Single Image Reflection Removal Exploiting Misaligned Training Data and Network Enhancements
(Supplementary Material)

Kaixuan Wei[1] Jiaolong Yang[2] Ying Fu[1*] David Wipf[2] Hua Huang[1]
[1]Beijing Institute of Technology [2]Microsoft Research

This supplementary material provides more details and results that were not included in the main paper due to the space limitations. The contents are organized as follows.

Section A details the procedures of human assessment in Section 4.4 of the main paper. Section B illustrates the complete overview of our newly collected unaligned dataset, followed by more visual results of real-world images in Section C.

A. Human Assessment on Unaligned Data

This section provides more details on how we conduct the user study to assess the model performance before and after finetuning with our alignment-invariant loss and unaligned training data. For each assessment, the user was presented with the reference R and the reconstructed images (A and B) by compared models. The user was asked to select the image that he/she considers more similar to the reference, as shown in Figure I.

We adopted two strategies to ensure the validation and reliability of this test. First, the output of a specific model was randomly placed, such that A or B could be either the result of the pretrained model or the finetuned version. Second, before the regular test, we indicated each user to complete a training phase. Ten standard training pairs (out of 50 testing pairs) with our predefined labels are presented in training phase. If the user selects the option that is entirely deviated from our label (*e.g.* selecting "B is obviously better" in Figure I), the system would require the user to reconsider the selection again. These labelled image pairs serve as our predefined evaluation baseline, as shown in Figure IV.

We quantify the human preference base on the following rules:

I. The option selected would be accounted for 2 point, as the finetuned result is obviously better.

II. 1 point, as the finetuned result is slightly better.

III. 0 point, as the "can't tell" option is selected.

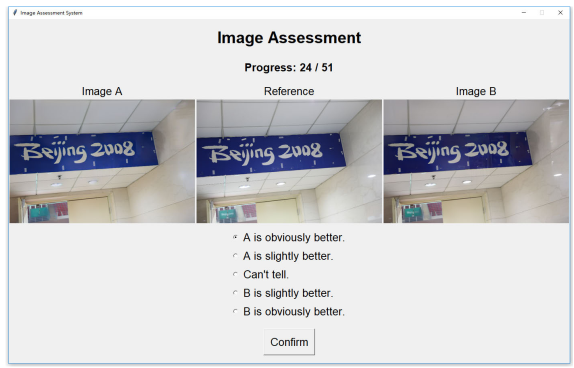


Figure I: Image assessment system interface

IV. -2, -1 points for the converse situations of rules 1 and 2 respectively.

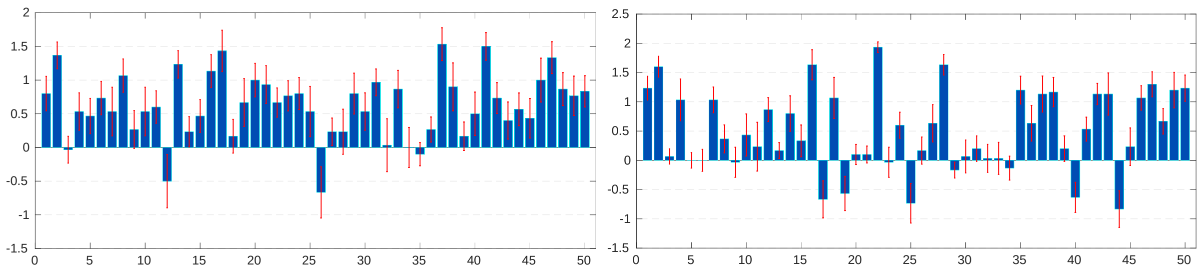
In Table 4 of the main paper, we displayed the (averaged) human preference scores of this user study. Here, we construct the corresponding 95% confidence interval as shown in Figure II. It can be observed that the endpoints of each confidence interval are not deviated too much from the sample mean, showing the reliability of the user study. More visual result comparison of our ERRNet models with and without training on unaligned data is shown in Figure III, which extends the Figure 7 in the main paper.

B. Overview of Dataset

The complete overview of our newly collected unaligned dataset is shown in Figure V, which extends the Figure 6 in the main paper.

C. More Qualitative Results

More visual results on real-world images are shown in Figure VI (3 pages). We compare our methods against state-of-the-art methods of Li and Brown (LB14) [2], Fan *et al.* (CEILNet) [1], Zhang *et al.* [4], and Yang *et al.* (BDN) [3]. This extends the Figure 5 of the main paper.



(a) BDN-F (b) ERRNet

Figure II: Human preference scores and their associated 95 % confidence interval (denoted by red error bar) of self-comparison experiments.

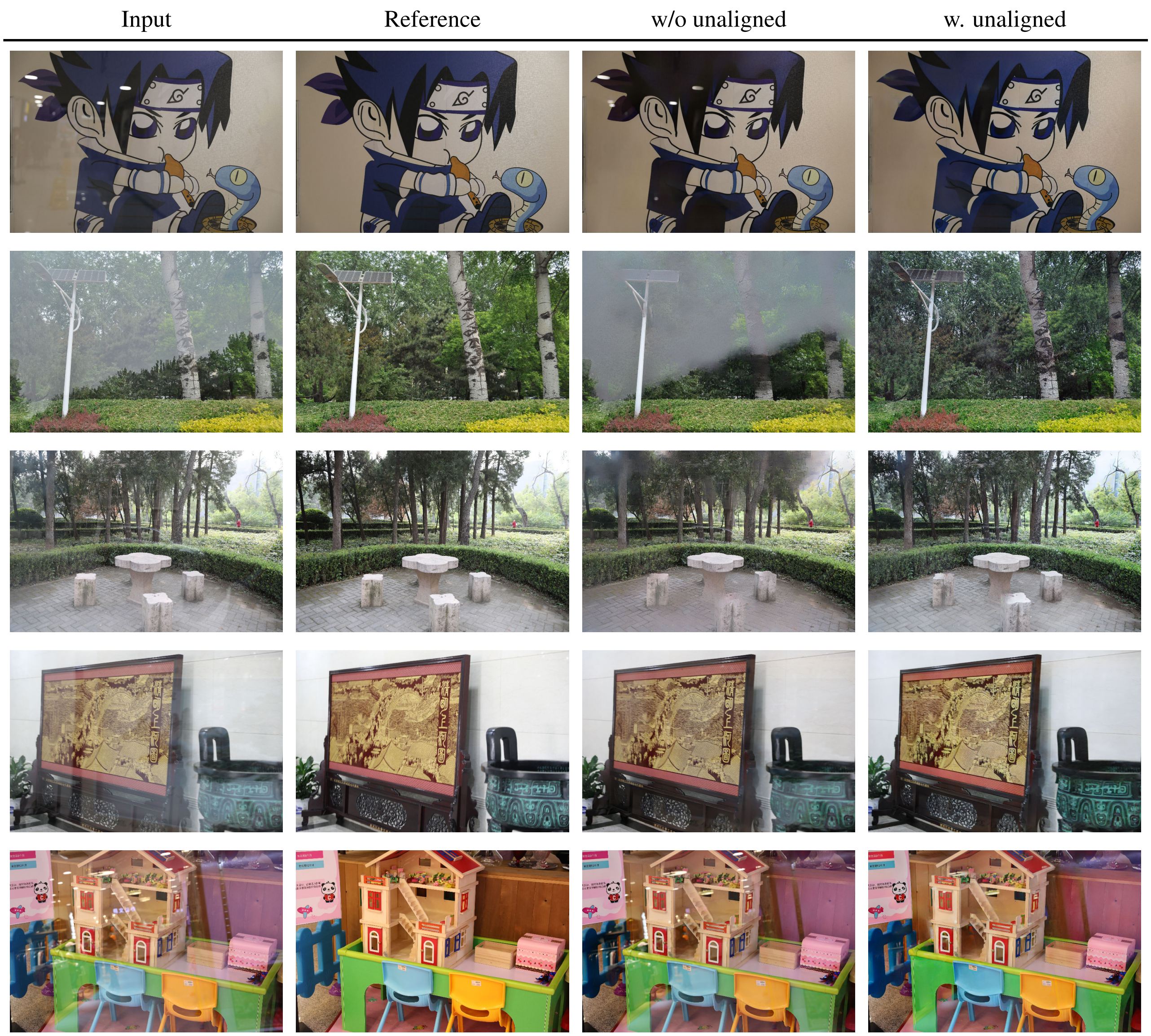


Figure III: Results of ERRNet that trained with and without unaligned data.

References

[1] Q. Fan, J. Yang, G. Hua, B. Chen, and D. Wipf. A generic deep architecture for single image reflection removal and image smoothing. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. 1

Figure IV: Ten standard training pairs with predefined labels

[2] Y. Li and M. S. Brown. Single image layer separation using relative smoothness. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2752–2759, 2014. 1

[3] J. Yang, D. Gong, L. Liu, and Q. Shi. Seeing deeply and bidirectionally: A deep learning approach for single image reflection removal. In *The European Conference on Computer Vision (ECCV)*, September 2018. 1

[4] X. Zhang, R. Ng, and Q. Chen. Single image reflection sep-

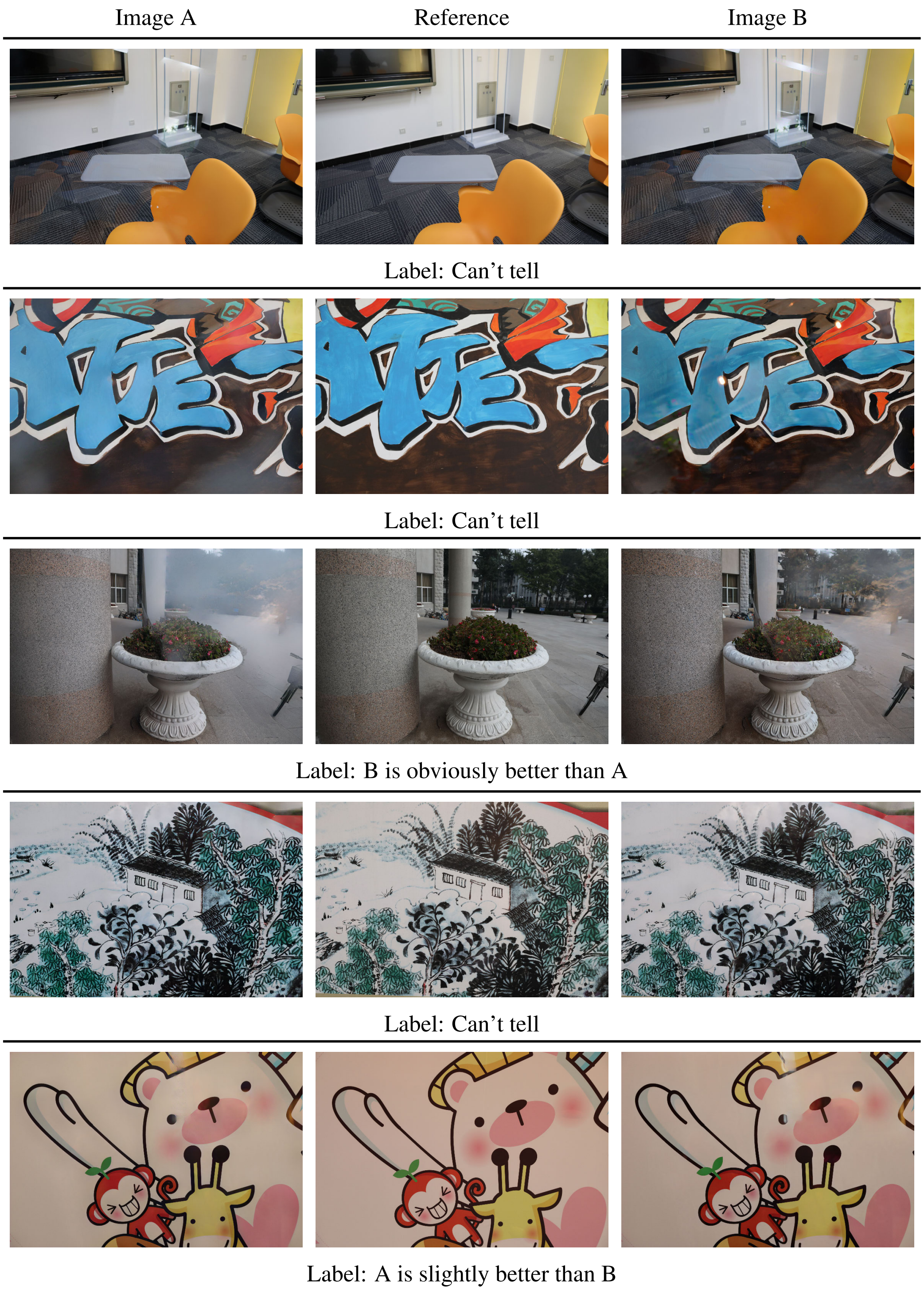


Figure IV: (Cont.) Ten standard training pairs with predefined labels

aration with perceptual losses. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 1

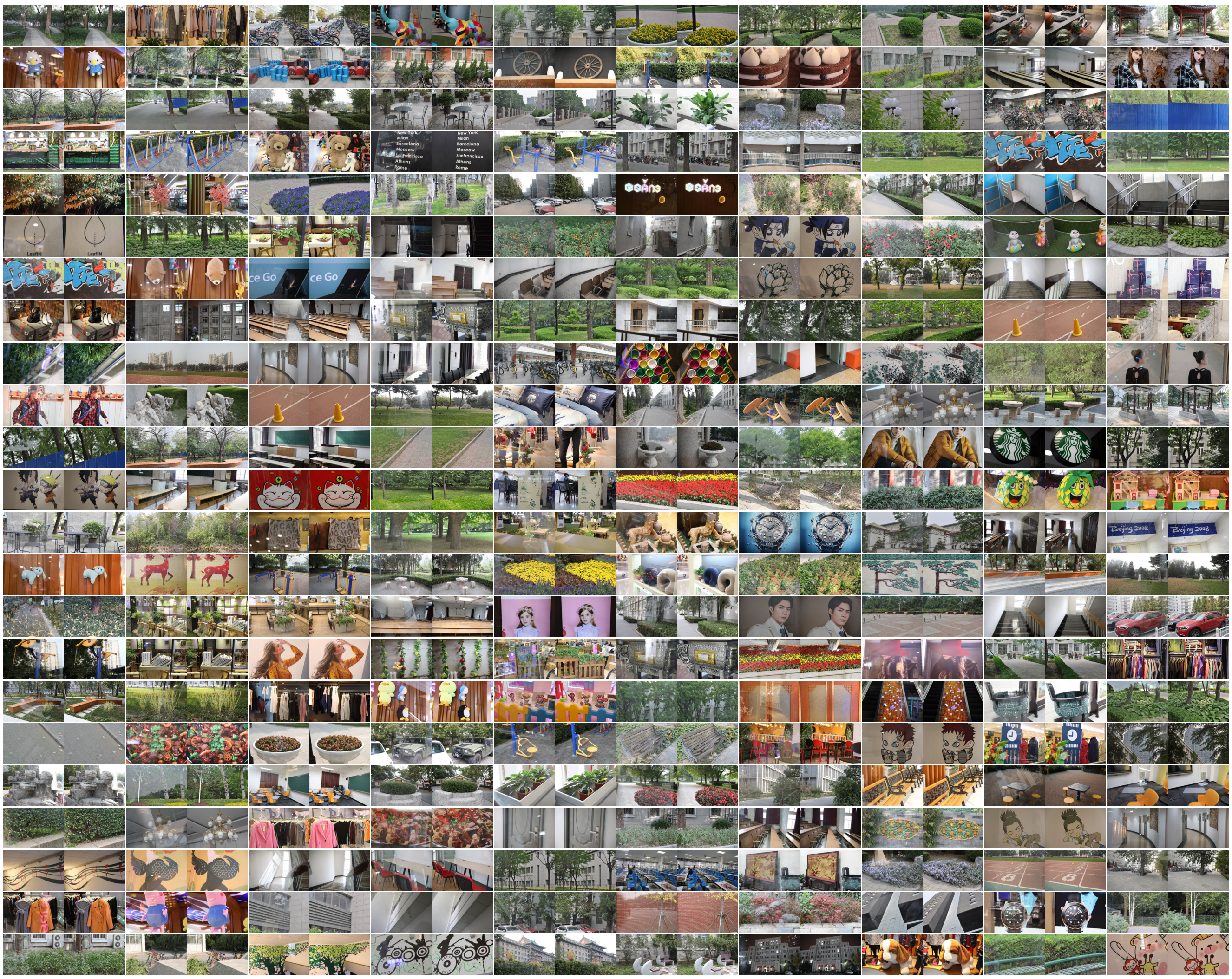


Figure V: Complete Overview of our unaligned dataset (DSLR part)

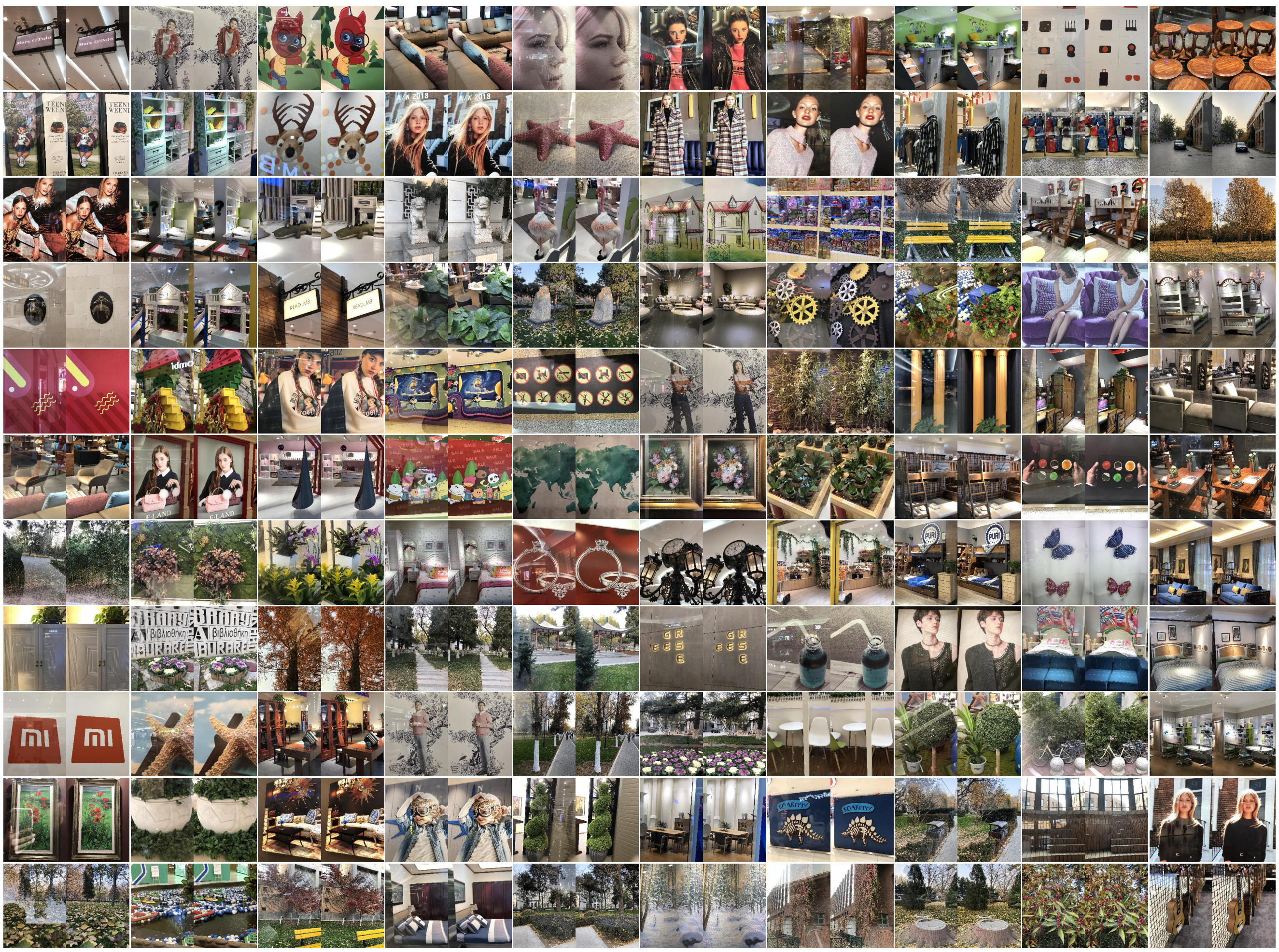


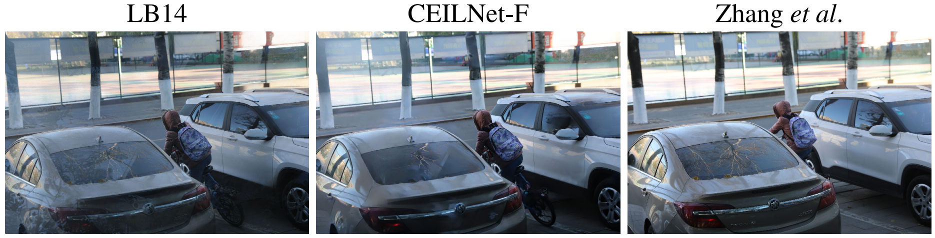
Figure V: Complete Overview of our unaligned dataset (smartphone part)

Figure VI: Qualitative comparisons on more real-world data

Input | LB14 | CEILNet-F | Zhang *et al.*

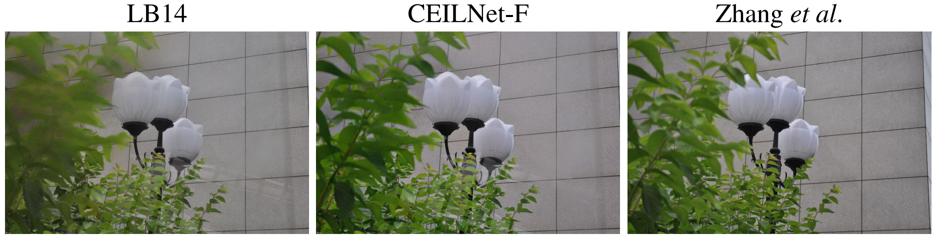
BDN-F | Ours | Reference

Input | LB14 | CEILNet-F | Zhang *et al.*



BDN-F | Ours | Reference

Input | LB14 | CEILNet-F | Zhang *et al.*



BDN-F | Ours | Reference

Figure VI: (Cont.) Qualitative comparisons on more real-world data

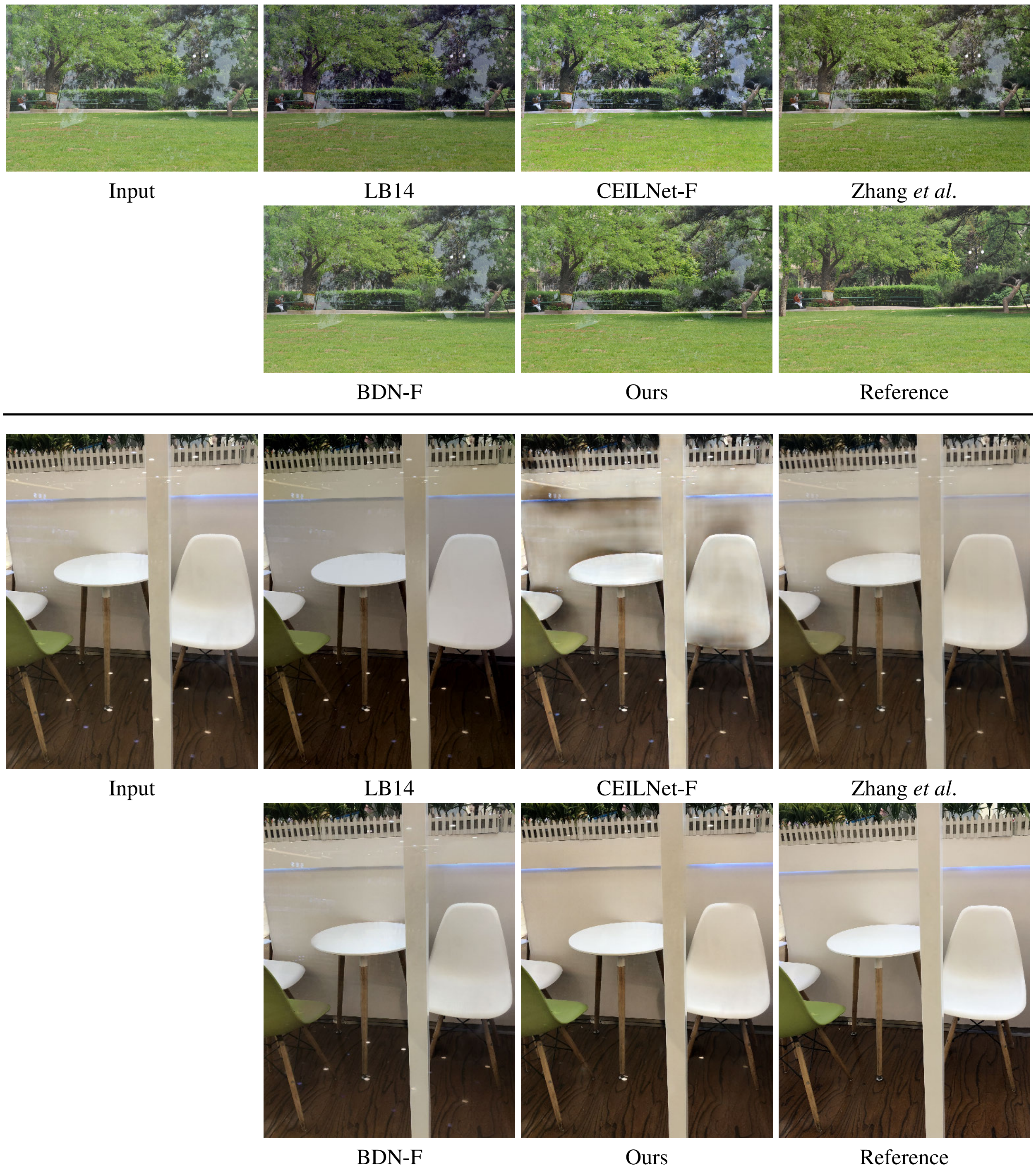


Figure VI: (Cont.) Qualitative comparisons on more real-world data