https://doi.org/10.2352/ISSN.2169-2629.2018.26.247

Deep-STRESS Capsule Video Endoscopy Image Enhancement

Ahmed Mohammed[1], Marius Pedersen[1], Øistein Hovde[2] and Sule Yildirim[1]
[1]Norwegian University of Science and Technology, Gjøvik, Norway
[2]University of Oslo, Oslo, Norway

Abstract

This paper proposes a unified framework for capsule video endoscopy image enhancement with an objective to enhance the diagnostic values of these images. The proposed method is based on a hybrid approach of deep learning and classical image processing techniques. Given an input image, it is decomposed spatially into multi-layer features. We estimate the base layer with pre-trained deep edge aware filters that are learned on the flicker dataset. The detail layers are estimated by the spatio-temporal retinex-inspired envelope with a stochastic sampling technique. The enhanced image is computed by a convex linear combination of the base and the detail layers giving detailed and shadow surface enhanced image. To show its potential, performance comparison between with and without the proposed image enhancement technique is shown using several video images obtained from capsule endoscopy for different parts of the digestive organ. Moreover, different learned filters such as Bilateral and L_0 norm are compared for enhancement using an objective image quality metric, BRISQUE, to show the generality of the proposed method.

Introduction

Colonoscopy is the ”gold standard” for colorectal cancer screening. During this procedure, a thin flexible tube called a colonoscope is passed into the rectum. A gastroenterologist guides the colonoscope to detect changes or abnormalities in the distal part of the small bowel (ileum), large intestine (colon) and rectum. Capsule video endoscopy is a distributive technology that is used to visualize the gastrointestinal (GI) tract. While it does not yet replace colonoscopy, it is gaining popularity as it is less invasive and more patient-friendly than colonoscopy. It is a swallow-able size electronic pill that takes pictures during its way through the GI tract. The capsule itself contains a CMOS image sensor with a light-emitting diode (LED) and lens for capturing images while passing through the GI tract as well as an ASIC transceiver with an antenna for transmitting captured image data to the receiving unit along with two batteries that provide power. Compared to the traditional colonoscopy, capsule endoscopy does not provide high image resolution and frame rates. The images lack apparent surface details and the directed LED light source on the capsule usually casts shadow on surface of tissues Fig. (1) . Moreover, the images are taken under low illumination, high compression ratio, and noisy medium. Image quality and enhancement of capsule images has attracted a lot of interest from numerous researchers, since it became commercially available in 2006. Research works on enhancement of capsule endoscopy image include vessel enhancement [1–4], removing or de-emphasizing specular reflections and illumination variation [5, 6], and color enhancement for a better visualization [7, 8]. However, despite these various methods that have been proposed recently to help gastroenterologists enhance the images, it is often unclear how these methods are related and when one method is preferable over another. Moreover, these approaches rely on different techniques without forming a unified framework for capsule endoscopy image enhancement.

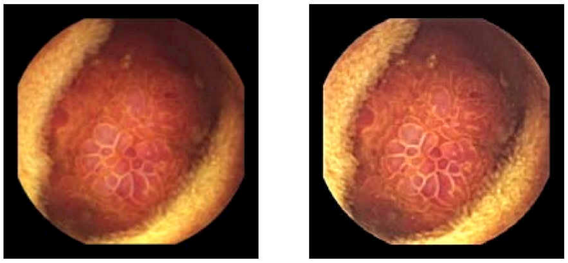

(a) Input Image (b) Enhanced image

Figure 1: A polyp with surface structure: among 7 learned filters, (b) shows the result of our proposed approach for with maximum BRISQUE score

In this paper, following the success of deep learning on various computer vision tasks such as classification, segmentation and image enhancement, we propose a unified approach to capsule image enhancement. Inspired by the recent progress in deep edge aware filters [9] and Spatio-Temporal Retinex-Inspired Envelope with Stochastic Sampling (STRESS) [10], we introduce an approach which combines edge-aware image processing with STRESS for capsule video image enhancement algorithm. The main contribution of this work can be summarized as follows. Firstly, we propose a unified framework that can enhance shadows and detailed structures for capsule endoscopy images. Using deep edge aware filters that are pre-trained on millions of flicker images for base layer and STRESS for detail layer, our method provides details of tissue surfaces and reveals shadow regions Fig. (1). Secondly, based on various learned filtering effects such as bilateral [11] and L_0 smoothing [12], our framework gives different effects to give a better visualization of tissue surfaces.

This paper is organized as follows: first we introduce earlier works for capsule image enhancement, then we present our framework along deep edge aware filtering and STRESS, followed by results and conclusion.

Background

The GI tract has different surface features that reflect or absorb the light from the capsule in different ways. The reflectance properties of a tissue surface depend on the particular material (blood, tumors, inflammation, etc.). Narrow band imaging (NBI) [2] uses this concept to enhance the mucosa and vasculature. This is done by using two set of filters that are interposed after the light

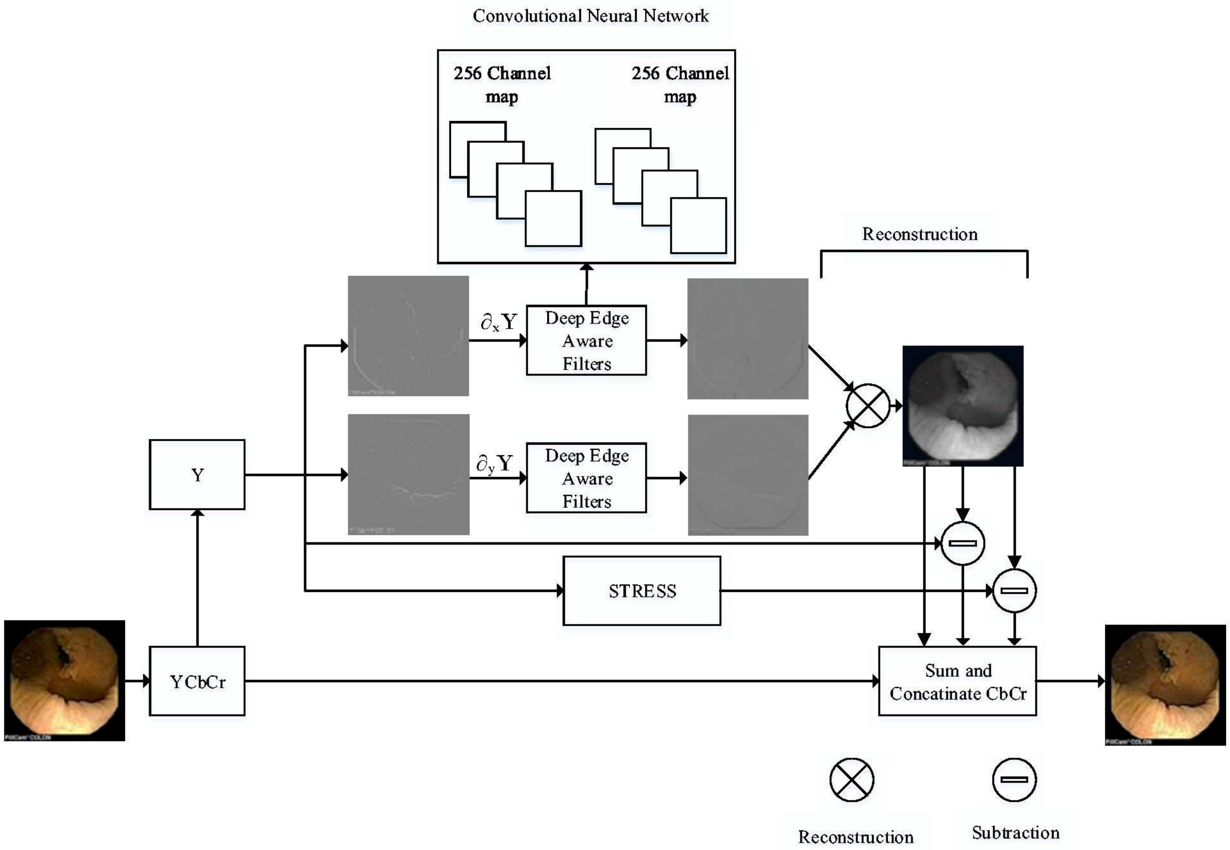


Figure 2: The proposed Deep-STRESS capsule video endoscopy image enhancement. Each deep edge aware filter is a two layer neural network with shared weights and takes gradient of the luminance channel as an input $(\partial_x Y, \partial_y Y)$ and gives smoothed gradients. The original input image is used for guided smoothed image reconstruction. The final image is computed using convex summation and is described under proposed method.

source to restrict the incident light to two narrow bands of wavelengths. Similarly, flexible spectral imaging color enhancement (FICE) [1] is a system that estimates the spectral reflectivity of the target tissue, reconstructs flexible spectral images with different wavelengths calculated from conventional white light (WL) images, and develops new flexible spectral images by selecting and reconstructing RGB wavelengths that emphasize the target [13]. These methods generally require modifying the camera design in former case or knowing spectral properties of the light source and camera characteristics in latter case. Among recent computational methods that are proposed for capsule endoscopy image enhancement include Rukundo et al. [3], where the authors proposed an algorithm that uses half-unit weighted-bilinear for darker areas and threshold weighted-bilinear method to avoid overexposure and enlargement of specular highlight spots while preserving the hue. Okuhata et al. [14] proposed a real-time image enhancement technique that is based on the variational approach of the Retinex theory. Moreover, in order to deal with uneven illumination and poor contrast and to reduce highlighted areas as much as possible, gamma correction, masking and histogram equalization, which are categorized into an image enhancement technique, can be found in the literature.

Deep edge aware filters

Let's denote the input color image as I and edge-aware filtering operator as $\psi(I)$. $\psi(I)$ could be any non-linear process that operates locally or globally in spatial image coordinate. The main objective is to approximate this operator by minimizing a loss function as defined in Eq. (1). Let D_N be the parameters of our deep neural network. Given T training images pairs of $(I, \psi(I))$, the parameters of the neural network can be estimated by minimizing the following function.

$$\frac{1}{T}\sum_i\left\{\frac{1}{2}\|D_N(\partial I_i)-\partial\psi(I_i)\|^2+\lambda\phi\left(D_N(\partial I_i)\right)\right\} \tag{1}$$

where $\phi(D_N(\partial I_i))$ is a sparse regularization term, i.e $\phi(z) = (z^2+\varepsilon^2)^{\frac{1}{2}}$ and λ is the regularization weight. Note that, the optimization is done using squared difference of gradient images rather color squared difference as the latter case results in blurring and contains unwanted details [9].

STRESS

The human visual system exploits the statistical structure of natural images in adjusting local contrast. For example, a relative bright region in a very bright part of an image can appear darker.

The main objective of the STRESS framework [10] is to estimate for each pixel, the local reference lightness and darkness points. Hence, the local reference lightness and darkness points makes two envelope functions, the maximum and minimum envelopes, completely containing the image signal. In order to estimate these envelops, let us assume for each pixel x_0 the maximum and minimum envelopes be E_{max} and E_{min}, respectively. In order to get a robust estimate for the envelopes, M number of pixels are sampled and this is repeated N number of iterations. Hence the maximum and minimum envelope pixel for each iteration n can be estimated.

$$\begin{aligned} I^n_{\max} &= \underset{i\in\{0,...,M\}}{\max} x^n_i \\ I^n_{\min} &= \underset{i\in\{0,...,M\}}{\min} x^n_i \end{aligned} \tag{2}$$

Similarly, the range R and relative value V of the pixel can be estimated as [10] Eq. (3)

$$V^n = \begin{cases} \frac{1}{2}, & \text{if } R_n = 0, \\ \frac{(x-I_{min})}{R^n}, & \text{otherwise} \end{cases} \text{and} \quad R^n = I^n_{\max} - I^n_{\min} \tag{3}$$

Taking the average of the relative values and range over N iterations, the average values are computed as follows,

$$\begin{aligned} \overline{v} &= \frac{1}{N}\sum_{n=1}^{N} V^n \\ \overline{r} &= \frac{1}{N}\sum_{n=1}^{N} R^n \end{aligned} \tag{4}$$

Given the average values, the local reference lightness and darkness points makes two envelope functions computed using Eq. (5).

$$\begin{aligned} E_{\min} &= x_0 - \overline{v}\overline{r} \\ E_{\max} &= x_0 + (1-\overline{v})\overline{r} = E_{\min} + \overline{r} \end{aligned} \tag{5}$$

Proposed Method

Given an input RGB channel image, it is converted to the YCbCr color space and enhancement is applied to the Y channel of the image. This is done to make sure that the color information is preserved as it is crucial for diagnosis [15]. The block diagram for the proposed method is shown on Fig. (2).

Deep model

In order to decompose the input image spatially into multilayer features, a pre-trained edge-aware filter trained on flickr dataset is used [9]. We used different pre-trained models that simulate a number of base layers including bilateral filter [16], weighted least square (WLS) smoothing [17], L_0 smoothing [12], rolling guidance filter [18], RTV texture smoothing [19] and iterative bilateral filtering [20]. The dataset contains randomly collected one million flicker images. The model is trained on a pair of input image and the output of the original filter method. The pretrained model is trained on dataset that contain enough structural variation for successful CNN training in multiple layers. The pretrained edge-aware filter model takes the gradient of luminance channel $(\partial_x Y, \partial_y Y)$ and outputs a smoothed gradient for each of x and y component.

Reconstruction

The final smooth image is reconstructed from the smoothed gradient output of the deep edge aware filter. The reconstruction considers structural information in the input image to guide smoothing in the gradient domain. Given the output of edge-aware filters, the smoothed version is computed by solving a quadratic loss function similar to [9]. For completeness, the loss function is given below in Eq. (6).

$$\|I_l - I_o\|^2 + \beta\left\{\|\partial_x I_l - D_N{}'(\partial_x I_o)\|^2 + \|\partial_y I_l - D_N{}'(\partial_y I_o)\|^2\right\} \tag{6}$$

where I_0 is the original image and I_l is as computed from any of the learned filters. ∂_x and ∂_y represent the partial derivative in x and y, respectively. The first term in Eq. (6) represents similarity between input and reconstructed smooth image, while the second term represents gradient constancy term and β is a parameter balancing the two losses. Eq. (6) is solved as described in [9].

Fusion

A directed LED light source mounted on capsule endoscopy casts shadow in the area projected by the tissue in the direction of directed light. Moreover, due to low lighting condition and camera sensor, capsule images lacks image details and contain self shadow regions. In order enhance the details and shadow regions, we fused lightness, detail and base layers to estimate the final enhanced image. The surface detail information that is provided by the deep learned filters and STRESS are complementary in that the former provides base layer keeping the general macro structures, while STRESS providing local lightness and darkness pixel information. This is due to stochastic sampling explore the image context around the pixel in search of the local reference for the pixel adjustment. Deep edge aware filters adjust pixels intensity to local mean and keeping edges. Given the above two local adjustment for the target pixel, the final pixel value is estimated using the difference in the intensity adjustment. The proposed framework is shown on Fig. (2). The surface details of the image can be estimated by removing the base layer from the original image and shadow details are computed by subtracting from the local contrast enhanced image. This can represented as

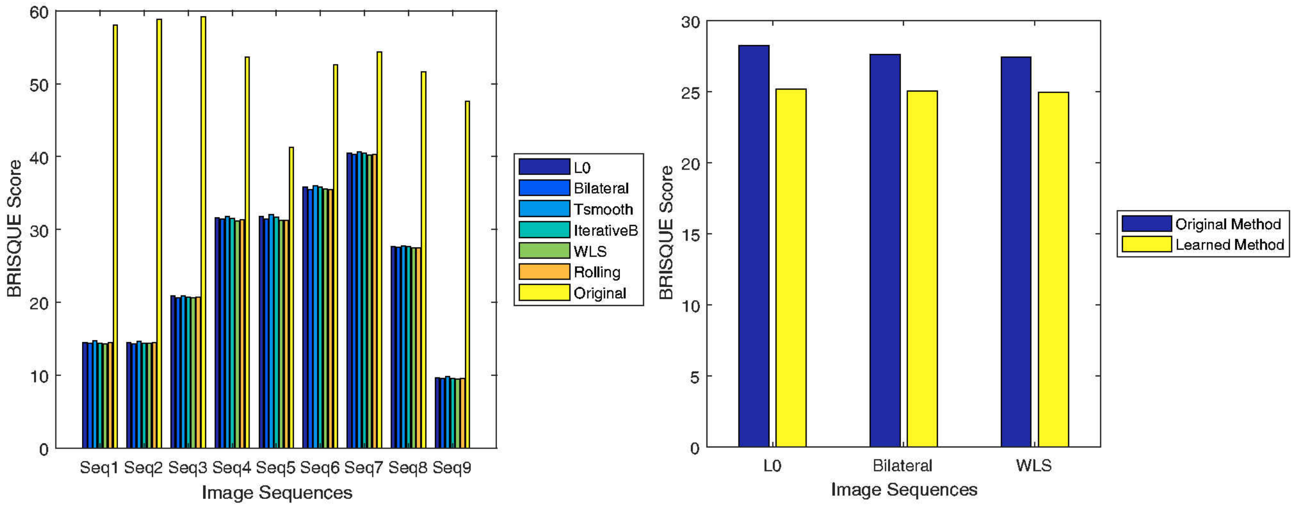
$$\begin{aligned} D_1 &= I_o - I_l \\ D_2 &= I_{ST} - I_l \end{aligned} \tag{7}$$

where I_0 is the original image and I_l is as computed from any of the learned filters. Hence D_1 contains detail image surface information. In the above Eq. (7), I_{ST} stands for local contrast enhanced image. The final detail and shadow information is computed by taking the convex linear combination of the D_1 and D_2.

Finally, the enhanced image is given by, summing the base layer and final detail and shadow information as in Eq. (8)

$$\begin{aligned} I_{enh} &= KD + I_{learned} \\ D &= \gamma D_1 + (1-\gamma)D_2 \end{aligned} \tag{8}$$

where γ is trade-off between detail layer D_1 and shadow layer that is computed from contrast enhanced version of our input image D_2 and K is a multiplier constant. While increasing γ gives more



(a) Learned filter comparison.

(b) Learned filter vs Original method.

Figure 3: Average BRISQUE score for each sequence in our test set (lower value indicates better quality). At least three frames are selected from nine sequences containing medical findings. The base layers are computed using pre-trained weights [9]. L0, Bilateral, Tsmooth, IterativeB , WLS, Rolling, Original stands for learned L_0 norm filter, bilateral, RTV texture smoothing, iterative bilateral, weighted least square, rolling guidance filter and input image respectively. Low BRISQUE score indicates higher image quality

details, the value of K controls the overall amplification of the detail and shadow layer.

For STRESS, we used 20 number of iterations and 250 samples. $K = 2.5$ and $\gamma = 0.7$ is set for all our experiment in Eq. (8).

Experimental Evaluation

A study done on objective image quality metric [21] suggests a metric that correspond to diagnostic qualities of capsule images. In their experiment conducted to assess image quality metrics for evaluation of capsule video endoscopy enhancement techniques [21], the BRISQUE [22] metric is found to have a good correlation with a subjective experiment conducted with a medical doctor. BRISQUE is a natural scene statistic-based distortion-generic blind/no-reference image quality metrics which operates in the spatial domain. BRISQUE does not compute distortion specific features such as blocking, blur, ringing, but rather uses scene statistics of locally normalized luminance coefficients to quantify possible losses of ”naturalness” in the image due to the presence of distortions [21]. In this section, we evaluate our proposed framework using different learned filters and give a visual comparison. In order to validate the proposed approach we used nine video samples from [23] and can be downloaded from `https://www.ntnu.edu/web/colourlab/software`. A total of 37 representative informative frames with pathologies are chosen from these sequences. 14 of these video frames contain polyps or tumors. The sample sequence contains images were captured by Given Imaging Pillcam COLON capsules and Mirocam capsules. The BRISQUE score for each of learned filters is shown on Fig. (3). Among the learned filters, rolling guidance filter performs better in-terms of average BRISQUE score over all sequences with maximum average difference of 0.4239 with the least performing learned filter, RTV texture smoothing. From Fig. (3), the learned filters capture similar base layer as measured by the BRISQUE metric. However, as the framework is flexible and it is possible to choose the best learned filter that gives the best visualization for the doctor. That is, by picking the best performing base layer, diagnostically useful information can be enhanced for the gastroenterologist in the sense of sharpness of image details, contrast, with original colors tones and no artifacts.

In addition, we compare the learned filters with their corresponding original proposed method for bilateral filter [16], weighted least square (WLS) smoothing [17], L_0 smoothing [12] algorithms. The result is summarized in Fig. (3b) and it can be seen that the learned filters give a better BRIQUE score compared to their corresponding original method. This is expected in that, the pre-trained models are trained using a optimal parameters of the original method. From Fig. (3b), we can conclude that under optimal parameters for classical Bilateral, WLS and L_0 filters, WLS and Bilateral filters gives the best result in-terms of BRIQUE score. Moreover, the six edge-aware filtering operators have a nearly identical trend on BRISQUE scores as they are trained using optimal parameters of their corresponding original method. The corresponding original filters also have nearly identical trend on BRISQUE as shown in Fig. (3a), under optimal parameters for each methods. It is important to mention that, if different effect is required, the pre-trained model should be retrained with new parameters for each of the six-filters. Sample visual results are provided in Fig. (4) for the best performing methods for each image. As can be seen on Fig. (4), our proposed approach enhances the detail of the tissue surfaces as well as distant parts without the image looking washed-out. Moreover, as it is evident from Fig. (5), showing histogram plots of the Y channel, the proposed method gives a trade-off result between local and global contrast enhancement methods, by shifting the dark

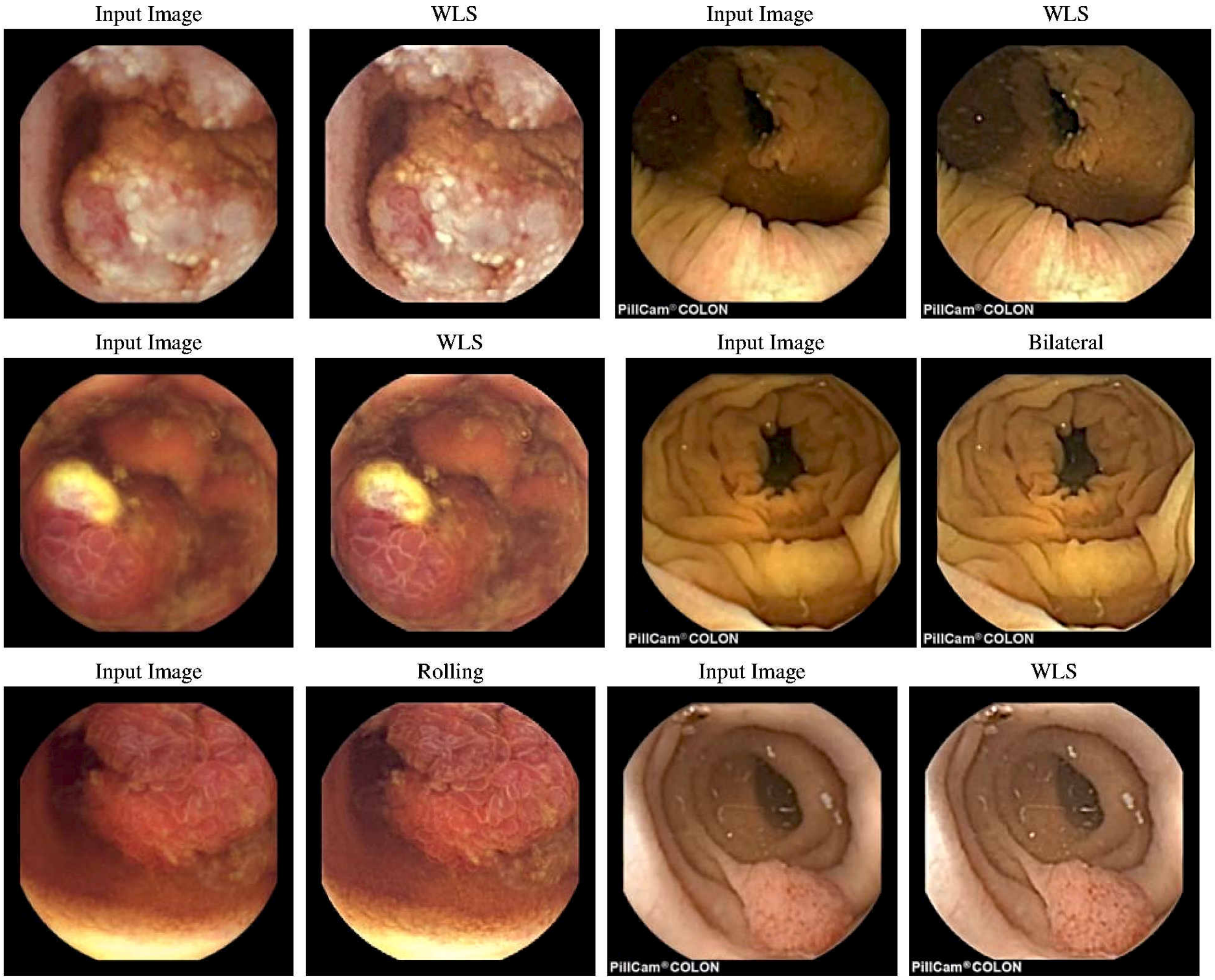
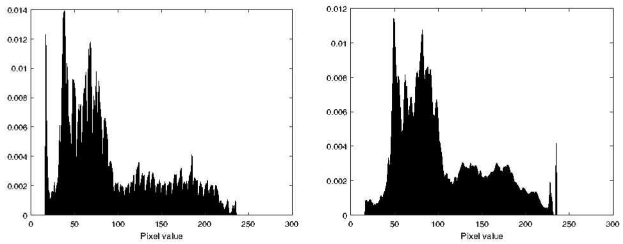


Figure 4: Visual comparison of best performing learned filters result in our framework. The first two columns are from Mirocam capsules and the the last two column image are from GivenImaging Pillcam COLON. Each image is enhanced with all the learned filters and the filter with maximum BRISQUE score is shown. As can be seen both the distant parts and details are clearly visible.

regions of the image to mean value 0.5 without saturation of the intensity values.

From the above results and approach, the deep neural network is able to model different filters in a unified framework. Despite the merits, it is important to mention that the model is pre-trained for a given set of parameters of each filter. For example, for bilateral filter the parameters are $\sigma_s = 7$ for spatial and $\sigma_r = 0.1$ for the range. Hence optimizing the parameters for best result requires further preprocessing of the input data with fine tuning the network.



(a) Input Image

(b) Enhanced Image

Figure 5: Histogram comparison for input image of Fig. (4) top-right. The histogram is computed after removing the bounding black region. The image is enhanced with WLS, histogram is computed only for Y channel.

Conclusion

In this paper, we proposed a unified approach for capsule video endoscopy enhancement. Our approach, Deep-STRESS, uses classical stochastic sampling contrast enhancement technique, STRESS and deep edge aware filters to enhance the images for the gastroenterologist in the sense of sharpness of image details, contrast, with original colors tones and no artifacts. Our preliminary results and evaluation show very promising results in terms of visual result and with flexible configuration for different base layers to give optimal result. The proposed approach requires no pre-processing or post-processing. Moreover, we showed the performance of the proposed method with a no-reference image quality metric BRISQUE, increasing the original image quality.

References

[1] Eiji Sakai, Hiroki Endo, Shingo Kato, Tetsuya Matsuura, Wataru Tomeno, Leo Taniguchi, Takashi Uchiyama, Yasuo Hata, Eiji Yamada, Hidenori Ohkubo, et al. Capsule endoscopy with flexible spectral imaging color enhancement reduces the bile pigment effect and improves the detectability of small bowel lesions. *BMC gastroenterology*, 12(1):83, 2012.

[2] R Lambert, K Kuznetsov, and JF Rey. Narrow-band imaging in digestive endoscopy. *The Scientific World Journal*, 7:449–465, 2007.

[3] Olivier Rukundo, Marius Pedersen, and Øistein Hovde. Advanced image enhancement method for distant vessels and structures in capsule endoscopy. *Computational and mathematical methods in medicine*, Article ID 9813165, 2017:13 pages.

[4] Ahmed Mohammed, Ivar Farup, Marius Pedersen, Øistein Hovde, and Sule Yildirim Yayilgan. Stochastic capsule endoscopy image enhancement. *Journal of Imaging*, 4(6), 2018.

[5] Hiroyuki Okuhata, Hajime Nakamura, Shinsuke Hara, Hiroshi Tsutsui, and Takao Onoye. Application of the real-time retinex image enhancement for endoscopic images. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 3407–3410. IEEE, 2013.

[6] Haiying Liu, W-S Lu, and Max Q-H Meng. De-blurring wireless capsule endoscopy images by total variation minimization. In *Communications, Computers and Signal Processing (PacRim), 2011 IEEE Pacific Rim Conference on*, pages 102–106. IEEE, August 2011.

[7] Hai Vu, Tomio Echigo, Keiko Yagi, Hirotoshi Okazaki, Yasuhiro Fujiwara, Yasushi Yagi, and Tetsuo Arakawa. Image-enhanced capsule endoscopy preserving the original color tones. In *International MICCAI Workshop on Computational and Clinical Challenges in Abdominal Imaging*, September.

[8] Mohammad S Imtiaz and Khan A Wahid. Color enhancement in endoscopic images using adaptive sigmoid function and space variant color reproduction. *Computational and mathematical methods in medicine*, 2015:1–19, 2015.

[9] Li Xu, Jimmy Ren, Qiong Yan, Renjie Liao, and Jiaya Jia. Deep edge-aware filters. In *International Conference on Machine Learning*, pages 1669–1678, France, July 2015.

[10] Øyvind Kolås, Ivar Farup, and Alessandro Rizzi. Spatio-temporal retinex-inspired envelope with stochastic sampling: a framework for spatial color algorithms. *Journal of Imaging Science and Technology*, 55(4):40503–1, 2011.

[11] Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, pages 839–846. IEEE, Jan 1998.

[12] Li Xu, Cewu Lu, Yi Xu, and Jiaya Jia. Image smoothing via l0 gradient minimization. In *ACM Transactions on Graphics (TOG)*, volume 30, pages 174–180. ACM, 2011.

[13] Yasushi Sato, Tamotsu Sagawa, Masahiro Hirakawa, Hiroyuki Ohnuma, Takahiro Osuga, Yutaka Okagawa, Fumito Tamura, Hiroto Horiguchi, Kohichi Takada, Tsuyoshi Hayashi, et al. Clinical utility of capsule endoscopy with flexible spectral imaging color enhancement for diagnosis of small bowel lesions. *Endoscopy international open*, 2(2):E80, 2014.

[14] Hiroyuki Okuhata, Hajime Nakamura, Shinsuke Hara, Hiroshi Tsutsui, and Takao Onoye. Application of the real-time retinex image enhancement for endoscopic images. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 3407–3410, Japan, July 2013. IEEE.

[15] Mihai Rimbaş, Lucian Negreanu, Lidia Ciobanu, Andreea Benguş, Cristiano Spada, Cristian Răsvan Băicuş, and Guido Costamagna. Is virtual chromoendoscopy useful in the evaluation of subtle ulcerative small-bowel lesions detected by video capsule endoscopy? *Endoscopy international open*, 3(6):E615, 2015.

[16] Sylvain Paris and Frédo Durand. A fast approximation of the bilateral filter using a signal processing approach. In *European conference on computer vision*, pages 568–580, Austria, May 2006. Springer.

[17] Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. In *ACM Transactions on Graphics (TOG)*, volume 27, pages 67–78. ACM, 2008.

[18] Qi Zhang, Xiaoyong Shen, Li Xu, and Jiaya Jia. Rolling guidance filter. In *European conference on computer vision*, pages 815–830, Germany, Sep 2014. Springer.

[19] Li Xu, Qiong Yan, Yang Xia, and Jiaya Jia. Structure extraction from texture via relative total variation. *ACM Transactions on Graphics (TOG)*, 31(6):139–142, 2012.

[20] Raanan Fattal, Maneesh Agrawala, and Szymon Rusinkiewicz. Multiscale shape and detail enhancement from multi-light image collections. *ACM Trans. Graph.*, 26(3):51–59, 2007.

[21] Marius Pedersen, Olga Cherepkova, and Ahmed Mohammed. Image quality metrics for the evaluation and optimization of capsule video endoscopy enhancement techniques. *Journal of Imaging Science and Technology*, 61(4):40402–1, 2017.

[22] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012.

[23] A. Koulaouzidis and D.K. Iakovidis. Kid: Koulaouzidis-iakovidis database for capsule endoscopy. `http://is-innovation.eu/kid`. (Accessed on Aug 2016).