
Input
BDN
ERRNet
IBCLN
Kim et al.
Chang et al.
Peng et al.
Ours

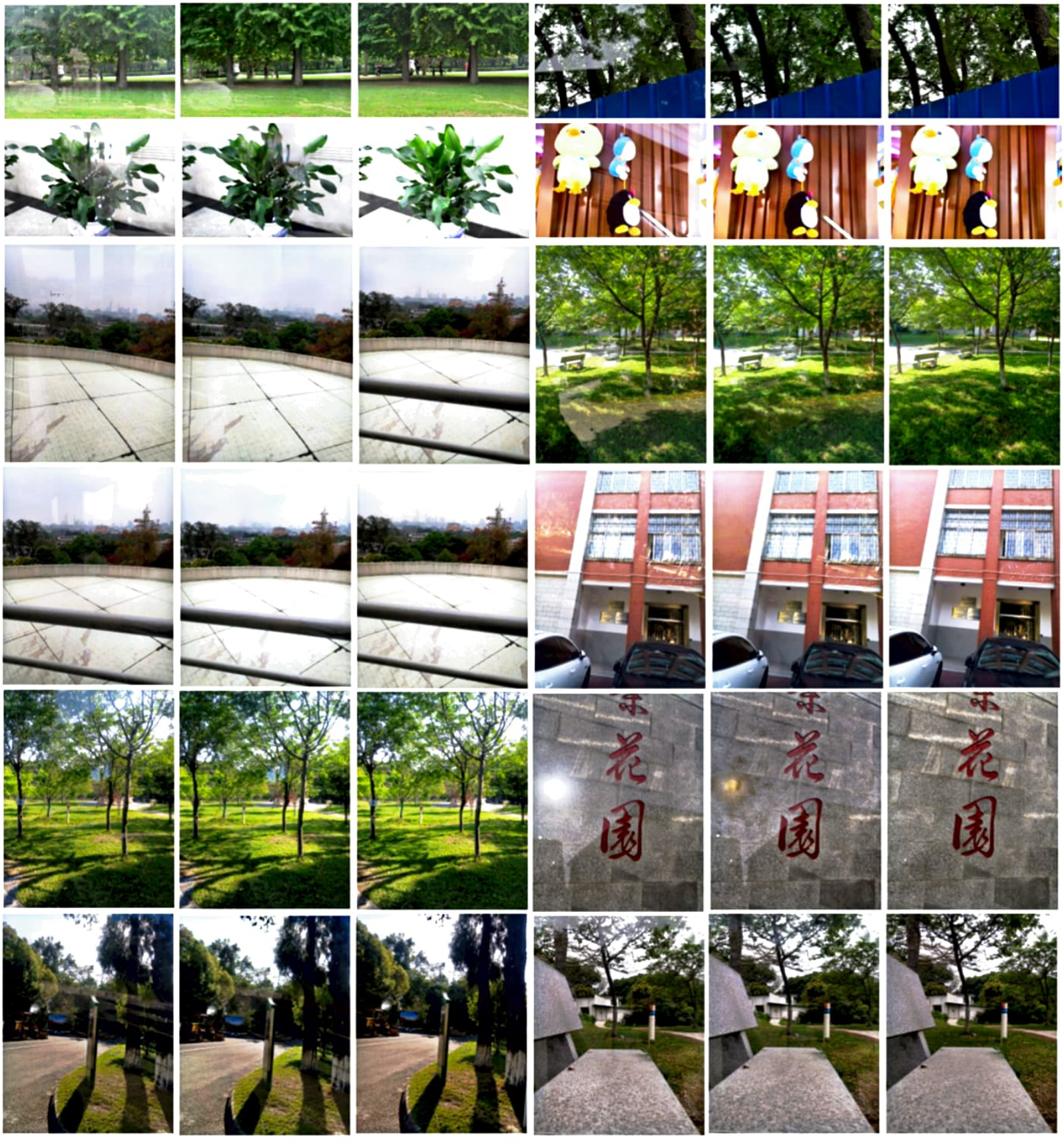
Synthesized input Ground truth I_B Ground truth I_R

(a) (b) (c) (d)

(a) M: Our ISP (b) M: Lightroom (c) M: Camera

(d) M: Our ISP1 (e) T: Our ISP1 (f) T: Our ISP2

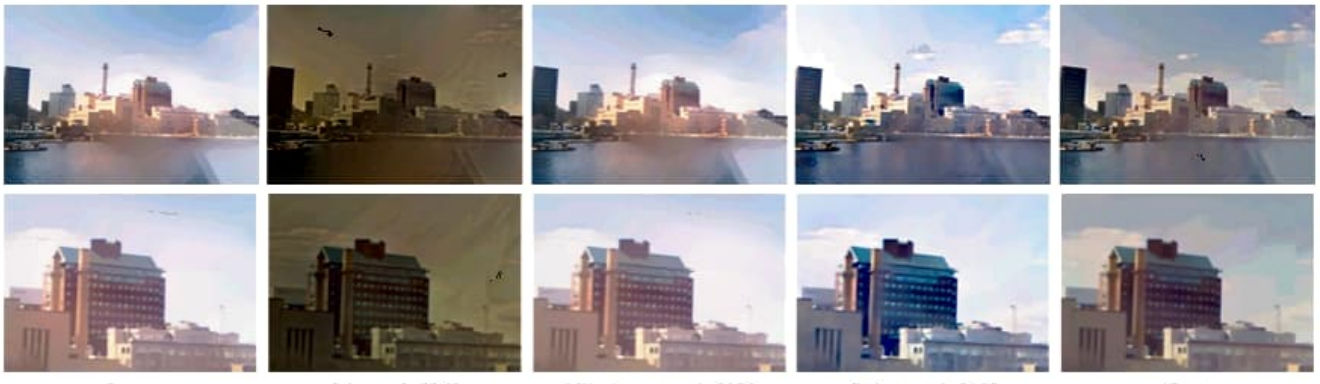
Figure 7. The first row shows that our ISP generates similar results compared with Lightroom and Camera Output. The second row shows that different ISPs can be applied to T to achieve similar results. Note that Lightroom is a professional ISP software supported by Adobe.



Input RABRRN Ground-truth Input RABRRN Ground-truth

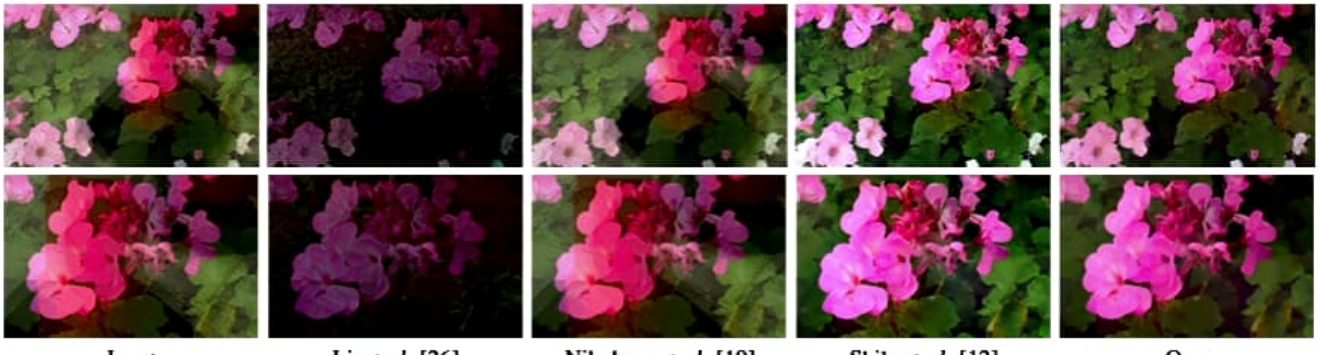
Input Li *et al.* [26] Nikolaos *et al.* [10] Shih *et al.* [12] Ours

Fig. 12: Results on image "Factory".



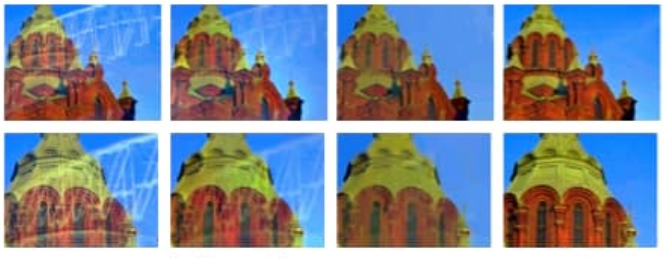
Input Li *et al.* [26] Nikolaos *et al.* [10] Shih *et al.* [12] Ours

Fig. 13: Results on image "Lake".



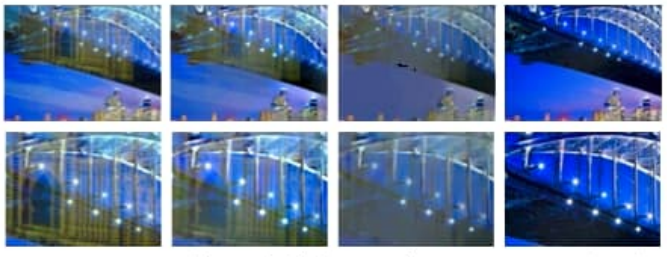
Input Li *et al.* [26] Nikolaos *et al.* [10] Shih *et al.* [12] Ours

Fig. 14: Reflection results on image "Flower".



Input Shih *et al.* [12] Ours Ground truth

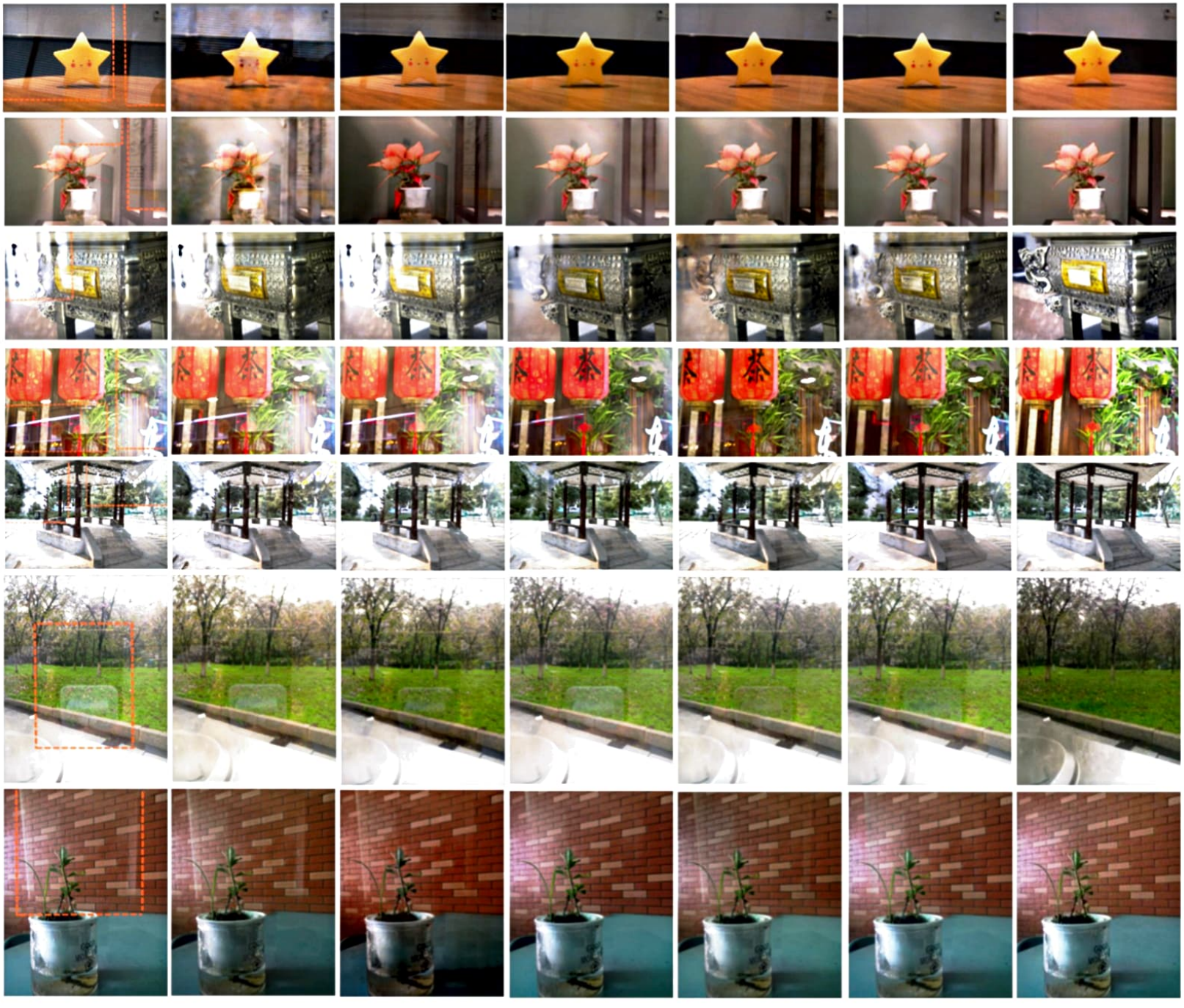
Fig. 15: Results on image "Church".



Input Shih *et al.* [12] Ours Ground truth

Fig. 16: Results on image "Bridge".

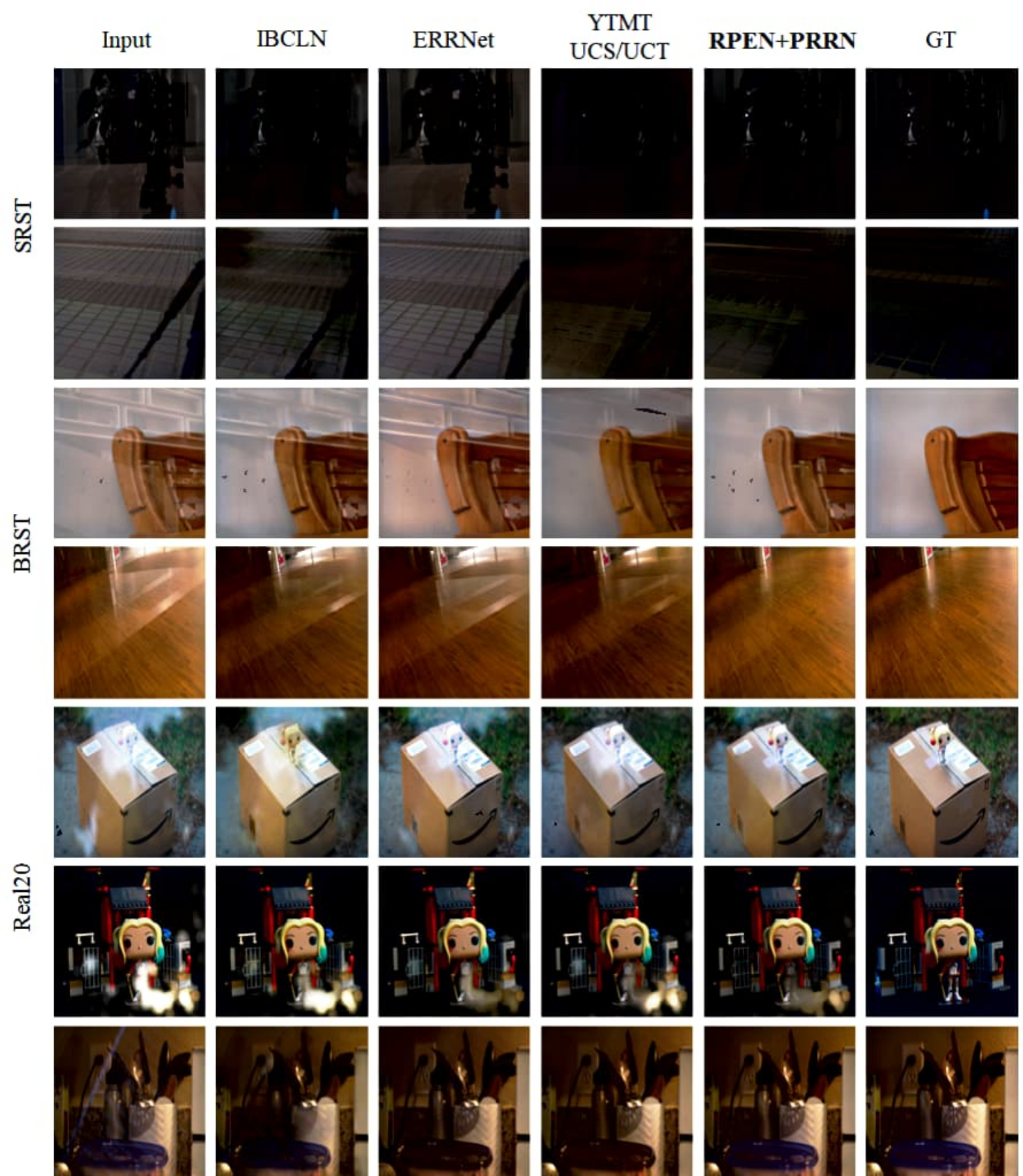

Input
BDN
ERRNet
IBCLN
Kim et al.
Chang et al.
Peng et al.
Ours
GT



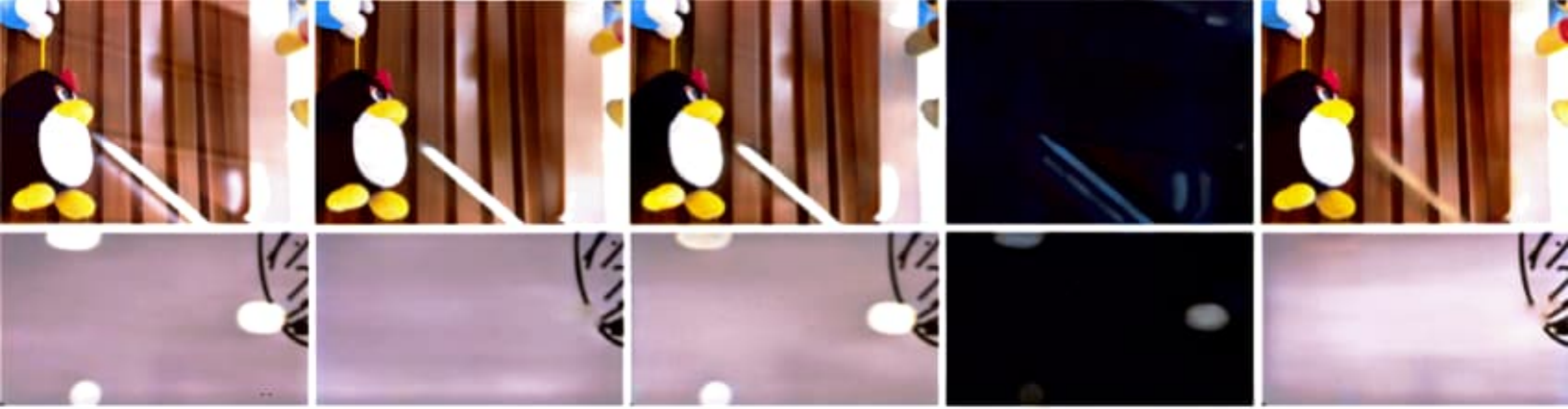
Ours-$M1$ Ours-$R1$ Ours-$M2$ Ours-$R2$

Figure 4. With the M-R pipeline proposed by Lei et al. [14], we can utilize a diverse set of glasses existing in our daily life (e.g., the curved and colored glass on the telephone booth, and the glass as a door).

Input | $\text{RAGNet}_{\mathbf{F}}$ | $\text{RAGNet}_{\mathbf{F} \circ \mathbf{M}}$ | $\text{RAGNet}_{\mathbf{F} \circ \mathbf{M}^{2c}}$ | RAGNet


Input
IBCLN
ERRNet
YTMT
UCS/UCT
RPEN+PRRN
GT
SRST
BRST
Real20


Input
RAGNet$_{I \to T}$
w/o $\mathbf{F}_{diff}$
w/o $\mathcal{L}_{\text{mask}}$
w/o $\mathcal{L}_{\text{mask}}^{diff}$
w/o $\mathcal{L}_{\text{mask}}^{reg}$
RAGNet



Input ERRNet IBCLN Predicted $\hat{B}$ Ours

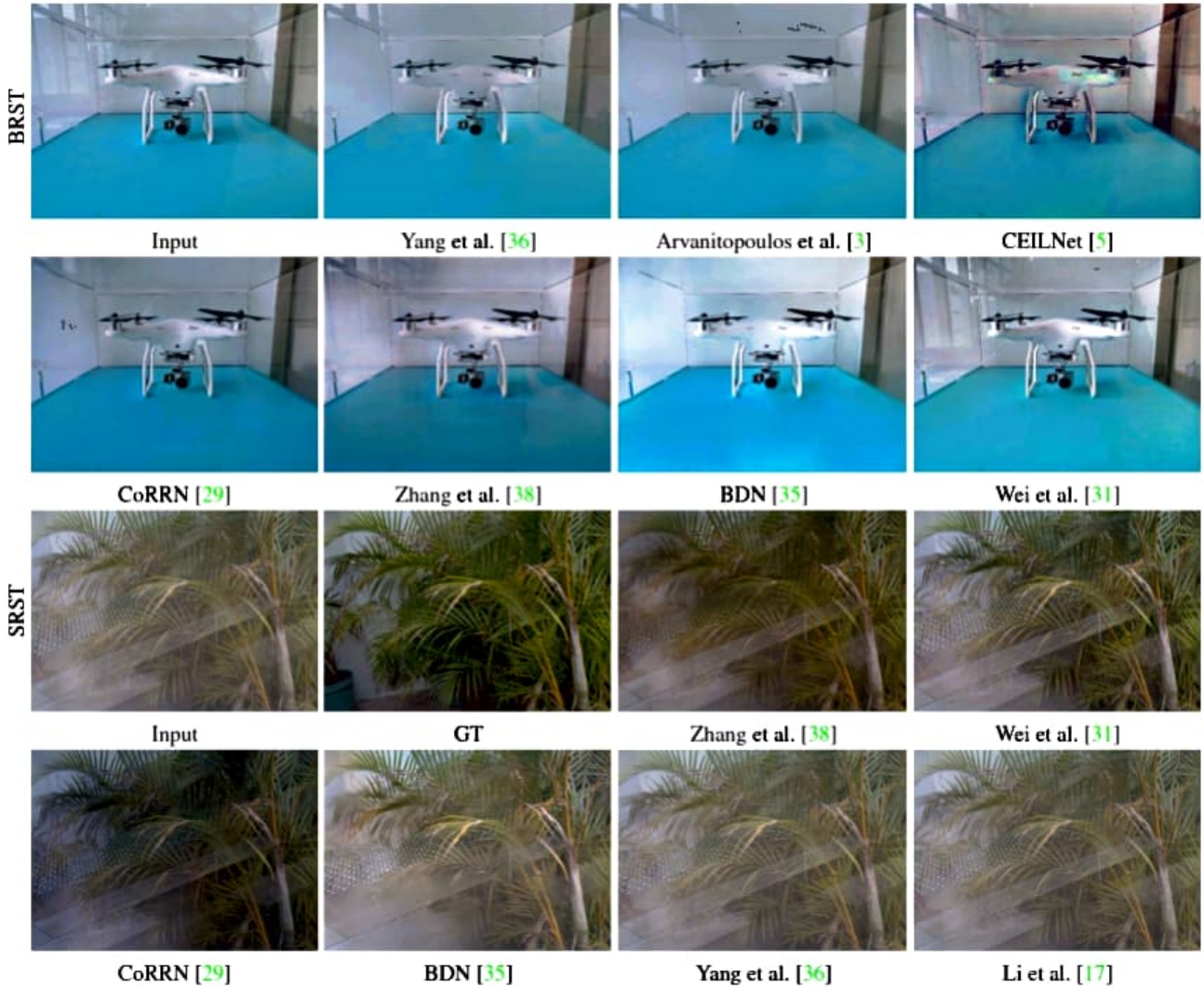


Figure 8. Most methods cannot remove the sharp reflection. This is probably because learning-based methods are trained on synthetic data where R is blurry. Learning free methods often assume reflection is blurry. However, sharp reflection is quite common in the real world. Figure best viewed in the electronic version.

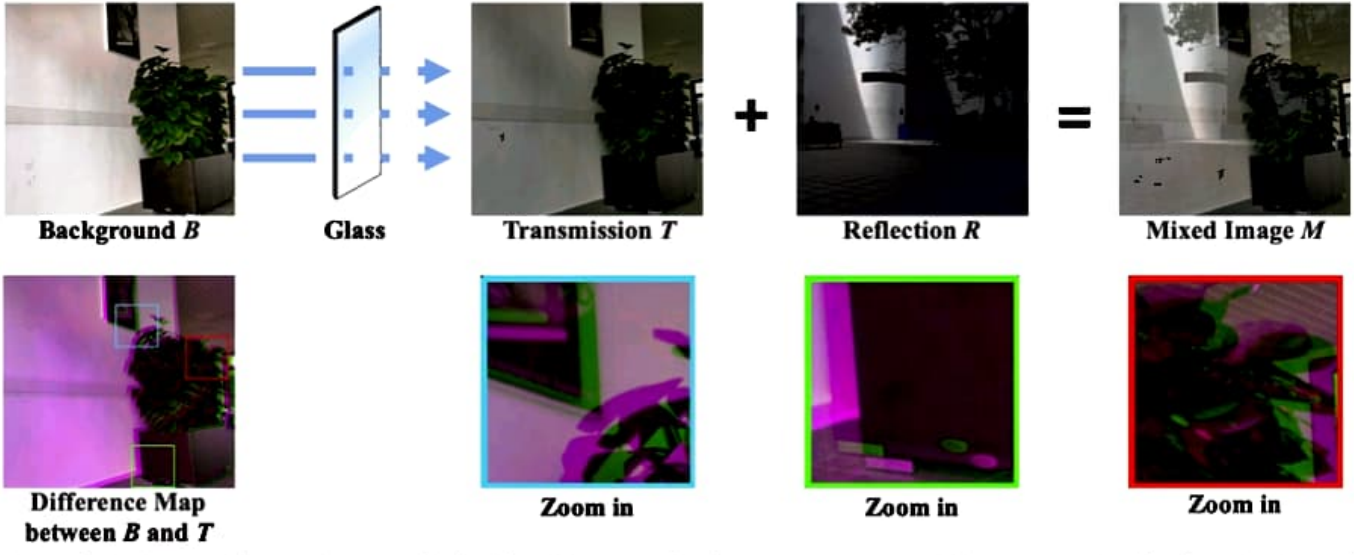


Figure 2. Due to refraction, spatial shift and intensity difference exist between B and T. The difference map visualizes the misalignment between B and T. The sum of the reflection R and the transmission T equals to the mixed image M in the raw data space.

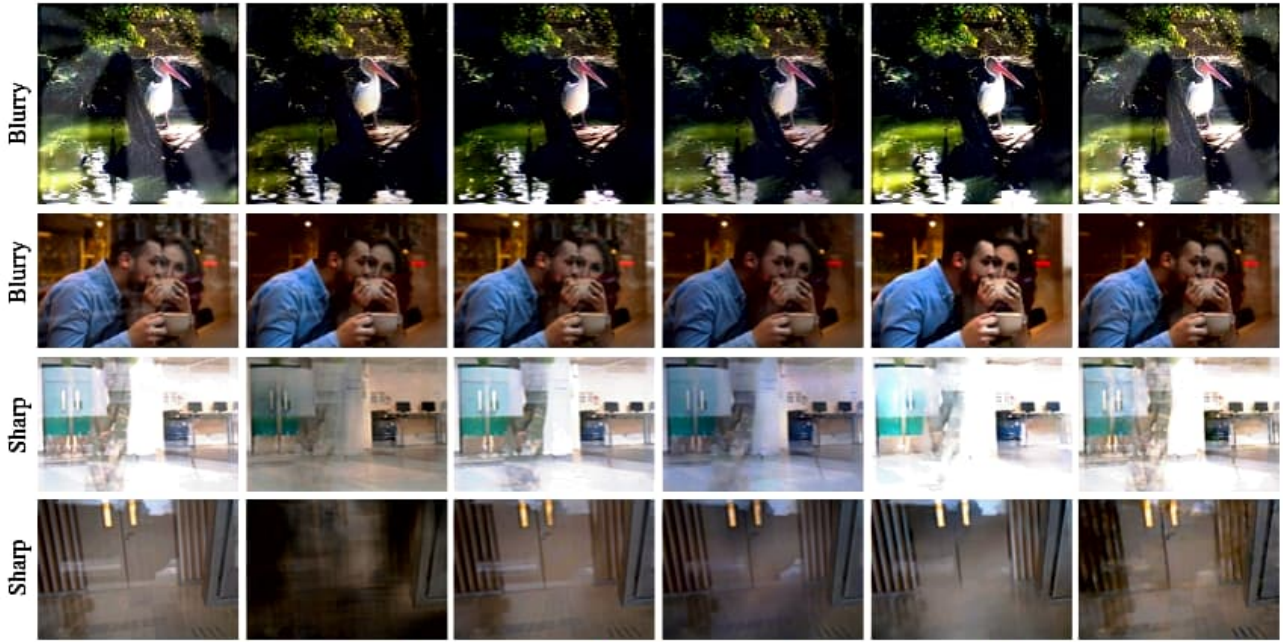


Figure 1. The performance of existing methods on different types of reflection is quite different. Most algorithms can remove the blurry reflection but cannot remove the sharp reflection well.

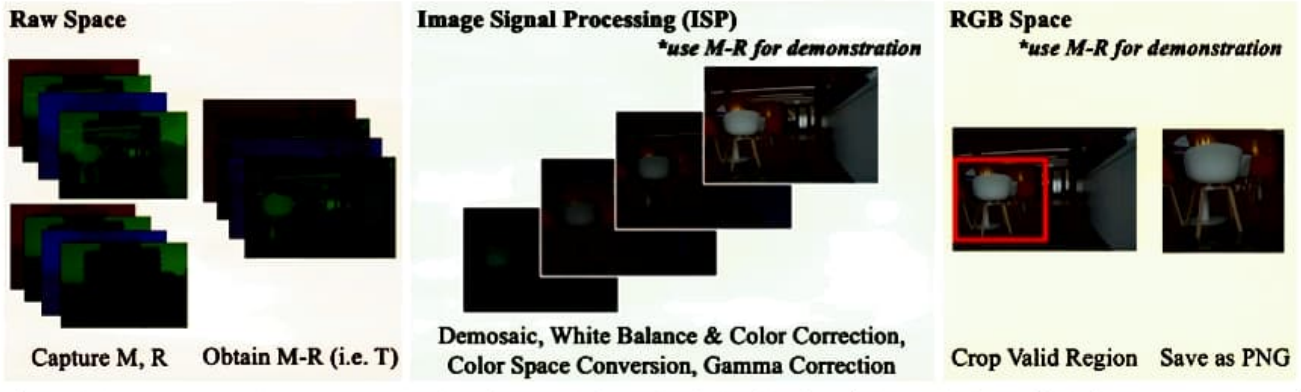
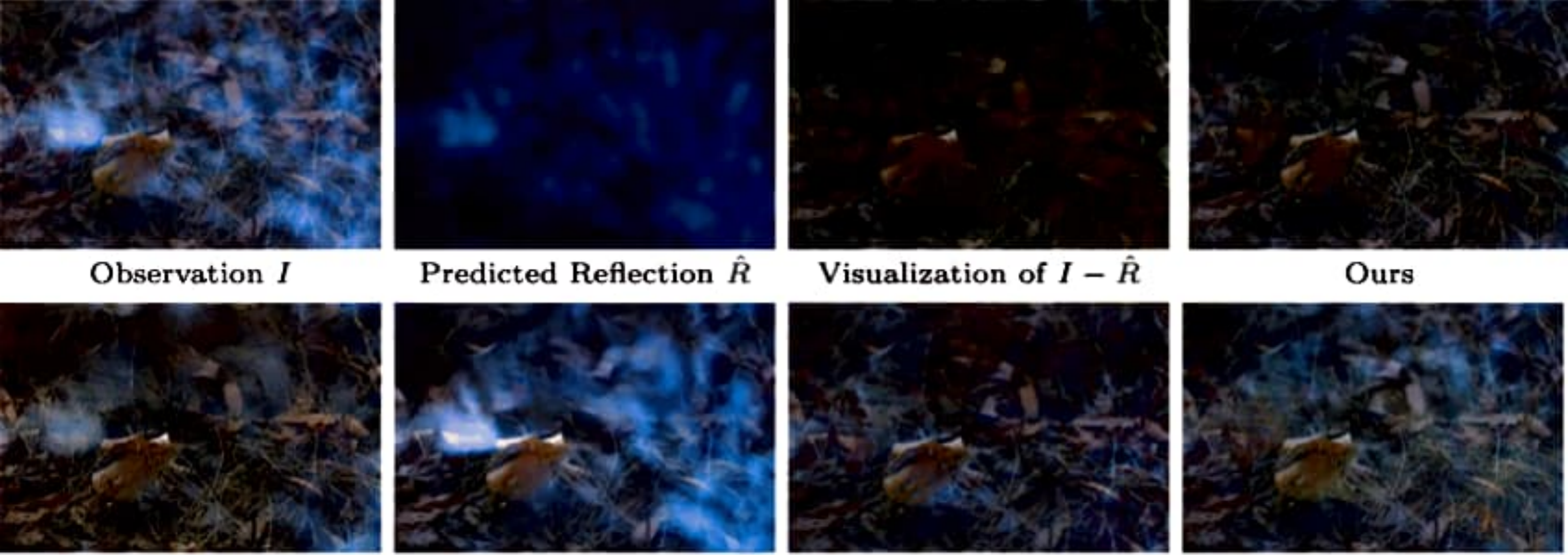
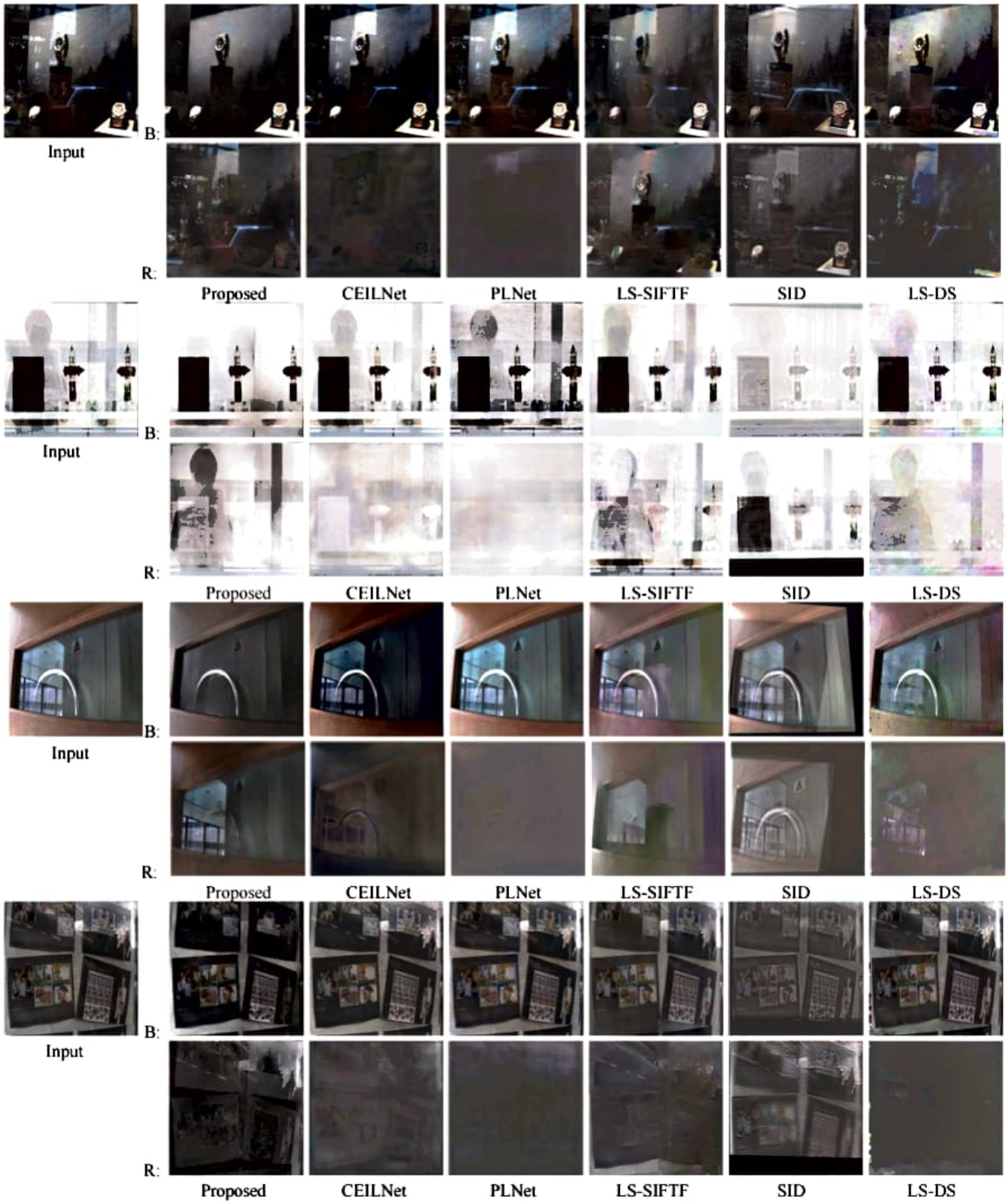


Figure 5. The post-processing pipeline. Ground-truth transmission T is obtained in the RAW space. Then all the RAW images are passed through an "ISP" to obtain the corresponding RGB images. Finally, the regions of interest regions are cropped out.


Observation I
Predicted Reflection $\hat{R}$
Visualization of $I - \hat{R}$
Ours
CEILNet [4]
BDN [9]
Lei et al. [13]
IBCLN [12]


Input
B:
R:
Proposed
CEILNet
PLNet
LS-SIFTF
SID
LS-DS
Input
B:
R:
Proposed
CEILNet
PLNet
LS-SIFTF
SID
LS-DS
Input
B:
R:
Proposed
CEILNet
PLNet
LS-SIFTF
SID
LS-DS
Input
B:
R:
Proposed
CEILNet
PLNet
LS-SIFTF
SID
LS-DS

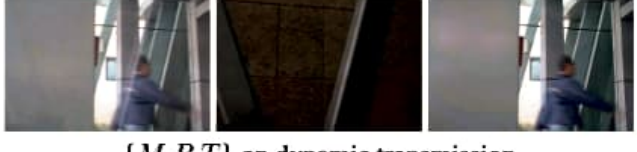
RGB M RGB R RGB $M-R$ Gamma $M-R$ Raw $M-R$

Figure 6. If $M-R$ is applied other than the raw data space, undesirable residuals will appear. "RGB $M-R$": do $M-R$ on RGB images. "Gamma $M-R$": use $M^{2.2}-R^{2.2}$ to reduce the impact of gamma correction. "Raw $M-R$": do $M-R$ on raw data.

$\{M,R,T\}$ on curved glass

$\{M,R,T\}$ on colored (red) glass

$\{M,R,T\}$ on dynamic transmission



$\{M,R,T\}$ on dynamic transmission

Figure 3. More examples about the data diversity. In addition to glass types, we are also able to capture dynamic scenes, which enriches the scene diversity.


Input
Shih *et al.* [12]
Ours
Ground truth

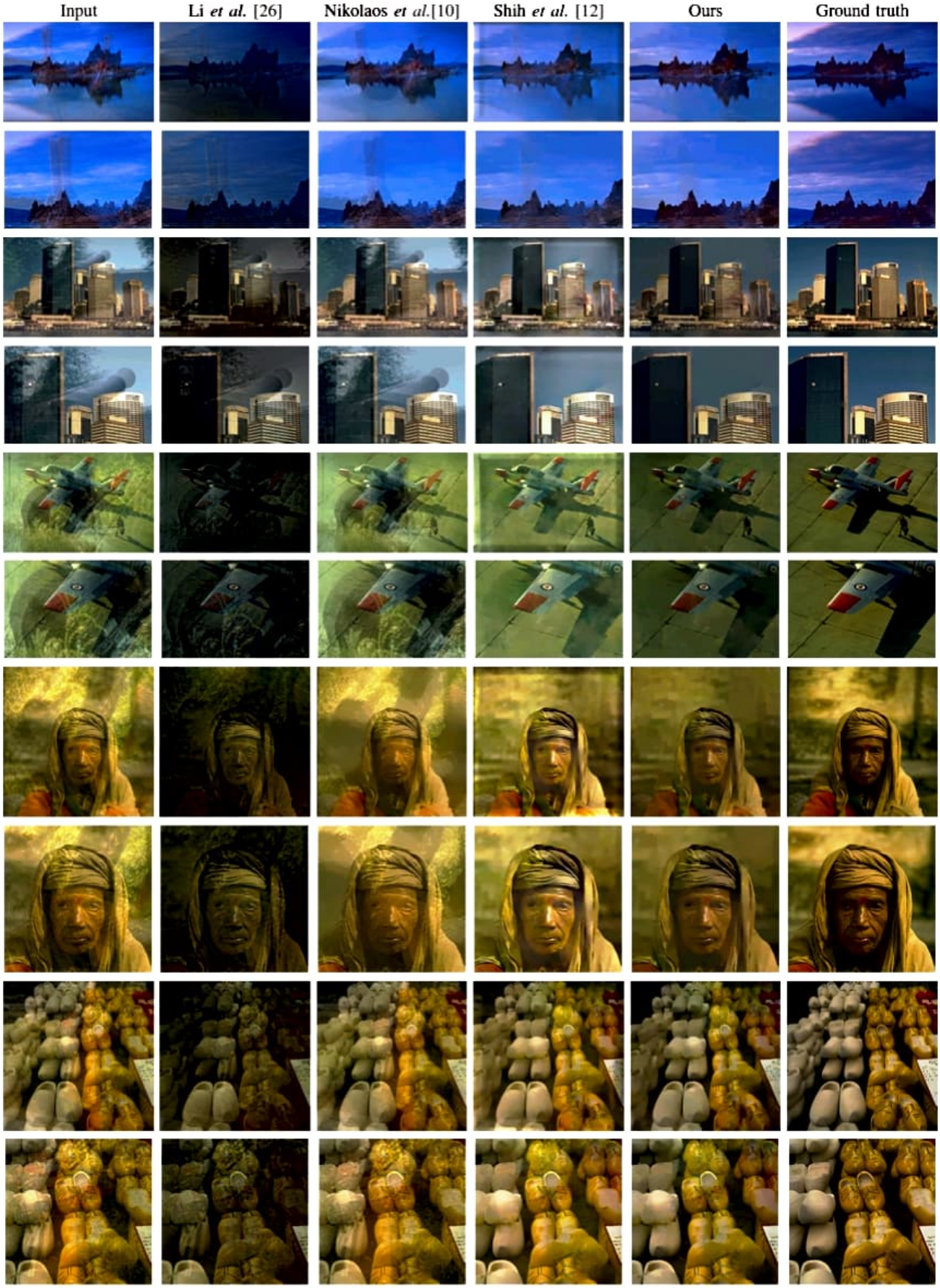


Fig. 11: The results on synthetic images. The odd/even rows show the whole/zoomed-in results.