Dual-Hybrid Attention Network for Specular Highlight Removal

Xiaojiao Guo*, Xuhang Chen*, Shenghong Luo, Shuqiang Wang†, Chi-Man Pun†

Department of Computer and Information Science, University of Macau

UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU
澳門大學

ACM Multimedia 2024
Melbourne, Australia

Abstract

Specular highlight removal plays a pivotal role in multimedia applications, as it enhances the quality and interpretability of images and videos, ultimately improving the performance of downstream tasks such as content-based retrieval, object recognition, and scene understanding. Despite significant advances in deep learning-based methods, current state-of-the-art approaches often rely on additional priors or supervision, limiting their practicality and generalization capability. In this paper, we propose the Dual-Hybrid Attention Network for Specular Highlight Removal (DHAN-SHR), an end-to-end network that introduces novel hybrid attention mechanisms to effectively capture and process information across different scales and domains without relying on additional priors or supervision. DHAN-SHR consists of two key components: the Adaptive Local Hybrid-Domain Dual Attention Transformer (L-HD-DAT) and the Adaptive Global Dual Attention Transformer (G-DAT). The L-HD-DAT captures local inter-channel and inter-pixel dependencies while incorporating spectral domain features, enabling the network to effectively model the complex interactions between specular highlights and the underlying surface properties. The G-DAT models global inter-channel relationships and long-distance pixel dependencies, allowing the network to propagate contextual information across the entire image and generate more coherent and consistent highlight-free results. To evaluate the performance of DHAN-SHR and facilitate future research in this area, we compile a large-scale benchmark dataset comprising a diverse range of images with varying levels of specular highlights.

Methods

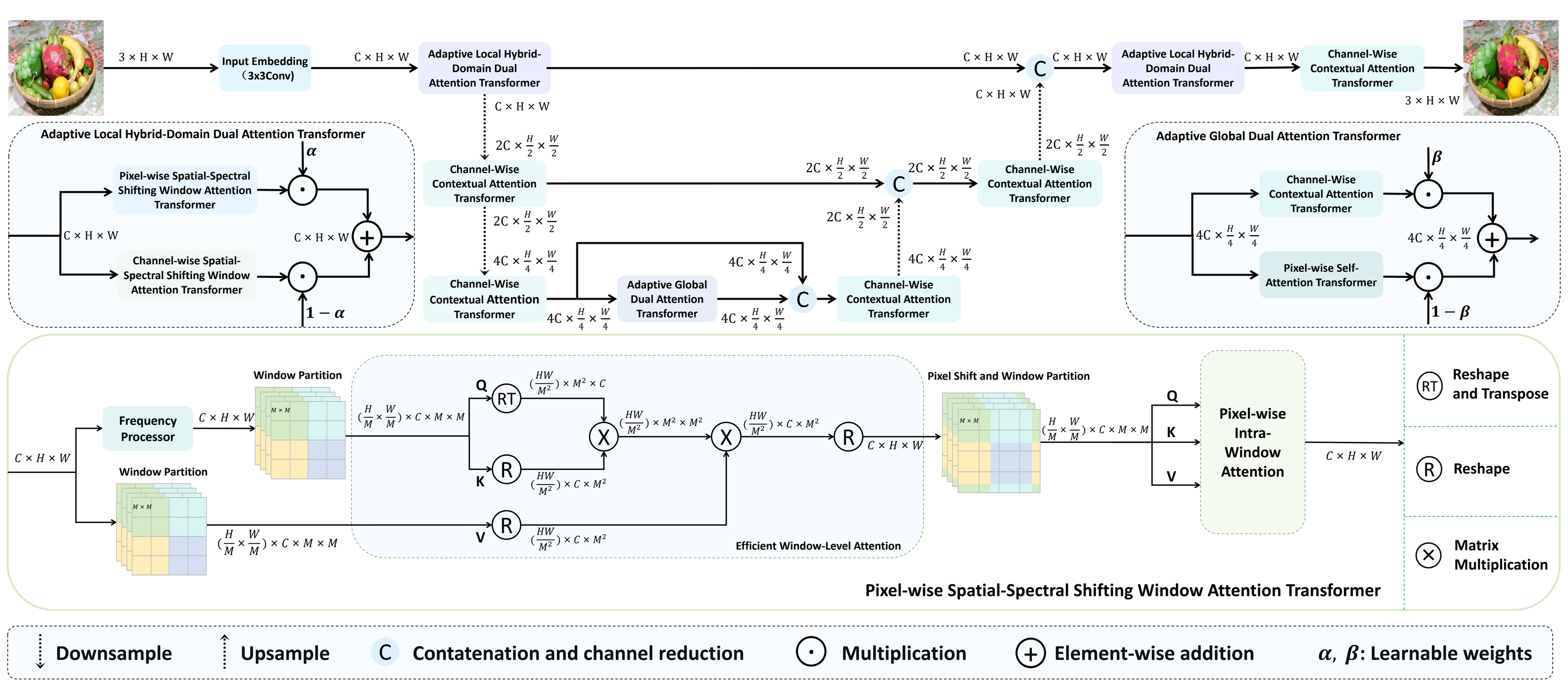

Figure 1:The overall architecture of our proposed Dual-Hybrid Attention Network for Specular Highlight Removal.

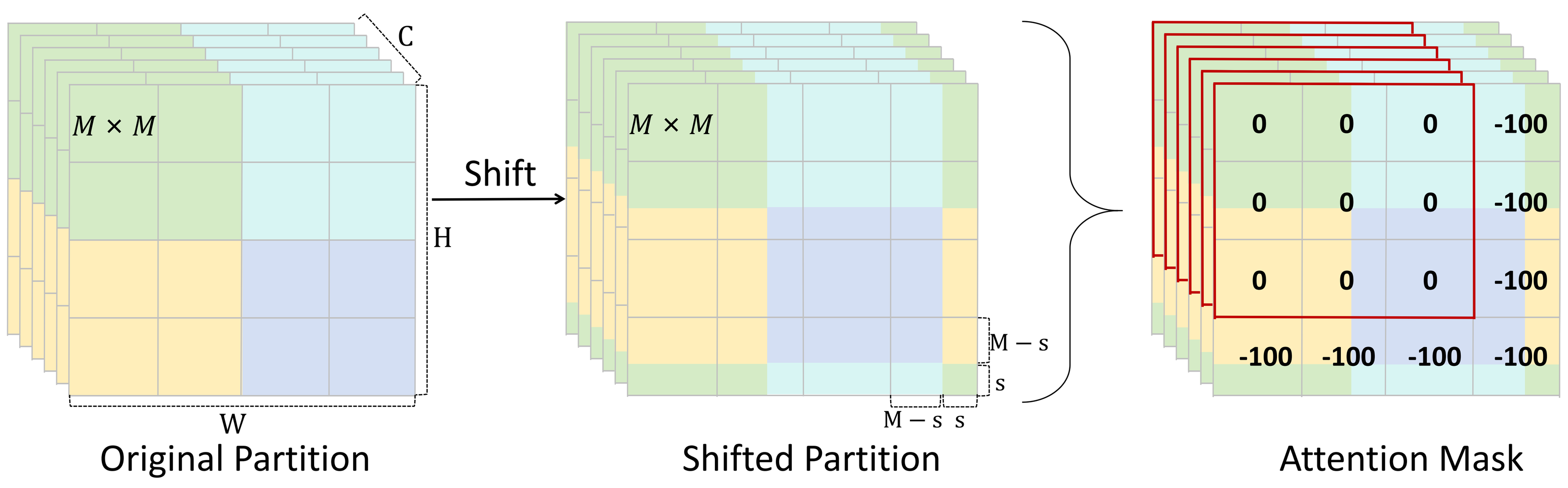

Figure 2:Illustration of the window shifting approach and the attention mask applied to the pixel-wise shifting window attention.

Corresponding Information

- cmpun@umac.mo
- https://www.cis.um.edu.mo/ cmpun/
- https://github.com/CXH-Research/DHAN-SHR

Results

Table 1:The quantitative comparison results, arranging traditional methods in the upper section and learning-based approaches below. The highest-performing results are emphasized in bold, while the second-best are underscored.

Method	PSD (947images)			SHIQ (1000images)			SSHR (1000images)		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Tan	5.44	0.218	0.746	5.47	0.483	0.823	10.87	0.778	0.357
Yoon	16.09	0.498	0.325	19.34	0.679	0.471	28.47	0.916	0.094
Shen	19.56	0.666	0.238	24.77	0.890	0.200	24.53	0.896	0.101
Shen	21.33	0.753	0.142	27.30	0.917	0.102	24.00	0.891	0.094
Yang	4.74	0.250	0.893	5.31	0.556	0.837	10.72	0.781	0.358
Shen	11.51	0.324	0.360	12.24	0.491	0.473	27.13	0.914	0.077
Akashi	17.48	0.565	0.334	21.78	0.700	0.460	29.46	0.924	0.076
Huo	20.16	0.767	0.182	23.80	0.909	0.154	18.62	0.804	0.281
Fu	15.24	0.688	0.146	16.40	0.724	0.306	26.15	0.910	0.076
Yamamoto	18.37	0.541	0.274	25.49	0.858	0.201	26.95	0.902	0.094
Saha	15.98	0.455	0.314	22.05	0.832	0.287	23.38	0.886	0.110
SLRR	13.25	0.571	0.235	14.74	0.724	0.283	26.16	0.916	0.060
JSHDR*	22.78	0.811	0.089	**37.97**	**0.980**	**0.034**	26.43	0.301	0.059
SpecularityNet	23.58	0.838	0.085	30.92	0.963	0.058	31.07	0.941	0.041
MG-CycleGAN	22.12	0.815	0.085	26.80	0.935	0.091	28.40	0.874	0.092
Wu	23.93	0.863	0.062	31.57	0.965	0.059	33.45	0.951	0.028
TSHRNet	23.30	0.826	0.097	34.57	0.972	0.044	33.32	0.950	0.036
AHA	20.79	0.845	0.084	21.42	0.903	0.165	31.57	0.944	0.035
Ours	**25.28**	**0.883**	**0.049**	33.81	0.975	0.039	**36.48**	**0.964**	**0.023**

* JSHDR's source code is not publicly available; the results are obtained from an executable file provided by its authors.

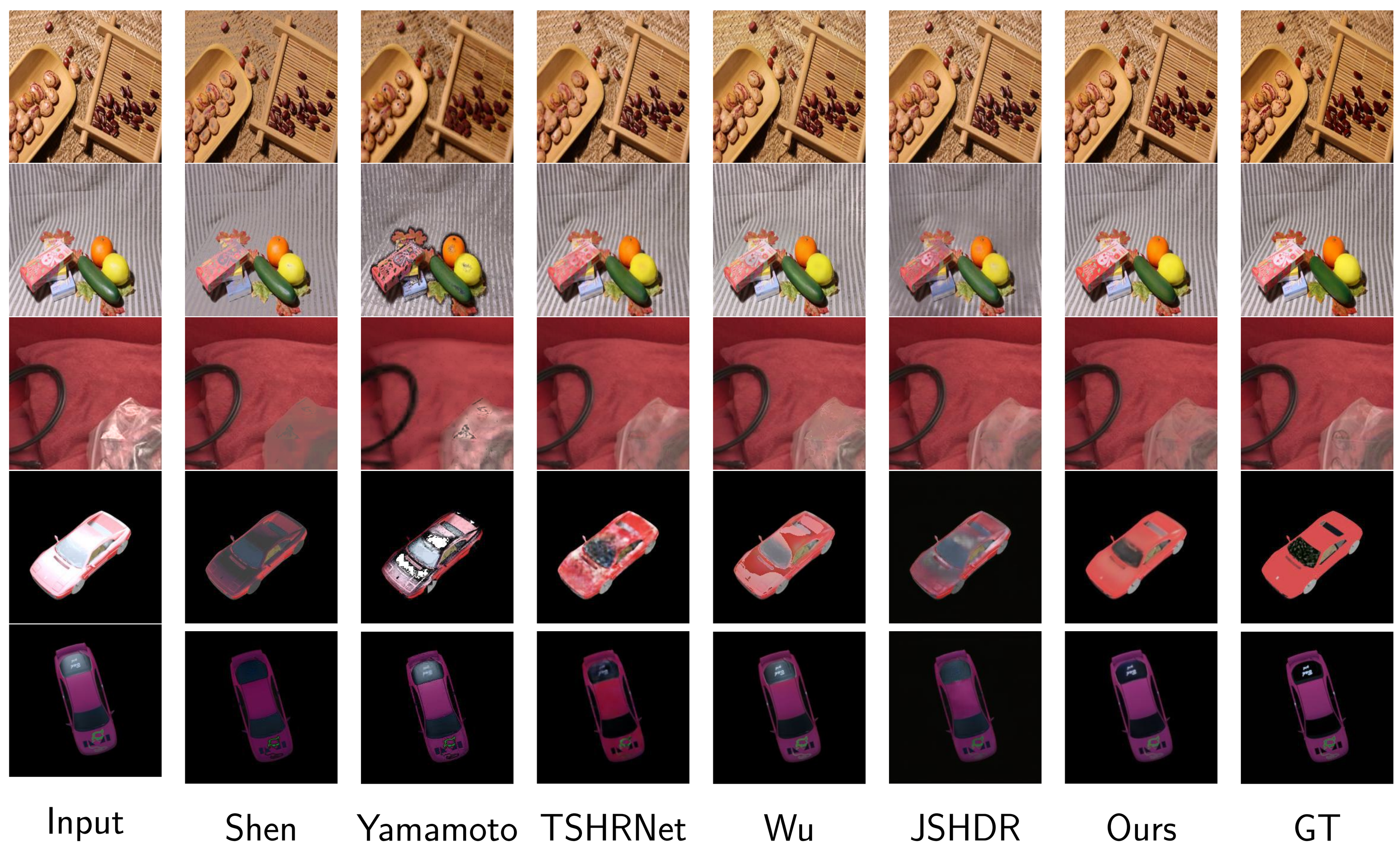

Figure 3:Visual comparative analysis of our method against leading SOTA approaches, highlighting our superior ability to remove specular highlights while preserving the original image's color tone, structure, and crucial details, such as text clarity on reflective surfaces.

Conclusion

DHAN-SHR leverages novel adaptive hybrid attention mechanisms, excelling at capturing both local and global dependencies, and at the same time, incorporating spectral domain features to effectively model complex interactions between specular highlights and surface properties. We assembled an extensive benchmark dataset combining images from three different highlight removal datasets. Experimental results demonstrate that DHAN-SHR outperforms 18 state-of-the-art methods across various test datasets, both quantitatively and qualitatively.

References

[1] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte. Swinir: Image restoration using swin transformer. In *ICCV*, pages 1833–1844, 2021.

[2] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *ICCV*, pages 10012–10022, 2021.