

Multi-Frequency and Smoke Attention-aware Learning based Diffusion Model for Removing Surgical Smoke

Hao Li¹, Xiangyu Zhai², Jie Xue¹, Changming Gu³, Baolong Tian³, Tingxuan Hong⁴, Bin Jin², Dengwang Li¹(✉), and Pu Huang¹(✉)

¹ Shandong Key Laboratory of Medical Physics and Image Processing, Shandong Normal University, Jinan, Shandong 250358, China

{dengwang, huangpu}@sdnu.edu.cn

² Department of Hepatobiliary Surgery, Second Hospital of Shandong University, Jinan, Shandong 250033, China

³ Optical and Digital Image Processing Division, Qingdao NovelBeam Technology Co., Ltd., Qingdao, Shandong 266100, China

⁴ Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin 300072, China

Abstract. Surgical smoke in laparoscopic surgery can deteriorate the visibility and pose hazards to surgeons, although medical devices for mechanical smoke evacuation worked well, its prolonged operative duration and thus restricted the efficiency. This work aims to simultaneously remove the surgical smoke and restore the true-to-live image colors with deep learning strategy to improve the surgical efficiency and safety. However, the deep network-based smoke removal remains a challenge due to: 1) higher frequency modes are hindered from being learned by spectral bias, 2) the distribution of surgical smoke is non-homogeneity. We propose the multi-frequency and smoke attention-aware learning-based diffusion model for removing surgical smoke. In this work, the frequency compensation strategy combines the multi-level frequency learning and contrast enhancement to integrates comprehensive features for learning mid-to-high frequency details that the smoke has obscured. The smoke attention learning employs the pixel-wise measurement and provides the diffusion model with complementary features about where smoke is present, which helps restore the smokeless regions during the inverse diffusion process. And the multi-task learning strategy incorporates L_1 loss, smoke perception loss, dark channel prior loss, and contrast enhancement loss to help the model optimization. Additionally, a paired smokeless/smoky dataset is simulated by a 3D smoke rendering engine. The experimental results show that the proposed method outperforms other state-of-the-art methods on both synthetic/real laparoscopic surgical images, with the potential to be embedded in laparoscopic devices for smoke removal.

Keywords: Laparoscopic surgery · Smoke attention · Diffusion model.

H. Li and X. Zhai contributed equally.

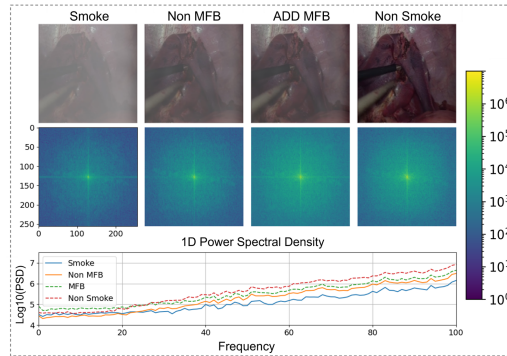


Fig. 1. Power spectral density (PSD) analysis on the predicted smoke-less image, smoke image and non-smoke image. The first row shows the smoke image, the predicted smoke-less image without/with multi-level frequency block (MFB), and the ground-truth smoke-less image. The 2nd row shows their 2D Fourier transforms results. The 3rd row shows the $\log_{10}(\text{PSD})$ curves.

1 Introduction

During laparoscopic surgery, surgeons use high-frequency electric and ultrasonic scalpels to cauterize human tissue for cutting operations, but this process generates surgical smoke [15]. As shown in Fig. 1, the surgical smoke can severely reduce visibility in the abdominal cavity and pose a hazard to surgeons, hence the urgent need to improve the quality of intraoperative images in clinical laparoscopic scenarios [21].

Traditional physical models such as: atmospheric scattering model (ASM) based, dark channel prior (DCP) [8] indicated that by combining DCP and dehazing model, the thickness of haze can be directly estimated and high-quality haze less images can be recovered. Surgery smoke is characterized by high variability and non-uniformity, deep learning-based algorithms, including generative adversarial networks (GANs) [11] and diffusion model [13] were studied for smoke removal. [19] proposed an improved GAN structure and called SSIM-PAN. [25] proposed DS-CycleGAN and it was constructed based on CycleGAN for smoke removal. [24] proposed MPR-Net, this model first learned the contextualized features using encoder-decoder architectures. [12] proposed CG-ASM driven semi-supervised learning framework for high-quality pixel-wise laparoscopic image enhancement. [17] proposed CGAN-DC and it embedded the dark channel into pix2pix. [16] proposed an end-to-end feature fusion attention network (FFA-Net) to restore haze less images. [9] proposed a learning module containing multi-level smoke features, and this module with a GAN network for smoke removal and image generation. [3] proposed a novel computational framework for unsupervised collaborative learning called De-smoke GCN. [5] proposed Cycle-Dehaze for dehazing. [2] proposed a novel nonlinear activation function in Dehaze-Net called Bilateral Rectified Linear Unit. However, the mode collapse

of GANs might result in the loss of image details and introduce distortions [4, 22]. Diffusion model had recently emerged as an alternative to GANs and was less prone to mode collapse [7].

We present the multi-frequency and smoke attention-aware learning-based diffusion model for removing surgical smoke. A paired smokeless/smoky dataset is simulated by a 3D smoke rendering engine. The main contributions of our study are as follows: 1)The frequency compensation strategy combines the multi-level frequency learning and contrast enhancement to integrate comprehensive features for adaptively learning the higher frequency details that the surgical smoke has obscured. 2)The smoke attention learning employs the pixel-wise measurement and provides the diffusion model with additional information about where smoke is present, which helps restore the smokeless regions during the inverse diffusion process. 3)The multi-task learning strategy incorporates L_1 loss DCP loss, smoke perception loss, and contrast enhancement loss to help the model optimization.

2 Method

The de-smoking problem of laparoscopic image can be formulated as a translation from the smoky image domain to the smokeless image domain. Given the input $(I_t^{sm}, I_t^{sl}, \beta_t)$, where I_t^{sl} (I_0^{sl} denotes the smokeless image) is the noisy image and I_t^{sm} (I_0^{sm} denotes the smoky image) is the conditional image, and β_t is the noise variance. The noise can be iteratively mapped onto a sequence of arbitrary images across T time steps, and the translation between the smoky image domain to the smokeless image domain can be modeled by Gaussian distributions under large T and small β_t .

During the translation, I_0^{sl} can be gradually obtained from I_t^{sm} together with Gaussian noise ϵ_t for $t=T, \dots, 1$ via the updating rule:

$$I_{t-1}^{sl} = \frac{1}{\sqrt{\alpha_t}} \left(I_t^{sl} - \frac{1 - \alpha_t}{\sqrt{1 - \beta_t}} f_\theta(I_t^{sm}, I_t^{sl}, \beta_t) \right) + \sqrt{1 - \alpha_t} \epsilon_t \quad (1)$$

where $\epsilon_t \sim \mathcal{N}(0, I)$, α_t controls the noise scale at each step t. f_θ is the generative model that is used to adjust the current noisy image I_t^{sl} to obtain a smokeless image I_{t-1}^{sl} with less noise at the next time step, and the inputs include the noisy image I_t^{sl} , the conditional image I_t^{sm} and the noise variance β . I_t^{sl} and I_t^{sm} can provide the context information.

2.1 Multilevel Frequency Learning

From the spatial domain, the image obscured by the smoke has smoother variations in chromaticity and brightness than that within the smokeless image. From the frequency domain (Fig. 1), more mid-to-high frequencies existed within the smokeless image, while the mid-to-high frequencies are found to be reduced in the smoky images.

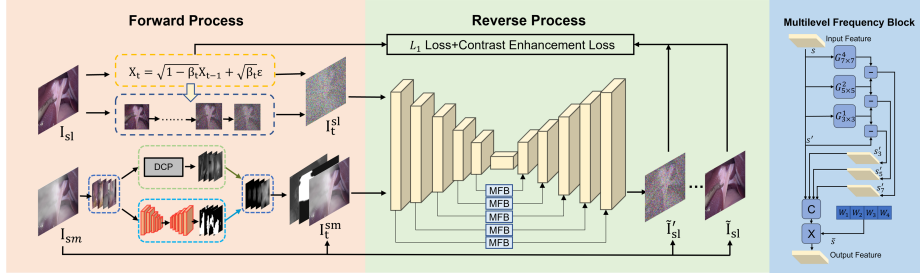


Fig. 2. The architecture of the proposed model, it includes the multilevel frequency learning block (MFB) and smoke attention learning block. Noise image I_t^{sl} and conditional image I_t^{sm} (the intersection of DCP maps, smoke masks and smoke image I_{sm}) are taken as inputs.

As shown in Fig. 2, f_θ is a symmetric U-shaped encoder-decoder structure, the detailed structure is shown in Appendix 1, the encoder consists of five convolutional blocks, given the feature map $x \in R^{H \times W \times C}$ of the last layer of each convolutional block, x is then passed through multi-level frequency block (MFB), which can emphasize the mid-to-high frequencies. MFB consists of three convolutional weights $G_{k \times k}^\sigma$ that transfer the mid-to-high frequencies from the encoder to the decoder via the skip connection. $G_{k \times k}^\sigma$ employs the Gaussian weights with $\sigma \in 1, 2, 4$ and $k \in 3, 5, 7$:

$$x_k = G_{k \times k}^\sigma * x \quad (2)$$

where $G_{k \times k}^\sigma$ is set to be non-learnable to keep the determined bandwidth unchanged. The mid-to-high frequencies are learned by subtracting the blurred feature maps from each other or from the original one. x' contains the mid-to-high frequency by concatenating these subtractions:

$$x' = [x, x - x_3, x_3 - x_5, x_5 - x_7] \quad (3)$$

Also, to allow the model adaptively to learn the relevant frequency bands, x' is aggregated via weighted averaging with learnable weights $W = [w_1, w_2, w_3, w_4]$. Meanwhile, to preserve its original signals, the identity mapping between x to x' is used, which is helpful during $t \rightarrow T$.

$$\bar{x} = \langle W, x' \rangle \quad (4)$$

where $\bar{x}$ is learned from MFB and is then transferred to the decoder via the skip connection. Furthermore, to prevent overfitting to artifacts during amplifying the high-frequency components, L_{ce} is introduced to balance the high-frequency between $\tilde{I}_{sl}$ and I_{sl} .

$$L_{ce} = abs \left(\log \left(\frac{Q_{I_{sl}}}{Q_{\tilde{I}_{sl}}} \right) \right) \quad (5)$$

where $Q_{I_{sl}}$ denotes the ratio of σ^2 to u , σ^2 and u denote the variance and the mean value of I_{sl} intensity, respectively. For $Q_{\tilde{I}_{sl}}$, the same principle applies.

Typically, σ^2 contributes significantly to its high-frequency, while u to the low-frequency. When noises dominate ($t \rightarrow T$), u closes to 0 while σ^2 closes to 1, L_{ce} tends to decrease. Conversely, when image details dominate ($t \rightarrow 1$), it forces σ^2 of $\tilde{I}_{sl}$ to be consistent with that of I_{sl} and avoiding being too large to result in artifacts.

2.2 Smoke Attention Learning

Considering the non-homogeneity property of smoke, the smoke attention (SA) learning is proposed to identify the areas within an image where surgical smoke is present and to estimate the density of the smoke in these regions, and the structure of SA network is shown in Appendix 2. Specifically, SA combines the smoke segmentation network (SSN) with the dark channel prior (DCP) module. SSN is a symmetric U-shaped network and it consists of encoder and decoder parts, it consists of contracting path (encoder stage) and expansive path (decoder stage). The encoder part of SSN is constructed based on the vanilla VGG16 network, except that the fully connected layers of VGG16 are removed, and the decoder part is inversely identical to the encoder. Also, the convolutional block attention module (CBAM) is introduced to enhance the perceptual ability of decoder for better capturing the smoke feature. Moreover, the SSN is trained independently before training the diffusion model.

For CBAM, both the channel and spatial attention is added after each group of convolutional blocks to recalibrate the feature maps by focusing on important features. Given the features map y , the channel attention $M_c(y)$ and spatial attention $M_s(y)$ are computed as:

$$M_c(y) = \text{Sigmoid} \left(\text{MLP}(\text{AvgPool}(y)) + \text{MLP}(\text{MaxPool}(y)) \right) \quad (6)$$

$$M_s(y) = \text{Sigmoid}(\text{Conv}^{7*7}([\text{AvgPool}(y), \text{MaxPool}(y)])) \quad (7)$$

$$y' = M_c \odot (M_s \odot y) \quad (8)$$

where Conv^{7*7} represents a convolution operation with kernel size of $7*7$, $[\quad, \quad]$ denotes the concatenation of the feature maps, $\odot$ denotes element-wise multiplication. And the refined feature map y' is obtained by applying the channel and spatial attention maps to the input feature map y .

The DCP module works jointly with SSN to enhance the smoke attention. The DCP covered by smoke has larger values than the dark channel without smoke, and the DCP tends to increase in the regions covered by smoke and decrease in the smokeless regions. Hence, DCP is sensitive to smoke images, and the combination of SSN and DCP can encode comprehensive smoke features to adaptively accommodate the complicated smoke.

2.3 Multi-task Learning

To adaptively optimize pixels in the smoke regions and prevent color distortion in the smokeless regions. The smoke perception is used to penalize the smoke

pixels and thus compensating the challenge of uneven smoke distribution. The smoke perception loss L_{sp} is defined as:

$$L_{sp} = \frac{1}{N} \left[\alpha_1 \sum_{x \notin S} I_d(x) + \sum_{x \in S} I_{sa} I_d(x) \right] \quad (9)$$

$$I_d(x) = |I_{sl}(x) - \tilde{I}_{sl}(x)| \quad (10)$$

where $\tilde{I}_{sl}$ and I_{sl} denote the synthetic smokeless image and the ground truth smokeless image. N denotes the total number of image pixels, S denotes the smoke attention, and x denotes the pixel index of the image. The penalty weight α_1 for smokeless region is empirically set to 0.2 to prevent large color distortion, and the attention penalty coefficient I_{sa} is restricted to $[0.5, 1]$.

To further improve the smoke removal, the DCP loss is introduced L_{dcp} is introduced and it forces the discriminators to take most of the smoke-covered pixels into account and facilitate the optimization of the generators. L_{dcp} is defined as:

$$L_{dcp} = \frac{1}{N} \sum_x [DCP(I_{sl}) - DCP(\tilde{I}_{sl})] \quad (11)$$

where DCP denotes the dark channel operation and N denotes the total number of image pixels. Herein, L_{dcp} is measured using its average value by cyclically traversing all pixels of the image, and it can capture the edges and contours of the smoke image, thereby preserving more detailed information.

To better capture the characteristics of the smoke and reduce the risk of overfitting, L_1 is defined as:

$$L_1 = \|f_\theta(I_t^{sm}, I_t^{sl}, \beta_t) - \varepsilon\|_1^1 \quad (12)$$

The multi-task objective function consists of smoke perception loss, DCP loss, and contrast enhancement loss, L_1 loss and thus the total loss L can be defined as:

$$L = L_1 + \lambda_1 L_{sp} + \lambda_2 L_{dcp} + \lambda_3 L_{ce} \quad (13)$$

For optimization, we empirically set $\lambda_1=0.8$, $\lambda_2=0.4$ and $\lambda_3=0.2$. Also, the performance comparison for parameter choice is shown in Appendix 4.

3 Experiment and Results

3.1 Datasets

In clinical practice, a paired smokeless/smoky laparoscopic surgical image can hardly be obtained. To compensate for this issue, the Blender [1] is used for smoke rendering to synthesize smoky images from its paired smokeless images. In this case, the real laparoscopic surgery dataset comes from both the Cholec80 dataset [20] and a real-world dataset collected from the Second Hospital of Shandong University.

Table 1. Quantitative comparison results of the proposed model with 10 other methods on a synthetic surgical dataset. The bold denotes best results.

Method	PSNR [10]	SSIM [18]	CIEDE-2000 [6]
SSIM-PAN [19]	23.358±1.684	0.846±0.050	6.856±2.720
DS-CycleGAN [25]	26.533±1.923	0.872±0.086	5.262±2.305
De-smoke GCN [3]	26.979±1.425	0.899±0.046	4.726±3.256
CG-ASM [12]	26.793±1.477	0.873±0.096	5.750±2.761
CGAN-DC [17]	26.640±1.381	0.879±0.037	5.848±2.478
MARS-GAN [9]	27.821±1.826	0.918±0.024	4.055±1.855
MPR-Net [24]	26.712±1.809	0.901±0.025	4.883±2.141
FFA-Net [16]	22.793±2.142	0.851±0.084	6.641±1.742
Cycle-Dehaze [5]	25.429±1.764	0.887±0.048	8.320±2.248
DCP [8]	15.747±1.380	0.728±0.045	12.451±4.520
Ours	28.314±1.921	0.929±0.032	3.994±1.901

3.2 Evaluation and Results

Synthetic smoke dataset As shown in Table 1, comparing to other methods, the proposed model achieves the best results in the three metrics, PSNR, SSIM, and CIEDE-2000. Comparing to the de-smoking model, the proposed model improves over De-smoke GCN by an average of 1.335, 0.030 in the PSNR and SSIM, and decreases by an average of 0.732 in the CIEDE-2000. Comparing to the de-hazing model, the proposed improves over MPR-Net by an average of 1.602, 0.028 in the PSNR and SSIM, and decreases by an average of 0.889 in the CIEDE-2000. Comparing to MARS-GAN (the 2nd best method), the proposed improves by an average of 0.493, 0.011 in the PSNR and SSIM and decreases by an average of 0.061 in the CIEDE-2000.

Real smoke dataset The proposed model achieves the best results with the BRISQUE [14], FADE [14], and CEIQ [23] metrics (Appendix 5). Comparing to the de-smoking model, the proposed model improves over De-smoke GCN by an average of 0.238 in the CEIQ, and decreases by an average of 2.666, 0.069 in the BRISQUE and FADE. Comparing to the de-hazing model, the proposed model improves over MPR-Net by an average of 0.318 in the CEIQ, and decreases by an average of 2.874, 0.080 in the BRISQUE and FADE. Comparing to MARS-GAN (the 2nd best method), the proposed model improves over MARS-GAN by an average of 0.015 in the CEIQ, and decreases by an average of 1.441, 0.002 in the BRISQUE and FADE.

3.3 Qualitative Results

As shown in Fig. 3, comparing to other models on the same real laparoscopic smoke dataset. Serious color distortion exists and the abdominal cavity is over-saturated using DCP and FFA-Net. The portion of the surface of the atraumatic

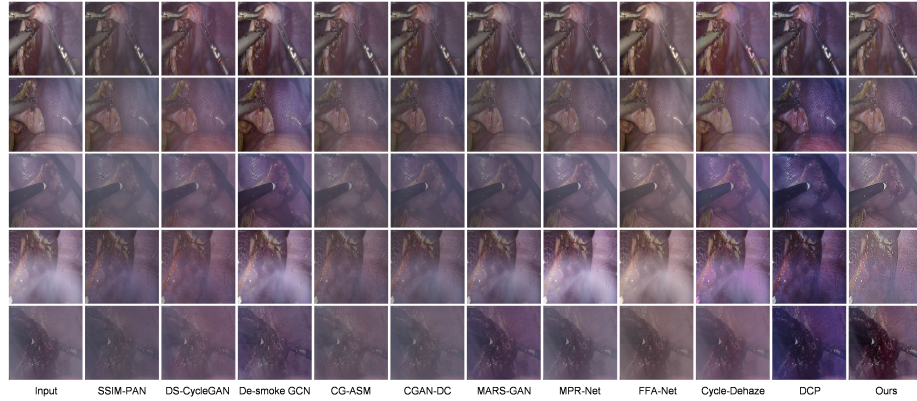


Fig. 3. The comparison of the proposed model with the comparative models originally used for de-smoking and de-hazing on the real laparoscopic surgical image dataset, and some randomly selected subjects are shown.

separation forceps seems to be eroded using Cycle-Dehaze. The smoke is not removed by MPR-Net, and the images get blurred. The images are prone to smooth and darkening using DS-CycleGAN and CG-ASM. And the de-smoking effect is not obvious using SSIM-PAN, De-smoke GCN, CGAN-DC, and MARS-GAN.

For Cholec80 dataset, the de-smoking of Cholec80 is shown in Appendix 3, which shows that the proposed model removes the smoke, as well as restores the true-to-life color. The images obtained from DS-CycleGAN look yellowish tint color, Cycle-Dehaze, DCP cause color discoloration and distortion, especially the images obtained from DCP get too dark. And De-smoke GCN, CG-ASM, CGAN-DC, MARS-GAN and MPR-Net do not well remove the smoke, SSIM-PAN and FFA-Net cause the luminance distortion.

4 Conclusion

In this study, we propose the multi-frequency and smoke attention-aware learning based diffusion model for removing surgical smoke. It incorporates the multilevel smoke attention-aware learning, multilevel frequency learning and multi-task learning into together, and the paired smokeless/smoky images are synthesized for model training. The qualitative and quantitative evaluation show that the proposed model achieves remarkable de-smoking performance and outperforms 10 compared models, and this has the potential to assist surgeons for efficient surgical smoke removal.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (62201330, 62271294), Shandong Province Major Technological Innovation (2022CXGC010502, 2023CXPT094), Shandong Province Scientific and Technological Achievement Transfer and Transformation Subsidy (Shandong-Chongqing Science and Technology Cooperation) Project (2022LYXZ027).

Disclosure of Interests. All authors disclosed no relevant relationships.

References

1. Baechler, O., Greer, X.: Blender 3D By Example: A project-based guide to learning the latest Blender 3D, EEVEE rendering engine, and Grease Pencil. Packt Publishing Ltd (2020)
2. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. *IEEE transactions on image processing* **25**(11), 5187–5198 (2016)
3. 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)
4. Du, Y., Jiang, Y., Tan, S., Wu, X., Dou, Q., Li, Z., Li, G., Wan, X.: Arsdm: colonoscopy images synthesis with adaptive refinement semantic diffusion models. In: *International conference on medical image computing and computer-assisted intervention*. pp. 339–349. Springer (2023)
5. Engin, D., Genç, A., Kemal Ekenel, H.: Cycle-dehaze: Enhanced cyclegan for single image dehazing. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. pp. 825–833 (2018)
6. Gómez-Polo, C., Muñoz, M.P., Luengo, M.C.L., Vicente, P., Galindo, P., Casado, A.M.M.: Comparison of the cielab and ciede2000 color difference formulas. *The Journal of prosthetic dentistry* **115**(1), 65–70 (2016)
7. Graham, M.S., Pinaya, W.H.L., Wright, P., Tudosi, P.D., Mah, Y.H., Teo, J.T., Jäger, H.R., Werring, D., Nachev, P., Ourselin, S., et al.: Unsupervised 3d out-of-distribution detection with latent diffusion models. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 446–456. Springer (2023)
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* (2023)
10. Hore, A., Ziou, D.: Image quality metrics: Psnr vs. ssim. In: *2010 20th international conference on pattern recognition*. pp. 2366–2369. IEEE (2010)
11. Huang, Y., Chen, X., Xu, L., Li, K.: Single image desmoking via attentive generative adversarial network for smoke detection process. *Fire Technology* pp. 1–20 (2021)
12. Li, W., Fan, J., Li, Y., Hao, P., Lin, Y., Fu, T., Ai, D., Song, H., Yang, J.: Endoscopy image enhancement method by generalized imaging defect models based adversarial training. *Physics in Medicine & Biology* **67**(9), 095016 (2022)
13. Macháček, R., Mozaffari, L., Sepasdar, Z., Parasa, S., Halvorsen, P., Riegler, M.A., Thambawita, V.: Mask-conditioned latent diffusion for generating gastrointestinal polyp images. *arXiv preprint arXiv:2304.05233* (2023)
14. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing* **21**(12), 4695–4708 (2012)

15. 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)
16. Qin, X., Wang, Z., Bai, Y., Xie, X., Jia, H.: Ffa-net: Feature fusion attention network for single image dehazing. In: *Proceedings of the AAAI conference on artificial intelligence*. vol. 34, pp. 11908–11915 (2020)
17. Salazar-Colores, S., Jiménez, H.M., Ortiz-Echeverri, C.J., Flores, G.: Desmoking laparoscopy surgery images using an image-to-image translation guided by an embedded dark channel. *IEEE Access* **8**, 208898–208909 (2020)
18. Setiadi, D.R.I.M.: Psnr vs ssim: imperceptibility quality assessment for image steganography. *Multimedia Tools and Applications* **80**(6), 8423–8444 (2021)
19. Sidorov, O., Wang, C., Cheikh, F.A.: Generative smoke removal. In: *Machine Learning for Health Workshop*. pp. 81–92. PMLR (2020)
20. Twinanda, A.P., Shehata, S., Mutter, D., Marescaux, J., De Mathelin, M., Padoy, N.: Endonet: a deep architecture for recognition tasks on laparoscopic videos. *IEEE transactions on medical imaging* **36**(1), 86–97 (2016)
21. Wang, F., Sun, X., Li, J.: Surgical smoke removal via residual swin transformer network. *International Journal of Computer Assisted Radiology and Surgery* pp. 1–11 (2023)
22. Wolleb, J., Bieder, F., Sandkühler, R., Cattin, P.C.: Diffusion models for medical anomaly detection. In: *International Conference on Medical image computing and computer-assisted intervention*. pp. 35–45. Springer (2022)
23. Yan, J., Li, J., Fu, X.: No-reference quality assessment of contrast-distorted images using contrast enhancement. *arXiv preprint arXiv:1904.08879* (2019)
24. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 14821–14831 (2021)
25. 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* (2022)