One-Step Specular Highlight Removal with Adapted Diffusion Models

Supplementary Material

1. Additional Ablation Study

In addition to the losses mentioned above, we tried $\mathcal{L}_{\text{pixel}}$ loss in combination with PSNR Loss and SSIM Losses. As can be seen in Table 1 $\mathcal{L}_{\text{pixel}} + \mathcal{L}_{\text{perc}}$ outperforms well on all metrics and PSNR and SSIM losses didn't contribute to the overall quality of scores.

PSNR Loss: Peak Signal-to-Noise Ratio (PSNR) is a widely used metric for measuring image reconstruction quality. It is defined as:

$$\text{PSNR} = 20 \cdot \log_{10}\left(\frac{\text{MAX}_D}{\sqrt{\text{MSE}}}\right) \tag{1}$$

where MAX_D is the maximum possible pixel value (e.g., 1 for normalized images, 255 for 8-bit images). MSE is the Mean Squared Error between the predicted image $\hat{\mathbf{D}}$ and the ground truth image $\mathbf{D}$, given by:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{D}_i - \hat{\mathbf{D}}_i)^2 \tag{2}$$

Since PSNR is a similarity metric (higher is better), we define the PSNR loss as its negative normalized form:

$$\mathcal{L}_{\text{psnr}} = 1 - \frac{\text{PSNR} - \text{PSNR}_{\min}}{\text{PSNR}_{\max} - \text{PSNR}_{\min}} \tag{3}$$

where $\text{PSNR}_{\min}$ and $\text{PSNR}_{\max}$ are the expected PSNR range (e.g., 10 dB to 50 dB).

SSIM Loss: Structural Similarity Index (SSIM) is designed to evaluate image similarity by considering luminance, contrast, and structural information. It is defined as:

$$\text{SSIM}(\mathbf{D}, \hat{\mathbf{D}}) = \frac{(2\mu_D\mu_{\hat{D}} + C_1)(2\sigma_{D\hat{D}} + C_2)}{(\mu_D^2 + \mu_{\hat{D}}^2 + C_1)(\sigma_D^2 + \sigma_{\hat{D}}^2 + C_2)} \tag{4}$$

where $\mu_D, \mu_{\hat{D}}$ are the means of $\mathbf{D}$ and $\hat{\mathbf{D}}$. $\sigma_D^2, \sigma_{\hat{D}}^2$ are the variances of $\mathbf{D}$ and $\hat{\mathbf{D}}$. $\sigma_{D\hat{D}}$ is the covariance between $\mathbf{D}$ and $\hat{\mathbf{D}}$. C_1, C_2 are small constants for numerical stability.

Since SSIM is a similarity metric (higher is better), we define the SSIM loss as its complement:

$$\mathcal{L}_{\text{ssim}} = 1 - \text{SSIM} \tag{5}$$

This loss encourages the model to maximize structural similarity between the predicted and ground truth images.

Table 1. Exploring the effect of loss functions on the model. Bold and underlined values represent the best and second best scores respectively.

Dataset	SHIQ		
Metric	PSNR ↑	SSIM ↑	MSE ↓
Loss Function			
$\mathcal{L}_1$	35.329	0.966	0.063
$\mathcal{L}_2$	34.381	0.962	0.073
$\mathcal{L}_{\text{pixel}}$	<u>35.383</u>	0.966	<u>0.060</u>
$\mathcal{L}_{\text{pixel}} + \mathcal{L}_{\text{perc}}$	**36.190**	**0.971**	**0.049**
$\mathcal{L}_{\text{pixel}} + \mathcal{L}_{\text{psnr}}$	35.223	<u>0.967</u>	0.062
$\mathcal{L}_{\text{pixel}} + \mathcal{L}_{\text{ssim}}$	34.105	0.965	0.086

2. Additional Visual Results on SHIQ Dataset

Figure 1 presents a comparison between our proposed method and the state-of-the-art model, TSHRNet, on real images from the SHIQ dataset. As depicted in the figure, the TSHRNet model struggles to restore fine details, resulting in noticeable degradation in image quality. Specifically, it tends to produce images with darker tones, fails to accurately reconstruct high-frequency details, and introduces a blurring effect, particularly in regions containing textual content. These limitations significantly impact the clarity and readability of the text within the images. In contrast, our proposed approach demonstrates superior performance by effectively preserving the original colors, enhancing sharpness, and maintaining the structural integrity of the images. As a result, our model produces outputs that are visually more accurate, detailed, and perceptually closer to the original real-world scenes.

3. Additional Visual Results on SSHR Dataset

Figure 2 shows images where our model produces better results on the SSHR dataset. In contrast, TSHRNet fails to remove the highlights effectively—it mistakenly removes non-highlighted white regions, introduces black shadows, and loses color fidelity. These results highlight the superiority of our approach in preserving both the structural and color details of the input images.

4. Additional Visual Results on PSD Dataset

Figure 3 illustrates the superior performance of ProbLoRA over LoRA on real images from the PSD dataset. In the first row, LoRA noticeably darkens the natural green color of the flower leaves and fails to remove the specular highlight on

Figure 1. A visual side-by-side comparison between our approach and top state-of-the-art methods on SHIQ dataset demonstrates that our technique more effectively removes specular highlights while maintaining the image's natural color, structure, and essential details, such as the clarity of text on reflective surfaces.

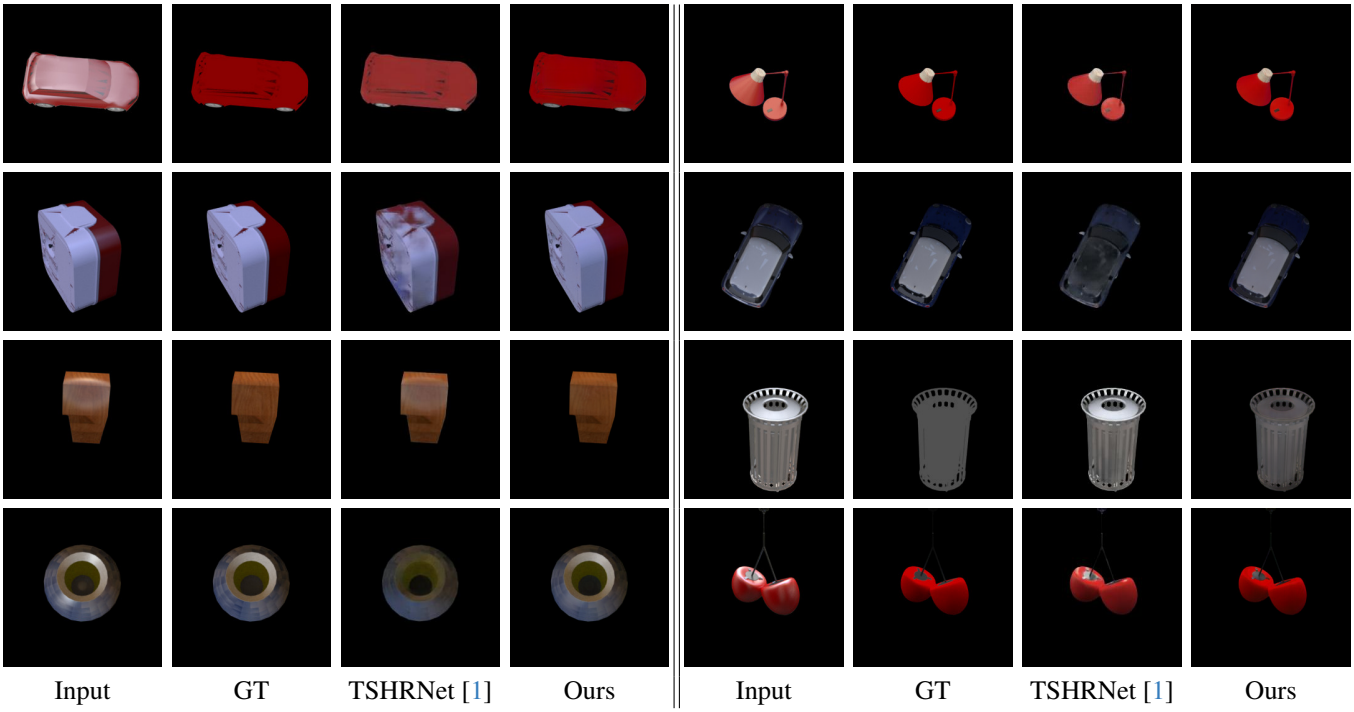

Figure 2. A side-by-side visual comparison between our approach and the state-of-the-art TSHRNet [1] on the SSHR dataset. Our method more effectively removes specular highlights while preserving the color, structure, and essential details of the image, accurately distinguishing between white regions and specular highlights.

Figure 3. A side-by-side visual comparison between LoRA and ProbLoRA on the PSD dataset.

the cucumber. In the second row, residual highlights remain on the surface of the grapes when using LoRA. The third row clearly shows a large specular region left uncorrected on the dragon fruit. In the fourth row, the text on the orange box appears significantly degraded in resolution when processed with LoRA.

References

[1] Gang Fu, Qing Zhang, Lei Zhu, Chunxia Xiao, and Ping Li. Towards high-quality specular highlight removal by leveraging large-scale synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12857–12865, 2023. 2