Asymmetric Color Transfer with Consistent Modality Learning Supplementary Material

Kaiwen Zheng* Jie Huang* Man Zhou Feng Zhao†
University of Science and Technology of China
{kezh,hj0117,manman}@mail.ustc.edu.cn, fzhao956@ustc.edu.cn

1. Discussion

Why Color Transfer instead of Super-resolution? Given a mono-color image pair, there are two methods that can be adopted to obtain high-quality images: stereo color transfer and stereo super-resolution (SR). As for why most studies based on mono-color system prefer color transfer, we have conducted comparative experiments to investigate and explain. In the experiment, we designed a reference-based stereo SR method based on the representative stereo SR method iPASSR [9] to transfer the detail information of mono image to color image. The visual comparison results are shown in Fig. 1. It can be seen that the image restored by the SR-based method is unstable, resulting in unacceptable visual effects. Especially in the occluded areas caused by stereo disparity, the SR-based method is difficult to infer the non-existent texture of the reference image, while color transfer-based methods can make reasonable inference for the occluded areas.

2. Generalization Experiments

In this section, we supplement more generalization experiments to prove the progressiveness of our method.

KITTI dataset. We combine the KITTI 2012 [4] and KITTI 2015 [6] testsets to construct a KITTI testset for comparing the generalization capability of all models trained on the Flickr1024 dataset. Specifically, we constructed the KITTI testset containing 160 pairs of stereo images for asymmetric color transfer using the same process as the Flickr1024 dataset. The Quantitative results shown in Table 3 indicates that our model provides the best generalization performance. In addition, we also supplemented the qualitative experimental results in Fig. 2.

Large-size generalization. Since the image size of the training set (160×480) is small, we further conducted a generalization experiment on the large-size (1024×640) testset. Compared with the previous SOTA method [8], we achieved better generalization effect, as shown in Table 2.

*Both authors contributed equally to this research.
†Corresponding author.

Table 1. Comparing the complexity of the model with previous SOTA [8], the experimental environment is the NVIDIA RTX3090 GPU, and the input image size is 160×480.

Method	Flops (G)	Runtime (ms)
Wang (2021)	103.7524	50
Ours	143.8076	90

3. Low-light Condition Application

Under low-light conditions, the color sensor will inevitably cause noise for better photosensitive effect. Consequently, multi-sensor system joint imaging is an essential technique to enhance the imaging quality of low-light scenes. To prove the efficiency and superiority of our method, we simulated the low-light dataset with noise-added to the Flickr1024 dataset. Fig. 3 shows the visualization results of color transfer under low-light conditions. In the quantitative experiment, the average PSNR / SSIM of our method is 29.39dB / 0.924, which is better than the 28.49dB / 0.913 of the previous SOTA method [8].

4. Model Complexity

Although the dual-branch information complementary learning mechanism allows the model to establish more accurate binocular correspondence, the modulation operation also increases the model complexity. As shown in Table 1, our method has higher flops than the previous SOTA method proposed by Wang et al. [8], which is a drawback of our method.

5. More Qualitative Comparisons

As show in Fig. 4, we present more qualitative comparative results over the Flickr1024 dataset.

References

[1] Canqiang Chen, Chunmei Qing, Xiangmin Xu, and Patrick Dickinson. Cross parallax attention network for stereo image

(a) Mono. image (b) Color image (c) Color Transfer (d) Super-resolution (e) Left-view GT (f) Right-view GT

Figure 1. Comparison results between stereo color transfer and stereo super-resolution. It can be seen that the super-resolution results are unstable, showing many blurred areas, while the result of color transfer presents a better visual effect.

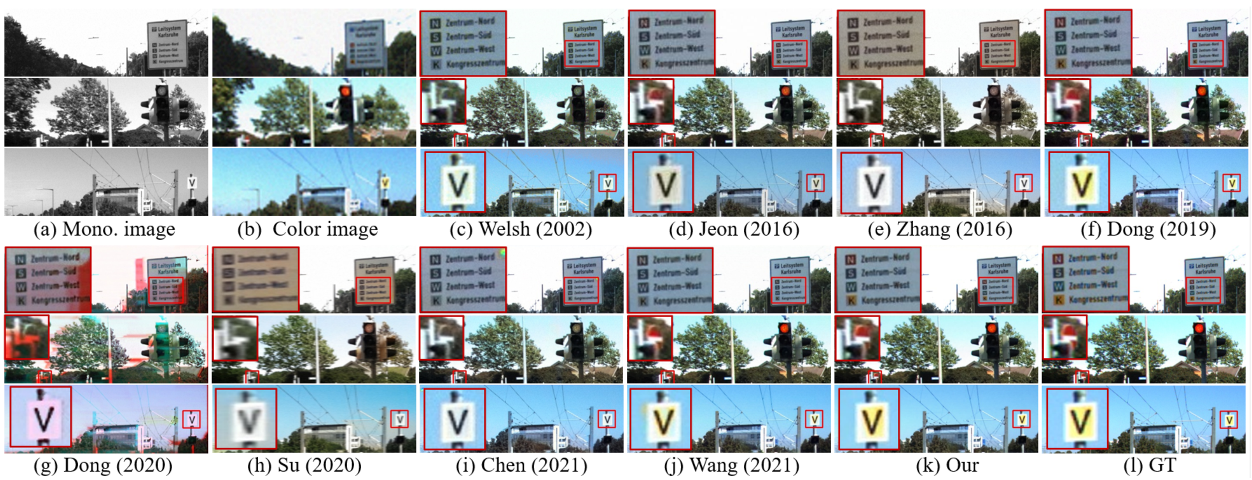

Figure 2. Visualization results on the KITTI dataset. As can be seen, our model gives the best generalization results.

Table 2. Generalization experiment over large size images.

Method	Flickr1024 (PSNR/SSIM)	
	160 × 480	1024 × 640
Wang (2021)	30.50dB/0941	25.29dB/0.910
Ours	31.10dB/0.948	27.10dB/0.930

super-resolution. *IEEE Transactions on Multimedia*, 2021. 3

[2] Xuan Dong, Weixin Li, Xiaojie Wang, and Yunhong Wang. Learning a deep convolutional network for colorization in monochrome-color dual-lens system. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 8255–8262, 2019. 3

[3] Xuan Dong, Weixin Li, Xiaojie Wang, and Yunhong Wang. Cycle-cnn for colorization towards real monochrome-color camera systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10721–10728, 2020. 3

[4] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, 2012. 1

[5] Hae-Gon Jeon, Joon-Young Lee, Sunghoon Im, Hyowon Ha, and In So Kweon. Stereo matching with color and monochrome cameras in low-light conditions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4086–4094, 2016. 3

[6] Moritz Menze and Andreas Geiger. Object scene flow for autonomous vehicles. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3061–3070, 2015. 1

[7] Jheng-Wei Su, Hung-Kuo Chu, and Jia-Bin Huang. Instance-aware image colorization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7968–7977, 2020. 3

[8] Yicheng Wang, Jiayong Peng, Yueyi Zhang, Shan Liu, Xiaoyan Sun, and Zhiwei Xiong. Asymmetric stereo color

Table 3. Quantitative generalization comparison results on KITTI testset.

Model	Setup1		Setup2		Setup3		#Param (M)
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	
Welsh (2002) [10]	25.76	0.831	25.41	0.789	25.31	0.777	-
Jeon (2016) [5]	28.06	0.901	27.81	0.865	27.62	0.862	-
Zhang (2016) [11]	25.76	0.896	25.42	0.861	25.42	0.861	32.60
Dong (2019) [2]	28.93	0.903	28.16	0.867	28.23	0.868	2.30
Su (2020) [7]	24.16	0.823	23.99	0.816	23.99	0.816	34.18
Dong (2020) [3]	18.58	0.780	18.82	0.761	18.61	0.758	0.12
Chen (2021) [1]	27.95	0.903	27.73	0.883	27.59	0.879	2.99
Wang (2021) [8]	29.11	0.909	28.77	0.892	28.69	0.890	1.35
Ours	**29.21**	**0.910**	**28.92**	**0.895**	**28.90**	**0.894**	1.09

(a) Mono. image (b) Color image (c) Wang (2021) (d) Our results (e) Ground Truth

Figure 3. Colorization over noise-added low-light scene. It can be seen that our method achieves accurate color transfer under extreme conditions, which is better than the previous state-of-the-art method [8].

transfer. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, pages 1–6, 2021. 1, 3

[9] Yingqian Wang, Xinyi Ying, Longguang Wang, Jungang Yang, Wei An, and Yulan Guo. Symmetric parallax attention for stereo image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 766–775, 2021. 1

[10] Tomihisa Welsh, Michael Ashikhmin, and Klaus Mueller. Transferring color to greyscale images. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, pages 277–280, 2002. 3

[11] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *Proceedings of the European Conference on Computer Vision*, pages 649–666. Springer, 2016. 3

(a) Mono. image (b) Color image (c) Welsh (2002) (d) Jeon (2016) (e) Zhang (2016) (f) Dong (2019)

(g) Dong (2020) (h) Su (2020) (i) Chen (2021) (j) Wang (2021) (k) Ours (l) GT

Figure 4. Visualization results on the Flickr1024 testset.