Supplementary Material:
Demosaicing a Time-Varying Exposures Array for Single-Shot HDR Imaging

1. Experimental Details of Comparison

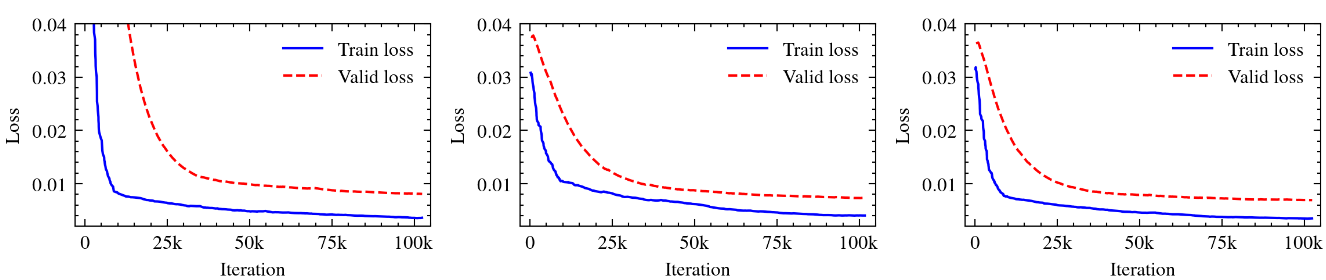
We compare our method's results with three different methods (Suda *et al*. [5], Yan *et al*. [6] and Liu *et al*. [3]) in this paper. Since Suda *et al*. [5] release the pretrained model only without any training code, we cannot adapt their model to the evaluation dataset. For a fair comparison with the other two methods [6, 3], we adapt the existing methods to have the same inputs and outputs as our network while maintaining the original structure of the HDR deghosting method. To make these methods adapt to the problem domain, we also retrain the modified network from scratch using the same training set and hyperparameters as our model.

Both compared methods receive images of three exposure times as input and produce an HDR image result with the same resolution as the input image. The resulting image of our method has a four times larger resolution than an image of a single exposure time. And thus, after upscaling each exposure time image by two times, we feed three feature vectors into the original network. After executing the original network for inference, we generate an output by upscaling by two times again to make their results' resolution match ours. Specifically, in the first upscale, we use Pixelshuffle [4] to extract upscaled shallow feature vectors and feed them as input to the network. We use the same upscale method in our network architecture. In the second upscale, we apply Pixelshuffle again to upscale the network just before the convolutional layer creates the final output image with the three channels. For the modified HDR deghosting network, we also concatenate image values normalized by the exposure time of each pixel without gamma correction to the subsampled images.

In Figure 1, we show the training and valid losses when training each modified network. Note that all three methods, including ours, are trained with the same number of iterations, approximately 100,000. We can observe that our network and both modified networks converge well.

2. Additional Results

In this document, we provide additional results of our method, compared with other methods (Suda *et al*. [5], Yan *et al*. [6] and Liu *et al*. [3]) for the test dataset, which consists of 40 HDR images obtained from the Stuttgart HDR image dataset [1]. All HDR images are tone-mapped adaptively by the μ-law [2] as clear as possible for visualization. We show the ground-truth images and then present three compared methods and our results of the synthetic dataset for each scene. Also, we visualize three compared methods' and our results of a real-world dataset captured with our camera setup.



Yan et al. Liu et al. Ours

Figure 1. Visualization of training and validation losses of three different methods (Yan *et al*. [6], Liu *et al*. [3] and ours) over 100,000 iterations.

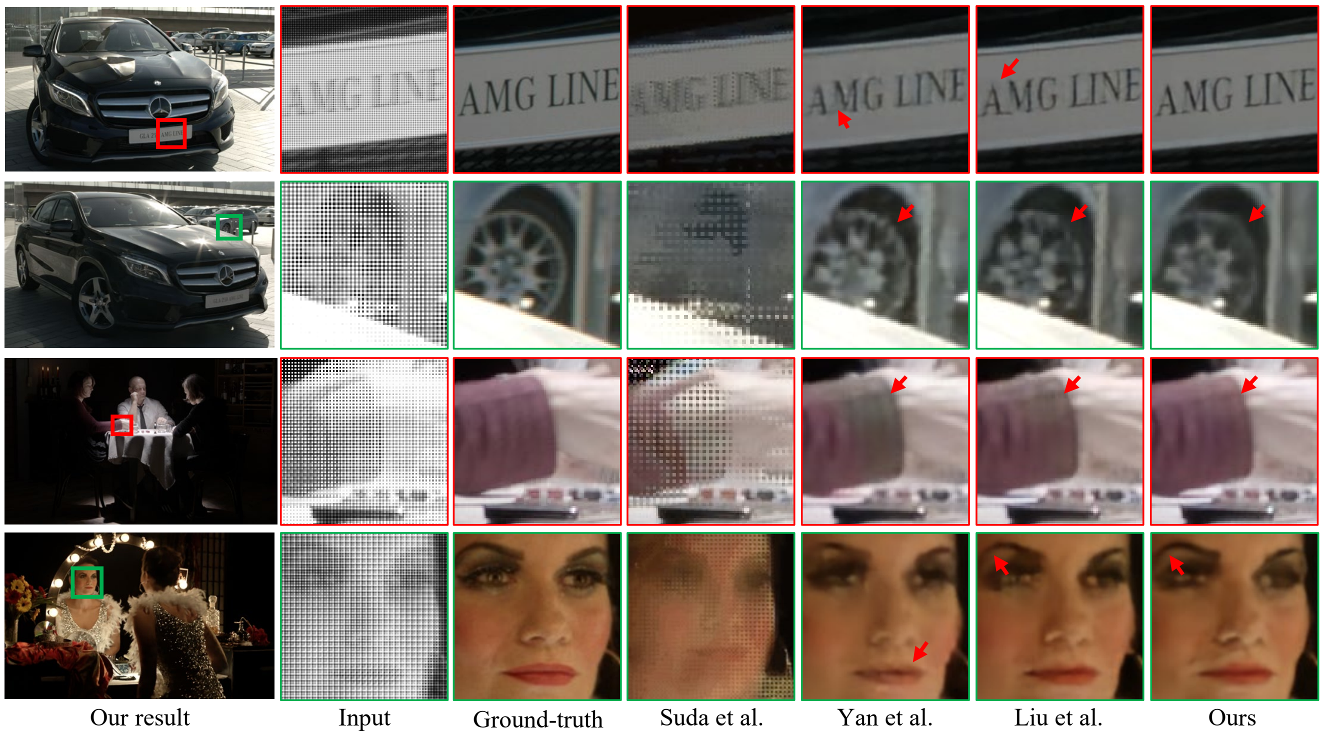


Figure 2. Synthetic dataset comparison of our method with Suda *et al*. [5], Yan *et al*. [6] and Liu *et al*. [3]. Our model outperforms baseline methods in terms of image structure and motion blur. The performance difference is clearly visible, particularly in the area with strong motion blur.

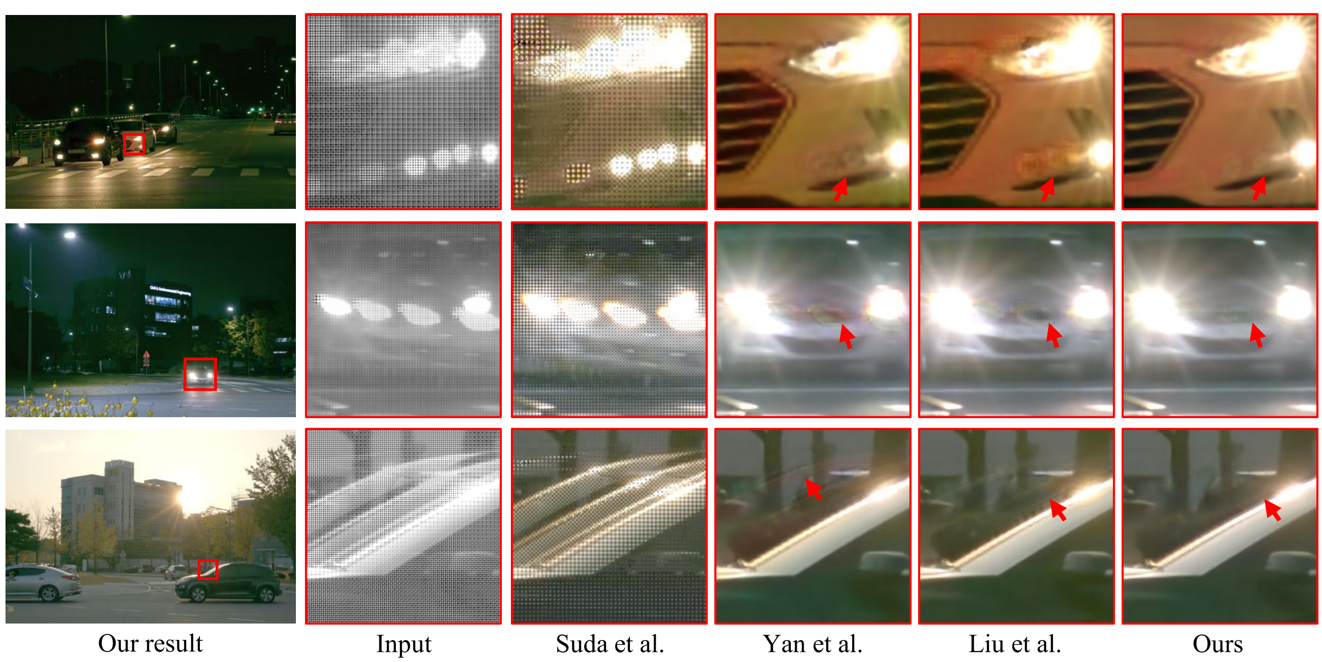


Figure 3. Real-world dataset comparison with Suda *et al*. [5], Yan *et al*. [6] and Liu *et al*. [3]. Three test scenes are presented in this figure. Even in these extreme environments, our approach successfully minimizes motion blur artifacts.

References

[1] Jan Froehlich, Stefan Grandinetti, Bernd Eberhardt, Simon Walter, Andreas Schilling, and Harald Brendel. Creating cinematic wide gamut hdr-video for the evaluation of tone mapping operators and hdr-displays. In *Digital photography X*, volume 9023, page 90230X. International Society for Optics and Photonics, 2014. 1

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

[3] Zhen Liu, Yinglong Wang, Bing Zeng, and Shuaicheng Liu. Ghost-free high dynamic range imaging with context-aware transformer. In *European Conference on Computer Vision*, pages 344–360. Springer, 2022. 1, 2

[4] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016. 1

[5] Takeru Suda, Masayuki Tanaka, Yusuke Monno, and Masatoshi Okutomi. Deep snapshot hdr imaging using multi-exposure color filter array. In *Proceedings of the Asian Conference on Computer Vision*, 2020. 1, 2

[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. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1751–1760, 2019. 1, 2