Supplementary Material for "Low-light Color Imaging via Dual Camera Acquisition"

Peiyao Guo[1][0000−0003−2887−3463] and Zhan Ma[1][0000−0003−3686−4057]

Nanjing University, Nanjing, China
peiyao@smail.nju.edu.cn, mazhan@nju.edu.cn

In this supplementary material, we provide specific technical details and extra results visualization for the main paper.

1 Network Architecture

We summarize each module of our pipeline in Table A. More details about PWCNet module could be found in [1][1]. Besides, the slicing layer interpolates the reshaped color coefficients (layer 30) with the guidance map (layer 32) in a bilateral-grid upsampling way [2]. Afterwards, the color coefficients at high resolution are applied to the input monochromic image $I^{\mathrm{Y}}_{v_2,s_H,l_H}$ in the form of affine combination via the applying coefficients layer for chrominance reconstruction.

2 Training Details

Details on training process are listed in Table B as the supplementary of Table 1 and section 4.1 in the body paragraphs. Losses in Table B are in the form as,

$$\text{L1 loss:} \quad \ell(y,\hat{y}) = \frac{\sum_{i=1}^{n} \mid y_i - \hat{y}_i \mid}{n}$$

$$\text{MSE loss:} \quad \ell(y,\hat{y}) = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}$$

$$\text{Cosine similarity loss:} \quad \text{similarity}(y,\hat{y}) = \frac{y \cdot \hat{y}}{\max(\|y\|_2 \cdot \|\hat{y}\|_2, \epsilon)},$$

$$\ell(y,\hat{y}) = mean(\frac{1 - \text{similarity}(y,\hat{y})}{2})$$

3 Ablation Study

We have divided the overall problem defined in this paper using three consecutive operational subtasks, i.e., RefEC for illumination compensation, RefColor for image alignment and color transfer, and RefSR for final spatial resolution enhancement. Visualizations with partial subtasks disabled are shown in Fig. 1 while Table C demonstrates the quantitative evaluation.

[1] https://github.com/NVlabs/PWC-Net

Table A. Network configuration for our workflow.

	Layer Description	Output Tensor Dim.
	Input color image $I^{\mathrm{YUV}}_{v_1,s_L,l_L}$	$3 \times \frac{h}{4} \times \frac{w}{4}$
	Input monochrome image $I^{\mathrm{Y}}_{v_2,s_H,l_H}$	$h \times w$
	RefEC net	
1	9×9 conv	$64 \times \frac{h}{4} \times \frac{w}{4}$
2-6	$(3 \times 3$ conv, ReLU$) \times 5$	$64 \times \frac{h}{4} \times \frac{w}{4}$
7	3×3 conv, Sigmoid, extra bias	$3 \times \frac{h}{4} \times \frac{w}{4}$
	RefColor net	
8	PWCNet module	$2 \times \frac{h}{16} \times \frac{w}{16}$
9	upsampling & warping layer	$3 \times \frac{h}{4} \times \frac{w}{4}$
10	9×9 conv	$64 \times \frac{h}{4} \times \frac{w}{4}$
11	3×3 conv, Batch Normalization , ReLU	$64 \times \frac{h}{4} \times \frac{w}{4}$
12	3×3 conv, Batch Normalizatoin	$64 \times \frac{h}{4} \times \frac{w}{4}$
13	skip connection between layer 10 and 12, ReLU	$64 \times \frac{h}{4} \times \frac{w}{4}$
24	repeat (layer 11,12,13) $\times$ 4	$64 \times \frac{h}{4} \times \frac{w}{4}$
25	3×3 conv, Sigmoid, extra bias	$3 \times \frac{h}{4} \times \frac{w}{4}$
	RefSR net	
26	3×3 conv (stride 2), ReLU	$8 \times \frac{h}{8} \times \frac{w}{8}$
27	3×3 conv (stride 2), ReLU	$16 \times \frac{h}{16} \times \frac{w}{16}$
28	3×3 conv (stride 2), ReLU	$32 \times \frac{h}{32} \times \frac{w}{32}$
29	3×3 conv (stride 2), ReLU	$64 \times \frac{h}{64} \times \frac{w}{64}$
30	global & local stream fusion with 1×1 conv, ReLU	$96 \times \frac{h}{64} \times \frac{w}{64}$
31	3×3 conv, ReLU	$16 \times h \times w$
32	1×1 conv, Tanh	$1 \times h \times w$
33	slicing layer	$12 \times h \times w$
34	applying coefficients layer	$3 \times h \times w$
35	3×3 conv, ReLU	$16 \times h \times w$
36	1×1 conv, Tanh	$1 \times h \times w$
37	skip connection between layer 25 and 36	$3 \times h \times w$
	Local stream in layer 30	
	3×3 conv, ReLU	$64 \times \frac{h}{64} \times \frac{w}{64}$
	3×3 conv, w/o ReLU	$64 \times \frac{h}{64} \times \frac{w}{64}$
	Global stream in layer 30	
	3×3 conv (stride 2), ReLU	$64 \times \frac{h}{128} \times \frac{w}{128}$
	3×3 conv (stride 2), ReLU	$64 \times \frac{h}{256} \times \frac{w}{256}$
	FC(1024,256)	256
	FC(256,128)	128
	FC(128,64)	64

Table B. Training details

Module	Input-color	Input-monochrome	GT-color	Loss
RefEC	(v_1, s_L, l_L)	(v_2, s_L, l_H)	(v_1, s_L, l_H)	L1+ CosineSimlarity loss
RefColor	(v_1, s_L, l_H)	(v_2, s_L, l_H)	(v_2, s_L, l_H)	L1+ CosineSimlarity loss
RefSR	(v_2, s_L, l_H)	(v_2, s_H, l_H)	(v_2, s_H, l_H)	MSE+ CosineSimlarity loss
Overall	(v_1, s_L, l_L)	(v_2, s_H, l_H)	(v_2, s_H, l_H)	L1+ CosineSimlarity loss
description	v, s, l represent the viewpoint, spatial resolution and light engergy specification of the image while detailed introduction is in the Table 1 of the manuscript. The loss in each task is calculated between the GT and corresponding predicition. For faster convergence, we use MSE loss in the RefSR.			

Table C. Ablation study

	PSNR_{YUV}/dB	PSNR_{UV}/dB	MS-SSIM
w/o RefEC	37.22	35.51	0.9773
w/o RefColor	31.94	30.20	0.9552
Overall	**38.60**	**38.91**	**0.9804**

4 User Study

We randomly pick up 5 cases for qualitative performance validation with 25 subjects. First, we investigate the user preference of reconstrcuted image quality at each intermediate step by asking the subjects to pickup their favorites. We enforce the random display order of algorithms used in each case. Results are shown in 2, revealing that our proposed algorithms are preferred in each subtask.

We further invite subjects to rate the final reconstructed images from 1 to 5 scales (e.g., representing the Bad, Poor, Fair, Good, Excellent levels) as suggested in ITU-R BT.500-11. As depicted in Fig. 3, the average score of reconstructed images reaches at 4.56, revealing the perceptual quality that is close to the "Excellent" scale. This validates the efficiency of our algorithm as well.

5 Results Visualization

In this section, we present more details of the overall test performance of our proposal. Fig. 4 illustrates the qualitative results on the simulated dataset (Middlebury2014 [3]). And Fig. 5 shows the complete performance on the captured scene via industrial cameras and Huawei P20, which supplements Fig. 8 in the main paper. These visualization envidences our workflow's efficiency in different low-light condition. As referred in the last paragraph of Section 4, the only supervision of the re-colored HSR monochromic image influences the separative performance on refEC, refColor and refSR, inducing the slightly over-saturated

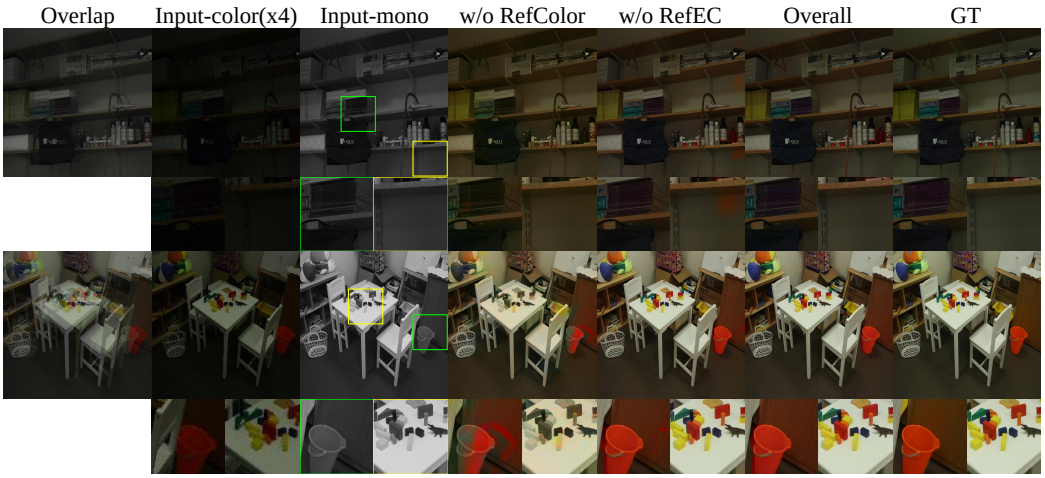

Fig. 1. Visual performance with partial subtasks disabled

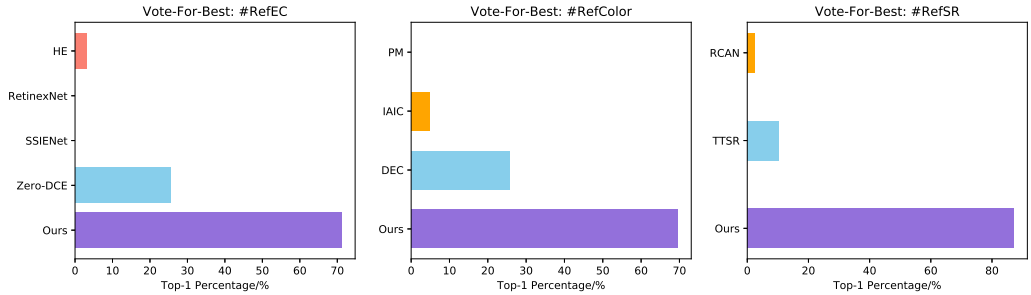

Fig. 2. Subjective preference distribution of each sub-task

color which is shown in 4^{th} row of Fig. 4. Besides, the color bleeding caused via the ill-measured gray channel correlation affects the final reconstruction's quality as shown in the third row of Fig. 4.

References

1. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018) 8934–8943
2. Gharbi, M., Chen, J., Barron, J.T., Hasinoff, S.W., Durand, F.: Deep bilateral learning for real-time image enhancement. ACM Transactions on Graphics (TOG) **36** (2017) 118
3. Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X., Westling, P.: High-resolution stereo datasets with subpixel-accurate ground truth.

Fig. 3. Subjective score distribution of 5 test cases

In Jiang, X., Hornegger, J., Koch, R., eds.: Pattern Recognition, Cham, Springer International Publishing (2014) 31–42

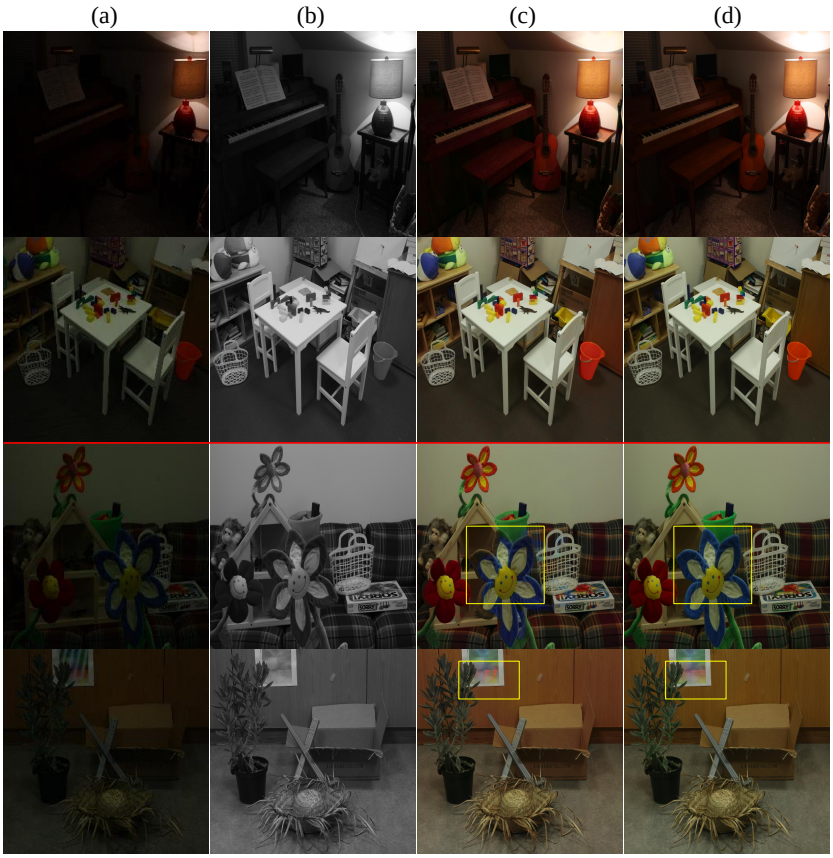

Fig. 4. Overall simulation on Middlebury2014: (a) is the input color image with poor illumination at the low resolution $I^{\mathrm{YUV}}_{v_1,s_L,l_L}$. ($\times 4$ for illusration) (b) represents the monochrome input at the high resolution $I^{\mathrm{Y}}_{v_2,s_H,l_H}$. (c) and (d) show the prediction and the ground truth of the color image $I^{\mathrm{YUV}}_{v_2,s_H,l_H}$.

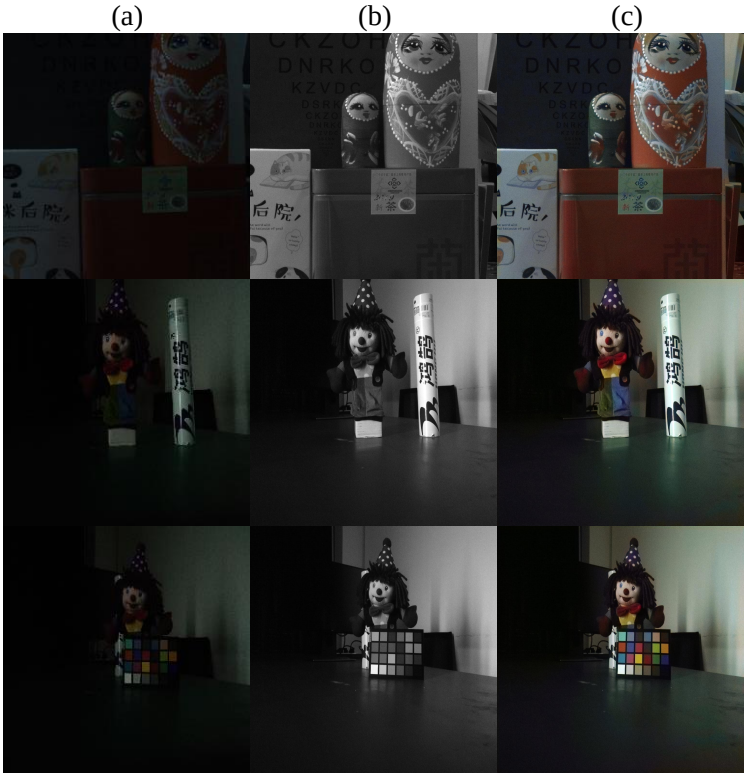

Fig. 5. Overall performance on captured scenes: (a) is the input color image with poor illumination at the low resolution $I^{\mathrm{YUV}}_{v_1,s_L,l_L}$. ($\times 4$ for illusration) (b) represents the monochrome input at the high resolution $I^{\mathrm{Y}}_{v_2,s_H,l_H}$. (c) presents the final reconstruction of the color image $I^{\mathrm{YUV}}_{v_2,s_H,l_H}$.