2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A Smoke Removal Method Based on Combined Data and Modified U-Net for Endoscopic Images

Longfei Ma, Han Song, Xinran Zhang and Hongen Liao*, *Senior Member, IEEE*

***Abstract*—In minimally invasive surgery, the ablation of human tissue will produce a lot of smoke, which will interfere with the surgeon's operation. We propose a smoke removal method based on combined data and modified U-net for endoscopic images. The real dataset and the synthetic dataset are built using a small amount of images with smoke. The real dataset is combined with the synthetic dataset successively. Qualitative evaluation shows that the quality of the output smoke-free image is the best when training using the combined data, compared to using only either the real dataset or the synthetic dataset above. Quantitative evaluation shows that the effect of smoke removal is still the best when training using the combined data in our method.**

***Clinical Relevance*— A real-time smoke removal method suitable for endoscopic surgery is proposed to help surgeons get clear images in real time and make the operation go smoothly.**

I. Introduction

Image smoke removal is a classic problem in computer vision. Its main purpose is to get the corresponding smoke-free image through a given image with smoke, so that people can see more clearly.

For the mathematical model of smoke in the image, Nayer and Narasimhan et al. [1,2] deduced the atmospheric scattering model in detail, and divided the influence of smoke on the image into two parts: direct attenuation and occlusion. The classic smoke removal method is based on prior knowledge of dark channel proposed by He et al. [3]. The prior knowledge comes from the regular characteristics obtained from the statistics of a large number of natural scene images. Using the commonly used smoke model in computer graphics and the prior knowledge of dark channel, the relationship of each pixel under two different conditions of without smoke and with smoke is found. Then the image with smoke is transformed into the smoke-free image.

With the development of artificial intelligence, many smoke removal methods based on deep learning are proposed. Cai et al. [4] used neural network to fit the conversion parameters in the traditional smoke mathematical model and got a better result than the traditional method. Dong et al. [5] combined the traditional prior knowledge of dark channel with the deep learning method, and proposed a method of learning dark channel and transmitting prior knowledge using neural network for the first time. Li et al. [6] directly adopted the end-to-end process, which directly generated smoke-free images from the images with smoke. The complex process of removing smoke was simulated by the network of deep learning, which simplified the algorithm process. Shao et al. [7] applied the smoke removal algorithm of natural image to the field of remote sensing images based on the framework method of deep learning. The mathematical model of smoke distribution was used to build a synthetic dataset to train the network.

In the medical field, although there are several researches related to endoscopic smoke removal, most of them are based on traditional image processing methods and their effects are not very well. Kotwal et al. [8] treated smoke removal and noise reduction as a Bayesian problem. A probability graph model is established for the removal of endoscope smoke by combining the prior model of undamaged color image and the transmission image of color attenuation caused by smoke. Baid et al. [9] also established a new Bayesian model for endoscopic smoke removal, special reflection removal and noise reduction. Tchaka et al. [10] explained the decline of visibility by introducing haze medium to simulate the image formation process from the perspective of image chromaticity. To improve the dark channel method, an adaptive dark channel prior method was used, which combines with histogram equalization to remove the smoke artifacts, restore the brightness image, and enhance the contrast and brightness of the final result.

Although there are several researches on smoke removal, there are still many problems in the field of endoscopic surgery. For example, the smoke removal method applied in the natural scene is directly applied in the surgical scene, which will cause obvious maladjustment. Some smoke removal algorithms fail when the smoke concentration is high and the scene is seriously occluded. Moreover, it is difficult to meet the requirements of real-time processing in the medical field.

II. Methods

A. Distribution and Mathematical Model of Smoke in Image

Nayer and Narasimhan gave the analysis of smoke scattering model in [1,2]. This model divides the influence of smoke into two parts: direct attenuation and occlusion, which are described by Koschmider formula, as formula (1).

$$\mathrm{I(x) = J(x) * t(x) + (1 - A) * t(x)} \tag{1}$$

Research supported by National Natural Science Foundation of China (82027807, 81771940, 81901844, 81930119) and Beijing Municipal Natural Science Foundation (7212202).

Longfei Ma, Han Song and Xinran Zhang are with the Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China (e-mail: malongfei@tsinghua.edu.cn; songhan920421@163.com; zhangxinran@tsinghua.edu.cn)

Hongen Liao is with the Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China (corresponding author, e-mail: liao@tsinghua.edu.cn)

Where I(x) is the image with smoke and J(x) is the image without smoke. A is the global atmospheric light, which is usually understood as a constant. t(x) is a physical quantity related to depth, which is specifically expressed as formula (2).

$$t(x) = e^{-\beta d(x)} \tag{2}$$

Where *d(x)* is the depth and *e* is the natural base. β is an adjustable parameter with a range of 0 to 1, which generally affects the color of the smoke observed in the image.

Formula (1) essentially describes the corresponding relationship between each pixel of the image with smoke and the image without smoke. Therefore, the smoke removal process of the image is equivalent to the solution process of the parameters such as t (x) and A in formula (1). Hence, a clean image without smoke can be recovered from the image with smoke.

A. Modified U-net Structure for Smoke Removal

We use an end-to-end process algorithm based on deep learning. For the Da Vinci robot surgery, we use the U-net [11] network structure commonly used in the medical image field and make appropriate improvements to U-net. U-net is a classic network, which is an encoder-decoder structure. The first half of the network is the feature extraction part, which is the encoder, reducing the resolution of the feature map layer by layer and increasing the number of channels. The second part is the decoder, which recovers the target image step by step from the feature image with low resolution and high channel number through up sampling layer by layer. The biggest advantage of U-net is that it puts forward a skip direct connection structure, which splices each feature map of the encoder and each layer of the decoder separately. This advantage is that it can better preserve the detail features of the image.

Our demand is to achieve smoke removal in the surgical scene, so the effectiveness and speed of the network are required. On the basis of ensuring the effectiveness to meet the requirements, we modify the U-net network structure, reducing the number of down sampling and improving the operation efficiency. Fig.1 is a schematic diagram of our network model. We use the end-to-end training method to generate smoke-free images directly from the images with smoke. Mean square error (MSE) is the loss constraint, as shown in formula (3).

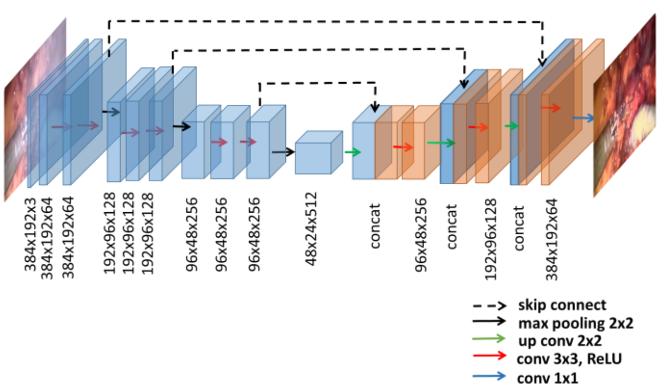

Fig.1. Schematic diagram of smoke removal network.

$$MSE = \frac{\sum \left| I_{(x,y)} - \widehat{I_{(x,y)}} \right|}{W * H} \tag{3}$$

B. Building Dataset Using Images with Smoke

In this study, we use the real images of kidney surgery in Da Vinci robotic surgery from Hamlyn Centre Laparoscopic dataset [12]. The dataset contains a total of 11 different scene sequences. We split the video into single frames, and select the clean image without smoke and the image with smoke respectively.

How to construct the image and its corresponding label (with or without smoke) from these images is the key to solve the problem of dataset establishment. Here, we use two different ways to build the dataset, and compare the effect of different methods in the later experiments.

The first method uses the adjacent frames of the video images and whether there is a smoke, as the data and the corresponding labels. Capture a video of ablation operation and find the close frame image before the start of ablation as the truth label without smoke. Then, after the start of ablation, we find about 5 frames of images with different smoke concentrations as the input data. These frames of images with smoke also correspond to a clean image as a label. However, due to the breathing of human beings, the shaking of endoscope and the movement of surgical instruments, the observed scene is a dynamic scene. As shown in Fig.2, even if an image with a relatively close frame number is selected, there is a phenomenon that the scene is dynamic. This will make the final image blurred when the pixel-to-pixel loss function is constrained. We will analyze this phenomenon in the evaluation.

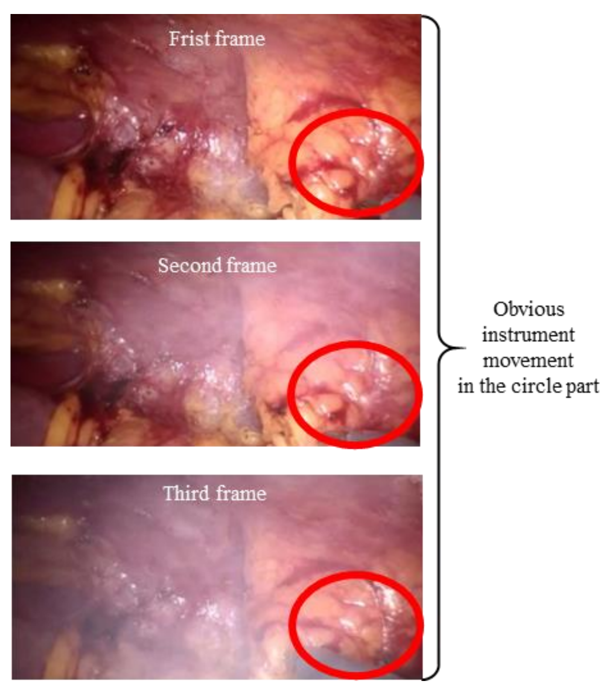

Fig.2. Construction of real dataset.

The second method is to use the smoke distribution formula (1) to generate smoke on clean images. The synthesized smoke image is used as the data, and the original clean image is used as the label. There is no pixel offset in the dynamic scene in the same image. Therefore, it can avoid the problem of blurring the output image after pixel-to-pixel supervision training. In the endoscopic surgery, smoke is generated due to ablation, and the process of generation is gradually increasing. It does not fill the whole picture instantaneously, but generates smoke locally and randomly. We use randomly selected regions to simulate the locality of smoke generation. At the same time, the smoke concentration is also different. Multiple images with different smoke concentrations on the same image are generated. As shown in Fig. 3, an example of generating multiple images with smoke for a clean image is given.

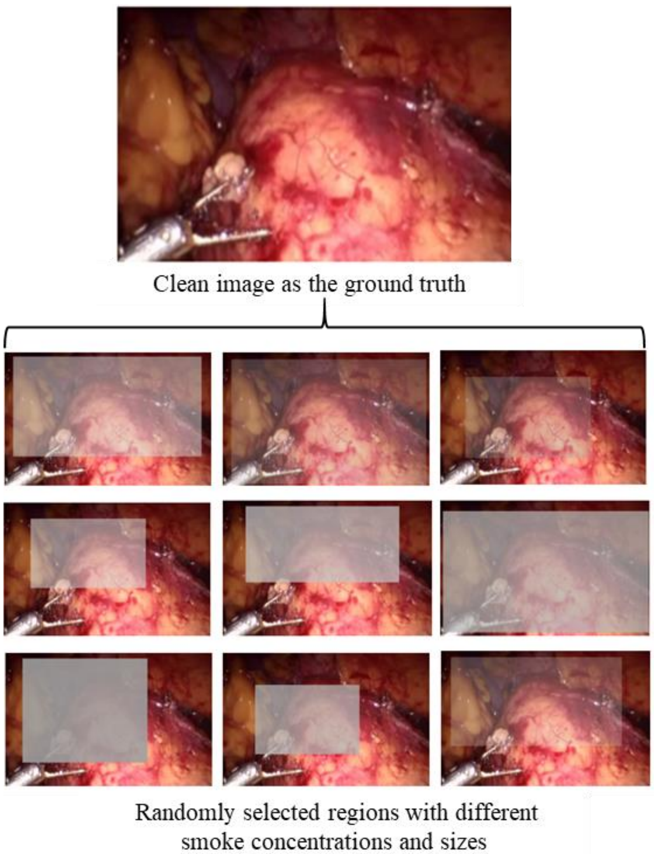

Fig.3. Generate multiple images with smoke from one clean image without smoke.

During the training, we take the images with smoke as the input of the network, and the output end uses the images without smoke as the supervision, which realizes the end-to-end smoke removal process. Real dataset (a total of 567 pairs of images) and synthetic dataset (a total of 24000 pairs of images) are built. The real dataset and the synthetic dataset are applied for training the network, successively. In the training process of the network, the same clean image without smoke corresponds to multiple images with smoke in both the real dataset and the synthetic dataset.

III. Experiments and Results

A. Qualitative Evaluation

Three training methods were compared: (I) only train the network using the real dataset; (II) only train the network using the synthetic dataset; (III) train using the real dataset and synthetic dataset, successively. The results of smoke removal are shown in Fig. 4. Qualitative evaluation shows that the quality of the output smoke-free image is the best when training using the real dataset and synthetic dataset successively, compared to using only one dataset above.

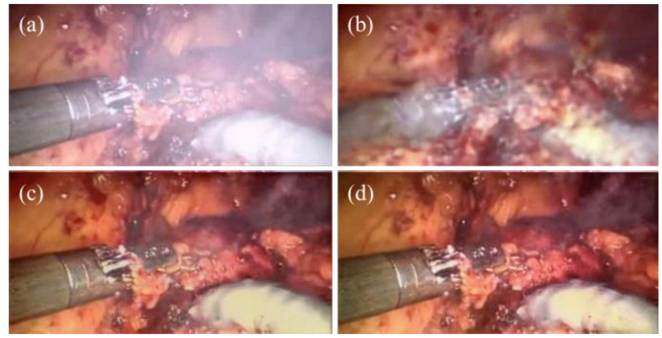

Fig.4. Qualitative evaluation using three training method: (a) the original image with smoke; (b) the output smoke-free image base on the real dataset; (c) the output smoke-free image base on the synthetic dataset; (d) the output smoke-free image base on the combination of the real dataset and synthetic dataset.

Because the data and the labels are dynamic scenes in real dataset, the output images cannot achieve pixel-to-pixel correspondence, but have a certain offset. Hence, the results are fuzzy and do not meet the requirements. Compared with the former method, the training results using the synthetic dataset are better: the effect of smoke removal is obvious, there is no serious color deviation, and the picture is clean and not blurred.

There are some differences between the synthetic smoke data and the real smoke data, so a group of experiment is added, that is, training on the real dataset first, and then training on the synthetic dataset. As shown in Fig.4, the effect of smoke removal is better, especially in some places with thick smoke.

B. Quantitative Evaluation

Because the result of the smoke removal is an image. The common parameters were used to evaluate the image consistency and the quality of smoke removal. If the real dataset is used for evaluation, there will be a certain pixel offset between the data and its corresponding real label. Hence, it is difficult to do quantitative evaluation based on pixels. Therefore, we selected the synthetic dataset for evaluation. As shown in Table 1, we evaluate the MSE, Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) of different training methods.

TABLE I. The Results of Quantitative Evaluation

	Synthetic Dataset	*R+S Dataset*
MSE	84.41	86.44
PSNR	28.90	28.79
SSIM	0.93	0.93

The smaller of MSE is the better; the higher of PSNR is the better; the value of SSIM closer to 1 is the better. We find that the performance based on the training of the synthetic dataset is almost the same with that on the real dataset plus synthetic dataset. However, combined with qualitative evaluation, the training effect is still the best on the real dataset plus synthetic dataset.

In addition to the quantitative evaluation of smoke removal, the operation efficiency of the network was also evaluated to ensure the real-time image processing. Our test environment includes Ubuntu 16.04 operating system, GPU 1080 graphics card. The resolution of endoscopic images are 384×192. In the test, the average processing time of an image was about 0.022 s. Hence, the theoretical video can reach about 44 frames, which fully meets the real-time requirements.

Smoke removal can be regarded as the noise removal of the image. The distribution histograms of the gray values of the original image with smoke and the clean image after removing smoke were calculated and shown in Fig. 5 (c), respectively. It can be seen from the histograms that the grey values of the smoke are roughly distributed between the gray values of 120 and 180, which is indicated by a red ellipse.

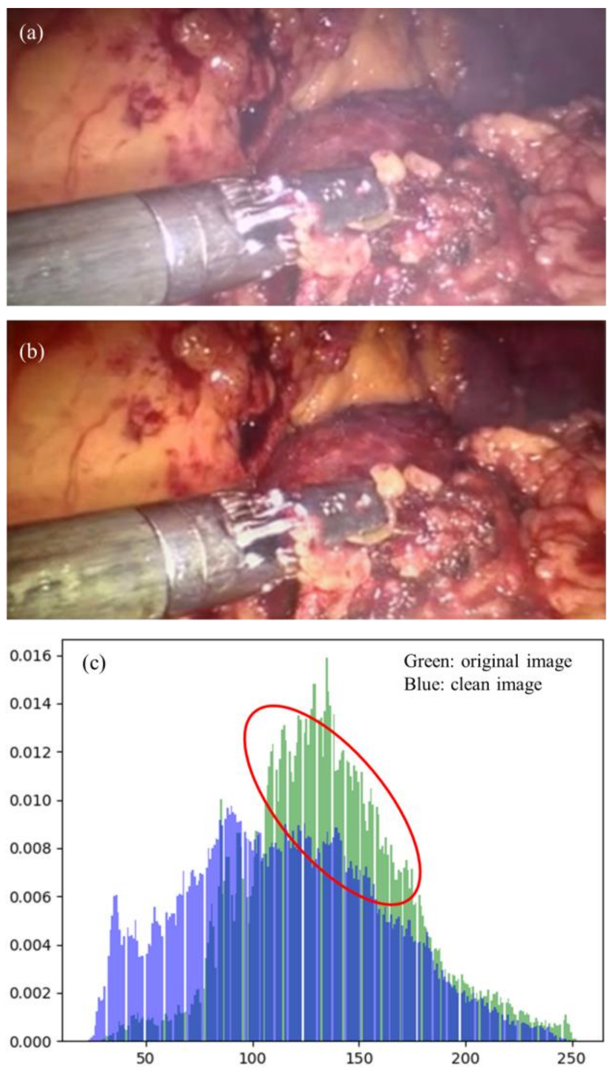

Fig.5. Analysis of the distribution histograms of the gray values: (a) the original image with smoke; (b) the clean image after removing smoke; (c) Green histogram is for the original image and blue histogram is for the clean image.

IV. Discussion and Conclusion

Our method is an end-to-end algorithm based on deep learning, which ensures the simplicity of the smoke removal process. The modified U-net structure can accurately fit the smoke removal process, which overcomes the problems of large smoke concentration and heavy color deviation.

Moreover, in order to improve the operation efficiency of the algorithm, we modify and simplify the traditional U-net structure to meet the speed requirements of real-time processing.

A small amount of endoscopic images are used to construct the real dataset and synthetic dataset for training the network. The real dataset is combined with the synthetic dataset successively. Qualitative and quantitative evaluations shows that the best effect of smoke removal can be realized when training using the combined data, compared to using only either the real dataset or the synthetic dataset above.

In conclusion, we propose a smoke removal method based on combined data and modified U-net for endoscopic images, which may be suitable for endoscopic surgery to help surgeons get clear images in real time and make the operation go smoothly. Moreover, the output smoke-free images can be further applied to the depth estimation of endoscopic images.

References

[1] Nayar S K, and Narasimhan S G, "Vision in bad weather." Computer Vision, 1999. The Proceedings of the Seventh IEEE, International Conference on IEEE, (1999).

[2] Narasimhan S G, and Nayar S K, "Vision and the Atmosphere." International Journal of Computer Vision, 48.3(2002):233-254.

[3] He K, Sun J and Tang X, "Single Image Haze Removal Using Dark Channel Prior." IEEE Transactions on Pattern Analysis & Machine Intelligence, 33.12(2011):2341-2353.

[4] Cai B, Xu X, Jia K, et al., "DehazeNet: An End-to-End System for Single Image Haze Removal." IEEE Transactions on Image Processing, 25.11(2016):5187-5198.

[5] Dong Y, and Jian S, "Proximal Dehaze-Net: A Prior Learning-Based Deep Network for Single Image Dehazing." Springer, Cham (2018).

[6] Li B, Peng X, Wang Z, et al., "AOD-Net: All-in-One Dehazing Network." 2017 IEEE International Conference on Computer Vision, (ICCV) IEEE, (2017).

[7] Shao S, Guo Y, Zhang Z, et al., "Single Remote Sensing Multispectral Image Dehazing Based on a Learning Framework." Mathematical Problems in Engineering 2019, PT.1(2019):1-8.

[8] Kotwal A, Bhalodia R, and Awate S P, "Joint desmoking and denoising of laparoscopy images." 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI 2016) IEEE, (2016).

[9] Baid A, Kotwal A, Bhalodia R, et al., "Joint desmoking, specularity removal, and denoising of laparoscopy images via graphical models and Bayesian inference." 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017) IEEE, (2017).

[10] Tchaka K, Pawar V M, and Stoyanov D, "Chromaticity based smoke removal in endoscopic images." Society of Photo-optical Instrumentation Engineers Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, (2017).

[11] Ronneberger O, Fischer P, and Brox T, "U-Net: Convolutional Networks for Biomedical Image Segmentation." Springer, Cham (2015), 234-241.

[12] Video P, and Description S, "Hamlyn Centre Laparoscopic / Endoscopic Video Datasets." (2014).