

DeSmoke-LAP: Improved Unpaired Image-to-Image Translation for Desmoking in Laproscopic Surgery

Yirou Pan^{1*}, Sophia Bano¹, Vasconcelos Francisco¹, Hyun Park², Taikyeong Ted. Jeong³ and Danail Stoyanov¹

¹Wellcome/EPSRC Centre for Interventional & Surgical Sciences, Department of Computer Science, University College London, UK.

²Department of Obstetrics and Gynecology, CHA Bundang Medical Center, CHA University, Seongnam, South Korea.

³School of Artificial Intelligence Convergence, Hallym University, Chuncheon, South Korea.

*Corresponding author(s). E-mail(s): yirou.pan.20@ucl.ac.uk;

Abstract

Purpose: Robotic-assisted laparoscopic surgery has become the trend in medicine thanks to its convenience and lower risk of infection against traditional open surgery. However, the visibility during these procedures may severely deteriorate due to the electrocauterisation which generates smoke in the operating cavity. This decreased visibility hinders the procedural time and surgical performance. Recent deep learning-based techniques have shown the potential for smoke and glare removal, but few targets laparoscopic videos. **Method:** We propose DeSmoke-LAP, a new method for removing smoke from real robotic laparoscopic hysterectomy videos. The proposed method is based on the unpaired image-to-image cycle-consistent generative adversarial network in which two novel loss functions, namely, inter-channel discrepancies and dark channel prior, are integrated to facilitate smoke removal while maintaining the true semantics and illumination of the scene. **Results:** DeSmoke-LAP is compared with several state-of-the-art desmoking methods both qualitatively and quantitatively using referenceless image quality metrics on 10 laproscopic hysterectomy videos through 5-fold cross-validation. **Conclusion:** DeSmoke-LAP outperformed existing methods and generated smoke-free images without applying ground truths (paired

2 *DeSmoke-LAP*

images) and the atmospheric scattering model. This shows distinctive achievement in dehazing in surgery, even in scenarios with partial inhomogenous smoke. Our code and hysterectomy dataset will be made publicly available at <https://github.com/yiroup20/DeSmoke-LAP>.

Keywords: De-smoking, Robotic-assisted laparoscopic hysterectomy, Deep learning, Generative adversarial network

1 Introduction

In laparoscopic surgery, the risks of bleeding can be reduced using instruments with electrocauterisation capabilities, in which a heating source is directly applied to tissue during or after dissection. Such electric instruments have been adapted to robotic-assisted surgery platforms such as the da Vinci Xi in the context of e.g. performing cholecystectomy and hysterectomy. One of the challenges in electrocauterisation is the production of smoke that hinders the visibility of the operative site through the laparoscopic camera. This may require the surgeon to stop any action and remove the instruments from their working channels until visibility is re-established. As demonstrated in [1], this tends to increase the operation time and leads to an increase in surgeon's anxiety. Figure 1 shows sample laparoscopic images with clear view (no smoke) and light, medium and high density of smoke.

A substantial number of computer vision techniques have been proposed in the past to restore visibility in hazy images. These include traditional computer vision methods, Generative Adversarial Networks (GANs) for paired image-to-image translation and cycle-consistent generative adversarial networks (cycleGANs) for unpaired image-to-image translation. Traditional methods use neural networks [2] or variational interference [3, 4] for image desmoking, whose generator is simply updated according to the provided database, leading to an unstable training model. On the other hand, GAN model updates the generator by the backpropagation from discriminator, which helps in obtaining more reliable results. Paired image-to-image translation GANs [5] require the same images with and without ground-truth hazy conditions during training, and thus rely on synthetic training data. In contrast, unpaired image-to-image translation GANs [6, 7] can be trained simply on arbitrary examples of clear and hazy images, without the need of ground-truths or generative physical models, thus offering more flexibility in terms of training data.

Several physical models, including atmospheric scattering model and dark channel prior, have been utilised to model smoke parameters efficiently [2, 5, 8, 9]. The purpose of the atmospheric scattering model is to simulate the smoke component by relating global atmospheric light to the transmission map [2, 8]. According to the characteristic of generated smoke, it can not be distributed uniformly, and can not be simply computed by the scattering model. The dark channel prior is shown to model haze [5], but no attempts have been done to investigate its use within a loss function to train a dehazing model using

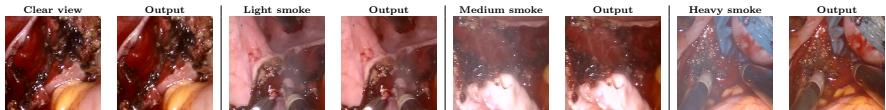


Fig. 1 Sample laparoscopic frames showing different grades of smoke in input and desmoked output from the proposed DeSmoke-LAP method.

unpaired real data. Due to the difficulty in obtaining paired images for real data, quantitative evaluation of these methods mostly rely on synthetic data.

In this paper, we propose DeSmoke-LAP, a dehazing technique to improve the visibility of laparoscopic scenes during electrocauterisation. We use an unpaired surgical image dehazing network which is based on CycleGAN [6]. Our proposed method enhances the CycleGAN network by introducing the inter-channel discrepancies and the dark channel prior as part of the loss function during network training. These losses help in modelling the different smoke components and lead to smoke-free images with visually higher quality. We created a dataset of clear view and hazy images from 10 laparoscopic hysterectomy videos and use cross-validation for the performance evaluation. Additionally, we perform validation on continuous clips, containing varying smoke density, from each video to assess real operation scenarios. In real-data, paired clear and hazy images are not available, therefore, we propose to utilise three existing referenceless metrics for the performance evaluation. Through both quantitative and qualitative comparative analysis with the existing methods, we show that our proposed method achieves better performance. The main contributions of this paper are as follows:

- We develop enhanced CycleGAN which focuses on smoke removal in laparoscopic surgery using unpaired image-to-image translation, without utilising atmospheric scattering models or ground-truths during model training.
- We introduce additional loss functions on inter-channel discrepancies and dark channel prior that allows qualifying remaining smoke component in the generated image, aiding cycle-consistency loss and adversarial loss.
- We introduce the use of referenceless image quality metrics for evaluation which are designed to measure image quality in the absence of ground truth.
- The utilised dataset that includes 6000 clear and hazy images extracted from 10 laparoscopic videos and 10 video clips will be made publicly available¹, providing a benchmark for unpaired laparoscopic image desmoking.

2 Proposed Method

The proposed DeSmoke-LAP model is designed for unpaired image-to-image translation in two domains based on the architecture of CycleGAN, where two additional loss functions are designed for inter-channel differences and dark channel prior (as shown in Fig. 2). These loss functions, added for discriminating, aim to capture the remaining smoke covered on the generated image and promote the optimisation of the generator in the next iteration.

¹Code and visual comparison on video clips: <https://github.com/yirou20/DeSmoke-LAP>

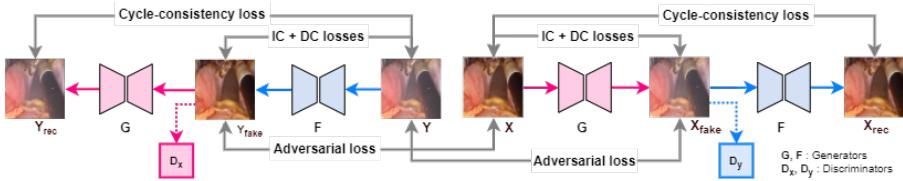


Fig. 2 An overview of the proposed DeSmoke-LAP method. The Inter-Channel (IC) discrepancies and Dark Channel (DC) prior are introduced to qualify the remaining smoke, aiding cycle-consistency and adversarial losses in smoke removal.

2.1 Cycle-Consistent Generative Adversarial Network

CycleGAN architecture forms the backbone of our proposed method, which is an improved GAN [10] that uses adversarial and cycle-consistency losses for unpaired image-to-image translation from source X to target Y domains. GAN is composed of generator and discriminator, where the purpose of the generator is to synthesise examples realistic enough to fool the discriminator, while the discriminator aims to correctly distinguish between real and synthetic generated images. The weights of these two models are updated dynamically to achieve a stabilised balance. Given unpaired clear (smoke-free) images $\{x_i\}_{i=1}^N$ where $x_i \in X$ and hazy images $\{y_j\}_{j=1}^M$ where $y_j \in Y$, the goal is to learn the mapping between X and Y . Two discriminators are implemented in the network, where D_X is applied to distinguish between clear images X and translated data from hazy images $F(Y)$ and D_Y distinguishes hazy images Y and translated data from clear image $F(X)$ (Fig. 2). The adversarial loss is introduced to measure the deviation between the translated image from one domain and the real sample in the other domain. It is applied to both generator and discriminator, where the discriminator aims to maximise the loss and the generator aims to minimise it.

$$\mathcal{L}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{data}(x)}[\log(1 - D_Y(G(x)))] \quad (1)$$

$$\mathcal{L}(F, D_X, X, Y) = \mathbb{E}_{x \sim p_{data}(x)}[\log D_X(x)] + \mathbb{E}_{y \sim p_{data}(y)}[\log(1 - D_X(F(y)))] \quad (2)$$

where $x \sim p_{data}(x)$ and $y \sim p_{data}(y)$ denote the data distributions of the two domains. The objective of generative adversarial loss is summarised as:

$$\mathcal{L}_{GAN}(G, F, D_X, D_Y) = \min_G \max_{D_Y} \mathcal{L}(G, D_Y, X, Y) + \min_F \max_{D_X} \mathcal{L}(F, D_X, Y, X) \quad (3)$$

The cycle consistency loss is evaluated to improve the functionality of generators, which aims to assess the difference between the real data in one domain and data that translated forward and back to the origin domain. The cycle consistency loss is used to judge the recovery with two translations, forward cycle consistency: $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ and backward cycle consistency: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$.

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)}[\|x - F(G(x))\|_1] + \mathbb{E}_{y \sim p_{data}(y)}[\|y - G(F(y))\|_1] \quad (4)$$

2.2 DeSmoke-LAP: Desmoking in Laparoscopic Surgery

CycleGAN alone cannot eliminate smoke from laparoscopic video frames since it does not learn to optimise the model using priors specific to the smoke.

Therefore, we propose DeSmoke-LAP, a desmoking approach for laparoscopic surgery that targets hazy-to-clear translation by introducing two additional loss functions, namely, inter-channel loss and dark channel loss, to the discriminator of each domain. These losses allow measuring the remaining smoke components in the generated image by evaluating the differences between images before and after processing them through the generator.

2.2.1 Inter-channel (IC) Loss

Inter-channel discrepancies [11, 12] describe the difference between any two channels of a pixel in the image. The absolute norm between any two channels in the pixel of the image is given by,

$$\Psi(P) = \|P_R - P_G\|_1 + \|P_G - P_B\|_1 + \|P_B - P_R\|_1, \quad (5)$$

where P denotes a pixel in the image, and P_R , P_G and P_D respectively represent the R, G, D channel of the pixel. The value of channels in the pixel is normalised between 0 and 1 in the calculation. Thus, the loss of an image can be measured by the mean value of norms for all pixels in that image.

$$I^{IC}(X) = \frac{1}{n} \sum_{i=1}^n \Psi(P_i), \quad (6)$$

where x corresponds to the selected image and n represents the total amount of pixels in that image, $P_{i \dots n} \in X$.

According to the observations of He et al. [9], the inter-channel difference of a pixel in equation 6 relates to the level of blur and a small value is obtained when there is heavy smoke covered on that pixel. The difference will be reflected in the discriminator and contribute to the development of the generator. Based on the analysis on the smoky section of images, a pixel with strong colour results in a value close to 1, otherwise, the value is close to 0. Therefore, 1 is considered as an approximate boundary in the calculation to ensure that the function results in large impacts to the generator if the divergence is small. Our network works between the clear and hazy domains, when performing hazy-to-clear translation, it is intended to generate a fake image with less smoke, hence, the inter-channel loss focuses on the hazy components. If the target of the translation is a hazy image, the corresponding discriminator is developed by the smoke-free sections. The inter-channel loss used in the network is given by,

$$\mathcal{L}_{IC} = f(I^{IC}(G(x))) + f(I^{IC}(F(y))), \quad (7)$$

where,

$$f(x) = \begin{cases} 1 - \|x - 1\|, & \text{clean} \rightarrow \text{hazy} \\ \|x - 1\|, & \text{hazy} \rightarrow \text{clean}. \end{cases} \quad (8)$$

2.2.2 Dark Channel (DC) Loss

Inspired by [5, 9, 13], we assess hazy components in the image using the dark channel prior, which measures the intensity of the image and reveals its luminous density. This is defined as the minimum value in the RGB channels,

$$I^{dark}(x) = \min_{y \in \Omega(x)} (\min_{c \in r,g,b} I^c(y)), \quad (9)$$

6 DeSmoke-LAP

where I^c is a colour channel of the arbitrary image I and $\Omega(x)$ is a local patch centred at x . If the image is smoke-free, $I^{dark} \rightarrow 0$. We observe that most of the pixels in a clean laparoscopic image have a low-density value, but few smoke-free pixels still output high value due to the luminance effect and light reflection. To fairly and reasonably access the dark channel of the sample input, we calculate its average density, preventing the influence from pixels with extra-high or extra-low value. Thus, the dark channel loss of an image is measured as the average value by looping through all pixels. It is also suggested to implement a step to obtain the refinement stage for the dark channel by applying the soft matting algorithm [9], thus the edge and profile of objects can be captured and highlighted maintaining more details in the image. This refinement algorithm is executed in our model as well. The DC loss is respectively added to the two discriminators, and if most of the sample is covered by smoke, the loss will be large, promoting the parameter optimisation of the generator.

$$\mathcal{L}_{DC} = \overline{I_{dark}(F(y))} + \overline{I_{dark}(G(x))} \quad (10)$$

2.2.3 Combined Loss Function

The full objective of loss function for the proposed Desmoke-LAP is given by,

$$\mathcal{L}(G, F, D_x, D_y) = \mathcal{L}_{CycleGAN}(G, F, D_x, D_y) + \mathcal{L}_{IC}(D_x, D_y) + \beta \cdot \mathcal{L}_{DC}(D_x, D_y), \quad (11)$$

where,

$$\mathcal{L}_{CycleGAN}(G, F, D_x, D_y) = \mathcal{L}_{GAN}(G, F, D_X, D_Y) + \mathcal{L}_{cyc}(G, F). \quad (12)$$

To regulate the DC loss, β is added in Eq. 11 where its value is selected through experimentation (refer to Sec. 4.2). G and F respectively stands for generators for domain X and Y , and D_x and D_y are two discriminators in two domains.

3 Referenceless Evaluation Metrics

Since all our data is unpaired and collected from real robotic surgery, ground-truth (paired clear and hazy) images are not available. Therefore, metrics commonly used in haze and smoke removal evaluation such as mean squared error (MSE), peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), etc, are not applicable. We rely on several referenceless image quality metrics for evaluating the performance of the resulting desmoke images. Three referenceless metrics are used to evaluate reconstructed images based on their own characteristics correlated to fog density, image blurriness and edge restoration, without taking ground-truths as reference. These metrics are briefly explained below:

Fog Aware Density Evaluator (FADE) [14] is used to compute the fog density of the image, where a higher FADE value means there is more fog covered on the image, resulting in blurry output. It is constructed in accordance with natural scene statistics (NSS) and fog aware statistical features.

Just Noticeable Blur Metric (JNBM) [15] measures the perceptual sharpness of the image, where a lower value results from low sharpness. It

focuses on the behaviour of the human visual system to sharpness at different contrast levels and accesses the blurriness of edges in the image.

Restoring Edge Assessment (REA) [16] assesses the edge restoration of the image, which differentiates between the original image and the image reconstructed after smoke removal. A higher value of REA indicates better restoration of the edges.

4 Dataset and Experimental Setup

4.1 Data Organisation

We collected 10 robot-assisted laparoscopic hysterectomy procedure recordings. Active instrument labels at the bottom of the video-feed assisted in manually annotating hazy and clear images. These videos were decomposed into frames at 1 fps. 300 clear and 300 hazy images per video were selected to form our dataset. In total, 3000 clear and 3000 hazy images were selected from 10 sampled videos, where the images were cropped to remove video display and resized to 720×540 pixel resolution while maintaining the aspect ratio. The organised data contain both inter-patient and intra-patient variabilities in the scene, adding diversity to the dataset. Intra-patient variability is experienced due to the movement of the camera in the surgical operating field. The collected images contain various levels of haze that were split into light, medium, heavy and partial smoke based on inhomogeneity in the scene. Moreover, a sequence of 50 frames is selected as a short clip from each video, such that these clips also capture frames with motion blur. These clips are used as part of the testing data to analyse consistency of desmoking algorithms across frames. The dataset summary is provided in the supplementary material Sec. 1.

4.2 Training Details

For all experiments, models are trained using an Nvidia 16GB V100 GPU and batch size of 4. The DeSmoke-LAP utilises a ResNet generator and a PatchGAN discriminator along with a least-squares GANs objective for each network, following the implementation by Zhu et al. [6]. The learning rate is set to 0.002 for the first 50 iterations, and linearly decay to 0 in the latter 50 epochs. To test and verify the superiority of the combined losses, the model is also trained with DC and IC losses independently. To control the effect of the dark channel prior efficiently, testing was completed on one fold with various values of β . The model trained with a larger β gave a lower FADE value indicated in Fig. 3. The median FADE value is lowest at $\beta = 0.05$, leading to outputs with less smoke. Thus, β in equation 11 is set to 0.05 for evaluation.

To investigate the performance of our proposed model, 5-fold cross-validation is used, with each fold containing image samples from 2 videos. A sequence of 50 continuous frames from each test video is also used to evaluate

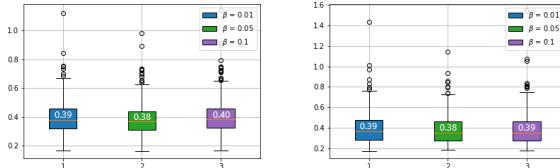


Fig. 3 FADE value of clear (left) and hazy (right) images in Fold 3 with various β .

the trained network. Data augmentation is utilised to enhance the generalisation capabilities of the proposed network. The data is cropped at random positions to 256×256 resolution, creating more patches before training.

We perform quantitative and qualitative comparisons of DeSmoke-LAP with CycleGAN [6], FastCUT [17], Cycle-Dehaze [7] and Colores et al. [5] methods. FastCUT [17] improves over CycleGAN by providing a faster training network for image translation, utilising the advantages of contrastive learning. Cycle-Dehaze [7] is an enhanced CycleGAN for image dehazing that employs cyclic perceptual-consistency loss to maintain the original structure of the image. Colores et al. [5] fused the dark channel prior with inputs before passing it to the generator for learning paired image-to-image translation. Experiments were performed with hazy images synthesised by adding smoke to the input image. Since our data is unpaired, retraining of this model is not possible. Therefore, we use the pre-trained Colores et al. [5] model on our dataset.

5 Results and Discussion

Table 1 presents the quantitative comparison of the proposed DeSmoke-LAP (IC+DC) with the state-of-the art models using 5-fold cross-validation reporting average FADE, JNBM and REA metrics and their standard deviation over all folds. The results over each fold are provided in the supplementary material Sec. 2. DeSmoke-LAP with either IC or DC loss results are also reported. Among the three variants of the DeSmoke-LAP models, the model with both IC and DC losses outperformed on all metrics, whereas the performance of the model with only one loss was attenuated on either haze or contrast levels. Focusing on quantitative results, DeSmoke-LAP performance was marginally lower than Colores et al. [5] and Cycle-Dehaze [7], though DeSmoke-LAP outperformed other traditional unpaired image-to-image translation methods [6, 17]. However, when visually analysing the methods under comparison (shown in Fig. 1-2 in Supplementary), we observe that though JNBM value of Colores et al. and Cycle-Dehaze and DeSmoke-LAP are comparable, DeSmoke-LAP removed smoke while retaining scene semantic, i.e. without overexposing and attenuating the image intensity. Besides, we observe that all referenceless metrics follow the same trend.

We also perform testing on short video clips extracted from each video. Evaluation results for clips per fold are reported in Supplementary material Sec. 3 and average results over all folds and clips are presented in Table 2. Similar observations as mentioned above are drawn when testing video clips.

Table 1 Quantitative comparison through 5-fold cross-validation on the organised clear and hazy images dataset. Mean and standard deviation of the 3 metrics are reported. Lower FADE, and higher JNBM and REA values are better.

	FADE		JNBM		REA	
	clear	hazy	clear	hazy	clear	hazy
Input	0.41±0.14	0.85±0.53	1.71±0.24	1.13±0.30	0.00	0.00
CycleGAN [6]	0.42±0.12	0.43±0.14	1.01±0.23	1.11±0.23	1.12±0.20	1.33±0.37
FastCUT [17]	0.61±0.22	0.81±0.21	1.19±0.27	1.11±0.25	2.00±0.58	2.34±0.55
Colores et al. [5]	0.31±0.08	0.40±0.12	1.27±0.27	1.19±0.26	2.57±0.55	3.09±0.56
Cycle-Dehaze [7]	0.28±0.20	0.29±0.08	2.10±0.30	2.11±0.22	1.65±0.42	1.77±0.55
DeSmoke-LAP (IC)	0.42±0.15	0.41±0.17	0.97±0.20	1.00±0.27	1.02±0.15	1.37±0.33
DeSmoke-LAP (DC)	0.41±0.15	0.41±0.15	0.99±0.25	1.11±0.24	1.06±0.12	1.38±0.31
DeSmoke-LAP (IC+DC)	0.41±0.14	0.41±0.14	1.10±0.20	1.13±0.26	1.09±0.20	1.41±0.30

Table 2 Comparative analysis using the video clips' dataset, reporting mean standard deviation of the 3 metrics. Lower FADE, and higher JNBM and REA values are better.

	FADE	JNBM	REA
Input	0.95±0.50	2.80±1.09	0.00
CycleGAN [6]	0.40±0.2	1.05±0.15	1.37±0.35
FastCUT [17]	0.59±0.27	1.11±0.20	1.13±0.18
Colores et al. [5]	0.36±0.09	1.18±0.20	5.60±1.75
Cycle-Dehaze [7]	0.28±0.05	2.03±0.19	1.60±0.34
DeSmoke-LAP (IC)	0.38±0.77	1.00±0.19	1.70±1.17
DeSmoke-LAP (DC)	0.37±0.83	1.07±0.18	1.74±1.00
DeSmoke-LAP (IC+DC)	0.36±0.79	1.08±0.15	1.79±1.23

Colores et al. [5] achieved overall best performance when analysing the referenceless metrics whereas our proposed method can still output fine results compared to other approaches. Focusing on FADE value, it shows that our method, Cycle-Dehaze and Colores's method perform well in both clear and hazy classes. Delving into JNBM and REA metrics, our method falls behind Colores's method, meaning that the testing outputs were of low sharpness and the performance of edge restoration is not outstanding. Since referenceless metrics are designed from natural images which are largely different from surgical images, these may not be true indicative of surgical images quality.

We further investigated the reliability and quality of desmoked images through qualitative comparison using the short video clips (Fig. 4). Results on complete video clips are provided on our Github¹ and in the supplementary video. To review the results meaningfully, we divided the testing images into three main groups on the basis of their density of smoke, which includes light, medium and heavy. Three samples were picked from each group for visual contrast in each fold (as shown in Supplementary Sec. 4). We observe the density of smoke in the recovered image and colour variation between the input and output and also measure the JNBM of individual images. Colour differences could affect the performance of surgery and the decision made by surgeons. Moreover, we also considered the reliability and harmony of the recovered image, meaning the synthetic data must look like real data.

All methods showed positive effects on desmoking in surgical images, except FastCUT [17] which failed to completely eliminate the smoke from light-hazy and partially hazy images (Fig. 4 and supplementary video¹). Cycle-Dehaze [7]

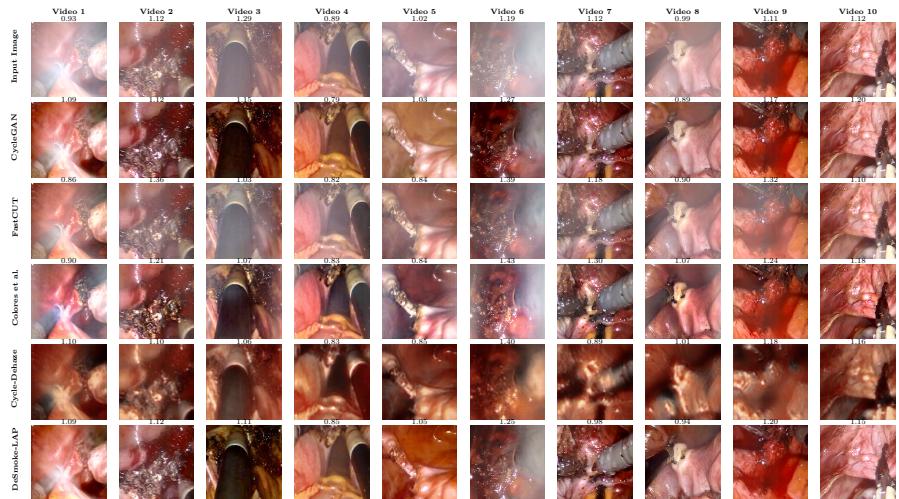


Fig. 4 Qualitative comparison of the DeSmoke-LAP with the existing approaches using representative frames from 10 video clips where JNBM value of each image is also displayed.

works well on dehazing but produced low-quality outputs that have negative effect on visualisation. The proposed DeSmoke-LAP visually outperformed Colores et al. [5] and especially took the lead in better optimisation of dark pixels where most smoke was detected and removed (see suppl. video). In terms of JNBM, Colores et al. appeared to be better or comparable with the DeSmoke-LAP because sometimes Colores et al. generate sharper images which are clipped at lower intensities, leading to information loss due to visually attenuated and less accurate desmoked images. Referenceless metrics fails in evaluating these details which are aesthetically not appealing during visualisation.

The proposed DeSmoke-LAP showed its strength in dealing with images covered with smoke partially. Referenceless metrics fail to describe detailed desmoked information of the image but give a brief summary of the performance, thus we have to rely on the visual evaluation whereas quantitative results act as the assistance. When Colores et al.'s method produced excellent quantitative outputs, images may not be dehazed uniformly, which suggests that only some parts of smoke are removed perfectly. Outputs by Cycle-Dehaze failed to meet the demand of high-quality and vivid vision needed for laparoscopic surgery. DeSmoke-LAP guarantees the smoke is processed in accordance with its blur level, and the coordination of the colour is not be affected dramatically. Future work involves retaining the original resolution of the laparoscopic video to obtain high-quality desmoked images that would be beneficial for clinical use and use of larger dataset for further improving the method's robustness.

6 Conclusion

Compared to traditional open surgery, laparoscopic robot-assisted surgery manages the operation through tiny incisions by robot arms, finding a wide application in medicine. However, smoke generated due to electrocauterisation

during laparoscopic surgery has been a potential risk to patients and surgeons. To address this issue, we proposed DeSmoke-LAP, a method for virtually removing smoke in laproscopic surgery for enhancing intraoperative imaging. DeSmoke-LAP performed unpaired image-to-image translation between hazy and clear images based on cycle-consistency generative adversarial network. Unlike existing image dehazing methods, DeSmoke-LAP does not rely on paired data with the smoke generated synthetically and atmospheric scattering model. Instead, we introduced two additional losses in the discriminator that assist to estimate the remaining smoke in the generated image by inter-channel discrepancies and dark channel prior. We quantitatively and qualitatively compared DeSmoke-LAP with the state-of-the-art image translation, dehazing and desmoking methods through 5-fold cross-validation. The trained models were also tested on video clips, and we observed that desmoked frames by DeSmoke-LAP appeared to be consistent and smooth throughout the duration of the clip, outperforming other methods. DeSmoke-LAP generated better quality, colour and contrast outputs without any clipping or attenuation, leading to visually meaningful desmoked results. The dataset and code will be made publicly available, providing a benchmark for desmoking in laproscopic surgery.

Declarations

Conflict of Interest The authors declare they have no conflict of interest.

Code availability Code to be released with this paper.

Ethics approval For this type of study, formal consent is not required.

Informed consent This articles does not contain patient data.

References

- [1] Gu, L., Liu, P., Jiang, C., Luo, M., Xu, Q.: Virtual digital defogging technology improves laparoscopic imaging quality. *Surgical innovation* **22**(2), 171–176 (2015)
- [2] Morales, P., Klinghoffer, T., Jae Lee, S.: Feature forwarding for efficient single image dehazing. In: IEEE Computer Vision and Pattern Recognition Workshops, pp. 0–0 (2019)
- [3] Cheng, X., Yang, B., Liu, G., Olofsson, T., Li, H.: A variational approach to atmospheric visibility estimation in the weather of fog and haze. *Sustainable cities and society* **39**, 215–224 (2018)
- [4] Zhou, H., Xiong, H., Li, C., Jiang, W., Lu, K., Chen, N., Liu, Y.: Single image dehazing based on weighted variational regularized model. *IEICE Transactions on Information and Systems* **104**(7), 961–969 (2021)
- [5] Salazar-Colores, S., Jiménez, H.M., Ortiz-Echeverri, C.J., Flores, G.: Desmoking laparoscopy surgery images using image-to-image translation

- guided by embedded dark channel. *IEEE Access* **8**, 208898–208909 (2020)
- [6] Zhu, J.-Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Computer Vision (ICCV), 2017 IEEE International Conference (2017)
 - [7] Engin, D., Genç, A., Kemal Ekenel, H.: Cycle-dehaze: Enhanced cycle-gan for single image dehazing. In: IEEE Computer Vision and Pattern Recognition Workshops, pp. 825–833 (2018)
 - [8] Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: IEEE Computer Vision and Pattern Recognition, pp. 3194–3203 (2018)
 - [9] 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)
 - [10] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. *Advances in neural information processing systems* **27** (2014)
 - [11] Kang, M., Jung, M.: A single image dehazing model using total variation and inter-channel correlation. *Multidimensional Systems and Signal Processing* **31**(2), 431–464 (2020)
 - [12] Wang, C., Cheikh, F.A., Kaaniche, M., Elle, O.J.: A smoke removal method for laparoscopic images. arXiv preprint arXiv:1803.08410 (2018)
 - [13] Peng, Y.-T., Cao, K., Cosman, P.C.: Generalization of the dark channel prior for single image restoration. *IEEE Transactions on Image Processing* **27**(6), 2856–2868 (2018)
 - [14] Choi, L.K., You, J., Bovik, A.C.: Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE Transactions on Image Processing* **24**(11), 3888–3901 (2015)
 - [15] Ferzli, R., Karam, L.J.: A no-reference objective image sharpness metric based on the notion of just noticeable blur (jnb). *IEEE transactions on image processing* **18**(4), 717–728 (2009)
 - [16] Hautière, N., Tarel, J.-P., Aubert, D., Dumont, E.: Blind contrast enhancement assessment by gradient ratioing at visible edges. *Image Analysis & Stereology Journal* **27**(2), 87–95 (2008)
 - [17] Park, T., Efros, A.A., Zhang, R., Zhu, J.-Y.: Contrastive learning for unpaired image-to-image translation. In: European Conference on Computer Vision, pp. 319–345 (2020). Springer