LHDR: HDR Reconstruction for Legacy Content using a Lightweight DNN, Supplementary Material

Cheng Guo[1,2][0000−0002−2660−2267] and Xiuhua Jiang[1,2]

[1] State Key Laboratory of Media Convergence and Communication, Communication University of China
[2] Peng Cheng Laboratory, Shenzhen, China
{guocheng, jiangxiuhua}@cuc.edu.cn

Abstract. This supplementary material provides additional content and discussion to complement the main manuscript. First, we explain how the pipeline model is derived into problem modeling. Second, we introduce implementation detail. Third, we provide more quantitative and visual results and analysis.

Keywords: High dynamic range · Legacy Content · Degradation model.

1 Deriving Problem Modeling from Pipeline Model

We take [1][3] as the prototype camera pipeline model, as Figure1. We will discuss if a specific step is related to SI-HDR, if so, further analyze its mathematical operations to show whether it (and its reverse operation) is global or local.

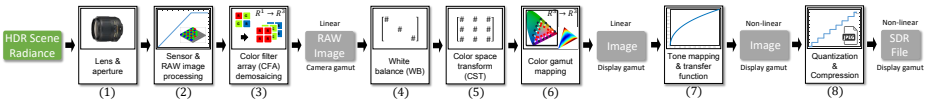


Fig. 1: Prototype camera pipeline model [1] consisting of 8 steps. Note that some operations e.g. white balance and exposure adjustment could be practically conducted in a different order.

Detailed operations in this precise pipeline are:

1. First, an optical system containing lens and aperture will ideally maintain the full range of scene radiance (green box in Figure1), what HDR images dedicate to faithfully record.

[3] Newest version: https://www.eecs.yorku.ca/~mbrown/ICCV2019_Brown.html.

2. Then, digital signal will be produced by sensor and its RAW processing: ISO gain, noise reduction, defective pixel mask, black light subtraction, linearization, lens flat-filed correction, exposure adjustment, etc. All those operations are to ensure the signal is numerically and spatially linear to scene radiance, which is the foundation of subsequent color manipulation. Camera noise and dynamic range truncation (leading to over&under-exposure) are sequentially introduced here due to sensor imperfection. Such degradations are severer for legacy imaging devices.
3. Mosaiced achromatic response produced by color filter array (CFA) will be demosaiced to RGB values in camera gamut primaries. SI-HDR has nothing to do with this $\mathbb{R}^1 \rightarrow \mathbb{R}^3$ mapping since HDR images are already demosaiced.
4. Step (4) is white balance, and could usually be ignored in SI-HDR too since most HDR images are already white balanced[4]. Rigorously, if not, the RAW response will be white balanced by applying a 3×3 diagonal matrix:

$$\mathbf{E}_{WB} = \mathbf{M}_{WB}\mathbf{E} \tag{1}$$

   where $\mathbf{E} = [R, G, B]^T, \mathbf{M}_{WB} = diag(k_R, k_G, k_B)$. Obviously, matrix transformation and its reverse operation are global where the result on single pixel is not affected by its neighbors.
5. Usually, SDR image is assumed in display gamut e.g. sRGB/BT.709, while HDR is in camera gamut RGB[5]. Hence, color space transform (CST) will be conducted to convert between different color spaces using a 3×3 matrix:

$$\mathbf{E}_{disp.\ gamut} = \mathbf{M}\mathbf{E}_{WB} \tag{2}$$

   where $M \in \mathbb{R}^{3\times3}$. It also belongs to global operation, similar as step (4).
6. Most display gamut is smaller than camera gamut, thus some RGB value in will fall outside the valid range after CST, i.e. $\exists\ \mathbf{E}_{disp.\ gamut} \notin [0, 1]$. The simplest and most common way is clamping those out-of-gamut (OOG) pixels to the boundary: i.e. hard-clipping [4]:

$$\mathbf{E}_{disp.\ gamut} = clamp(\mathbf{E}_{disp.\ gamut}, 0, 1) \tag{3}$$

   but this involves multiple-to-one $\mathbb{R}^3 \rightarrow \mathbb{R}^3$ mapping thus its reverse operation in SI-HDR is no longer global. Fortunately, in natural scenes, the transition from non-OOG to OOG pixels is usually spatially continuous, therefore it's easy for DNN to infer OOG value from its neighbor. Reverting gamut mapping will become a sheer global operation only when gamut soft-mapping[6] i.e. completely one-to-one mapping is applied.

[4] Real-world off-the-shelf HDR images are almost generated by the multi-exposure fusion(MEF) of RAW images. The latter are white-balanced in most cases.

[5] For example, HDR images in [2,3] i.e. our training set is in specified camera RAW RGB primaries. Steps (5) and (6) can be ignored in SI-HDR only when assuming HDR and SDR are in same RGB primaries (gamut).

[6] In Figure1(6), we show an example replacement map of gamut soft-mapping, darker color indicates longer distance from its original position in xy chromaticity diagram.

7. Then, non-linearity is globally added to each pixel. Though there is a standardized gamma2.2 or BT.1886 [5] opto-electronic transfer function (OETF), the SDR non-linearity is usually the combination of specific camera response function (CRF) and possible aesthetic tone mapping (curve adjustment).
8. Finally, 8-bit quantization and compression (usually JPEG) are applied to get a distribution-ready file. Multiple artifacts are introduced here: quantization artifact and blocking artifact by the 8×8 JPEG quantization block, etc. Recovering those artifacts involves the help of adjacent local pattern.

Ground truth (GT) HDR image could be treated as the scene radiance to be simulated shot i.e. the start of the camera pipeline, assuming it has successfully recorded the full luminance range. In this case, the camera pipeline is equivalent to the HDR-to-SDR degradation, hence SI-HDR need to recover all degradation introduced there. As analyzed above, pipeline steps (1) and (3) are unrelated to SI-HDR; both (4), (5), (7) and their reverse ops are global; and reverting (2), (6) and (8) belongs to local operation. Finally, those degradations/operations are summarized into 6-step problem modeling by excluding unrelated operations. The remainder of derivation can be found in main paper.

2 Implementation Detail

2.1 DNN Structure

In Figure3 of main paper, 'k3s1n32' stands for a 2D convolutional layer with 3×3 kernel, stride = 1, number of filters (nf) = 32, and group number = 4 if 'g4' is appended, and 'FC' means fully-connected layer.

In the large-scale branch of local network, we use 'PixelShuffle' [6] upsampling at decoder end, and 'SqEx' (squeeze-and-excitation) block [7] at the bottleneck to endow the DNN with channel attention on intermediate deep feature.

2.2 Degradations to Simulated Legacy SDR

Camera noise is already contained in SDR images from NTIRE [8] dataset. For noise-free SDR images in Fairchild [2] dataset, we first linearize the normalized nonlinear input SDR ($\mathbf{x}'$), and then simulated it to camera RAW gamut (RGB primaries) before adding noise:

$$\mathbf{x}_{cam.\ gamut} = \mathbf{M}(\mathbf{x}'^{1/0.45}),\ \mathbf{x} = \begin{bmatrix} R \\ G \\ B \end{bmatrix},\ \mathbf{M} = \begin{bmatrix} 0.6313 & 0.2708 & 0.0979 \\ 0.0368 & 0.7931 & 0.1701 \\ 0.0174 & 0.1488 & 0.8338 \end{bmatrix} \tag{4}$$

Here, matrix $\mathbf{M}$ determines what camera gamut to be converted to. Since the RAW gamut of real cameras largely diversifies, for simplicity, we assume a fixed one i.e. Arri ALEXA Wide Gamut. Then, noise described in main paper is independently added to each RGB channel of $\mathbf{x}_{cam.\ gamut}$. After this, we convert linear RGB in camera gamut back to nonlinear sRGB using:

$$\mathbf{x}'_{w.\ n.} = (\mathbf{M}^{-1}\mathbf{x}_{cam.\ gamut\ w.\ n.})^{0.45} \tag{5}$$

According to the pipeline model, compression is added lastly on noise-affected sRGB image $\mathbf{x}'_{w.\ n.}$. The first JPEG compression with QF $\sim U(60, 80)$ is applied to the whole SDR image, and the second JPEG compression with fixed 75 QF is applied when cropping training patches: For each HDR-SDR pair in training set, we first rescale then with $\times 0.5$, $\times 0.75$ and $\times 1$ factor, respectively. We randomly crop 1(for NTIRE [8]) or 5(for Fairchild [2]) patch-pair(s) sized 600×600 at each scale, and then save with fixed 75 QF JPEG compression. This 'rescale-and-random-patch' double JPEG could better simulate the multiple internet transmission than current approach e.g. staggered double JPEG on same scale [9].

2.3 Training Detail

First, we pre-process label HDR since original HDR images from NTIRE [8] and Fairchild [2] dataset are aligned differently: HDR images in [8] were normalized and transferred into gamma2.2 nonlinear before 16-bit .png storage, while some HDR images from [2] linearly record absolute luminance in .exr encapsulation. They have to be aligned consistently for DNN training. Hence, we linearize all HDR images from [8], and normalize those from [2] according to their max value.

Then, on each HDR-SDR patch-pair sized 600×600, a smaller patch sized 200×200 is again cropped at different random positions every time the training dataloader is called. Augmentation including orthogonal rotation and flipping are applied, and we set batch size = 8. Parameters of adaptive moment estimation (AdaM) optimizer are $\beta_1 = 0.9,\ \beta_2 = 0.999$. The total number of iterations is set to 1.2×10^6, and it takes about 3 days to finish the training on the desktop computer with i7-4790k CPU and GTX1080 GPU.

3 Experiment Configuration and Result

3.1 Experiment Configuration

Among all competitors, HDRCNN [10] provide 2 extra checkpoints, we used the one trained with JPEG compressed SDR. Also, due to their training set, the output HDR from HDRUNet [11] are stored in gamma2.2 nonlinearity, hence we linearized them for a fair comparison. Meanwhile, since HDRCNN/DHDR [12] use UNet (encoder-decoder) of 6/8 levels, a $\mathbb{R}^{3\times 1080\times 1920}$ input is unacceptable since its height or width could not be divided by $2^{6-1}/2^{8-1}$. Hence, we zero-pad input SDR during their inference phase and delete those padded pixels as the final result. Also, when counting their MACs, we choose input $\mathbb{R}^{3\times 1088/1024\times 1920}$ with a similar number of elements for HDRCNN/DHDR. Finally, result HDR ($\bar{\mathbf{y}}$) from all competitors is normalized according to their maximum value.

Note that metric VDP [13] require HDR images in absolute luminance, we therefore assume a peak luminance of $1000cd/m^2$ for all normalized output HDR ($\bar{\mathbf{y}}$) and GT HDR ($\mathbf{y}$), i.e.

$$\bar{\mathbf{y}}_{lum.} = 1000 \times \frac{\bar{\mathbf{y}}}{max(\bar{\mathbf{y}})},\ \mathbf{y}_{lum.} = 1000 \times \frac{\mathbf{y}}{max(\mathbf{y})} \tag{6}$$

were used to compute VDP. Also, we set its 'viewing condition' as 'side-by-side', $0.5m$ from 24" 1920×1080 monitor ($32.45pixel/°$).

3.2 Ablation Studies Configuration

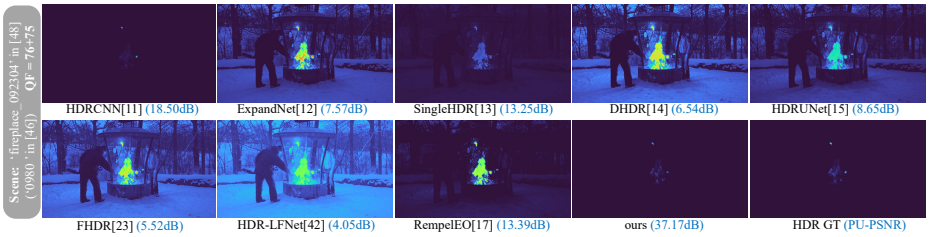
When testing the impact of conventional degradations, we are supposed to remove all camera noise for training SDR. Therefore, for NTIRE [8] dataset where noise is already contained in SDR, we managed to find their original HDR frames from the footage of HdM-HDR [3] dataset, and degraded them to clean version using same HDR-to-SDR degradation as in [14].

When examining the effect of HDR-exclusive degradation (over and under-exposure), we changed the original training set [2, 8] to HDR-LFNet [15]. For this dataset, we ultilized 2480 pairs of 600×600 patches.

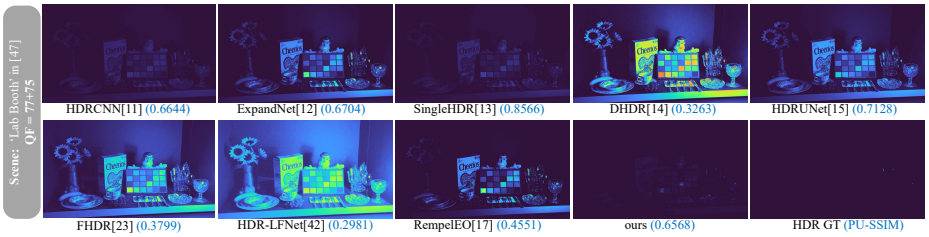
3.3 More Visual Comparison

Since the pixel energy/value of normalized linear HDR image is more concentrated in lower part, HDR images will appear dim if directly visualized. To simulate the HDR viewing on conventional SDR display, we tone-map all GT&result HDR images by MATLAB function `localtonemap()` with default parameters.

We provide more visual comparison for the following reason: Better metrics only represent method's capability of generating HDR with closer luminance distribution with GT, could not yet manifest its degradation (over&under-exposure,



(a) As seen, higher metrics mainly represent closer estimated luminance with GT, meaning a method is better for image-based lighting (IBL) application.



(b) Our method performs normally under this scene, however, still gets a plausible score. This means pixel value exerts an undue impact even on structure-related SSIM.

Fig. 2: Recovered luminance of outdoor and indoor scenes.

noise, and compression) recovery ability. This phenomenon was proven in main paper, [16] and [17], and further confirmed by Figure2.

Visual comparisons are provided in Figure3-5 with analysis on their title. Take Figure3(red arrow) and Figure4(green arrow) for example, while other DNNs are able to recover at least a few lost information, the output over-exposed area from ExpandNet [18] and HDR-LFNet [15] do not share much difference with SDR. The cause for ExpandNet is its 'Trad. TMO' 'degradation' model with insufficient degradation ability. For HDR-LFNet, it's because their DNN just polishes the result of traditional expansion operators (EOs), from where the lost information was never recovered.

Since our method is designed to additionally tackle legacy SDR, we start assess the capability of removing noise and compression artifact. As far as denoising is concerned, HDRUNet [11] and ours are better since they're trained so. In Figure3(green arrow), 4(red arrow) and 5, SingleHDR outperforms other methods not trained to denoise and decompression, because their degradation model actually contains noise and compression, but to a lesser degree[7] than ours.

So far, conclusion can be drawn that: (1) Better performance of SingleHDR, HDRUNet and our method on noise-and-compression-affected area confirm the importance of conventional degradation on joint-task SI-HDR, and (2) the lack of information recovery ability of ExpandNet further proves the significance of HDR-exclusive degradation on universal SI-HDR.

References

1. Karaimer, H.C., Brown, M.S.: A software platform for manipulating the camera imaging pipeline. In: Proc. ECCV. (2016) 429–444
2. Fairchild, M.D.: The hdr photographic survey. In: Color and imaging conference. Volume 2007. (2007) 233–238
3. Froehlich, J., Grandinetti, S., et al.: Creating cinematic wide gamut hdr-video for the evaluation of tone mapping operators and hdr-displays. Digital photography X **9023** (2014) 279–288
4. ITU: Report ITU-R BT.2407-0: Colour gamut conversion from Recommendation ITU-R BT.2020 to Recommendation ITU-R BT.709. (2017)
5. ITU: Recommendation ITU-R BT.1886: Reference electro-optical transfer function for flat panel displays used in HDTV studio production. (2011)
6. Shi, W., Caballero, J., Huszár, F., et al.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proc. CVPR. (2016) 1874–1883
7. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proc. CVPR. (2018) 7132–7141
8. Pérez-Pellitero, E., et al.: Ntire 2021 challenge on high dynamic range imaging: Dataset, methods and results. In: Proc. CVPR. (2021) 691–700

[7] Their: {JPEG QF $\sim U(85, 100)$, Poisson-Gaussian noise w. $\sigma_p \sim U(0, 0.0013) + \sigma_g \sim U(0, 0.0005)$}, ours: {JPEG QF $\sim U(60, 80) + 75$, noise with $\sigma \sim U(0.001, 0.003)$}. We did not consider SingleHDR as joint-denoise&decompression since (1) the extent of degradation is small, and (2) the author didn't hold such motive.

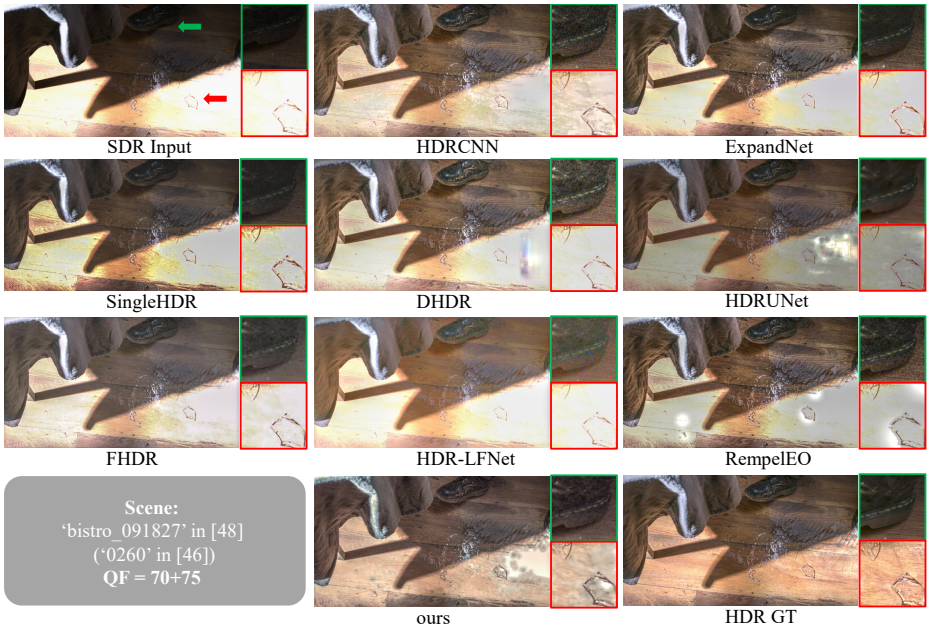


Fig. 3: When recovering large area of over-exposure (red arrow), ours and HDR-CNN [10] are relatively better, DHDR [12] and HDRUNet [11] produce strange pattern, while the rest methods almost fail to hallucinate lost information.

9. Jiang, J., Zhang, K., Timofte, R.: Towards flexible blind jpeg artifacts removal. In: Proc. ICCV. (2021) 4997–5006
10. Eilertsen, G., Kronander, J., et al.: Hdr image reconstruction from a single exposure using deep cnns. ACM Trans. Graph. **36** (2017) 1–15
11. Chen, X., Liu, Y., et al.: Hdrunet: Single image hdr reconstruction with denoising and dequantization. In: Proc. CVPR. (2021) 354–363
12. Santos, M.S., Ren, T.I., Kalantari, N.K.: Single image hdr reconstruction using a cnn with masked features and perceptual loss. ACM Trans. Graph. **39** (2020) 80–1
13. Wolski, K., Giunchi, D., et al.: Dataset and metrics for predicting local visible differences. ACM Trans. Graph. **37** (2018) 1–14
14. Kalantari, N.K., Ramamoorthi, R.: Deep hdr video from sequences with alternating exposures. In: Comput. graph. Forum. Volume 38. (2019) 193–205
15. Chambe, M., Kijak, E., et al.: Hdr-lfnet: Inverse tone mapping using fusion network. hal preprint:03618267 (2022)
16. Eilertsen, G., Hajisharif, S., et al.: How to cheat with metrics in single-image hdr reconstruction. In: Proc. ICCV. (2021) 3998–4007
17. Hanji, P., Mantiuk, R., Eilertsen, G., Hajisharif, S., Unger, J.: Comparison of single image hdr reconstruction methodsthe caveats of quality assessment. In: Proc. SIGGRAPH '22. (2022) 1–8
18. Marnerides, D., Bashford-Rogers, T., et al.: Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. Comput. Graph. Forum **37** (2018) 37–49

Fig. 4: Methods' performance on over-exposure (green arrow) consistent with Figure3, while only ours and HDRUNet [11] suppress the noise and compression in dark area (red arrow).

Fig. 5: When it comes to severer noise and compression, all methods fail. Ours is sightly better in that artifact is 'smoothed', while HDRUNet [11] tend to overprocess.