Supplementary Material for:

Unpaired Learning for High Dynamic Range Image Tone Mapping

Yael Vinker Inbar Huberman-Spiegelglas Raanan Fattal

Content:

Appendix

The FID measures the distance between the distribution of two image classes by modeling their neural activations using multivariate Gaussians and measuring their Fréchet distance. More specifically, the 2048-dimensional activation vectors produced at the last layer of the Inception-v3 classification network are used. Fitting a Gaussian stably is typically done with using 10,000 samples or more. Since each image produces one sample vector, this is the number of images needed from each class in order to estimate the FID.

Datasets of HDR images do not offer these numbers of samples and do not permit the use of the FID. To overcome this limitation we consider the pooled activations in the last Inception layer of the Inception Module B. This layer consists of a spatial array of 8×8 pixels, each containing an 768-dimensional activation vector. Thus, compared to the original FID where every image provides a single sample vector, switching to this layer results in 64 samples per image.

We term this modified distance *pixFID* and compare it to the original FID in Figure 1. We used only 1,000 images to evaluate the pixFID in this comparison, next to 50,000 images used to evaluate the FID. The comparison shows the response of these distances to three types of disturbances that were applied to natural images. In case of the Gaussian blur and the additive noise, the pixFID followed the trend of the FID and increased as the disturbance magnitude increased, implying its ability to identify such low-level discrepancies.

In case of a geometric swirl distortion, which does not change the color and shape distribution of small image patches, the pixFID showed an inferior sensitivity to the transformation magnitude. This behaviour is consistent with the fact that activations from a lower level in the network correspond to smaller receptive field and hence to a visual sensitivity to smaller scale features in the image.

Nevertheless, HDR tone-mapping procedures typically does not suffer from geometric or semantic distortions, but from low-level effects such as loss of local contrast, exaggerated details, and edge-related halos.

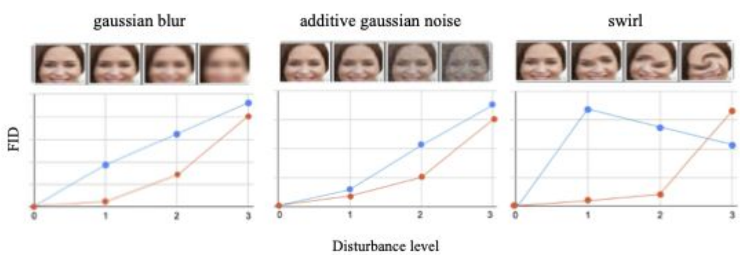
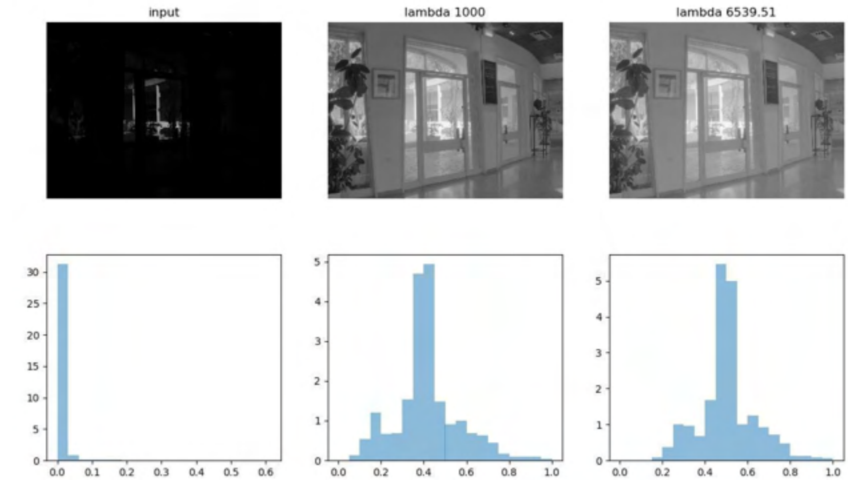


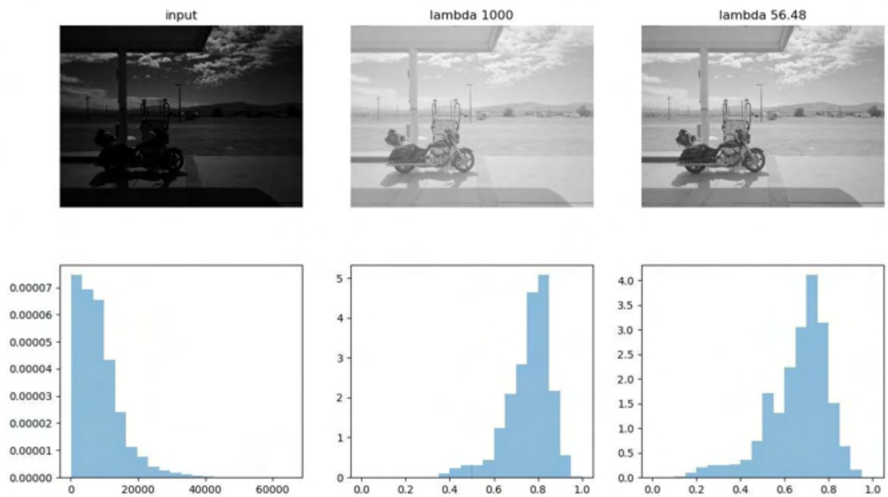
Figure 1. The response of the FID, red plot, and the pixFID, blue plot, to three types of image disturbances are shown. The FID is computed over 50,000 images whereas the pixFID over 1,000 images.

Ablation Study - Additional Tests and Images

Compression Level Estimation

Below we compare the results produced by our per-image compression factor estimation, next to using a fixed value = 1000 as done in “Deep High Dynamic Range Imaging of Dynamic Scenes” SIGGRAPH 2017. This fixed value (center column) led to a slight over compression in the luminance map of the Motorcycle and a clear under compression in the Belgium House image below. In both these examples, our image dependent factor (right column) brought most of the pixels to a visible range and maintained a consistent level of image contrasts.





As expected, also the final images produced by training and inferring by applying a fixed level of compression show an inconsistent appearance due to the different dynamic range they had at the input.

Augmenting the Skip-Connection with Square-Root Transformation

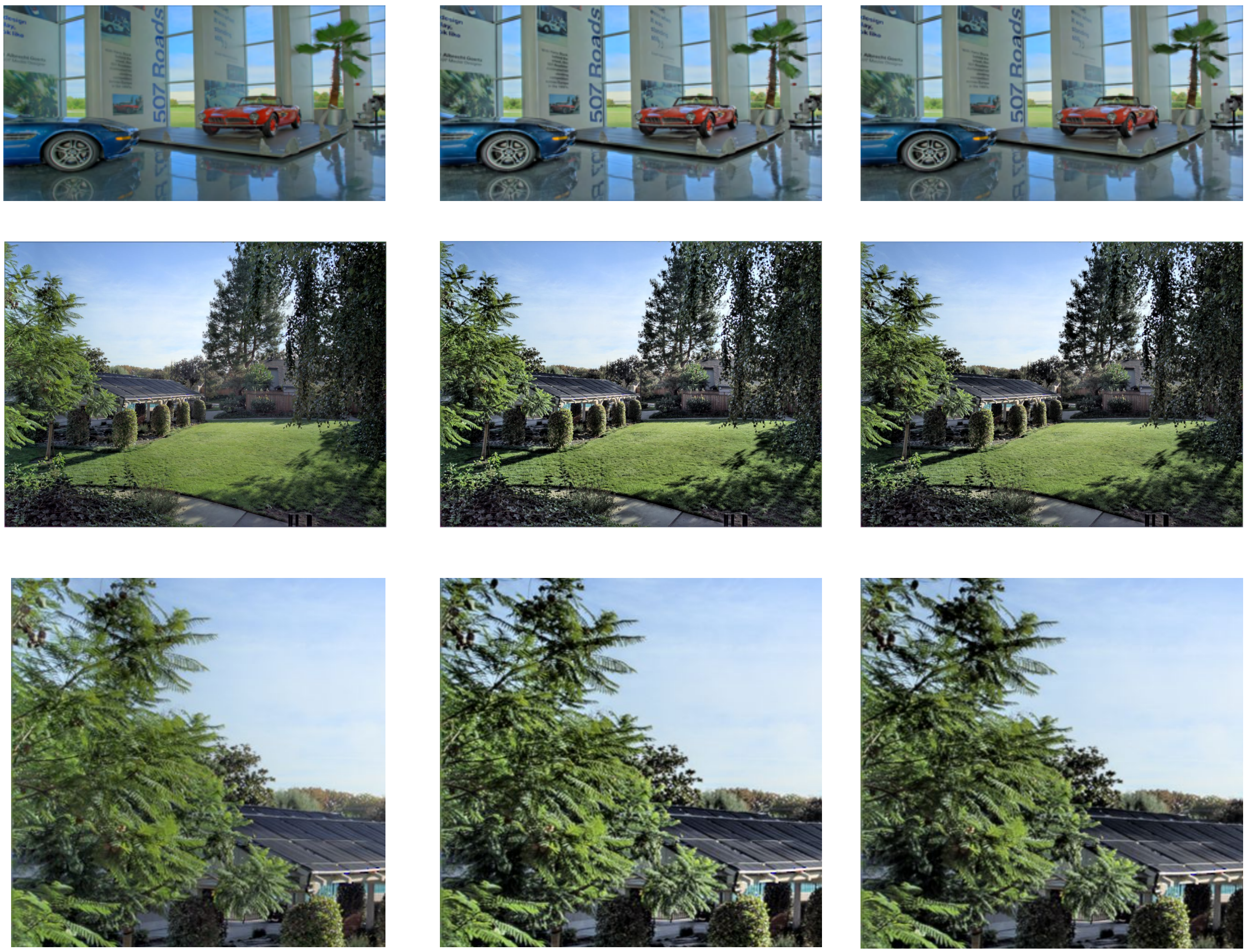
The following table shows the images resulting with and without augmenting the skip connections with a smooth transformation. The lack of one (left column) shows a coarse action over the image with noticeable over - and undershooting effects. The use of a square-root (right column) almost completely eliminates these effects.

No trans. Square root trans.

Effect of the Discriminators at Different Image Scales

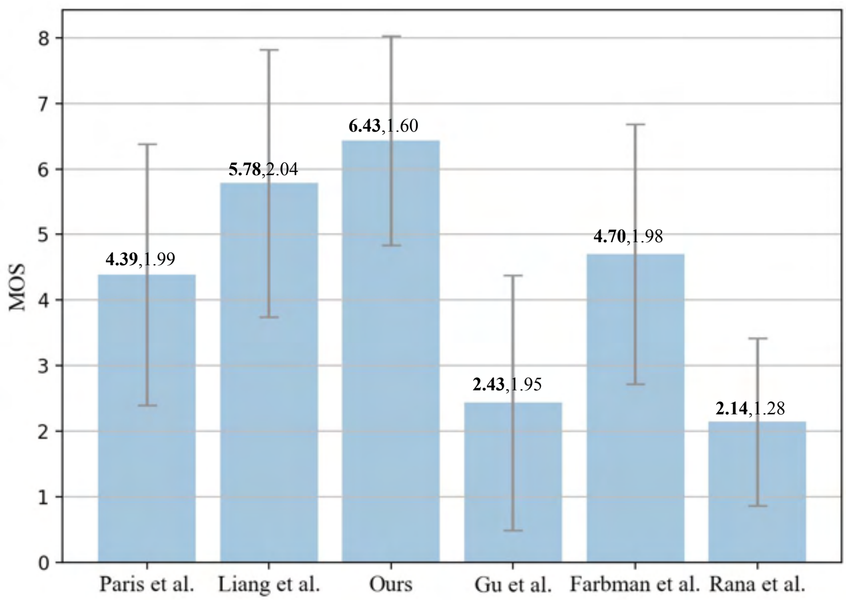
The following table shows the results obtained by training our network using discriminator at different scales. From this comparison it is clear that when omitting discriminators at certain scales (coarse scale or two in the right columns of the table), the network does not reproduce the missing (by the prior TRC step) contrasts at these scales.

One scale | Two scales | Three scales (ours)



User Study

Graph reports the results of the user study we conducted



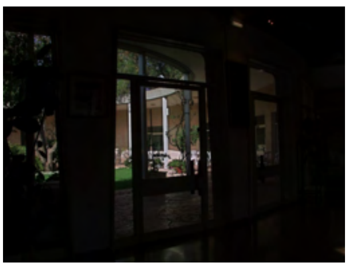
In order to further assess the visual quality of our results, we conducted a user-study that follows the protocol used in Liang[10]. Specifically, it consists of eight participants who were asked to rate between 1 to 8 images produced by the methods in Liang[10], Paris[7], Rana[12], Gu[8], Farbman[9] as well as ours. The images were displayed next to each other (at a random order). The participants were naive as to the purpose of the test and were asked to score the images based on how realistic and unprocessed thy appear.
The graph shows that our method obtained the highest mean score 6.43 with a tolerate std of 1.6. The mean opinion score (in bold) and std of each method are shown in the graph.

Results by Low Light Enhancement Methods

While Low light methods also manipulate the image brightness, they are not adapted to operate on images with very high dynamic range.

Jiang et al. [16]

Guo and Li et al. [15]



Additional Visual Comparisons

from the HDR+ dataset [2]

Gu [8]

Farbman [9]

Liang [10]

Ours

image name: merged_33TJ_20150612_201525_012

Paris [7]

Shibata [5]

Liang [10]

Ours

image name: merged_6FHF_20150331_121129_927

Gu [8]

Shan [3]

Liang [10]

Ours

image name: merged_33TJ_20150723_154227_435

Gu [8]

Farbman [9]

Liang [10]

Ours

image name: merged_5a9e_20150403_125941_989

Additional Visual Comparisons

from the HDR Survey dataset [14]

Farbman [9]

Paris [7]

Shibata [5]

Liang [10]

Rana [12]

Ours

Farbman [9]

Paris [7]

Shibata [5]

Liang [10]

Gu [8]

Ours

Farbman [9]

Paris [7]

Shibata [5]

Liang [10]

Rana [12]

Ours

Farbman [9]

Paris [7]

Shibata [5]

Liang [10]

Rana [12]

Ours

Farbman [9]

Shan [3]

Shibata [5]

Ferradans [4]

Mai [6]

Ours

Farbman [9]

Paris [7]

Shibata [5]

Liang [10]

Rana [12]

Ours

Farbman [9]

Paris [7]

Shibata [5]

Liang [10]

Rana [12]

Ours

Farbman [9]

Paris [7]

Shibata [5]

Liang [10]

Rana [12]

Ours

Farbman [9]

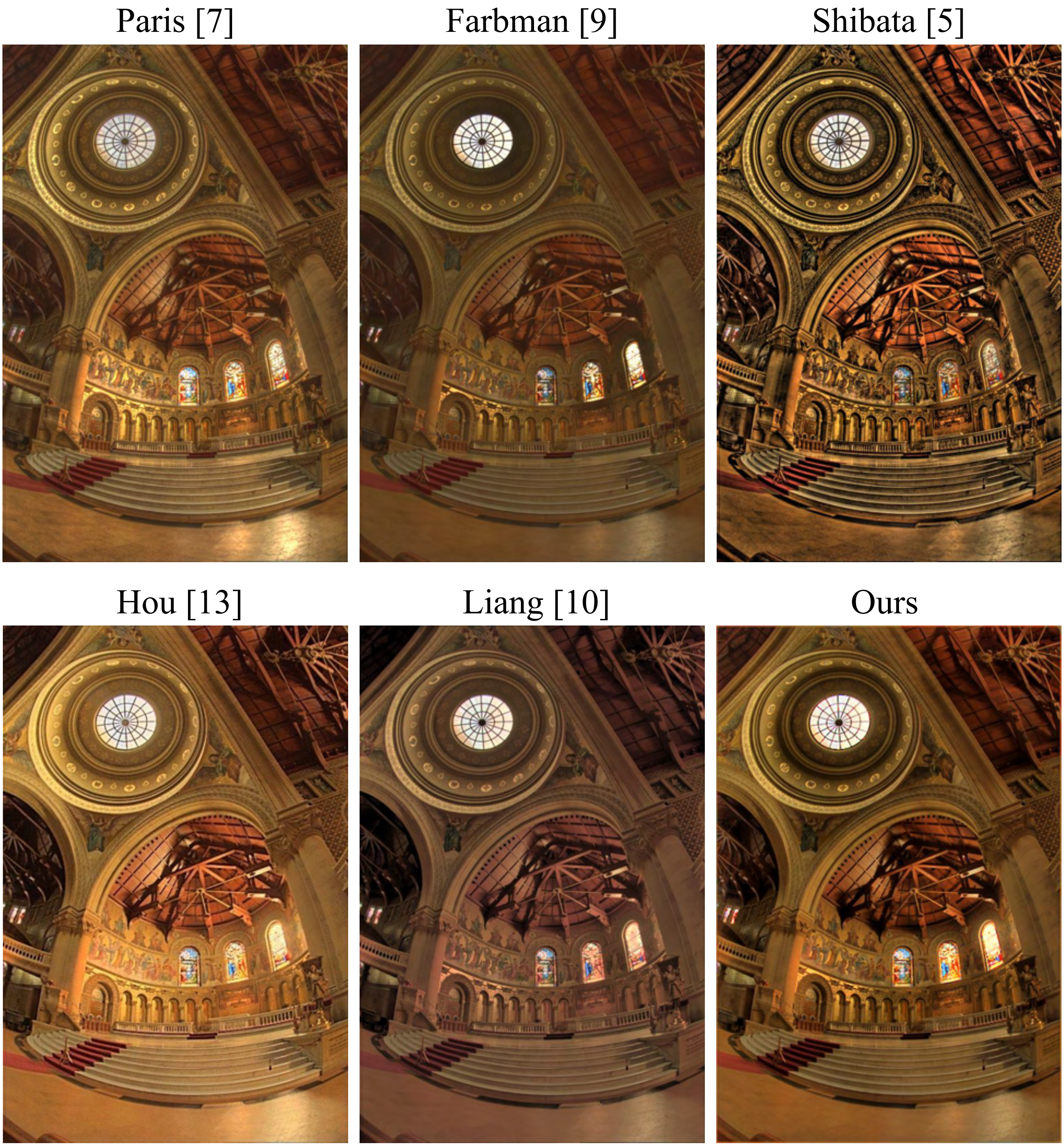
Paris [7]

Shibata [5]

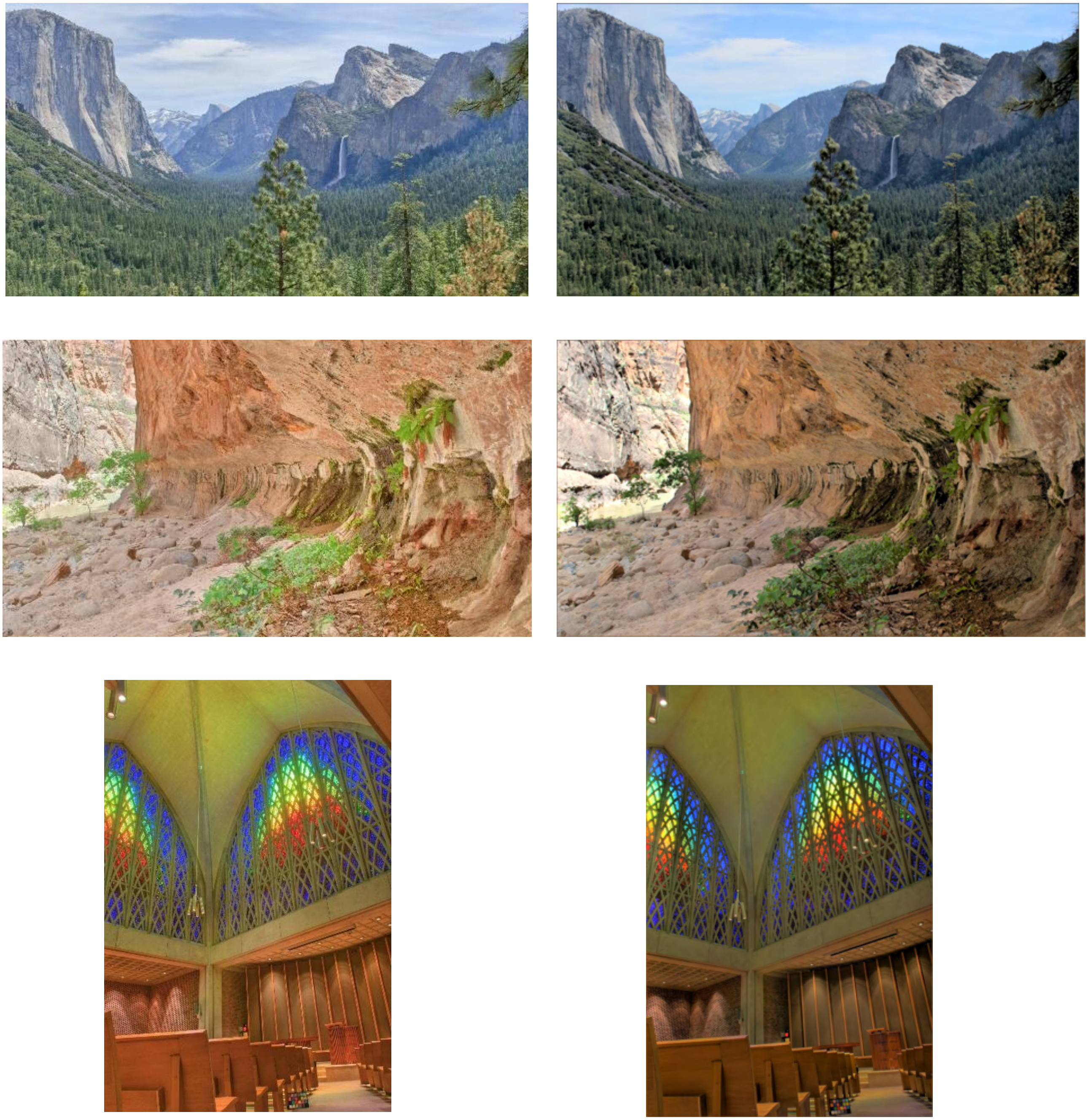
Liang [10]

Rana [12]

Ours


Paris [7]
Farbman [9]
Shibata [5]
Hou [13]
Liang [10]
Ours

Zhang [11] | Ours




Zhang [11]
Ours

Examples Training Images

Examples of LDR images from the DIV2K dataset [1]

Images from the tone-mapping pipeline

LDR images

Tone mapping network input (Yc)

Network output (N(Yc))

Images from the tone-mapping pipeline

LDR images

Tone mapping network input (Yc)

Network output (N(Yc))

Images from the tone-mapping pipeline

LDR images

Tone mapping network input (Yc)

Network output (N(Yc))

Network’s action: Yc (shown in color) and the network’s output on it are shown.

color input our result

color input our result

color input our result

Fail Cases

Over contrast due to the adversarial loss in the coarsest scale, which encourage higher contrasts

By omitting this term we can reduce this exaggerated appearance

References

1. Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017
2. Samuel W. Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T. Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 35(6), 2016.
3. Qi Shan, Jiaya Jia, and Michael S Brown. Globally optimized linear windowed tone mapping. *IEEE transactions on visualization and computer graphics*, 16(4):663–675, 2009.
4. Sira Ferradans, Marcelo Bertalmio, Edoardo Provenzi, and Vincent Caselles. An analysis of visual adaptation and contrast perception for tone mapping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):2002–2012, 2011
5. Takashi Shibata, Tanaka Masayuki, and Masatoshi Okutomi. Gradient-domain image reconstruction framework with intensity-range and base-structure constraints. pages 2745–2753, 06 2016.
6. Zicong Mai, Hassan Mansour, Rafal Mantiuk, Panos Nasiopoulos, Rabab Ward, and Wolfgang Heidrich. Optimizing a tone curve for backward-compatible high dynamic range image and video compression. *IEEE transactions on image processing*, 20(6):1558–1571, 2010
7. Sylvain Paris, Samuel W. Hasinoff, and Jan Kautz. Local laplacian filters: Edge-aware image processing with a laplacian pyramid. *ACM Trans. Graph.*, 30(4), 2011
8. B. Gu, W. Li, M. Zhu, and M. Wang. Local edge-preserving multiscale decomposition for high dynamic range image tone mapping. *IEEE Transactions on Image Processing*, 22(1):70–79, 2013.
9. Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2008)*, 27(3):to appear, Aug. 2008
10. Zhetong Liang, Jun Xu, David Zhang, Zisheng Cao, and Lei Zhang. A hybrid l1-l0 layer decomposition model for tone mapping. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
11. N. Zhang, C. Wang, Y. Zhao, and R. Wang. Deep tone map- ping network in hsv color space. In *2019 IEEE Visual Communications and Image Processing (VCIP)*, pages 1–4, 2019.
12. A. Rana, P. Singh, G. Valenzise, F. Dufaux, N. Komodakis, and A. Smolic. Deep tone mapping operator for high dynamic range images. *IEEE Transactions on Image Processing*, 29:1285–1298, 2020.
13. Xianxu Hou, Jiang Duan, and Guoping Qiu. Deep feature consistent deep image transformations: Downscaling, decolorization and HDR tone mapping. *CoRR*, abs/1707.09482, 2017.
14. M.D. Fairchild. The hdr photographic survey. pages 233– 238, 01 2007.
15. Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vi- sion and Pattern Recognition (CVPR)*, June 2020
16. Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE Trans. Image Process* 2021