Proceedings of The 6th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/2026.31969

Smoke Removal in Laparoscopic Surgical Videos Using Smoke-Free Prior

Siyue Chen[1], Rui Li[1], Miaomiao Li[2], Yao Lu[1*]

[1]School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China
[2]Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Department of Medical Imaging, Shandong Medical College, Jinan, China
**Corresponding Author. Email: luyao23@mail.sysu.edu.cn*

Abstract. During laparoscopic surgery, the smoke generated by instruments can obstruct the surgical field of view. The existing surgical smoke removal methods have not fully utilized the temporal redundant information in video data. To address this, we propose DeSmoke-SFP, a new framework that leverages smoke-free prior for smoke removal. Our approach integrates a video predictor to predict anatomical structures from preceding clear frames, generating a temporally informed, smoke-free prediction for a given smoky input. The predictive result is then fused with the current smoky frame by an image desmoker within an improved Cycle-GAN architecture, designed for unpaired image-to-image translation. In addition, a smoke feature loss based on dark channel prior is used to model the smoke in detail and guide the architecture to remove smoke. We train and evaluate our method on a dataset of real laparoscopic videos. Extensive experiments demonstrate that the proposed method outperforms state-of-the-art approaches, particularly in heavily smoke-affected regions. Quantitative results on severe smoke frames show superior performance in metrics: FADE (0.195), NIQE (5.152), and REA (1.061). Ablation studies validate the effectiveness of video prediction and smoke feature loss in the method for restoring anatomical details and improving surgical visibility.

Keywords: Laparoscopic Surgery Video, Smoke Removal, Medical Image Analysis, Deep Learning

1. Introduction

During laparoscopic surgery, the smoke generated by instruments will obstruct surgical view and interfere the surgical process [1,2]. Different from natural haze scenes where the smoke is global, the smoke in the surgical scene exhibits strong spatiotemporal localization characteristics: it appears, spreads unevenly within the frame, and dissipates over time. This spatiotemporal locality information allows anatomical structures covered by smoke to be found before few frames, which we named it "smoke-free prior" [3].

This paper proposes a new smoke removal framework (DeSmoke-SFP) based on an improved Cycle-GAN architecture, designed for unpaired translation between smoky and clear surgical video

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.31969

frames. The model incorporates a video predictor to propagate smoke-free prior from preceding clear frames, and an image desmoker that integrates this smoke-free prior with spatial anatomical features from smoky frames to generate smoke-free outputs. The model integrate the dark channel prior directly into the loss function, which effectively constrains the network to restore images in a physically-grounded manner. The framework is trained and validated on an unpaired dataset of laparoscopic videos. The main contributions are as followed:

(1) A novel method leveraging semantic data from smoke-free frames to remove surgical smoke.

(2) An unpaired data approach for smoke removal, better reflecting real-world conditions than synthetic dataset methods.

(3) A dark channel prior-based loss function designed to model smoke distribution and guide the model in leveraging smoke-free information for effective smoke removal.

2. Literature review

The existing smoke removal methods can be divided into single-image-based methods and video-based methods. Researchers have focused largely on single-image-based methods, leaving video-based methods relatively understudied in comparison.

The single-image-based methods can be generally devided into physics-based or deep learning methods. A common physical model for this problem is the atmospheric scattering model [4]. The dark channel prior has been widely adopted for dehazing [5], assuming globally uniform smoke distribution. However, physics-based methods will introduce color distortion and artifacts under the complex illumination typical of surgical conditions. Deep learning based methods, particularly generative adversarial networks [6], have been widely adopted. Some methods synthesize paired training data using tools like Blender to generate surgical smoke [7,8]. However, the complex light field in an operating room is highly complex, meaning this model does not translate well to real surgical settings. Unpaired learning strategies [9,10], such as those integrating CycleGAN [11] architectures, avoid relying on synthetic data; nevertheless, it challenge to recover fine anatomical details in dense smoke due to the inherent limitations of single-frame information.

Many video desmoking methods use a foreground-background fusion strategy, processing elements separately before combining them. In processing natural videos, common techniques typically rely on operations like frame differencing [12] or employ enhancement algorithms such as CLAHE [13] in their pipelines. However, these methods are not applicable to surgical conditions, which have unclear foreground-background distinctions and unique smoke dynamics. Therefore, utilizing time redundant information to achieve smoke removal in surgical scenarios remains challenging.

3. Methodology

3.1. Overview

The overall architecture of our algorithm is illustrated in Figure 1. We maintain two sequences: a smoke-free frame sequence C and a smoky frame sequence S. At each time step t, for a smoky frame s_t, we employ a video predictor (VP) using the smoke-free frames $\{c_{t-k}, c_{t-k+1}, \ldots, c_{t-1}\}$ (k is the required input sequence length for the video predictor) to generate a predicted smoke-free frame c_t^{vp}. We then integrate this predicted frame with the smoky frame s_t using a image desmoker (F) to produce $\widehat{c_t}$. Finally, $\widehat{c_t}$ is added to the smoke-free frame sequence C.

DOI: 10.54254/2755-2721/2026.31969

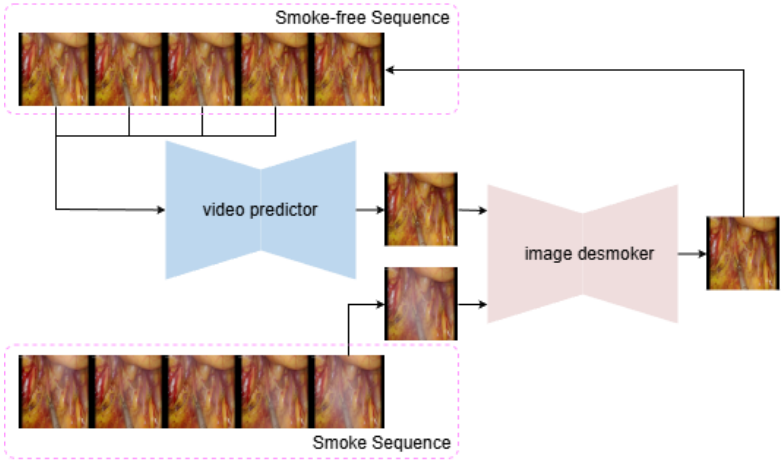

Figure 1. The overview of proposed method

3.2. Video predictor

The video predictor is designed to forecast the anatomical structures in the (k+1)-th frame based on the first k input frames. This process involves analyzing spatiotemporal features from the historical sequence to accurately project and localize anatomical details in subsequent frames.

It adopts an Encoder-Translator-Decoder architecture. The encoder is used to extract spatial features. The translator learns the temporal evolution. And the decoder integrates spatiotemporal information and predicts future frames.

During training process, we utilize smoke-free video clips, each containing k+1 frames. The first k frames serve as input, while the (k+1)-th frame is used as the ground truth for the output. The optimization objective of the video predictor is to make the predicted output c_t^{vp} as close as possible to the true output, i.e., $c_t^{vp} \approx c_t$. We achieve this goal by optimizing the network using Mean Squared Error (MSE) loss. Based on experimental results and a comprehensive consideration of both temporal performance and output quality, the parameter k is set to 4.

3.3. Image desmoker

Given unpaired data consisting of pairs of smoke-free frames and their predictions by the video predictor denoted as $\{c_i, c_i^{vp}\}_{i=1}^{N}$ where $c_i \in C$ and $c_i^{vp} \in C^{vp}$, and pairs of smoky frames and their smoke-free predictions by the video predictor, denoted as $\left\{s_j, s_j^{vp}\right\}_{j=1}^{M}$ where $s_j \in S$ and $s_j^{vp} \in S^{vp}$ (since the first k frames before each smoky frame are smoke-free, the predictions for smoky frames are also smoke-free), we represent the data distributions as $c, c^{vp} \sim p_{data}\left(c, c^{vp}\right)$ and $s, s^{vp} \sim p_{data}\left(s, s^{vp}\right)$. Our goal is to find two mappings: the first mapping is an image desmoker F that transforms the set of smoky frames and their smoke-free predictions $\{\mathrm{S}, S^{vp}\}$ into smoke-free frames C ; the second mapping is an image smoke generator G that transforms smoke-free frames C into smoky frames S .

DOI: 10.54254/2755-2721/2026.31969

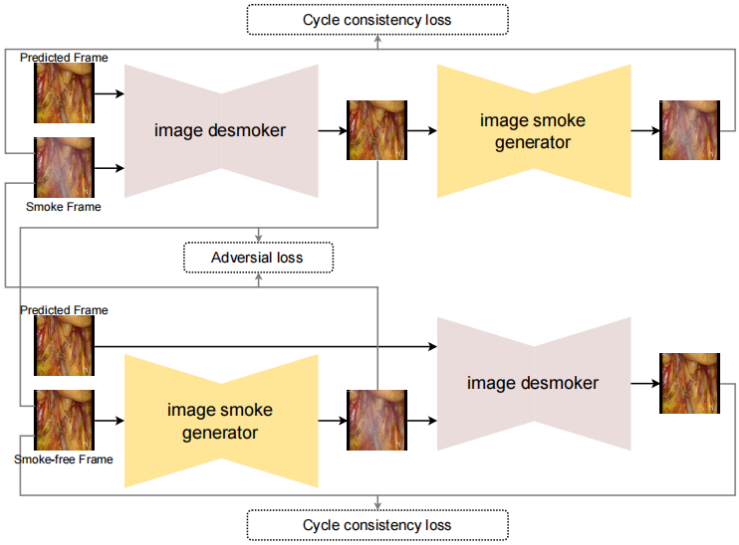

Figure 2. The overview of the proposed image desmoker's training process

The overview of our framework is illustrated in Figure 2. On one hand, we feed both the smoky frame s and its smoke-free prediction s^{vp} into the image desmoker to obtain a fake smoke-free frame s^{fake}. We then apply the image smoke generator to the smoke-free frame s^{fake} to map it back to the original smoky domain. On the other hand, we input smoke-free frame c into the image smoke generator to generate a fake smoky frame c^{fake}, which is then fed into the image desmoker to recover the original smoke-free frame. Additionally, since the smoke-free frames are sampled from a continuous sequence of smoke-free video, the predicted frame c^{vp} generated by the video predictor using the preceding k smoke-free frames should belong to the same image domain as s^{vp}. Therefore, we input both c^{fake} and c^{vp} into the image desmoker for smoke removal, ensuring they are mapped back to the original smoke-free domain.

The framework includes two discriminators, D_C and D_S. The discriminator D_C distinguishes between real smoke-free frames $\{\mathrm{c}\}$ and the desmoked frames $\{\mathrm{F}(s, s^{vp})\}$ produced by the image smoke desmoker. Meanwhile, D_S distinguishes between real smoky frames $\{\mathrm{s}\}$ and the smoke-added frames $\{\mathrm{G}(c)\}$ generated by the image smoke generator.

The model's optimization combines adversarial loss and cycle-consistency loss. The adversarial loss aligns generated and target domain distributions, with the discriminator maximizing and the generator minimizing it. Cycle consistency ensures that translating back and forth reconstructs the original input: $c \to G(c) \to F(G(c)) \approx c$ and $s \to F(s) \to G(F(s)) \approx s$. The total loss integrates both components:

$$\mathscr{L}(F, D_C, C, S) = \mathbb{E}_{c,c^{vp}\sim p_{data}(c,c^{vp})}[\log D_C(c)] + \mathbb{E}_{s,s^{vp}\sim p_{data}(s,s^{vp})}[\log(1 - D_C(F(s, s^{vp})))], \quad (1)$$

$$\mathscr{L}(G, D_S, C, S) = \mathbb{E}_{s,s^{vp}\sim p_{data}(s,s^{vp})}[\log D_S(s)] + \mathbb{E}_{c,c^{vp}\sim p_{data}(c,c^{vp})}[\log(1 - D_S(G(c)))], \quad (2)$$

DOI: 10.54254/2755-2721/2026.31969

$$\mathscr{L}_{GAN}(G,F,D_C,D_S)=\min_{F}\max_{D_C}\mathscr{L}(F,D_C,C,S)+\min_{G}\max_{D_S}\mathscr{L}(G,D_S,C,S), \tag{3}$$

$$\mathscr{L}_{cyc}(G,F)=\mathbb{E}_{c,c^{vp}\sim p_{data}(c,c^{vp})}\left[||F(G(c),c^{vp})-c||_1\right]+\mathbb{E}_{s,s^{vp}\sim p_{data}(s,s^{vp})}\left[||G(F(s,s^{vp}))-s||_1\right], \tag{4}$$

$$\mathscr{L}(G,F,D_C,D_S)=\mathscr{L}_{GAN}(G,F,D_C,D_S)+\alpha\cdot\mathscr{L}_{cyc}(G,F). \tag{5}$$

3.4. Smoke feature loss

We propose a smoke feature loss aimed at guiding the image desmoker F to properly leverage the information from the predicted smoke-free frame s^{vp} to perform smoke removal.

$$M(s,s^{vp})=\mathscr{J}^{dark}(s)-\mathscr{J}^{dark}(s^{vp}), \tag{6}$$

$$\mathscr{L}_{SF}(F)=\mathbb{E}_{s,s^{vp}\sim p_{data}(s,s^{vp})}\frac{||M(s,s^{vp})\odot|F(s,s^{vp})-s^{vp}|||_1}{||M(s,s^{vp})||_1}, \tag{7}$$

where $\odot$ represents the Hadamard product of two matrices, and $J^{dark}(s)$ and $J^{dark}(s^{vp})$ denote the dark channels [5] of images s and s^{vp}, which is significantly higher in smoky regions due to light scattering:

$$\mathscr{J}^{dark}(x)=\min_{y\in\Omega(x)}\left(\min_{c\in\{\mathrm{r,g,b}\}}\mathscr{J}^{c}(y)\right). \tag{8}$$

In conjunction with the generative adversarial loss and the cycle consistency loss mentioned previously, the ultimate optimization goal of our network is:

$$\mathscr{L}(G,F,D_C,D_S)=\mathscr{L}_{GAN}(G,F,D_C,D_S)+\alpha\cdot\mathscr{L}_{cyc}(G,F)+\beta\cdot\mathscr{L}_{SF}(F). \tag{9}$$

where α and β are the weighting coefficients. After the experiment, we set both coefficients to 0.1.

3.5. Smoke removal in the smoky video

Given a smoky frame s_{t_0} at time t_0, we first feed the preceding k smoke-free frames $\{c_{t_0-k},\ldots,c_{t_0-1}\}$ into the video predictor(VP) to predict a temporally-informed prediction $s^{vp}_{t_0}=VP(\{c_{t_0-k},c_{t_0-k+1},\ldots,c_{t_0-1}\})$. The image desmoker(F) then takes both the smoky frame

DOI: 10.54254/2755-2721/2026.31969

s_{t_0} and its predicted smoke-free version $s_{t_0}^{vp}$ as input, producing the final desmoked output $\widehat{c_{t_0}} = F\left(s_{t_0}, s_{t_0}^{vp}\right)$.

This process continues recursively for each subsequent smoky frame s_t, where the latest k available smoke-free frames (including previously desmoked outputs) are used by the video predictor to provide contextual semantic information. The resulting smoke-free frame $\widehat{c_t}$ is then appended to the sequence, updating it to $\left\{c_{t-k+1}, \ldots, c_{t-1}, \widehat{c_t}\right\}$ for processing the next frame. This iterative approach progressively enhances smoke removal by leveraging both original smoky content and predicted clear temporal information.

3.6. Evaluation indicators

As our dataset contains authentic surgical videos without ground truth, reference-based metrics (e.g., SSIM, PSNR) are unsuitable. We instead adopt no-reference metrics: Fog Aware Density Evaluator (FADE) [14], Natural Image Quality Evaluator (NIQE) [15] and Restoring Edge Assessment (REA) [16].

FADE [14] quantifies fog density (higher values indicate more fog). NIQE [15] blindly assesses image quality (lower scores are better). REA [16] measures edge restoration performance (higher values indicate better recovery).

4. Experiments

4.1. Datasets

We constructed a novel laparoscopic video dataset from patients at Shandong Cancer Hospital and Institute, with videos sampled at 30fps and cropped to 256×256. Precise smoke event annotation yielded 481 segments containing 4 pre-smoke frames and full smoke evolution sequences (10–80 frames each), along with 495 smoke-free segments (5 frames each) for training both the image desmoker and the video predictor.

Table 1. The quantitative comparison between our method and state-of-the-art approaches. Metrics with ↑/↓ arrows prefer higher/lower scores respectively. Bold-face indicates statistically significant best results

	Input	DCP	DeSmoke-LAP	Proposed(single-frame)	Proposed (w/o SF)	Proposed
FADE↓	0.538±0.171	0.325±0.040	0.201±0.031	0.212±0.035	0.203±0.031	0.195±0.044
NIQE↓	5.602±1.045	5.596±0.962	5.162±0.823	5.404±0.452	5.329±0.847	5.152±0.794
REA↑	0.00	0.361±0.256	1.005±0.531	0.945±0.539	0.923±0.506	1.061±0.487

DOI: 10.54254/2755-2721/2026.31969

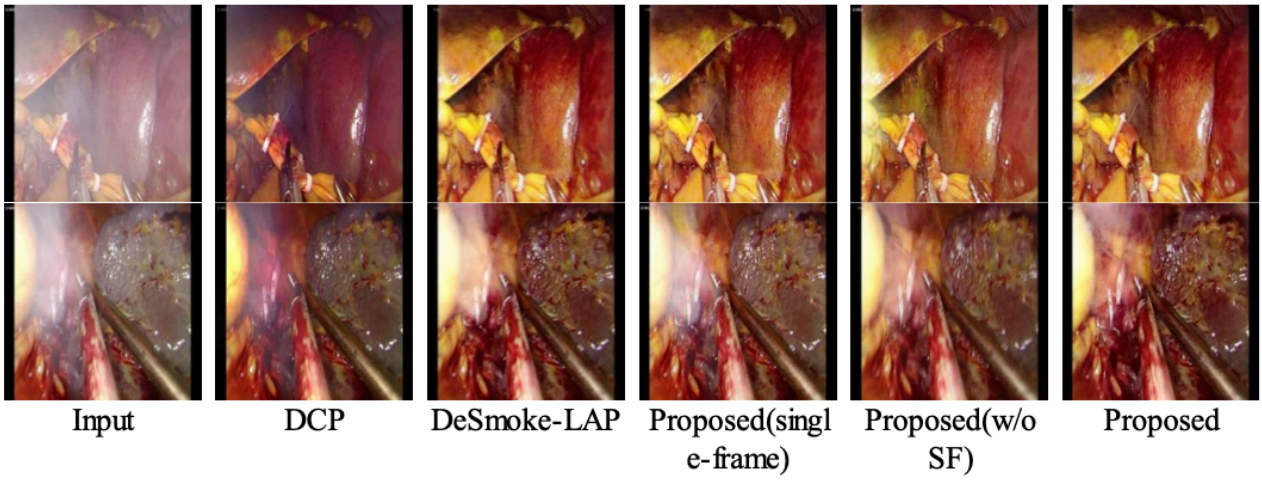

Figure 3. Qualitative comparison between proposed method and state-of-the-art approaches

4.2. Comparison with state-of-the-art methods

Since the real dataset lacks paired smoke-free and smoky data, methods relying on paired references are inapplicable. Therefore, we only compare with unpaired approaches, including representative physics-based and deep learning-based methods [5,10].

Quantitative Comparison. As shown in Table 1, our method outperforms others under heavy smoke conditions. DCP [5] reduced FADE from 0.538 to 0.325, with minor improvements in NIQE and REA. DeSmoke-LAP [10] further lowered FADE to 0.201 and NIQE to 5.162, while increasing REA by 1.005. By using smoke-free prior, our approach achieves the best results: FADE decreases to 0.195, NIQE to 5.152, and REA improves by 1.061. These results confirm our method's superior performance in smoke reduction, image restoration, and edge recovery.

Qualitative Comparison. Figure 3 qualitatively compares various methods. The first row displays low-smoky cases: DeSmoke-LAP [10] removes haze effectively, whereas DCP [5] shows color distortion. The second row presents heavily smoke-affected cases that fully obscures anatomical structures, single-frame approaches [5,10] inadequately eliminate smoke. With the using of smoke-free prior, our method can better restore smoke-affected anatomical details.

4.3. Ablation study

We conduct ablation studies by progressively adding modules: (1) image desmoker with traditional CycleGAN framework (single-frame); (2) image desmoker with framework introduced in Sec. 3.3 (w/o SF); (3) image desmoker with framework in Sec 3.3 and loss function in Sec 3.4 (proposed).

Quantitative Comparison. As shown in Table 1, the image desmoker baseline achieves a FADE of 0.212, confirming single-frame effectiveness. Adding the predictive result reduces the FADE score to 0.203. The full model achieves the best FADE (0.195), validating the proposed fusion strategy.

Qualitative Comparison. Figure 3 shows that the baseline struggles in smoked regions. After incorporating the prediction results, significant improvements were observed in smoked regions. The addition of a smoke feature loss to guide the image desmoker using the video predictor's output further enhanced performance in these areas.

DOI: 10.54254/2755-2721/2026.31969

5. Conclusion

We propose DeSmoke-SFP, a novel framework that enhances single-frame desmoking using smoke-free prior from preceding clear frame. A video predictor extracts smoke-free prior, which are adaptively fused with spatial features by an enhanced CycleGAN-based image desmoker. A dark channel prior-based loss function guides the image desmoker to leverage the smoke-free prior. Our method outperforms state-of-the-art approaches, reducing the FADE metric by 0.006 compared to the best single-frame baseline and effectively restoring anatomical structures in smoky regions. Ablation studies confirm the contribution of each component, with the full model achieving optimal performance (FADE: 0.195).

Acknowledgement

This work was supported in part by the National Key R&D Program of China under Grant 2023YFE0204300; in part by the R&D project of Pazhou Lab(HuangPu) under Grant 2023K0606; in part by the National Natural Science Foundation of China under Grants 82441027 and 62371476; in part by the Guangzhou Science and Technology Bureau under Grant 2023B03J1237; in part by the Health Research Major Projects of Hunan Health Commission under Grant W20241010; and in part by the Guangdong Province Key Laboratory of Computational Science at the Sun Yat-sen University under Grant 2020B1212060032.

References

[1] Ulmer B. C. (2008). The hazards of surgical smoke. AORN Journal, 87(4), 721–738. https: //doi.org/10.1016/j.aorn.2007.10.012

[2] Zhou, Y., Hu, Z., Xuan, Z., Wang, Y., & Hu, X. (2024). 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. https: //doi.org/10.1109/TCBB.2022.3204673

[3] Limchantra, I. V., Fong, Y., & Melstrom, K. A. (2019). Surgical Smoke Exposure in Operating Room Personnel: A Review. JAMA surgery, 154(10), 960–967. https: //doi.org/10.1001/jamasurg.2019.2515

[4] Narasimhan, S. G., & Nayar, S. K. (2003). Contrast restoration of weather degraded images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(6), 713–724. https: //doi.org/10.1109/TPAMI.2003.1201821

[5] He, K., Sun, J., & Tang, X. (2009). Single image haze removal using dark channel prior. In 2009 IEEE Conference on Computer Vision and Pattern Recognition(pp. 1956-1963). IEEE. https: //doi.org/10.1109/CVPR.2009.5206515

[6] Abdel-Basset, M., Moustafa, N., & Hawash, H. (2023). Generative adversarial networks (GANs). In Deep learning approaches for security threats in IoT environments(pp. 271–285). John Wiley & Sons. https: //doi.org/10.1002/9781119884170.ch12

[7] Sidorov, O., Wang, C., & Cheikh, F. A. (2019). Generative smoke removal. In A. V. Dalca, M. B. A. McDermott, E. Alsentzer, S. G. Finlayson, M. Oberst, F. Falck, & B. K. Beaulieu-Jones (Eds.), Proceedings of the Machine Learning for Health Workshop at the 33rd Conference on Neural Information Processing Systems(Vol. 116, pp. 81–92). PMLR.

[8] Hong, T., Huang, P., Zhai, X., Gu, C., Tian, B., Jin, B., & Li, D. (2023). MARS-GAN: Multilevel-feature-learning attention-aware based generative adversarial network for removing surgical smoke. IEEE Transactions on Medical Imaging, 42(8), 2299–2312. https: //doi.org/10.1109/TMI.2023.3245298

[9] Engin, D., Genç, A., & Ekenel, H. K. (2018). Cycle-Dehaze: Enhanced CycleGAN for single image dehazing. In 2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)(pp. 825–833). IEEE Computer Society. https: //doi.org/10.1109/CVPRW.2018.00127

[10] Pan, Y., Bano, S., Vasconcelos, F., Park, H., Jeong, T. T., & Stoyanov, D. (2022). 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. https: //doi.org/10.1007/s11548-022-02595-2

[11] Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In 2017 IEEE International Conference on Computer Vision (ICCV)(pp. 2242–2251). IEEE Computer Society. https: //doi.org/10.1109/ICCV.2017.244

DOI: 10.54254/2755-2721/2026.31969

[12] Xu, Z., Liu, X., & Chen, X. (2009). Fog removal from video sequences using contrast limited adaptive histogram equalization. In 2009 International Conference on Computational Intelligence and Software Engineering(pp. 1-4). IEEE. https: //doi.org/10.1109/CISE.2009.5366207
[13] John, J., & Wilscy, M. (2008). Enhancement of weather degraded video sequences using wavelet fusion. In 2008 7th IEEE International Conference on Cybernetic Intelligent Systems(pp. 1-6). IEEE. https: //doi.org/10.1109/UKRICIS.2008.4798926
[14] Choi, L. K., You, J., & Bovik, A. C. (2015). Referenceless prediction of perceptual fog density and perceptual image defogging. IEEE Transactions on Image Processing, 24(11), 3888–3901. https: //doi.org/10.1109/TIP.2015.2456502
[15] Mittal, A., Soundararajan, R., & Bovik, A. C. (2013). Making a "completely blind" image quality analyzer. IEEE Signal Processing Letters, 20(3), 209–212. https: //doi.org/10.1109/LSP.2012.2227726
[16] Hautière, N., Tarel, J.-P., & Aubert, D. (2007). Blind contrast restoration assessment by gradient ratioing at visible edges. In 2007 IEEE International Conference on Image Processing(Vol. 3, pp. III-361–III-364). IEEE. https: //doi.org/10.1109/ICIP.2007.4379281