

Supplementary Material

1. Quantitative Evaluations on Cube dataset

Table 1 lists our results (based on global fitting) on Cube dataset [3]. Cube datasets contains 1365 exclusively outdoor images taken by Canon EOS 550D camera in parts of Croatia, Slovenia, and Austria during various seasons. All static methods are reported using the best results from the several parameters tested based. The implementation of FFCC [5] and SqueezeNet-FC4 [9] are based on their released code ¹². In the lower right corner of each image in Cube dataset, the SpyderCube calibration [3] object is placed. It is two neutral 18% gray faces are used to determine the ground truth illumination for each image. In all dataset, when images contain two distinct illuminants, one of them is usually dominant so that the uniform illumination assumption effectively remains valid. To correctly identify the dominant illumination, it is two possible chromatically adapted versions were manually checked for each image to create final ground truth illumination. For more detail about Cube dataset, please refer to ³.

Method	Mean	Med.	Tri.	Best 25%	Worst 25%	G.M.
White-Patch [6]	6.58	4.48	5.27	1.18	15.23	4.88
Gray-world [7]	3.75	2.91	3.15	0.69	8.18	2.87
Color Tiger [3]	2.94	2.59	2.66	0.61	5.88	2.35
Shades-of-Gray [8]	2.58	1.79	1.95	0.38	6.19	1.84
2nd-order Gray-Edge [10]	2.49	1.60	1.80	0.49	6.00	1.84
1st-order Gray-Edge [10]	2.45	1.58	1.81	0.48	5.89	1.81
General Gray-World [4]	2.50	1.61	1.79	0.37	6.23	1.76
Restricted Color Tiger [3]	1.64	0.82	1.05	0.24	4.37	1.08
Color Dog [1]	1.50	0.81	0.99	0.27	3.86	1.05
Smart Color Cat [2]	1.49	0.88	1.06	0.24	3.75	1.04
SqueezeNet-FC4 [9]	1.45	0.88	1.02	0.22	3.26	0.81
FFCC [5]	1.43	0.86	1.00	0.21	3.50	0.79
Our method	1.39	0.75	0.88	0.14	3.23	0.67

Table 1. Results on the Cube dataset. For each evaluation metric, the best result is highlighted with bold type.

2. Variants of Architectures

Besides, we explore the proposed framework with various network input size and kernel size. We summarize the comparison among variants on MIO dataset in Table 2, MIO dataset contains self-made massive indoor and outdoor

¹<https://github.com/google/ffcc>

²<https://github.com/yuanming-hu/fc4>

³https://ipg.fer.hr/ipg/resources/color_constancy

Structure input size / kernel size	Mean	Med.	Tri.	Best 25%	Worst 25%	G.M.
32x32x3/ 1x1x1x3	1.28	1.01	1.19	0.48	2.90	0.93
32x32x3 / 1x1x3x3	0.88	0.59	0.63	0.25	2.21	0.54
64x64x3 / 1x1x1x3	1.09	0.76	0.79	0.35	2.20	0.69
64x64x3 / 1x1x3x3	0.78	0.61	0.63	0.23	1.69	0.54
64x64x3 / 3x3x1x3	1.01	0.74	0.78	0.36	2.31	0.70
64x64x3 / 3x3x3x3	0.74	0.59	0.62	0.24	1.62	0.55
128x128x3 / 1x1x1x3	1.23	0.91	0.99	0.40	2.61	0.89
128x128x3 / 1x1x3x3	0.98	0.68	0.74	0.24	2.29	0.64
128x128x3 / 3x3x1x3	0.93	0.75	0.79	0.32	2.19	0.73
128x128x3 / 3x3x3x3	0.81	0.56	0.61	0.24	1.95	0.57
128x128x3 / 5x5x1x3	0.93	0.75	0.78	0.34	1.97	0.69
128x128x3 / 5x5x3x3	0.72	0.59	0.60	0.23	1.61	0.53

Table 2. Results of variant architectures on MIO dataset. For each evaluation metric, the best result is highlighted with bold type.

images (13k) captured from Sony DSC-RX100M3 camera. The ground truth is obtained from the meta-file generated by DSC-RX100M3 camera. We split the training, validation and test sets as 9k, 2k, 2k images respectively.

We compare the different network architectures on three settings: $32 \times 32 \times 3$, $64 \times 64 \times 3$ and $128 \times 128 \times 3$. For input size with $32 \times 32 \times 3$ and $64 \times 64 \times 3$, the network architecture only has one pair of downsampling-upsampling module. For input size with $128 \times 128 \times 3$, two pairs of downsampling-upsampling module are used. Channel size will be doubled when downsampling and halved when upsampling. Our best results is obtained under input size of $128 \times 128 \times 3$ with filter size of $5 \times 5 \times 3 \times 3$, Figure 1 and 2 show some examples on Gehler, NUS-8, Cube, and MIO dataset. A gamma correction of $\gamma = 1/2.2$ to linear RGB images is applied for display. Moreover, benefit from our efficient design of network architecture, the architecture with input size $64 \times 64 \times 3$ and $32 \times 32 \times 3$ also exhibit the comparable performance which requires much less computational cost. e.g. [9] uses $512 \times 512 \times 3$ as input.

References

- [1] N. Banic and S. Loncaric. Color dog-guiding the global illumination estimation to better accuracy. In *VISAPP (1)*, pages 129–135, 2015.
- [2] N. Banić and S. Lončarić. Using the red chromaticity for illumination estimation. In *Image and Signal Processing and Analysis (ISPA), 2015 9th International Symposium on*, pages 131–136. IEEE, 2015.

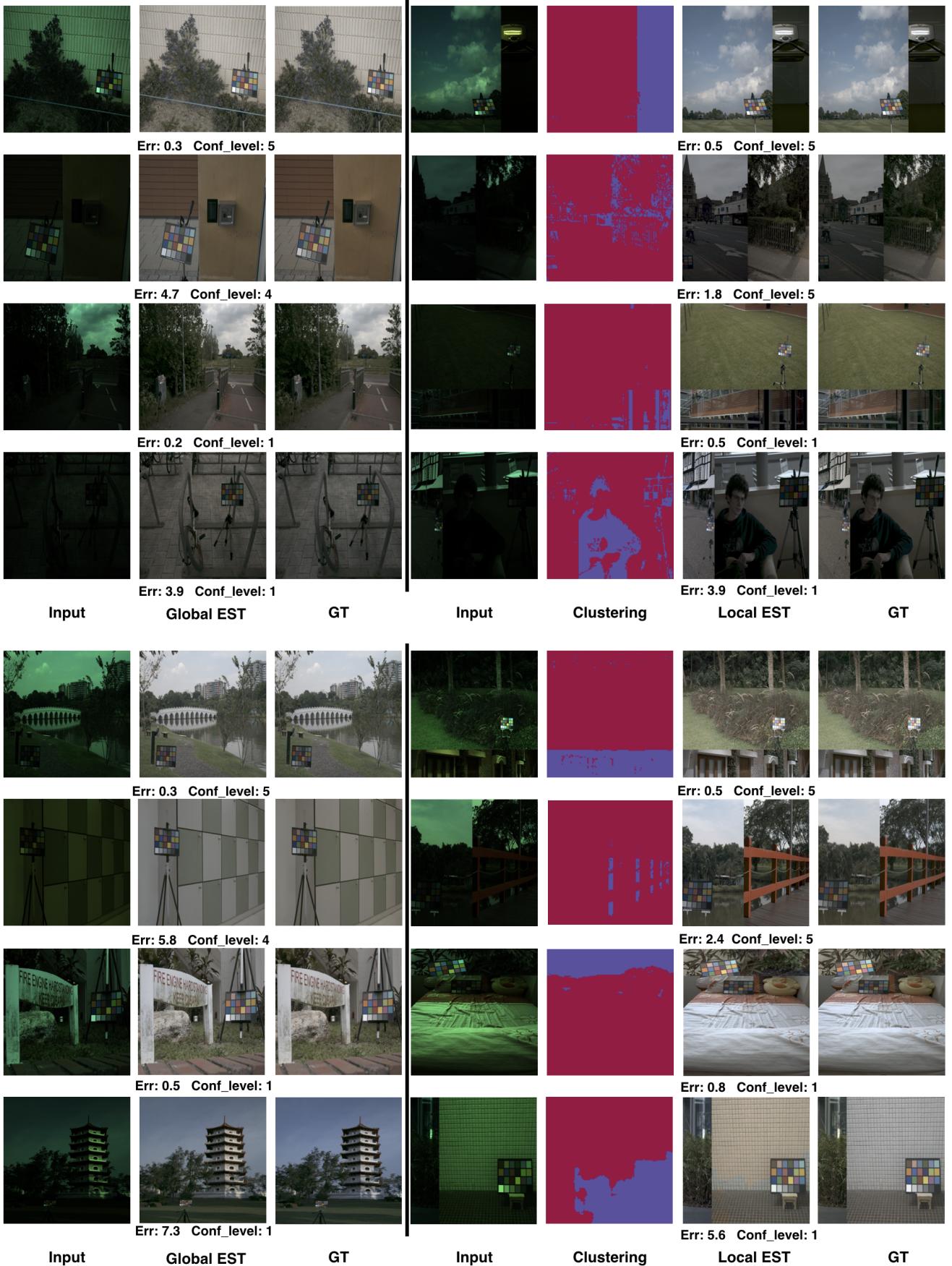


Figure 1. Examples of single (**left**) and multi-illuminant (**right**) White Balance (WB) results on Gehler and NUS-8 dataset. Single illuminant WB is based on global fitting. The top-down order is organized as high confidence, small error; high confidence, big error; low confidence, small error; low confidence big error.

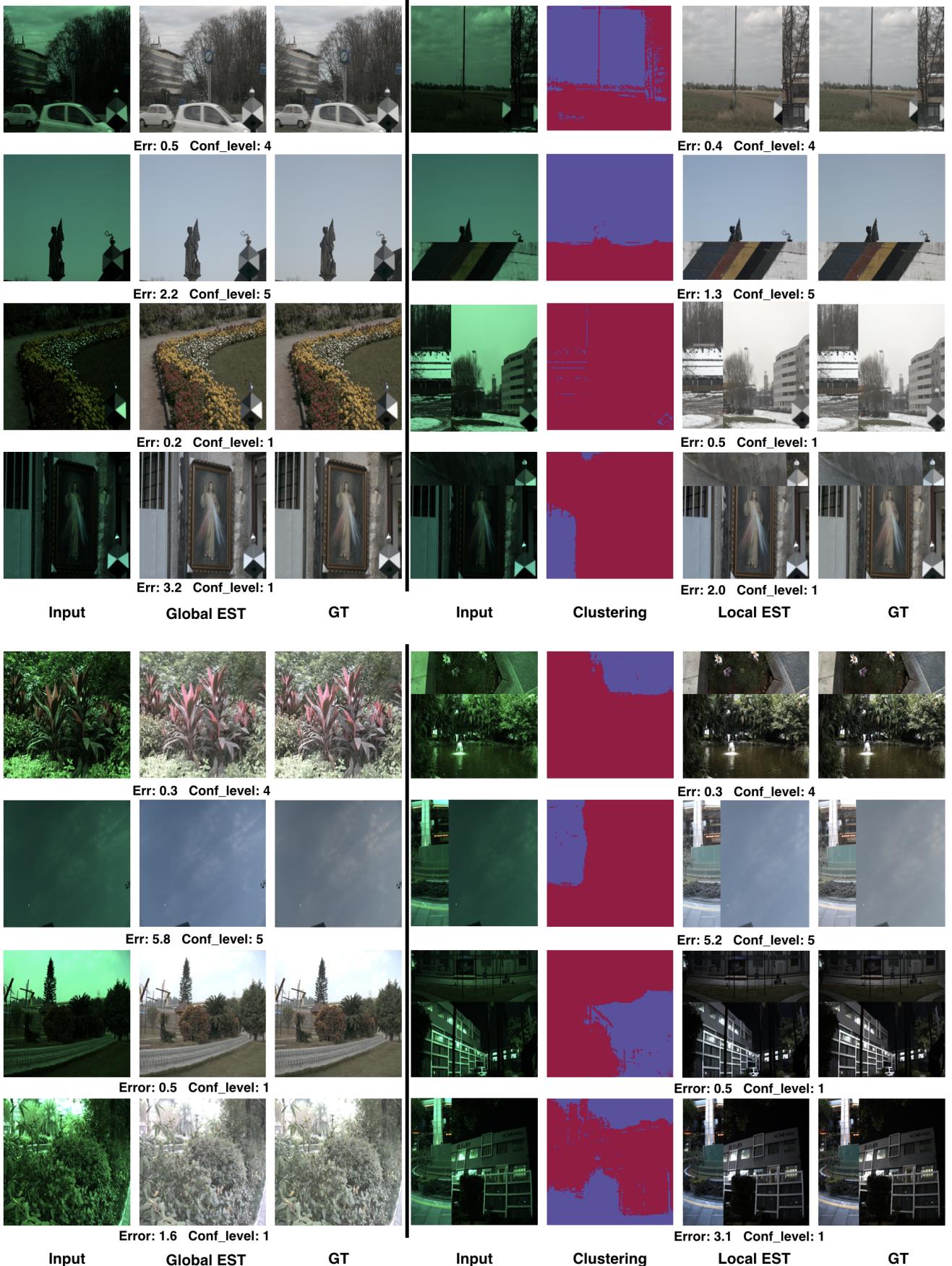


Figure 2. Examples of single (**left**) and multi-illuminant (**right**) WB results on Cube and MIO dataset, Single illuminant WB is based on global fitting. The top-down order is organized as high confidence, small error; high confidence, big error; low confidence, small error; low confidence big error.

- [3] N. Banic and S. Loncaric. Unsupervised learning for color constancy. *CoRR*, abs/1712.00436, 2017.
- [4] K. Barnard, V. Cardei, and B. Funt. A comparison of computational color constancy algorithms. i: Methodology and experiments with synthesized data. *IEEE transactions on Image Processing*, 11(9):972–984, 2002.
- [5] J. T. Barron and Y. Tsai. Fast fourier color constancy. *CoRR*, abs/1611.07596, 2016.
- [6] D. H. Brainard and B. A. Wandell. Analysis of the retinex theory of color vision. *JOSA A*, 3(10):1651–1661, 1986.
- [7] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1):1–26, 1980.
- [8] G. D. Finlayson and E. Trezzi. Shades of gray and colour constancy. In *Color and Imaging Conference*, volume 2004, pages 37–41. Society for Imaging Science and Technology, 2004.
- [9] Y. Hu, B. Wang, and S. Lin. Fc 4: Fully convolutional color constancy with confidence-weighted pooling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4085–4094, 2017.
- [10] J. Van De Weijer, T. Gevers, and A. Gijsenij. Edge-based color constancy. *IEEE Transactions on image processing*, 16(9):2207–2214, 2007.