MICCAI

A Two-Stage Method for Specular Highlight Detection and Removal in Medical Images

Zefeng Li[1], Mingyue Cui[1], Daosong Hu[1], Jin Gong[2], Jingchong Weng[1], Zeyu Zhang[3], Lele Tian[4], Mengran Li[1], and Kai Huang[1✉]

[1] Sun Yat-sen University, Guangzhou, China
[2] The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, China
[3] City University of Macau, Macau, China
[4] Macao Polytechnic University, Macao, China
huangk36@mail2.sysu.edu.cn

Abstract. In minimally invasive surgeries, such as endoscopic and ophthalmic procedures, specular highlights on tissue and instrument surfaces can obscure critical details, compromising surgical safety and precision. Traditional methods rely on color segmentation and filtering optimization but are highly sensitive to lighting variations and produce suboptimal restoration. While deep learning enhances detection robustness, its effectiveness is constrained by the scarcity of annotated medical data and unnatural boundary transitions in restored regions. To address these challenges, this paper proposes a two-stage hierarchical network framework. First, a Hierarchical Feature Attention Network (HFA-Net) is designed, integrating spatial-shift segmented attention (S^2MLP), dual-flow attention (DFA), multi-scale feature fusion (SFF), and partial mask convolution (PMConv) to achieve precise detection and removal of specular highlights. Second, a large-mask inpainting model (LaMa) is introduced, utilizing dilated mask expansion to enhance contextual awareness and improve texture consistency in the restored regions. To address the scarcity of medical highlight datasets, we construct four specialized datasets covering various surgical scenarios, including ophthalmic injections and instrument reflections, while also incorporating publicly available data to enhance model generalization. Experimental results demonstrate that the proposed method outperforms existing approaches across six datasets in terms of detection accuracy and restoration quality, particularly excelling in complex textures and natural boundary transitions. Our code is available at https://github.com/tkllndxn/highlight-removal.

Keywords: Specular highlight detection and removal · Medical image inpainting · Two-stage strategy network.

1 Introduction

In modern surgical procedures, particularly in endoscopic and ophthalmic surgeries, precise visual information is crucial for surgical success. However, intense

illumination often induces specular highlights in moist and smooth human tissues and surgical instruments [20]. These highlights obscure critical details, impair surgeons' visual perception, and reduce surgical precision and safety [13]. Therefore, the detection and removal of specular highlights have become a significant research focus. Traditional specular highlight detection methods primarily rely on color threshold segmentation [12, 22], which analyzes color channels to locate highlight regions. However, these approaches are highly sensitive to lighting variations, exhibit inconsistent performance in complex scenarios, and require extensive parameter tuning. In contrast, deep learning-based approaches provide more robust highlight detection and improved accuracy [8, 18]. Specular highlight removal methods generally fall into filtering-based and optimization-based techniques. Filtering-based methods suppress highlights via various filters [13, 19], yet they often fail to restore fine textures, leading to inconsistencies. Optimization-based methods typically seek an optimal balance between highlight removal effectiveness and image quality preservation by carefully designing objective functions. However, these methods exhibit limitations in handling complex textures or highly reflective surfaces. Recently, deep learning-based approaches have been introduced for specular highlight removal [9, 24], demonstrating superior generalization capabilities and enhanced restoration quality. However, these methods heavily rely on large-scale training datasets, which are relatively scarce in the medical imaging domain, making it challenging to accurately learn the structural features of highlight regions. Therefore, the primary challenge in current specular highlight removal research lies in accurately detecting highlight regions and generating realistic restoration results, ensuring a natural transition between highlight and non-highlight regions. In response to the challenges mentioned above, the main contributions of this paper are as follows:

Construction of Medical Scene Datasets: To address the scarcity of medical highlight data, we construct four medical-scene highlight datasets covering typical surgical scenarios (ocular surface, fundus, surgical instruments, and specular highlights on endoscopic tissue surfaces.). Additionally, two publicly available datasets are incorporated to enhance experimental comparability and validate the model's generalizability, thereby improving adaptability across different surgical environments.

Improved Accuracy in Specular Highlight Detection and Removal: We propose a hierarchical feature attention network(HFA-Net) that effectively distinguishes specular highlights from normal white regions, improving detection accuracy. Simultaneously, our approach removes highlights while preserving image structures, reducing edge blurring and information loss.

A more realistic restoration effect: This paper adopts a two-stage strategy, where the first stage detects and removes specular highlights, and the second stage restores the highlight-free regions, ensuring smoother transitions and improved texture consistency. Experimental results demonstrate that the proposed method outperforms existing approaches in both subjective visual quality and objective evaluation metrics.
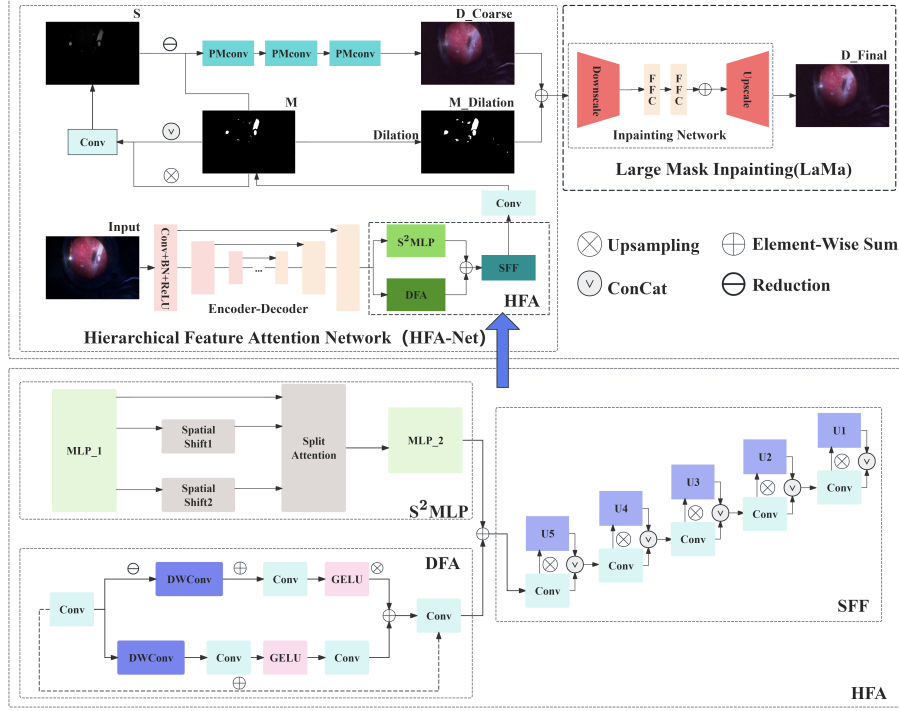
Fig. 1. The overall architecture of the proposed two-stage specular highlight detection and removal network. The lower half of the image shows the HFA module.

2   Method

Improved Dichromatic Reflection Model. The dichromatic reflection model is a widely used reflection decomposition framework in highlight removal and image restoration tasks [15]. To adapt this model for medical imaging scenarios, we propose an improved Dichromatic Reflection Model and adopt a two-stage specular highlight detection and removal strategy. A highlight mask M is introduced to distinguish between highlight and non-highlight regions, formulated as:

$$D_{Coarse} = I \odot (1 - M) + (I - S) \odot M \tag{1}$$

where D_{Coarse} denotes the highlight-free image obtained in the first stage, I denotes the input image, S denotes the specular reflection component in the highlight regions, and $\odot$ represents the Hadamard product. This formulation serves as the foundation for the first-stage network in specular highlight detection and removal.

HFA-Net and Lama (Large Mask Inpainting). To address specular highlights in medical images, we adopt a two-stage strategy for precise detection

and removal, as illustrated in Figure 1. In the first stage, we propose the HFA-Net, which employs an encoder-decoder architecture integrating the HFA module, convolutional blocks, and Partial Mask Convolution (PMConv) for highlight detection and preliminary removal. The HFA module consists of Spatial Shift and Split Attention MLP (S²MLP) [23], Dual-Flow Attention (DFA), and Scale-Feature Fusion (SFF). The S²MLP enhances global feature modeling through spatial shift and split attention mechanisms but has limitations in extracting fine-grained details. The DFA incorporates a dual-branch depthwise separable convolution mechanism to effectively capture local details [5], complementing the limitations of S²MLP. The SFF integrates multi-scale features through up-sampled concatenation and convolutional operations, enhancing the analysis of complex structures and boundary regions. To reduce background interference and computational cost, we improve the Partial Convolution method [4] by introducing PMConv, which performs convolution only in highlight regions. After feature extraction, the model generates the highlight mask M, and computes the specular component S, Finally, it restores the preliminary highlight-free image D_{Coarse} using PMConv.

To address the issue of abrupt boundaries caused by feature discrepancies between highlighted and non-highlighted regions (e.g., the unnatural transition between normal tissue and pathological areas), we employ LaMa to refine the preliminary result. Specifically, LaMa takes as input the mask M generated and D_{Coarse} by HFA-Net, and leverages its large receptive field architecture to enhance the continuity of transitions between regions(For further details, see [17]). Furthermore, to mitigate the mismatch between the inpainted region and the surrounding tissue—resulting from HFA-Net's complete boundary detection but insufficient contextual information, we apply a morphological dilation to the mask M to generate an expanded mask, $M_{Dilation}$. This expanded mask, together with D_{Coarse}, is then fed into LaMa. By increasing the receptive field, the model is guided to integrate additional structural and contextual features, thereby enhancing the textural consistency between the inpainted region and its surrounding tissue.

Loss Functions. To optimize specular highlight detection and removal in the first stage, we design corresponding loss functions to improve detection accuracy and removal effectiveness. For highlight detection, we employ Binary Cross-Entropy (BCE) loss to minimize the discrepancy between the predicted highlight mask $\hat{M}$ and the ground truth mask M_{gt}:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^{N} \left[M_{gt,i} \log(\hat{M}_i) + (1 - M_{gt,i}) \log(1 - \hat{M}_i) \right] \tag{2}$$

For highlight removal, we adopt L_2 loss (Mean Squared Error, MSE) to measure the pixel-wise difference between the restored image D and the ground truth highlight-free image I_{clean}:

$$\mathcal{L}_{L_2} = \frac{1}{N} \sum_{i=1}^{N} \left(D_i - I_{clean,i} \right)^2 \tag{3}$$
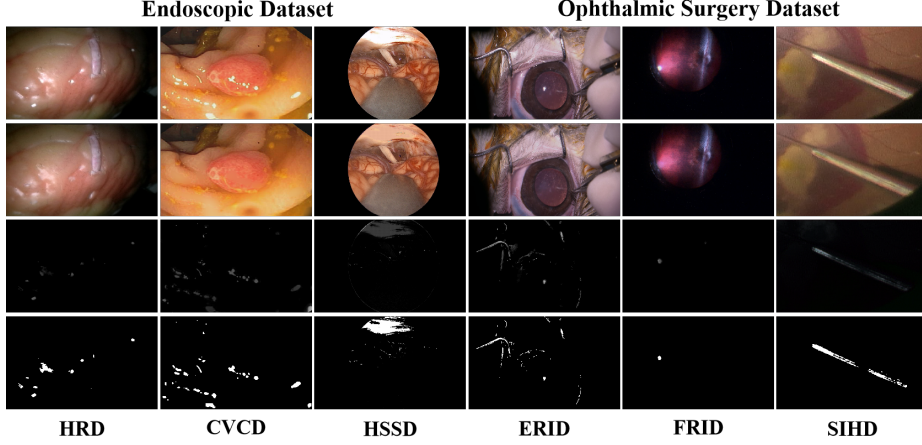
Fig. 2. Examples of dataset samples, covering endoscopic and ophthalmic surgical scenarios, including two public datasets and four custom-built datasets.

The total loss function integrates both specular highlight detection and removal losses to optimize overall performance:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{\mathrm{BCE}} + \lambda_2 \mathcal{L}_{L_2} \tag{4}$$

where λ_1 and λ_2 balance the contributions of specular highlight detection and removal. In our experiments, we empirically set $\lambda_1 = 1$ and $\lambda_2 = 10$.

3   Experiments

Dataset. As illustrated in Figure 2, to evaluate the generalizability of our method across diverse surgical scenarios, we constructed four task-specific datasets from endoscopic and ophthalmic surgical video streams following the approach in [7]: the External Retinal Injection Dataset (ERID), Fundus Retinal Injection Dataset (FRID), Surgical Instrument Highlight Dataset (SIHD), and Hemifacial Spasm Surgery Dataset (HSSD). Additionally, we incorporated two publicly available datasets: the CVC-Clinic Dataset (CVCD) [3] and the Hamlyn Rectified Dataset(HRD) [14]. All datasets were uniformly preprocessed and partitioned into training and testing subsets using a 75%/25% split, with no further fine-tuning or separate validation sets. Each task-specific dataset comprises over 10,000 samples, and each sample includes four image types: the specular image, the highlight-free image, the highlight intensity map, and the highlight mask.

Implementation details. The algorithm was implemented in PyTorch using an NVIDIA RTX 4070 GPU. The network was trained for 100 epochs with a batch size of 8 and an initial learning rate of 1e-5, using the ADAM optimizer [11]. Learning rate decay was applied periodically to prevent overfitting.
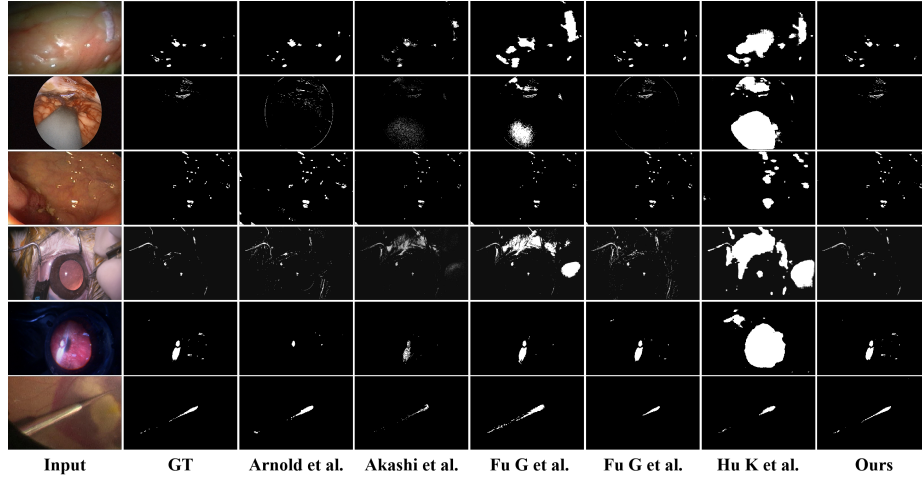
Fig. 3. Visual comparison of our method with different highlight detection methods on different datasets.

**Table 1.** Specular highlight detection performance on the FRID and HRD datasets.

Method	FRID			HRD		
	Precision	Recall	IoU	Precision	Recall	IoU
Arnold et al. [2]	0.1518	0.3335	0.1165	0.6633	0.7917	0.5647
Akashi et al. [1]	0.4922	0.8692	0.4583	0.4291	0.5201	0.3074
Fu G et al. [6]	0.8615	0.7288	0.6524	0.1484	0.6200	0.1360
Fu G et al. [7]	0.8809	0.9279	0.8245	0.7919	**0.9086**	0.7335
Hu K et al. [10]	0.2998	0.1627	0.1179	0.2836	0.3781	0.1934
Ours	**0.9324**	**0.9832**	**0.9178**	**0.9671**	0.8587	**0.8343**

Specular Highlight Detection Comparison and Evaluation. We conducted comparative experiments with the methods of Arnold et al. [2], Fu G et al. [6,7], Akashi et al. [1], and Hu K et al. [10]. The experiments were conducted on six different datasets. As shown in Figure 3, our method demonstrates superior performance in boundary misjudgment, white region confusion, and specular highlight integrity. Existing methods tend to misclassify specular highlight regions or lose some highlight information in complex scenarios. In contrast, our method ensures precise highlight region detection while effectively reducing false detections through more accurate feature extraction and attention mechanisms, making it applicable to various types of image data. As shown in Table 1 and Table 2, our method maintains high performance in terms of Precision, Recall, and Intersection over Union (IoU), as well as Accuracy (ACC) and Matthews Correlation Coefficient (MCC), demonstrating a stronger ability to distinguish specular highlight regions from non-specular areas.

Specular Highlight Removal Comparison and Evaluation. As shown in Figure 4, we compare our method with those of Shen and Zheng [16], Fu G et

Table 2. ACC and MCC of specular highlight detection across six datasets.

|	HRD		HSSD		CVCD		ERID		FRID		SIHD	
Method	Acc	MCC	Acc	MCC	Acc	MCC	Acc	MCC	Acc	MCC	Acc	MCC
Arnold et al. [2]	0.94	0.72	0.89	0.71	0.95	0.68	0.94	0.76	0.69	0.37	0.94	0.74
Akashi et al. [1]	0.94	0.46	0.38	0.10	0.94	0.77	0.55	0.14	0.49	0.65	0.80	0.54
Fu G et al. [6]	0.74	0.35	0.50	0.63	0.97	0.54	0.48	0.54	0.96	0.54	0.94	0.79
Fu G et al. [7]	0.95	0.84	0.90	0.80	0.97	0.91	0.80	0.31	0.95	0.90	0.70	**0.84**
Hu K et al. [10]	0.69	0.27	0.50	0.06	0.90	0.40	0.27	0.14	0.23	0.24	0.92	0.65
Ours	**0.96**	**0.88**	**0.97**	**0.82**	**0.98**	**0.97**	**0.95**	**0.84**	**0.98**	**0.98**	**0.97**	0.83
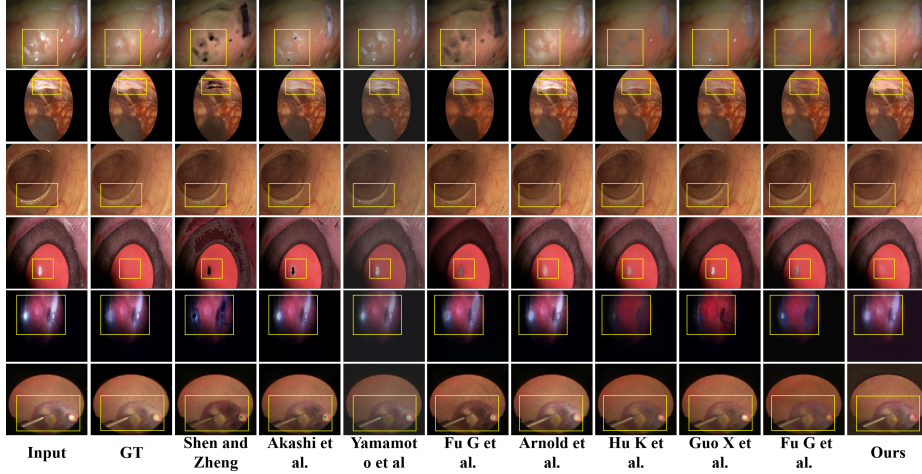
Fig. 4. Visual comparison of our framework against other methods.

al. [6,7], Akashi et al. [1], Yamamoto et al. [21] , Arnold et al. [2] , Guo X et al. [9] , and Hu K et al. [10] on six different datasets. Our method demonstrates superior performance in artifact suppression, background consistency, and specular highlight texture restoration, avoiding common issues in existing methods such as black artifacts, background inconsistency, and unnatural texture reconstruction. As shown in Table 3, our method achieves higher SSIM and PSNR scores than existing approaches, demonstrating enhanced specular highlight removal and visual quality restoration. Compared with traditional methods [1, 2, 6, 16, 21], our method effectively removes highlights while better preserving image details and structural information, avoiding excessive smoothing and color distortion. In comparison to deep learning-based methods [7, 9, 10], our model removes complex highlights while effectively reducing artifacts and maintaining structural integrity.

Ablation study. To verify the effectiveness of the proposed network architecture in the highlight removal task, we designed five baseline ablation experiments. As shown in Figure 5, Net5 achieves the most realistic visual performance in highlight removal. **Net1** adopts a basic encoder-decoder architecture,

Table 3. Quantitative results of our framework against other methods.

	HSSD		ERID		FRID		SIHD		HRD		CVCD	
Method	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Shen and Zheng [16]	24.25	0.69	30.76	0.66	28.48	0.32	24.75	0.65	24.76	0.69	32.96	0.74
Akashi et al. [1]	26.47	0.71	33.16	0.82	28.62	0.33	30.26	0.62	23.67	0.78	33.28	0.72
Yamamoto et al. [21]	28.13	0.47	29.05	0.80	31.41	0.84	25.39	0.84	26.23	0.72	29.27	0.63
Fu G et al. [6]	31.26	0.82	28.83	0.73	28.86	0.40	25.03	0.62	29.06	0.65	31.95	0.83
Arnold et al. [2]	30.12	0.86	33.14	0.84	29.01	0.38	27.93	0.67	32.28	0.86	32.26	0.88
Hu K et al. [10]	32.43	0.89	32.24	0.82	27.99	0.34	28.51	0.72	28.30	0.84	31.83	0.82
Guo X et al. [9]	29.43	0.84	26.43	0.77	23.94	0.78	23.58	0.62	28.78	0.83	24.43	0.73
Fu G et al. [7]	30.73	0.44	32.37	0.84	26.47	0.64	28.09	0.74	28.84	0.86	34.02	0.85
Ours	**35.78**	**0.91**	**38.85**	**0.90**	**34.26**	**0.88**	**31.55**	**0.88**	**33.42**	**0.89**	**37.13**	**0.94**

Input   Net1   Net2   Net3   Net4   Net5
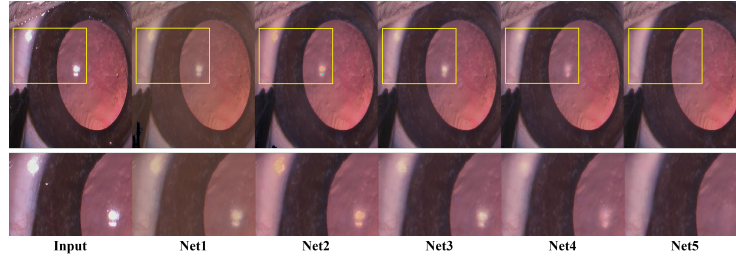
Fig. 5. Ablation study for our method. The second row presents a magnified view of the region highlighted by the yellow box.

extracting M and D_{Coarse} through convolution and PMconv. However, due to the absence of advanced attention mechanisms, its capability in detail recovery is limited. **Net2** introduces S^{2}MLP and DFA to enhance feature extraction. **Net3** further incorporates SFF for multi-scale feature fusion, enhancing boundary region analysis and making highlight removal more natural. **Net4** integrates the LaMa network to refine M and D_{Coarse}, improving visual consistency, though the enhancement over Net3 is relatively limited. **Net5** further expands M, ultimately achieving the highest PSNR and SSIM, ensuring optimal highlight removal and texture restoration, as shown in Table 4.

Table 4. The PSNR and SSIM results of our ablation study.

Metric	Net1	Net2	Net3	Net4	Net5
PSNR	27.80	35.20	38.75	39.20	**43.20**
SSIM	0.71	0.93	0.94	0.94	**0.98**

4   Conclusion

This paper proposes a hierarchical feature attention-based two-stage method for specular highlight detection and removal, effectively improving detection accu-

racy and restoration naturalness. To address the scarcity of medical highlight data, we construct four medical highlight datasets and incorporate publicly available datasets to enhance model generalization. We then introduce HFA-Net, integrating S²MLP, DFA, and SFF, along with PMConv for highlight detection and preliminary removal. Under the two-stage restoration strategy, the preliminary restored image D_{Coarse} is further refined using LaMa, with mask expansion to enhance contextual awareness and texture consistency. Overall, our results demonstrate that the proposed method surpasses existing approaches in both visual quality and objective metrics. Future work will focus on designing lightweight models for real-time deployment and exploring domain adaptation techniques to enhance generalization across diverse medical imaging conditions.

Disclosure of Interests. The authors have no competing interests

References

1. Akashi, Y., Okatani, T.: Separation of reflection components by sparse non-negative matrix factorization. In: Computer Vision–ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part V 12. pp. 611–625. Springer (2015)
2. Arnold, M., Ghosh, A., Ameling, S., Lacey, G.: Automatic segmentation and inpainting of specular highlights for endoscopic imaging. EURASIP Journal on Image and Video Processing **2010**, 1–12 (2010)
3. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized medical imaging and graphics **43**, 99–111 (2015)
4. Chen, J., Kao, S.h., He, H., Zhuo, W., Wen, S., Lee, C.H., Chan, S.H.G.: Run, don't walk: chasing higher flops for faster neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12021–12031 (2023)
5. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1251–1258 (2017)
6. Fu, G., Zhang, Q., Song, C., Lin, Q., Xiao, C.: Specular highlight removal for real-world images. In: Computer graphics forum. vol. 38, pp. 253–263. Wiley Online Library (2019)
7. Fu, G., Zhang, Q., Zhu, L., Li, P., Xiao, C.: A multi-task network for joint specular highlight detection and removal. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7752–7761 (2021)
8. Fu, G., Zhang, Q., Zhu, L., Lin, Q., Wang, Y., Fan, S., Xiao, C.: Towards high-resolution specular highlight detection. International Journal of Computer Vision **132**(1), 95–117 (2024)

9. Guo, X., Chen, X., Luo, S., Wang, S., Pun, C.M.: Dual-hybrid attention network for specular highlight removal. In: Proceedings of the 32nd ACM International Conference on Multimedia. pp. 10173–10181 (2024)

10. Hu, K., Huang, Z., Wang, X.: Highlight removal network based on an improved dichromatic reflection model. In: ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 2645–2649. IEEE (2024)

11. Kinga, D., Adam, J.B., et al.: A method for stochastic optimization. In: International conference on learning representations (ICLR). vol. 5, p. 6. San Diego, California; (2015)

12. Nie, C., Xu, C., Li, Z., Chu, L., Hu, Y.: Specular reflections detection and removal for endoscopic images based on brightness classification. Sensors **23**(2), 974 (2023)

13. Pan, J., Li, R., Liu, H., Hu, Y., Zheng, W., Yan, B., Yang, Y., Xiao, Y.: Highlight removal for endoscopic images based on accelerated adaptive non-convex rpca decomposition. Computer Methods and Programs in Biomedicine **228**, 107240 (2023)

14. Recasens, D., Lamarca, J., Fácil, J.M., Montiel, J., Civera, J.: Endo-depth-and-motion: Reconstruction and tracking in endoscopic videos using depth networks and photometric constraints. IEEE Robotics and Automation Letters **6**(4), 7225–7232 (2021)

15. Shafer, S.A.: Using color to separate reflection components. Color Research & Application **10**(4), 210–218 (1985)

16. Shen, H.L., Zheng, Z.H.: Real-time highlight removal using intensity ratio. Applied optics **52**(19), 4483–4493 (2013)

17. Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., Kong, N., Goka, H., Park, K., Lempitsky, V.: Resolution-robust large mask inpainting with fourier convolutions. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 2149–2159 (2022)

18. Wu, Z., Guo, J., Zhuang, C., Xiao, J., Yan, D.M., Zhang, X.: Joint specular highlight detection and removal in single images via unet-transformer. Computational Visual Media **9**(1), 141–154 (2023)

19. Xia, W., Chen, E.C., Pautler, S.E., Peters, T.M.: A global optimization method for specular highlight removal from a single image. IEEE Access **7**, 125976–125990 (2019)

20. Xiang, S., Wei, L., Tang, S.: Fusion of unet and criminisi algorithms for endoscopic specular highlight removal. In: 2023 38th Youth Academic Annual Conference of Chinese Association of Automation (YAC). pp. 1093–1097. IEEE (2023)

21. Yamamoto, T., Nakazawa, A.: General improvement method of specular component separation using high-emphasis filter and similarity function. ITE Transactions on Media Technology and Applications **7**(2), 92–102 (2019)

22. Yu, B., Chen, W., Zhong, Q., Zhang, H.: Specular highlight detection based on color distribution for endoscopic images. Frontiers in Physics **8**, 616930 (2021)

23. Yu, T., Li, X., Cai, Y., Sun, M., Li, P.: S²-mlpv2: improved spatial-shift mlp architecture for vision. arXiv preprint arXiv:2108.01072 (2021)

24. Zhang, C., Liu, Y., Wang, K., Tian, J.: Specular highlight removal for endoscopic images using partial attention network. Physics in Medicine & Biology **68**(22), 225009 (2023)