



Unsupervised smoke to desmoked laparoscopic surgery images using contrast driven Cyclic-DesmokeGAN

Vishal Venkatesh ^a, Neeraj Sharma ^b, Vivek Srivastava ^c, Munendra Singh ^{a,*}

^a Department of Mechatronics Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

^b School of Biomedical Engineering, Indian Institute of Technology (Banaras Hindu University), Varanasi 221005, India

^c Department of General Surgery, Institute of Medical Sciences, Banaras Hindu University, Varanasi 221005, India

ARTICLE INFO

Keywords:

GANs
Desmoking
Laparoscopic surgery
Enhancement

ABSTRACT

In laparoscopic surgery, energized dissecting devices and laser ablation causes smoke, which degrades the visual quality of the operative field. This paper proposes an unsupervised approach to desmoke laparoscopic images called Cyclic-DesmokeGAN. In the generator, multi-scale residual blocks help to alleviate the smoke component at multiple scales, while refinement module helps to obtain desmoked images with sharper boundaries. As the presence of smoke degrades contrast and fine structure, the proposed method utilizes high boost filtered image at each encoder layer. The contrast loss improves overall contrast, thereby reducing the smoke, while Unsharp Regularization loss helps to stabilize the network. The proposed Cyclic-DesmokeGAN is tested on 200 smoke images obtained from Cholec80 dataset consisting of videos of cholecystectomy surgeries. The results depict effectiveness, as proposed approach achieved 3.47 ± 0.09 Contrast-Distorted Images Quality, 4.15 ± 0.74 Naturalness Image Quality Evaluator, and 0.23 ± 0.00 Fog Aware Density Evaluator, these indexes are best in comparison to other state-of-the-art methods.

1. Introduction

Laparoscopy is a type of surgery, which is usually performed in the abdomen or pelvis using minor incisions. The name laparoscopy surgery came from the laparoscope that has a tiny camera and light source at its end. This setup helps surgeons to visualize the surgical activity on a monitor. The laparoscopic processors are capable to produce high definition digital video outputs [1]. However, the generation of smoke in laparoscopic surgery reduces surgeon's visibility [2], and it also results in the poor performance of computer vision algorithms for various tasks like depth estimation, semantic segmentation and localization of tissues and the tools. In such a situation, smoke-free visualization of laparoscopic surgery is a prime concern. The smoke extraction using evacuation technique [3] is the popular approach to address the problem of smoke. On the other hand, retrospective smoke removal and enhancement methods for the laparoscopic images/videos do not add additional constraints like hardware and equipment installations, hence are very effective to produce better information of the anatomy of the interest. Previously, methods relevant to defogging and dehazing have been applied for the task of desmoking the visualization but due to the assumptions underlying the problem, optimal results have not been achieved. Fig. 1 shows the overview of laparoscopic surgery and the effect of smoke on the visualization.

Further, the artefacts are also generated during laparoscopic surgery which includes specular reflections, illumination changes, blood and surgical smoke. One of the crucial concerns while developing algorithms for desmoking is its ability not to lose colour contrast, structural information and can be embedded in the computer vision pipeline of the image-guided surgical systems.

In the present paper, the problem of desmoking focuses on removing the smoke component and this should result in the enhancement of contrast and fine edges in the image. As the paired real-world image data is not available for the considered task, drawing the inspiration from Generative Adversarial Networks (GANs) in Image-to-Image Translation tasks, the present study formulates the single image desmoking problem as a translation from smoke images domain to smoke-free images domain. This paper presents an unsupervised learning approach, which directly works on the unpaired laparoscopic images, and hence, the proposed approach inherently removes the issue of simulation to real-world domain adaptation. The study addresses the following challenges: recovering structure, details and boundaries, which has been degraded due to the smoke. The study also helps to realize stability in the network as it enhances the reliability and safety of the method for the task of image synthesis in medical imaging. In summary, the contributions of the paper are as follows:

* Corresponding author.

E-mail address: munendra107@gmail.com (M. Singh).

<https://doi.org/10.1016/j.complbiomed.2020.103873>

Received 29 February 2020; Received in revised form 10 June 2020; Accepted 20 June 2020

Available online 24 June 2020

0010-4825/© 2020 Elsevier Ltd. All rights reserved.

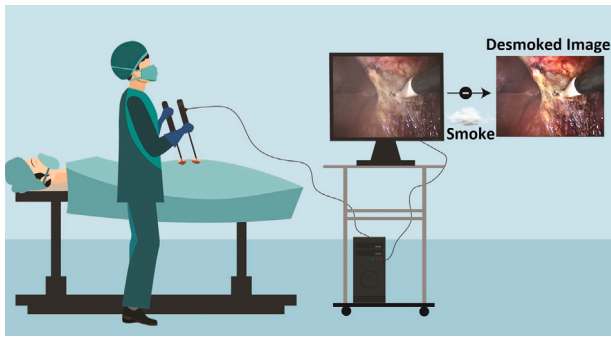


Fig. 1. Semantic of operative field affected by the smoke in laparoscopic procedure.

1. An unsupervised learning approach that builds on a new generator architecture, consisting of multi-scale residual blocks to alleviate the smoke component at multiple scales and the refinement module to refine the output feature map, capture the sharper boundaries.
2. Concatenation of the images from unsharp masking at each encoder layer improves the high-frequency information and fine details in the generated image.
3. In addition to the adversarial and cycle-consistency losses, two new losses are proposed (i) Contrast loss, which reduces the low contrast smoke component and helps to enhance the overall contrast, and (ii) Unsharp Regularization loss, which reduces the oscillations in the generator loss and penalizes the network from divergence.
4. The proposed method is end-to-end trainable and can be directly inculcated as a pre-processing step in the pipeline of the computer vision systems. Further, extensive experimentation on real-world surgical smoke dataset along with the qualitative and quantitative comparison with other state-of-the-art methods for desmoking to depict the capability of the proposed Cyclic-DesmokeGAN.

The organization of the paper is as follows. Section 2 reviews the related work on desmoking methods for laparoscopic images and Deep learning methods like GANs and Image-to-Image Translation. Section 3 examines the details of the proposed method along with its fundamental blocks, architecture designs, and loss functions. Section 4 shows the experimentation of the proposed method on real laparoscopic images and evaluates the results in comparison with other desmoking methods. Section 5 presents a discussion on the findings. Finally, Section 6 presents the conclusion, lacking capabilities and future scope.

2. Related work

Loukas et al. suggested the detection of smoke as the first step of information retrieval and employed the Support Vector Machine (SVM) to classify the laparoscopic frames with and without smoke [4]. Further, the quality of features is one of the important aspect in improving the effectiveness of classification and hence, Jalal et al. proposed ten visual features for smoke detection in laparoscopic images [5]. In [6], weighted least squares based a new enhancement method was proposed, which improved the quality of the image and segregated the smoke and non-smoke laparoscopic images efficiently.

The present study deals with the next step of information retrieval and focuses on the removal of smoke rather than the detection of smoke. Kotwal et al. utilized the probabilistic graphical model that performs joint desmoking and denoising operation while considering the transmission-map and uncorrupted colour image as the prior models [7]. As an extension of this method, an algorithm for specular removal, along with desmoking and denoising has been proposed [8].

Recently, Wang et al. assumed the smoke component has low contrast and low inter-channel differences and devised a variational method to solve the problem of desmoking [9]. This method has depicted the results having the enhancement in the image, however, it is based on certain assumptions such as smoke being a local phenomenon of varying depth and heterogeneous at times.

Researchers have viewed desmoking problem similar to dehazing and defogging. In [10], K. Tchaka et al. has utilized Dark Channel Prior (DCP) approach [11], which was originally used for dehazing and suggested a refined DCP in combination with histogram equalization to reduce the smoke from laparoscopic images. The DCP based method processes the patches and assumed constant air-light throughout the patch, to address this issue a per-pixel computation based method has been proposed, which also minimized the artefact using the multi-scale approach [12]. The morphological reconstruction along with DCP has been proposed to reduce the environmental pollution effects [13]. This technique is computationally less expensive in comparison to the original DCP. Recently, a robust dehazing technique has been proposed, which calculate transmission values using colour ellipsoid geometry and does not require any post processing [14]. Further, Luo et al. proposed a visibility-driven fusion defogging framework that aims at enhancing the contrast and at the same time it recovers the luminance of the surgical field [15]. In [16] a defogging method is proposed to enhance the maritime surveillance images, and utilized illumination decomposition, which removes the glow effect of the environment.

The recent advancements in computer vision due to deep convolutional neural networks (CNN) has produced impressive results for the task of dehazing [17], and de-raining [18]. Similar deep learning based approaches have been utilized for the task of smoke removal in laparoscopic images as well. In [19], Sabri et al. utilized the synthetic surgical smoke images dataset consisting of varying smoke densities and applied convolutional neural networks to remove smoke in a supervised setting. In [20], a preliminary study was conducted, which was the enhancement of the CycleGAN framework [21] for the task of desmoking and yielded good results but critical problems like mode collapse, diminishing gradients problems that arise in GANs were not discussed.

There is unprecedented growth in the number of innovative model architectures, frameworks and objective/loss functions. The advent of the Generative Adversarial Networks (GANs) by Goodfellow et al. in 2014 [22] led to the drastic improvement in the synthesis of images in recent times. The fundamental game-theoretic approach between the generator and the discriminator networks enable to learn the distribution of the training images, while the adversarial loss with min-max optimization framework helps to produce images that are sharper, more realistic and visually appealing. The performance of GANs depends on the ability of the generator network to produce fake images in such a way that the discriminator network is not able to distinguish them from the real images. This process enables the framework to produce the results that are visually appealing and realistic. In the view of GANs capability, researchers continue exploring it for the different applications [23–25].

The application of Image-to-Image translation methods has resulted in the achievement of outstanding outcomes [21]. The Image-to-Image Translation method can be divided into two categories, one with paired data and another with unpaired data. A translation function is learned to map an image of a particular domain to another desired domain. Isola et al. proposed the Pix2Pix framework that builds on a conditional generative adversarial network using paired data [26]. Similar approaches can be seen in the case of generation semantic layouts from input images [27], photographs from sketches [28]. These methods rely on the availability of paired data, however, it is difficult to obtain the required paired data for every problem. Hence, an unpaired Image-to-Image translation was proposed that aimed at learning the relationship between two data domains: X and Y . CycleGAN [21] and Unsupervised

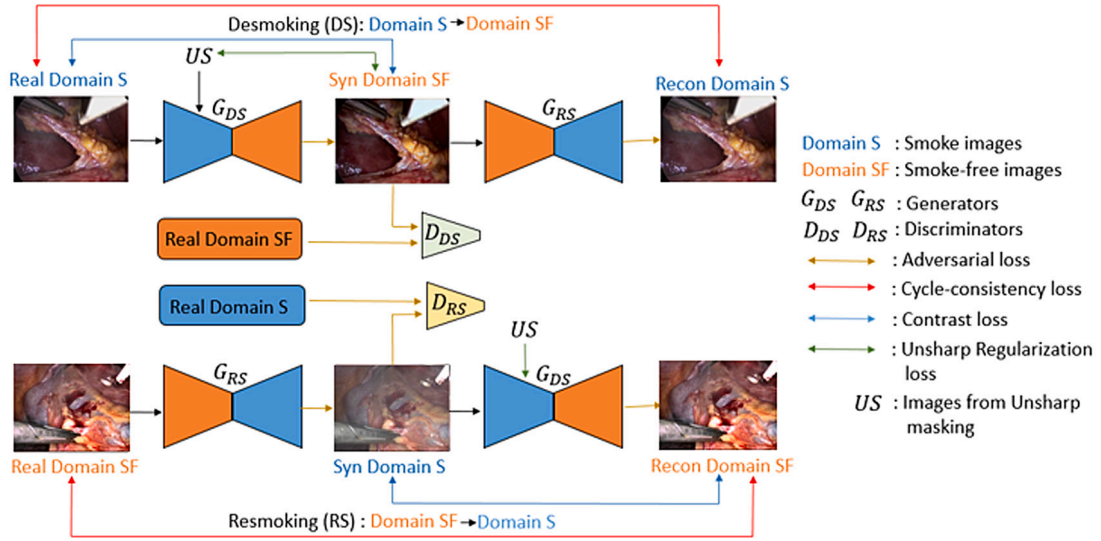


Fig. 2. The framework of the proposed approach consists of two generator networks G_{DS} and G_{RS} , and two discriminator networks D_{DS} and D_{RS} . The mapping function for desmoking (DS) translates an image from the smoke domain (S) to the smoke-free domain (SF), while the resmoking (RS) translates an image from the smoke-free domain (SF) to smoke domain (S). Hence, it is a two-way translation with losses as represented in the figure.

Image-to-image Translation (UNIT) [29] were among the initial methods that depicted impressive results in the unsupervised setting. Since then, various improvements have been observed in these methods and their application to a wide range of problems. In medical imaging, these methods have been appreciated for CT-to-MRI synthesis [30], motion artefact correction [31], and low-dose CT denoising [32].

3. Method

The first part of this section introduces the multi-scale residual blocks and refinement module, while the latter part presents the loss functions for the smoke removal of laparoscopic surgery images. Fig. 2 depicts the framework, which aims to learn two mapping functions. One of the mapping function translates the image from smoke to smoke-free domain, on the other hand, second mapping function translates the image from smoke-free to smoke domain. There are two generator networks: G_{DS} and G_{RS} . Generator G_{DS} generates synthetic desmoked images, while, G_{RS} generates synthetic resmoked images. Further, the two discriminator networks are D_{DS} and D_{RS} . The D_{DS} helps to distinguish between the synthetic desmoked and real smoke-free images, while, D_{RS} helps to distinguish between the synthetic resmoked and real smoke images. The complete training loss consists of four loss functions: an adversarial loss, cycle-consistency loss, contrast loss and unsharp regularization loss. The present paper further describes a new generator architecture along with its fundamental blocks.

3.1. Network architecture

3.1.1. Multi-scale residual blocks

Inspired by the ability of multi-scale feature extraction at tasks like semantic segmentation [33], image super-resolution [34], and image deblurring [35] to obtain features at different scales, the present study adopts the following method for the application of smoke removal task. As smoke can be a local phenomenon or can cover the entire image, it is necessary to have a robust method to capture features that can differentiate the smoke component in the laparoscopy image. In order to capture the features at multiple scales, the present study performs convolution operation with different kernels i.e. 3×3 , 5×5 and 7×7 . The method is pictorially depicted in Fig. 3, where the input feature map is convolved using 3×3 , 5×5 and 7×7 sequentially. The feature maps of each different kernel convolution are then concatenated followed by parallel convolution with the three kernels of different

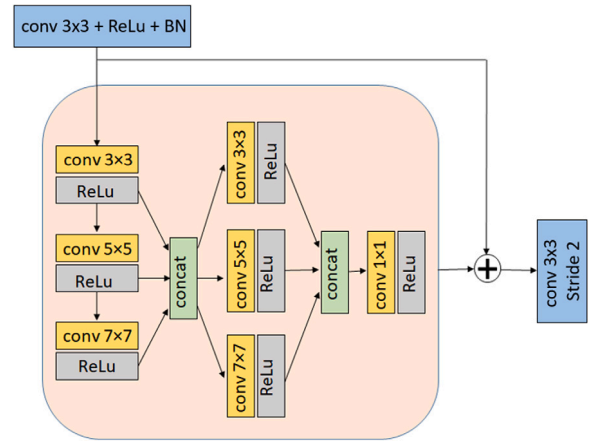


Fig. 3. The pictorial representation of the multi-scale residual block. The approach presented in block performs feature extraction by sequential and parallel set of convolutions using multiple kernel sizes and makes it more efficient by performing residual learning.

sizes. The output of parallel convolutions is concatenated and the channel depth is controlled by a 1×1 convolution. As residual learning has enabled to make networks deeper and efficient, we apply the same technique to the multi-scale feature extraction and this combination results in multi-scale residual blocks.

3.1.2. Refinement module

Wang et al. have shown the application of Dense Upsampling Convolution (DUC) to generate more detailed information for the task of semantic segmentation [36], which was not possible by conventional upsampling methods. Inspired by a similar approach, the present paper proposes a refinement module that helps to obtain sharper boundary details and the better structure. The module takes input feature maps from different encoder layers and the dimension is denoted as $h \times w \times c$, where h is the height, w is the width and c is the channel depth. Based on the downsampling factor (d) of the input feature map with respect to the input image of size $H \times W \times C$ where $h = H/d$ and $w = W/d$, the present study performs a convolution operation to increase the depth of the feature map to obtain the dimension of $h \times w \times (d^2 \times 16)$.

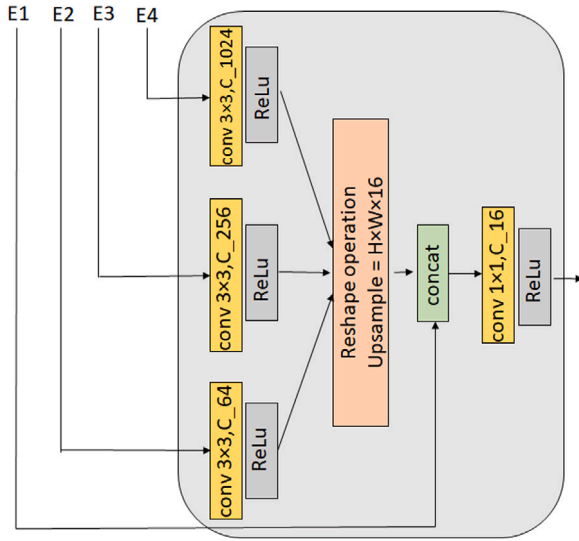


Fig. 4. The refinement module takes in feature maps from various encoder layers and performs convolution and reshaping operations to realize the actual image spatial dimensions. The features maps are then concatenated and convolved with a 1×1 kernel to control the channel dimension. The features responsible to generate sharp boundaries are expected to present in the output feature map of the module. Note, the C_{XX} depict a 3×3 convolution with particular XX filter depths.

Further, a reshaping operation is performed to obtain a feature map of the dimension of $H \times W \times 16$. The output feature maps of different downsampling steps are concatenated and the later channel depth is reduced to 16 by 1×1 convolution as shown in Fig. 4. This feature map is responsible for forcing the network to generate sharper images with better structure.

3.1.3. Generator architecture

Fig. 5 illustrates the architecture of the generator, which follows the encoder–decoder structure. The encoder consists of multi-scale residual blocks and downsampling by four stride-2 convolutions. The bottleneck

comprises of six residual dense blocks (RDB) [37]. On the other hand, the decoder consists of transposed convolutions for upsampling along with concatenation of skip connections, followed by convolution. The output of the last decoder block gets summed with feature map from the refinement module followed by convolution operations to realize the input image dimensions.

The generator G_{DS} produces synthetic desmoked images. In order to enhance the ability of the generator to produce images with better fine details and edge information, we consecutively concatenate different downsampled images obtained from unsharp masking as shown in Fig. 6 with the input feature maps at the respective encoder layers. This is represented as US in the encoder layer in Fig. 5. The generator G_{RS} helps to produce synthetic resmoked images and the encoder layers of the network do not contain the concatenation of images obtained from unsharp masking with input feature maps. The unsharp masking derives the sharp image as follows:

$$g(x, y) = f(x, y) - f_{smooth}(x, y) \quad (1)$$

$$f_{sharp}(x, y) = f(x, y) + k \times g(x, y) \quad (2)$$

where $f(x, y)$ is the input image, $f_{smooth}(x, y)$ is the smoothed image obtained by convolution operation and $g(x, y)$ is the image with high frequency information. This high-frequency information is multiplied by an amount k and added to the original image resulting in $f_{sharp}(x, y)$ that is expected to be sharper and contains fine details and better edge information. The utilization of the sharp images obtained from unsharp masking offers features that are responsible for the improvement in the high-frequency image elements and the overall contrast in the image, hence suppressing the low contrast slow varying smoke component. Each layer of the encoder has been downsampled by a factor of 2 with respect to the previous layer, the dimensions of the sharp images are in consistency with the dimensions of particular encoder layers so that channel-wise concatenation is possible. The kernel dimensions used for the generation of the sharp images are 24, 12, 6 and 3 with different downsampled images of 1, 1/2, 1/4, 1/8 dimension values with respect to the original image. The suitable value chosen for k is 1.25 based on experimental analysis, please refer to Table S3(b) in the SUPPLEMENTARY MATERIAL: A for the analysis.

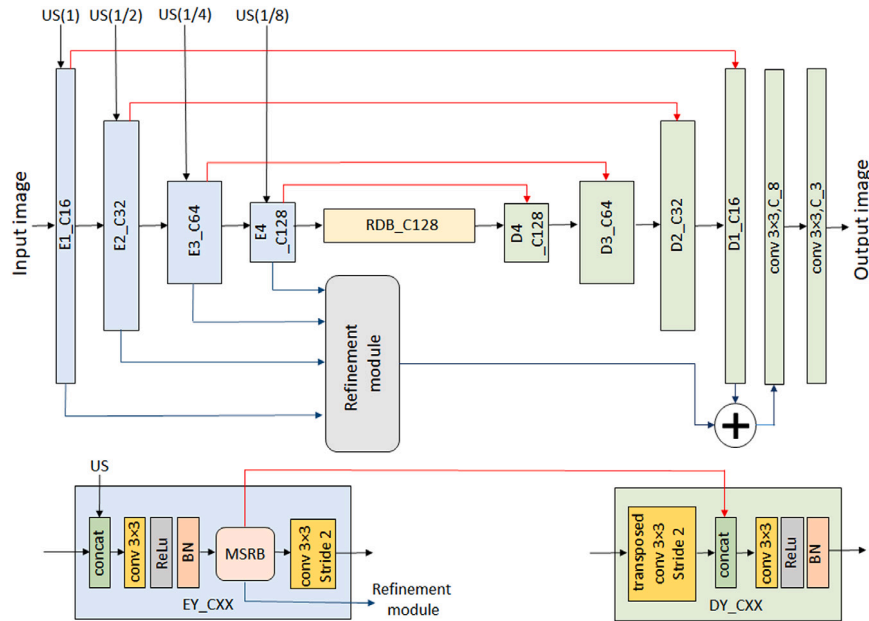


Fig. 5. The generator architecture consists of multi-scale residual blocks at each encoder layer, a bottleneck consisting of six residual dense blocks, decoder layer consisting of transposed convolutions and concatenations with skip connections, refinement module and few convolutional layers to generate the image of the input dimensions. The generator G_{DS} performs concatenation of the images obtained from unsharp masking with the input feature map, whereas G_{RS} does not perform this operation.

Table 1

Experimental results for different number of multi-scale residual blocks (MSRB) in each encoder layer. (High Value of CEIQ, Less Value of FADE, and Less Value of NIQE indicates Better Image Quality). The metric values denote the mean and standard deviation obtained from k -fold cross-validation.

| No of blocks | CEIQ | NIQE | FADE |
|--------------|-----------------------------------|-----------------------------------|-----------------------------------|
| 3 | 3.41 ± 0.09 | 5.22 ± 0.62 | 0.25 ± 0.04 |
| 2 | 3.37 ± 0.1 | 4.64 ± 0.66 | 0.25 ± 0.07 |
| 1 | 3.44 ± 0.06 | 4.17 ± 0.89 | 0.23 ± 0.09 |
| w/o MSRB | 3.40 ± 0.08 | 4.21 ± 0.74 | 0.30 ± 0.06 |

Table 2

Experimental results for different kernel sizes used in the refinement module.

| Kernel size | CEIQ | NIQE | FADE |
|--------------|-----------------------------------|-----------------------------------|-----------------------------------|
| 7×7 | 3.43 ± 0.09 | 4.58 ± 0.89 | 0.25 ± 0.06 |
| 5×5 | 3.34 ± 0.09 | 4.16 ± 0.58 | 0.26 ± 0.07 |
| 3×3 | 3.45 ± 0.08 | 4.15 ± 0.60 | 0.22 ± 0.06 |

3.2. Selection of the generator network parameters

The proposed approach contains architectural contributions like the multi-scale residual blocks (MSRB) and the refinement module. We conduct the experimental analysis for different configurations based on the number of blocks and kernel size. The present study started with the selection of appropriate multi-scale residual blocks while the keeping constant kernel size as 3×3 . There is an increase in the number of parameters of the method, as with the increase in the number of blocks, and the best image quality values are obtained by the network having one MSRB in the encoder layer as shown in Table 1. This result is relevant and desirable as it also favours the real-time application of the method given the lesser number of parameters. Further, we analyse the effect of different kernel sizes in the refinement module, while fixing the number of MSRB as one. The quantitative results shown in Table 2 reveals that the kernel size of 3×3 performs best in terms of image quality metric values.

3.3. Discriminator architecture

For the discriminators D_{DS} and DS , the present work utilizes similar architecture as proposed in [21] to classify the 70×70 overlapping image patches [23,38]. The discriminator architecture can be further referred to [21].

3.4. Loss function

The present study formulates the weighted loss function, which is a combination of the adversarial, cycle-consistency, contrast and unsharp regularization loss. The detail of the loss function is as follows:

3.4.1. Adversarial loss

The adversarial loss [22] is applied for both mapping functions. For mapping function, $DS : S \rightarrow SF$ domains, the generator and discriminator are G_{DS} and D_{DS} respectively. The adversarial loss function is expressed as follows:

$$L_{GAN}(G_{DS}, D_{DS}, S, SF) = E_{sf \sim p_{data}(sf)} [\log D_{DS}(sf)] + E_{s \sim p_{data}(s)} [\log (1 - D_{DS}(G_{DS}(s)))] \quad (3)$$

where the generator G_{DS} attempts to generate synthetic desmoked images that are similar to the smoke-free domain while the discriminator D_{DS} aims to distinguish the $G_{DS}(s)$ from real samples sf . The G_{DS} tries to minimize the D_{DS} ability to distinguish the images, while the D_{DS} maximizes the objective, i.e. $\min_{G_{DS}} \max_{D_{DS}} L_{GAN}(G_{DS}, D_{DS}, S, SF)$. The other mapping function, $RS : SF \rightarrow S$ follows the similar adversarial loss, i.e. $\min_{G_S} \max_{D_S} L_{GAN}(G_S, D_S, SF, S)$.

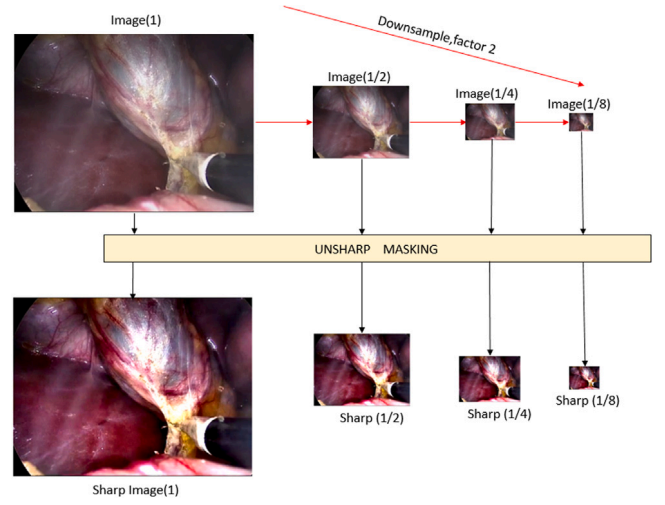


Fig. 6. The images obtained from unsharp masking is concatenated with input feature maps at different encoder layers.

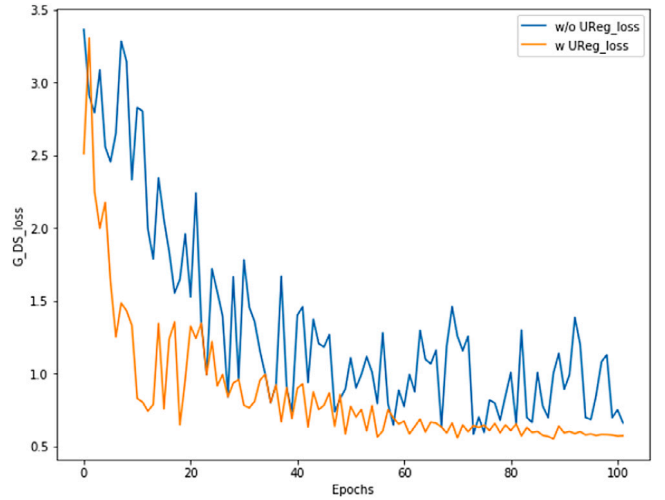


Fig. 7. The generator loss curves denoting the averaged loss values of the steps for every epoch with and without Unsharp Regularization loss.

3.4.2. Cycle consistency loss

In order to ensure the two mapping functions, DS and RS produce images that closely map to the respective domains, a cycle-consistency loss is introduced. The loss forces the generator G_{DS} and G_S to reconstruct images $G_{DS}(G_S(sf))$ and $G_S(G_{DS}(s))$ to be identical to the input samples of domain SF and S . The loss is given as follows:

$$L_{cyclic}(G_{DS}, G_S) = \|G_{DS}(G_S(sf)) - sf\|_1 + \|G_S(G_{DS}(s)) - s\|_1 \quad (4)$$

3.4.3. Contrast loss

The quality of an image can be defined as the ratio of variance (σ^2) and mean (μ) of the image intensity [39]. Further, the contrast enhancement factor (CEF) states the level of contrast enhancement in the reconstructed image over the input image. The value of CEF greater than one depicts the improvement in contrast of processed image over the input image. The formula of CEF is given as follows:

$$CEF = \frac{Q_{sf}}{Q_s}, \text{ where } Q = \frac{\sigma^2}{\mu} \quad (5)$$

$$L_{contrast}(G_{DS}) = 1/CEF$$

Table 3Determination of weight parameter for cycle-consistency loss(w_1).

| w_1 | CEIQ | NIQE | FADE |
|-------|--------------------|--------------------|--------------------|
| 5 | 3.43 ± 0.08 | 4.01 ± 0.62 | 0.23 ± 0.05 |
| 10 | 3.32 ± 0.10 | 4.05 ± 0.57 | 0.27 ± 0.07 |
| 25 | 3.15 ± 0.13 | 4.65 ± 0.64 | 0.26 ± 0.08 |
| 100 | 3.05 ± 0.11 | 4.34 ± 0.56 | 0.24 ± 0.06 |

Table 4Determination of weight parameter for contrast loss(w_2).

| w_2 | CEIQ | NIQE | FADE |
|-------|--------------------|--------------------|--------------------|
| 0.01 | 3.42 ± 0.86 | 4.09 ± 0.61 | 0.28 ± 0.07 |
| 0.025 | 3.37 ± 0.09 | 4.06 ± 0.85 | 0.24 ± 0.06 |
| 0.05 | 3.42 ± 0.08 | 4.21 ± 0.54 | 0.26 ± 0.06 |
| 0.1 | 3.43 ± 0.08 | 4.58 ± 0.88 | 0.26 ± 0.06 |

As smoke is assumed to be phenomenon that degrades the contrast of an image. The present study aims to reduce the smoke by inculcating inverse of the CEF as a loss function. On the minimization of this loss we can maximize the overall contrast of the generated image. We apply this loss in the desmoking operation in order to obtain better visibility of the tissues, and surgical tools.

3.4.4. Unsharp regularization loss

Several general regularization techniques have been utilized to stabilize the GANs training [40]. The problems of GANs such as mode collapse, non-convergence and diminishing gradients are well known and this reduces the reliability of the method in medical imaging. In order to help stabilize the GANs training, we propose regularization technique called the Unsharp Regularization loss that is relevant for the task of desmoking in images. This loss reduces the oscillations in the generator loss and penalizes the network from diverging and helps to generate consistent, stable results. The loss function is given as follows:

$$L_{UReg}(G_{DS}) = \|(G_{DS}(s)) - f(s)\|_1 \quad (6)$$

where $f(s)$ is set as the images obtained from unsharp masking corresponding to input smoke (s) images for 20% of the smoke domain data applicable for every epoch at its regular intervals. The rest 80% of the smoke domain data for every epoch, $f(s)$ is set as $G_{DS}(s)$. The image dimension of $f(s)$ is the same as that of the smoke image. The direct comparison of desmoked images with the images obtained from unsharp masking for 20% portion of every epoch in regular intervals helps to preserve the colour of the desmoked image and also helps to obtain convergence, while for the rest 80% portion of every epoch, the value of the loss is zero and this promotes the adversarial loss to learn an objective function that translates smoke to desmoked images. The effect of the regularization loss on the stability and convergence is evident in Fig. 7. The loss is also advantageous as it helps control the network from hallucinating/synthesizing new information and helps maintain consistency in the structural details, information in addition to preserving the colour. Experimental analysis on selecting the appropriate ratio for images obtained from unsharp masking and desmoked images for regularization loss can be found in the Table S3 (a) of SUPPLEMENTARY MATERIAL: A.

3.4.5. Aggregate loss function

The complete loss functions for the desmoking and resmoking networks are given as follows:

$$L(G_{DS}, D_{DS}) = L_{GAN}(G_{DS}, D_{DS}) + w_1 \times L_{cyclic}(G_{DS}, G_S) + w_2 \times L_{contrast}(G_{DS}) + w_3 \times L_{UReg}(G_{DS}) \quad (7)$$

$$L(G_S, D_S) = L_{GAN}(G_S, D_S) + w_1 \times L_{cyclic}(G_{DS}, G_S) \quad (8)$$

Table 5Determination of weight parameter for unsharp regularization loss(w_3).

| w_3 | CEIQ | NIQE | FADE |
|-------|--------------------|--------------------|--------------------|
| 5 | 3.29 ± 0.10 | 4.92 ± 0.84 | 0.28 ± 0.08 |
| 10 | 3.51 ± 0.08 | 4.01 ± 0.56 | 0.23 ± 0.58 |
| 25 | 3.42 ± 0.08 | 4.27 ± 0.70 | 0.24 ± 0.57 |
| 100 | 3.465 ± 0.80 | 3.93 ± 0.61 | 0.24 ± 0.51 |

Table 6

Ablation study for the proposed Cyclic-DesmokeGAN on image quality of laparoscopic images. (high value of CEIQ, less value of NIQE and less value of FADE indicates better image quality).

| Image | Method | CEIQ | NIQE | FADE |
|-------------|--------------------------------|--------------|--------------|--------------|
| 1 | Original input image | 3.375 | 4.920 | 0.784 |
| | Base network | 3.345 | 4.410 | 0.611 |
| | Base network+MSRB | 3.354 | 4.863 | 0.575 |
| | Base network+MSRB+RM | 3.399 | 4.520 | 0.462 |
| | Proposed network+Contrast loss | 3.436 | 3.943 | 0.341 |
| | Proposed method | 3.500 | 4.036 | 0.325 |
| 2 | Original input image | 3.460 | 4.242 | 0.474 |
| | Base network | 3.461 | 4.272 | 0.433 |
| | Base network+MSRB | 3.466 | 4.662 | 0.380 |
| | Base network+MSRB+RM | 3.523 | 4.094 | 0.287 |
| | Proposed network+Contrast loss | 3.532 | 4.064 | 0.241 |
| | Proposed method | 3.566 | 3.727 | 0.239 |
| Mean of val | Original input image | 3.285 | 4.511 | 0.549 |
| | Base network | 3.300 | 4.247 | 0.488 |
| | Base network+MSRB | 3.304 | 4.731 | 0.424 |
| | Base network+MSRB+RM | 3.361 | 4.487 | 0.319 |
| | Proposed network+Contrast loss | 3.383 | 4.060 | 0.251 |
| | Proposed method | 3.435 | 4.051 | 0.233 |

The overall objective function consists of four losses: adversarial, cycle-consistency, contrast and unsharp regularization loss. Analytical selection of the weights for different loss functions helps to obtain the best image quality values with maximum desmoking. First, experimentation is done for the cycle-consistency loss, we started with the weights for contrast and unsharp regularization losses and fixed them as 0.1 and 5 respectively. The experimental results depict the weight parameter of 5 to be the best performing among the range of different weight parameters as shown in Table 3. Further, setting the weight parameter of cycle-consistency loss as 5 and the unsharp regularization as 5, we analyse the weight parameters: 0.01, 0.025, 0.05 and 0.1 for the contrast loss. The contrast loss is susceptible to the instability and oscillating nature of the adversarial loss, we have chosen a lower range of weight parameters for this experimentation as compared to that of other losses. The weight parameter 0.025 for the contrast loss yields the better image quality values as shown in Table 4. Finally, on setting the weight parameters for cycle-consistency and contrast loss to be 5 and 0.025 we experimented on the different weight parameters: 5, 10, 25, and 100 for the unsharp regularization loss. Table 5 shows that the weight parameter 10 can be the appropriate choice for unsharp regularization loss. Tables 3–5 depict the quantitative analysis of the experimentation performed to obtain the weight parameters for the loss functions. The final weight parameters set for the cycle-consistency (w_1), contrast (w_2) and unsharp regularization (w_3) losses are 5, 0.025 and 10 respectively.

3.5. Evaluation

For quantitative analysis of the generated images, the present study does not use popular methods like structure similarity index and peak signal to noise ratio, as the ground truth smoke-free laparoscopic images are not available. In this view, Table 6 quantitatively shows the utility of different components in the proposed method to improve the image quality in terms of (i) Contrast-Distorted Images Quality (CEIQ) [41], which measures the contrast distortion in the images,

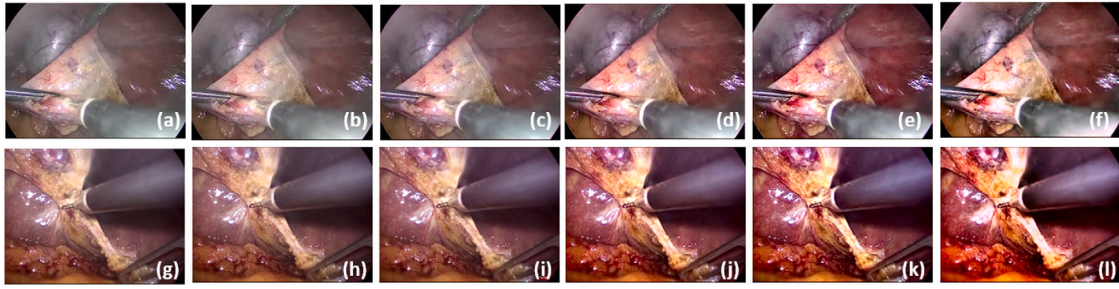


Fig. 8. Qualitative results of ablation study. (a):smoke images, (b):images obtained from base network, (c): images obtained from base network with multi-scale residual blocks, (d):images obtained from base network with MSRB's and refinement module, (e):images obtained from proposed network with addition of contrast loss, (f): images obtained from the proposed method.

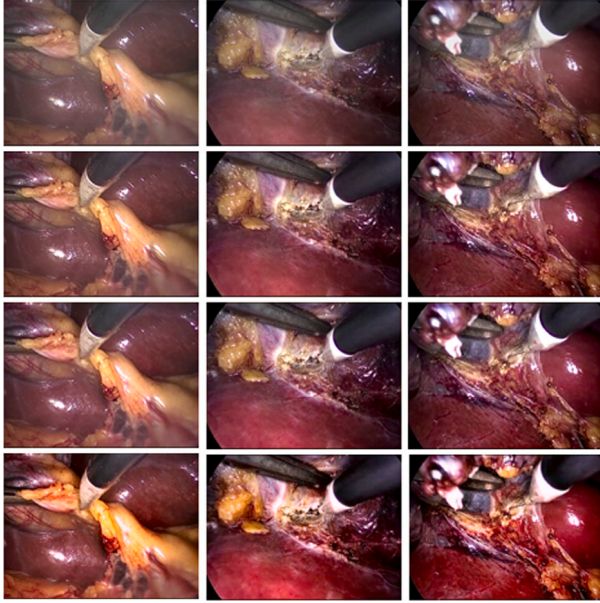


Fig. 9. Comparative study of the proposed method with different variants of CycleGAN. First row:smoke images, second row:images obtained from CycleGAN with U-Net, third row:images obtained from CycleGAN with modified ResNet, fourth row:images obtained from proposed method.

(ii) Naturalness Image Quality Evaluator (NIQE) [42], assesses image quality without knowledge of anticipated distortions, (iii) Fog Aware Density Evaluator (FADE) [43], which have been found useful in the quantification of fog removal from the images. The value of FADE indicates the density of the fog, which has similar characteristics to smoke, hence lesser value of FADE indicates better smoke reduction. The less value of NIQE suggest that the generated images have better perceptual quality and naturalness in colours, while the higher value of CEIQ indicates better contrast.

This section first depicts the importance of each key component in the proposed method. The base network comprises of CycleGAN with adversarial and cycle-consistency losses. The second column of Fig. 8 shows the base network partially reduces the smoke component and fails at capturing sharp, fine structural details in the images. The third column depicts the base network with the addition of the multi-scale residual blocks in the encoder layers, this results in better smoke reduction. Further, the addition of the refinement module to the network with MSRB's results in the generation of images with sharper boundaries and fine structural information and this evident in the fourth column. The utility of the contrast loss is noticeable in the fifth column images, as it enables to explicitly control the overall contrast in the desmoked image. Finally, column six denotes the desmoked image obtained from the proposed method that also contains the unsharp

regularization loss in the objective function. The stabilizing effect of this loss is evident in Fig. 7.

4. Experimental results

4.1. Dataset

We utilize the Cholec80 dataset [44] that contains eighty videos of cholecystectomy surgeries that are manually labelled as smoke/non-smoke image sequence by [45]. These videos are converted into approximately 100K annotated images. For the present study, we have selected 1200 and 200 pair of smoke, smoke-free images as our training set and testing set respectively (The dataset split analysis can be found in Table S3 (c) of SUPPLEMENTARY MATERIAL: A). The camera arrangement leads to the black corners in the images. The cropping out black corners avoids network to learn unnecessary information, hence, we have cropped all training and testing images from its boundary, which reduced to size 240×320 . Further, to ensure network learning on a diverse set of images, we have selected images to have varying levels of smoke at different depths.

4.2. Implementation details

In [46], the stability during the learning process of Least Squares GANs was found to be better than regular GANs. Inspired by that approach, the training procedure is stabilized by utilization of the least-squared loss instead of the log-likelihood loss. The generator and discriminator networks minimize the $E_{s \sim p_{data}(s)}[(D_{DS}(G_{DS}(s)) - 1)^2]$ and $E_{sf \sim p_{data}(sf)}[(D_{DS}(sf) - 1)^2] + E_{s \sim p_{data}(s)}[(D_{DS}(G_{DS}(s)))^2]$ objective functions respectively. The ADAM optimizer is used with the learning rate of 0.0001 and is trained for 100 epochs with a batch size of one. All the hyperparameters, weight parameters and experimental analysis were based on k-fold cross-validation, with k-fold set to 5. The Tensorflow framework and NVidia K80 GPU offered by Google Colaboratory [47] was used for the study. For detailed information on training methodology, please refer Section S.1 of SUPPLEMENTARY MATERIAL: A.

4.3. Results

This study demonstrates the advantage of the proposed method in comparison to CycleGAN with modified ResNet [37] and U-Net [48] as the generators. The U-Net architecture based CycleGAN relies on normal convolution blocks and does not contain any specific architectural highlights. The results obtained with this method are sub-optimal as shown in the second row of Fig. 9 and Table 7. The ResNet based CycleGAN does not include multi-scale feature extraction or technique to generate sharper images, the results in the third row of Fig. 9 shows that the overall smoke content is not completely reduced and there still exists smoke at some portions of the images. The proposed method has a few architectural improvements like the multi-scale residual

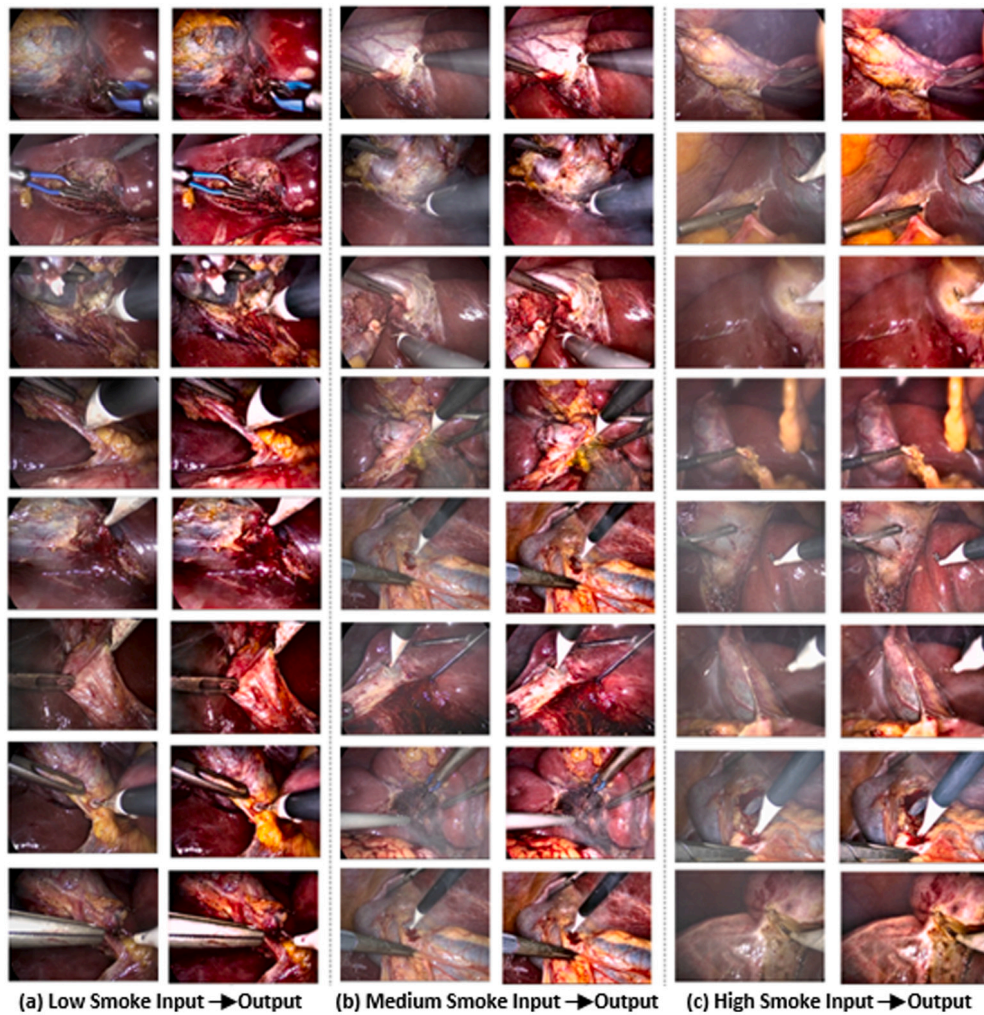


Fig. 10. Smoke removal performance of proposed Cyclic-DesmokeGAN on laparoscopic images with different level of smoke. (a) Low smoke input and corresponding output images, (b) medium smoke and corresponding output images (c) high smoke and corresponding output images.

Table 7

Comparative study of the proposed method different variants of CycleGAN (Mean of Image Quality Metrics).

| Method | CEIQ | NIQE | FADE |
|-------------------------------|--------------|--------------|--------------|
| Original input images | 3.285 | 4.511 | 0.549 |
| CycleGAN with U-Net | 3.357 | 4.349 | 0.338 |
| CycleGAN with modified ResNet | 3.365 | 4.124 | 0.323 |
| Proposed method | 3.454 | 4.055 | 0.229 |

blocks, refinement module along with new losses like the contrast, unsharp regularization loss that helps to obtain images with lesser smoke content, better contrast and fine structural information. This is evident in the qualitative and quantitative results shown in Fig. 9 and Table 7 respectively.

Fig. 10 demonstrates the smoke removal capability of the proposed approach at different densities of smoke. The images are categorized in three different levels of smoke i.e. low, medium and high based on FADE value. We manually observed and classify images as follows: (a) low smoke (images with $FADE < 0.4$), (b) medium smoke (images with $0.4 < FADE < 0.6$) (c) high smoke (images with $FADE > 0.6$). The reduction of smoke is quite evident in the respective second column images in each case, at the same time the colour of input images remain preserved in the output images, the effect of smoke removal in terms of better fine details, structural information and colour contrast. Fig. 11 shows the comparative performance of the Colour Ellipsoid

Table 8

Comparison study of the proposed approach with other state-of-the-art methods on Test data.

| Image | Method | CEIQ | NIQE | FADE |
|-------|------------------------|-------------------------------------|-------------------------------------|--------------------------------------|
| Mean | Original input image | 3.286 ± 0.124 | 4.510 ± 1.057 | 0.553 ± 0.171 |
| ± | Colour Ellipsoid Prior | 2.732 ± 0.282 | 4.794 ± 2.049 | 0.306 ± 0.214 |
| S.D | Dark Channel Prior | 2.948 ± 0.137 | 4.231 ± 0.828 | 0.324 ± 0.059 |
| | Non-local Dehazing | 3.133 ± 0.154 | 4.363 ± 0.976 | 0.240 ± 0.071 |
| | DehazeNet | 3.116 ± 0.179 | 4.463 ± 1.013 | 0.324 ± 0.106 |
| | CycleGAN | 3.365 ± 0.114 | 4.188 ± 0.660 | 0.360 ± 0.100 |
| | Cycle-Dehaze | 3.400 ± 0.104 | 4.298 ± 0.763 | 0.295 ± 0.077 |
| | Cyclic-DesmokeGAN | 3.475 ± 0.099 | 4.153 ± 0.746 | 0.237 ± 0.0055 |

Prior [14], Dark Channel Prior (DCP) [11], DehazeNet [17], Non-local Dehaze [49], CycleGAN [21], Cycle-Dehaze [50] and proposed Cyclic-DesmokeGAN. This comparison is performed on Test data, which is not used during the training process. Further, the performance evaluation is given in Table 8, where, the proposed method produces the best values of image quality indexes i.e. maximum CEIQ, and least FADE and NIQE.

Further, to show the generalization ability of the proposed network, we have tested the Cyclic-DesmokeGAN approach on completely different datasets: (i) Natural foggy Image dataset, (ii) Inhomogeneity artefact MRI data. For more details please follow section S. 6. of SUPPLEMENTARY MATERIAL: B.

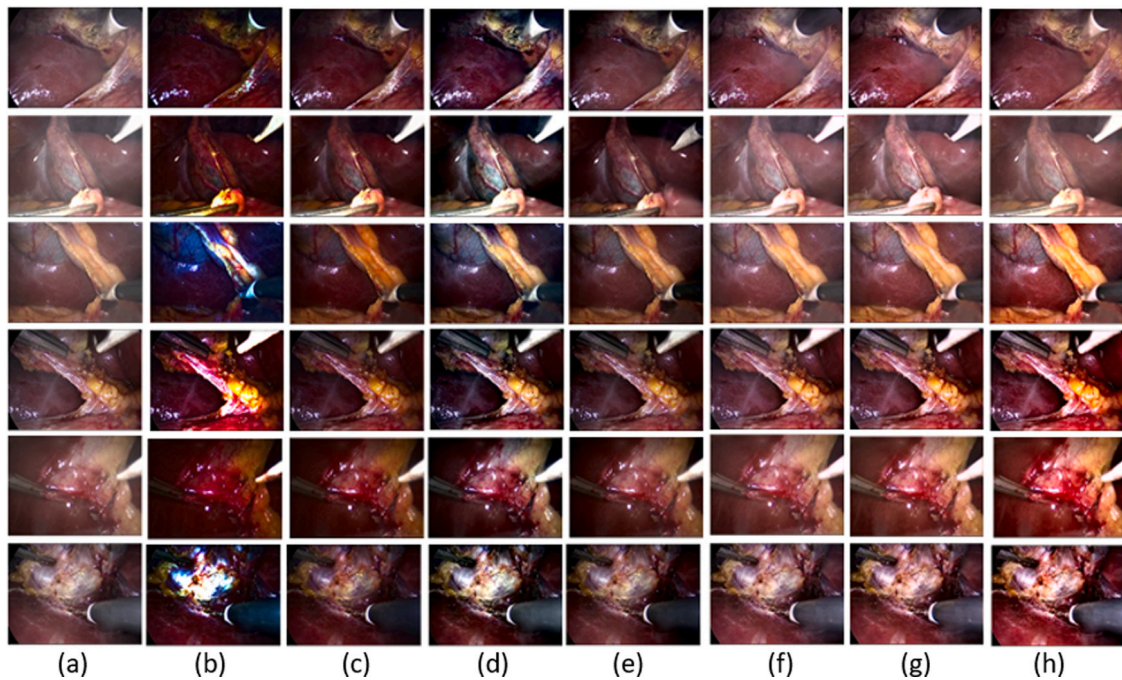


Fig. 11. Qualitative analysis of the comparison study of proposed method with other state-of-the-art methods for desmoking. (a) randomly selected test input images, (b) Colour Ellipsoid Prior [14], (c) Dark Channel Prior [11], (d) Non-Local Image Dehazing [49], (e) DehazeNet [17], (f) CycleGAN [21], (g) Cycle-Dehaze [50] and (h) proposed Cyclic-DesmokeGAN.

5. Discussion

It is evident from the results shown in Fig. 8 and Table 6 that the different modules of the proposed Cyclic-DesmokeGAN optimally contribute to the task of desmoking. The multiscale residual blocks are useful at selecting the features at multiple scales. The refinement module helps to recover more structural information and concatenation of images obtained from unsharp masking at the encoder layers improves the high-frequency information such as edges and fine details. The complete architecture trained with the overall objective function depicts the enhancement in the image in terms of contrast and colour preservation. The concatenation of images obtained from unsharp masking at each encoder layer and contrast loss is quite beneficial and help the network to learn to reduce the low contrast smoke component. The inculcation of the unsharp regularization loss keeps a check on the instability that arises from generative adversarial networks for the task of desmoking in laparoscopic images.

The quantitative analysis as shown in Table 6, it is observed that the NIQE value increases with the addition of the MSRB's. This undesirable rise in the NIQE value is due to the image translation to the other domain since the features of the generated image do not compare well with the features derived from a corpus of natural, undistorted images that the NIQE metric tries to statistically examine for. On the addition of the refinement module, the overall structural information is improved and this results in features that are comparable to the features derived from a corpus of natural images, hence obtaining a desirable lower NIQE value.

On comparison with other methods of desmoking, proposed CyclicDesmokeGAN produced images with lesser distortion (high colour distortion seen in CEP [14]), blurring (seen in DCP [11]), and darkness (seen in Non-local dehazing [49], and DehazeNet [17]), as shown in Fig. 11. The CycleGAN [21] and Cycle-Dehaze [50] techniques also show similar result but comparatively the proposed method image has better contrast and less smoke in it. Table 8 suggested that proposed approach obtained laparoscopic images with better contrast, colour and have the least smoke, as it obtained the highest CEIQ, least NIQE and least FADE metric values. The desmoked images have

become brighter and sharper using the proposed method. There is no appreciable distortion of structures visualized in the output image. In a few cases, where smoke is highly localized and looks like the structural part of the laparoscopic image, the performance of the proposed Cyclic-DesmokeGAN degraded, however, it still achieved better results as compared to the state-of-the-art methods.

6. Conclusion

This paper presented a new unsupervised learning method for removal of smoke in the laparoscopic images. The proposed generator architecture consisting of multiscale residual blocks, which helped to alleviate the smoke component at different scales, while the proposed refinement module helped to refine and recover more detailed edges, and structure. This generator network has proven to produce desmoked images robustly. Further, in addition to adversarial and cycle-consistency losses, the proposed contrast loss has resulted in the enhancement of the image in terms of overall contrast, while unsharp regularization loss helps to stabilize the network and also preserves the colour. The proposed Cyclic-DesmokeGAN is qualitatively and quantitatively compared with Colour Ellipsoid Prior, DCP, Non-Local Image Dehazing, DehazeNet, CycleGAN, and Cycle-Dehaze methods. The comparative results have shown a clear edge of the proposed method over the other methods for the smoke removal of laparoscopic surgery images.

The surgical smoke removal relies on mechanical solutions that still have a lag time, having a digital visualization of the frames that automatically removes smoke will be of great use for the practitioners and surgeons. Hence, this work can be extended to develop a system that can perform in real-time. The tasks of smoke detection and smoke removal are related to each other, hence employing multi-task learning as part of future work to jointly detect smoke and remove it will lead to more accurate systems as it relies on innovative techniques like parameter sharing between related tasks etc. Further, the Spatio-temporal consistency between the sequences of frames can also be studied, which would ensure that the level of smoke removal is not dependent on its amount but just on its presence in that frame.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

A.1. No-reference image quality indexes used in the study

A.1.1. Contrast enhancement based image quality measure (CEIQ)

This image quality index predict the image quality using Support Vector Regression (SVR), which uses the information of input image and histogram equalized processed contrast enhanced image. The learning of SVR was performed using only five input features: (i) structure similarity index measure between original input image and contrast enhanced image, (ii) histogram based entropy of input image, (iii) histogram based entropy of contrast enhanced image, (iv) cross entropy of the original image and the enhanced image, (v) cross entropy of the enhanced image and the original image. The CEIQ is efficient to predict the quality of contrast altered image. Its high value indicates enhanced image with less distortion because of the contrast.

A.1.2. Natural image quality evaluator (NIQE)

This model predict the quality of image without knowledge of distortion and human opinion score. It uses the information of statistical regularities derived from the images. NIQE takes help from natural scene statistic (NSS) model to derive quality aware features from a corpus of natural images and these feature fit to multivariate Gaussian model, called learned MVG model.

The MVG probability density is given as follows:

$$MVG(f) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(f - v)^T (\Sigma)^{-1} (f - v)\right] \quad (9)$$

where f is the local statistical features derived from foggy images, d dimensions, v is the mean and Σ is covariance matrix of the MVG model.

Eventually, the quality of input image is calculated in terms of distance between learned MVG model and MVG fitted model obtained from the input image. where v_1, v_2 are the mean vectors and Σ_1, Σ_2 are the covariance matrix of the learned MVG model and input image's MVG model respectively.

$$D(v_1, v_2, \Sigma_1, \Sigma_2) = \sqrt{((v_1 - v_2)^T (\frac{\Sigma_1 + \Sigma_2}{2})^{-1} (v_1 - v_2))} \quad (10)$$

A.1.3. Fog Aware Density Evaluator (FADE)

This image quality index spatially designed to estimates the density of fog or visibility in the natural foggy images without any reference. FADE takes help from NSS model to derive local statistical features and mean subtracted contrast normalized (MSCN) coefficient of foggy and fog free images. Finally, MVG model used to predict the fog density based on the distance. It has been found that this image quality index is correlated well with the human perception in case of foggy images.

A.2. Loss curves comparison

The generator loss values versus the epochs graph depicted in Fig. 12 for CycleGAN, Cycle-Dehaze and our proposed Cyclic-DesmokeGAN indicates that our method achieves lower loss values and is more stable in comparison to the other two unsupervised methods. This leads better convergence and helps obtain results that are consistent and stable.

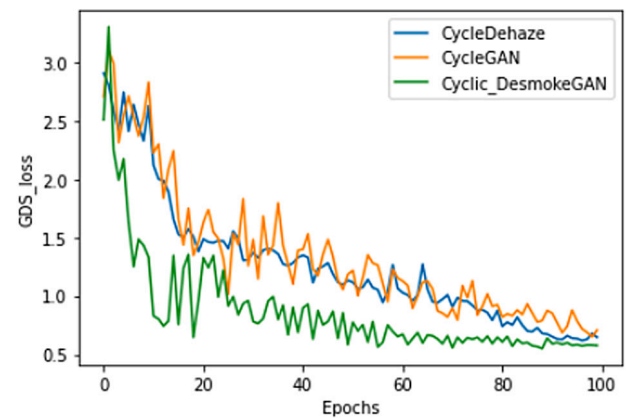


Fig. 12. The generator loss curves denoting the averaged loss values of the steps for every epoch for CycleGAN, Cycle-Dehaze and the proposed Cyclic-DesmokeGAN.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.compbiomed.2020.103873>.

References

- [1] P.H. Cosman, C.J. Shearer, T.J. Hugh, A.V. Biankin, N.D. Merrett, A novel approach to high definition, high-contrast video capture in abdominal surgery, *Ann. Surg.* 245 (4) (2007) 533.
- [2] N. Lawrentschuk, N.E. Fleshner, D.M. Bolton, Laparoscopic lens fogging: a review of etiology and methods to maintain a clear visual field, *J. Endourol.* 24 (6) (2010) 905–913.
- [3] H. Takahashi, M. Yamasaki, M. Hirota, Y. Miyazaki, J.H. Moon, Y. Souma, M. Mori, Y. Doki, K. Nakajima, Automatic smoke evacuation in laparoscopic surgery: a simplified method for objective evaluation, *Surg. Endosc.* 27 (8) (2013) 2980–2987.
- [4] C. Loukas, E. Georgiou, Smoke detection in endoscopic surgery videos: a first step towards retrieval of semantic events, *Int. J. Med. Robot. Comput. Assist. Surg.* 11 (1) (2015) 80–94.
- [5] N.A. Jalal, T.A. Alshirbaji, L. Mündermann, K. Möller, Features for detecting smoke in laparoscopic videos, *Curr. Dir. Biomed. Eng.* 3 (2) (2017) 521–524.
- [6] C. Wang, V. Sharma, Y. Fan, F.A. Cheikh, A. Beghdadi, O.J. Elle, R. Stiefelhagen, Can image enhancement be beneficial to find smoke images in laparoscopic surgery? in: *Color and Imaging Conference, Vol. 2018, No. 1, Society for Imaging Science and Technology*, 2018, pp. 163–168.
- [7] A. Kotwal, R. Bhalodia, S.P. Awate, Joint desmoking and denoising of laparoscopy images, in: *2016 IEEE 13th International Symposium on Biomedical Imaging, ISBI, IEEE*, 2016, pp. 1050–1054.
- [8] A. Baid, A. Kotwal, R. Bhalodia, S. Merchant, S.P. Awate, Joint desmoking, specular removal, and denoising of laparoscopy images via graphical models and Bayesian inference, in: *2017 IEEE 14th International Symposium on Biomedical Imaging, ISBI 2017, IEEE*, 2017, pp. 732–736.
- [9] C. Wang, F.A. Cheikh, M. Kaaniche, A. Beghdadi, O.J. Elle, Variational based smoke removal in laparoscopic images, *Biomed. Eng. Online* 17 (1) (2018) 139.
- [10] K. Tchaka, V.M. Pawar, D. Stoyanov, Chromaticity based smoke removal in endoscopic images, in: *Medical Imaging 2017: Image Processing, Vol. 10133, International Society for Optics and Photonics*, 2017, 101331M.
- [11] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (12) (2010) 2341–2353.
- [12] C.O. Ancuti, C. Ancuti, Single image dehazing by multi-scale fusion, *IEEE Trans. Image Process.* 22 (8) (2013) 3271–3282.
- [13] S. Salazar-Colores, E. Cabal-Yepez, J.M. Ramos-Arrequin, G. Botella, L.M. Ledesma-Carrillo, S. Ledesma, A fast image dehazing algorithm using morphological reconstruction, *IEEE Trans. Image Process.* 28 (5) (2018) 2357–2366.
- [14] T.M. Bui, W. Kim, Single image dehazing using color ellipsoid prior, *IEEE Trans. Image Process.* 27 (2) (2017) 999–1009.
- [15] X. Luo, A.J. McLeod, S.E. Pautler, C.M. Schlachta, T.M. Peters, Vision-based surgical field defogging, *IEEE Trans. Med. Imag.* 36 (10) (2017) 2021–2030.
- [16] H.-M. Hu, Q. Guo, J. Zheng, H. Wang, B. Li, Single image defogging based on illumination decomposition for visual maritime surveillance, *IEEE Trans. Image Process.* 28 (6) (2019) 2882–2897.
- [17] B. Cai, X. Xu, K. Jia, C. Qing, D. Tao, Dehazenet: An end-to-end system for single image haze removal, *IEEE Trans. Image Process.* 25 (11) (2016) 5187–5198.

- [18] X. Fu, J. Huang, X. Ding, Y. Liao, J. Paisley, Clearing the skies: A deep network architecture for single-image rain removal, *IEEE Trans. Image Process.* 26 (6) (2017) 2944–2956.
- [19] S. Bolkar, C. Wang, F.A. Cheikh, S. Yildirim, Deep smoke removal from minimally invasive surgery videos, in: 2018 25th IEEE International Conference on Image Processing, ICIP, IEEE, 2018, pp. 3403–3407.
- [20] V. Vishal, N. Sharma, M. Singh, Guided unsupervised desmoking of laparoscopic images using cycle-desmoke, in: OR 2.0 Context-Aware Operating Theaters and Machine Learning in Clinical Neuroimaging, Springer, 2019, pp. 21–28.
- [21] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.
- [22] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.
- [23] H. Zhang, V. Sindagi, V.M. Patel, Image de-raining using a conditional generative adversarial network, 2017, arXiv preprint [arXiv:1701.05957](https://arxiv.org/abs/1701.05957).
- [24] H. Zhang, V.M. Patel, Densely connected pyramid dehazing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3194–3203.
- [25] W. Liu, X. Hou, J. Duan, G. Qiu, End-to-end single image fog removal using enhanced cycle consistent adversarial networks, 2019, arXiv preprint [arXiv:1902.01374](https://arxiv.org/abs/1902.01374).
- [26] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.
- [27] L. Karacan, Z. Akata, A. Erdem, E. Erdem, Learning to generate images of outdoor scenes from attributes and semantic layouts, 2016, arXiv preprint [arXiv:1612.00215](https://arxiv.org/abs/1612.00215).
- [28] P. Sangkloy, J. Lu, C. Fang, F. Yu, J. Hays, Scribbler: Controlling deep image synthesis with sketch and color, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5400–5409.
- [29] M.-Y. Liu, T. Breuel, J. Kautz, Unsupervised image-to-image translation networks, in: *Advances in Neural Information Processing Systems*, 2017, pp. 700–708.
- [30] H. Yang, J. Sun, A. Carass, C. Zhao, J. Lee, Z. Xu, J. Prince, Unpaired brain mr-to-ct synthesis using a structure-constrained cyclegan, in: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2018, pp. 174–182.
- [31] I. Fantini, L. Rittner, C. Yasuda, R. Lotufo, Automatic detection of motion artifacts on MRI using deep CNN, in: 2018 International Workshop on Pattern Recognition in Neuroimaging, PRNI, IEEE, 2018, pp. 1–4.
- [32] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M.K. Kalra, Y. Zhang, L. Sun, G. Wang, Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss, *IEEE Trans. Med. Imag.* 37 (6) (2018) 1348–1357.
- [33] M. Khened, V.A. Kollerathu, G. Krishnamurthi, Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers, *Med. Image Anal.* 51 (2019) 21–45.
- [34] J. Li, F. Fang, K. Mei, G. Zhang, Multi-scale residual network for image super-resolution, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 517–532.
- [35] S. Nah, T. Hyun Kim, K. Mu Lee, Deep multi-scale convolutional neural network for dynamic scene deblurring, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3883–3891.
- [36] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, G. Cottrell, Understanding convolution for semantic segmentation, in: 2018 IEEE Winter Conference on Applications of Computer Vision, WACV, IEEE, 2018, pp. 1451–1460.
- [37] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual dense network for image restoration, 2018, arXiv preprint [arXiv:1812.10477](https://arxiv.org/abs/1812.10477).
- [38] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al., Photo-realistic single image super-resolution using a generative adversarial network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4681–4690.
- [39] M. Singh, A. Verma, N. Sharma, Optimized multistable stochastic resonance for the enhancement of pituitary microadenoma in MRI, *IEEE J. Biomed. Health Inform.* 22 (3) (2018) 862–873.
- [40] K. Roth, A. Lucchi, S. Nowozin, T. Hofmann, Stabilizing training of generative adversarial networks through regularization, in: *Advances in Neural Information Processing Systems*, 2017, pp. 2018–2028.
- [41] J. Yan, J. Li, X. Fu, No-reference quality assessment of contrast-distorted images using contrast enhancement, 2019, arXiv preprint [arXiv:1904.08879](https://arxiv.org/abs/1904.08879).
- [42] A. Mittal, R. Soundararajan, A.C. Bovik, Making a “completely blind” image quality analyzer, *IEEE Signal Process. Lett.* 20 (3) (2012) 209–212.
- [43] L.K. Choi, J. You, A.C. Bovik, Referenceless prediction of perceptual fog density and perceptual image defogging, *IEEE Trans. Image Process.* 24 (11) (2015) 3888–3901.
- [44] A.P. Twinanda, S. Shehata, D. Mutter, J. Marescaux, M. De Mathelin, N. Padoy, Endonet: A deep architecture for recognition tasks on laparoscopic videos, *IEEE Trans. Med. Imag.* 36 (1) (2016) 86–97.
- [45] A. Leibetseder, M.J. Primus, S. Petschmann, K. Schoeffmann, Real-time image-based smoke detection in endoscopic videos, in: Proceedings of the on Thematic Workshops of ACM Multimedia 2017, ACM, 2017, pp. 296–304.
- [46] X. Mao, Q. Li, H. Xie, R.Y. Lau, Z. Wang, S. Paul Smolley, Least squares generative adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2794–2802.
- [47] Google, Colaboratory, <https://colab.research.google.com>.
- [48] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2015, pp. 234–241.
- [49] D. Berman, S. Avidan, et al., Non-local image dehazing, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1674–1682.
- [50] D. Engin, A. Genç, H. Kemal Ekenel, Cycle-dehaze: Enhanced cyclegan for single image dehazing, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 825–833.