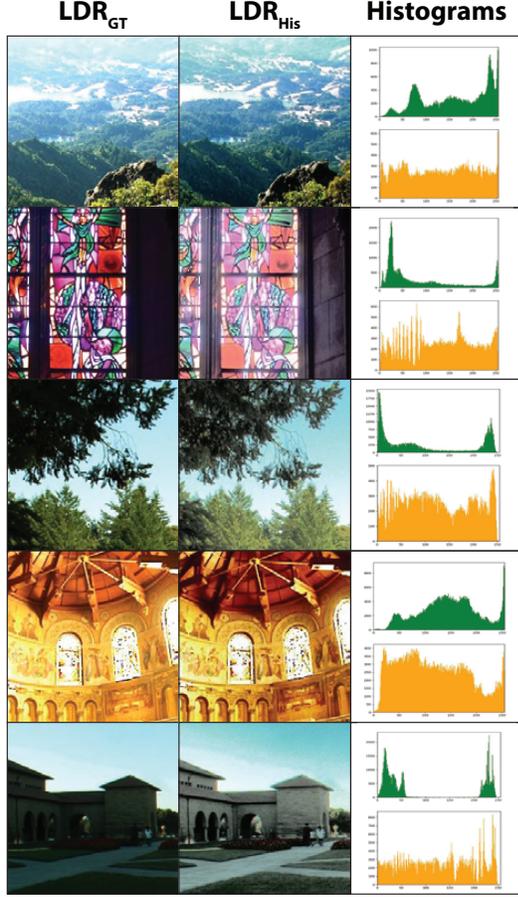


Fig. 3: Original and histogram-equalized LDR images. The green and orange histograms are for the original LDR_{GT} and histogram-equalized LDR_{His} images, respectively.

ages. Tone-mapping is based on the μ -law [38] and mitigates the influence of high intensity pixel values in the HDR images. The tone-mapping operator $T(\cdot)$ is formulated as:

$$T(\text{HDR}) = \frac{\log(1 + \mu\text{HDR})}{\log(1 + \mu)}, \quad (2)$$

where HDR^{TM} is the output of the tone-mapping operation $T(\text{HDR})$ and μ specifies the amount of compression, which is set to 5000 following [36].

We designed a novel loss function consisting of five components. The L1 (\mathcal{L}_{L1}) and Perceptual (\mathcal{L}_{VGG}) losses are commonly used in LDR to HDR reconstruction methods. Apart from these losses, we introduce a loss based on Weber's law (\mathcal{L}_W), MS-SSIM metric (\mathcal{L}_{SIM}), and Color information (\mathcal{L}_C) using the ΔE^* score. The final loss is calculated as:

$$\mathcal{L} = \alpha\mathcal{L}_{L1} + \beta\mathcal{L}_{VGG} + \delta\mathcal{L}_W + \gamma\mathcal{L}_{SIM} + \lambda\mathcal{L}_C, \quad (3)$$

where α , β , δ , γ , and λ are scaling factors which are empirically set to 0.18, 0.5, 0.82, 0.80, and 0.82, respectively.

To facilitate the presentation, we denote $\text{HDR}_{GT}^{\text{TM}}$ by X and $\text{HDR}_{\text{Rec}}^{\text{TM}}$ by Y in all loss functions. The \mathcal{L}_{L1} loss is the Mean Absolute Error between the pixels of X and Y :

$$\mathcal{L}_{L1} = \frac{1}{n} \sum_{t=1}^n \|X_t - Y_t\|, \quad (4)$$

where n (in all loss functions) is the total number of images.

The \mathcal{L}_{VGG} loss is also calculated between X and Y and uses $f_{VGG}(\cdot)$ [39] and a pre-trained VGG19 [40] model:

$$\mathcal{L}_{VGG} = \frac{1}{n} \sum_{t=1}^n f_{VGG}(X_t, Y_t). \quad (5)$$

The \mathcal{L}_W loss uses Weber's law [30] of perceptual image meaningfulness, *i.e.*, the background light influences the human's perception and response toward the intensity fluctuation of visual signals. The Weber's law based PSNR_W induces the psychology of the human vision into the perceptual image quality evaluation and is defined as:

$$\text{PSNR}_W = 10 \log_{10} \frac{(2^{\text{bitDepth}} - 1)^2}{\frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} w_{i,j}^2 (X_{i,j} - Y_{i,j})^2}, \quad (6)$$

where $w_{i,j}$ is calculated with $0.02 \times (2^{\text{bitDepth}} - X_{i,j})$ using the Weber's law and 0.02 is set using Weber's fraction [30]. M and N are image width and height in terms of pixel. The bitDepth is the depth of the pixel intensity values in the tone-mapped HDR images, in our case $\text{bitDepth} = 8$. Then, \mathcal{L}_W is defined as:

$$\mathcal{L}_W = \frac{1}{n} \sum_{t=1}^n \frac{1}{\text{PSNR}_{W_t}}. \quad (7)$$

The \mathcal{L}_{SIM} loss is derived from the MS-SSIM metric for image comparison. It operates on multiple scales using a multi-stage process of image sub-sampling. It is a better representation of structural similarity of two images than simple SSIM. The MS-SSIM metric is calculated as:

$$\text{MS-SSIM} = l_M^\eta \prod_{k=1}^M cs_k^{\tau_k}, \quad (8)$$

where l_M^η and $cs_k^{\tau_k}$ are defined below, M is scales of operation, and η and τ_k are set to 1 in our implementation. The luminance comparison between X and Y is calculated as:

$$l = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1}, \quad (9)$$

and the comparison of the contrast and the structure between X and Y is jointly calculated as:

$$cs = \frac{2\sigma_{XY} + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2}. \quad (10)$$

Table 1: Within-dataset performance of the proposed HistoHDR-Net and state-of-the-art. The best results are in **bold** and the second best are underlined.

Method	City Scene [41]			HDR-Synth & HDR-Real [5]		
	PSNR \uparrow	SSIM \uparrow	HDR-VDP-2 \uparrow	PSNR \uparrow	SSIM \uparrow	HDR-VDP-2 \uparrow
FHDR [36]	32.51	0.90	67.23	17.11	0.71	66.72
SingleHDR [5]	33.42	0.91	68.22	26.33	0.85	68.21
ArtHDR-Net [6]	35.12	<u>0.93</u>	<u>69.31</u>	33.45	0.88	<u>68.37</u>
Diffusion-based [21]	36.07	<u>0.93</u>	68.51	33.52	<u>0.90</u>	68.21
HistoHDR-Net	<u>35.14</u>	0.94	69.32	<u>33.48</u>	0.91	69.24

Table 2: Cross-dataset evaluation of the proposed HistoHDR-Net and state-of-the-art on unseen dataset (HDR-Eye [26]). The best results are in **bold** and the second best underlined.

Method	PSNR \uparrow	SSIM \uparrow	HDR-VDP-2 \uparrow
FHDR [36]	33.07	0.90	67.88
SingleHDR [5]	33.78	0.91	68.51
ArtHDR-Net [6]	35.81	0.94	<u>69.13</u>
Diffusion-based [21]	36.46	<u>0.95</u>	69.01
HistoHDR-Net	<u>36.21</u>	0.96	69.38

μ_X and μ_Y are the pixel sample means, σ_X^2 and σ_Y^2 are the variances, and σ_{XY} is the covariance. The variables $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$ are used to gain stability in the case of small denominator division, where K_1 and K_2 are set to 0.01 and 0.03, respectively, as per the original metric definition. $L = 2^{\text{bitDepth}} - 1$ is the representation of the dynamic range of the image pixels. Then, the \mathcal{L}_{SIM} loss is defined as:

$$\mathcal{L}_{SIM} = \frac{1}{n} \sum_{t=1}^n (1 - \text{MS-SSIM}_t). \quad (11)$$

The \mathcal{L}_C loss is calculated on the basis of the ΔE^* score and is defined as:

$$\mathcal{L}_C = \frac{1}{n} \sum_{t=1}^n \frac{\sqrt{\sum_{m=1}^P (X_m - Y_m)^2}}{2^{\text{bitDepth}} - 1}, \quad (12)$$

where P is the total number of pixels in the images. X_m and Y_m are the color intensities of pixel m in the corresponding image. $\sum_{m=1}^P (\cdot)$ calculates the sum of the squared differences along the color channels of the images.

3. EXPERIMENTS

Implementation. We trained the proposed method on an Ubuntu 20.04.6 LTS workstation with Intel® Xeon® CPU E5-2687W v3 @ 3.10GHz (20 CPU cores), 126 GB RAM (+ 2 GB swap memory), NVIDIA GeForce GTX 1080 GPU (having 8 GB memory), and 1.4 TB SSD. The model was trained for 200 epochs. We used a batch size of 10 and Adam

optimizer [42]. The learning rate was set to $1e-4$, i.e., 0.0001 at first and adjusted to decay later on.

Datasets. We used a mix of real and synthetic datasets to evaluate the proposed method. The City Scene dataset [41] includes 20K LDR/HDR image pairs with diverse scenes from indoor and outdoor environments, objects, and buildings. HDR-Synth & HDR-Real dataset [5] contains real LDR/HDR image pairs (9785 in total) and synthetic pairs (around 500). All images have been pre-processed to 512×512 resolution. These datasets were split into 80% train and 20% test splits and used in the within-dataset performance evaluation. We also consider a third dataset, HDR-Eye [26] having 46 real HDR images with resolution 512×512 for cross-dataset evaluation. For all ablation studies, we selected a subset of 5000 random images from City Scene and HDR-Synth & HDR-Real datasets.

Evaluation. We use High Dynamic Range Visual Differences Predictor (HDR-VDP-2) [43, 44] for human level perceptual image similarity comparison. Structural similarity of the images is measured with Structural Similarity Index Measure (SSIM) [31, 32]. We also use the common Peak Signal-to-Noise Ratio in dB (PSNR) [45] metric for pixel level similarity. We calculate the HDR-VDP-2 and SSIM scores on the ground truth and reconstructed HDR images in the linear domain. The PSNR value is obtained on the μ -law based tone-mapped ground truth and reconstructed HDR images. We evaluate our method and compare it with four different state-of-the-art methods including, FHDR [36], SingleHDR [5], Diffusion-based [21] and ArtHDR-Net [6].

4. RESULTS AND ABLATIONS

Quantitative Results. Table 1 summarizes the within-dataset quantitative results. We see that the proposed method outperforms all of the selected state-of-the-art methods in terms of SSIM and HDR-VDP-2. However, the diffusion-based model [21] outperforms our model in terms of PSNR score on both datasets. Table 2 reports the results for cross-dataset evaluation. The observed trend is similar to the within-dataset evaluation - our method outperforms all of the selected state-of-the-art methods in terms of SSIM and HDR-VDP-2 and second best in terms of PSNR.

Qualitative Results. Fig. 4 illustrates that the visual quality of our method is the closest to the ground truth HDR images in terms of color, contrast, saturation and brightness. We can also see the foreground and background objects with clarity. We display the HDR_{GT} and HDR_{Rec} from all the methods using Reinhard’s tone-mapping algorithm [46].

Architecture Ablation. We studied the impact of different fusion techniques (i.e., element-wise addition and concatenation) of the original and histogram-equalized LDR images including, early fusion (input data), late fusion (encoder feature maps), and attention mechanism. In the case of element-wise addition, we distort the information in the original LDR

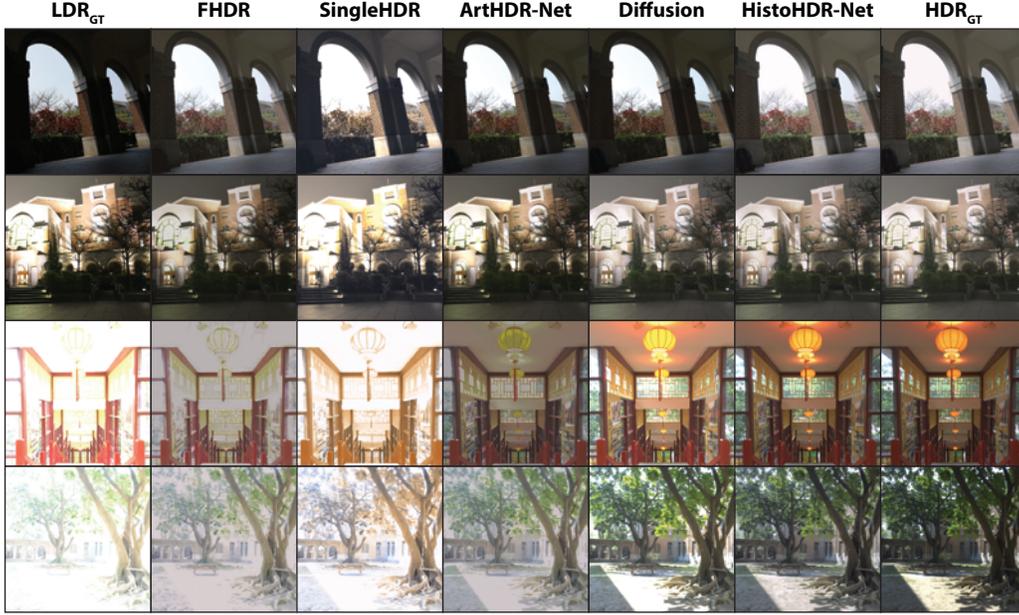


Fig. 4: HDR images generated by the proposed HistoHDR-Net and state-of-the-art methods.

Table 3: Architecture ablation results for different data fusion and attention approaches. ‘+’ denotes element-wise addition and ‘ \oplus ’ denotes concatenation. The best results are in **bold**.

Variant	PSNR \uparrow	SSIM \uparrow	HDR-VDP-2 \uparrow
LDR _{GT} -only	24.40	0.79	64.21
LDR _{His} -only	24.11	0.82	65.02
LDR _{GT} + LDR _{His}	26.38	0.85	65.69
LDR _{GT} \oplus LDR _{His}	28.11	0.87	66.12
$f_{LDR}^{GT} + f_{LDR}^{His}$	30.19	0.88	67.74
$f_{LDR}^{GT} \oplus f_{LDR}^{His}$	31.43	0.90	68.01
$f_{att}(f_{LDR}^{GT} + f_{LDR}^{His})$	33.57	0.91	68.11
$f_{att}(f_{LDR}^{GT} \oplus f_{LDR}^{His})$	35.30	0.93	69.31

with the histogram-equalized LDR images. In contrast, concatenation allows for information sharing between the original LDR and equalized LDR images. The results in Table 3 demonstrate how the fusion techniques and attention mechanism improve the performance of the model compared to the baseline implementation with only LDR images or equalized LDR images as input. The first and second rows depict the cases without fusion and attention. We see LDR_{His} is slightly better than LDR_{GT}. The third and fourth rows display the early fusion results, which are slightly better than the previous case. Histogram-equalized LDR images lead to a significant improvement when used in late fusion (fifth and sixth rows). When self-attention is introduced to the late fusion outputs (seventh and eighth rows), we see further improvements.

Loss Ablation. The proposed loss function has five components out of which the first two are commonly used in image reconstruction tasks. Table 4 shows the contribution of each

Table 4: Loss ablation results for different loss components. The best results are in **bold**.

Loss components	PSNR \uparrow	SSIM \uparrow	HDR-VDP-2 \uparrow
\mathcal{L}_{L1}	33.43	0.912	67.83
$\mathcal{L}_{L1}, \mathcal{L}_W$	34.22	0.917	68.11
$\mathcal{L}_{L1}, \mathcal{L}_{VGG}$	33.72	0.915	67.91
$\mathcal{L}_{L1}, \mathcal{L}_{VGG}, \mathcal{L}_W$	34.58	0.922	68.22
$\mathcal{L}_{L1}, \mathcal{L}_{VGG}, \mathcal{L}_W, \mathcal{L}_{SIM}$	34.98	0.927	68.85
$\mathcal{L}_{L1}, \mathcal{L}_{VGG}, \mathcal{L}_W, \mathcal{L}_{SIM}, \mathcal{L}_C$	35.30	0.930	69.31

Table 5: Encoder ablation results for different feature extraction backbones. The best results are in **bold**.

Backbone	PSNR \uparrow	SSIM \uparrow	HDR-VDP-2 \uparrow
MobileNet [47]	32.45	0.895	67.01
InceptionV3 [48]	33.10	0.899	67.68
VGG19 [40]	35.41	0.927	69.23
ResNet50 [29]	35.30	0.930	69.31

of the loss components. We see that \mathcal{L}_{L1} and \mathcal{L}_{VGG} achieve moderate scores and the actual improvement is achieved with the addition of \mathcal{L}_W , \mathcal{L}_{SIM} , and \mathcal{L}_C (fourth, fifth, and sixth rows). We do not use L2 loss because it is too sensitive to noise [33]. We also perform a separate test on the combination of \mathcal{L}_{L1} and \mathcal{L}_W (second row). We observe that the PSNR, SSIM and HDR-VDR-2 scores are better than the combination of \mathcal{L}_{L1} and \mathcal{L}_{VGG} by a margin of 1.5%, 0.2%, and 0.3%, respectively. Since, both \mathcal{L}_{VGG} and \mathcal{L}_W are perceptual losses, we can consider \mathcal{L}_W as an alternative to \mathcal{L}_{VGG} in HDR image reconstruction tasks.

Encoder Ablation. We also study different choices of the feature extraction block which is important for HDR image

reconstruction. We examine the most common feature extraction backbones including, MobileNet [47], InceptionV3 [48], ResNet50 [29], and VGG19 [40]. The results are summarized in Table 5. We see that ResNet50 attains the best results in terms of SSIM and HDR-VDP-2. However, VGG19 reaches better performance in the case of PSNR.

5. CONCLUSION

Histogram-equalized LDR images play a significant role in improving the performance of Convolutional Neural Network based models for HDR image reconstruction. Data fusion and self-attention mechanisms further improve the performance. The proposed loss function that employs Weber’s law, MS-SSIM metric, and ΔE^* color difference further improves the reconstructed HDR images resulting in a state-of-the-art performance. Future work includes the study of histogram matching [24] that allows us to control the shape of the histogram. The prospects of applying similar approach with video HDR reconstruction using extensions such as video transformers or timesformers can be studied in future works.

6. REFERENCES

- [1] Lin Wang and Kuk-Jin Yoon, “Deep learning for hdr imaging: State-of-the-art and future trends,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 12, pp. 8874–8895, 2021.
- [2] Xueyu Han, Ishtiaq Rasool Khan, and Susanto Rahardja, “High dynamic range image tone mapping: Literature review and performance benchmark,” *Digital Signal Processing*, p. 104015, 2023.
- [3] Rufeng Chen, Bolun Zheng, Hua Zhang, Quan Chen, Chenggang Yan, Gregory Slabaugh, and Shanxin Yuan, “Improving dynamic hdr imaging with fusion transformer,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023, vol. 37, pp. 340–349.
- [4] Xiangyu Chen, Zhengwen Zhang, Jimmy S Ren, Lynhoo Tian, Yu Qiao, and Chao Dong, “A new journey from sdrtv to hdrtv,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 4500–4509.
- [5] Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang, “Single-image hdr reconstruction by learning to reverse the camera pipeline,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1651–1660.
- [6] Hrishav Bakul Barua, Ganesh Krishnasamy, KokSheik Wong, Kalin Stefanov, and Abhinav Dhall, “Arthdr-net: Perceptually realistic and accurate hdr content creation,” in *2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 2023, pp. 806–812.
- [7] Marcel Santana Santos, Tsang Ing Ren, and Nima Khademi Kalantari, “Single image hdr reconstruction using a cnn with masked features and perceptual loss,” *ACM Transactions on Graphics (TOG)*, vol. 39, no. 4, pp. 80–1, 2020.
- [8] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K Mantiuk, and Jonas Unger, “Hdr image reconstruction from a single exposure using deep cnns,” *ACM transactions on graphics (TOG)*, vol. 36, no. 6, pp. 1–15, 2017.
- [9] Jinghui Li and Peiyu Fang, “Hdrnet: Single-image-based hdr reconstruction using channel attention cnn,” in *Proceedings of the 2019 4th International Conference on Multimedia Systems and Signal Processing*, 2019, pp. 119–124.
- [10] Phuoc-Hieu Le, Quynh Le, Rang Nguyen, and Binh-Son Hua, “Single-image hdr reconstruction by multi-exposure generation,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 4063–4072.
- [11] B-C Guo and C-H Lin, “Single-image hdr reconstruction based on two-stage gan structure,” in *2023 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2023, pp. 91–95.
- [12] Yunhao Zou, Chenggang Yan, and Ying Fu, “Rawhdr: High dynamic range image reconstruction from a single raw image,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 12334–12344.
- [13] Cheng Guo and Xiuhua Jiang, “Lhdr: Hdr reconstruction for legacy content using a lightweight dnn,” in *Proceedings of the Asian Conference on Computer Vision (ACCV)*, 2022, pp. 3155–3171.
- [14] Yuzhen Niu, Jianbin Wu, Wenxi Liu, Wenzhong Guo, and Rynson WH Lau, “Hdr-gan: Hdr image reconstruction from multi-exposed ldr images with large motions,” *IEEE Transactions on Image Processing*, vol. 30, pp. 3885–3896, 2021.
- [15] Jianrui Cai, Shuhang Gu, and Lei Zhang, “Learning a deep single image contrast enhancer from multi-exposure images,” *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 2049–2062, 2018.
- [16] Haoyu Ren, Yi Fan, and Stephen Huang, “Robust real-world image enhancement based on multi-exposure ldr images,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 1715–1723.
- [17] Zhilu Zhang, Haoyu Wang, Shuai Liu, Xiaotao Wang, Lei Lei, and Wangmeng Zuo, “Self-supervised high dynamic range imaging with multi-exposure images in dynamic scenes,” in *ICLR*, 2024.
- [18] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng, “Nerf: Representing scenes as neural radiance fields for view synthesis,” *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.
- [19] Xin Huang, Qi Zhang, Ying Feng, Hongdong Li, Xuan Wang, and Qing Wang, “Hdr-nerf: High dynamic range neural radiance fields,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 18398–18408.
- [20] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P Srinivasan, and Jonathan T Barron, “Nerf in the dark: High dynamic range view synthesis from noisy raw images,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 16190–16199.

- [21] Dwip Dalal, Gautam Vashishtha, Prajwal Singh, and Shanmuganathan Raman, "Single image ldr to hdr conversion using conditional diffusion," in *2023 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2023, pp. 3533–3537.
- [22] Shaoliang Yang, Dongming Zhou, Jinde Cao, and Yanbu Guo, "Lightingnet: An integrated learning method for low-light image enhancement," *IEEE Transactions on Computational Imaging*, vol. 9, pp. 29–42, 2023.
- [23] Hanbyol Jang, Kihun Bang, Jinseong Jang, and Dosik Hwang, "Dynamic range expansion using cumulative histogram learning for high dynamic range image generation," *IEEE Access*, vol. 8, pp. 38554–38567, 2020.
- [24] Dinu Coltuc, Philippe Bolon, and J-M Chassery, "Exact histogram specification," *IEEE Transactions on Image processing*, vol. 15, no. 5, pp. 1143–1152, 2006.
- [25] Xueyu Han and Susanto Rahardja, "High dynamic range image tone mapping based on layer decomposition and image fusion," in *2023 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2023, pp. 221–225.
- [26] Hiromi Nemoto, Pavel Korshunov, Philippe Hanhart, and Touradj Ebrahimi, "Visual attention in ldr and hdr images," in *9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, 2015, number CONF.
- [27] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena, "Self-attention generative adversarial networks," in *International conference on machine learning*. PMLR, 2019, pp. 7354–7363.
- [28] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang, "Attention-guided network for ghost-free high dynamic range imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1751–1760.
- [29] Seungjun Shin, Kyeongbo Kong, and Woo-Jin Song, "Cnn-based ldr-to-hdr conversion system," in *2018 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2018, pp. 1–2.
- [30] Jie Jia and Hae-Kwang Kim, "Weber's law based psnr for perceptual image quality evaluation," .
- [31] Zhou Wang and Alan C Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE signal processing magazine*, vol. 26, no. 1, pp. 98–117, 2009.
- [32] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [33] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on computational imaging*, vol. 3, no. 1, pp. 47–57, 2016.
- [34] Gaurav Sharma, Wencheng Wu, and Edul N Dalal, "The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," *Color Research & Application*, vol. 30, no. 1, pp. 21–30, 2005.
- [35] David H Brainard et al., "Color appearance and color difference specification," *The science of color*, vol. 2, no. 191–216, pp. 5, 2003.
- [36] Zeeshan Khan, Mukul Khanna, and Shanmuganathan Raman, "Fhdr: Hdr image reconstruction from a single ldr image using feedback network," in *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2019, pp. 1–5.
- [37] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.
- [38] Takao Jinno, Hironori Kaida, Xinwei Xue, Nicola Adami, and Masahiro Okuda, " μ -law based hdr coding and its error analysis," *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 94, no. 3, pp. 972–978, 2011.
- [39] Justin Johson, Alexandre Alahi, and Li Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," Tech. Rep., Springer, 2016.
- [40] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [41] Jinsong Zhang and Jean-François Lalonde, "Learning high dynamic range from outdoor panoramas," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 4519–4528.
- [42] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [43] Rafal Mantiuk, Kil Joong Kim, Allan G Rempel, and Wolfgang Heidrich, "Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Transactions on graphics (TOG)*, vol. 30, no. 4, pp. 1–14, 2011.
- [44] Manish Narwaria, Rafal K Mantiuk, Mattheiu Perreira Da Silva, and Patrick Le Callet, "Hdr-vdp-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images," *Journal of Electronic Imaging*, vol. 24, no. 1, pp. 010501–010501, 2015.
- [45] Prateek Gupta, Priyanka Srivastava, Satyam Bhardwaj, and Vikrant Bhateja, "A modified psnr metric based on hvs for quality assessment of color images," in *2011 International Conference on Communication and Industrial Application*. IEEE, 2011, pp. 1–4.
- [46] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda, "Photographic tone reproduction for digital images," in *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pp. 661–670. 2023.
- [47] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
- [48] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.