

A Categorized Reflection Removal Dataset with Diverse Real-world Scenes

Chenyang Lei^{1*}

Xuhua Huang^{1,2*}

Chenyang Qi¹

Yankun Zhao¹

Wenxiu Sun³

Qiong Yan³

Qifeng Chen¹

¹HKUST ²CMU ³SenseTime

Abstract

Due to the lack of a large-scale reflection removal dataset with diverse real-world scenes, many existing reflection removal methods are trained on synthetic data plus a small amount of real-world data, which makes it difficult to evaluate the strengths or weaknesses of different reflection removal methods thoroughly. Furthermore, existing real-world benchmarks and datasets do not categorize image data based on the types and appearances of reflection (e.g., smoothness, intensity), making it hard to analyze reflection removal methods. Hence, we construct a new reflection removal dataset that is categorized, diverse, and real-world (CDR). A pipeline based on RAW data is used to capture perfectly aligned input images and transmission images. The dataset is constructed using diverse glass types under various environments to ensure diversity. By analyzing several reflection removal methods and conducting extensive experiments on our dataset, we show that state-of-the-art reflection removal methods generally perform well on blurry reflection but fail in obtaining satisfying performance on other types of real-world reflection. We believe our dataset can help develop novel methods to remove real-world reflection better. Our dataset is available at <https://alexzhao-hugga.github.io/Real-World-Reflection-Removal/>.

1. Introduction

Reflection removal is a task to remove undesirable reflection artifacts from a photograph. Existing deep learning based approaches for reflection removal have demonstrated superior performance on synthetic data, but we find that their performance degrades severely in diverse real-world data, as shown in Fig. 1. Most existing learning-based methods [5, 38] are trained on synthetic data created under various assumptions. Hence, their performance is limited by the domain gap between real-world and synthetic data. The assumptions to create the synthetic data are often simplified,

and thus these approaches have sub-optimal performance on real-world data.

The existing real-world benchmark dataset SIR² facilitates the research in reflection removal, but it has weaknesses as reported in the paper [28]. In this dataset, most images only contain flat objects or small object (i.e., postcards and solid objects) under controlled lighting. These images do not represent the scenes in our daily live because object distance, scales, and natural illumination variation are quite diverse in the wild. In the SIR² dataset, only collected 55 pairs of images with ground-truth transmission in the wild. There are also other real-world datasets [38, 15] with high-quality ground truth but they have a small number of images.

To address the limitations of existing reflection removal datasets, we present a large reflection removal dataset CDR that contains diverse scene in the real world. Compared with prior work, CDR has several advantages, as shown in Table 1. In terms of the number of real-world images, ours is much larger and more diverse than existing ones. We construct the dataset following several principles to ensure image quality and diversity. First, we capture our data in the wild since it is similar to images captured in daily life. Second, to ensure perfect alignment between transmission and input mixed images, we obtain the transmission by subtracting reflection from the mixed image in the raw data space [14]. Third, we capture the images using different glasses in diverse scenes to ensure the diversity of our dataset. We believe that the performance of a reflection removal method is related to the smoothness of reflection, and thus we carefully categorize the collected data into different types of reflection. We split the data based on the smoothness of reflection (i.e., sharp reflection or blurry reflection). In our experiments, all existing methods are sensitive to the smoothness of reflection.

We hope the CDR dataset can facilitate the research in reflection removal. The CDR dataset can provide a more extensive evaluation for reflection removal methods. In addition, the detailed categorization can help analyze the bottlenecks of existing methods.

*Equal contribution

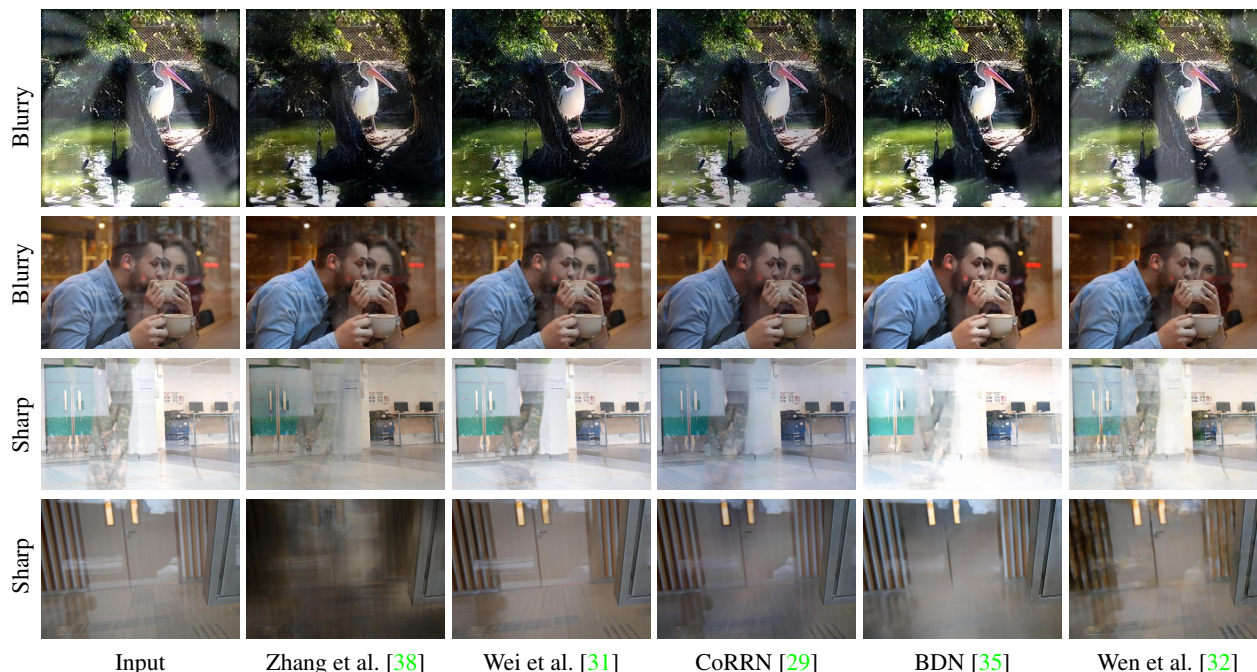


Figure 1. The performance of existing methods on different types of reflection is quite different. Most algorithms can remove the blurry reflection but cannot remove the sharp reflection well.

2. Background

2.1. Single Image Reflection Removal

Most single image reflection removal methods [5, 38, 35, 36] rely on various assumptions. Considering image gradients, Arvanitopoulos et al. [3] propose the idea of suppressing the reflection, and Yang et al. [36] propose a faster method based on convex optimization. These methods fail to remove sharp reflection. Under the assumption that transmission is in focus, Punnapurath et al. [22] design a method based on dual-pixel camera input. CEILNet [5], Zhang et al. [38], and BDN [35] assume that reflection is out of focus and synthesize images to train their neural networks. CEILNet [5] estimates target edges first and uses it as guidance to predict the transmission layer. Zhang et al. [38] use perceptual and adversarial losses to capture the difference between reflection and transmission. BDN [35] estimates the reflection images, which is then used to estimate the transmission layer. These methods [38, 5, 35] work well when reflection is more defocused than transmission but fail otherwise. For these deep learning based approaches, training data is critical for good performance. To bridge the gap between synthetic and real-world data, Zhang et al. [38] and Wei et al. [31] collected some real-world images for training. However, their images have misalignment issues between transmission and input images, and the dataset size is small. Wei et al. [31] propose to use high-level features that are less sensitive to small misalignment to calculate losses. To obtain more realistic and

diverse data, Wen et al. [32] and Ma et al. [20] propose to synthesize data using a deep neural network and achieve better performance and generalization. Kim et al. [11] propose a physics based method to render the reflection and achieve better performance than using synthetic images.

2.2. Reflection Removal with Multiple Images

Utilizing multiple images as input provides additional information, which makes it possible to relax some strict assumptions used in prior work. A number of approaches [27, 24, 23, 16, 8, 9, 34, 26, 2, 18] exploit the relative motion between reflection and transmission with multiple images captured with camera movement to remove reflection.

Some other methods may take a sequence of images under specific conditions or camera settings. For example, pairs of flash and no-flash images [1, 13], near-infrared cameras [10], light field cameras [30], dual-pixel cameras [22], and polarization cameras [19, 21, 25, 12, 6] can be used.

Different from supervised learning models, DoubleDIP [7] separates a mixed image into reflection and transmission layers based on internal self-similarities in multiple superpositions, in an unsupervised fashion.

Although the methods based on multiple images do not rely on strict assumptions for appearances of reflections (e.g., blurry reflection, ghosting effects), they need additional requirements on data [34] or special devices [14], which may prevent them from broader applications. Therefore, in this work, we focus on building a dataset and eval-

	Glass type	Categorized	Scene-level data	Alignment	Reflection image	Curve glass	Data type	Training set
SIR ² -Wild [28]	3	No	55	Calibrated	Yes	No	RGB	No
Zhang et al. [38]	1	No	<110	Calibrated	No	No	RGB	No
Nature [15]	2	No	< 220	Misalignment	No	No	RGB	No
Ours	>200	Yes	1,063	Perfect	Yes	Yes	RGB&Raw	Yes

Table 1. The comparisons between our data and existing datasets [28, 38, 15]. Scene-level data: the data that is captured in the wild instead of lab environments.

uating single image reflection removal methods.

2.3. SIR² benchmark dataset

Wan et al. [28] propose the SIR² benchmark dataset for single image reflection removal. This dataset contains images taken in controlled scenes and in the wild. To solve the misalignment problem, they calibrate the alignment between the mixed image M and the background B . The dataset was captured with three glasses of different thicknesses, various combinations of aperture sizes, and different exposure times to improve the diversity in this dataset.

As described by Wan et al. [28], most of the objects in their controlled scene contain only flat objects (postcard) or objects with similar scales (solid objects). However, real-world scenes contains objects at different depths, and the natural environment illumination also varies greatly, while the controlled scenes are mostly captured in an indoor office environment. To address this limitation, 55 pairs of images with ground-truth reflection and transmission are captured in the wild, but 55 pairs are far from large scale. Also, this dataset does not provide a standard split among the training set, the validation set, and the test set.

3. CDR Dataset

In this section, we describe the features of our CDR dataset for reflection removal, which stands for ‘‘Categorized, Diverse, and Real-world.’’ A triplet $\{M, R, T\}$ is collected in each scene where M is the mixed image, R is the reflection image, and T is the transmission image.

This dataset is collected mainly by three cameras: a DSLR (Digital Single-Lens Reflex) camera Canon EOS 50D, a MILC (Mirrorless Interchangeable Lens Cameras) Nikon Z6, and a smartphone camera Huawei Mate30. In total, we provide 1,063 triplets in our dataset.

We collected the real-world data in the wild. Compared with data captured in a controlled environment, real-world data in the wild contains objects of various distances and illumination variation, as reported by SIR² [28]. We improve our dataset in various aspects. Table 1 summarizes the main differences between the CDR dataset and existing datasets [28, 38, 15]. Specifically, our main advantages are listed as followed:

(a) Data categorization. We notice that the performance of a reflection removal model is related to the appearance of reflection. To facilitate in-depth analysis, we split all the images according to the reflection types.

(b) Perfect alignment. We provide perfect alignment between the mixed image M and transmission T . Existing datasets [28, 38] has the misalignment issue between M and T . The misalignment issue does not only degrade a model’s performance but also makes evaluation less accurate. A reflection removal model trained on misaligned paired data often generate blurry images [31]. Note that a single pixel shift in an image can affect evaluation metrics PSNR and SSIM significantly.

(c) Diversity. We provide much more diverse data by utilizing different types of glasses existing in our daily life, including colored and curved glasses shown in Fig. 3. We capture various objects in different environments and lighting conditions with three different types of cameras. As mentioned before, when we capture images in the wild, we also guarantee diversity of the smoothness and intensity of reflection.

(d) Large-scale data. Our dataset is significantly larger than existing datasets or benchmarks. In total, our dataset contains 1,063 triplets M, R, T . We believe our evaluation is more accurate because there is no misalignment between the mixed image and the ground truth.

(e) Other advantages. Compared with existing datasets [38, 31, 15], we provide reflection image since many works have demonstrate the effectiveness of using reflection [35, 14]. We also provide raw images instead of only RGB images for future study in reflection removal.

Our CDR dataset is publicly available, which can be used for training and evaluation. The detailed categories can help researchers understand the strengths and weaknesses of existing methods. We hope it can accelerate the research in reflection removal.

3.1. Data Acquisition

In order to collect diverse data with perfect alignment in the wild, we adopt the M-R pipeline proposed by Lei et al. [14] where T is obtained by $M - R$. Different from Lei et al. [14] which implements $M - R$ on raw data obtained by a polarization sensor, we use normal cameras that provide

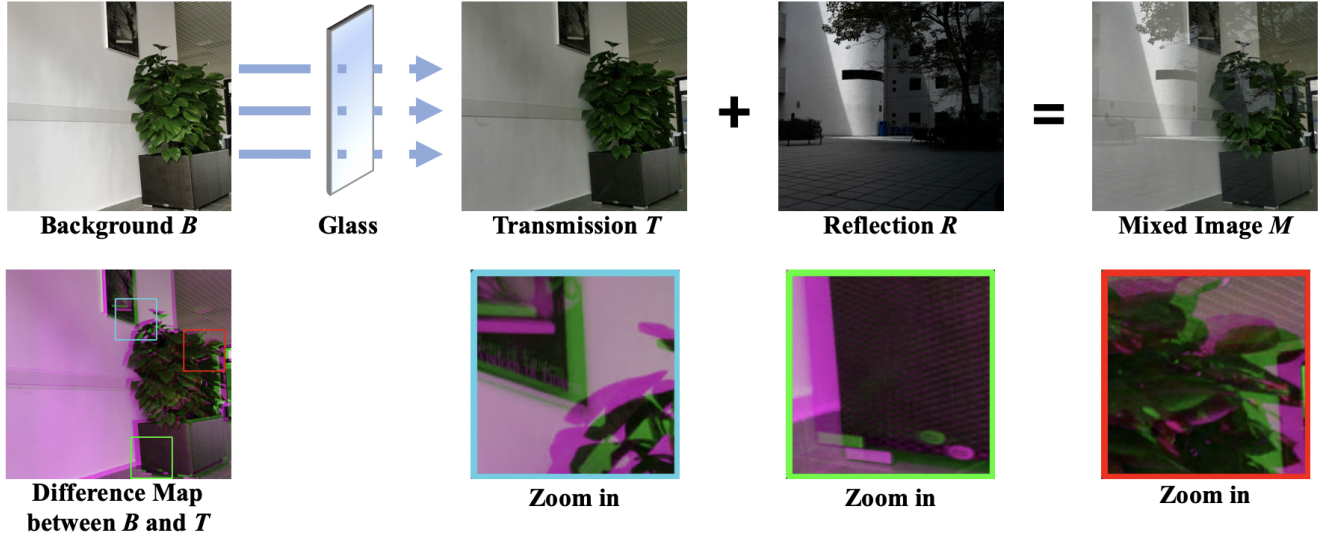


Figure 2. Due to refraction, spatial shift and intensity difference exist between B and T . The difference map visualizes the misalignment between B and T . The sum of the reflection R and the transmission T equals to the mixed image M in the raw data space.

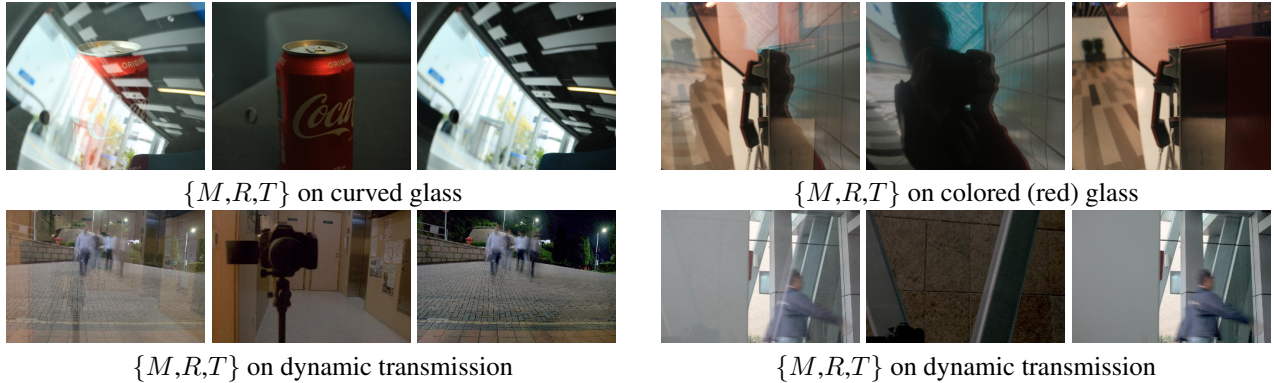


Figure 3. More examples about the data diversity. In addition to glass types, we are also able to capture dynamic scenes, which enriches the scene diversity.

the raw data to construct the CDR dataset.

Fig. 4 shows our data collection pipeline. The first step is to have an appropriate glass. Since we do not need to remove the glass to obtain background B as other methods do [28, 31, 38], we can utilize immovable glasses in the real world. As the second step, we use a piece of black cloth behind the glass to block transmission to obtain the reflection R . In the end, we remove the cloth to collect the mixed image M . Please refer to Fig. 3 for some captured example iamges.

There are some details in real capturing progress:

- To ensure perfect alignment between M and R , we use a tripod to fix the camera to take images.
- To ensure the exposure is consistent between M and R , all the cameras are set to the manual mode with a fixed camera setting including ISO and exposure time.

- To reduce the noise level in M and R as much as possible, we capture data with a long exposure time and a small ISO.
- The objects in the reflection (not transmission) need be static in both M and R to ensure the perfect alignment in $M - R$.

3.2. Post Processing

Fig. 5 shows the overall pipeline of our post-processing step. With raw M and raw R , we calculate raw T by $T = M - R$ in the raw data space. Note that $M - R$ should be implemented in the RGB space because the linearity between light intensities and RGB values does not hold, as shown in Fig. 6. In $T = M - R$, negative values may appear due to noise, and they are set to zeros directly. The black level of a camera is added back to ensure its ISP can be applied to T . So far, the raw data format of T is the same as M and R .

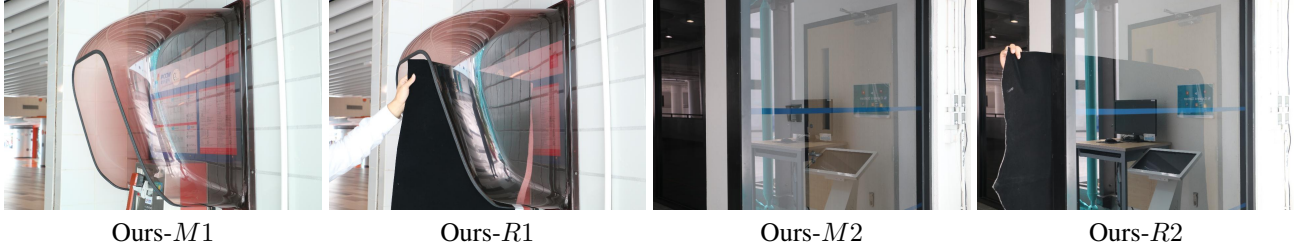


Figure 4. With the M-R pipeline proposed by Lei et al. [14], we can utilize a diverse set of glasses existing in our daily life (e.g., the curved and colored glass on the telephone booth, and the glass as a door).

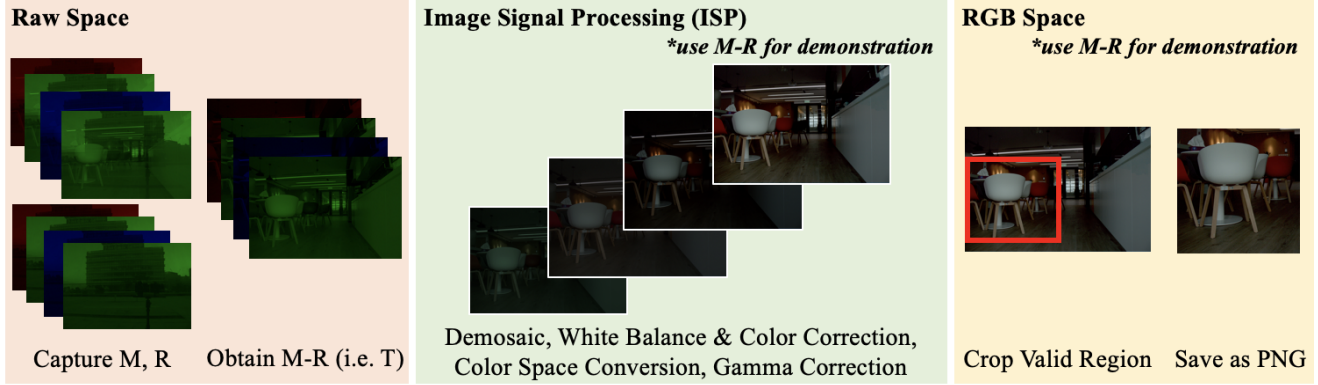


Figure 5. The post-processing pipeline. Ground-truth transmission T is obtained in the RAW space. Then all the RAW images are passed through an “ISP” to obtain the corresponding RGB images. Finally, the regions of interest regions are cropped out.

Since most existing reflection removal algorithms adopt RGB images as input, we need to convert raw images from the raw data space to RGB space. However, the camera default ISP is not public. Therefore, we implement our own image signal processing (ISP) pipeline to generate RGB images for M , T , and R using the same ISP. As for the metadata of T , we simply apply the metadata of M to T directly.

After obtaining RGB T images, we need to crop the area of interest for T . In the dataset, we eliminate the triplets in the following two cases: (1) If multiple glasses exist behind the covered glass (i.e., glasses exist in the background), the obtained T still has glasses. (2) If there is no glass in M , then no black cloth exists to cover the transmission. In this case, the obtained T will equal 0, which contradicts the ground truth.

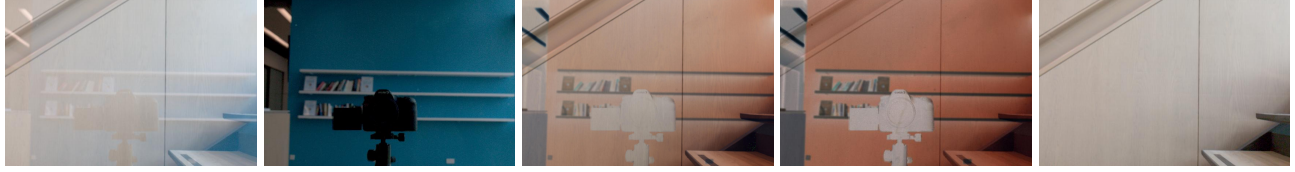
Different from SIR^2 that classify images as bright or dark scenes according to absolute intensity, focus, or thickness of glass, we categorize images according to three criteria: relative intensity, smoothness between R and T and the ghosting effect. First, we observe that the impact of reflection is decided by the relative intensity instead of absolute intensity. We calculate the mean intensity ratio for each pair of R and T and categorize them as weak reflection, moderate reflection or strong reflection. Second, we sort the data based on the smoothness of reflection and transmission: BRST (blurry reflection and sharp trans-

mission), SRST (sharp reflection and sharp transmission), BRBT (blurry reflection and blurry transmission). In fact, the BRST type has been generally used in previous work. At last, we pick up the data which has ghosting effect as a class.

4. Experiments

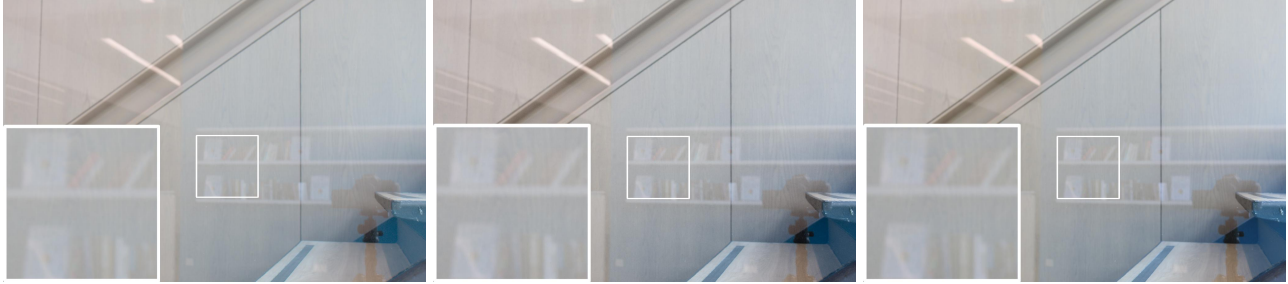
We first describe our experimental setup. The images from all three cameras are provided for training, which can reduce the impacts of the domain gap. Following previous work [28], we choose PSNR, SSIM and NCC as our main evaluation metrics.

The baselines for comparison in our experiment include CEILNet [5], CoRRN [29], BDN [35], Zhang et al. [38], Wei et al. [31], Yang et al. [36], Arvanitopoulos et al. [3], Li et al. [17], IBCLN [15], and Kim et al. [11]. All these methods take a single RGB image as input. For learning-based methods, we use their pre-trained models by default as some methods do not provide training code. For Yang et al. [36], different thresholds might result in different output images. Therefore, except for the original threshold, multiple thresholds are used for comparison, and the best result is reported.



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

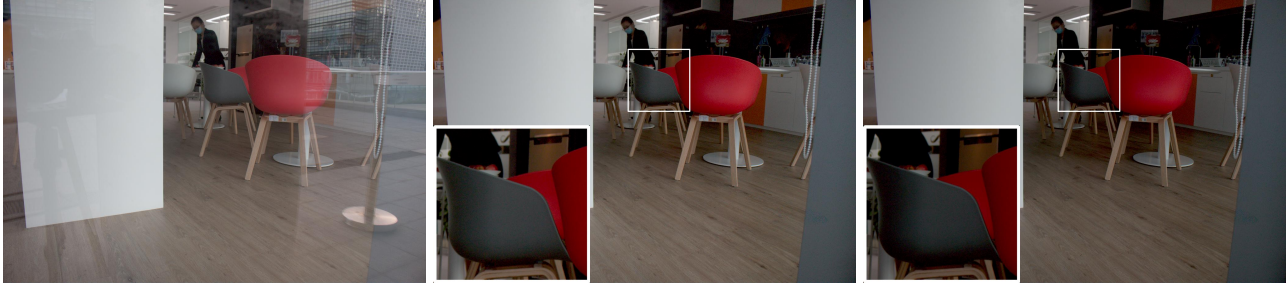
Figure 6. If $M - R$ is applied other than the raw data space, undesirable residuals will appear. “RGB $M - R$ ”: do $M - R$ on RGB images. “Gamma $M - R$ ”: use $M^{2.2} - R^{2.2}$ to reduce the impact of gamma correction. “Raw $M - R$ ”: do $M - R$ on raw data.



(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.

4.1. Evaluation

One interesting observation is that the performance of most existing single image reflection removal methods is highly related to the type of reflection.

Quantitative results. Table 2 shows the performance of the evaluated methods in terms of PSNR, SSIM, and NCC. We also split data according to smoothness, relative intensity, and ghosting effect of reflection and report separate results to analyze the impact of different image patterns. We find that learning free methods [3, 36] achieve good performance in cases that follow their model assumptions. Although these methods usually rank poorly on average, they can perform quite well when the reflection is blurry, weak, or has ghosting effect.

In the following, we analyze the impact of different factors on the results.

1) Impact of real data. The methods [31, 29] trained on real data achieve better performance on general real-scene data. Since Kim et al. [11] synthesized physically-based data for training, they also achieve good performance on

real-world data. These data include SRST data, reflection with moderate intensity or strong intensity. As we can see, these methods [11, 31, 29] are often the first, second, and third best-performing methods.

2) The smoothness of reflection/transmission. The results on different smoothness of reflection and transmission are consistent with previously adopted assumptions. All methods perform relatively well the BRST (blurry reflection and sharp transmission) set. However, when the assumption does not hold (e.g., on SRST set), the performance degrades heavily. This phenomenon occurs to all evaluated algorithms. Note that in our dataset, almost 50% percents of M are SRST.

3) Ghosting effect. Another assumption about reflection is the ghosting effect. Although most deep learning methods do not synthesize ghosting reflection data, they still achieve better performance on such kind of data.

4) Reflection Intensity. The difficulty of reflection removal increases significantly when the intensity of the reflection increases. However, the importance of reflection

	All			SRST			BRST			Non-ghosting		
	PSNR	SSIM	NCC	PSNR	SSIM	NCC	PSNR	SSIM	NCC	PSNR	SSIM	NCC
Li et al. [17]	12.73	0.650	0.721	12.26	0.565	0.644	13.19	0.723	0.789	12.56	0.624	0.703
Arvan. et al. [3]	19.63	0.753	0.788	18.24	<u>0.680</u>	0.691	20.91	0.816	0.873	19.00	0.727	0.752
Yang et al. [36]	19.42	0.767	0.782	18.10	<u>0.680</u>	0.676	20.65	0.841	0.874	18.78	0.738	0.744
CEILNet [5]	17.96	0.708	0.757	16.17	0.596	0.654	19.49	0.802	0.847	17.24	0.673	0.720
Zhang et al. [38]	15.20	0.694	0.703	13.52	0.590	0.612	16.58	0.780	0.785	14.48	0.662	0.677
BDN [35]	18.97	0.758	0.745	19.04	0.713	0.642	19.06	0.799	0.836	18.62	0.733	0.698
Wei et al. [31]	21.01	0.762	0.756	19.52	0.672	0.631	22.36	<u>0.839</u>	0.864	20.50	0.731	0.713
CoRRN [29]	<u>20.22</u>	0.774	0.764	20.32	0.699	<u>0.656</u>	20.08	0.838	0.859	<u>20.37</u>	0.750	0.723
IBCLN [15]	19.85	<u>0.764</u>	0.735	18.33	0.671	0.613	<u>21.14</u>	0.842	0.846	19.23	<u>0.735</u>	0.687
Kim et al. [11]	21.00	0.760	<u>0.769</u>	<u>19.27</u>	0.676	0.654	22.61	0.833	<u>0.871</u>	20.42	0.731	<u>0.726</u>

	Weak R			Moderate R			Strong R			Ghosting		
	PSNR	SSIM	NCC	PSNR	SSIM	NCC	PSNR	SSIM	NCC	PSNR	SSIM	NCC
Li et al. [17]	14.36	0.779	0.841	12.47	0.636	0.709	8.89	0.309	0.401	13.36	0.742	0.785
Arvan. et al. [3]	<u>23.52</u>	0.878	0.941	18.43	0.744	0.765	13.56	0.397	0.423	21.88	0.844	0.919
Yang et al. [36]	23.18	0.903	<u>0.937</u>	18.28	0.754	0.755	13.50	0.402	0.425	21.72	0.870	<u>0.917</u>
CEILNet [5]	21.34	0.862	0.910	17.02	0.685	0.731	12.06	0.341	0.397	20.51	0.836	0.886
Zhang et al. [38]	17.20	0.827	0.822	15.10	0.685	0.688	9.33	0.311	0.402	17.81	0.806	0.797
BDN [35]	21.10	0.867	0.909	18.25	0.746	0.711	16.15	0.485	0.411	20.20	0.850	0.909
Wei et al. [31]	24.89	0.901	0.929	<u>19.42</u>	0.737	0.714	17.00	0.450	0.423	22.80	0.871	0.908
CoRRN [29]	20.50	0.890	0.928	21.01	0.768	<u>0.735</u>	15.12	<u>0.433</u>	0.394	19.70	0.861	0.911
IBCLN [15]	23.17	<u>0.899</u>	0.908	18.98	<u>0.752</u>	0.714	13.81	0.395	0.290	<u>22.07</u>	<u>0.867</u>	0.906
Kim et al. [11]	25.03	0.897	0.946	19.66	0.740	0.730	<u>15.25</u>	0.431	0.416	23.10	0.865	0.925

Table 2. Quantitative results for different methods on our dataset. A detailed analysis is presented in the paper. The first, second, and third best results are marked by bold font, bold font with underline, and underline only.

removal also increases in this case. When reflection is too weak, we even do not need to remove it because it is almost invisible. When the reflection is moderate or strong, the image quality suffers heavily.

Qualitative results. We present qualitative results in Fig. 8 to analyze the results further. The perceptual performance is consistent with the quantitative results on different reflection types.

The performance on SRST is not satisfactory. For SRST, most methods cannot remove most reflection. For BRST, most benchmarked methods can remove the reflection when the reflection is weak. Learning free methods achieve good performance in this case [36, 3]. However, it is quite rare when the reflection is both blurry and weak.

Another problem that occurs commonly is the degradation of image quality as shown in Fig. 8. In some cases, a method may remove the reflection nearly completely, but transmission is modified at the same time.

4.2. Open Problems and Discussion

From the quantitative evaluation and perceptual results, we find that state-of-the-art single image reflection removal methods are still far from perfect, although they have achieved great performance on synthetic data or real-world data in a controlled environment.

From our experiment, we find that evaluation on synthetic data is flawed. The improvement of results on synthetic data or real-world data in a controlled environment cannot represent real improvement. If we want to apply the reflection removal method in our daily life, we should aim to achieve excellent performance on real-world data collected in the wild.

We believe we should relax strong assumptions in reflection removal. Strong assumptions may make the method perform well on a certain type of reflection but fail on other types. As analyzed above, most methods can achieve satisfactory results on blurry reflection but have poor performance on sharp reflection. It is definitely a difficult task to distinguish between the sharp reflection and transmission.

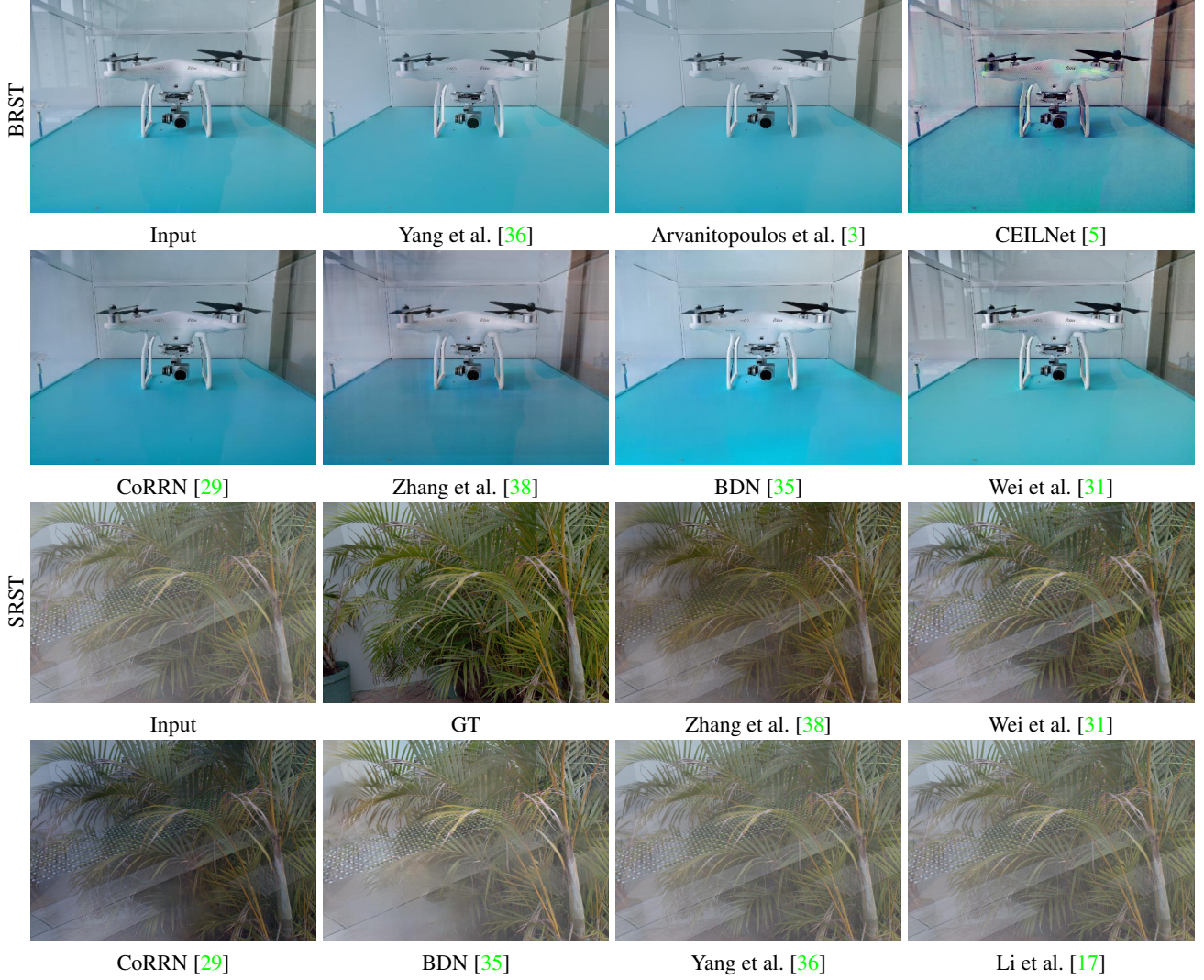


Figure 8. Most methods cannot remove the sharp reflection. This is probably because learning-based methods are trained on synthetic data where R is blurry. Learning free methods often assume reflection is blurry. However, sharp reflection is quite common in the real world. Figure best viewed in the electronic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

4.3. Reflection Removal on Raw Images

It is unclear whether single image reflection removal can benefit from raw images. However, we keep the raw data in our dataset since applying raw images has achieved amazing results on low-level computer vision tasks, including low-light image enhancement [4], super-resolution [33], image denoising [37] and ISP [?, ?]. We leave the study for raw images on single image reflection removal to future work. To the best of our knowledge, our dataset is the first dataset containing raw images for single image reflection removal.

5. Conclusion

In this work, we propose a new dataset CDR for single image reflection removal. Compared with other reflection removal datasets, our dataset is categorized according to reflection types, has the perfect alignment, and contains diverse scenes. We carefully categorize the captured images into different classes and analyze the performance of state-of-the-art methods. The experimental results show that the performance of these state-of-the-art methods is highly related to the appearance and intensity of reflection. The methods cannot achieve top performance when the pre-adopted assumptions do not hold on real-world images. We believe researchers can utilize our benchmark to do research on real-world data in the wild. In addition to RGB images, the raw data is also provided for future study.

References

- [1] Amit Agrawal, Ramesh Raskar, Shree K. Nayar, and Yuanzhen Li. Removing photography artifacts using gradient projection and flash-exposure sampling. *TOG*, 2005. 2
- [2] Jean-Baptiste Alayrac, Joao Carreira, and Andrew Zisserman. The visual centrifuge: Model-free layered video representations. In *CVPR*, 2019. 2
- [3] Nikolaos Arvanitopoulos, Radhakrishna Achanta, and Sabine Susstrunk. Single image reflection suppression. In *CVPR*, 2017. 2, 5, 6, 7, 8
- [4] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *CVPR*, 2018. 8
- [5] Qingnan Fan, Jiaolong Yang, Gang Hua, Baoquan Chen, and David Wipf. A generic deep architecture for single image reflection removal and image smoothing. In *ICCV*, 2017. 1, 2, 5, 7, 8
- [6] H. Farid and E. H. Adelson. Separating reflections and lighting using independent components analysis. In *CVPR*, 1999. 2
- [7] Yossi Gandelsman, Assaf Shocher, and Michal Irani. "double-dip": Unsupervised image decomposition via coupled deep-image-priors. In *CVPR*, 2019. 2
- [8] Xiaojie Guo, Xiaochun Cao, and Yi Ma. Robust separation of reflection from multiple images. In *CVPR*, 2014. 2
- [9] Byeong-Ju Han and Jae-Young Sim. Reflection removal using low-rank matrix completion. In *CVPR*, 2017. 2
- [10] Y. Hong, Y. Lyu, S. Li, and B. Shi. Near-infrared image guided reflection removal. In *ICME*, 2020. 2
- [11] Soomin Kim, Yuchi Huo, and Sung-Eui Yoon. Single image reflection removal with physically-based training images. In *CVPR*, 2020. 2, 5, 6, 7
- [12] Naejin Kong, Yu-Wing Tai, and Joseph S. Shin. A physically-based approach to reflection separation: from physical modeling to constrained optimization. *TPAMI*, 2014. 2
- [13] Chenyang Lei and Qifeng Chen. Robust reflection removal with reflection-free flash-only cues. In *CVPR*, 2021. 2
- [14] Chenyang Lei, Xuhua Huang, Mengdi Zhang, Wenxiu Sun, Qiong Yan, and Qifeng Chen. Polarized reflection removal with perfect alignment in the wild. In *CVPR*, 2020. 1, 2, 3, 5
- [15] Chao Li, Yixiao Yang, Kun He, Stephen Lin, and John E. Hopcroft. Single image reflection removal through cascaded refinement. In *CVPR*, 2020. 1, 3, 5, 7
- [16] Yu Li and Michael S Brown. Exploiting reflection change for automatic reflection removal. In *ICCV*, 2013. 2
- [17] Yu Li and Michael S Brown. Single image layer separation using relative smoothness. In *CVPR*, 2014. 5, 7, 8
- [18] Yu-Lun Liu, Wei-Sheng Lai, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Learning to see through obstructions. In *CVPR*, 2020. 2
- [19] Youwei Lyu, Zhaopeng Cui, Si Li, Marc Pollefeys, and Boxin Shi. Reflection separation using a pair of unpolarized and polarized images. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. 2
- [20] Daiqian Ma, Renjie Wan, Boxin Shi, Alex C. Kot, and Ling-Yu Duan. Learning to jointly generate and separate reflections. In *ICCV*, 2019. 2
- [21] Wieschollek Patrick, Gallo Orazio, Gu Jinwei, and Kautz Jan. Separating reflection and transmission images in the wild. In *ECCV*, 2018. 2
- [22] Abhijith Punnappurath and Michael S. Brown. Reflection removal using a dual-pixel sensor. In *CVPR*, 2019. 2
- [23] Bernard Sarel and Michal Irani. Separating transparent layers through layer information exchange. In *ECCV*, 2004. 2
- [24] Bernard Sarel and Michal Irani. Separating transparent layers of repetitive dynamic behaviors. In *ICCV*, 2005. 2
- [25] Yoav Schechner, Joseph Shamir, and Nahum Kiryati. Polarization and statistical analysis of scenes containing a semireflector. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 2000. 2
- [26] Chao Sun, Shuaicheng Liu, Taotao Yang, Bing Zeng, Zhengning Wang, and Guanghui Liu. Automatic reflection removal using gradient intensity and motion cues. In *ACM-MM*, 2016. 2
- [27] Richard Szeliski, Shai Avidan, and P Anandan. Layer extraction from multiple images containing reflections and transparency. In *CVPR*, 2000. 2
- [28] Renjie Wan, Boxin Shi, Ling-Yu Duan, Ah-Hwee Tan, and Alex C Kot. Benchmarking single-image reflection removal algorithms. In *ICCV*, 2017. 1, 3, 4, 5
- [29] Renjie Wan, Boxin Shi, Haoliang Li, Ling-Yu Duan, Ah-Hwee Tan, and Alex Kot Chichung. Corrn: Cooperative reflection removal network. *TPAMI*, 2019. 2, 5, 6, 7, 8
- [30] Qiaosong Wang, Haiting Lin, Yi Ma, Sing Bing Kang, and Jingyi Yu. Automatic layer separation using light field imaging. *arXiv preprint arXiv:1506.04721*, 2015. 2
- [31] Kaixuan Wei, Jiaolong Yang, Ying Fu, David Wipf, and Hua Huang. Single image reflection removal exploiting misaligned training data and network enhancements. In *CVPR*, 2019. 2, 3, 4, 5, 6, 7, 8
- [32] Qiang Wen, Yinjie Tan, Jing Qin, Wenxi Liu, Guoqiang Han, and Shengfeng He. Single image reflection removal beyond linearity. In *CVPR*, 2019. 2
- [33] Xiangyu Xu, Yongrui Ma, and Wenxiu Sun. Towards real scene super-resolution with raw images. In *CVPR*, 2019. 8
- [34] Tianfan Xue, Michael Rubinstein, Ce Liu, and William T Freeman. A computational approach for obstruction-free photography. *TOG*, 2015. 2
- [35] Jie Yang, Dong Gong, Lingqiao Liu, and Qinfeng Shi. Seeing deeply and bidirectionally: a deep learning approach for single image reflection removal. In *ECCV*, 2018. 2, 3, 5, 7, 8
- [36] Yang Yang, Wenye Ma, Yin Zheng, Jian-Feng Cai, and Weiyu Xu. Fast single image reflection suppression via convex optimization. In *CVPR*, 2019. 2, 5, 6, 7, 8
- [37] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Cycleisp: Real image restoration via improved data synthesis. In *CVPR*, 2020. 8
- [38] Xuaner Zhang, Ren Ng, and Qifeng Chen. Single image reflection separation with perceptual losses. In *CVPR*, 2018. 1, 2, 3, 4, 5, 7, 8