IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 17, 2020, accepted October 21, 2020, date of publication November 17, 2020, date of current version December 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3038437

Desmoking Laparoscopy Surgery Images Using an Image-to-Image Translation Guided by an Embedded Dark Channel

SEBASTIÁN SALAZAR-COLORES[1], HUGO MORENO JIMÉNEZ[1], CÉSAR JAVIER ORTIZ-ECHEVERRI[2], (Member, IEEE), AND GERARDO FLORES[1]

[1]Centro de Investigaciones en Óptica (CIO), León 37150, Mexico
[2]Facultad de informática, Universidad Autónoma de Querétaro (UAQ), Querétaro 76230, Mexico

Corresponding author: Gerardo Flores (gflores@cio.mx)

ABSTRACT In this paper, a method to remove the smoke effects in laparoscopic images is presented. The proposed method is based on an image-to-image conditional generative adversarial network endowed with a dark channel's embedded guide mask. The obtained experimental results were evaluated and quantitatively compared with desmoking state-of-art methods using the Peak Signal-to-Noise Ratio (PSNR) metrics and Structural Similarity (SSIM) index. Those results throw an improved performance compared with relevant works. Also, the processing time required by our method is 92 frames per second; a processing time that sets the foundation for a possible real-time implementation in a more modest embedded system.

INDEX TERMS Laparoscopy, image smoke removal, conditional generative adversarial network, dark channel.

I. INTRODUCTION

In laparoscopic surgery, the image's visibility can be severely degraded by the smoke caused by the CO_2 injection and dissection tools, thus reducing the visibility of organs and tissues [1]. This lack of visibility increases the surgery time and even the probability of the surgeon's mistakes, then producing negative consequences on the patient's health [2], [3]. Clinical studies have shown that digital smoke removal reduces surgery operative time and surgeons anxiety during surgical operation [2], [4]. With that in mind, recently, researchers have been conducted works to reduce or even avoid the smoke effects. Commonly, surgeons perform the smoke removal process with medical instruments [5]. However, this way results in being costly and impractical, requiring another option to solve the problem. This is how the image processing and deep learning techniques raise as an alternative to this and related medical problems [2], [6]–[21]. In the literature, we can find three main approaches for addressing the smoke removal problem:

(a) Traditional image processing techniques [7]–[11],
(b) Physics-model-based methods, especially the atmospheric scattering model and the dark channel prior (DCP) [2], [12], [13],
(c) Artificial intelligence (AI) methods based on convolutional neural networks and generative networks [15]–[19], [22].

Approaches (a) and (b) have been widely studied in the literature, while approach (c) is practically unexplored. The technique (c) AI, plus (a) image processing, is used in this paper. In particular, we based our method on a conditional generative adversarial network (cGAN) plus a guide mask based on a dark channel. Fig. 1 depicts an example of the proposed approach.

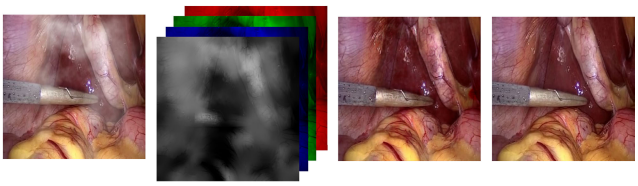

FIGURE 1. The proposed approach for solving the digital smoke removal problem in laparoscopic surgery: (a) input image with simulated smoke; (b) input image with its corresponding embedded dark channel; (c) output of our method; and (d) ground-truth.

The associate editor coordinating the review of this manuscript and approving it for publication was Jonghoon Kim.

A. RELATED WORK

In this section, we describe the most important works that deal with the problem of desmoking images. We classify them by using the three approaches above. We begin by presenting those works that use physics-model-based techniques.

In the literature, image-processing-based techniques are commonly used in solving the desmoking problem. The most relevant works are as follows. In [7] and [8], joint image desmoking and denoising of laparoscopic images is formulated as a Bayesian inference problem. Then, in [9], the method first recovers the visibility and enhances the contrast of hazy images, then luminance of enhanced images is fused in the gradient domain by solving the Poisson equation in the frequency domain. The authors of [10] use a physical model of gas scattering and Lagrangian methods to remove unwanted smoke components. The work of [11] proposes an algorithm to remove the dehazing artifacts by optimizing the transmission map; the method is based on the estimated reflectance map and the structure-guided transmission map ℓ_0.

The physics-model-based approach estimates the transmission map and atmospheric lighting. This approach includes the methods based on the atmospheric scattering model and dark channel prior (DCP). In this regard, [12] presents a color attenuation prior to haze removal. Their method is based on a linear model that learns the parameters employing a supervised learning method. Whereas, [13] proposes an adapted dark-channel prior method combined with histogram equalization to remove smoke artifacts, thus recovering the radiance image. The authors of [14] assume that the smoke veil has low contrast and low inter-channel differences. Then, a cost function is defined based on prior knowledge and solved using an augmented Lagrangian method. On the other hand, in technological regard, the authors of [2] have implemented a digital system based on the dark channel prior.

Since the smoke patterns are complex and hard to categorize, the two approaches mentioned above generally produces inconsistent results, altering and saturating colors in no-smoke regions. As a solution, the artificial-intelligence-based approaches arise, showing superior results in dehazing-like problems. The state-of-the-art in artificial-intelligence-based solutions is as follows. In [15] a deep neural network is implemented in real-time to enhance the quality of surgery video frames trying to eliminate the smoke presented in the image. The authors of [16] propose an unsupervised image-to-image translation with a generative adversarial network (GAN) architecture; they use a perceptual image quality metric for the loss function. In [17], a convolutional neural network (CNN) takes the smoke image and its pyramidal decomposition as inputs, and then, it directly outputs a smoke-free image without relying on any physical model and with no estimation of intermediate parameters. Also, in [18], it is implemented an end-to-end network called Cycle-Desmoke. Such a network uses a *generator* architecture with two-loss functions: a guided-unsharp upsample loss function and an adversarial and cycle-consistency loss function. In [19], the authors implement an unsupervised framework for learning smoke removal that uses a fully convolutional encoder-decoder network to generate the same size desmoked image. In [22], an unsupervised deep learning technique based on a GAN converts laparoscopic images from the smoke domain to the smoke-free domain. The network comprises a *generator* architecture endowed with an encoder-decoder structure composed of multi-scale feature extraction at each encoder block. Then, it obtains a robust deep representation map to reduce the image's smoke component.

With the state-of-the-art mentioned above, we propose a method of desmoking laparoscopic surgery images based on the pix2pix architecture presented in [23]. The pix2pix has shown promising results in the task of desmoking laparoscopic images [24], [25]. Next, we present the motivation for our work.

B. MOTIVATION

In laparoscopic surgery, surgery is performed through small incisions using long instruments and video cameras. To create a working and viewing space in the abdomen, carbon dioxide (CO2) is insufflated to separate the abdominal wall from internal organs [26], [27]. The inconvenience is that carbon dioxide causes a natural lack of visibility in laparoscopy surgery images. This can produce adverse effects during the surgery, such as surgeon stress, increased surgery time, and decision-making errors. Thus, the motivation to conduct this research is to solve the lack of visibility in laparoscopy surgery images without any mechanical device intervention.

Despite the conducted efforts to generate smoke-free laparoscopic images with no mechanical device intervention, current methods fail in crucial factors to be applicable and useful in surgery. Such methods generally have the following issues: high processing times, changes in color, and the emergence of artifacts in the image [17].

C. CONTRIBUTION

As we have mentioned above, in our approach, we use AI methods plus image processing techniques. In this regard, previous works have shown that embedding additional information into a neural network can increase its performance [2]–[4]. Then, our hypothesis is based on the idea that *the use of a state-of-the-art generative neural network, together with additional embedded information (in our case, dark channel map), can improve the desmoking task in laparoscopic images*. In particular, embedded image masks can provide additional information to GANs to enhance the network's output [28]–[30]. Furthermore, the dark channel intensity is a good indicator of smoke level presence in images [31]. Thus, we propose a combination of a conditional generative adversarial network and an embedded dark channel mask. The proposed network is endowed with four input and output channels, where the useful information of the output channel

is the smoke-free image formed by the first three channels. In particular, our proposal is based on the pix2pix architecture together with two main characteristics: a) an embedded dark channel map is utilized as a guide mask in the input image, and b) a modification to the $L1$ loss function that does not consider the fourth channel when the loss is minimized. The dark channel mask gives concentration smoke data to the conditional generative adversarial network to identify the restoration level needed in each image region. Obtained results in synthetic images show that the proposed approach can remove local smoke and recover realistic tissue colors without affecting non-smoke areas. According to the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index, our method outperforms five state-of-the-art methods, including the image-to-image conditional generative model with no embedded mask (Pix2Pix). Obtained results can be reproduced with the source code and trained models available at our GitHub site presented in Section III. In summary, the main contribution is summarized as follows:

- The proposal of a desmoking method is applied to laparoscopic images. According to the given results and comparisons, such a method results in being fast and accurate.
- We provide qualitative and quantitative comparisons of our approach against the state-of-art.
- We liberate the code of the proposed method.

D. CONTENT

The remainder of this paper is as follows. In Section II it is presented the fundamental idea around our proposed method, including the foundations that support the method. Section III exposes the setup used in the design of the experiment and the obtained results; also, comparatives with six state-of-the-art methods are given. Finally, conclusions and a perspective of future research are presented in Section IV.

II. METHODS

In a smoked image, there exists a relation between the presence of smoke and the dark channel [32]. Motivated by that fact, this work's fundamental idea is as follows: by embedding a dark channel mask into the cGAN's input, we can obtain vital information to improve the desmoking performance in laparoscopic images. This section sets forth such an idea by developing the theoretical background, framework, and architecture description of our approach.

A. PROPOSED METHOD

The following researches strongly inspire our proposed method:

- In [33] is proposed a method based on a convolutional neural network oriented to classify tasks that receive two inputs: an RGB image and its corresponding salience map (S) (a salience map is an image that tries to predict human fixations). The obtained results of this method show that performance improves when meaningful data are incorporated.
- In [28]–[30], the authors demonstrate that object features such as shape, color, or even pose can be embedded from input masks in generative adversarial networks.

From the above, we propose a framework described by the following steps and summarized in Fig. 4.

1) We obtain an estimate of the dark channel map; this is performed through (1) and is depicted in the upper part of Fig. 4.
2) The estimated dark channel map is refined using *the Fast Guided Filter*, see (2). Please refer to the upper part of Fig. 4.
3) Next, the dark channel map is concatenated to the RGB channels; please refer to the upper-middle part of Fig. 4.
4) Employing the above information, we train a conditional Generative Adversarial Network (cGAN) utilizing a modified $L1$ loss function. A scheme of this network is shown in Fig. 4.
5) Finally, the cGAN is applied to generate an image free of smoke with few computing resources and a minimum amount of time. The results are depicted in Figs. 7 and 8.

From the above, we divide our approach into two parts: the dark channel and the conditional generative adversarial network. We start with the former.

B. THE DARK CHANNEL

1) THE RELATION BETWEEN THE DARK CHANNEL AND LAPAROSCOPIC IMAGES

For each pixel (x, y) the dark channel (I^{dark}) is defined as

$$I^{\text{dark}}(x, y) = \min_{c \in \{R,G,B\}} \left(\min_{z \in \Omega(x,y)} I^c(z) \right), \tag{1}$$

where $\Omega(x, y)$ is a kernel (usually squared) centered in the (x, y) position; $I^c(z)$ are the elements of the smoked laparoscopic image I in the positions $z \in \Omega(x, y)$; and c represent each *RGB*. In [32], [34], it is reported that the dark channel tends to have low values in the vast majority of regions without haze or smoke. This means that $I^{\text{dark}} \rightarrow 0$ holds. This statistical fact is called *dark channel prior*. In laparoscopic images, the tissues and organs are commonly bright and colorful, similar to outdoor natural images. Hence, the dark channel is still valid to discriminate between smoke and clear images. Fig. 2 depicts the dark channel map of laparoscopic images when there is smoke and when not. Smoke and clear regions tend to have high and low-intensity values, respectively.

2) DARK CHANNEL REFINEMENT

In dark channel computation, the implemented kernels $\Omega(x, y)$ are squared patches. Hence, the dark channel is not precisely aligned to the corresponding image I, provoking a spatial imprecision. For alleviating this, it is mandatory to integrate a *refinement stage* for the dark channel maps. In this regard, several methods have been proposed, such as

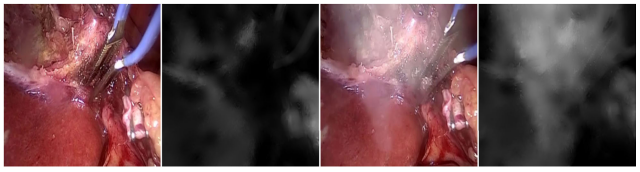

FIGURE 2. The dark channel in clear and smoked images: (a) clear laparoscopic image; (b) dark channel of (a); (c) smoked laparoscopic image; and (d) dark channel of (c).

[35] and [36]. In order to obtain an accurate refined dark channel map showed in (1), we apply a *guided filter* to I^{dark}, which is an edge-preserving smoothing filter based on a linear model in a region [37]. Such a filter is given as follows

$$I_{\text{ref}}^{\text{dark}}(z) = a(x,y)I(z) + b(x,y), \quad \forall z \in \Omega(x,y), \tag{2}$$

where $I_{\text{ref}}^{\text{dark}}$ is the filtering output dark channel; I is the guidance image; and z is the position of a pixel in the local squared window Ω of size $s \times s$ and centered in (x, y). The $a(x, y)$ and $b(x, y)$ parameters from linear model (2) are defined as

$$a(x,y) = \frac{\frac{1}{|\Omega|}\sum_{(z)\in\Omega(x,y)} I(z)I^{\text{dark}}(z) - \mu(x,y)\overline{I^{\text{dark}}(x,y)}}{\sigma(x,y)^2 + \epsilon}, \tag{3a}$$

$$b(x,y) = \overline{I^{\text{dark}}(x,y)} - a(x,y)\mu(x,y), \tag{3b}$$

where $\mu(x, y)$ and $\sigma(x, y)$ are the mean and variance of I in $\Omega(x, y)$; $\overline{I^{\text{dark}}(x, y)}$ is the mean of $I^{\text{dark}}(x, y)$ in Ω_k; and ϵ is parameter that regulates the smoothness degree. The computed and refined dark channel $I_{\text{ref}}^{\text{dark}}(z)$ is stacked into the smoked laparoscopic image I, as it is shown in Fig. 3.

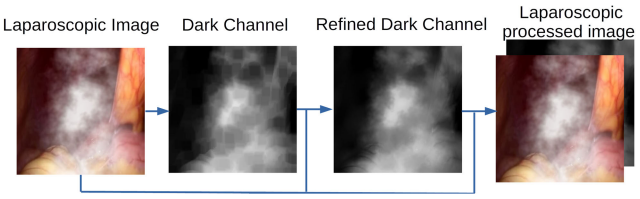

FIGURE 3. This figure shows the process of estimation, refinement, and embedding of the dark channel into the smoked laparoscopic image.

Once the refined dark channel is embedded, the processed images are given as inputs to the proposed cGAN. This is explained next.

C. CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS

1) DEFINITION

Generative adversarial networks (GANs) [38] are unsupervised generative models that learn a mapping $\mathbf{G} : r \rightarrow o$, where r is a random noise vector, and o is an output image. Such a mapping is implemented by a system of two competitor neural networks: a *generator* (G) and a *discriminator* (D). The *generator* is commonly formed by convolutional and deconvolutional layers, designed to produce outputs that cannot be distinguished from real images. The *discriminator* is given by a CNN trained to perform its best in detecting the counterfeits made by G. The conditional generative adversarial networks (cGAN's), unlike the GAN, additionally uses a conditional input vector i in its mapping, i.e. $\mathbf{G} : i, r \rightarrow o$ [23]. In this work, we implement a generator based on a U-Net architecture with input images of 256×256 pixels.[1] The discriminator is a Convolutional Neural Network of seven layers. The hyper-parameters of G and D are detailed in Tabs. 1 and 2.

Notice the cGAN approach is currently state-of-the-art in most tasks that can be reduced to an image-to-image translation problem [39]–[42].

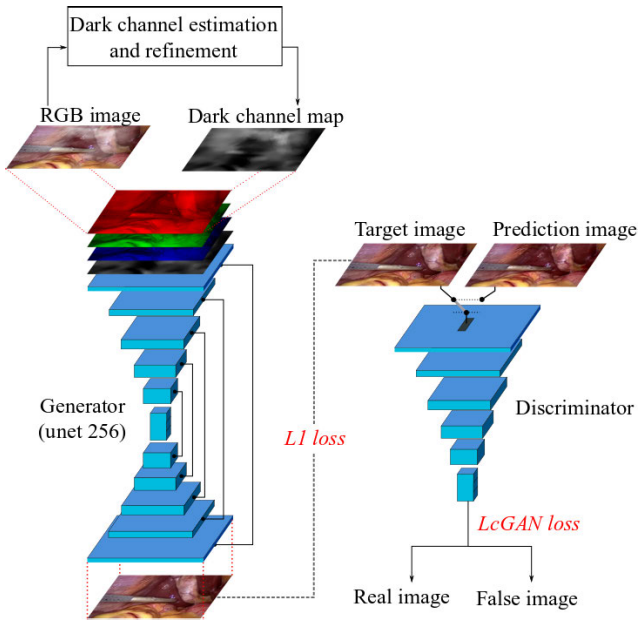

FIGURE 4. Diagram of the proposed approach based on the dark channel estimation and a generative adversarial network. The input is the refined dark channel map stacked into the smoked RGB image. Such RGB image is processed by the *generator* to obtain the de-smoked laparoscopic images. The *discriminator* is applied only in the training stage.

2) OBJECTIVE FUNCTION

This work's loss function is based on a combination of a cGAN's objective function and a $L1$ loss function. As in [23], the objective function of the cGAN is given as

$$\mathcal{L}_{cGAN}(G,D) = \log D(I,J) + \log(1 - D(I, G(I,r))), \tag{4}$$

where G is the *generator*; D is the *discriminator*; I is the smoked image; J is the target image; r is a random noise vector; and $G(I, r)$ are the restored image.

Since the laparoscopic images are only defined in the RGB channels, the fourth channel's output value is only used as a guide, and it is not relevant for the network performance.

[1] U-Net's architecture is similar to an Auto-Encoder. Both of them have two networks: encoder and decoder. The difference between U-Net and AutoEncoder is that U-Net skips connections between encoder layers and decoder layers, while Auto-Encoder does not. The U-Net and Auto-Encoder have a stage where the inputs' features are described: the latent space and bottleneck layer, respectively.

Then, we propose that the $L1$ loss function be focused only on the RGB channels, i.e.

$$\mathcal{L}_{L1}(G) = \sum_{c \in R,G,B} \sum_{x,y} ||J^c - [G(I,z)]^c||, \tag{5}$$

where (x, y) represents the pixel positions; recall that c represents each channel component of the *RGB*.

The final objective function used in this work is based on the previous loss functions, and it is given by

$$G* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G,D) + \lambda \mathcal{L}_{L1}(G), \tag{6}$$

where $\lambda > 0$ is the weight of $L1$ loss function.

III. EXPERIMENTS

This section presents the setup of the used datasets, the architecture, some comparisons, and the quantitative and qualitative obtained results.

A. SETUP

The experiments were implemented in a Ryzen Threadripper processor with 128 GB of RAM and an Nvidia RTX 2080 Ti card. The used operating system is Linux Ubuntu 18.10 endowed with Python 3.7, OpenCV libraries, Keras 2.3.0, Tensorflow 1.15, and Pytorch 1.4.0 framework.

B. DATASETS

The images used in this work were acquired from videos of the Cholec80 dataset [43], which contains 80 videos of cholecystectomy surgeries performed by 13 surgeons. From the available videos, two datasets were created: a train and a test dataset. For generating the training dataset, 20, 000 representative images without smoke were extracted from 50 of the available videos. Synthetic random smoke was added as it is described in Section III-C. This procedure is illustrated in Fig. 5.

Similarly, to measure and compare the performance with the state-of-art methods, a test dataset was created with 2, 398 representative images extracted from the five videos above. It is highlighted that the videos used in the test dataset are different from those used in the training dataset.

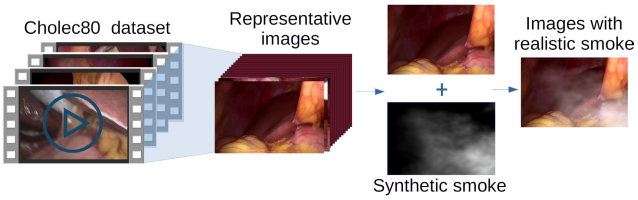

FIGURE 5. The process to generate the dataset.

C. SMOKE SIMULATION

A realistic simulation of heterogeneous smoke is vital to train and test the developed model. In this work, smoke was simulated using Python-Clouds library,[2] which is based on Perlin noise. Each simulation was saved in a normalized synthetic smoke mask $m(x, y)$. Smoke was added to the laparoscopic images using the atmospheric scattering model:

$$I^c(x,y) = J^c(x,y)\, t(x,y) + (1 - t(x,y))A^c, \tag{7}$$

where the transmission map $t(x, y)$ is computed as the complement of $m(x, y)$ as follows

$$t(x,y) = \text{smoke}_{\text{intensity}}(1 - m(x,y)), \tag{8}$$

where c represents each of the *RGB* channels; $\text{smoke}_{\text{intensity}}$, $(0 \leq \text{smoke}_{\text{intensity}} \leq 1)$, is the intensity value; A is a normalized *RGB* atmospheric light color, in which we suppose a white atmospheric light $A = [1 \quad 1 \quad 1]$; $J(x, y)$ is a clear laparoscopic image; and $I(x, y)$ is the image with simulated smoke with dimensions of 256×256 pixels.

[2]Link to Python-Clouds library

TABLE 1. Architecture of the *generator G*. Capital letter *C* denotes (Convolution -> BatchNorm -> Leaky ReLU); *CTD* denotes (Deconvolution -> BatchNorm->ReLU->Dropout rate 50%); and *CT* denotes (Deconvolution -> BatchNorm -> ReLU).

Layer	Conv	Kernel	Strides	Definition	Size	Skip connection
1	64	4	2	C	(128, 128, 64)	→ 16
2	128	4	2	C	(64, 64, 128)	→ 15
3	256	4	2	C	(32, 32, 256)	→ 14
4	512	4	2	C	(16, 16, 512)	→ 13
5	512	4	2	C	(8, 8, 512)	→ 12
6	512	4	2	C	(4, 4, 512)	→ 11
7	512	4	2	C	(2, 2, 512)	→ 10
8	512	4	2	C	(1, 1, 512)	n/a
9	512	4	2	CTD	(1, 1, 1024)	n/a
10	1024	4	2	CTD	(2,2, 1024)	7 →
11	1024	4	2	CTD	(4,4, 1024)	6 →
12	1024	4	2	CT	(8, 8, 1024)	5 →
13	1024	4	2	CT	(16, 16, 1024)	4 →
14	512	4	2	CT	(32, 32, 512)	3 →
15	256	4	2	CT	(64, 64, 256)	2 →
16	128	4	2	CT	(128, 128, 128)	1 →
17	n/a	4	2	tanh	(256, 256, 3)	n/a

D. cGan ARCHITECTURE

For this research, the cGAN is configured as follows. The generator is based on a U-Net architecture with images of input of 256^2 pixels. The discriminator is a Convolutional Neural Network of seven layers. Both architectures, the generator, and discriminator have a stage where the inputs' features are described: the latent space and the bottleneck layer, respectively. In Tabs. 1 and 2 the hyper-parameters of G and D are detailed.

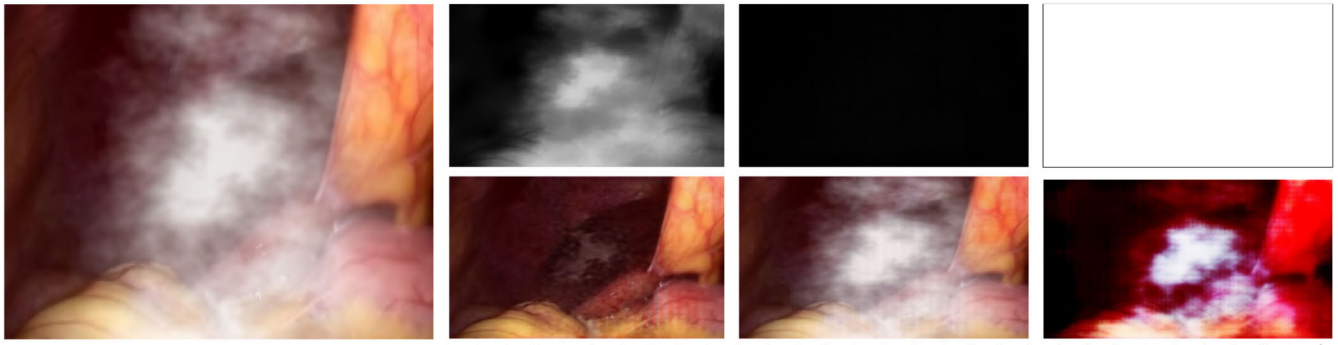

FIGURE 6. Effect of the dark channel-mask on image restoration. (a) Input image with synthetic smoke; (b) the dark channel-mask and its corresponding output from our method; (c) embedded dark channel-mask with modified values to zero and the network's output; and (d) embedded dark channel-mask with values modified to one, and the network's output.

TABLE 2. The architecture of the employed discriminator *D*.

Layer	Conv	kernel	Strides	Definition	Size
1	64	4	2	(Conv ->BatchNorm ->Leaky ReLU)	(128, 128, 64)
2	128	4	2	(Conv ->BatchNorm ->Leaky ReLU)	(64, 64, 128)
3	256	4	2	(Conv ->BatchNorm ->Leaky ReLU)	(32, 32, 256)
4	0	0	0	(ZeroPadding2D)	(34, 34, 256)
5	512	4	1	(Conv)	(31, 31, 512)
6	0	0	0	(BatchNorm ->Leaky ReLU->ZeroPadding)	(33, 33, 512)
7	1	4	1	(Conv)	(30, 30, 1)

E. TRAINING

The proposed method was trained using a synthetic dataset. We used a batch size of 16, a learning rate of 0.0002, and a resolution of 256 × 256 pixels for the input and output images. Weights were initialized using a Gaussian distribution with zero mean and a standard deviation of 0.02. Adaptable Momentum (ADAM) was used as an optimization function. The network was trained for 50 epochs. In particular, epoch 35 showed the best performance in our approach, while epoch 15 showed the best for the pix2pix alone. Pytorch framework was used in each model. The model was trained on two NVIDIA RTX 2080 Ti during one day. More details can be found in our source code.[3]

F. THE DARK CHANNEL EMBEDDED MASK's EFFECT

To corroborate the influence of the dark channel embedded mask in our trained model, we run the proposed network with three different manipulated embedded masks values: 1) dark channel-mask (Fig. 6b); 2) 0's (Fig. 6c); and 3) 1's (Fig. 6d). In Fig. 6, an example of the experiment is shown. As it is expected, for the relation between the dark channel and laparoscopic images, when the dark channel has low values, the restoration tends to be imperceptible (see Fig. 6c). In contrast, when dark channel values are high, the performed restoration saturates the pixels (see Fig. 6d).

[3]Link to our code via github

G. COMPARISONS

To analyze the compared methods' performance under different smoke intensities, we define three smoke levels as a function of the parameter $\text{smoke}_{\text{intensity}}$ in (8): low ($\text{smoke}_{\text{intensity}} \in [0.40, 0.55]$), medium ($\text{smoke}_{\text{intensity}} \in [0.6, 0.75]$), and high ($\text{smoke}_{\text{intensity}} \in [0.80, 0.95]$). The particular value of $\text{smoke}_{\text{intensity}}$ is randomly chosen, provided that such a value belongs to any of the previously defined sets.

To demonstrate the capabilities of our approach, we have compared it against five state-of-the-art methods. The first method is [15]; it uses a convolutional architecture with multi-scale kernels. The second is [19], where the authors use an unsupervised learning approach based on U-Net structure. Next, we compare our result with [17], a CNN-based approach where a Laplacian image pyramid decomposition input strategy is performed. Finally, we compare our work with two approaches based on a cGAN pix2pix: [23], and [24]. These comparisons permit us to observe our proposed impact based on the dark channel embedded mask plus a cGAN.

1) QUALITATIVE COMPARISON

The comparison results against the five previous works are summarized in Fig. 7. From left to right we show in: (a) the nine input images with synthetic CO_2; (b) the respective ground-truth; (c) the results of Bolkar *et al.* [15]; (d) the results of Chen *et al.* [19]; (e) the results of Wang *et al.* [17]; (f) the output of the Pix2pix [23]; (f) the output of Salazar *et al.* [23]; and finally (g), the results of our method.

From the comparisons, it can be seen easily that Bolkar *et al.* [15] (Fig. 7c) present the lowest performance compared to the rest of the works. Chen *et al.* (Fig. 7d), Salazar *et al.* (Fig. 7e), and Isola *et al.* [23] (Fig. 7f) achieved a good reduction of synthetic gas. However, the output images' color is affected. In the results of our method, shown in Fig. 7g), the images are nearly close to the ground-truth since the images preserve the original color and hold more details in the recovered zones.

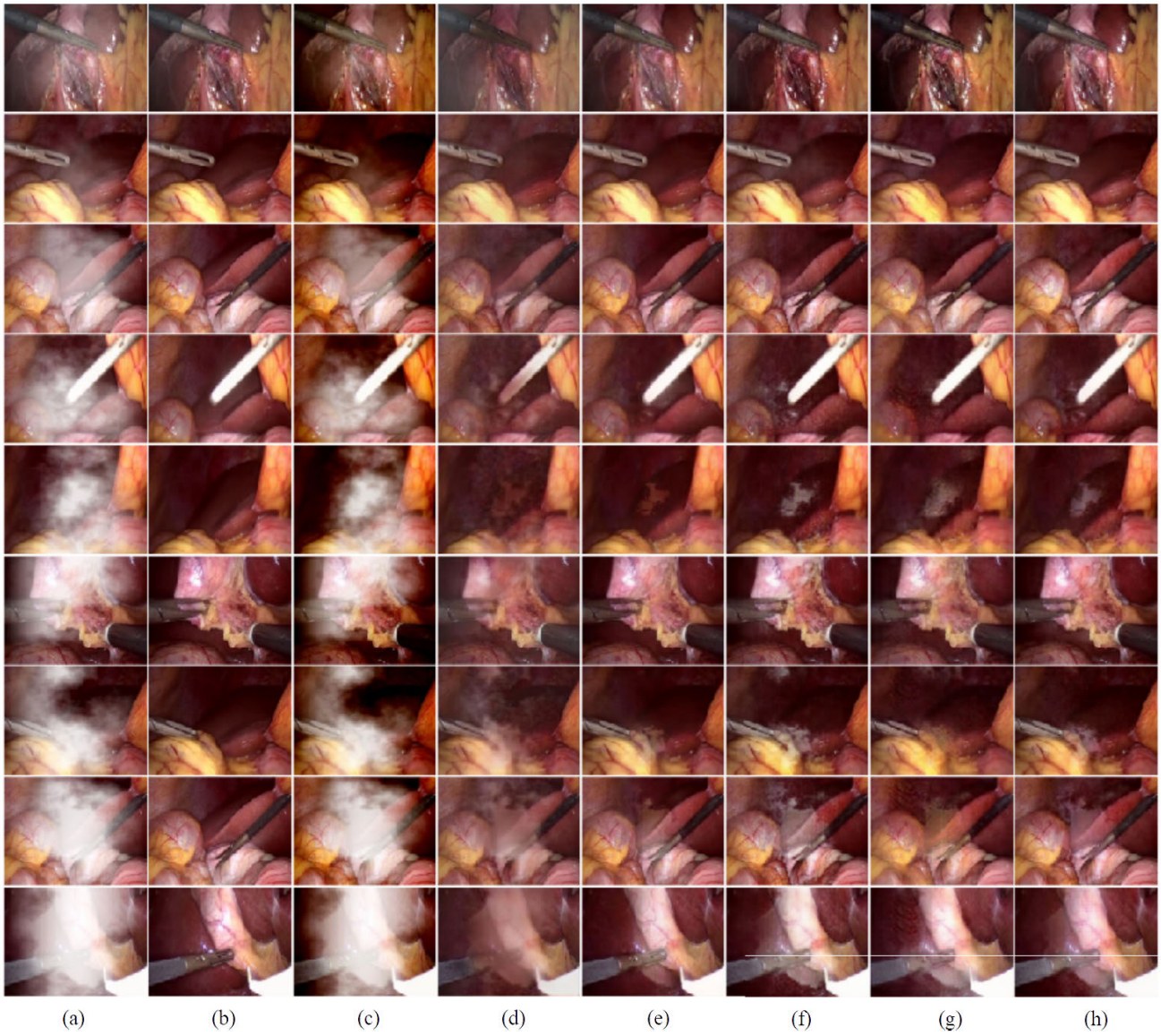

FIGURE 7. Comparison experiments in synthetic laparoscopic images. (a) Input images with synthetic smoke, (b) ground-truth, (c) Bolkar *et al.* [15], (d) Chen *et al.* [19], (e) Wang *et al.* [17], (f) Isola *et al.* [23], (g) Salazar *et al.* [24], and (h) our method.

Real laparoscopic smoked images are depicted in Fig. 8. From left to right, we show: (a) five input images with smoke; and from (b) to (i) the outputs of the current state-of-the-art methods. The method of Bolkar *et al.* [15] seems to mitigate the effect of the gas marginally. Nevertheless, it presents a major color saturation where darker regions look black, as one can see in Fig. 7b). For the second group, Chen *et al.* [19] (Fig. 7c) show a slight improvement over the methods of the first group; however, it presents opaque colors and poorly defined textures. Wang *et al.* [17] (Fig. 7d), Pix2pix [23], (Fig. 7e), and Salazar *et al.* [24], (Fig. 7f) show satisfactory results in terms of gas reduction, color, and object details. Finally, our method (Fig. 7g) presents a minor gas influence among all the images, preserving fine details and the original colors.

The results mentioned above are consistent with the quantitative analysis presented next.

2) QUANTITATIVE COMPARISON

Results of compared methods are quantitative evaluated using four metrics commonly used in the literature: peak signal-to-noise ratio (PSNR), structural similarity (SSIM) index, chrominance feature similarity (FSIMc), and universal quality index (UQI).

PSNR is a quantitative measure of the restoration quality and is defined as

$$\mathrm{PSNR}(I, J) = 10\log_{10}\left(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}}\right) = 20\log_{10}\left(\frac{\mathrm{MAX}_I}{\sqrt{\mathrm{MSE}}}\right), \quad (9)$$

where $\mathrm{MAX} = 2^B - 1$ in which B is the number of bits used in the image, and MSE is the mean squared error. The MSE

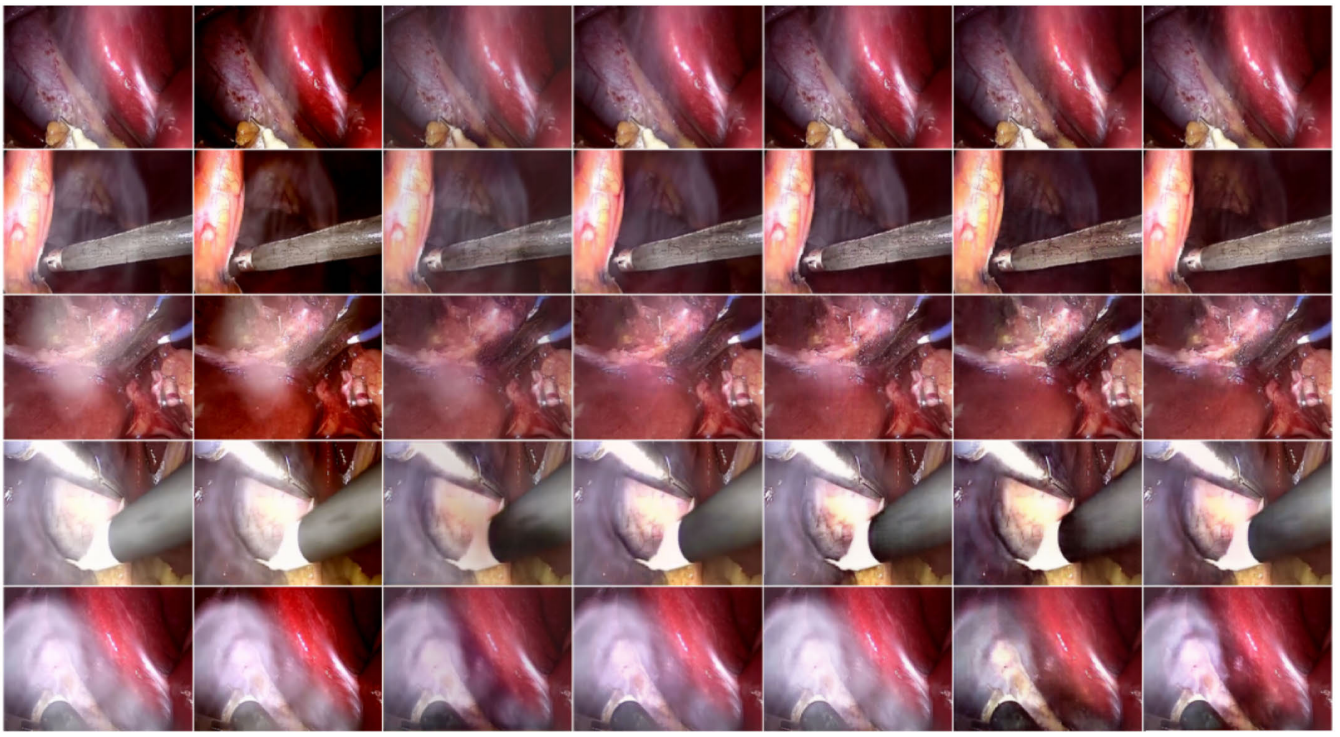

FIGURE 8. Comparison with five state-of-the-art methods in real laparoscopic images. (a) Input images with smoke, (b) Bolkar *et al.* [15], (c) Chen *et al.* [19], (d) Wang *et al.* [17], (e) Isola *et al.* [23], (f) Salazar *et al.* [24], and (g) our method.

for two monochrome images, I and J of size $m \times n$, is given by

$$\text{MSE}(I, J) = \frac{1}{mn} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} ||I(x, y) - J(x, y)||^2. \quad (10)$$

High values of PSNR indicate better restorations. The MPEG committee establishes an informal threshold of PSNR $= 0.5$ decibels (dB) to decide whether a PSNR value is significant, which means that differences in quality are visible [44].

SSIM index is a perceptual image similarity metric. It is proposed as an alternative to MSE and PSNR indexes since it has been shown to perform better in the subjective assessment. SSIM index is defined between -1 and 1, where -1 represents a total anti-correlation, 0 no correlation, and 1 a total correlation between images. For the original and reconstructed images I and J, SSIM index is defined as follows

$$\text{SSIM}(I, J) = \frac{(2\mu_I\mu_J + C_1)(2\sigma_{IJ} + C_2)}{(\mu_I^2 + \mu_J^2 + C_1)(\sigma_I^2 + \sigma_J^2 + C_2)}, \quad (11)$$

where μ is the mean; σ is the variance; σ_{IJ} is the images' covariance; and C_1, C_2 are two variables to stabilize the division with a weak denominator.

The Feature Similarity Index for Image Quality Assessment (FSIM) is based on two criteria: the phase congruency (PC) that reflects the image behavior in the frequency domain, and the gradient magnitude (GM) that reflects the maximum intensity variation regardless of orientation [45]. FSIM is defined as

$$\text{FSIM}(I, J) = \frac{\sum_{x,y} S_L PC_m}{\sum_{x,y} PC_m}, \quad (12)$$

where S_L is the combination of similarities between PC_I and PC_J, namely $S_{PC}(I, J)$; and between GM_I and GM_J, namely $S_{GM}(I, J)$; and $PC_m = max(PC_I, PC_J)$. On the other hand, the version of FSIM oriented to color images FSIM_C is defined as:

$$\text{FSIM}_C(I, J) = \frac{\sum_{x,y} S_L [S_{Lu} S_{Ch}]^{\lambda} PC_m}{\sum_{x,y} PC_m}, \quad (13)$$

where λ is the parameter used to adjust the importance of the chromatic components; Lu is the luminance; (Ch) is the chrominance; and S_{Lu}, S_{Ch} are the similarity maps of Lu and Ch, respectively.

The UQI (universal quality index) is designed by modeling any image distortion as a combination of three factors: loss of correlation, luminance distortion, and contrast distortion. For original I and test images J, UQI is defined as

$$\text{UQI}(I, J) = \frac{4\sigma_{IJ}\overline{I}\,\overline{J}}{(\sigma_I^2 + \sigma_J^2)[(\overline{I})^2 + (\overline{J})^2]}, \quad (14)$$

where $\overline{I}$, and $\overline{J}$ are the mean of original and test images, respectively; σ_I^2, and σ_J^2 are the variances, and σ_{IJ} is the standard deviation [46].

Tab. 3 shows the overall average of the SSIM index, PSNR, FSIM_C, and UQI, where our method yields the values of

0.93, 27.63 dB, 0.97, and 0.95, respectively. According to Tab. 3 the average of our approach outperforms state-of-the-art methods, except for the UQI metric, where our approach performance is equaled by Wang *et al.* method [17].

TABLE 3. Performance comparison of the state-of-the-art methods according to SSIM index, PSNR (in dB), FSIMc, and UQI average metrics.

Methods	SSIM index	PSNR	$FSIM_C$	UQI
Bolkar et al. [15]	0.64	18.34	0.92	0.63
Chen et al. [19]	0.77	20.93	0.91	0.86
Wang et al. [17]	0.87	27.38	0.95	**0.95**
Isola et al. [23]	0.91	27.10	0.96	0.94
Salazar et al. [24]	0.80	24.50	0.92	0.89
Our proposed method	**0.93**	**27.63**	**0.97**	**0.95**

To further analyze our method's performance under different smoke intensities, in Figs. 9, 10, 11, and 12 the SSIM index, PSNR, FSIMc, and UQI metrics are shown for three different densities of synthetic smoke: (a) low; (b) medium; and (c) high density.

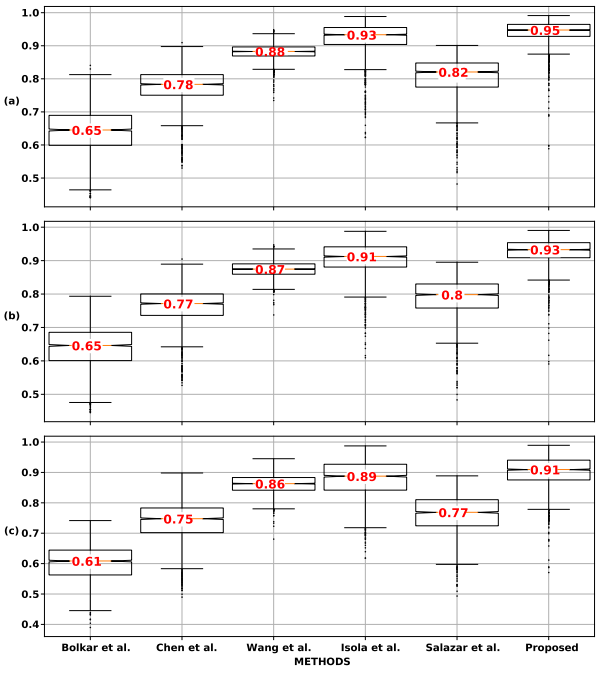

FIGURE 9. Performance comparison of the state-of-the-art methods according to the SSIM index on different smoke intensities: (a) low, (b) medium, and (c) high.

The reader can see that our proposed method presents the highest values in all the smoke conditions: low, medium, and high for SSIM index, PSNR, and $FSIM_C$; except in case of low smoke for UQI and high smoke for PSNR and UQI. In such cases, the algorithm developed by Wang *et al.* [17] presents values of 0.96, 26.75 dB, and 0.95, while our method has a value of 0.95, 26.29 dB, and 0.94. We highlight the dispersion reduction in our method for the three smoke

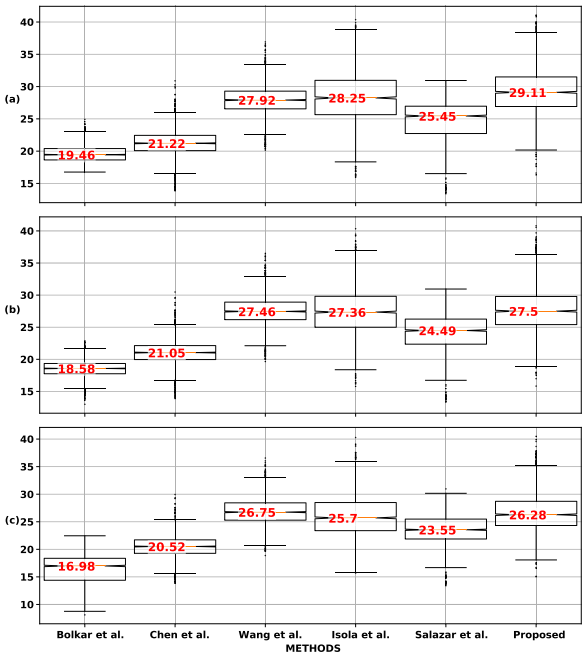

FIGURE 10. Performance comparison of the state-of-the-art methods according to the PSNR on different smoke intensities: (a) low, (b) medium, and (c) high.

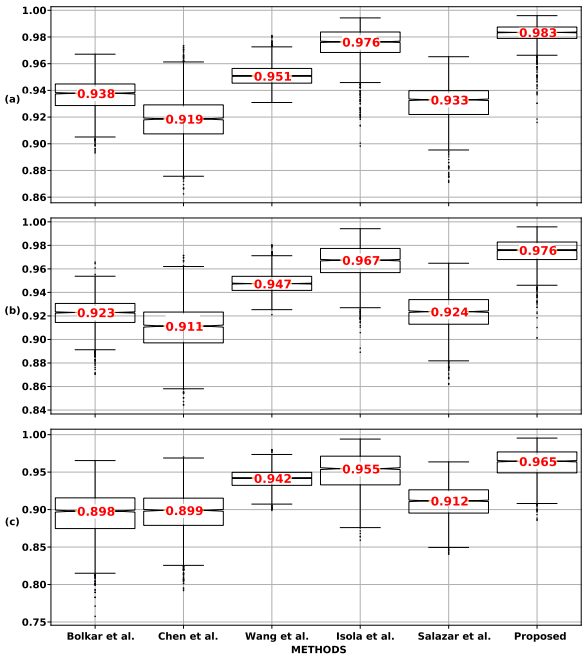

FIGURE 11. Performance comparison of the state-of-the-art methods according to the $FSIM_c$ on different smoke intensities: (a) low, (b) medium, and (c) high.

intensities w.r.t. the pix2pix results. This can be seen in the metrics of SSIM index, PSNR, FSIMc, and UQI shown in Figs. 9, 10, 11, and 12, respectively.

3) PROCESSING SPEED COMPARISON

Since the main goal is to support surgeons during laparoscopic surgery, the processing speed is crucial. Tab. 4 shows

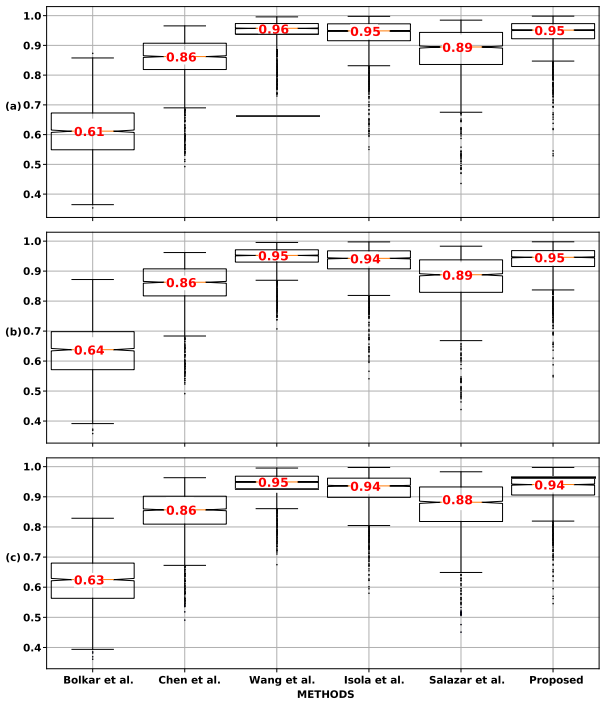

FIGURE 12. Performance comparison of the state-of-the-art methods according to the UQI on different smoke intensities: (a) low, (b) medium, and (c) high.

TABLE 4. Processing speed comparison in frames per second (fps).

Methods	Processing speed (fps)	Platform
Bolkar et al. [15]	32.40	Python (Caffe)
Chen et al. [19]	89.14	Python (Tensor Flow)
Wang et al. [17]	24.00	Python (Keras)
Isola et al. [23]	120.0	Python (Pytorch)
Salazar et al. [23]	90.0	Python (Pytorch)
Our proposed method	92.19	Python (Pytorch)

the compared methods and their respective average processing time in frames per second (fps). Since our method computes the dark channel, the processing speed is a little higher than the pix2pix based (Isola *et al.*), in which it is evident that the dark channel is not computed. This is shown in Tab. 4. It is also noted in Tab. 4 that our method has a better performance than the state-of-art methods. Since our method outperforms by three times the typical video frame rate of 30 fps in the experimental platform, it is possible to implement our algorithm in a real-time system and even in a low-cost device. We highlight that our method has not been optimized. In [4] it is shown a demonstrative video of the proposed method in real-world laparoscopic images.

[4]Link to a demonstrative video of our method.

IV. CONCLUSION

The lack of visibility in laparoscopic surgery caused by CO_2 increases the surgical operation time and the possibility of surgery errors. Therefore, a method to remove the smoke effects is necessary to support the surgeons, thus reducing risks for the patient and increasing medical specialists' efficiency. In this work, it is proposed the use of the dark channel embedded in the input of a *generator* network.

The addition of a dark-channel-guide contributes to identifying regions with the presence of smoke, focusing on the restoration of those regions. Qualitative evaluations in synthetic and real images showed that this new approach generates output images with better contrast and color preservation than the other five state-of-the-art methods. On the other hand, quantitative evaluations based on PSNR and SSIM index for three different synthetic smoke densities showed that the proposed approach outperformed all the five algorithms mentioned in the paper. Moreover, the processing time of 92.19 frames per second shows that the proposed architecture can be easily implemented into a system for real-time applications used in medical devices.

One of the main limitations of the proposed approach is the quality and amount of information incorporated into the embedded channel since if the additional channel does not provide reliable information, the network's performance could be affected.

As shown in the manuscript, incorporating a dark channel embedded in a cGAN network increases its performance by slightly reducing its speed. However, the concept of an embedded channel could be applied to other kinds of neural network architectures and even other approaches, such as classification or even regression tasks.

The work of Hugo Alberto Moreno was supported by the National Council for Science and Technology (CONACYT) for his master's studies.

REFERENCES

[1] S. Sauerland, T. Jaschinski, and E. A. Neugebauer, "Laparoscopic versus open surgery for suspected appendicitis," *Cochrane Database Systematic Rev.*, no. 11, pp. 1–164, 2018.

[2] L. Gu, P. Liu, C. Jiang, M. Luo, and Q. Xu, "Virtual digital defogging technology improves laparoscopic imaging quality," *Surgical Innov.*, vol. 22, no. 2, pp. 171–176, Apr. 2015.

[3] K. Y. Hahn, D. W. Kang, Z. A. M. Azman, S.-Y. Kim, and S.-H. Kim, "Removal of hazardous surgical smoke using a Built-in-Filter trocar: A study in laparoscopic rectal resection," *Surgical Laparoscopy, Endoscopy Percutaneous Techn.*, vol. 27, no. 5, pp. 341–345, Oct. 2017.

[4] N. Lawrentschuk, N. E. Fleshner, and D. M. Bolton, "Laparoscopic lens fogging: A review of etiology and methods to maintain a clear visual field," *J. Endourology*, vol. 24, no. 6, pp. 905–913, Jun. 2010.

[5] J. H. Goodson, M. M. Judy, and R. A. Moses, "Laser smoke evacuation system and method," U.S. Patent 4 735 603, Apr. 5, 1988.

[6] T. R. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P. K. R. Maddikunta, I.-H. Ra, and M. Alazab, "Early detection of diabetic retinopathy using PCA-firefly based deep learning model," *Electronics*, vol. 9, no. 2, p. 274, Feb. 2020.

[7] A. Kotwal, R. Bhalodia, and S. P. Awate, "Joint desmoking and denoising of laparoscopy images," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 1050–1054.

[8] A. Baid, A. Kotwal, R. Bhalodia, S. N. Merchant, and S. P. Awate, "Joint desmoking, specularity removal, and denoising of laparoscopy images via graphical models and Bayesian inference," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 732–736.
[9] X. Luo, A. J. McLeod, S. E. Pautler, C. M. Schlachta, and T. M. Peters, "Vision-based surgical field defogging," *IEEE Trans. Med. Imag.*, vol. 36, no. 10, pp. 2021–2030, Oct. 2017.
[10] C. Wang, F. Alaya Cheikh, M. Kaaniche, and O. Jacob Elle, "A smoke removal method for laparoscopic images," 2018, *arXiv:1803.08410*. [Online]. Available: http://arxiv.org/abs/1803.08410
[11] J. Shin, M. Kim, J. Paik, and S. Lee, "Radiance–reflectance combined optimization and structure-guided ℓ_0-norm for single image dehazing," *IEEE Trans. Multimedia*, vol. 22, no. 1, pp. 30–44, Jan. 2020.
[12] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3522–3533, Nov. 2015.
[13] K. Tchaka, V. M. Pawar, and D. Stoyanov, "Chromaticity based smoke removal in endoscopic images," *Proc. SPIE*, vol. 10133, Feb. 2017, Art. no. 101331M.
[14] C. Wang, F. Alaya Cheikh, M. Kaaniche, A. Beghdadi, and O. J. Elle, "Variational based smoke removal in laparoscopic images," *Biomed. Eng. OnLine*, vol. 17, no. 1, p. 139, Dec. 2018.
[15] S. Bolkar, C. Wang, F. A. Cheikh, and S. Yildirim, "Deep smoke removal from minimally invasive surgery videos," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 3403–3407.
[16] O. Sidorov, C. Wang, and F. A. Cheikh, "Generative smoke removal," 2019, *arXiv:1902.00311*. [Online]. Available: http://arxiv.org/abs/1902.00311
[17] C. Wang, A. K. Mohammed, F. A. Cheikh, A. Beghdadi, and O. J. Elle, "Multiscale deep desmoking for laparoscopic surgery," *Proc. SPIE*, vol. 10949, Mar. 2019, Art. no. 109491Y.
[18] V. Vishal, N. Sharma, and 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*, 2019, pp. 21–28.
[19] L. Chen, W. Tang, and W. N. John, "Unsupervised learning of surgical smoke removal from simulation," in *Proc. Hamlyn Symp. Med. Robot.*, Jun. 2018, pp. 1–2.
[20] A. Singh, A. Chandra, R. Kumar, K. Singh, and N. Dey, "Dark channel processing for medical image enhancement," in *Proc. IEEE Int. WIE Conf. Electr. Comput. Eng. (WIECON-ECE)*, Nov. 2019, pp. 1–6.
[21] N. Dey, "Uneven illumination correction of digital images: A survey of the state-of-the-art," *Optik*, vol. 183, pp. 483–495, Apr. 2019.
[22] V. Vishal, V. Venkatesh, K. Lochan, N. Sharma, and M. Singh, "Unsupervised desmoking of laparoscopy images using multi-scale desmokenet," in *Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst.* Cham, Switzerland: Springer, 2020, pp. 421–432.
[23] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.
[24] S. Salazar-Colores, H. Alberto-Moreno, G. Flores, and C. J. Ortiz-Echeverri, "Laparoscopy surgery CO_2 removal via generative adversary network and dark channel prior," 2019, *arXiv:1909.12314*. [Online]. Available: https://arxiv.org/abs/1909.12314
[25] L. Chen, W. Tang, N. W. John, T. R. Wan, and J. J. Zhang, "De-smokeGCN: Generative cooperative networks for joint surgical smoke detection and removal," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1615–1625, May 2020.
[26] D. W. Birch, J. T. Dang, N. J. Switzer, N. Manouchehri, X. Shi, G. Hadi, and S. Karmali, "Heated insufflation with or without humidification for laparoscopic abdominal surgery," *Cochrane Database Systematic Rev.*, no. 10, pp. 1–85, 2016.
[27] M. Perrin and A. Fletcher, "Laparoscopic abdominal surgery," in *Continuing Education in Anaesthesia Critical Care & Pain*, vol. 4. New York, NY, USA: Oxford, Aug. 2004, pp. 107–110.
[28] Y. Ren, Z. Zhu, Y. Li, and J. Lo, "Mask embedding in conditional GAN for guided synthesis of high resolution images," 2019, *arXiv:1907.01710*. [Online]. Available: http://arxiv.org/abs/1907.01710
[29] G. Yildirim, C. Seward, and U. Bergmann, "Disentangling multiple conditional inputs in GANs," 2018, *arXiv:1806.07819*. [Online]. Available: http://arxiv.org/abs/1806.07819
[30] S. Gu, J. Bao, H. Yang, D. Chen, F. Wen, and L. Yuan, "Mask-guided portrait editing with conditional GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3436–3445.
[31] Y. Xu, J. Wen, L. Fei, and Z. Zhang, "Review of video and image defogging algorithms and related studies on image restoration and enhancement," *IEEE Access*, vol. 4, pp. 165–188, 2016.
[32] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.
[33] F. Murabito, C. Spampinato, S. Palazzo, D. Giordano, K. Pogorelov, and M. Riegler, "Top-down saliency detection driven by visual classification," *Comput. Vis. Image Understand.*, vol. 172, pp. 67–76, Jul. 2018.
[34] A. Chandra, A. Singh, R. Kumar, and N. Dey, "Dehazing of aerial images by dark channel and gamma correction," in *Proc. 3rd Int. Conf. Workshops Recent Adv. Innov. Eng. (ICRAIE)*, Nov. 2018, pp. 1–7.
[35] J. Pang, O. C. Au, and Z. Guo, "Improved single image Dehazing using guided filter," in *Proc. APSIPA ASC*, Oct. 2011, pp. 1–4.
[36] S. Lee, S. Yun, J.-H. Nam, C. S. Won, and S.-W. Jung, "A review on dark channel prior based image Dehazing algorithms," *EURASIP J. Image Video Process.*, vol. 2016, no. 1, p. 4, Dec. 2016.
[37] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Computer Vision—ECCV*, K. Daniilidis, P. Maragos, and N. Paragios, Eds. Berlin, Germany: Springer, 2010, pp. 1–14.
[38] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[39] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4401–4410.
[40] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3943–3956, Nov. 2020.
[41] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, G. Liu, A. Tao, J. Kautz, and B. Catanzaro, "Video-to-video synthesis," 2018, *arXiv:1808.06601*. [Online]. Available: http://arxiv.org/abs/1808.06601
[42] Y. Zhang, Z. Gan, K. Fan, Z. Chen, R. Henao, D. Shen, and L. Carin, "Adversarial feature matching for text generation," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 4006–4015.
[43] A. P. Twinanda, S. Shehata, D. Mutter, J. Marescaux, M. de Mathelin, and N. Padoy, "EndoNet: A deep architecture for recognition tasks on laparoscopic videos," *IEEE Trans. Med. Imag.*, vol. 36, no. 1, pp. 86–97, Jan. 2017.
[44] D. Salomon, *Data Compression: Complete Reference*. New York, NY, USA: Springer, 2004.
[45] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
[46] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

SEBASTIÁN SALAZAR-COLORES received the Ph.D. degree in computer science from the Universidad Autónoma de Querétaro, in 2019. He is currently working with the Centro de Investigaciones en Óptica, as a Researcher Associate. His research interests include deep learning, computer vision, and signal processing.

HUGO MORENO JIMÉNEZ received the Robotics Engineering degree from the Universidad Politécnica del Bicentenario, Guanajuato, Mexico, in 2016. He is currently pursuing the M.S. degree in optomecatronic with the Centro de Investigaciones en Óptica, A.C. His research interests include digital image processing and automation.

CÉSAR JAVIER ORTIZ-ECHEVERRI (Member, IEEE) received the bachelor's degree in electronic engineering and the M.Sc. degree in material science from the Universidad del Quindío, Colombia, in 2015, and the Ph.D. degree in computer science from the Universidad Autónoma de Querétaro, in 2020. His research interests include digital image, and signal processing mainly focused on musical information retrieval, and biomedical applications.

GERARDO FLORES received the B.S. degree (Hons.) in electronic engineering from the Instituto Tecnológico de Saltillo, Mexico, in 2007, the M.S. degree in automatic control from CINVESTAV-IPN, Mexico, in 2010, and the Ph.D. degree in systems and information technology from the Heudiasyc Laboratory, Université de Technologie de Compiègne-Sorbonne Universités, France, in October 2014. Since August 2016, he has been a full-time Researcher and the Head of the Perception and Robotics LAB with the Center for Research in Optics, León, Mexico. His current research interests include theoretical and practical problems arising from the development of autonomous robotic and vision systems. Since 2020, he has been an Associate Editor of *Mathematical Problems in Engineering*.