LoHi-Net: A Low-to-High Hierarchical Wavelet Network for Laparoscopic Surgical Smoke Removal

Shiwei Wu[1], Xiyuan Hu[2](✉), Xiaobo Zhu[3], Dongyang Gao[1], Song Zhang[3], and Hui Yan[1]

[1] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[2] School of Computer Science, Beijing University of Technology, Beijing 100124, China
huxiyuan@bjut.edu.cn
[3] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract. In minimally invasive surgery, surgical smoke generated by electrocautery devices degrades the laparoscopic view, obstructing the surgeon's visual field and increasing operative risks. Existing smoke removal methods often overlook the inherent spectral discrepancies between global smoke artifacts and local scene details. Consequently, these approaches struggle to strike a balance between removing heavy smoke and preserving fine surgical textures. Based on wavelet decomposition analysis, we observe a critical phenomenon: replacing the low-frequency component of a smoke-degraded image with that of a clear reference effectively eliminates visual smoke, while residual errors are predominantly concentrated in high-frequency components. This indicates that surgical smoke degradation is structurally decoupled: global smoke dominates low frequencies, while texture degradation resides in high frequencies. Motivated by this observation, we propose LoHi-Net, a Low-to-High Hierarchical Wavelet Network. Unlike conventional holistic or independent processing schemes, LoHi-Net adopts a hierarchical guided restoration mechanism. Specifically, it first restores low-frequency structural components to eliminate global smoke effects, which then serve as a structural prior to guide the precise refinement of high-frequency textures. To maintain model compactness while capturing the distinct characteristics of each frequency band,we introduce specialized Mamba variants into the architecture for hierarchical modeling. Extensive experiments on real-world datasets demonstrate that LoHi-Net achieves state-of-the-art performance, effectively recovering high-fidelity surgical scenes with compact model capacity.

Keywords: Laparoscopic surgery · Smoke removal · Wavelet transform · Mamba.

1 Introduction

During laparoscopic surgery, energy devices such as electrocautery or ultrasonic scalpels are routinely employed for tissue dissection. Tissue cauterization during this process produces surgical smoke, which degrades visual clarity and may compromise surgical safety [13]. Computer vision and deep learning-based surgical smoke removal solutions have emerged as cost-effective, low-risk, and efficient alternatives to hardware-based smoke evacuators [15].

Traditional smoke removal methods often rely on the atmospheric scattering model, estimating physical parameters via hand-crafted priors (e.g., Dark Channel Prior [8]) to recover clear images. In contrast, deep learning-based methods learn smoke removal patterns directly from data, bypassing the need for hand-crafted priors. CycleGAN-based [14, 26] methods approach smoke removal as an unpaired image-to-image translation task, eliminating the need for paired training data. Conversely, De-smokeGCN [1] and MARS-GAN [9] employ synthesized smoke to construct paired datasets, enabling the end-to-end training of CNN-based de-smoking models. Wang et al. [21] propose a Swin Transformer-based U-shaped architecture, which enhances feature representation by combining local detail extraction with global context modeling. Furthermore, SVP-Net [20] and SurgiATM [15] incorporate physical models to bolster the robustness of smoke removal networks against complex degradation. Recently, Li et al. [11] introduced Diffusion Models to this task, proposing a framework based on multilevel frequency learning and multilevel smoke attention-aware learning. EndoIR [2] jointly extracts spatial and frequency features to generate information recovery prompts.

Beyond the medical imaging domain, several studies offer complementary insights. SGDN [6] leverages the structural properties of YCbCr color space to guide RGB dehazing, while ConvIR [4] demonstrates that optimized convolutional networks can rival Transformers across various restoration tasks. In addition, frequency-aware restoration methods have shown promising performance by exploiting spectral representations through frequency modulation, prompting mechanisms, or feature mixing [5, 25, 10]. However, these approaches primarily use frequency information as auxiliary guidance and generally treat different frequency components in a unified manner, without explicitly considering the different restoration characteristics of low- and high-frequency components. As a result, they fail to explicitly decouple global smoke from local details, leading to a trade-off where removing heavy smoke often hinders the precise recovery of high-frequency surgical details.

To further investigate the frequency characteristics of surgical smoke degradation, we perform a simple yet illustrative analysis based on wavelet decomposition as shown in Fig. 1. Specifically, smoky and smokeless images are decomposed into low- and high-frequency components using wavelet transform. By exchanging the low-frequency components (LL) between the two images and reconstructing them via inverse wavelet transform, the originally smoke-degraded image becomes visually clean. However, the error map between the reconstructed image and the original smokeless image reveals a noticeable loss of high-frequency

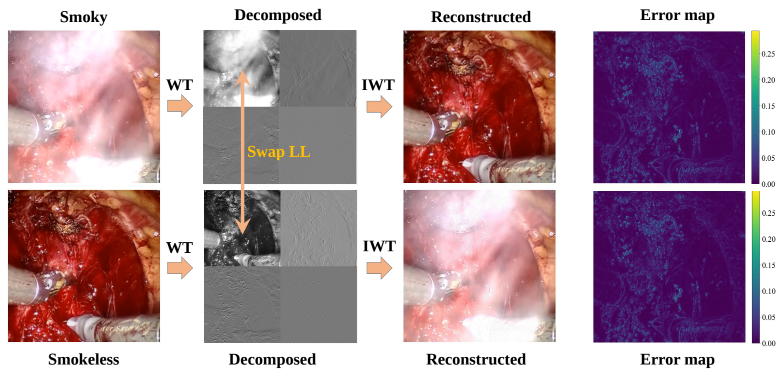

Fig. 1. Wavelet-based frequency-domain analysis reveals that surgical smoke exhibits dominant low-frequency global degradation and leads to the suppression of high-frequency structural details. WT and IWT denote the Discrete Wavelet Transform and the Inverse Wavelet Transform, respectively.

details. This observation indicates that surgical smoke primarily induces a global degradation concentrated in low-frequency components, while the degradation of fine textures and structural details manifests in high-frequency regions. Inspired by this observation, we propose LoHi-Net, a Low-to-High Hierarchical Wavelet Network surgical smoke removal. The main contributions of this work are summarized as follows: 1) We propose a hierarchical guided restoration block that prioritizes low-frequency optimization to suppress global smoke interference, and subsequently guides high-frequency reconstruction to preserve fine surgical details. 2) We design a Low-to-High (L2H) structural guidance block, which extracts reliable structural priors from relatively clean low-frequency features to facilitate accurate and stable high-frequency detail recovery. 3) We build upon the Visual Mamba architecture [16, 12] and extend it with a hierarchical design to explicitly model frequency-specific representations, enabling lightweight and effective surgical smoke removal with competitive performance.

2 Methods

The proposed Low-to-High Hierarchical Wavelet Network adopts a U-Net–style encoder–decoder architecture, as shown in Fig. 2. Wavelet transform and inverse wavelet transform (WT/IWT) are used as core operators to jointly achieve frequency decomposition and scale-aware downsampling/upsampling, enabling unified modeling of multi-scale and multi-frequency information. In shallow layers, Def-Mamba [12] is introduced to effectively model long-range spatial dependencies. In deeper layers, the Hierarchical Guided Restoration (HGR) Block is employed to perform guided restoration of wavelet sub-bands in the frequency domain, facilitating structural reconstruction. Furthermore, multi-level wavelet-decomposed inputs and decoder-side multi-level supervision are incorporated to

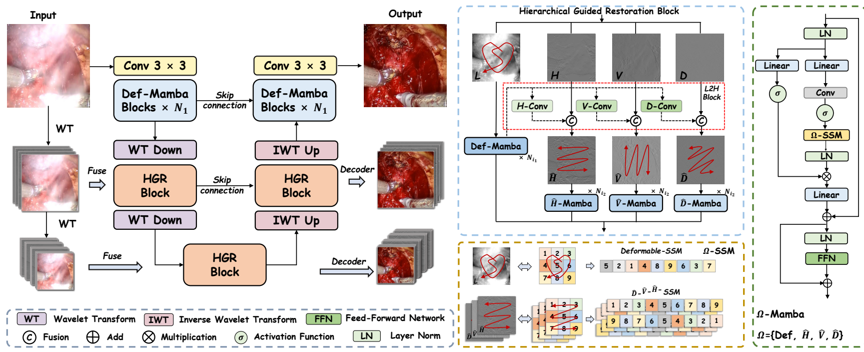

Fig. 2. The architecture of the proposed LoHi-Net.

preserve information across scales and improve the quality of the reconstructed smokeless image.

2.1 Hierarchical Guided Restoration Block

Given a smoke-degraded image $I_s \in \mathbb{R}^{H\times W\times C}$, we first apply the wavelet transform to decompose it into four frequency components:

$$\{L, H, V, D\} = WT(I_s) \in \mathbb{R}^{\frac{H}{2}\times\frac{W}{2}\times C}, \tag{1}$$

where L denotes the low-frequency component, and H, V and D represent the horizontal, vertical, and diagonal high-frequency components, respectively. To reconstruct a clean image $I_c \in \mathbb{R}^{H\times W\times C}$, the corresponding clean wavelet components $\{L', H', V', D'\}$ need to be predicted and then recombined via the inverse wavelet transform, i.e.,

$$I_c = IWT([L', H', V', D']). \tag{2}$$

Based on the previous frequency analysis, surgical smoke affects low- and high-frequency components differently, leading to distinct degradation patterns across sub-bands. Uniform treatment ignores these heterogeneous properties, while independent processing lacks cross-band interaction. Therefore, to obtain high-quality restored components $\{L', H', V', D'\}$, the proposed Hierarchical Guided Restoration (HGR) Block combines a low-to-high structural guidance mechanism with sub-band-specific Mamba variants to achieve frequency-aware hierarchical restoration.

2.2 Low-to-High Structural Guidance

Since the low-frequency component dominates the image energy and preserves the primary spatial structure, we leverage the restored clean low-frequency component L' as a structural prior to guide the enhancement of high-frequency sub-bands $\{H, V, D\}$.

Specifically, to achieve precise structural alignment, we design three types of convolutional operators to extract directional structural features from L'. This structure-guided process is formulated as:

$$\widehat{W} = \mathcal{F}(W\text{-}Conv(L'), W), \quad W \in \{H, V, D\}, \quad \widehat{W} \in \{\widehat{H}, \widehat{V}, \widehat{D}\}, \tag{3}$$

where $\mathcal{F}(\cdot, \cdot)$ denotes the fusion operation, and $\widehat{W}$ represents the enhanced high-frequency component. W-$Conv$ denotes the convolution operation parameterized by the kernel associated with sub-band W. The kernel (e.g., $\mathbf{K}_H$, $\mathbf{K}_V$, and $\mathbf{K}_D$) is designed to align with the directional characteristics of the wavelet decomposition, serving as structural priors. The corresponding convolution kernels are defined as follows:

$$\mathbf{K}_H = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}, \mathbf{K}_V = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}, \mathbf{K}_D = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{4}$$

$\mathbf{K}_H$ and $\mathbf{K}_V$ are first-order directional differences along two spatial axes. $\mathbf{K}_D$ is a Laplacian-like second-order residual operator capturing mixed-axis high-frequency variations, providing complementary structural cues for the diagonal sub-band D. This Low-to-High (L2H) structural guidance mechanism enables cross-band information interaction. Benefiting from the pre-restored and relatively clean guidance source L', the mechanism introduces structural cues into high-frequency components without significantly amplifying noise artifacts.

2.3 Mamba-based Frequency Modeling

State Space Models (SSMs) have emerged as compelling alternatives to CNNs and Transformers for modeling long-range dependencies, characterized by their linear scalability with respect to sequence length [7]. Recent advancements in this domain have demonstrated significant improvements in image restoration tasks [16, 19]. The discretized SSM can be formulated as a linear recurrence system:

$$h_t = \bar{\mathbf{A}} h_{t-1} + \bar{\mathbf{B}} x_t, \quad y_t = \mathbf{C} h_t + \mathbf{D} x_t, \tag{5}$$

where x_t denotes the serialized 1-D input sequence, y_t represents the output, and $\bar{\mathbf{A}}, \bar{\mathbf{B}}, \mathbf{C}, \mathbf{D}$ are the learnable system matrices.

Mamba stands out as a structured SSM designed for efficient hardware-aware processing. Inspired by its success, we introduce distinct Mamba variants tailored to the heterogeneous properties of different frequency components to achieve frequency-aware modeling.

Given the dynamic and complex nature of smoke degradation in the low-frequency component, we employ Def-Mamba [12] to model the low-frequency sub-band L. Unlike standard fixed scanning paths, Def-Mamba utilizes a learnable deformable scanning mechanism to serialize the input features. This enables the model to dynamically capture intrinsic spatial dependencies, thereby effectively identifying and eliminating smoke interference. To further enhance the processing of high-frequency components, we design direction-customized scanning mechanisms specifically tailored to the orientation of each sub-band. This results in three specialized variants: $\widehat{H}$-Mamba, $\widehat{V}$-Mamba, and $\widehat{D}$-Mamba, which align the sequence modeling with the inherent directional characteristics of the wavelet coefficients.

Consequently, the restoration process of the wavelet coefficients can be formulated as:

$$W' = \Omega\text{-}Mamba(Y), \quad Y \in \{L, \widehat{H}, \widehat{V}, \widehat{D}\}, \tag{6}$$

where $W' \in \{L', H', V', D'\}$ denotes the restored components, and Ω represents the specific scanning configuration for input Y. As illustrated in Fig. 2, the Ω-Mamba block comprises Layer Normalization (LN), followed by a Visual State Space Module (VSSM) with residual connections, and incorporates a Feed-Forward Network (FFN) to enhance information flow. Specifically, to accommodate the distinct properties of each sub-band, the Ω-SSM variants in VSSM are configured with deformable scanning for L, and horizontal, vertical, and diagonal scanning for $\widehat{H}, \widehat{V}$, and $\widehat{D}$, respectively.

2.4 Loss Function

To ensure faithful reconstruction while preserving fine-grained details, we optimize the network using a composite loss that jointly constrains predictions in the spatial (wavelet) domain and the global frequency domain across multiple scales. The total loss $\mathcal{L}$ is defined as:

$$\mathcal{L} = \sum_{i=0}^{3} \left(\|\hat{I}_c^i - I_c^i\|_1 + \lambda \|FFT(\hat{I}_c^i) - FFT(I_c^i)\|_1 \right), \tag{7}$$

where $FFT(\cdot)$ represents the Fast Fourier Transform operation. I_c^0 and $\hat{I}_c^0$ refer to the final restored image and the ground truth, respectively, while I_c^i and $\hat{I}_c^i$ for $i > 0$ represent the ground-truth and predicted wavelet coefficients obtained from the i-level wavelet decomposition. The first term minimizes pixel/coefficient-wise reconstruction errors, while the second term ensures global spectral fidelity. We empirically set $\lambda = 0.1$ in this work.

3 Experiments

3.1 Datasets and Experimental Setting

To comprehensively evaluate the effectiveness of LoHi-Net, we conduct training and evaluation on two real-world laparoscopic smoke removal datasets. The first dataset is VASST-desmoke [24, 23], a recently released large-scale dataset containing 3,464 paired smoke/smoke-free images from 21 surgical scenes. We randomly split it at the sequence/scene level, with 70% for training and 30% for testing, to avoid temporal leakage. The second is LSVD [22], a real surgical video desmoking dataset. Its training set consists of 416 video clips, and the test set

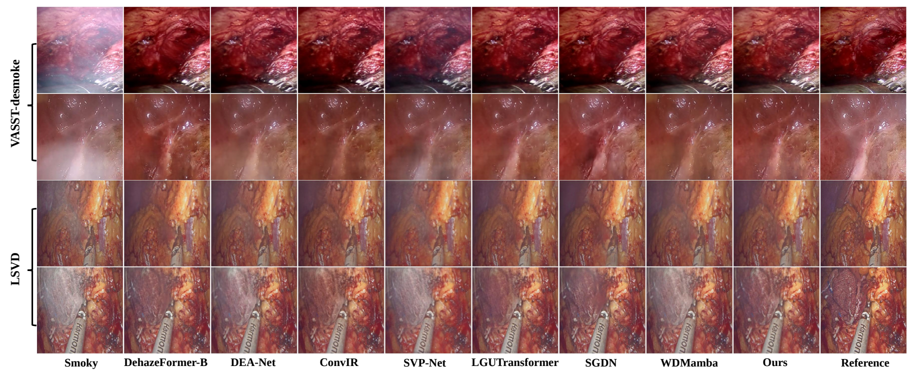

Fig. 3. Visual comparison results on the VASST-desmoke and LSVD datasets.

contains 70 clips, with each clip ranging from 20 to 50 frames in length. Following the training and evaluation protocol described in [22], we utilize a pre-trained PWC-Net [18] to register the pre-smoke frame (the relatively clean first frame of a video clip) with the network-predicted frames, using the registered images as the smoke-free ground truth.

For comparative analysis, we select seven state-of-the-art algorithms, including two specialized desmoking algorithms (SVP-Net [20] and LGUTransformer [21]), four dehazing methods (DehazeFormer-B [17], DEA-Net [3], SGDN [6], and WDMamba [19]), and one general image restoration method (ConvIR-S [4]). We include these dehazing approaches as baselines because they incorporate advanced architectural designs that are often transferable to the surgical smoke removal task. The evaluation metrics employed are three commonly used full-reference image quality indicators: SSIM, PSNR, and CIEDE-2000.

To ensure a fair comparison, all baseline models are retrained on the aforementioned two dataset. All models are optimized using the AdamW optimizer with a batch size of 4. The learning rate follows a cosine annealing strategy, decaying from an initial 1×10^{-4} to 1×10^{-6}. Regarding the specific datasets, we train the models for 350 epochs on VASST-desmoke and 200 epochs on LSVD. All experiments are conducted on a single NVIDIA GeForce RTX 4090 GPU.

3.2 Comparison on VASST-desmoke and LSVD Datasets

As shown in Table. 1, our method achieves the best overall performance on both datasets, attaining the highest SSIM and PSNR scores while yielding the lowest CIEDE values. These results indicate superior structural restoration and color fidelity compared with all competing methods. Owing to the low-to-high hierarchical guided restoration strategy, the proposed model maintains high restoration quality with low computational complexity, requiring only 1.38M parameters and 14.07G MACs.

Qualitative results in Fig. 3 further validate these findings. Our method effectively removes smoke while preserving fine tissue textures without over-

Table 1. Quantitative comparison of various methods on the VASST-desmoke and LSVD datasets. ↓ denotes that lower values are better, while ↑ indicates that higher values are better. Underlined values highlight second-best results, and bold values indicate the best performance.

Methods	**VASST-desmoke**			**LSVD**			**Overhead**	
	SSIM↑	PSNR↑	CIEDE↓	SSIM↑	PSNR↑	CIEDE↓	Params↓	MACs↓
DehazeFormer-B	0.886	26.599	4.195	0.796	26.340	4.330	2.52M	19.76G
DEA-Net	0.880	26.472	4.197	0.792	26.022	4.472	3.65M	24.68G
ConvIR-S	0.874	25.690	4.627	0.789	26.112	4.512	5.53M	42.23G
SVP-Net	0.854	25.051	5.027	0.755	25.186	4.919	17.46M	152.67G
LGUTransformer	0.880	26.329	4.347	0.792	26.305	4.285	5.54M	32.45G
SGDN	0.886	26.519	4.075	0.775	25.727	4.535	11.09M	41.16G
WDMamba	0.882	26.484	4.193	0.787	25.627	4.688	11.25M	29.74G
Ours	**0.889**	**27.071**	**3.872**	**0.798**	**26.505**	**4.164**	**1.38M**	**14.07G**

Table 2. Ablation experiment study of LoHi-Net.

Methods	**SSIM↑**	**PSNR↑**	**CIEDE↓**	**Params**	**MACs**
Non-Hierarchical	0.887	26.423	4.270	2.40M	18.82G
Same SSM	0.885	26.357	4.217	1.37M	16.66G
Same Depth	0.886	26.439	4.125	1.17M	13.21G
w/o L2H	0.882	26.148	4.426	1.28M	13.38G
Ours	**0.889**	**27.071**	**3.872**	**1.38M**	**14.07G**

smoothing or color distortion. In contrast, some methods, such as SVP-Net and WDmamba, tend to leave residual smoke or introduce artifacts, particularly in regions with dense smoke or complex structures. Overall, these results demonstrate the effectiveness and robustness of the proposed approach on the VASST-desmoke and LSVD datasets.

3.3 Ablation Studies

In Table. 2, we present the ablation study conducted to verify the effectiveness of the core components within LoHi-Net. We evaluated the following variants: 1) Non-Hierarchical: A variant that processes all wavelet components using DefMamba blocks directly, without the hierarchical processing strategy; 2) Same SSM: A variant where the same scanning mechanism is applied to all components; 3) Same Depth: Since our original design assigns a deeper network to low-frequency components to ensure thorough removal of interference, we compared it against a variant where low-frequency and high-frequency components share the same module depth; 4) w/o L2H: A variant without the Low-to-High Structural guidance mechanism for high-frequency components. The ablation results demonstrate the significance of each contribution to the overall performance of the model.

4 Conclusion

In this paper, we propose LoHi-Net, a novel laparoscopic smoke removal network based on wavelet transform and the Low-to-High hierarchical guided restoration strategy. Experiments on the VASST-desmoke and LSVD datasets demonstrate that LoHi-Net outperforms state-of-the-art methods in both quantitative metrics and visual quality, effectively removing smoke interference while accurately preserving tissue details. Furthermore, with only 1.38M parameters, the model achieves a balance between performance and computational cost, demonstrating significant potential for clinical application.

Acknowledgments. This work was supported by Beijing Natural Science Foundation (L247034).

Disclosure of Interests. All authors disclosed no relevant relationships.

References

1. Chen, L., Tang, W., John, N.W., Wan, T.R., Zhang, J.J.: De-smokegcn: generative cooperative networks for joint surgical smoke detection and removal. IEEE transactions on medical imaging **39**(5), 1615–1625 (2019)
2. Chen, T., Ma, X., Bai, L., Wang, W., Sun, Y., Zhou, L.: Endoir: Degradation-agnostic all-in-one endoscopic image restoration via noise-aware routing diffusion. arXiv preprint arXiv:2511.05873 (2025)
3. Chen, Z., He, Z., Lu, Z.M.: Dea-net: Single image dehazing based on detail-enhanced convolution and content-guided attention. IEEE Transactions on Image Processing **33**, 1002–1015 (2024)
4. Cui, Y., Ren, W., Cao, X., Knoll, A.: Revitalizing convolutional network for image restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(12), 9423–9438 (2024)
5. Cui, Y., Zamir, S.W., Khan, S., Knoll, A., Shah, M., Khan, F.: Adair: Adaptive all-in-one image restoration via frequency mining and modulation. In: International Conference on Learning Representations. vol. 2025, pp. 101306–101327 (2025)
6. Fang, W., Fan, J., Zheng, Y., Weng, J., Tai, Y., Li, J.: Guided real image dehazing using ycbcr color space. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 2906–2914 (2025)
7. Gu, A., Goel, K., Ré, C.: Efficiently modeling long sequences with structured state spaces. arXiv preprint arXiv:2111.00396 (2021)
8. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE transactions on pattern analysis and machine intelligence **33**(12), 2341–2353 (2010)
9. Hong, T., Huang, P., Zhai, X., Gu, C., Tian, B., Jin, B., Li, D.: Mars-gan: multilevel-feature-learning attention-aware based generative adversarial network for removing surgical smoke. IEEE Transactions on Medical Imaging **42**(8), 2299–2312 (2023)
10. Jiang, X., Zhang, X., Gao, N., Deng, Y.: When fast fourier transform meets transformer for image restoration. In: European conference on computer vision. pp. 381–402. Springer (2024)

11. Li, H., Zhai, X., Xue, J., Gu, C., Tian, B., Hong, T., Jin, B., Li, D., Huang, P.: Multi-frequency and smoke attention-aware learning based diffusion model for removing surgical smoke. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 47–56. Springer (2024)
12. Liu, L., Zhang, M., Yin, J., Liu, T., Ji, W., Piao, Y., Lu, H.: Defmamba: Deformable visual state space model. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 8838–8847 (2025)
13. Luo, X., McLeod, A.J., Pautler, S.E., Schlachta, C.M., Peters, T.M.: Vision-based surgical field defogging. IEEE transactions on medical imaging **36**(10), 2021–2030 (2017)
14. Pan, Y., Bano, S., Vasconcelos, F., Park, H., Jeong, T.T., Stoyanov, D.: Desmoke-lap: improved unpaired image-to-image translation for desmoking in laparoscopic surgery. International Journal of Computer Assisted Radiology and Surgery **17**(5), 885–893 (2022)
15. Sheng, M., Fan, J., Liu, D., Zheng, G., Kikinis, R., Cai, W.: Surgiatm: A physics-guided plug-and-play model for deep learning-based smoke removal in laparoscopic surgery. arXiv preprint arXiv:2511.05059 (2025)
16. Shi, Y., Xia, B., Jin, X., Wang, X., Zhao, T., Xia, X., Xiao, X., Yang, W.: Vmambair: Visual state space model for image restoration. IEEE Transactions on Circuits and Systems for Video Technology (2025)
17. Song, Y., He, Z., Qian, H., Du, X.: Vision transformers for single image dehazing. IEEE Transactions on Image Processing **32**, 1927–1941 (2023)
18. Sun, D., Yang, X., Liu, M.Y., Kautz, J.: Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8934–8943 (2018)
19. Sun, J., Liu, H., Wang, Y., Zhang, X.P., Wei, M.: Wdmamba: When wavelet degradation prior meets vision mamba for image dehazing. IEEE Transactions on Circuits and Systems for Video Technology pp. 1–1 (2025). https://doi.org/10.1109/TCSVT.2025.3614173
20. Wang, C., Zhao, M., Zhou, C., Dong, N., Khan, Z.A., Zhao, X., Cheikh, F.A., Beghdadi, A., Chen, S.: Smoke veil prior regularized surgical field desmoking without paired in-vivo data. Computers in Biology and Medicine **168**, 107761 (2024)
21. Wang, W., Liu, F., Hao, J., Yu, X., Zhang, B., Shi, C.: Desmoking of the endoscopic surgery images based on a local-global u-shaped transformer model. IEEE Transactions on Medical Robotics and Bionics (2024)
22. Wu, R., Zhang, Z., Zhang, S., Gou, L., Chen, H., Zhang, L., Chen, H., Zuo, W.: Self-supervised video desmoking for laparoscopic surgery. In: European Conference on Computer Vision. pp. 307–324. Springer (2024)
23. Xia, W., Fan, V., Peters, T., Chen, E.C.S.: A New Benchmark In Vivo Paired Dataset for Laparoscopic Image De-smoking . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. vol. LNCS 15001. Springer Nature Switzerland (October 2024)
24. Xia, W., Peters, T.M., Fan, V., Sthanunathan, H., Qi, O., Chen, E.C.S.: In vivo laparoscopic image de-smoking dataset, evaluation, and beyond. IEEE Transactions on Medical Imaging **44**(12), 4841–4853 (2025). https://doi.org/10.1109/TMI.2025.3584641
25. Zhou, S., Pan, J., Shi, J., Chen, D., Qu, L., Yang, J.: Seeing the unseen: A frequency prompt guided transformer for image restoration. In: European Conference on Computer Vision. pp. 246–264. Springer (2024)

26. Zhou, Y., Hu, Z., Xuan, Z., Wang, Y., Hu, X.: Synchronizing detection and removal of smoke in endoscopic images with cyclic consistency adversarial nets. IEEE/ACM Transactions on Computational Biology and Bioinformatics **21**(4), 670–680 (2022)