Multiscale deep desmoking for laparoscopic surgery

Congcong Wang, Ahmed Kedir Mohammed, Faouzi Alaya Cheikh, Azeddine Beghdadi, Ole Jacob Elle

Multiscale deep desmoking for laparoscopic surgery

Congcong Wang[*][a], Ahmed Kedir Mohammed[*][a], Faouzi Alaya Cheikh[a], Azeddine Beghdadi[b], and Ole Jacob Elle[c,d]

[a]Norwegian Colour and Visual Computing Lab, Norwegian University of Science and Technology, Gjøvik, Norway.
[b]L2TI-Institut Galilée, Université Paris 13, Sorbonne Paris Cité, Villetaneuse, France.
[c]The Intervention Centre, Oslo University Hospital, Oslo, Norway.
[d]The Department of Informatics, University of Oslo, Oslo, Norway.

ABSTRACT

In minimally invasive surgery, smoke generated by such as electrocautery and laser ablation deteriorates image quality severely. This creates discomfortable view for the surgeon which may increase surgical risk and degrade the performance of computer assisted surgery algorithms such as segmentation, reconstruction, tracking, etc. Therefore, real-time smoke removal is required to keep a clear field of view. In this paper, we propose a real-time smoke removal approach based on Convolutional Neural Network (CNN). An encoder-decoder architecture with Laplacian image pyramid decomposition input strategy is proposed. This is an end-to-end network which takes the smoke image and its Laplacian image pyramid decomposition as inputs, and outputs a smoke free image directly without relying on any physical models or estimation of intermediate parameters. This design can be further embedded to deep learning based follow-up image guided surgery processes such as segmentation and tracking tasks easily. A dataset with synthetic smoke images generated from Blender and Adobe Photoshop is employed for training the network. The result is evaluated quantitatively on synthetic images and qualitatively on a laparoscopic dataset degraded with real smoke. Our proposed method can eliminate smoke effectively while preserving the original colors and reaches 26 fps for a video of size 512×512 on our training machine. The obtained results not only demonstrate the efficiency and effectiveness of the proposed CNN structure, but also prove the potency of training the network on synthetic dataset.

Keywords: Image pyramid decomposition, smoke removal, convolutional neural networks, laparoscopic surgery, synthetic dataset

1. INTRODUCTION

For a laparoscopic surgery, the abdominal cavity is visualized by some specialized instruments such as ultrasonic probe and telescope. Those instruments are inserted to the abdomen through small incisions. Thus, videos/images acquired by the laparoscope is one of the main data modality during a surgery. It not only gives a intuitive visualization of the patients abdomen but also is an informative input for computer vision based navigation systems.[1–3]

However, the image quality can be degraded by artifacts during the surgery. Those artifacts include blood, illumination change, specular reflections, noise, smoke, etc. In particular, smoke, which is caused by such as laser tissue ablation or electrocautery, can significantly deteriorate the image quality for large areas of the scene. Therefore, it is important to remove smoke by physical smoke evacuation solutions [4] (e.g., the Laparoshield Laparoscopic Smoke Filtration System [*]) and by image processing algorithms.[5] This paper aims at a real-time effective smoke removal approach via deep learning in order to maintain a clear surgery view from the harmful surgical smoke.

---

[*] Contributed equally to this work
Further author information:
E-mail: congcong.wang@ntnu.no, Telephone: +47 47731808
[*]https://shop.pall.com/us/en/medical/surgical-gas-and-smoke-filtration/surgical-gas-and-smoke-filtration

The reminder of this paper is organized as follows. First, in Sec. 2 , we review the related work on laparoscopic image desmoking methods. Next, in Sec 3, the proposed method is presented. The experimental results and analysis are then discussed in Sec 4. Finally, conclusions are drawn in Sec 5.

2. RELATED WORK

Image processing based smoke removal is a recent topic and there are few works proposed.[5–11] In these papers, the approaches can be classified into traditional approaches[5–9] and deep learning approaches[10,11] which will be discussed in this section.

Traditional approaches: In these methods, the smoke removal problem is developed based on the atmospheric scattering model[12] as presented by Eq. (1).

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

where $\mathbf{I}$ is the observed haze image, $\mathbf{J}$ is the haze-free image, t is the medium transmission map and $\mathbf{A}$ is the global atmospheric light.

In,[6] denoising and desmoking problem are formulated to a Bayesian inference problem, where the uncorrupted image is modeled by a Markov Random Field (MRF), and maximum-a-posteriori (MAP) is applied to find the solution. This work is then extended for denoising, desmoking and specularity removal in[7] where texture preserving prior is introduced to preserve contrast and natural texture. In,[8] a dark channel prior dehazing method (DCP) proposed in [13] is adapted for desmoking purpose. When original DCP is directly applied, color distortion is introduced because of the underlying assumptions for dehazing are improper for desmoking. In order to overcome the color distortion, Tchaka *et al.*[8] propose to threshold the dark channel values by an empirically selected constant or to refine the dark channel values by putting higher emphasis on values close to 0.5 and decreasing the emphasis according to the value distance to 0.5. Then the desmoked image's contrast is enhanced by a histogram equalization process. In order to evaluate the performance quantitatively, the authors choose one smoke free frame at the beginning of the sequence as groundtruth, then compute the MSE (mean square error) between the chosen groundtruth image and the following frames. The proposed approach obtains better visual quality but reduces less smoke compared to the original DCP method. In,[9] in order to avoid estimating the transmission map t which is a challenging problem, Luo *et al.* propose a bilateral of bilateral grid (BBG) method to estimate the atmospheric veil which is defined as $\mathbf{A}(1 - t(x,y))$ from Eq.(1) in.[14] Then together with an estimated $\mathbf{A}$, the smoke free image $\mathbf{J}$ is recovered by inversing the physical model of Eq. (1). To further correct the color infidelity and dark pixels, contrast enhancement and luminance fusion scheme are introduced. No-reference image quality assessment metrics to account for naturalness[15] and sharpness[16] are proposed to evaluate the performance. The results show equivalent sharpness and better naturalness compare to other previous proposed dehazing methods. In,[5] following the same definition of atmospheric veil as in,[9,14] Wang *et al.* propose to estimate the smoke veil by a variational method based on the observation that smoke veil has low contrast and low inter-channel differences. The obtained smoke veil is then subtracted from the original degraded image, resulting in the direct attenuation part (defined as $\mathbf{J}(x,y)t(x,y)$ from Eq. (1)). Finally, the smoke free image is computed using a linear intensity transformation of the direct attenuation part. The results show that the smoke is removed effectively while keeping perceptual visual quality.

Deep learning approaches: The first deep learning desmoking approach is proposed in,[10] Bolkar *et al.* propose to generate a synthetic dataset by Perlin noise,[17] then the dataset is used to fine tuning a dehazing network AOD-Net.[18] The computational speed of this method reaches 20 fps for size of 512× 512 color videos. Later in,[11] a Blender generated synthetic dataset is proposed and used to train a cGAN (conditional Generative Adversarial Network) structure. Both of the methods achieve fast computational speed with fine results, however, further study is required to improve the smoke removal ability.

Although the atmospheric model shows its success for dehazing[13,19] and its promising performance for desmoking,[5,9] there are some assumptions underlying it. The methods will fail when the assumptions are violated. Such as, in the case of dense and heterogeneous smoke, the performance is degenerated as also pointed out in the literatures.[5,9] Therefore, instead of estimating transmission map t or other inter-mediate parameters of the atmospheric physical model, we directly estimate smoke free images from degraded smoke images by a CNN approach which will be introduced in the following section.
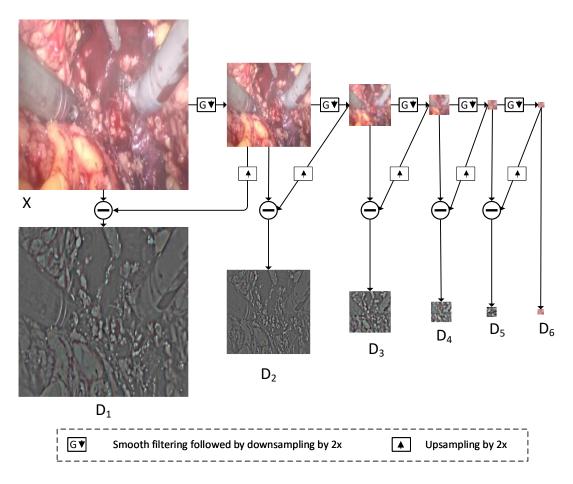
Figure 1. Image Pyramid Decomposition.

3. METHOD

In this section, our proposed method is presented: starting with an image pyramid decomposition strategy, then the CNN (Convolutional Neural Network) structure.

3.1 Image Pyramid Decomposition

Image pyramid decomposition is a multi-resolution representation of images and it can perfectly reconstruct original images.[20] It is widely used in computer vision community, such as dehazing,[14] image blending,[20] super resolution,[21] etc. Figure 1 illustrates the decomposition of an image into a N levels pyramid, with $N = 6$ in the figure.

For level n of the pyramid, the image is decomposed into a low frequency base layer and a high frequency detail layer by using guided filter for smoothing:[22, 23]

$$D_n(\mathbf{X}) = B_n(\mathbf{X}) - upsample(B_{n+1}(\mathbf{X})), \tag{2}$$

where $\mathbf{X}$ is the original smoke image, n indicates the level of the pyramid with $n = 1, ..., N$. For decomposition level n, $D_n(\mathbf{X})$ is the detail layer and the last layer is defined as $D_N(\mathbf{X}) = B_N(\mathbf{X})$, $B_n(\mathbf{X})$ is the base layer defined as $B_n(\mathbf{X}) = downsample(guided(B_{n-1}(\mathbf{X}))$ where $guided$ means low-pass filtering and $B_1(\mathbf{X}) = \mathbf{X}$. A nearest neighbor interpolation is applied for upsampling. The layers D_n is the Laplacian pyramid representation of the original image $\mathbf{X}$ which contains all the information needed to reconstruct $\mathbf{X}$. Note that the last layer $D_6(\mathbf{X}) = B_6(\mathbf{X})$ as illustrated in Figure 1.
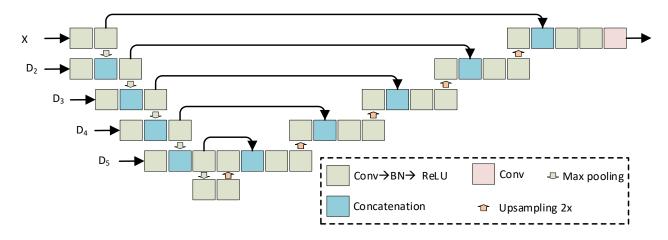
Figure 2. Overview of the Proposed CNN architecture.

3.2 Network Structure

An encoder-decoder network is adopted in this paper and the network structure is shown in Figure 2.

Encoder: The encoder of our proposed architecture includes five downsampling (max pooling) operations and blocks of "Convolution (Conv)→ BN (BatchNorm) → ReLU (Rectified Linear Unit)". The downsampling operations cause spatial information loss. Therefore, in addition to input the original degraded image, the Laplacian detail layers of the original images are inputted to the corresponding levels of the encoder as shown in Figure 2. The original smoke image X is inputted to level 1 of the network, detail layers D_2, D_3, D_4, D_5 are inputted to level $2, 3, 4, 5$ of the encoder separately.

Decoder: The decoder of the network consists of 5 upsampling operations and blocks of "Conv→ BN → BeLU". They are stacked as illustrated in Figure 2. Besides, skip-connections between the encoder and decoder are introduced for information sharing from encoder to decoder.

This structure is inspired by the Laplacian Pyramid[24] and its application in the multi-scale fusion dehazing approaches[14, 25] and Y-Net.[26] Our proposed architecture is similar to the encoder-decoder network U-Net[27] and differs mainly in the two following aspects: (1) the image pyramid representation is encoded into the network to compensate for the information loss caused by max pooling operations; (2) the decoder network is designed to be deeper at each level compared to the encoder one. The information of each layer of the decoder includes up-sampled information from neighbor layer and information from the skip connection, which not only contains information from the down-sampling operation but also includes the information from pyramid inputs. A deeper design is required to help the network learn better features from the corresponding layers' information.

4. EXPERIMENTS

In this section, we demonstrate the proposed method on a public dataset.[28] We first introduce the dataset and implementation details of our proposed method. Then, we evaluate and compare our method with other state-of-the-art approaches. The evaluation starts with an objective evaluation achieved through the use of full reference image quality metrics on synthetic dataset, followed by a qualitative visual inspection of the results on real smoke images and then a computational speed analysis, ended with a comparison to U-Net.

4.1 Training Dataset

In vivo porcine procedure dataset collected in partial nephrectomy in Da Vinci surgery from Hamlyn Centre Laparoscopic / Endoscopic Video Dataset Page[28, 29] is used for training. This dataset is originally used for training and testing deep learning based disparity estimation which contains 20,000 rectified stereo image pairs from 11 video sequences. In,[11] smoke free images are selected manually from the left view of the original training

dataset, then synthetic smoke images are generated with Blender[†],[11] three different smoke density levels are generated for the selected 7553 smoke free images. In addition to this Blender generated images, we synthesize the selected images using Adobe Photoshop[‡] by adding rendered clouds with three different opacity: 50%, 75% and 100%. Thus, in total 67,977 smoke images and the corresponding 7553 ground truth images are included. We randomly selected 80% of the smoke images to train the network and 20% for validation.

4.2 Implementation Details

Our model is implemented on Tensorflow [§] and Keras [¶] library with a single NVIDIA 12GB Titan X GPU. Mean Squared Error (MSE) loss function is employed. We first resize all images into fixed dimensions with spatial size of 512×512 before feeding them to the network and the inputs are normalized to [0, 1] in intensity. We use Adam as the optimizer with batch size 5 and learning rate is set to 0.001. We monitor PSNR loss and use early-stop criteria (with patience of 8 and min_delta of 0.0001) on the validation set error. The parameters for guided filter are: window radius $r = 3$ and regularization parameter $\epsilon = 1.0$

4.3 Experimental Results

We compare the proposed approach with four state-of-the-art methods. The first one is the popular dark channel prior dehazing method (DCP).[13] The second one is a laparoscopic smoke removal method based on the dark channel prior, we implement the first step of the original method which aims for removing smoke, this step is designated by R-DCP-1. Besides, the whole algorithm is implemented and denoted as R-DCP.[8] The main difference is the latter includes an image contrast enhancement step. The third one EVID is a variational framework based dehazing method.[30] Since the assumptions underlying the atmospheric physical model may be violated for laparoscopic desmoking purpose, compared to physical model based dehazing methods, EVID depends only mildly on physical considerations, so it is more suitable for desmoking purpose. The fourth one is a recent proposed laparoscopic desmoking method (VAR).[5]

4.3.1 Quantitative evaluation on synthetic dataset

In this evaluation, 100 smoke free images are selected manually from the left view of the original test dataset from Hamlyn Centre Laparoscopic / Endoscopic Video Dataset Page[29] used in.[28] Following the same procedure of generating training dataset by Adobe Photoshop, we render the three different opacity level (50%, 75% and 100%) and generate 300 images for test. They are designated as *low*, *medium* and *high* separately, which indicates the rendered smoke density (opacity levels). Full-reference image quality metrics SSIM[31] and PSNR are employed for the validation.

Table 1 illustrates the mean and standard deviation of the metrics scores for the different methods and the smoke images. Our proposed method apparently achieves the best performance in terms of SSIM and PSNR.

Table 1. Average and standard deviation results for each evaluation metric.

		Smoke images	DCP[13]	R-DCP[8]	EVID[30]	VAR[5]	Proposed
	low	15.87 ± 1.17	15.25 ± 1.52	18.59 ± 1.60	20.90 ± 1.50	16.90 ± 2.08	**28.58 ± 1.84**
PSNR	*medium*	12.14 ± 1.03	16.08 ± 1.55	17.32 ± 1.14	20.63 ± 1.65	16.71 ± 1.83	**27.91 ± 1.69**
	high	9.81 ± 1.21	17.00 ± 1.53	15.53 ± 1.33	18.26 ± 2.07	15.72 ± 1.70	**26.92 ± 1.65**
	low	0.88 ± 0.03	0.85 ± 0.04	0.76 ± 0.03	0.92 ± 0.03	0.90 ± 0.03	**0.99 ± 0.01**
SSIM[31]	*medium*	0.77 ± 0.04	0.86 ± 0.05	0.68 ± 0.03	0.93 ± 0.03	0.88 ± 0.04	**0.98 ± 0.01**
	high	0.65 ± 0.07	0.87 ± 0.05	0.58 ± 0.05	0.91 ± 0.04	0.83 ± 0.06	**0.98 ± 0.01**

[†]https://www.blender.org/

[‡] https://www.adobe.com/no/products/photoshop.html

[§]https://www.tensorflow.org/

[¶]https://keras.io/

4.3.2 Qualitative evaluation on real images

Although the proposed method achieves good performance on synthetic dataset. It is necessary to evaluate the generalization ability of the network on real smoke images. 300 smoke images are selected manually from the left view of the original test dataset[28] for validation.

As the ground-truth information for a smoked laparoscopic image is not available, and there are no gold standards for evaluating desmoked images' quality by no-reference image quality metrics, in this part, the result is evaluated subjectively.

| (a) | (b) | (c) | (d) | (e) | (f) | (g) |

Figure 3. Subjective results. (a) Input smoke laparoscopic images and the obtained desmoked ones using: (b) DCP;[13] (c) R-DCP-1;[8] (d) R-DCP;[8] (e) EVID;[30] (f) VAR[5] and (g) proposed method.

Figure 3 illustrates the results of the different approaches with different smoke distributions: low smoke density (first row) which happens when smoke fade in/out, moderate and smooth (second row) smoke density, moderate and not smooth (third row) smoke density, high and not smooth (fourth row) smoke density, high and smooth (fifth row) smoke density. For first row, all the methods can remove the smoke effectively, however, EVID, DCP and VAR cause color distortion. R-DCP-1 overcomes the color distortion problem compared to the original DCP method, and R-DCP enhances the contrast compared to R-DCP-1. Our proposed method shows an equal ability to eliminate smoke and a better performance for preserving color. For moderate density smoke, EVID, VAR and the proposed method show better smoke removal results while our method has the best color restoration performance. All the methods obtain worse results when the smoke becomes not smooth. For dense smoke, all the approaches fail to eliminate the smoke especially with dense and non smooth smoke, however, the proposed method reaches the best perceptual image quality by introducing minimum color distortion. In conclusion, our proposed method can remove the smoke effectively without causing color distortion compared to the other approaches. However, the difference between real smoke images and synthetic images affects the performance of the proposed approach, this can be improved by improving the synthetic smoke dataset.

4.3.3 Run time comparison

Table 2 shows the average run time per image obtained on the 300 real smoke images dataset. All the methods are run on the same machine of Windows operation system (with a single NVIDIA GeForce GTX 980 GPU). Our method illustrates its efficiency thanks to the deep learning framework. Moreover, on our training machine, the speed is 0.038 second per frame (26 fps). Note that all the implementations are run as it is, no extra optimization is applied for faster computation, which may introduce some unfair comparison. For example,

Matlab implementations may be slower than a C++ version, parallel computing techniques can improve some methods' speed. All the comparison methods are applied on the original video size (384 × 192), except our proposed method is tested on resized video (512 × 512). Despite of the comparison difficulties because of different implementation platform, the results demonstrate our proposed method's promising efficiency.

Table 2. Computation time comparison per image.

Methods	Computation time (seconds)	Platform
DCP[13]	0.2167	Matlab
R-DCP[8]	0.2407	Matlab
EVID[30]	3.7433	C++ based executable, with a Matlab wrapper
VAR[5]	0.7417	Matlab
Proposed	0.0689	Python

4.3.4 Comparison to U-Net

We first quantitatively compare U-Net[27] and our proposed method on the synthetic test dataset. U-Net's PSNR and SSIM scores are: 28.29±1.92, 27.56±1.80 , 26.54±1.72 and 0.98±0.01, 0.98±0.01, 0.97±0.01 respectively. The scores of our approach (shown in Table 1) is slightly better. Figure 4 shows result obtain form the synthesized images. For dense smoke region (highlighted by blue rectangles), the proposed method achieves better results. Moreover, Figure 5 illustrates the result obtained on some real smoke images, the proposed approach has a better smoke removal ability.

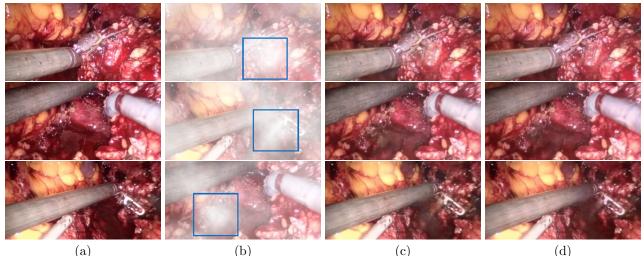| (a) | (b) | (c) | (d) |

Figure 4. Subjective results of synthetic dataset. (a) Smoke free images; (b) Synthesized smoke images by Adobe Photoshop; (c) Desmoked images by U-Net;[27] (d) Desmoked images by proposed method.

5. CONCLUSION

A fast and effective smoke removal method is demanded to assist surgeons and image guided surgery systems for laparoscopic surgery. In this paper, a CNN based smoke removal approach is proposed. Although the network is trained only on synthetic data, the experimental results show that our method outperforms the state-of-the-art ones in terms of perceptual image quality and computational speed. In the future, we plan to further simulate more realistic training dataset to improve the results especially for dense and heterogeneous smoke.
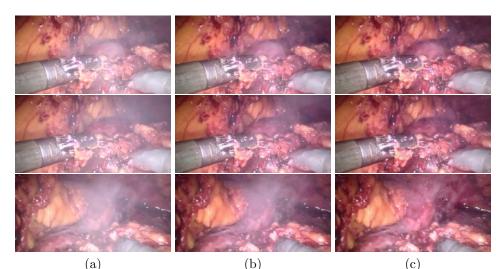
Figure 5. Subjective results of real smoke images. (a) Smoke images; (b) Desmoked images by U-Net;[27] (c) Desmoked images by proposed method.

ACKNOWLEDGMENTS

REFERENCES

[1] Stoyanov, D., "Surgical vision," *Annals of biomedical engineering* **40**(2), 332–345 (2012).

[2] Wang, C., Alaya Cheikh, F., Kaaniche, M., and Elle, O. J., "Liver surface reconstruction for image guided surgery," in [*Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*], **10576**, 105762H, International Society for Optics and Photonics (2018).

[3] Wang, C., Palomar, R., and Alaya Cheikh, F., "Stereo video analysis for instrument tracking in image-guided surgery," in [*European Workshop on Visual Information Processing (EUVIP)*], 1–6, IEEE (2014).

[4] Lawrentschuk, N., Fleshner, N. E., and Bolton, D. M., "Laparoscopic lens fogging: a review of etiology and methods to maintain a clear visual field," *Journal of endourology* **24**(6), 905–913 (2010).

[5] Wang, C., Alaya Cheikh, F., Kaaniche, M., Beghdadi, A., and Elle, O. J., "Variational based smoke removal in laparoscopic images," *Biomedical engineering online* **17**(1), 139 (2018).

[6] Kotwal, A., Bhalodia, R., and Awate, S. P., "Joint desmoking and denoising of laparoscopy images," in [*international Symposium on Biomedical Imaging (ISBI)*], 1050–1054, IEEE (2016).

[7] Baid, A., Kotwal, A., Bhalodia, R., Merchant, S. N., and Awate, S. P., "Joint desmoking, specularity removal, and denoising of laparoscopy images via graphical models and bayesian inference," in [*international Symposium on Biomedical Imaging (ISBI)*], 732–736, IEEE (2017).

[8] Tchakaa, K., Pawara, V. M., and Stoyanova, D., "Chromaticity based smoke removal in endoscopic images," in [*Proc. of SPIE Vol*], **10133**, 101331M–1 (2017).

[9] Luo, X., McLeod, A. J., Pautler, S. E., Schlachta, C. M., and Peters, T. M., "Vision-based surgical field defogging," *IEEE transactions on medical imaging* **36**(10), 2021–2030 (2017).

[10] Bolkar, S., Wang, C., Cheikh, F. A., and Yildirim, S., "Deep smoke removal from minimally invasive surgery videos," in [*International Conference on Image Processing (ICIP)*], 3403–3407, IEEE (2018).

[11] Chen, L., Wen, T., and John, N., "Unsupervised learning of surgical smoke removal from simulation," in [*Hamlyn Symposium on Medical Robotics*], Imperial College London (2018).

[12] Narasimhan, S. G. and Nayar, S. K., "Vision and the atmosphere," *International Journal of Computer Vision* **48**(3), 233–254 (2002).

[13] He, K., Sun, J., and Tang, X., "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence* **33**(12), 2341–2353 (2011).

[14] Ancuti, C. O. and Ancuti, C., "Single image dehazing by multi-scale fusion," *IEEE Transactions on Image Processing* **22**(8), 3271–3282 (2013).

[15] Yeganeh, H. and Wang, Z., "Objective quality assessment of tone-mapped images," *IEEE Transactions on Image Processing* **22**(2), 657–667 (2013).

[16] Bahrami, K. and Kot, A. C., "A fast approach for no-reference image sharpness assessment based on maximum local variation," *IEEE Signal Processing Letters* **21**(6), 751–755 (2014).

[17] Perlin, K., "Improving noise," in [*ACM Transactions on Graphics (TOG)*], **21**(3), 681–682, ACM (2002).

[18] Li, B., Peng, X., Wang, Z., Xu, J., and Feng, D., "Aod-net: All-in-one dehazing network," in [*International Conference on Computer Vision (ICCV)*], 4770–4778, IEEE (2017).

[19] Xu, Y., Wen, J., Fei, L., and Zhang, Z., "Review of video and image defogging algorithms and related studies on image restoration and enhancement," *IEEE Access* **4**, 165–188 (2016).

[20] Burt, P. and Adelson, E., "The laplacian pyramid as a compact image code," *IEEE Transactions on communications* **31**(4), 532–540 (1983).

[21] Lai, W.-S., Huang, J.-B., Ahuja, N., and Yang, M.-H., "Fast and accurate image super-resolution with deep laplacian pyramid networks," *IEEE transactions on pattern analysis and machine intelligence* (2018).

[22] He, K., Sun, J., and Tang, X., "Guided image filtering," *IEEE transactions on pattern analysis and machine intelligence* (6), 1397–1409 (2013).

[23] Fu, X., Huang, J., Ding, X., Liao, Y., and Paisley, J., "Clearing the skies: A deep network architecture for single-image rain removal," *IEEE Transactions on Image Processing* **26**(6), 2944–2956 (2017).

[24] Burt, P. J. and Adelson, E. H., "The laplacian pyramid as a compact image code," in [*Readings in Computer Vision*], 671–679, Elsevier (1987).

[25] Galdran, A., "Image dehazing by artificial multiple-exposure image fusion," *Signal Processing* **149**, 135–147 (2018).

[26] Mohammed, A., Sule, Y., Ivar, F., Marius, P., and Hovde, Ø., "Y-net: A deep convolutional neural network for polyp detection," in [*British Machine Vision Conference(BMVC)*], BMVA (2018).

[27] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*International Conference on Medical image computing and computer-assisted intervention (MICCAI)*], 234–241, Springer (2015).

[28] Ye, M., Johns, E., Handa, A., Zhang, L., Pratt, P., and Yang, G. Z., "Self-supervised siamese learning on stereo image pairs for depth estimation in robotic surgery," in [*Hamlyn Symposium on Medical Robotics*], Imperial College London (2017).

[29] London, I. C., "Hamlyn centre laparoscopic / endoscopic video datasets." http://hamlyn.doc.ic.ac.uk/vision/. Accessed: 2017-10.

[30] Galdran, A., Vazquez-Corral, J., Pardo, D., and Bertalmío, M., "Enhanced variational image dehazing," *SIAM Journal on Imaging Sciences* **8**(3), 1519–1546 (2015).

[31] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing* **13**(4), 600–612 (2004).