

# Single-Image Specular Highlight Removal via Real-World Dataset Construction

Zhongqi Wu, Chuanqing Zhuang, Jian Shi, Jianwei Guo, Jun Xiao, Xiaopeng Zhang, Dong-Ming Yan

This supplementary material provides additional results of our method. In Section I, we describe the detailed formula derivation for obtaining specular-free images, and also give more analysis of the dataset. In Section II, additional comparison with representative state-of-the-art methods are presented.

## I. DATASET

### A. Theoretical background

The specular reflection is governed by the well-known Fresnel equation [2] that describes the reflection and transmission of light (or electromagnetic radiation in general) when incident on an interface between different optical media. It is assumed that the incident wave shoots from the medium with a refractive index of  $n_1$  to the medium with a refractive index of  $n_2$ . The incident angle is  $\theta_1$ , the refraction angle is  $\theta_2$ , the electric vector is  $E_1$ , and the wave vector is  $k_1$  (see Fig. 1). The electric vector and wave vector of reflected wave and refracted wave are denoted as  $E_2$ ,  $k_2$ ,  $E_3$  and  $k_3$ , respectively. The direction of the polarized light (vibration vector) with its electric field along the plane of incidence is denoted as p-polarized, while light whose electric field is normal to the plane of incidence is called s-polarized.

Specifically, at the incident point, the relationship between the reflected instantaneous electric vector and the incident electric vector can be expressed by the following Fresnel equation (regardless of refraction):

$$\frac{E_{s2}}{E_{s1}} = \frac{n_1 \cos \theta_1 - n_2 \cos \theta_2}{n_1 \cos \theta_1 + n_2 \cos \theta_2} = -\frac{\sin(\theta_1 - \theta_2)}{\sin(\theta_1 + \theta_2)}, \quad (1)$$

$$\frac{E_{p2}}{E_{p1}} = \frac{n_2 \cos \theta_1 - n_1 \cos \theta_2}{n_2 \cos \theta_1 + n_1 \cos \theta_2} = \frac{\tan(\theta_1 - \theta_2)}{\tan(\theta_1 + \theta_2)}. \quad (2)$$

When the incident light is linearly polarized along any direction, the angles between p-polarization and the polarization direction of the incident light (see Fig. 2), reflected light and refracted light are  $\alpha_1$ ,  $\alpha_2$  and  $\alpha_3$ , respectively. We get:

$$\tan \alpha_1 = \frac{E_{s1}}{E_{p1}}; \tan \alpha_2 = \frac{E_{s2}}{E_{p2}}; \tan \alpha_3 = \frac{E_{s3}}{E_{p3}}. \quad (3)$$

Z. Wu, J. Shi, J. Guo, X. Zhang, D.-M. Yan are with NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China. (Zhongqi Wu and Chuanqing Zhuang are joint first authors. Jianwei Guo and Jun Xiao are the corresponding authors. E-mail: jianwei.guo@nlpr.ia.ac.cn, xiaojun@ucas.ac.cn)

C. Zhuang and J. Xiao are with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China.

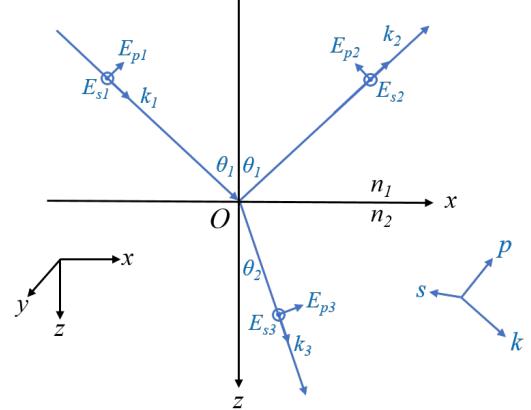


Fig. 1. Decomposition of light vector.

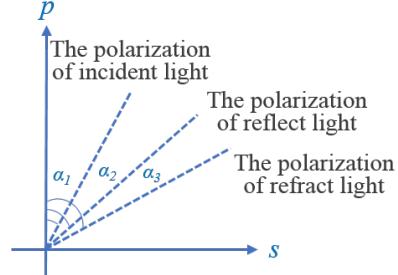


Fig. 2. Schematic of polarization of light.

From the law of reflection,

$$n_1 \sin \theta_1 = n_2 \sin \theta_2. \quad (4)$$

Regardless of the total reflection, then

$$\cos \theta_2 = \sqrt{1 - \sin^2 \theta_2} = \sqrt{n_2^2 - n_1^2 \sin^2 \theta_1}. \quad (5)$$

Combining Eqs (3), (4) with (5) and using the Fresnel equation, we obtain:

$$\tan \alpha_2 = \frac{\cos \theta_1 \sqrt{n_2^2 - n_1^2 \sin^2 \theta_1} + n_1 \sin^2 \theta_1}{\cos \theta_1 \sqrt{n_2^2 - n_1^2 \sin^2 \theta_1} - n_1 \sin^2 \theta_1} \tan \alpha_1, \quad (6)$$

$$\tan \alpha_3 = [\cos \theta_1 \sqrt{n_2^2 - n_1^2 \sin^2 \theta_1} + n_1 \sin^2 \theta_1] \tan \alpha_1. \quad (7)$$

It can be known from the above formula that both reflected light and refracted light are linearly polarized light, and the azimuth of vibration changes with the incident angle  $\theta_1$ .

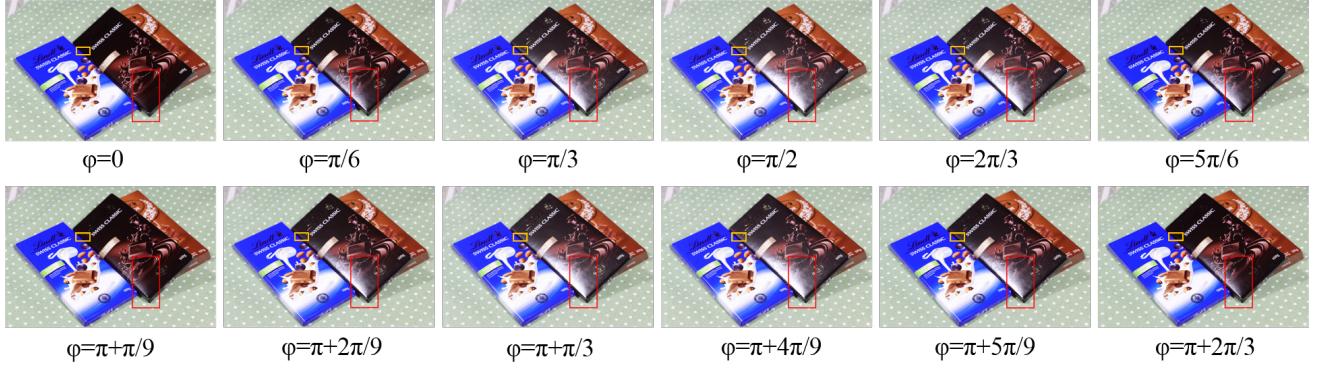


Fig. 3. A group of specular highlight images with 12 fixed polarization angles.

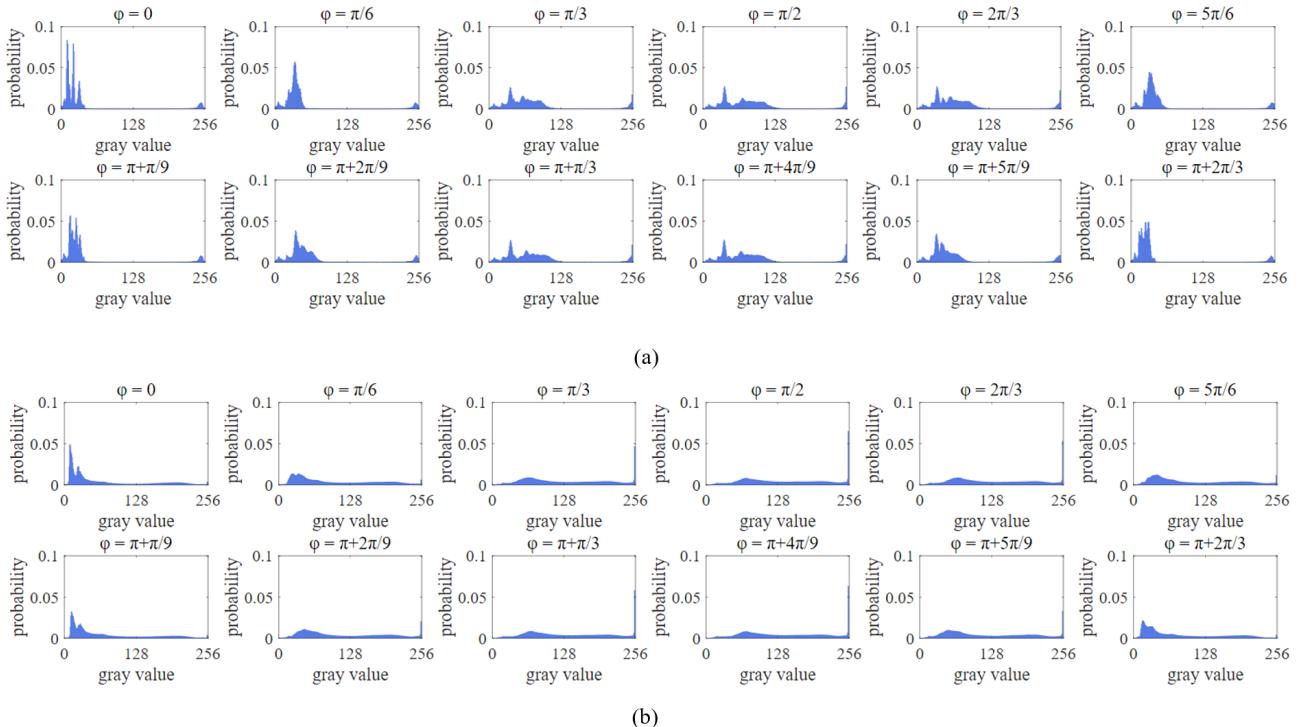


Fig. 4. Histograms of light intensity in local windows of specular highlight images with 12 fixed polarization angles. (a) and (b) show the information of yellow and red local windows marked in Fig. 3, respectively.

### B. Cross polarization for imaging

The polarization directions of the specular reflection are related to the surface of the objects. However, these specular reflections tend to travel in different directions because of the various surface normal of objects. Consequently, according to the law of light reflection, when the locations of the light source, the objects and the camera are determined, the light emitted from the light source and reflected by the object should meet the reflection law to reach the camera, which limits the camera to receive only a few specular reflections (i.e., specular highlights in the image) generated on the object surface, and their polarization directions are basically same in this case. In practice, when we captured the ground truth, the positions of the light source, the objects and the camera are fixed, and

when one or more of them are adjusted, we will change the polarization angle of the polarizer fixed in front of the camera to remove the specular highlights and get the ground truth. Thus, although we cannot find a fixed polarization angle for all specular highlights, we could almost find a good angle by rotating the polarizer for a specific specular highlight.

According to the discussion above, the highlights are linearly polarized and their polarization directions are basically same. Therefore, when the angle  $\phi$  between the polarization direction of reflected light and the axis of the CPL is  $\pi/2$ , the specularity of the reflected light can be removed.

In order to be closer to the actual situation, we took photos under natural light for testing. Our network processes images in linear space. Although natural light is not polarized light, it can be regarded as the superposition of polarized

TABLE I  
QUANTITATIVE COMPARISON ON OUR DATASET (SEE FIG. 5). THE BEST RESULT OF EACH MEASUREMENT IS MARKED IN **BOLD** FONT.

Scenes	Methods	MSE/ $1e^{-2}\downarrow$	SSIM $\uparrow$	PSNR $\uparrow$
<i>Chocolate</i>	Ours	<b>0.10</b>	<b>0.9876</b>	<b>28.4635</b>
	Multi-class GAN [4]	0.16	0.9728	26.3812
	Spec-CGAN[3]	0.14	0.9315	26.7707
	Shen <i>et al.</i> 2008 [6]	0.15	0.8251	25.2515
	Shen <i>et al.</i> 2009 [5]	0.30	0.9758	23.8261
	Akashi <i>et al.</i> 2014 [1]	0.62	0.8009	20.7315
	Yamamoto <i>et al.</i> 2017 [7]	15.45	0.3599	6.9627
<i>Balls</i>	Ours	<b>0.03</b>	<b>0.9946</b>	<b>32.7073</b>
	Multi-class GAN [4]	0.08	0.9833	28.9527
	Spec-CGAN[3]	0.08	0.9771	27.8457
	Shen <i>et al.</i> 2008 [6]	0.12	0.9032	24.8676
	Shen <i>et al.</i> 2009 [5]	0.13	0.9797	26.9133
	Akashi <i>et al.</i> 2014 [1]	1.51	0.8877	16.7794
	Yamamoto <i>et al.</i> 2017 [7]	7.30	0.7531	11.3651
<i>Toys</i>	Ours	<b>0.15</b>	<b>0.9849</b>	<b>26.5272</b>
	Multi-class GAN [4]	0.16	0.9801	26.1796
	Spec-CGAN[3]	0.31	0.9559	22.8340
	Shen <i>et al.</i> 2008 [6]	0.55	0.8679	19.3007
	Shen <i>et al.</i> 2009 [5]	0.40	0.9731	22.4474
	Akashi <i>et al.</i> 2014 [1]	0.86	0.9381	22.6288
	Yamamoto <i>et al.</i> 2017 [7]	3.28	0.8358	14.8415
<i>Beans</i>	Ours	<b>0.07</b>	<b>0.9850</b>	<b>29.6032</b>
	Multi-class GAN [4]	0.23	0.9636	24.4730
	Spec-CGAN[3]	0.19	0.9467	24.3184
	Shen <i>et al.</i> 2008 [6]	0.21	0.8816	21.2695
	Shen <i>et al.</i> 2009 [5]	0.70	0.9095	20.2650
	Akashi <i>et al.</i> 2014 [1]	0.42	0.9469	21.1798
	Yamamoto <i>et al.</i> 2017 [7]	0.85	0.9028	19.4611
<i>Fruits</i>	Ours	<b>0.10</b>	<b>0.9730</b>	<b>27.9719</b>
	Multi-class GAN [4]	0.29	0.9290	22.8184
	Spec-CGAN[3]	0.26	0.9104	22.9580
	Shen <i>et al.</i> 2008 [6]	1.20	0.9032	16.6899
	Shen <i>et al.</i> 2009 [5]	1.28	0.8121	17.2048
	Akashi <i>et al.</i> 2014 [1]	2.14	0.6640	15.8418
	Yamamoto <i>et al.</i> 2017 [7]	12.14	0.5488	7.8656
<i>All test set</i>	Ours	<b>0.14</b>	<b>0.9916</b>	<b>30.4694</b>
	Multi-class GAN [4]	0.50	0.8550	23.5240
	Spec-CGAN[3]	0.36	0.9172	25.7082
	Shen <i>et al.</i> 2008 [6]	4.01	0.7462	18.3307
	Shen <i>et al.</i> 2009 [5]	1.07	0.8826	20.6226
	Akashi <i>et al.</i> 2014 [1]	3.13	0.6697	15.6855
	Yamamoto <i>et al.</i> 2017 [7]	8.46	0.6264	11.8587

light. When the linearly polarized light is used as the light, the diffuse reflection is still close to natural light due to the random diffuse scattering process, and its intensity just becomes weaker. According to our results on natural images in the wild, networks trained on our data would not lead overfitting.

### C. Dataset analysis

We use different objects and backgrounds to build our dataset. There is no overlap of objects and tablecloths between the training and test set. We also carefully control the experimental conditions such as the number of lights, light intensity and object size in each scene to ensure they have similar distributions across the training and the testing set.

**About fixed polarization angles.** Fig. 3 shows 12 images photographed with 12 fixed polarization angles in the same scene. To further display the diversity of polarization angles, we convert the original color images to grayscale images and draw the histogram information in two local windows, as shown in Fig. 4, where the local windows are marked with

TABLE II  
QUANTITATIVE COMPARISON ON OUR DATASET (SEE FIG. 6). THE BEST RESULT OF EACH MEASUREMENT IS MARKED IN **BOLD** FONT.

Scenes	Methods	MSE/ $1e^{-2}\downarrow$	SSIM $\uparrow$	PSNR $\uparrow$
<i>Box</i>	Ours	<b>0.12</b>	<b>0.9759</b>	<b>29.3924</b>
	Multi-class GAN [4]	0.18	0.9481	27.3460
	Spec-CGAN[3]	0.30	0.9155	25.2466
	Shen <i>et al.</i> 2008 [6]	0.73	0.7790	21.3683
	Shen <i>et al.</i> 2009 [5]	0.68	0.9030	21.7034
	Akashi <i>et al.</i> 2014 [1]	2.95	0.6052	15.3081
	Yamamoto <i>et al.</i> 2017 [7]	13.56	0.4626	8.6771
<i>Toys</i>	Ours	0.59	<b>0.9454</b>	22.3244
	Multi-class GAN [4]	<b>0.26</b>	0.9448	<b>25.8046</b>
	Spec-CGAN[3]	0.60	0.9450	22.1832
	Shen <i>et al.</i> 2008 [6]	0.94	0.8251	20.2715
	Shen <i>et al.</i> 2009 [5]	0.96	0.9148	20.1699
	Akashi <i>et al.</i> 2014 [1]	3.63	0.8030	14.4067
	Yamamoto <i>et al.</i> 2017 [7]	7.30	0.7531	11.3651
<i>Vase</i>	Ours	<b>0.04</b>	<b>0.9948</b>	<b>34.0177</b>
	Multi-class GAN [4]	0.09	0.9747	30.6834
	Spec-CGAN[3]	0.21	0.9530	26.8488
	Shen <i>et al.</i> 2008 [6]	0.35	0.8816	24.5772
	Shen <i>et al.</i> 2009 [5]	0.27	0.9698	25.7602
	Akashi <i>et al.</i> 2014 [1]	0.84	0.9167	20.7354
	Yamamoto <i>et al.</i> 2017 [7]	3.28	0.8358	14.8415
<i>Fruits</i>	Ours	<b>0.15</b>	<b>0.9749</b>	<b>28.1471</b>
	Multi-class GAN [4]	0.27	0.9502	25.6594
	Spec-CGAN[3]	0.45	0.9272	23.4736
	Shen <i>et al.</i> 2008 [6]	2.19	0.6856	16.5862
	Shen <i>et al.</i> 2009 [5]	2.65	0.7627	15.7736
	Akashi <i>et al.</i> 2014 [1]	2.54	0.7771	15.9580
	Yamamoto <i>et al.</i> 2017 [7]	10.49	0.6275	9.7919
<i>Flowers</i>	Ours	<b>0.08</b>	<b>0.9935</b>	<b>31.2307</b>
	Multi-class GAN [4]	0.10	0.9846	29.9598
	Spec-CGAN[3]	0.27	0.9627	25.7242
	Shen <i>et al.</i> 2008 [6]	0.74	0.8685	21.3289
	Shen <i>et al.</i> 2009 [5]	0.29	0.9794	25.3812
	Akashi <i>et al.</i> 2014 [1]	0.63	0.9513	22.0060
	Yamamoto <i>et al.</i> 2017 [7]	0.45	0.9610	22.7072

bounding boxes in Fig. 3. The horizontal axis is the gray value, and the vertical axis is the corresponding distribution density in the local window.

## II. ADDITIONAL EXPERIMENTAL RESULTS

We now give more comparisons against various highlight removal competitors, including traditional approaches (*i.e.*, Shen *et al.* 2008 [5], Shen *et al.* 2009 [5], Akashi *et al.* 2014 [1], Yamamoto *et al.* 2017 [7]) and state-of-the-art learning-based approaches (*i.e.*, Spec-CGAN[3] and Multi-class GAN [4]).

In Fig. 5 we provide the complete comparison on the testing set of our dataset used in the paper. Fig. 6 shows more examples of our testing set. Tables I and II report the corresponding quantitative comparisons. In Fig. 7 we show more comparisons on natural images. These experiments further demonstrate that our approach has significantly better performance than previous methods.

## REFERENCES

- [1] Yasuhiro Akashi and Takayuki Okatani. Separation of reflection components by sparse non-negative matrix factorization. In *Asian Conference on Computer Vision*, pages 611–625. Springer, 2014. 3, 4, 5, 6
- [2] Max Born and Emil Wolf. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013. 1



Fig. 5. Visual comparison on the testing set of our dataset. From top to bottom, the scenes we selected are *chocolate, balls, toys, beans, and fruits*. (a) Input, (b) ground-truth, (c) our results, (d)-(i) are the results of Multi-class GAN [4], Spec-CGAN [3], Shen *et al.* 2008 [5], Shen *et al.* 2009 [5], Akashi *et al.* 2014 [1], Yamamoto *et al.* 2017 [7], respectively.

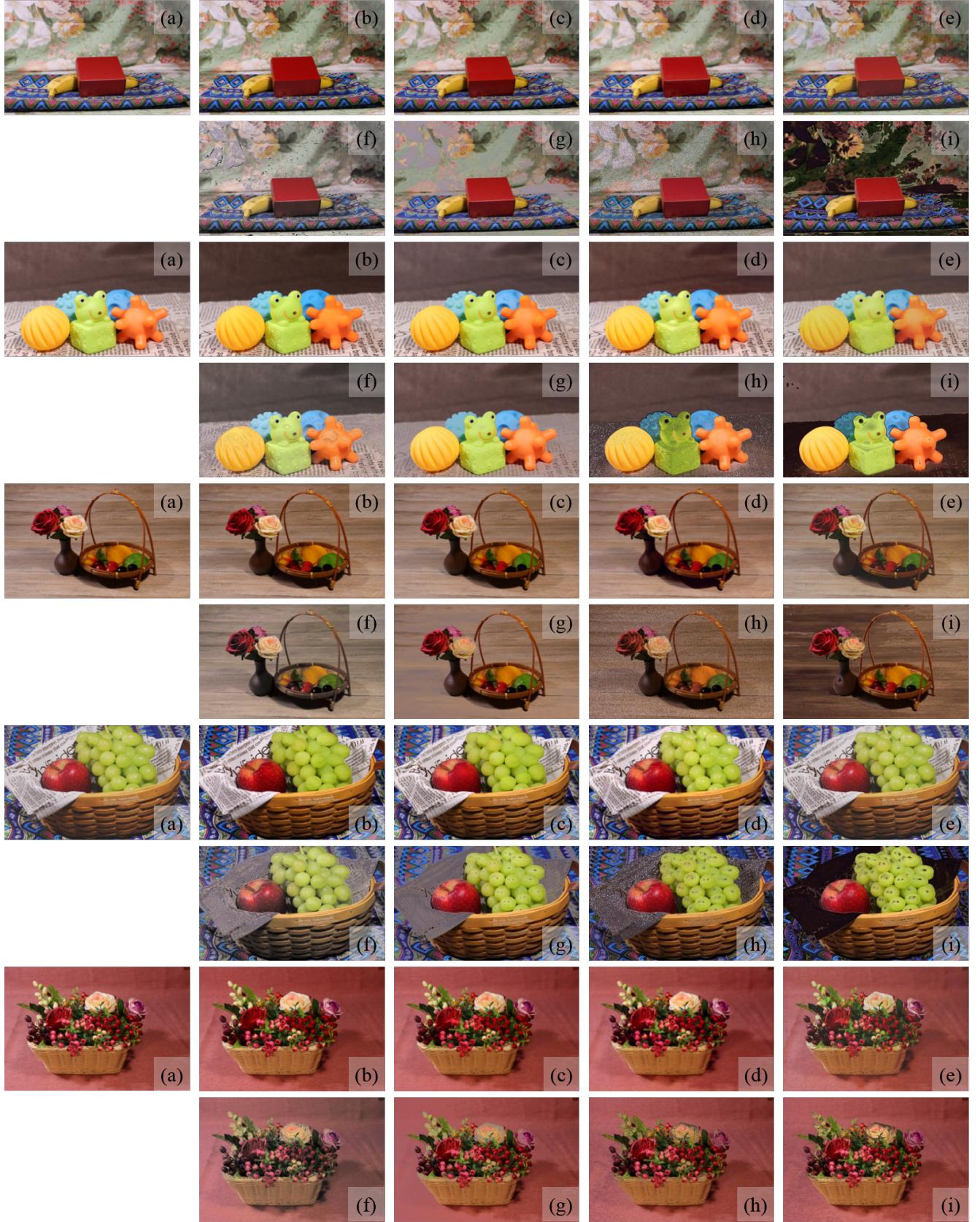


Fig. 6. Visual comparison on more testing set of our dataset. From top to bottom, the scenes we selected are *box*, *toys*, *vase*, *beans*, *fruits*, and *flowers*. (a) Input, (b) ground-truth, (c) our results, (d)-(i) are the results of Multi-class GAN [4], Spec-CGAN [3], Shen *et al.* 2008 [5], Shen *et al.* 2009 [5], Akashi *et al.* 2014 [1], Yamamoto *et al.* 2017 [7], respectively.

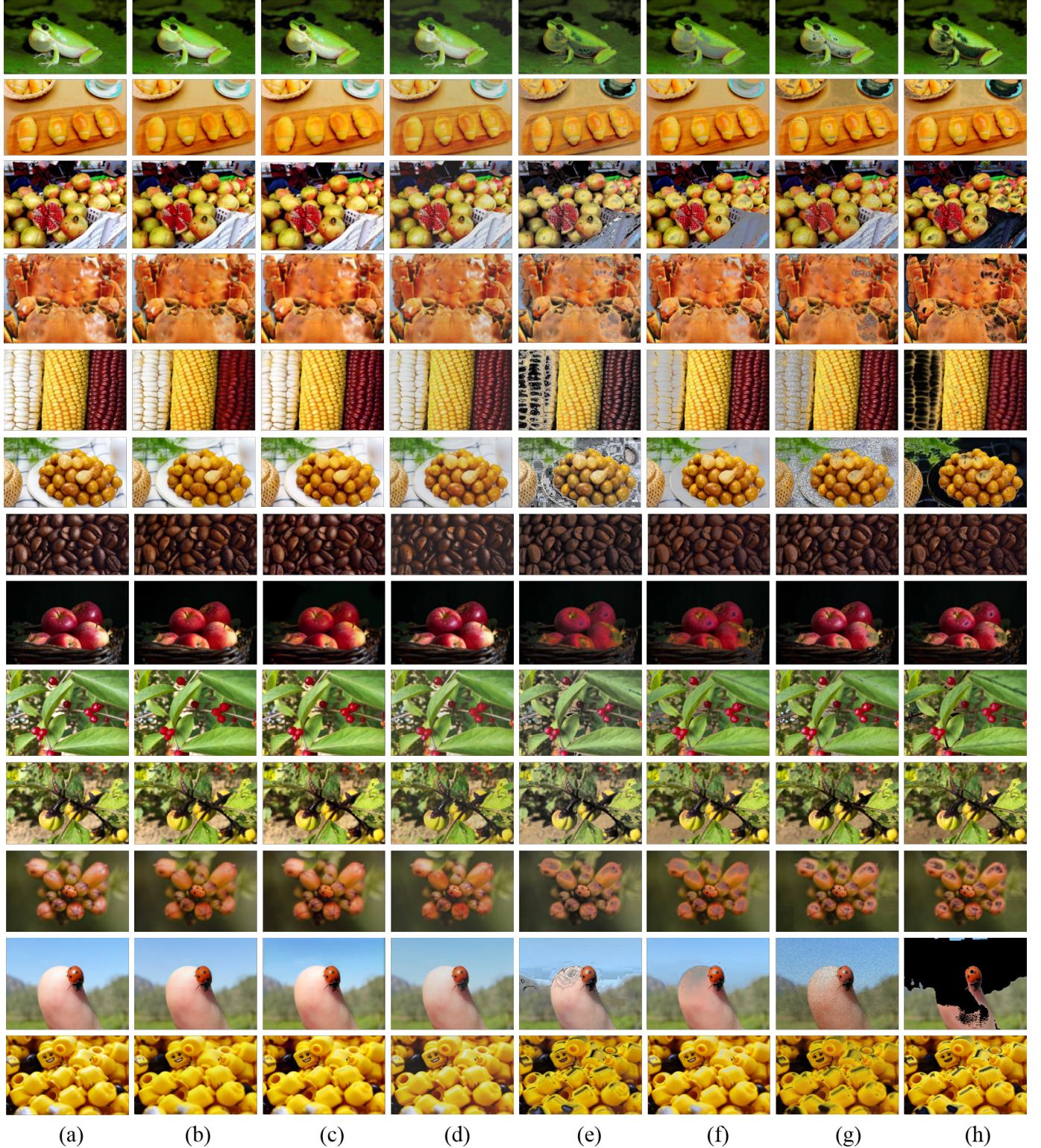


Fig. 7. Visual comparison on natural images in the wild. (a) Input, (b) ours, (c)-(h) are the results of Multi-class GAN [4], Spec-CGAN [3], Shen *et al.* 2008 [5], Shen *et al.* 2009 [5], Akashi *et al.* 2014 [1], Yamamoto *et al.* 2017 [7], respectively.

- [3] Isabel Funke, Sebastian Bodenstedt, Carina Riediger, Jürgen Weitz, and Stefanie Speidel. Generative adversarial networks for specular highlight removal in endoscopic images. In *Medical Imaging: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 10576, page 1057604, 2018. [3](#), [4](#), [5](#), [6](#)
- [4] John Lin, Mohamed El Amine Seddik, Mohamed Tamaazousti, Youssef

- Tamaazousti, and Adrien Bartoli. Deep multi-class adversarial specularity removal. In *Scandinavian Conference on Image Analysis*, pages 3–15. Springer, 2019. [3](#), [4](#), [5](#), [6](#)
- [5] Hui-Liang Shen and Qing-Yuan Cai. Simple and efficient method for specularity removal in an image. *Applied optics*, 48(14):2711–2719, 2009. [3](#), [4](#), [5](#), [6](#)

- [6] Hui-Liang Shen, Hong-Gang Zhang, Si-Jie Shao, and John H Xin. Chromaticity-based separation of reflection components in a single image. *Pattern Recognition*, 41(8):2461–2469, 2008. [3](#)
- [7] Takahisa Yamamoto, Toshihiro Kitajima, and Ryota Kawauchi. Efficient improvement method for separation of reflection components based on an energy function. In *IEEE international conference on image processing (ICIP)*, pages 4222–4226. IEEE, 2017. [3](#), [4](#), [5](#), [6](#)