Progressive Frequency-Aware Network for Laparoscopic Image Desmoking

XIAMEN • 2023

Code Available

Jiale Zhang*, Wenfeng Huang*, Xiangyun Liao, and Qiong Wang
* Indicates equal contribution
Guangdong Provincial Key Laboratory of Computer Vision and Virtual Reality Technology,
Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences

INTRODUCTION

Laparoscopy

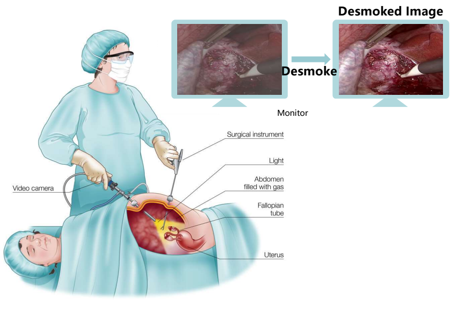

Fig. 1. Semantic of operative field affected by the smoke in laparoscopic procedure

- Smoke is generated by laser ablation and cauterization.
- Smoke fills abdomen, which presences challenges visibility and safety.

Difficulties

Traditional Theory-Based Method

- Low efficacy

Deep Learning-Based Method

- Necessity for extensive training data
- Having large parameter counts that unsuitable for medical devices

OUR CONTRIBUTIONS

- Progressive Frequency-Aware Net (PFAN) is proposed by focusing on the image frequency domain, **integrating high and low-frequency features efficiently**.
- The model achieves **a favorable performance-to-complexity balance**.

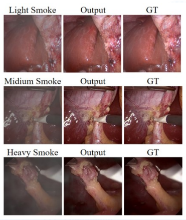

Fig. 2. Desmoking Result

METHODOLOGY Ⅰ

Framework

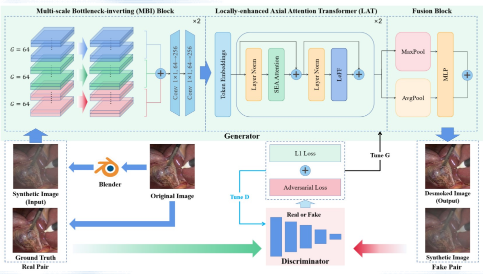

Fig. 3. The flowchart of PFAN illustrates a framework consisting of a generator network (G) and a discriminator network (D). Within this proposed approach, the generator G incorporates **Multi-scale Bottleneck-Inverting (MBI) Blocks** and **Locally-Enhanced Axial Attention Transformer (LAT) Blocks**.

- A CNN-ViT-based approach within GAN architecture
- Information **extraction progressively in the frequency domain by MBI Blocks and LAT Blocks**
- A graphics rendering engine integrated into our learning framework to generate paired training data without manual labeling

MBI Block

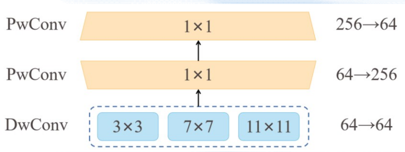

Fig. 4. The schematic illustration of the proposed **Multi-scale Bottleneck-Inverting (MBI) Block**

The MBI Block is designed to efficiently extract high-frequency features, drawing inspiration from Inception[1], ConvNext[2] and so on.

METHODOLOGY Ⅱ

LAT Block

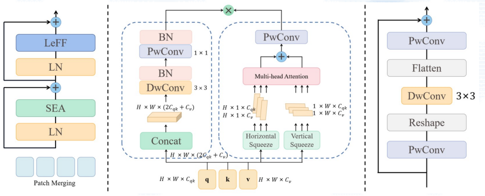

Fig. 5. Left: the schematic illustration of the proposed Locally-Enhanced Axial Attention Transformer (LAT) Block. Middle: Squeeze-Enhanced Axial Attention Layer. Right: Locally-Enhanced Feed-Forward network.

The LAT Block captures long-range dependencies and global low-frequency information with low parameter counts.

DATASETS

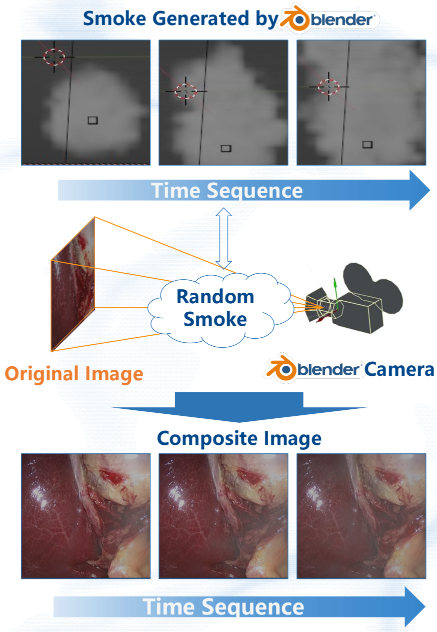

Fig. 6. Synthetic smoke generation method

We used images from the Cholec80[3] dataset and sampled 1,500 images at 20-second intervals from videos, selecting 660 representative smoke-free images. As detailed above, we added synthetic random smoke, yielding 660 image pairs.

RESULTS Ⅰ

Quantitative results

Model		Parameters↓	PSNR↑	SSIM↑	CIEDE2000↓
DCP		/	27.6250	0.5528	35.9952
CycleGAN	U-Net	54414K	28.7449	0.7621	10.3298
CycleGAN	ResNet6	7841K	29.0250	0.7826	9.5821
CycleGAN	ResNet9	11383K	29.0926	0.7802	9.2868
Pix2Pix	U-Net	54414K	29.2967	0.7073	8.8060
Pix2Pix	ResNet6	7841K	29.8249	0.8358	6.9364
Pix2Pix	ResNet9	11383K	29.8721	0.8417	6.7046
Pix2Pix	Uformer[4]	85605K	29.7030	0.8026	8.0602
Ablation Models					
w/o Multi-scale		613K	29.9970	0.8692	6.9362
w/o Fusion Block		629K	29.4425	0.7814	8.1200
w/o MBI		540K	29.7599	0.9029	6.9149
w/o LAT		90K	28.8936	0.7857	10.1284
Ours		**629K**	**30.4873**	**0.9061**	**5.4988**

Table. 1. Quantitative results. The best and second-best results are highlighted and underlined, respective.
PSNR, the Peak Signal-to-Noise Ratio, quantifies the difference between a reconstructed and original image. **SSIM**, Structure Similarity Index Method, is a perception based model, used to measure the similarity between two images. **CIEDE2000** represents color reconstruction accuracy for the human visual system.

The higher PSNR and SSIM, the lower CIEDE2000[5] indicate that the estimated smokefree images are similar to the real smoke-free images, which means a better desmoking capability.

RESULTS Ⅱ

Quantitative results

Smoke Density		Light Smoke			Medium Smoke			Heavy Smoke		
Model		PSNR↑	SSIM↑	CIEDE2000↓	PSNR↑	SSIM↑	CIEDE2000↓	PSNR↑	SSIM↑	CIEDE2000↓
DCP		27.6611	0.6215	30.1270	27.6811	0.5887	32.9143	27.6944	0.5807	33.8072
CycleGAN	U-Net	29.0426	0.7778	8.5370	28.9490	0.7607	10.7167	28.8837	0.7639	10.7521
CycleGAN	ResNet6	29.0713	0.7958	8.2868	28.7621	0.7741	11.7635	28.7647	0.7755	11.6661
CycleGAN	ResNet9	29.3232	0.8002	7.8017	28.7466	0.7650	11.9671	28.9379	0.7711	10.8202
Pix2Pix	U-Net	29.2652	0.7270	8.9004	29.4071	0.7119	9.1812	29.4474	0.7199	8.9037
Pix2Pix	ResNet6	29.9776	0.8404	6.6498	30.1833	0.8288	6.8033	30.2138	0.8344	6.2970
Pix2Pix	ResNet9	29.9492	0.8484	6.6610	30.1498	0.8372	**6.7079**	30.3287	0.8434	6.7079
Ours		**30.1209**	**0.8856**	**6.5182**	**30.2740**	**0.8704**	6.8001	**30.5223**	**0.8762**	**6.1147**

Table. 2. Quantitative comparison between SOTAs under different smoke densities

Qualtitive results

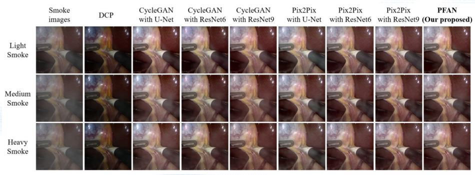

Fig. 7. Qualitative comparison between SOTAs under different smoke densities

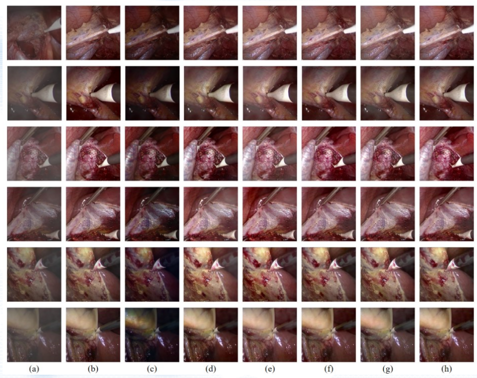

Fig. 8. Comparison experiments between SOTAs. (a) Input (b) Ground Truth, (c) Dark Channel Prior(DCP) [6] (d) CycleGAN + ResNet, (e) CycleGAN + U-Net, (f) Pix2Pix + ResNet, (g) Pix2Pix + U-Net, and (h) Ours

LIMITATIONS

- The method does not account for external factors like water vapor and pure white gauze that can degrade image quality.

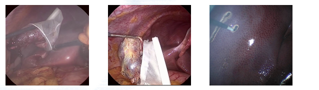

Fig. 9. Laparoscopic images with pure white gauze

- It may introduce temporal discontinuity in video desmoking tasks due to fluctuations in smoke density.

REFERENCES

[1] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1–9 (2015)
[2] Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: [Con_x0002_vNeXt CVPR22] A ConvNet for the 2020s. Cvpr pp. 11976–11986 (2022), http://arxiv.org/abs/2201.03545
[3] Twinanda, A.P., Shehata, S., Mutter, D., Marescaux, J., De Mathelin, M., Padoy, N.: EndoNet: A Deep Architecture for Recognition Tasks on Laparo_x0002_scopic Videos. IEEE Transactions on Medical Imaging 36(1), 86–97 (2017).
https://doi.org/10.1109/TMI.2016.2593957
[4] Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., Li, H.: Uformer: A General U-Shaped Transformer for Image Restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
[5] Luo, M.R., Cui, G., Rigg, B.: The development of the cie 2000 colour-difference for_x0002_mula: Ciede2000. Color Research & Application: Endorsed by Inter-Society Color
Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Fran¸cais de la Couleur 26(5), 340–350 (2001)
[6] He, K., Sun, J., TTang, X.: Single image haze removal using dark channel prior. 2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009 (January 2011), 1956–1963 (2009). https://doi.org/10.1109/CVPRW.2009.5206515

Code, and data are available:

https://github.com/jlzcode/PFAN

Acknowledgement:
This research was supported by multiple grants, including:
The National Key Research and Development Program of China (2020YFB1313900), National Natural Science Foundation of China (62072452), Shenzhen Science and Technology Program(JCYJ20200109115627045, JCYJ20220818101408019, JCYJ20200109115201707) and Regional Joint Fund of Guangdong (2021B1515120011).