Alignment-free HDR Deghosting with Semantics Consistent Transformer
- Supplementary Material -

Steven Tel[1,5*] Zongwei Wu[2,5*] Yulun Zhang[3 †] Barthélémy Heyrman[1]
Cédric Demonceaux[4,5] Radu Timofte[2] Dominique Ginhac[5]
[1] University of Burgundy, ImViA [2] Computer Vision Lab, CAIDAS & IFI, University of Würzburg
[3] CVL, ETH Zürich [4] University of Lorraine, CNRS, Inria, Loria [5] University of Burgundy, CNRS, ICB
{steven.tel; barthelemy.heyrman; cedric.demonceaux; dginhac}@u-bourgogne.fr, {zongwei.wu; radu.timofte}@uni-wuerzburg.de, yulun100@gmail.com

Abstract

In the supplementary material, we provide in Section 1 the ablation study on the hyperparameters, in Section 2 a detailed study on the existing benchmark, and in Section 3 the generalization capability with qualitative comparisons, and in Section 4 the full quantitative comparisons on our dataset.

1. Ablation Studies on the Number of Layers

In this section, we conduct studies to analyze the influence of hyperparameters, *i.e.*, numbers of attention blocks, on the HDR deghosting performance. We have three hyperparameters $(N_L; N_G; N_S)$, as shown in Figure 2 of the manuscript. N_L stands for the number of attention layers, N_G stands for the number of global spatial attention blocks, and N_S stands for the number of semantic-consistent attention blocks. The quantitative results under different settings can be found in Table 1. We can conclude that our method performs well with a large variation of hyperparameters.

Table 1. Ablation study on the hyperparameters. The comparison is conducted on the Kalantari *et al*. dataset [2].

N_L	N_G	N_S	Mb	μ-PSNR	l-PSNR	μ-SSIM	l-SSIM
4	6	4	29	**44.49**	42.29	**0.9924**	**0.9887**
4	7	4	31	43.59	42.25	0.9912	0.9885
4	5	4	21	43.85	42.21	0.9912	0.9883
4	4	4	21	43.81	42.08	0.9913	0.9886
4	6	2	28	43.82	**42.30**	0.9915	0.9882
4	6	5	30	43.76	42.18	0.9913	0.9880
4	5	5	23	43.57	41.09	0.9910	0.9879
5	6	4	37	43.91	42.05	0.9911	0.9881
3	6	4	23	43.23	41.43	0.9910	0.9878
2	6	4	15	43.62	41.42	0.9909	0.9874

2. Study on Kalantari *et al*. [2] Testing Samples

In this section, we first conduct a detailed analysis of the current benchmark Kalantari *et al*. In Figure 2 we plot the general performance by averaging the l-PSNR and μ-PSNR scores of SOTA methods [4, 3], including ours, on each Kalantari's testing sample. It can be seen that for samples 7 to 10, HDR deghosting networks produce low linear l-PSNR values (in blue color). However, after tone mapping, the μ-PSNR (in orange color) becomes significantly higher, leading to a large but abnormal gap (in red color) between these two metrics. We think that this phenomenon may be majorly coupled with the presence of over-exposed/underexposed regions in the predicted images or ground truth.

Table 2. Proportion of over-exposed pixels in Kalantari *et al*. [2] testing samples.

Sample ID	Num. over-exposed pixels	Proportion %
1	35	< 1
2	5	< 1
3	1060	< 1
4	10793	< 1
5	251	< 1
6	5685	< 1
7	187506	12.50
8	111788	7.45
9	213696	14.25
10	458368	30.56
11	14035	< 1
12	1060	< 1
13	1	< 1
14	251	< 1
15	5644	< 1

To validate our initial guess, we evaluate the number of over-exposed pixels in the corresponding ground truth. In

*Both authors contributed equally to this research.

†Corresponding Author: Yulun Zhang

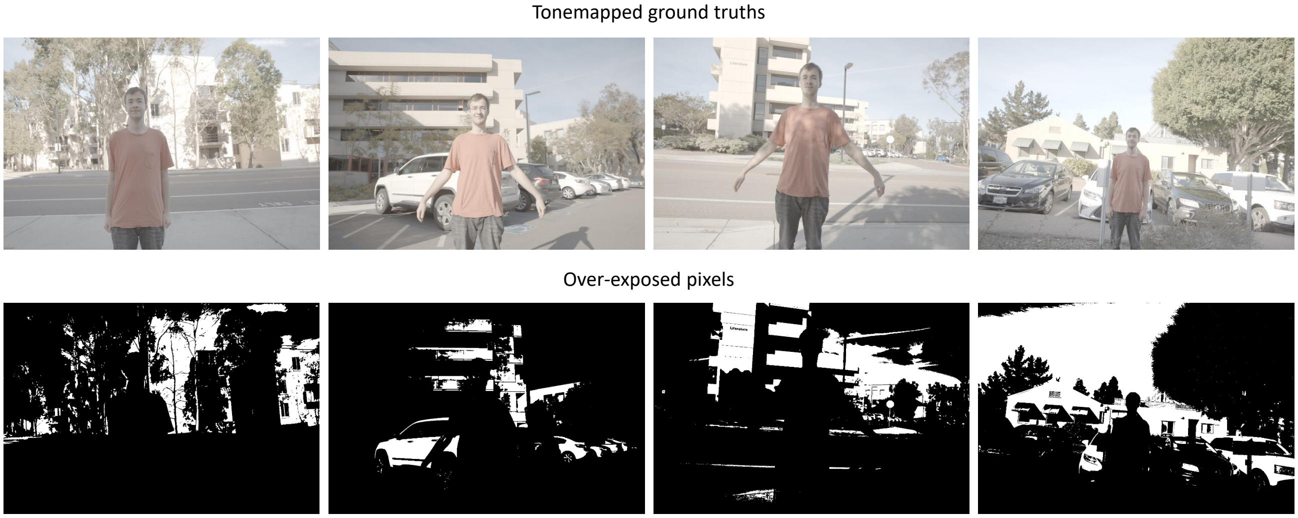

Figure 1. Exposition masks for Kalantari [2] testing samples where state-of-the-art deghosting methods [4, 3] yield unsatisfactory results in the linear domain. The exposition masks use white pixels to denote over-exposed regions and black pixels to represent reasonably-exposed regions. It can be observed that most of the backgrounds are over-exposed.

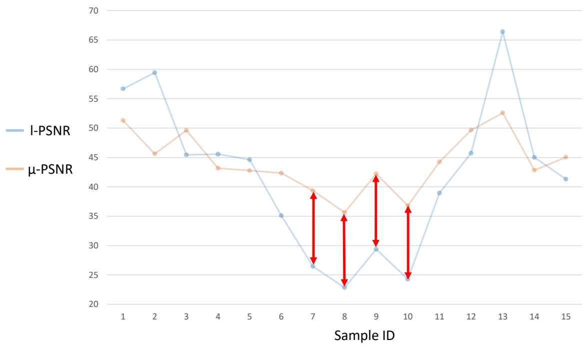

Figure 2. Average l-PSNR and μ-PSNR scores from SOTA methods, including ours, on Kalantari [2] testing samples. For samples 7 to 10, the general performance is poor in the linear domain. Meanwhile, after the tonemapping function, the μ-PSNR score becomes significantly higher, yielding an undesirable but important gap, in red color, between these metrics. This phenomenon may be caused due to the large number of over-exposed pixels in these samples. Please zoom in for more details.

our study, we define a pixel as over-exposed if its value is greater than 95% of the maximum value that can be encoded in the HDR ground truth. The proportion of over-exposed pixels in each sample can be found in Table 2. The visualization can be found in Figure 4. It can be seen that samples 7 to 10 contain more over-exposed pixels than others. The corresponding over-exposed mask can be found in Figure 1 for these samples. It can be seen that a large number of the background pixels are over-exposed. We think that the over-exposition may be linked to several factors: (1) a too-long medium exposure time making the reference image already over-exposed; (2) too-small differences in the exposure time during input LDR images, making it impossible to cover a larger dynamic range; (3) the conventional 3 input LDR images may not be sufficient.

Therefore, while collecting our dataset, we followed a very rigorous processing as described in our main manuscript. Following the same protocol, we show in Figure 5 that none of our samples contains more than 2% of over-exposed values. In addition, even though our dataset follows the conventional setting with 3 LDR inputs, we will release the whole bracket of 9 input exposures. In such a case, future works can benefit from a larger dynamic range to design better deghosting methods.

3. Qualitative Evaluation on Unsupervised Benchmarks

In order to verify the generalizability of our method, we conducted evaluations on the datasets proposed by Sen *et al.* [5]. All the compared networks are trained using our proposed dataset. As depicted in Figure 3, all other methods [6, 4, 3] exhibit distortion in over-exposed areas, while our network is able to reproduce the texture of the piano scores book most accurately.

4. Quantitative Performance on Our Dataset

The full performance can be found in Table 3.

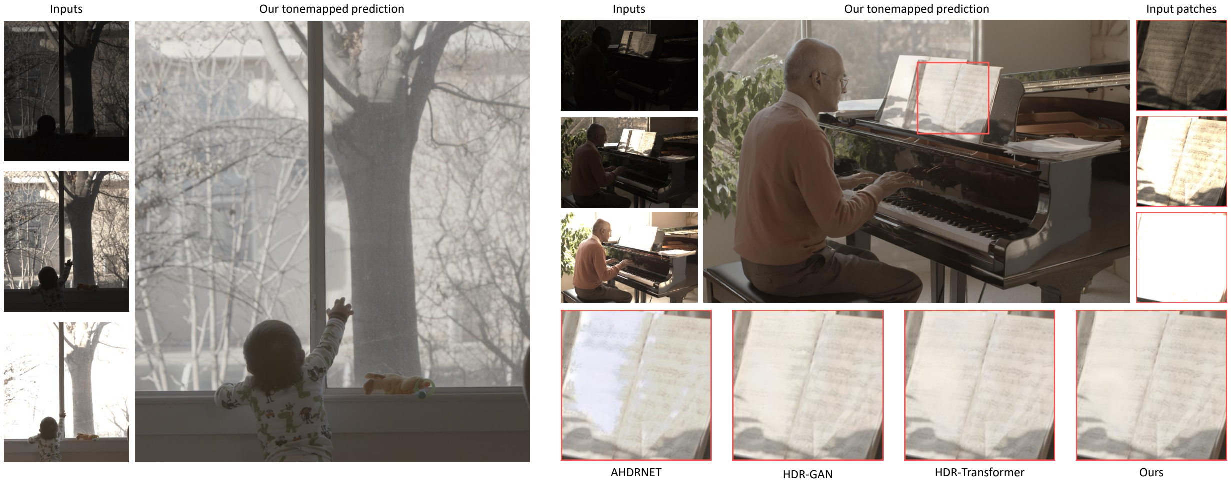

Figure 3. Evaluation of the generalizability of our solution using the Sen *et al*. [5] unsupervised dataset. All the compared networks are trained with our proposed dataset. It can be seen that our network reproduces a better texture of the piano scores book. Please zoom in for more details.

Table 3. Quantitative comparison with state-of-the-art methods on our proposed dataset. l-PSNR and l-SSIM are computed in the linear domain while μ-PSNR and μ-SSIM are computed after μ-law tone mapping. PU-PSNR and PU-SSIM are calculated by applying the encoding function proposed in [1]. The compared methods are trained through their official implementation.

Method	μ-PSNR	PU-PSNR	l-PSNR	μ-SSIM	PU-SSIM	l-SSIM	HDR-VDP2
NHDRRNet	36.68	37.06	39.61	0.9590	0.9777	0.9853	65.41
DHDRNet	40.05	40.47	43.37	0.9794	0.9889	0.9924	67.09
AHDRNet	42.08	42.30	45.30	0.9837	0.9919	0.9943	68.80
HDR-Transformer	42.39	42.65	46.35	0.9844	0.9920	0.9948	69.23
SCTNet (Ours)	**42.55**	**42.80**	**47.51**	**0.9850**	**0.9924**	**0.9952**	**70.66**

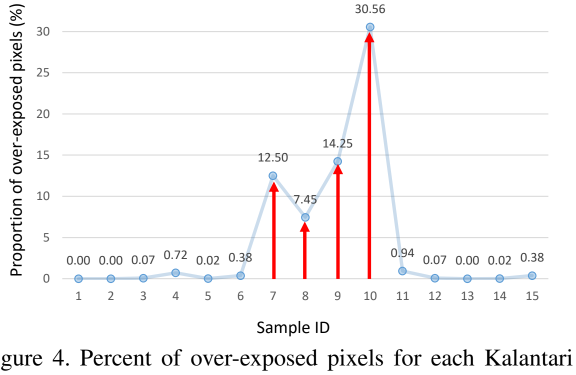

Figure 4. Percent of over-exposed pixels for each Kalantari [2] testing sample. We can find the anomalies in samples 7 to 10 where more than 5% of pixels are over-exposed.

References

[1] Maryam Azimi et al. Pu21: A novel perceptually uniform encoding for adapting existing quality metrics for hdr. In *2021 Picture Coding Symposium (PCS)*, pages 1–5. IEEE, 2021. 3

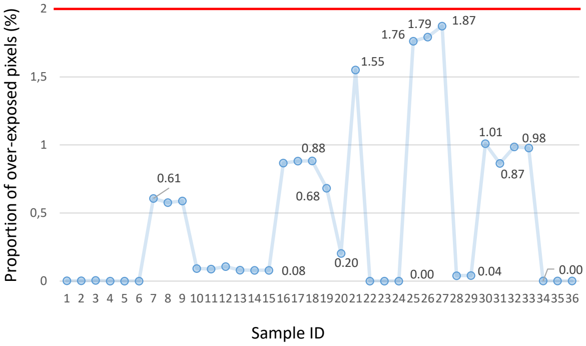

Figure 5. Compared to the existing benchmark, the proportion of over-exposed pixels is consistently lower than 2% in our dataset.

[2] Nima Khademi Kalantari, Ravi Ramamoorthi, et al. Deep high dynamic range imaging of dynamic scenes. *ACM TOG*, 36(4):144–1, 2017. 1, 2, 3

[3] Zhen Liu, Yinglong Wang, Bing Zeng, and Shuaicheng Liu. Ghost-free high dynamic range imaging with context-aware transformer. In *ECCV*, 2022. 1, 2

[4] 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 TIP*, 30:3885–3896, 2021. 1, 2

[5] Pradeep Sen, Nima Khademi Kalantari, Maziar Yaesoubi, Soheil Darabi, Dan B Goldman, and Eli Shechtman. Robust Patch-Based HDR Reconstruction of Dynamic Scenes. *ACM TOG*, 31(6):203:1–203:11, 2012. 2, 3

[6] 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. *CVPR*, 2019. 2