

Supplementary Materials: Dual-Hybrid Attention Network for Specular Highlight Removal

Xiaojiao Guo^{*†}
University of Macau
Macau, China
yc27441@um.edu.mo

Xuhang Chen^{*‡§}
Huizhou University
Huizhou, China
xuhangc@hzu.edu.cn

Shenghong Luo
University of Macau
Macau, China
ykd823332077@gmail.com

Shuqiang Wang[¶]
Shenzhen Institute of Advanced
Technology, Chinese Academy of
Sciences
Shenzhen, China
sq.wang@siat.ac.cn

Chi-Man Pun[¶]
University of Macau
Macau, China
cmpun@umac.mo

1 Computational Cost

The computational cost comparison of our methods with those of other deep learning techniques is shown in Table 1. Although the multiple attention mechanisms and U-shaped architecture of our model result in longer training and inference times, our model remains computationally efficient compared to other deep learning methods.

2 Supplementary Description of the Dataset

For the PSD and SHIQ datasets, the official dataset splits were used. However, due to the large size of the SSHR test split (18,000 samples), a random subset of 1,000 samples was selected to maintain dataset size parity with PSD (947 samples) and SHIQ (1,000 samples) and mitigate potential bias. This sampling strategy does not impact the evaluation results, as evidenced by the minimal deviation observed between the results obtained using the selected subset and the complete SSHR test set, as shown in Table 2.

3 Full Comparisons

Figure 1, 2, 3, and 4 provide a comprehensive visual comparison of all the methods discussed in the main paper, offering a more extensive performance evaluation. Overall, our method surpasses both traditional and deep learning specular highlight removal methods, excelling not only in effective highlight removal but also in the visual quality.

4 User Study

Although metrics such as Naturalness Image Quality Evaluator (NIQE) offers valuable insights into the visual quality of images, there remains a lack of unreferenced metrics specifically designed to evaluate the results of highlight removal. To address this gap and further validate the effectiveness of our DHAN-SHR in practical applications, we conducted a user study. This study aims to assess the perceptual quality of images processed by our method in

Table 1: Computational cost for deep learning methods. The batch size for evaluating the training time per iteration is uniformly set to 2.

| Method | MACs(G)↓ | Params(M)↓ | Infer Time(ms)↓ | Train Time(ms/iter)↓ |
|------------------|----------|------------|-----------------|----------------------|
| SpecularityNet | 212.87 | 17.00 | 18.33 | 99.02 |
| MG-CycleGAN | 178.07 | 12.78 | 115.13 | 213.24 |
| Unet-Transformer | 101.63 | 53.39 | 15.45 | 72.59 |
| TSHRNet | 72.76 | 116.99 | 12.00 | 35.90 |
| AHA | 92.40 | 35.86 | 11.26 | 56.07 |
| Ours | 65.69 | 4.53 | 43.82 | 119.84 |

Table 2: Metric comparison: official SSHR test split vs. randomly selected 1,000-group subset.

| Method | SSHR(Subset) | | | SSHR(Full) | | | PSNR Deviation |
|------------------|--------------|-------|--------|------------|-------|--------|----------------|
| | PSNR↑ | SSIM↑ | LPIPS↓ | PSNR↑ | SSIM↑ | LPIPS↓ | |
| Tan et al. | 10.87 | 0.778 | 0.357 | 10.78 | 0.775 | 0.359 | 0.82% |
| Yoon et al. | 28.47 | 0.916 | 0.094 | 28.32 | 0.914 | 0.094 | 0.52% |
| Shen et al. | 24.53 | 0.896 | 0.101 | 24.59 | 0.895 | 0.100 | 0.26% |
| Shen et al. | 24.00 | 0.891 | 0.094 | 24.28 | 0.892 | 0.092 | 1.16% |
| Yang et al. | 10.72 | 0.781 | 0.358 | 10.64 | 0.778 | 0.360 | 0.72% |
| Shen et al. | 27.13 | 0.914 | 0.077 | 27.20 | 0.913 | 0.077 | 0.25% |
| Akashi et al. | 29.46 | 0.924 | 0.076 | 29.48 | 0.923 | 0.076 | 0.04% |
| Huo et al. | 18.62 | 0.804 | 0.281 | 18.59 | 0.802 | 0.280 | 0.17% |
| Fu et al. | 26.15 | 0.910 | 0.076 | 26.25 | 0.909 | 0.076 | 0.41% |
| Yamamoto et al. | 26.95 | 0.902 | 0.094 | 27.04 | 0.902 | 0.093 | 0.31% |
| Saha et al. | 23.38 | 0.886 | 0.110 | 23.48 | 0.886 | 0.108 | 0.45% |
| SLRR | 26.16 | 0.916 | 0.060 | 26.34 | 0.916 | 0.059 | 0.67% |
| JSHDR | 26.43 | 0.301 | 0.059 | 26.60 | 0.304 | 0.058 | 0.66% |
| SpecularityNet | 31.07 | 0.941 | 0.041 | 30.92 | 0.940 | 0.042 | 0.47% |
| MG-CycleGAN | 28.40 | 0.874 | 0.092 | 28.24 | 0.872 | 0.092 | 0.58% |
| Unet-Transformer | 33.45 | 0.951 | 0.028 | 33.27 | 0.949 | 0.029 | 0.55% |
| TSHRNet | 33.32 | 0.950 | 0.036 | 33.14 | 0.948 | 0.036 | 0.55% |
| AHA | 31.57 | 0.944 | 0.035 | 31.61 | 0.943 | 0.036 | 0.12% |
| Ours | 36.48 | 0.964 | 0.023 | 36.29 | 0.962 | 0.024 | 0.52% |

comparison to other state-of-the-art techniques. It focuses not only on evaluating the effectiveness of highlight removal but also on the overall quality of the output images, providing a comprehensive analysis of our method’s performance.

Observing that deep learning methods significantly outperform traditional approaches, we limited participant evaluation to our method versus seven other learning-based methods to maintain the

^{*}Both authors contributed equally to this research.

[†]Also with Baoshan University.

[‡]Also with University of Macau.

[§]Also with Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences.

[¶]Corresponding authors.

focus and reduce the burden on our study participants. We invited 20 participants to evaluate the visual quality of highlight removal images. To ensure a comprehensive assessment, we randomly selected 10 images from each of the three test sets (PSD, SHIQ, and SSHR), resulting in a total of 240 images for evaluation.

We have meticulously prepared the user study form, as shown on the last page, in a Word document format and ensuring that there is no compression of the images. The images were organized into groups, each consisting of one original input image with specular highlights and eight corresponding highlight removal results, including our method and the seven other learning-based methods. The methods were anonymized to prevent bias, and the order of the methods within each group was randomized.

Participants were provided with a scoring table for each group of images, where they rated the eight methods based on the following criteria:

- (1) **Highlight Reflection Area Detection Ability:** Assessing the effectiveness and accuracy of detecting highlight areas.
- (2) **Highlight Removal Effect:** Evaluating the naturalness of highlight removal and the absence of color distortion.
- (3) **Texture Restoration Level:** Assessing the consistency of texture in the highlight-removed area with nearby regions.
- (4) **Diffuse Area Visual Quality:** Evaluating whether the diffuse areas were altered.

The scoring scale ranged from 1 (worst) to 5 (best), allowing participants to capture a spectrum of perceptible quality levels in the highlight removal results:

- 1 (Poor): The image significantly falls short in the specific criterion, marked by noticeable issues or distortions.
- 2 (Fair): The image, despite visible flaws, exhibits some elements of acceptable quality.
- 3 (Average): The image is satisfactory overall, with most elements adequately processed.
- 4 (Good): The image is well-processed, presenting only minor imperfections.
- 5 (Excellent): The image excels in the criterion, demonstrating exceptional quality.

Participants were instructed to assign a score for each criterion independently, ensuring a thorough evaluation of the various aspects of highlight removal. During the evaluation process, participants were able to zoom in on the images for a more detailed examination. For each de-highlighted image, we presented the original alongside the outputs from the eight methods, anonymizing the method names to prevent bias.

The final score for each image was determined by calculating the mean of the scores across the four criteria, with each criterion being equally weighted. This approach ensured a balanced and comprehensive assessment of each highlight removal result's overall quality. Table 3 presents the final user study scores, illustrating that our method consistently achieves the highest average score across all three test sets.

5 Performance in Videos

To validate our model's performance in video processing, we tested it on self-captured videos and those downloaded from the Internet. The results demonstrate that our model effectively removes specular

Table 3: Comparison of user study scores with seven learning-based methods. The highest-scored results are highlighted in bold, while the second-best are underlined for emphasis.

| Method | PSD | SHIQ | SSHR |
|--------------------|-------------|-------------|-------------|
| SLRR [22] | 1.65 | 1.83 | 3.09 |
| JSHDR [6] | 3.64 | <u>4.8</u> | 3.06 |
| SpecularityNet [5] | <u>3.95</u> | 3.96 | 3.72 |
| MG-CycleGAN [26] | 3.26 | 3.21 | 2.46 |
| Wu [25] | 3.65 | 3.95 | <u>4.25</u> |
| TSHRNet [7] | <u>3.95</u> | 4.53 | 3.79 |
| AHA [28] | 3.03 | 2.38 | 4.09 |
| Ours | 4.41 | 4.92 | 4.77 |

highlights from videos. We have included three sample videos in the "videos" folder to showcase the impressive specular highlight removal effects achieved by our model.



Figure 1: Comprehensive visual comparison. (a) Input specular highlight image, (b) Tan [10], (c) Yoon [31], (d) Shen [11], (e) Shen [12], (f) Yang [13], (g) Shen [14], (h) Akashi [15], (i) Huo [32], (j) Fu [18], (k) Yamamoto [19], (l) Saha [20], (m) SLRR [22], (n) JSHDR [6], (o) SpecularityNet [5], (p) MG-CycleGAN [26], (q) Wu [25], (r) TSHRNet [7], (s) AHA [28], (t) Ours, (u) GT diffuse image. The reader is encouraged to zoom-in.

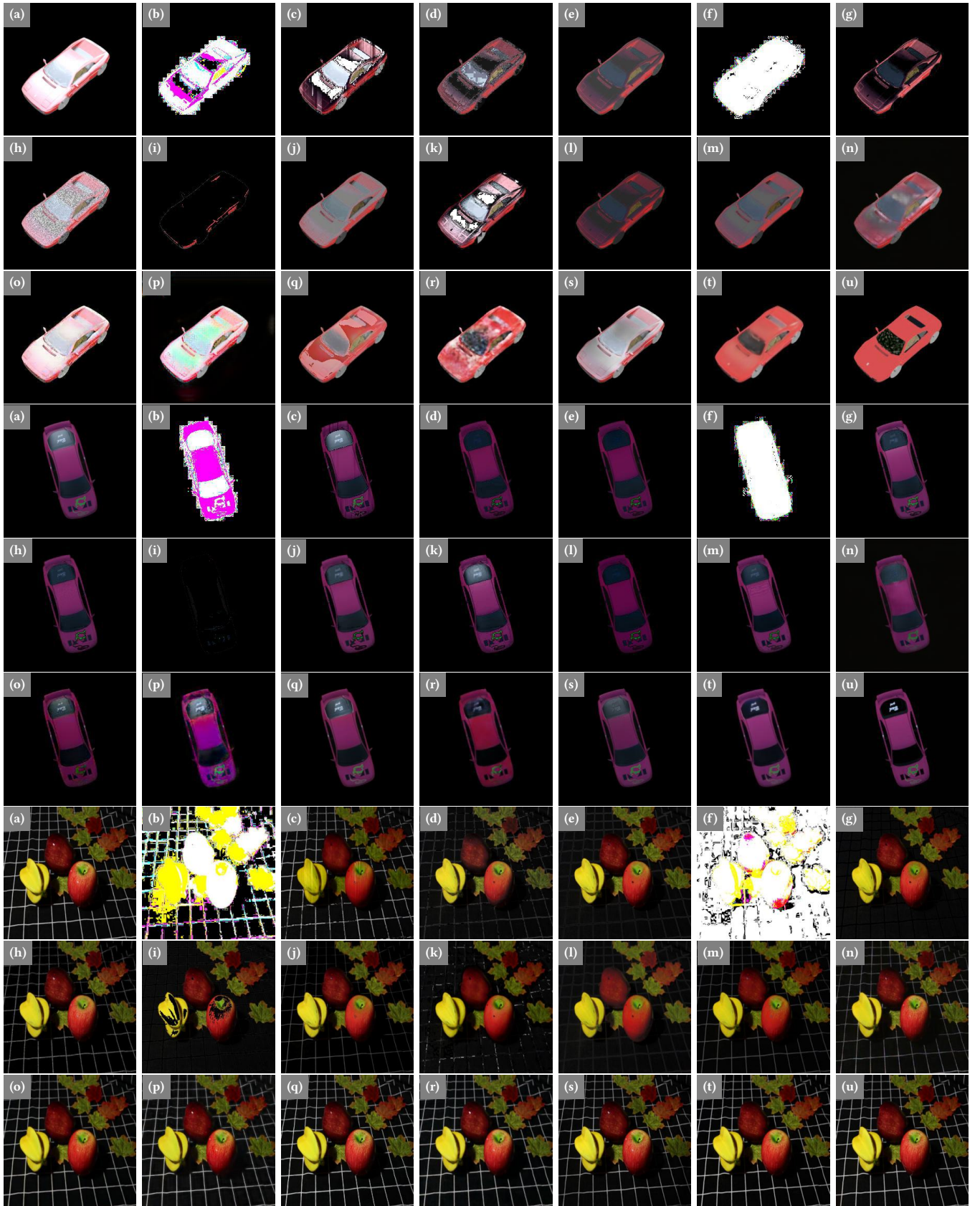


Figure 2: Comprehensive visual comparison. (a) Input specular highlight image, (b) Tan [10], (c) Yoon [31], (d) Shen [11], (e) Shen [12], (f) Yang [13], (g) Shen [14], (h) Akashi [15], (i) Huo [32], (j) Fu [18], (k) Yamamoto [19], (l) Saha [20], (m) SLRR [22], (n) JSHDR [6], (o) SpecularityNet [5], (p) MG-CycleGAN [26], (q) Wu [25], (r) TSHRNet [7], (s) AHA [28], (t) Ours, (u) GT diffuse image. The reader is encouraged to zoom-in.



Figure 3: Comprehensive visual comparison. (a) Input specular highlight image, (b) Tan [10], (c) Yoon [31], (d) Shen [11], (e) Shen [12], (f) Yang [13], (g) Shen [14], (h) Akashi [15], (i) Huo [32], (j) Fu [18], (k) Yamamoto [19], (l) Saha [20], (m) SLRR [22], (n) JSHDR [6], (o) SpecularityNet [5], (p) MG-CycleGAN [26], (q) Wu [25], (r) TSHRNet [7], (s) AHA [28], (t) Ours, (u) GT diffuse image. The reader is encouraged to zoom-in.



Figure 4: Comprehensive visual comparison. (a) Input specular highlight image, (b) Tan [10], (c) Yoon [31], (d) Shen [11], (e) Shen [12], (f) Yang [13], (g) Shen [14], (h) Akashi [15], (i) Huo [32], (j) Fu [18], (k) Yamamoto [19], (l) Saha [20], (m) SLRR [22], (n) JSHDR [6], (o) SpecularityNet [5], (p) MG-CycleGAN [26], (q) Wu [25], (r) TSHRNet [7], (s) AHA [28], (t) Ours, (u) GT diffuse image. The reader is encouraged to zoom-in.

Specular Highlight Removal Methods Evaluation Form

Purpose: This study aims to assess the effectiveness of various specular highlight removal methods. Your feedback will help improve the quality of specular highlight removal techniques.

Instructions:

- You will see an original image with specular highlights followed by its processed versions.
- Please rate each processed image based on the criteria provided.
- Use the scale from 1 (Poor) to 5 (Excellent) for your rating.

Evaluation Criteria:

- (1) **Highlight Reflection Area Detection Ability:** Assessing the effectiveness and accuracy of detecting highlight areas.
- (2) **Highlight Removal Effect:** Evaluating the naturalness of highlight removal and the absence of color distortion.
- (3) **Texture Restoration Level:** Assessing the consistency of texture in the highlight-removed area with nearby regions.
- (4) **Diffuse Area Visual Quality:** Evaluating whether the diffusion areas were altered.

Image Evaluation: (**Below is a demonstration of one group of images to serve as an example for the evaluation process.)



Input



Method 1



Method 2



Method 3



Method 4



Method 5



Method 6



Method 7



Method 8

| Method | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-----------------------------|---|---|---|---|---|---|---|---|
| Highlight Detection Ability | | | | | | | | |
| Highlight Removal Effect | | | | | | | | |
| Texture Restoration Level | | | | | | | | |
| Diffuse Area Visual Quality | | | | | | | | |

Groups 2 to 30 have been omitted in this section for brevity.

Thank You Note: Thank you for your participation. Your insights are invaluable to us.