Supplemental Document on
Exposure-Aware Dynamic Weighted Learning for Single-Shot HDR Imaging

An Gia Vien[0000−0003−0067−0285] and Chul Lee[0000−0001−9329−7365]

Department of Multimedia Engineering, Dongguk University, Seoul, Korea
viengiaan@mme.dongguk.edu, chullee@dongguk.edu

1 Network Details

1.1 DFN

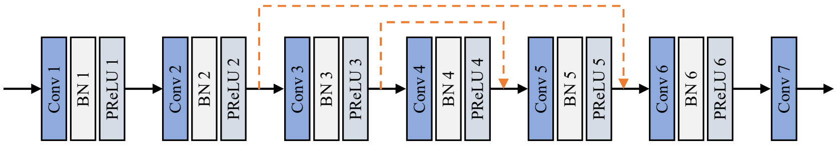

Fig. S-1. Architecture of DFN in DINet and FusionNet.

Fig. S-1 shows the architecture of DFN in DINet and FusionNet in Fig. 2 in the main paper. It consists of seven convolutional layers with symmetric skip connections. All convolutional layers have 3×3 kernels. The numbers of feature channels of the first six convolutional layers are 32, and that of the last convolutional layer (Conv7) depends on the number of input channels, *i.e.*, 36 for DINet and 18 for FusionNet. Each convolutional layer is followed by a batch normalization (BN) [3] and PReLU activation function, except for Conv7.

1.2 Weight Learning Network

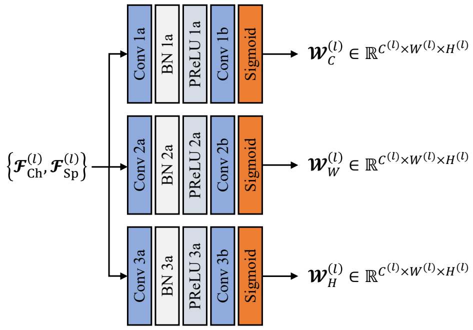
$\{\boldsymbol{\mathcal{F}}_{\mathrm{Ch}}^{(l)}, \boldsymbol{\mathcal{F}}_{\mathrm{Sp}}^{(l)}\}$
Conv 1a
BN 1a
PReLU 1a
Conv 1b
Sigmoid
$\boldsymbol{\mathcal{W}}_C^{(l)} \in \mathbb{R}^{C^{(l)} \times W^{(l)} \times H^{(l)}}$
Conv 2a
BN 2a
PReLU 2a
Conv 2b
Sigmoid
$\boldsymbol{\mathcal{W}}_W^{(l)} \in \mathbb{R}^{C^{(l)} \times W^{(l)} \times H^{(l)}}$
Conv 3a
BN 3a
PReLU 3a
Conv 3b
Sigmoid
$\boldsymbol{\mathcal{W}}_H^{(l)} \in \mathbb{R}^{C^{(l)} \times W^{(l)} \times H^{(l)}}$

Fig. S-2. Architecture of the weight learning network in the MEF block.

Fig. S-2 shows the architecture of the weight learning network in the MEF block in Fig. 3 in the main paper. It consists of three branches to generate three weight maps $\boldsymbol{\mathcal{W}}_C^{(l)}, \boldsymbol{\mathcal{W}}_W^{(l)}$, and $\boldsymbol{\mathcal{W}}_H^{(l)}$, respectively. Each branch has two convolutional layers. The first convolutional layer is followed by a BN and PReLU activation function. The sigmoid activation function is used to generate weight values in the range of $[0, 1]$. All convolutional layers have 3×3 kernels with the same number of feature channels at each scale l, *i.e.*, $C^{(1)} = 128$, $C^{(2)} = 256$, and $C^{(3)} = 512$.

1.3 Discriminator Network D

Table S-1. Architecture of the discriminator network D.

Layer	# of channels	Kernel size	Stride	Padding size
Conv1_0	128	3 × 3	1 × 1	1 × 1
SN1_0	–	–	–	–
LeakyReLU ($\alpha = 0.2$)	–	–	–	–
Residual block	128	–	–	–
Conv2_0	256	2 × 2	2 × 2	–
SN2_0	–	–	–	–
LeakyReLU ($\alpha = 0.2$)	–	–	–	–
Residual block	256	–	–	–
Conv3_0	512	2 × 2	2 × 2	–
SN3_0	–	–	–	–
LeakyReLU ($\alpha = 0.2$)	–	–	–	–
Conv3_1	512	3 × 3	1 × 1	1 × 1
SN3_1	–	–	–	–
LeakyReLU ($\alpha = 0.2$)	–	–	–	–
Conv3_2	512	3 × 3	1 × 1	1 × 1
SN3_2	–	–	–	–
LeakyReLU ($\alpha = 0.2$)	–	–	–	–
Conv3_3	512	3 × 3	1 × 1	1 × 1
SN3_3	–	–	–	–
LeakyReLU ($\alpha = 0.2$)	–	–	–	–
Global average pooling	512	–	–	–
Conv4_0	1	1 × 1	1 × 1	–
SN4_0	–	–	–	–
Tanh	–	–	–	–

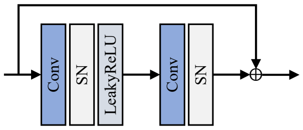

Fig. S-3. Architecture of the residual block in the discriminator network D.

Table S-1 lists the details of the discriminator network D to compute the adversarial loss $\mathcal{L}_{\text{Adv}}$ in (18) in the main paper. We employ spectral normalization (SN) [6] after each convolutional layer to stabilize the training. We use residual blocks [2] in Figure S-3 after Conv1_0 and Conv2_0 in Table S-1 to increase the receptive field. All convolutional layers in the residual block have 3 × 3 kernels.

2 Datasets

2.1 HDM-HDR[1]

Fig. S-4 shows 12 randomly selected frames from the HDM-HDR dataset, which are used for the test.

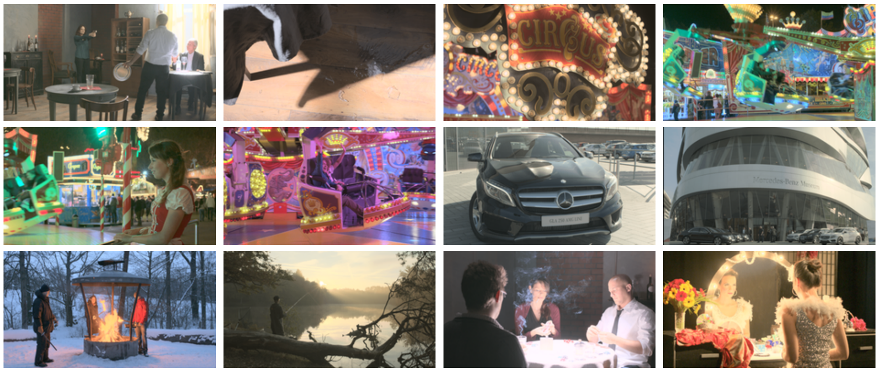

Fig. S-4. HDR frame selected from the HDM-HDR dataset for the test.

2.2 HDRv [5]

Fig. S-5 shows 16 randomly selected frames from the HDRv dataset, which are used for the test.

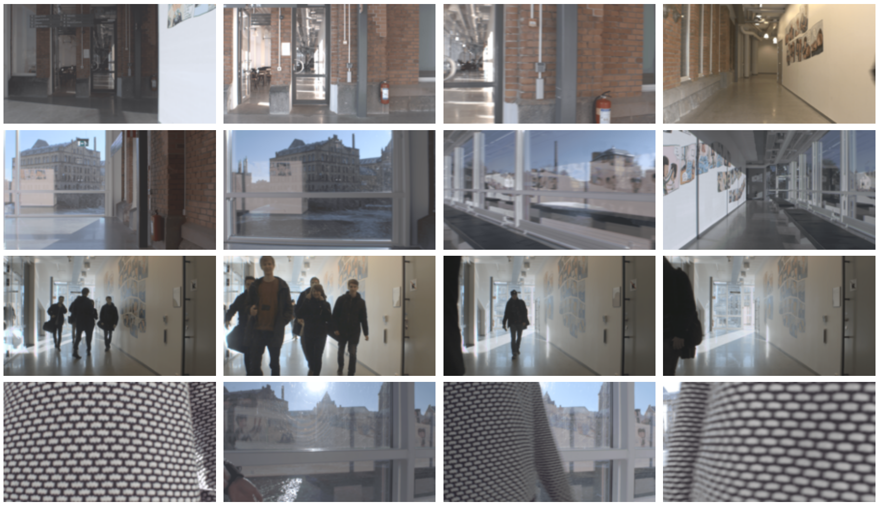

Fig. S-5. HDR images selected from the HDRv dataset for the test.

[1] https://www.hdm-stuttgart.de/vmlab/hdm-hdr-2014

3 More Experimental Results

We provide more comparative results on the Kalantari's [4], HDM-HDR,[2] HDR-Eye,[3] and HDRv [5] datasets.

3.1 Kalantari's Dataset

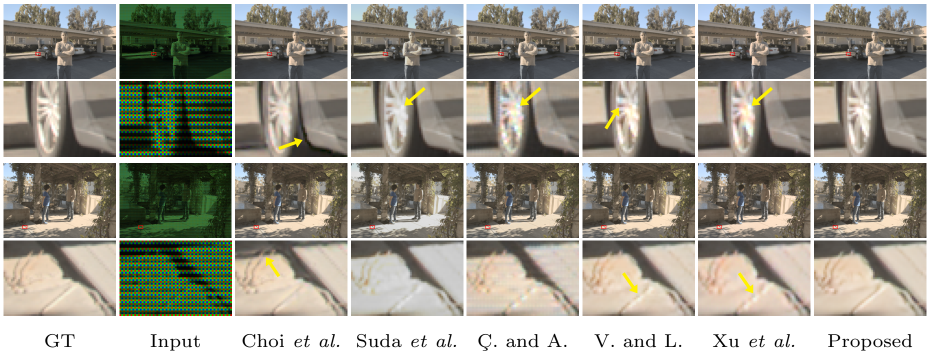

Fig. S-6. Qualitative comparison of synthesized HDR images and their magnified parts on the Kalantari's dataset.

3.2 HDM-HDR

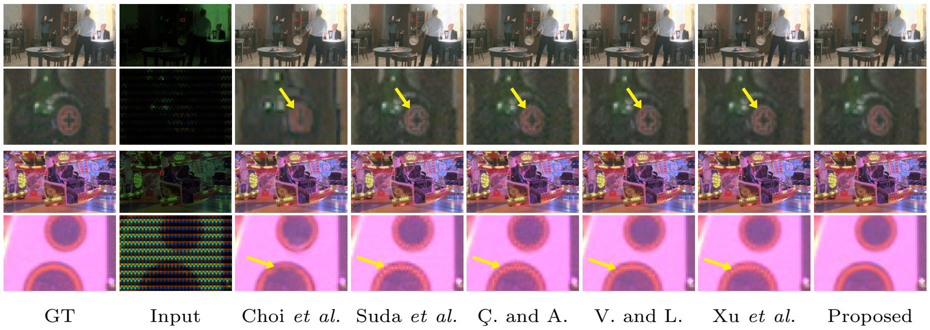

Fig. S-7. Qualitative comparison of synthesized HDR images and their magnified parts on HDM-HDR.

[2] https://www.hdm-stuttgart.de/vmlab/hdm-hdr-2014
[3] https://mmspg.epfl.ch/hdr-eye

3.3 HDR-Eye

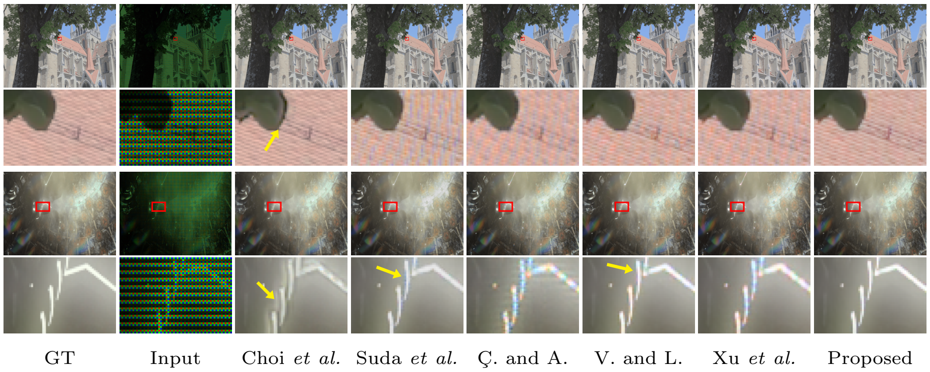

Fig. S-8. Qualitative comparison of synthesized HDR images and their magnified parts on HDR-Eye.

3.4 HDRv

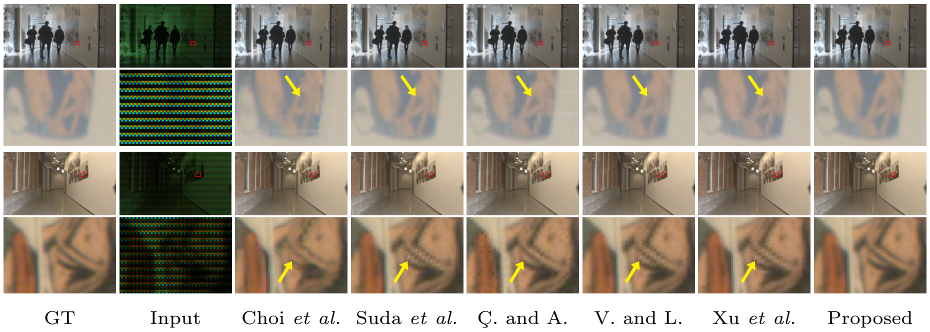

Fig. S-9. Qualitative comparison of synthesized HDR images and their magnified parts on HDRv.

4 DRIM Assessments

We also provide the distortion maps using the dynamic range independent quality metric (DRIM) [1] for the test images. DRIM estimates the probability of the differences between two images in each local region being noticed by viewers. DRIM provides three types of differences: loss of visible contrast (green), amplification of invisible contrast (blue), and reversal of visible contrast (red). In Figs. S-10–S-13, the proposed algorithm yields significantly less differences in all the changes than the conventional algorithms.

4.1 Kalantari's Dataset

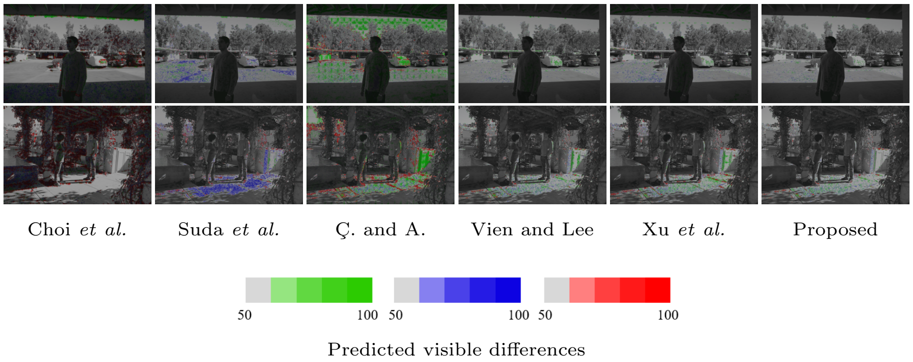

Fig. S-10. DRIM assessment on the Kalantari's dataset.

4.2 HDM-HDR

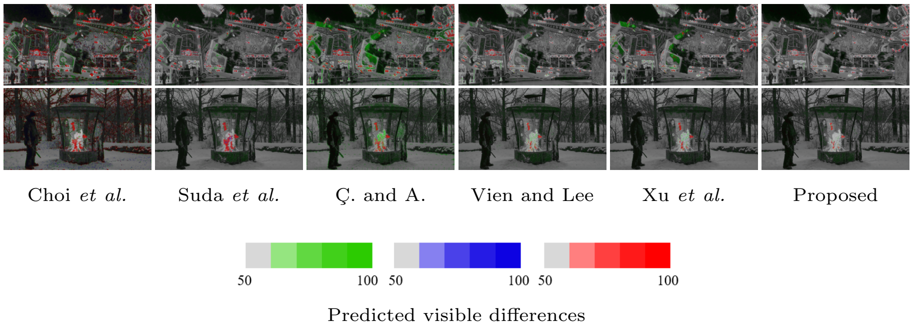

Fig. S-11. DRIM assessment on HDM-HDR.

4.3 HDR-Eye

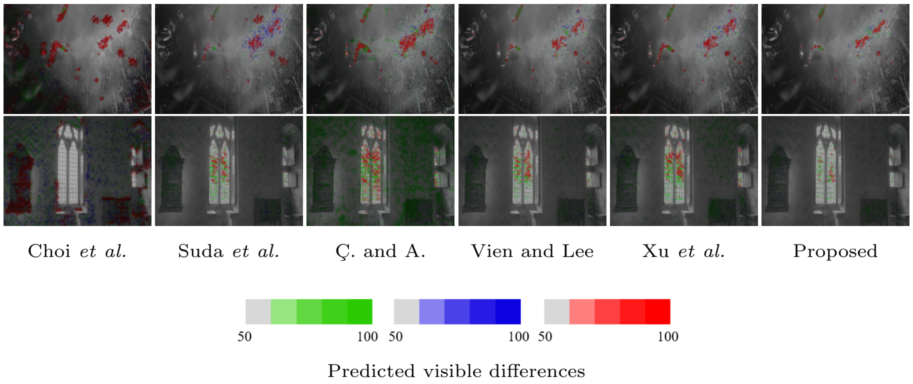

Fig. S-12. DRIM assessment on HDR-Eye.

4.4 HDRv

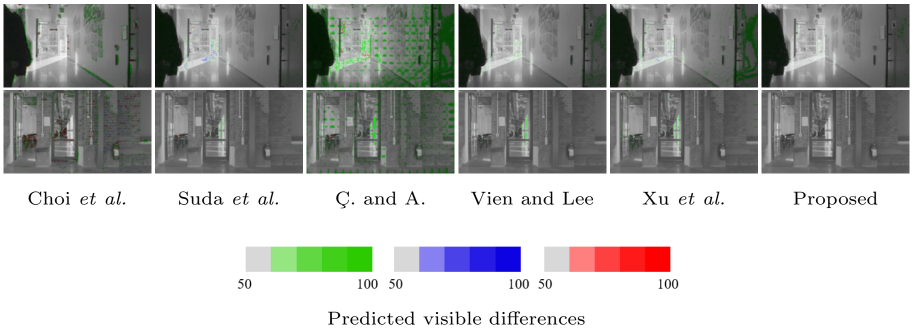

Fig. S-13. DRIM assessment on HDRv.

5 More Analyses

5.1 Necessity of Subnetworks

Table S-2. Effectiveness of multi-domain learning on the synthesis performance.

Model	pu-MSSSIM	pu-PSNR	log-PSNR	HDR-VDP		HDR-VQM
				Q	P	
U-Net	0.9963	45.48	42.44	73.46	0.4242	0.9692
U-Net with **M**	0.9963	45.62	42.38	73.57	0.4097	0.9692
Proposed	**0.9969**	**46.17**	**43.04**	**74.03**	**0.3889**	**0.9718**

Because of the diversity in shapes and sizes of poorly exposed regions, a single network may be ineffective in recovering those regions, as experimentally observed and described in the main paper. We address this issue by using multi-domain learning and combining the outputs of the two subnetworks having complementary information.

Table S-2 demonstrates the necessity of multi-domain learning. A single U-Net exhibits the worst performance, and the addition of **M** fed into U-Net improves the performance. The proposed algorithm with multi-domain learning outperforms the single U-Net with large margins, which confirms the effectiveness of the proposed network architecture.

5.2 Short- and Long-Exposures

Table S-3. Effectiveness of different EV spacing on the synthesis performance.

EV	Avg. DR	pu-MSSSIM	pu-PSNR	log-PSNR	HDR-VDP		HDR-VQM
					Q	P	
$\{-1, +1\}$	226.3	**0.9969**	**46.17**	**43.04**	**74.03**	**0.3889**	**0.9718**
$\{-2, +2\}$	230.3	0.9956	45.10	42.44	73.53	0.4344	0.9710
$\{-3, +3\}$	238.8	0.9917	42.41	40.65	71.73	0.5551	0.9697

The proposed algorithm used the EV spacing $\{-1, +1\}$, because it has been commonly used in single-shot HDR imaging [4, 44, 50]. To further analyze the effects of different short- and long-exposures on the synthesis performance, we evaluate the proposed algorithm with different EV spacing. Table S-3 shows that as the EV spacing increases, higher dynamic range (DR) can be recovered while the synthesis performance decreases.

5.3 Necessity of Filter Learning in DFN

Table S-4. Effectiveness of filter learning in DFN on the restoration performance.

	pu-MSSSIM	pu-PSNR	log-PSNR
Direct learning	0.9957	43.37	43.65
Filter learning (Proposed)	**0.9969**	**44.84**	**45.11**

Filter learning with DFNs can better exploit the information of missing regions in an SVE image during both interpolation in DINet and fusion in FusionNet, providing spatially consistent results, than direct image learning. Table S-4 compares the performance of DINet for different learning strategies.

5.4 Necessity of Multi-scale MEF

Table S-5. Effectiveness of multi-scale MEF on the restoration performance.

	pu-MSSSIM	pu-PSNR	log-PSNR
Single MEF	0.9962	46.55	46.68
Multi-scale MEF (Proposed)	**0.9965**	**46.96**	**47.01**

If an input image contains large missing regions, a single MEF block only in the first layer is ineffective to extract good features because of small receptive fields, degrading the synthesis performance. Table S-5 shows that multi-scale MEF yields higher scores than a single MEF.

5.5 Demosaicing

Table S-6. Impacts of the network architecture design.

	pu-MSSSIM	pu-PSNR	log-PSNR	HDR-VDP		HDR-VQM
				Q	*P*	
Vien [44]	0.9964	45.10	42.22	73.72	0.3930	0.9696
Xu [50]	0.9957	44.62	42.01	73.00	0.5593	0.9700
Proposed	**0.9969**	**46.17**	**43.04**	**74.03**	**0.3889**	**0.9718**
Proposed*	0.9965	45.80	42.64	73.79	0.4100	0.9705

Table S-6 includes the performance of the proposed algorithm, where FusionNet directly outputs the demosaiced image, denoted by Proposed*. It is worth pointing out that Proposed* outperforms the two best-performing conventional algorithms, which output demosaiced images. This confirms that the choice of demosaicing algorithm affects the performance as we stated in the main paper.

5.6 Execution Times

Table S-7. Analysis of execution time of each component in the proposed algorithm.

	DINet	ExRNet	FusionNet	Demosaicing	Total
GPU	21.56	63.47	23.25	104.17	212.45
CPU	30.17	334.59	30.07	302.56	697.39

Table S-8. Execution time comparison of the proposed algorithm with the conventional algorithms.

	Choi [5]	Suda [40]	Çoğalan [6]	Vien [44]	Xu [50]	Proposed
GPU	-	74.44	20.64	20.86	38.36	212.45
CPU	741.46	210.88	41.36	172.95	92.86	697.39

Tables S-7 and S-8 compare the average execution times in seconds of the proposed algorithm to process the test images in the Kalantari's dataset. Note that about half of execution time is consumed by demosaicing.

References

1. Aydın, T.O., Mantiuk, R., Myszkowski, K., Seidel, H.P.: Dynamic range independent image quality assessment. ACM Trans. Graph. **27**(3), 69:1–10 (Aug 2008) 7
2. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proc. CVPR. pp. 770–778 (Jun 2016) 3
3. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Proc. ICML. pp. 448–456 (Jul 2015) 1
4. Kalantari, N.K., Ramamoorthi, R.: Deep high dynamic range imaging of dynamic scenes. ACM Trans. Graph. **36**(4), 144:1–144:12 (Jul 2017) 5
5. Kronander, J., Gustavson, S., Bonnet, G., Unger, J.: Unified HDR reconstruction from raw CFA data. In: Proc. ICCP. pp. 1–9 (Apr 2013) 4, 5
6. Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y.: Spectral normalization for generative adversarial networks. In: Proc. ICLR (Apr 2018) 3