Guided Unsupervised Desmoking of Laparoscopic Images Using Cycle-Desmoke

V. Vishal[1], Neeraj Sharma[2], and Munendra Singh[1(✉)]

[1] Department of Mechatronics Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, Karnataka, India
vishalv7197@gmail.com, munendra107@gmail.com
[2] School of Biomedical Engineering,
Indian Institute of Technology (Banaras Hindu University),
Varanasi 221005, India
neeraj.bme@iitbhu.ac.in

Abstract. The generation of smoke in laparoscopic surgery due to laser ablation and cauterization causes deterioration in the visual quality of the operative field. In order to reduce the effect of smoke, the present paper proposes an end-to-end network, called Cycle-Desmoke. The network enhances the CycleGAN framework by adoption of a new generator architecture and addition of new Guided-Unsharp Upsample loss in combination to adversarial and cycle-consistency loss. The Atrous Convolution Feature Extraction Module present in the encoder blocks of the generator helps distinguishing smoke by capturing features at multiple scales by the use of kernels with different receptive fields. Further, the use of Guided-Unsharp Upsample loss supervises the upsampling process of the feature maps and helps improve the contrast of the desmoked image. The network performs robust unsupervised Image-to-Image Translation from smoke domain to smoke-free domain. The public Cholec80 dataset is used to evaluate the performance of the proposed method. Quantitative and qualitative comparative analysis of the proposed method over the state-of-the-methods reveals the effectiveness of the method at the task of smoke removal and enhancement of the image.

Keywords: Smoke removal · Image enhancement · Laparoscopic surgery

1 Introduction

In laparoscopic surgery, the visualization of the operative field is of great utility for the surgeon as well as for the computer-assistive algorithms such as segmentation and detection of different tissues and surgical instruments. The generation of several artefacts such as noise, abrupt illumination changes, specular reflections and smoke in laparoscopic surgery degrades the quality of the visualization and

hampers the efficiency of the surgeons and image-guided navigational systems. In this proposed work, we focus on the task of smoke removal in laparoscopy images using a computational, data driven approach. There exist several smoke evacuation techniques [1] that help to remove smoke but they rely on additional hardware instalments and have constraints. Our method directly works on the image data and helps improve the visualization by removing the smoke component and leads to enhancement of the image.

There exist several methods for the task of desmoking in laparoscopic images. Conventional methods represented the problem to be similar to dehazing, defogging and adopted the atmospheric scattering model to represent the phenomenon. He et al. [2] proposed a single image dehazing method using dark channel prior. The prior information utilized was based on the occurrence of some pixels in the local patches whose intensity are to be very low in at least one colour channel. In [3], Wang et al. assumed smoke to have low inter-channel and low contrast and proposed a variational method to estimate the smoke veil for desmoking. Although these methods led to enhancement of the image, they still lacked the ability to semantically distinguish and remove the smoke component in an image robustly. In recent times,several Deep learning methods have been proposed [5–8] for desmoking. Kotwal et al. [4] proposed a deep learning approach for desmoking on a synthetically generated dataset by transfer learning the task of smoke removal by using the AOD-Net. Wang et al. [7] performed desmoking by proposing a new Laplacian image pyramid decomposition input strategy on a synthetic dataset and evaluated the performance of the method on real smoke dataset. These new methods have outperformed the conventional methods by adopting the data-driven approach.

The proposed work focuses on translating a laparoscopy image from smoke domain to smoke-free domain. The method enhances the CycleGAN [9] framework used for unsupervised Image-to-Image Translation. The main contributions of the work are:

1. A new generator architecture that consists of Atrous Convolution Feature Extraction Module (ACFEM) at each encoder block, that helps to capture features at multiple scales by the use of kernels of different receptive fields. The upsampling operation is performed by means of pixel shuffle, leading to efficient transfer of the features in the network.
2. The use of unsharp images of the smoke images in the Guided-Unsharp Upsample loss in addition to the adversarial and cycle-consistency loss helps supervise the upsampling operation and also helps in contrast enhancement of the desmoked image.
3. The proposed end-to-end network performs unsupervised Image-to-Image translation from smoke to smoke-free domain in an unpaired manner without the need for synthetic ground truth data, hence removing the need for simulation to real-world domain adaptation.

2 Method

This section explains the loss function and network architectures of the generator and the discriminator. The Cycle-Desmoke framework is derived from the CycleGAN [9] framework. It consists of two generator networks G_S, G_{D_S} that generate synthetic smoke and desmoke images respectively and two discriminator network D_S and D_{D_S} that help to distinguish the synthetic smoke and desmoked images from the real smoke and smoke-free images respectively.

2.1 Guided-Unsharp Upsample Loss

The CycleGAN architecture utilizes the adversarial and cycle-consistency losses to perform unpaired image-to-image translation. The adversarial loss helps produce images of high perceptual quality by adopting a min-max optimization between the generator and discriminator networks. While, the cycle-consistency loss employs a *L1-norm* between the input and the reconstructed images to constrain the generated synthetic images to match the desired domain. Although these losses bring about image translation, they do not utilize the features information in the network. Hence, to guide the features in the network we introduce the Guided-Unsharp Upsample loss.

The upsampling operation helps realise desired dimensions for a feature map after the feature map has undergone reduction in spatial dimension after certain number of downsampling operations. In the proposed work, we utilize pixel shuffle to upsample the feature maps. A supervision for the upsampling operation is of great utility as it helps guide the network to accurately predict the desired image. This also helps in refinement of features in the upsampling operations. The unsharp masking technique helps increase the high frequency components and sharpens the images, highlighting fine details and edges. As smoke reduces the contrast of the image, unsharp masking works as a local contrast enhancement technique. The formulation of the technique is given as:

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

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

Where $f(x,y)$ is the image, $f_{smooth}(x,y)$ is the image obtained after smoothing/blurring by convolution operation and $g(x,y)$ is the image that contains high frequency information. The $f_{sharp}(x,y)$ image is realised on adding original image with weighted $g(x,y)$ with amount k. The enhancement of contrast during the process desmoking helps reduce the low contrast smoke component. Hence, the comparison of unsharp images with the prediction from different decoder levels, help guide the upsampling process. This loss is applicable to the generator network responsible for desmoking. The loss is given as:

$$L_{GUU}(G_{DS}) = \sum_j ||Y_d - Y_{us}|| \tag{3}$$

where, Y_d and Y_{us} are the prediction and unsharp images at particular decoder block j.
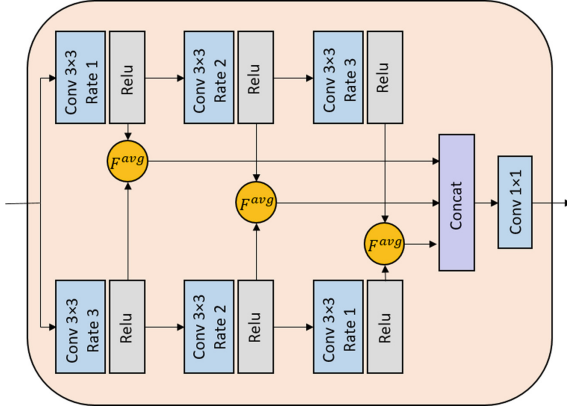
Fig. 1. Representation of Atrous Convolution Feature Extraction Module. The rate denotes the dilation rate of the convolution operation.

2.2   Aggregate Loss Function

The Cycle-Desmoke has an additional loss to the loss function in CycleGAN architecture. The loss function of CycleGAN is denoted as $L_{CycleGAN}(G_S, G_{DS}, D_S, D_{DS})$.

The complete loss function for training the framework is given as:

$$L(G_S, G_{DS}, D_S, D_{DS}) = L_{CycleGAN}(G_S, G_{DS}, D_S, D_{DS}) + \\ \alpha \times L_{GUU}(G_{DS}) \tag{4}$$

The term α controls the effect of the Guided-Unsharp Upsample loss.

2.3   Atrous Convolution Feature Extraction Module

The use of atrous convolutions to control the receptive field has resulted in remarkable results at tasks like semantic segmentation [10] and object detection [11]. The atrous convolutions allows to vary the dimension of the receptive field of the kernel without increasing the number of parameters as it pads zeros between kernel values. In context of the smoke removal problem, the occurrence of smoke can be either heterogeneous or homogeneous in the image, hence a robust feature extraction to capture features at multiple scales is essential. The Atrous Convolution Feature Extraction Module (ACFEM) employs a convolution 3×3 kernel with three different rates of dilation, i.e 1,2 and 3 and the receptive field of the atrous kernels match the dimension of 3×3, 5×5 and 7×7 kernels respectively. Figure 1, pictorially represents the flow of the feature maps across the different atrous convolutions. The flow of features across two branches, one with reducing receptive field and the other with increasing receptive field, helps in capturing a diverse set of features and the F_{avg}, average of the input feature maps helps to obtain the optimal features from both the branches. Convolution 1×1 kernel is
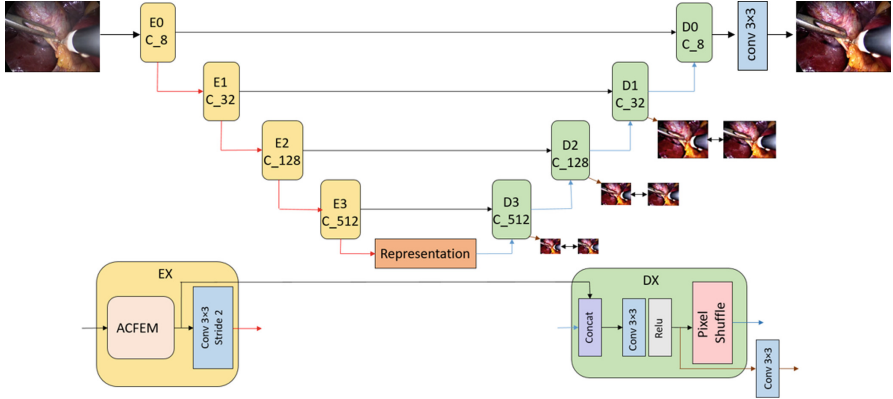
Fig. 2. The representation of the Generator network G_{DS} responsible for desmoking. G_S is similar to G_{DS} with the exception of the Guided-Unsharp Upsample loss setup

used to control the dimension of the output feature map. If the channel dimension of the input feature map is M, the atrous convolutions and convolution 1×1 kernel maintain the same channel depth and the output feature map has channel dimension of M. Hence, ACFEM helps in capturing features effective at distinguishing the smoke component in the image and performs efficient feature extraction.

### 2.4 Generator and Discriminator Networks

The generator network is represented in Fig. 2. It consists of an encoder-decoder structure. Each encoder block consists of Atrous Convolution Feature Extraction Module (ACFEM) and a 3×3 convolution with stride 2 to downsample the feature map by a factor of two. There exists four encoder blocks and a deep representation bottleneck followed by four decoder blocks. Corresponding encoder and decoder blocks are connected via skip connections. The feature map at a decoder block after convolution operation proceeds to pixel shuffle [12] for upsampling and convolution operation that outputs a prediction image that gets compared with the unsharp image at each decoder block except the last one.

The discriminator in CycleGAN [9] is utilized as the discriminator network for Cycle-Desmoke. The network utilizes 70×70 overlapping image patches to distinguish smoke images from smoke-free images.

3 Experimentation and Results

3.1 Dataset and Implementation Details

The dataset [13] used for the present study consists of 100K smoke/non-smoke images extracted from the Cholec80 dataset [14]. The training and test set consists of 1200 and 200 unpaired set of smoke and smoke-free images. The dimension of the image is maintained as 240×320 in order to remove the black corner
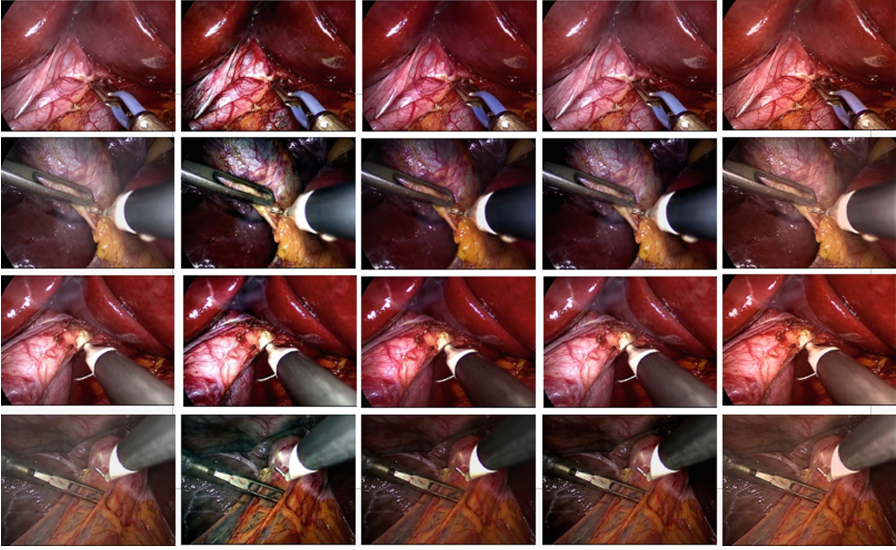
Fig. 3. Qualitative evaluation of smoke removal on randomly selected images from test set. First column: smoke images, second column: Non-Local Dehaze, third column: DCP, fourth column: DehazeNet, fifth column: Proposed method

details and enable the network to learn only the information in the operative field. The images in the smoke domain contains smoke of varying levels and depth and this ensures the network learns on different smoke levels.

The network is end-to-end trained with a learning rate of 0.0001 for the first 100 epochs and then the learning rate is linearly decayed to zero till 200 epochs. ADAM optimizer is used to optimize the generator and discriminator networks. The term α is set to 0.5 in the loss function. The convolution kernel dimension used for obtaining the unsharp images is 9×9 and the amount of sharpening i.e term k is set to 1.5. The tensorflow framework was used to train the network on a single NVidia Tesla T4 GPU.

3.2   Results

The qualitative and quantitative comparative analysis of the proposed Cycel-Desmoke is performed with state-of-the-art methods like Non-Local Dehaze [15], Dark channel prior (DCP) [2] and DehazeNet [16]. It is observed that the state-of-the-art methods although remove smoke to a certain extent, they lack the ability to maintain the colour consistency with respect to the smoke-free domain. The Non-Local Dehaze over saturates the color, causing difficulty in accurate differentiation of tissues. On the other hand, DCP seems to produce lower contrast images compared to the proposed method, while DehazeNet fails at removing smoke that is heterogeneous in nature. Hence, the lacking capabilities of other methods is efficiently handled by the proposed Cycle-Desmoke, that generates

images with good contrast, colour consistency and robust smoke removal for both heterogeneous and homogeneous smoke (Fig. 3). The quantitative metrics used to denote the performance of smoke removal are, BRISQE [17]: Blind/Referenceless Image Spatial Quality Evaluator, PIQUE [18]: Perception-based Image QUality Evaluator, and CEIQ [19]: Quality Assessment of Contrast-Distorted Images. Lower values of BRISQE and PIQUE, higher values of CEIQ denote better image quality. It is evident from Table 1, that the proposed method obtains the best metric values and outperforms the other methods.

Table 1. Quantitative evaluation of smoke removal of other methods and the proposed method. The analysis has been performed on the test set. SD denotes Standard deviation

Method	Nonlocal Dehaze			DCP			DehazeNet			Proposed		
Image quality	BRISQE	PIQUE	CEIQ	BRISQE	PIQUE	CEIQ	BRISQE	PIQUE	CEIQ	BRISQE	PIQUE	CEIQ
Mean	19.06	36.18	3.15	20.95	32.54	2.95	21.07	33.78	3.15	17.71	25.71	3.29
SD	5.64	4.63	0.14	5.40	5.25	0.14	5.26	5.67	0.16	5.64	4.09	0.11

4   Conclusion

In this work, we proposed an end-to-end network called Cycle-Desmoke that relies on a new generator architecture that consists of Atrous Convolution Feature Extraction Module (ACFEM) that helped in alleviating the smoke component at multiple scales and ensures the performance is analogous for both heterogeneous and homogeneous smoke. The use of Guide-Unsharp Upsample loss in addition to the cycle-consistency and adversarial loss helped to enhance the contrast of the desmoked image and also recover fine details. The quantitative and qualitative analysis of proposed method with other state-of-the-art methods depicts considerable improvement in terms of smoke removal and image quality as well. This work focuses primarily on single-image desmoking, it would be advantageous to utilize the spatial-temporal relationship between each frame in the video sequence to supervise the network to perform smoke removal. Hence, having a digital solution to remove surgical smoke in laparoscopic surgery would not only prove beneficial for practitioners, surgeons but also help improve the efficiency of computer-assistive algorithms.

References

1. Takahashi, H., et al.: Automatic smoke evacuation in laparoscopic surgery: a simplified method for objective evaluation. Surg. Endosc. **27**(8), 2980–2987 (2013)
2. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Trans. Pattern Anal. Mach. Intell. **33**(12), 2341–2353 (2010)
3. Wang, C., Cheikh, F.A., Kaaniche, M., Beghdadi, A., Elle, O.J.: Variational based smoke removal in laparoscopic images. Biomed. Eng. Online **17**(1), 139 (2018)

4. Kotwal, A., Bhalodia, R., Awate, S.P.: Joint desmoking and denoising of laparoscopy images. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1050–1054. IEEE (2016)

5. Bolkar, S., Wang, C., Cheikh, F.A., Yildirim, S.: Deep smoke removal from minimally invasive surgery videos. In: 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 3403–3407. IEEE (2018)

6. Sidorov, O., Wang, C., Cheikh, F.A.: Generative Smoke Removal. arXiv preprint arXiv:1902.00311 (2019)

7. Wang, C., Mohammed, A.K., Cheikh, F.A., Beghdadi, A., Elle, O.J.: Multiscale deep desmoking for laparoscopic surgery. In: Medical Imaging 2019: Image Processing, vol. 10949, p. 109491Y. International Society for Optics and Photonics (2019)

8. Chen, L., Tang, W., John, W.N.: Unsupervised learning of surgical smoke removal from simulation (2018)

9. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)

10. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 833–851. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_49

11. Guan, T., Zhu, H.: Atrous faster R-CNN for small scale object detection. In: 2017 2nd International Conference on Multimedia and Image Processing (ICMIP), pp. 16–21. IEEE (2017)

12. Shi, W., et al.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1874–1883 (2016)

13. Leibetseder, A., Primus, M.J., Petscharnig, S., Schoeffmann, K.: Real-time image-based smoke detection in endoscopic videos. In: Proceedings of the on Thematic Workshops of ACM Multimedia 2017, pp. 296–304. ACM (2017)

14. Twinanda, A.P., Shehata, S., Mutter, D., Marescaux, J., De Mathelin, M., Padoy, N.: EndoNet: a deep architecture for recognition tasks on laparoscopic videos. IEEE Trans. Med. Imaging **36**(1), 86–97 (2016)

15. Berman, D., Avidan, S.: Non-local image dehazing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1674–1682 (2016)

16. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: DehazeNet: an end-to-end system for single image haze removal. IEEE Trans. Image Process. **25**(11), 5187–5198 (2016)

17. Mittal, A., Moorthy, A.K., Bovik, C.: No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. **21**(12), 4695–4708 (2012)

18. Venkatanath, N., Praneeth, D., Bh, M.C., Channappayya, S.S., Medasani, S.S.: Blind image quality evaluation using perception based features. In: 2015 Twenty First National Conference on Communications (NCC), pp. 1–6. IEEE (2015)

19. Yan, J., Li, J., Fu, X.: No-reference quality assessment of contrast-distorted images using contrast enhancement. arXiv preprint arXiv:1904.08879 (2019)