

Supplementary Material – CycleISP: Real Image Restoration via Improved Data Synthesis

Syed Waqas Zamir¹ Aditya Arora¹ Salman Khan¹ Munawar Hayat¹

Fahad Shahbaz Khan¹ Ming-Hsuan Yang^{2,3} Ling Shao¹

¹Inception Institute of Artificial Intelligence, UAE

²University of California, Merced ³Google Research

Table 1: Ablation: the proposed denoising network is trained using the real RAW images from the training set of SIDD [1]. Evaluation is performed on the RAW images from the validation set of SIDD.

# RRG	# DAB	# channels	SA	CA		PSNR
4	8	64	✓	✓		51.85
4	8	64		✓		51.52
4	8	64	✓			51.63
5	10	64	✓	✓		51.97
4	8	128	✓	✓		52.03

1. Additional Ablations

In the main manuscript, we provide ablation study for the RAW2RGB branch of CycleISP. The RAW2RGB branch takes a clean RAW image and transforms it to a clean sRGB image (when we keep the noise injection module switched off). Those ablation experiments show the impact of each component on the quality of synthesized clean sRGB data.

Denoising network: Here, we ablate the effect of individual contributions to the proposed denoising network. We train our network for the task of RAW denoising using the *real* training data of SIDD [1]. Since the purpose is to examine the performance of the proposed denoising network, we do not include any synthetic data (generated with the CycleISP) for this ablation study. All the denoising experiments in the paper are performed using the network that consists of 4 recursive residual groups (RRGs), where each RRG further contains 8 dual attention blocks (DABs); we termed it as *base* network settings. The number of output channels for all the convolutional layers in RRG modules is set to $C = 64$, except for the first layer of the channel attention branch that has $\frac{C}{r}$ channel dimension. Where r is set to 16 in our experiments. Table 1 summarizes the results. The first row shows that we obtain PSNR of 51.85 dB with the base network settings. Removing either the spatial attention (SA) branch or the channel attention (CA) branch reduces the network performance (see row 2 and row 3).

Table 2: Ablation: incorporating real images (from the SIDD [1]) and synthetic images (generated with our CycleISP) for training the proposed denoising network. Evaluation is performed on the RAW images from the validation set of SIDD [1].

Real Data	Synthetic Data	PSNR
✓		51.85
	✓	50.45
✓	✓	52.41

Furthermore, the increase in the depth and width of the network improves accuracy in terms of PSNR as compared to the base settings (see rows 4 and 5), but at a high computational cost and a large memory footprint. We decide to use 4 RRGs and 8 DABs (base settings) in the main paper so as to have a good trade-off between the speed and accuracy.

Combining synthetic and real data: The SIDD benchmark dataset [1] contains 320 real (high-resolution) image pairs for training and 1280 image pairs for validation. The denoising results we present in the paper are obtained using the network that is trained on a combined dataset, incorporating both real images from the SIDD [1] and synthetic images generated with the CycleISP. Here, we perform two additional experiments to evaluate the performance of the denoising network: (1) training network only on the real data from SIDD [1], and (2) training network only on the synthetic data generated with the CycleISP. Table 2 shows that the network trained only on real data or synthetic data yields results with PSNR 51.85 dB (row 1) and 50.45 dB (row 2), respectively. Whereas the network trained on the dataset containing both real and synthetic images is much more effective in removing noise (see row 3).

Figure 1c depicts that the network trained on real data cannot completely remove noise, and produces image with splotchy texture and less neat structural content. It may be partially due to the small training dataset, and partially because of the imperfect ground-truth images. For instance, we can see in Figure 1b that the ground-truth image is not



(a) Noisy (b) Ground-truth (c) Real (d) Synthetic (e) Real+Synthetic

Figure 1: Ablation to study the effect of real and synthetic data on the network performance. (a) Noisy input image. (b) (Nearly) noise-free reference image. (c) Output the network trained on real data from SIDD [1]. (d) Output of the network trained on synthetic data generated with the proposed CycleISP. (e) Output of the network trained on combined dataset, incorporating both real images from the SIDD [1] and synthetic images generated with the CycleISP.

completely noise-free.

While the network which is trained on synthetic data generate images that are sharp and noise-free, it often produces dark image values which are not perceptually-faithful to the ground-truth. For example, see the alphabets ‘E’ and ‘X’ in Figure 1d. It may be due to the noise model that we use in the noise injection module of the CycleISP. We use the same procedure for noise modeling as in [3]. Although the noise modeling method of [3] is reasonable, it does not take into account the effect of clipping operation on the image values. Therefore, by incorporating more sophisticated noise modeling techniques (*e.g.*, [6]) in the noise injection module, we might be able to further improve the performance of the denoising network. Finally, Figure 1e shows that when the network is trained on a combined dataset (that includes both real and synthetic images), it effectively removes noise and generates results that are closer to the ground-truth in appearance.

2. Results for Denoising

In this section we provide additional results for denoising the sRGB images and RAW images.

Denoising sRGB images of SIDD [1]: Figures 2, 3 and 4 show the results produced by our method and those of the state-of-the-art. Our method performs favorably against the RIDNet [2] and CBDNet [7]. Notably, our method generates clean, sharp, and artifact-free results, while preserving the true image details.

Denoising RAW images of DND [11]: Figure 5 illustrates the results of different algorithms. Our method yields images with better visual quality and PSNR as compared to other competing approaches [3, 12, 4].

3. Color Matching for Stereoscopic Cinema

In Figure 6 and Figure 7, we compare the results of matching colors of the source view to the target view by

using the proposed method and three other techniques [9, 10, 13].

References

- [1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *CVPR*, 2018. 1, 2, 3, 4, 5
- [2] Saeed Anwar and Nick Barnes. Real image denoising with feature attention. *ICCV*, 2019. 2, 3, 4, 5
- [3] Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet, and Jonathan T Barron. Unprocessing images for learned raw denoising. In *CVPR*, 2019. 2, 6
- [4] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *TIP*, 2007. 2, 6
- [5] <https://noise.visinf.tu-darmstadt.de/benchmark/>, 2017. [Online; accessed 15-Nov-2019]. 6
- [6] Alessandro Foi. Clipped noisy images: Heteroskedastic modeling and practical denoising. *Signal Processing*, 2009. 2
- [7] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *CVPR*, 2019. 2, 3, 4, 5
- [8] Hakki Can Karaimer and Michael S Brown. A software platform for manipulating the camera imaging pipeline. In *ECCV*, 2016. 6
- [9] Hiroaki Kotera. A scene-referred color transfer for pleasant imaging on display. In *ICIP*, 2005. 2, 7, 8
- [10] François Pitié, Anil C Kokaram, and Rozenn Dahyot. Automated colour grading using colour distribution transfer. *Trans. on CVIU*, 2007. 2, 7, 8
- [11] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *CVPR*, 2017. 2, 6
- [12] Tobias Plötz and Stefan Roth. Neural nearest neighbors networks. In *NeurIPS*, 2018. 2, 6
- [13] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *Trans. on Computer graphics and applications*, 2001. 2, 7, 8

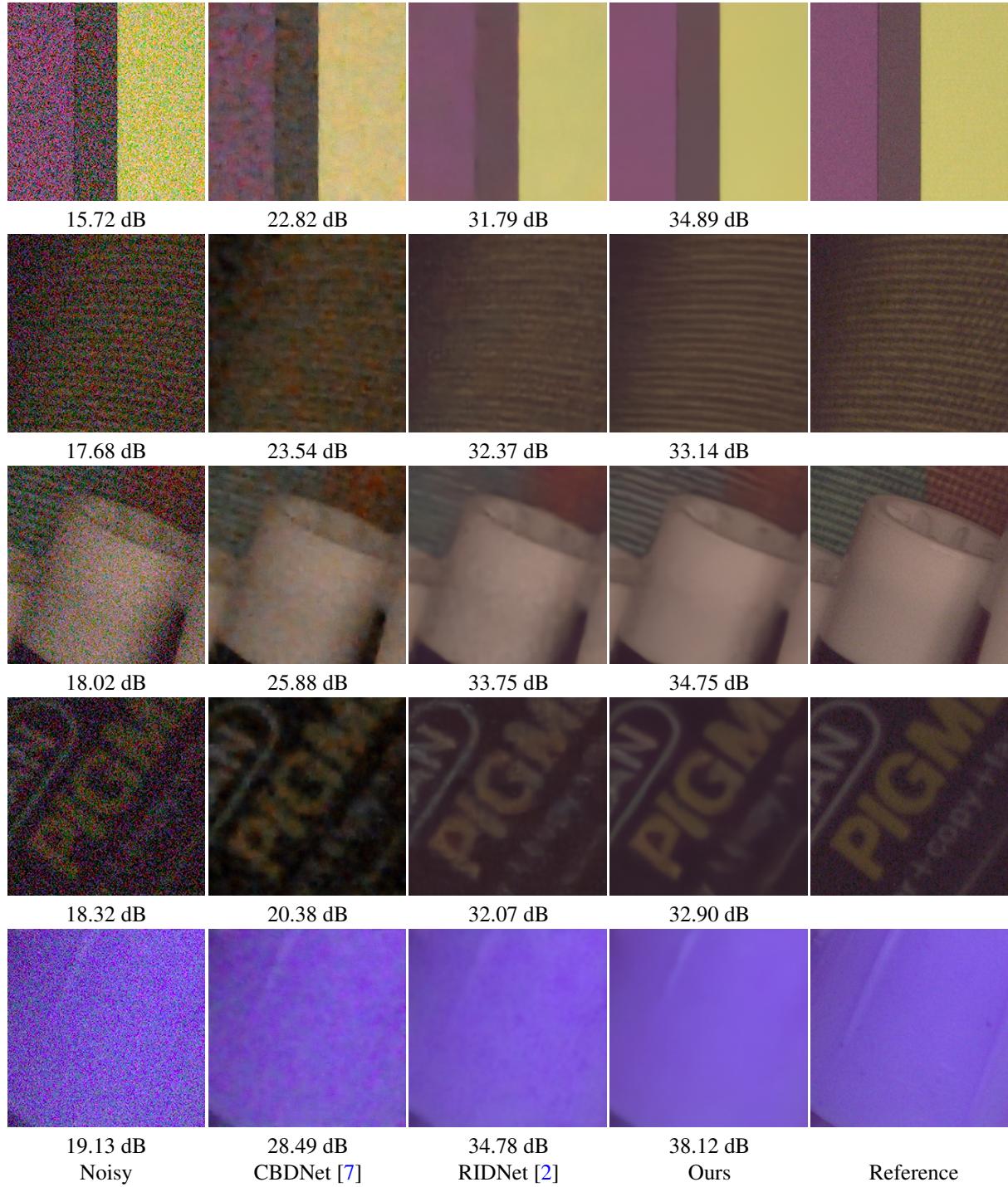


Figure 2: Denoising sRGB images from the SIDD benchmark dataset [1]. RIDNet [2] and CBDNet [7] generates images with splotchy texture (see row 3 and row 4). Our method preserves the right image structure (see row 2 and row 3) and reproduces images with better color and without chroma artifacts (see rows 1 and 5).

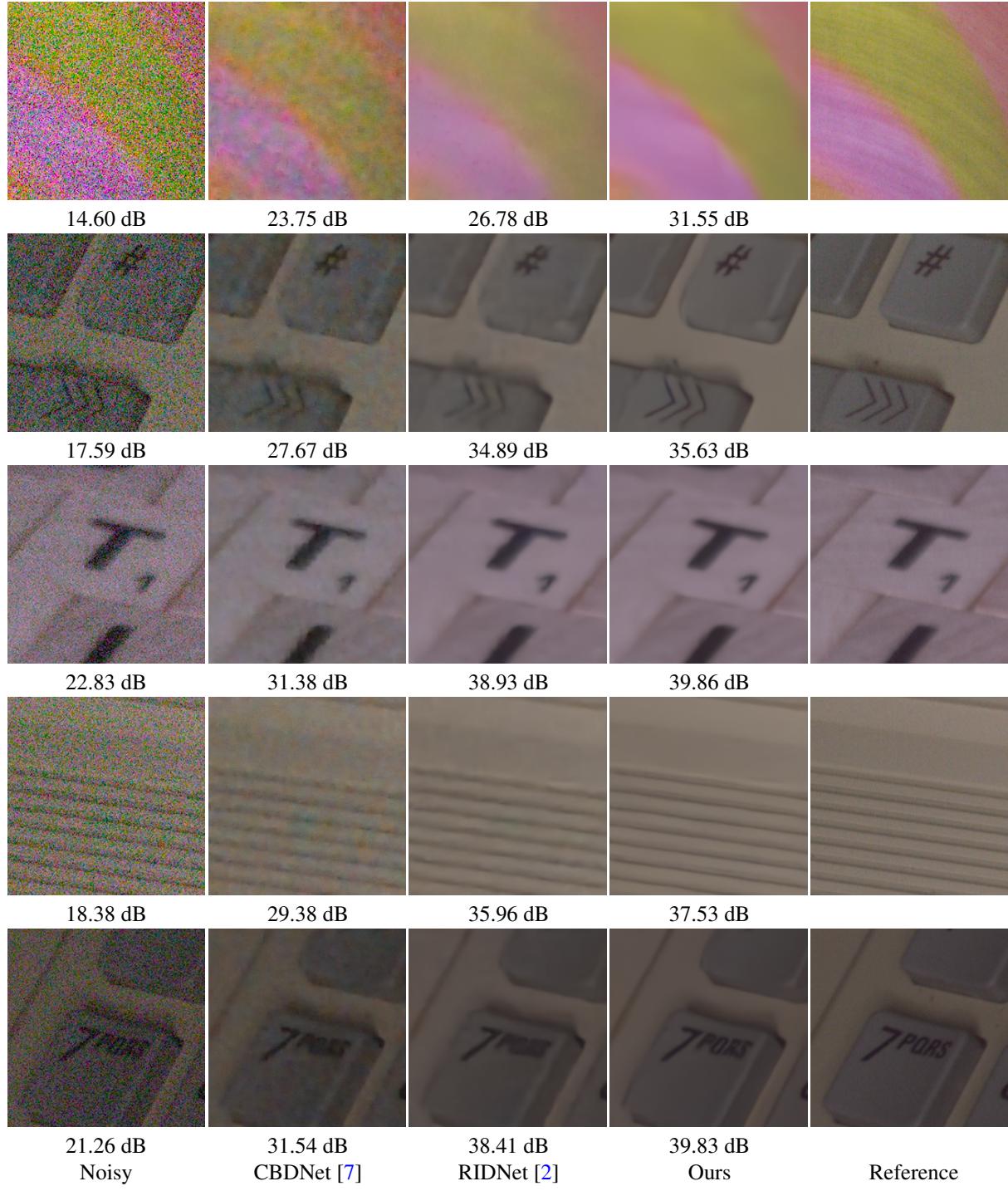


Figure 3: Denoising sRGB images from the SIDD benchmark dataset [1]. Compared to the state-of-the-art [2, 7], our method provides results that are visually pleasant and closer to the ground-truth, as well as have better PSNR.

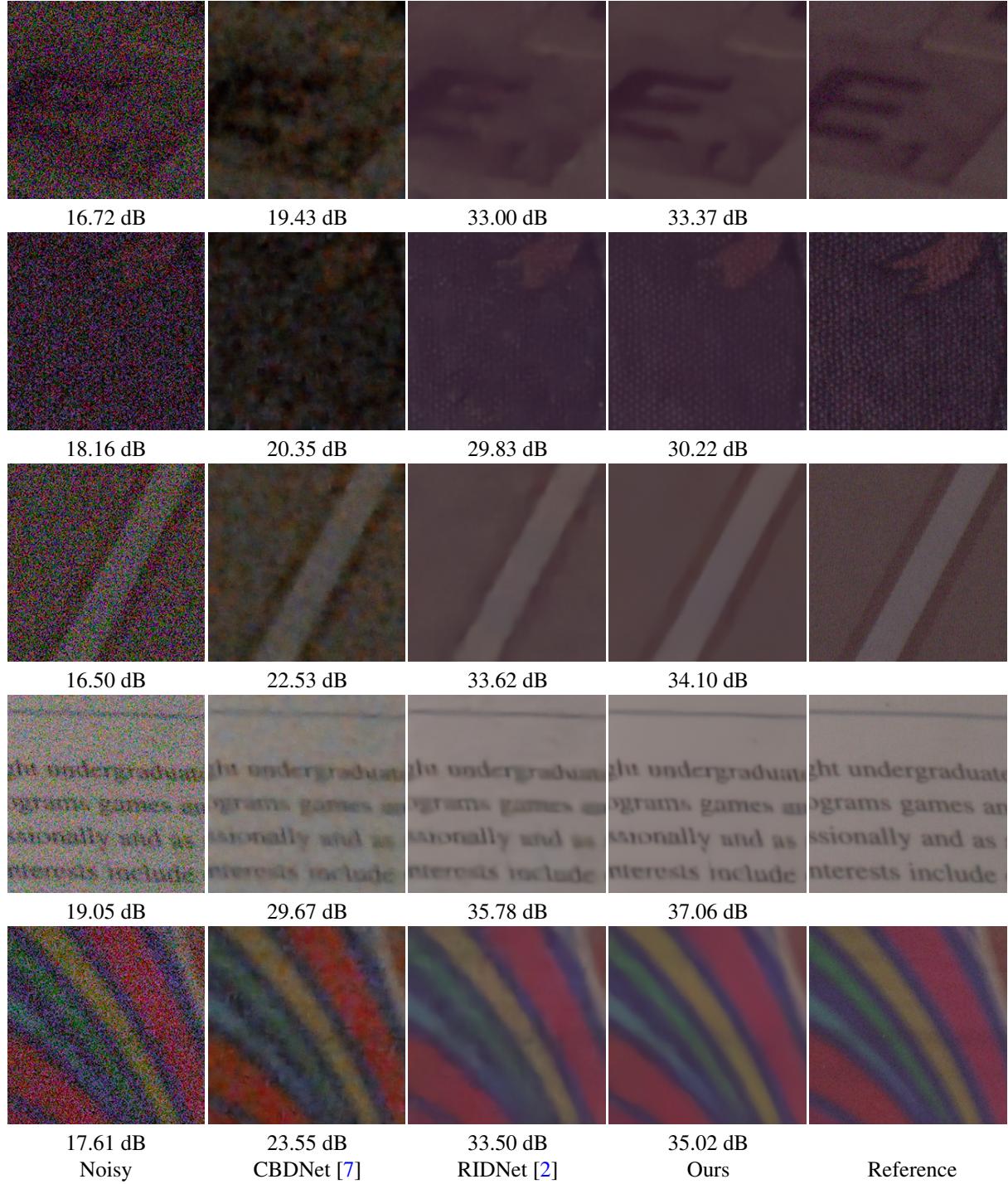


Figure 4: Denoising results of different methods on challenging sRGB images from the SIDD dataset [1].



Figure 5: Denoising RAW images from the DND benchmark dataset [11]. The PSNR scores for all competing methods are obtained from the website of the DND evaluation server [5]. For better visualization, RAW images are converted to the sRGB color space by the server [5] using the camera imaging pipeline of [8].



(a) Target view. (PSNR)



(b) Source view. 30.01 dB



(c) Reinhard *et al.* [13]. 34.44 dB



(d) Kotera [9]. 32.48 dB



(e) Pitié *et al.* [10]. 34.59 dB



(f) Ours. 36.86 dB

Figure 6: Example of color matching in stereo pairs. The colors of source view are matched to the target view by different algorithms. Compare the sky and clouds. Images are property of Mammoth HD Inc.



(a) Target view. (PSNR)



(b) Source view. 39.56 dB



(c) Reinhard *et al.* [13]. 23.29 dB



(d) Kotera [9]. 39.61 dB



(e) Pitié *et al.* [10]. 39.80 dB



(f) Ours. **43.62 dB**

Figure 7: Example of color matching in stereo pairs. The colors of source view are matched to the target view by different algorithms. Compare the cheek of the hippo. Images are property of Mammoth HD Inc.