Colorectal polyp region extraction using saliency detection network with neutrosophic enhancement

Keli Hu, Liping Zhao, Sheng Feng, Shengdong Zhang, Qianwei Zhou, Xiaozhi Gao, Yanhui Guo

PII: S0010-4825(22)00534-0

DOI: https://doi.org/10.1016/j.compbiomed.2022.105760

Reference: CBM 105760

To appear in: *Computers in Biology and Medicine*

Received Date: 27 February 2022

Revised Date: 2 June 2022

Accepted Date: 18 June 2022

Please cite this article as: K. Hu, L. Zhao, S. Feng, S. Zhang, Q. Zhou, X. Gao, Y. Guo, Colorectal polyp region extraction using saliency detection network with neutrosophic enhancement, *Computers in Biology and Medicine* (2022), doi: https://doi.org/10.1016/j.compbiomed.2022.105760.

Colorectal Polyp Region Extraction Using Saliency Detection Network with Neutrosophic Enhancement

Keli Hu[a,b], Liping Zhao[a], Sheng Feng[a], Shengdong Zhang[a], Qianwei Zhou*[c],

Xiaozhi Gao[d] , Yanhui Guo[e]

a Department of Computer Science and Engineering, Shaoxing University, Shaoxing 312000, PR China

b Cancer Center, Department of Gastroenterology, Zhejiang Provincial People's Hospital (Affiliated People's Hospital, Hangzhou Medical College), Hangzhou 310014, PR China.

c College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, PR China

d University of Eastern Finland, Finland

e University of Illinois Springfield, One University Plaza, Springfield, IL, 62703, USA

E-mail: zqw@zjut.edu.cn

***Abstract*—Colorectal polyp recognition is crucial for early colorectal cancer detection and treatment. Colonoscopy is always employed for colorectal polyp scanning. However, one out of four polyps may be ignored, due to the similarity of polyp and normal tissue. In this paper, we present a novel method called NeutSS-PLP for polyp region extraction in colonoscopy images using a short connected saliency detection network with neutrosophic enhancement. We first utilize the neutrosophic theory to enhance the quality of specular reflections detection in the colonoscopy images. We develop the local and global threshold criteria in the single-valued neutrosophic set (SVNS) domain and define the corresponding *T* (Truth), *I* (Indeterminacy), and *F* (Falsity) functions for each criterion. The well-built neutrosophic images are processed and employed for specular reflection detection and suppressing. Next, we introduce two-level short connections into the saliency detection network, aiming to take advantage of the multi-level and multi-scale features extracted from different stages of the network. Experimental results conducted on two public colorectal polyp datasets achieve 0.877 and 0.9135 mIoU for polyp extraction respectively, and our method performs better compared with several *state-of-the-art* saliency networks and semantic segmentation networks, which demonstrate the effectiveness of applying the saliency detection mechanism for colorectal polyp region extraction.**

***Index Terms*—Colorectal polyp, Polyp recognition, Polyp segmentation, Saliency detection, Short connection.**

I. INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer diagnosed in both genders in the United States[1], and it is estimated that there will be 104,270 new cases of colon cancer, and 45,230 new cases of rectal cancer in 2021. In addition, CRC is the second most common cause of cancer deaths in the United States, and it is expected to cause about 52,980 deaths during 2021. According to the reports published by Globocan'2020, CRC is also the third most common cancer in both genders across the globe[2] and the second most prevailing type in terms of mortality globally [10]. The total number of deaths caused by cancer is 9.9 million in 2020, and CRC-related deaths take up 9.4% [10]. Colorectal polyps are the precursors for CRC, and they are abnormal tissues growing in the mucosal of the colon. A polyp may develop into cancer after 10 to 15 years if untreated [13]. Early diagnosis and proper treatment are crucial for reducing the probability of CRC. Each 1.0% increase in the polyp detection rate will lead to a 3% reduction in the risk of CRC [14]. Colonoscopy is the gold standard method for investigating the gastrointestinal tract. It is employed for discovering polyps and resecting them before they transform into cancer. However, coloscopy is a highly operator-dependent procedure, and one out of four polyps may be ignored in a single colonoscopy due to the human factors like clinicians' skills or subjectivity [15]. Additionally, evidence shows that missing cancer or incomplete resected lesions are the two key factors when post-colonoscopy cancer occurs [16]. Extracting the polyp regions highly depends on doctors' attention and practical experiences, and it is very time-consuming. In addition to polyp detection, accurately segmenting the polyp becomes another crucial issue for

[1] https://www.cancer.org/cancer/colon-rectal-cancer/about/key-statistics/
[2] https://gco.iarc.fr/today/home

improving the prognosis and survival rate of CRC. Therefore, a computer-aided fully automated system with the ability for accurate polyp extraction is highly desirable. Such a polyp extraction system can provide an efficient way to reduce medical errors and release clinicians from the heavy workload.

Automation of colorectal polyp region extraction is a quite demanding task with two main challenges. Firstly, the interclass difference is small. The appearance inside the polyp area is similar to the normal area of the intestinal wall in colonoscopy images, and there exists limited interclass variation between polyps and folds or vessels [17]. In addition, the boundary between the polyp and the surrounding mucosa is sometimes very confusing, which considerably increases the difficulty of polyp region extraction. Secondly, the intraclass variation of polyps is rather large in terms of locations, sizes, and morphologies. The problem of specular reflections further aggravates this challenging issue [18]. Moreover, bubbles and fecal particles challenge the effectiveness of the automatic extraction of the polyp region.

Many efforts have been made to tackle the aforementioned issues. In the early stage, hand-engineered methods [19-23] are utilized to model the discriminative features and improve the performances of the polyp detection approaches. Features like boundary information, context, color, shape, and texture are frequently employed to feed the recognition model. The hand-engineered characterizations highly depend on the experience of scientists, it is quite challenging to design an adequate feature vector for robust polyp description[21]. In recent years, deep learning-based approaches have shown prominent performance in lots of fields, such as semantic relation extraction [24], medical lesion segmentation [25, 26], and natural image analysis[4, 7, 27]. Benefiting from the deep learning technology, robust features can be automatically learned by a deep network for specific tasks. The convolutional neural network (CNN) based strategy is widely adopted in the problem of extracting more powerful features for polyp region extraction [28]. Currently, most existing methods treat the polyp region extraction as an image segmentation issue [29-34]. Though such a scheme achieves a large gain than the traditional mechanism which employs hand-engineered methods, there still exist many problems left to tackle, and its performance is still far from that of professional doctors. Meanwhile, the saliency scheme shows its effectiveness for polyp region extraction [18, 35, 36], these results prompt us to introduce the CNN-based saliency detection models into such a challenging issue. To fully utilize features in different levels, as well as different stages of a network, we propose a saliency detection network with two-level short connections to handle the colorectal polyp region extraction problem.

Due to the light from the colonoscopy and the shiny surface of the GI tract, specular reflections usually exist on the wall of the GI tract. These reflections may result in the misdiagnosis of polyps [31]. Thus, lots of efforts have been done to remove the specular highlights at the preprocessing stage when the computer-aided mechanism is considered for polyp detection [18, 19, 21, 22, 31]. Specular detection is the key step for the removal of the specular highlights. Currently, only simple features are considered. For instance, the source image was usually converted into a specific color space. The intensity in the grayscale domain, the

saturation channel in HSI or HSV color space, and the value element in HSV color space were frequently employed. However, the indeterminacy attribution has not been considered yet, and a simple thresholding scheme was usually used for detecting the regions of specular reflections. Neutrosophic (NS) theory is a new branch of philosophy with the ability to deal with the origin, nature, and scope of neutralities [37]. Lots of evidence have shown its effectiveness in modeling indeterminate information [38]. To facilitate the detection of the specular region and better suppress the specular reflections, we employ the Single-Valued Neutrosophic Set (SVNS)-based multi-criteria decision-making scheme to fully utilize the color feature, as well as the indeterminate information.

In this work, we present an effective polyp region extraction strategy (NeutSS-PLP for short) by employing the scheme of saliency detection and the theory of neutrosophic. The key contributions of our work are given as follows:

- We define the local and global threshold criteria as well as the corresponding *T*, *I*, and *F* functions to transfer the specular detection problem to the SVNS-based multi-criteria decision-making problem.
- We apply the CNN-based saliency detection network to deal with the colorectal polyp region extraction problem.
- We propose a saliency detection-based polyp extraction network with two-level short connections to take advantage of the multi-level and multi-scale features extracted from different stages of the network.

The remainder of this paper is organized as follows. Section II illustrates related work. Section III explains the details of the SVNS-based specular reflection suppressing and the architecture of the NeutSS-PLP network based on the saliency detection with two-level short connections. Section IV presents and discussed the experimental results. Finally, some remarks and conclusions are given in Section V.

II. Related Work

A. *Hand-engineered Methods for Polyp Extraction*

Hand-engineered methods are widely employed to model the polyp features distinct from its surroundings in the early stage. Ganz *et al.* [19] adopt the prior knowledge of the shape information of the polyps into a boundary detection and segmentation algorithm. Since the specular reflections may distract the boundary detector, the saturation and value element in HSV color space are jointly employed for extracting the regions of specular highlights, and a simple thresholding strategy is applied in [19]. The images after inpainting are used for the next stages of the approach. The boundary information is also introduced into the work of Bernal *et al.* [20] for polyp localization. The valley information is integrated to foster a better boundary of the polyps associated with the concave and continuous attribution in [20]. In addition to the shape information, Tajbakhsh *et al.* [21] utilize the context information to remove the non-polyp structure. The problem of specular reflections is also considered in [21]. Unlike the strategy in [19], Tajbakhsh *et al.* use the knowledge of the distinct appearance around the polyp the specular highlights boundaries. Gonzalez *et al.* [22] propose a novel method for polyp segmentation by jointly employing the characterization of the shape, color,

and curvature of edges. The threshold calculated by the average of the original image in grayscale is applied for extracting the specular region by thresholding in [22]. The shape, color, and texture information are considered in [23] for polyp extraction. Different from the aforementioned work, a unified bottom-up and top-down saliency approach [18] is proposed for polyp detection, and it can yield a good performance, which proves that the scheme of saliency detection is suitable for handling the automatic polyp recognition problem. For achieving even better performances, the specular reflections detection and image inpainting are also introduced into the step of image painting in [18]. Similar to the works in [19, 22], the thresholding in HSI color space on the channel of saturation and intensity is applied in the work of Yuan *et al.* [18] for specular reflections detection. Though the hand-engineered methods have achieved great improvement, they are far from sufficient to discriminate the polyps from the challenging surroundings [21].

B. Deep Learning for Polyp Segmentation

Due to the power of feature representation and analysis of the deep learning model, lots of deep learning-based methods are proposed for extracting the polyp region. Brandao *et al.* [29] introduced the fully-connected convolution network (FCN) into the polyp region extraction issue by converting AlexNet, VGG, and ResNets into FCNs. In addition, the generated depth information is fed into the networks with the RGB information. Evidence in [29] proved the depth information boosts performance. SegNet was applied in the work of Wang *et al.* [30] for polyps extraction. Considering the problem of object rotation, Yuan *et al.* [39] proposed RIIS-DenseNet for polyp recognition with a joint loss function by applying the rotation-invariance and the image similarity regularization constraints. Several CNN-based approaches show that error predictions may occur in the inference of specular highlights, such as the ResNet50-FCN [29] and DeepLab v3+ [40] networks. Recently, several complicated CNN-based schemes have been proposed for achieving more precise pixel-wise polyp segmentation. A dual-tree wavelet pooled CNN combined with a new method of the region-based level-set was proposed in [31], and an approach of specular reflections detection and suppressing was employed in the preprocessing step to decrease the probability of miss-diagnosis of polyps. Fan *et al.* [32] introduce the attention mechanism into PraNet. A global map was generated in [32] by utilizing the high-level features, and the global map combined with the boundary cues is employed for guiding the polyp extraction. The boundary cues were also considered in the work of Fang *et al.* [33] for polyp segmentation. Fang *et al.* [33] utilize a shared encoder and two mutual constrained decoders of area-boundary branches to sufficiently exploit multi-scale information. To further improve the polyp segmentation accuracy by sematic guiding, the extension of the faster R-CNN is applied for predicting the polyp proposal, and the semantic information generated in the stage of polyp proposal is transferred to the task of segmentation to facilitate polyp extraction results. Jha *et al.* [34] introduce ResUNet into the polyp extraction issue and proved the conditional random field and the test-time augmentation can facilitate the performance. All the above work for the polyp extraction belong to the class of CNN-based segmentation mechanism, and there still exists a big gap between current performance than that of a professional physician. Different from the schemes

mentioned above, we introduce the saliency detection mechanism into this issue, and results show the effectiveness when employing a saliency detection network.

C. *Saliency Analysis for Polyp Extraction*

Since the polyp can be regarded as the salient object in colonoscopy images, several relevant papers have been published [18, 35, 36]. Sasmal *et al.* [35] use the visual saliency map for the rough detection of the polyp region, and the modified particle filter tracking mechanism based on the generated saliency information is employed for the task of polyp extraction in a colonoscopic video. The saliency map applied in [35] is predicted on the basis of the fusion of the frequency, color, and location prior information. A traditional saliency detection algorithm is adopted by Deeba *et al.* [36] for saliency region detection. The candidate saliency proposals were then utilized for the final classification step by extracting the HOG feature in the corresponding regions in [36]. These works have proved the potential effectiveness of the saliency detection mechanism for the problem of computer aided polyp region extraction. The CNN-based saliency detection networks have achieved great progress than the traditional ones [41]. However, to our best knowledge, the CNN-based saliency detection network has not been deployed to deal with such a challenging polyp extraction issue. Instead of directly utilizing a traditional network, we propose a new saliency detection network for polyp extraction, which utilizes two-level short connections to take advantage of the multi-level and multi-scale features extracted from different stages of a network.

D. *Neutrosophic Set*

SVNS is a subset of the neutrosophic set [42], which provides an additional probability to represent uncertainty, imprecise, incomplete, and inconsistent information in the real world. NS has been widely applied in a lot of fields to deal with uncertain problems, such as image segmentation [43, 44], object tracking [45, 46], medical analysis [47-49], sentiment analysis [50], and multi-focus image fusion [51]. Both NS-based decision-making [52] and similarity measure [49, 53] are frequently employed for dealing with information fusion, as well as finding the solution deemed to be satisfactory. To transfer the information into the NS domain, the corresponding *T* (Truth), *I* (Indeterminacy), and *F* (Falsity) functions are usually defined [43-45, 47, 48]. Due to the ability of NS in tackling uncertainty problems, it has shown its effectiveness when confronting noisy circumstances [47]. To better reduce the inference caused by the specular reflections, and improve the performance of specular reflections detection and suppressing, the SVNS is introduced into this work for fusing the color feature and the indeterminate information.

III. METHODS

The structure of the proposed NeutSS-PLP is shown in Fig. 1. To decrease the interference of the specular reflections to the polyp detection net, a new method for specular reflection detection based on SVNS is developed. Furthermore, to reduce the risk of losing detailed information about the polyp after specular reflections suppression, both the source image and the image after the

suppression of specular reflections are employed to construct the training set. In each iteration, images are randomly selected for training. To speed up the extraction of the polyp region, only the source image is employed for testing. So that the cost of specular reflections suppressing can be saved during the inference. Additionally, to fully use the features extracted from different levels and scales, we have developed a saliency detection network with two-level short connections. All these strategies are employed in our work, aiming at achieving enhanced performances in dealing with normal colonoscopy images.

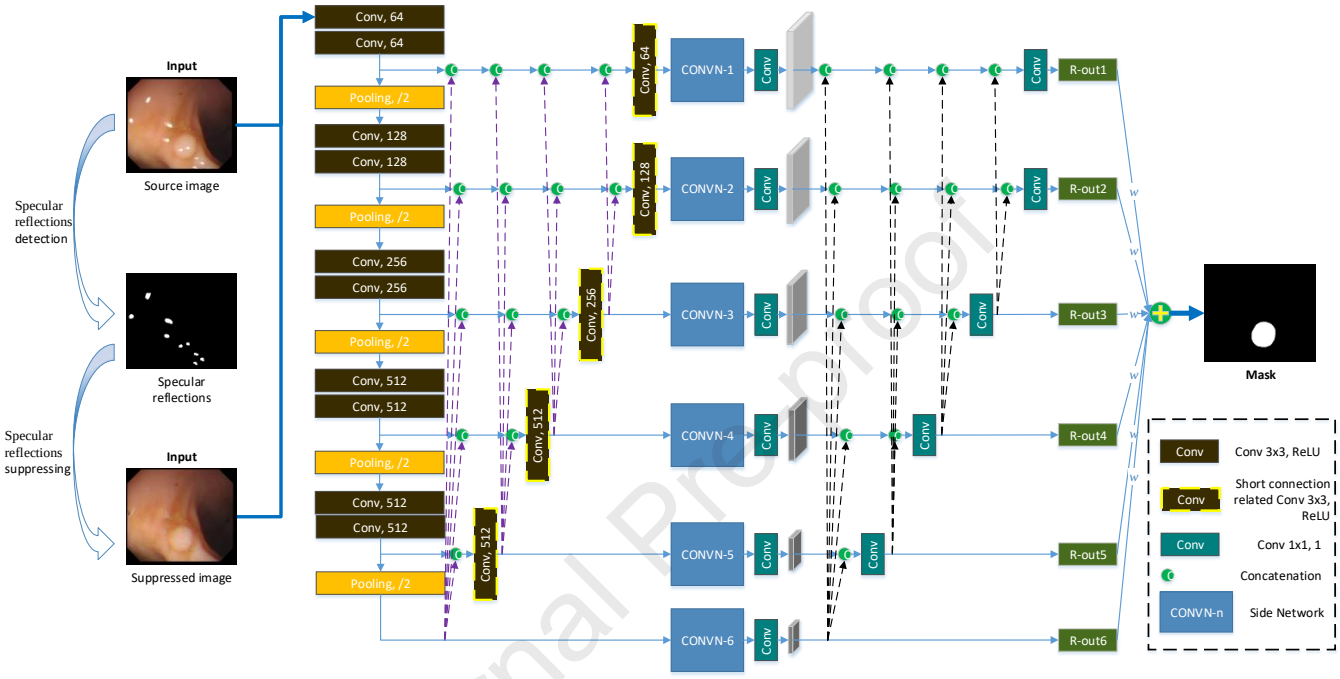


Fig. 1. Architecture of the proposed NeutSS-PLP for polyp region extraction. It is comprised of two stages: specular reflections detection and suppression stage, and polyp region extraction stage. The VGG net[2] is employed as the backbone. Two-level short connections are introduced into the saliency detection network, namely NeutSS-PLP net. Each dashed line with an arrow is selectable, which means each dashed line can be deleted, and each remaining dashed line means a short connection. The purple dashed lines correspond to low-level short connections, and the black dashed lines correspond to high-level short connections. The dashed bounding box means that the convolutional layer only exists when there are corresponding short connections.

A. Specular Reflections Suppressing

Specular reflection frequently exists in colonoscopy images, and the saliency detection network may be disturbed if the images with high reflectivity are directly fed into it. Therefore, we try to suppress the specular reflections by recovering the corresponding regions, as seen the suppressed image in Fig. 1. Specifically, we first employ the SVNS scheme for the task of specular reflections detection, then recover the detected specular regions by using a method of dynamic searching based inpainting.

1) **SVNS**

The NS theory is proposed by Smarandache [37], which is useful for handling fuzzy information, especially for indeterminacy information. However, it is very difficult to utilize traditional NS theory to deal with practical problems due to each membership degree for an alternative in the traditional NS theory is a non-standard interval. The corresponding membership degree is with a specific value in the SVNS theory, which brings the NS theory into practical problems. Suppose A is an alternative in the SVNS

domain, it can be represented as

$$A=\left\{\left\langle C_i, T_{Ci}, I_{Ci}, F_{Ci}\right\rangle\right\}, i=1,...,n\,, \tag{1}$$

where C_i is the i-th criterion, T_{Ci} denotes the degree of A satisfying with C_i, I_{Ci} indicates the indeterminacy of A when the criterion C_i, F_{Ci} presents the degree to which A does not satisfy C_i. All these three memberships locate in the range of [0,1].

The cosine similarity score [54] has been widely used for evaluating the similarity between two alternatives in the SVNS domain. Suppose the alternative $A'=\left\{\left\langle C_i, T'_{Ci}, I'_{Ci}, F'_{Ci}\right\rangle\right\}$, the corresponding cosine similarity can be calculated by

$$Cs=\sum_{i=1}^{n} w_i \frac{T_{Ci}T'_{Ci}+I_{Ci}I'_{Ci}+F_{Ci}F'_{Ci}}{\sqrt{T_{Ci}{}^2+I_{Ci}{}^2+F_{Ci}{}^2}\sqrt{T'^2_{Ci}+I'^2_{Ci}+F'^2_{Ci}}}\,, \tag{2}$$

where w_i is the weight for the i-th criterion, $w_i\in[0,1]$, $\sum_{i=1}^{n} w_i=1$.

2) **Specular Reflections Detection via SVNS**

As shown in Fig. 2(a), pixels with high values of intensity are likely to be in specular regions. However, it is difficult to judge the attribution of the pixels near the edges of regions. To fully utilize such kind of information and achieve a better performance of the specular detector, indeterminacy information is considered. We treat the specular reflections detection task as a multi-criteria decision-making problem. For the proposition of specular pixel with relatively high intensity, T_l, I_l, and F_l represent the probabilities when the proposition is true, indeterminate, and false, respectively. Considering the local suppression criterion, the calculation of these three probabilities of pixel **x** is defined as follows:

$$T_l(\mathbf{x})=\frac{Fmid(\mathbf{x})-Fmid_{\min}}{Fmid_{\max}-Fmid_{\min}}\,, \tag{3}$$

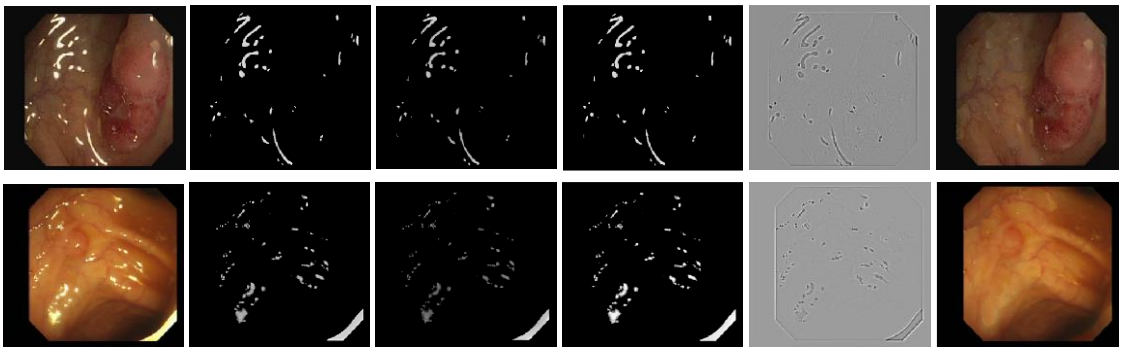
$$I_l(\mathbf{x})=\frac{Gl(\mathbf{x})-Gl_{\min}}{Gl_{\max}-Gl_{\min}}\,, \tag{4}$$

$$F_l(\mathbf{x})=1-T_l(\mathbf{x})\,, \tag{5}$$

where $Fmid(\mathbf{x})$ is the intensity value at the position of **x** on the image after intensity thresholding and differential calculation using median filter processing, $Fmid_{min}$ and $Fmid_{max}$ are the minimum and maximum values on the image $Fmid$, respectively. The Laplacian of Gaussian filtering is known as an intuitive procedure for extracting robust edge information. The parameter Gl in Eq. (4) is the gradient magnitude value at location **x** after the Laplacian of Gaussian filtering, and Gl_{min} and Gl_{max} are the corresponding minima and maxima values on the image Gl. $Fmid(\mathbf{x})$ is calculated by

$$Fmid(\mathbf{x})=\begin{cases}\left|G(\mathbf{x})-Gmid(\mathbf{x})\right| & else\\ 0 & if\,(G(\mathbf{x})<t_0)\end{cases}, \tag{6}$$

where G is the source colonoscopy image in grayscale, $G(\mathbf{x})$ is the intensity at location **x**, $Gmid(\mathbf{x})$ is the intensity at location **x** after



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

Fig. 2. Illustration of SVNS-based specular reflections suppression. Each colonoscopy image has one polyp. (a) Original images, (b) Neutrosophic images, (c) The truth-images based on the local suppression criterion, (d) The truth-images based on the global suppressing criterion, (e) The indeterminacy-images, (f) Images after specular reflections suppressing.

median filtering on G, and the fixed threshold t_0 is employed to filter out pixels that do not likely belong to bright spots. To keep the high-intensity attribution of the pixel located in the region of specular reflection, the median filter is of relatively large size. The truth-images calculated by using Eq. (3) and the indeterminacy-images calculated by Eq. (4) are shown in Figs. 2(c) and (e), respectively.

Considering the specular reflections always occupy relatively small areas in a colonoscopy image. We propose the global criterion for the decision-making problem in the SVSN architecture. The corresponding three membership functions T_g, I_g, and F_g are defined as follows:

$$T_g(\mathbf{x}) = \frac{Fm(\mathbf{x}) - Fm_{\min}}{Fm_{\max} - Fm_{\min}}, \tag{7}$$

$$I_g(\mathbf{x}) = I_l(\mathbf{x}), \tag{8}$$

$$F_g(\mathbf{x}) = 1 - T_g(\mathbf{x}), \tag{9}$$

Fm can be calculated by

$$Fm(\mathbf{x}) = \begin{cases} |G(\mathbf{x}) - mv| & else \\ 0 & if\,(G(\mathbf{x}) < t_0) \end{cases}, \tag{10}$$

$$mv = \frac{1}{|\mathbf{P}|}\sum_{\mathbf{x},\mathbf{x}\in\mathbf{P}} G(\mathbf{x}), \mathbf{P} = \left\{\mathbf{x}\middle|\mathrm{x}, G(\mathbf{x}) > t_1\right\}, \tag{11}$$

where t_1 is a fixed threshold for filtering out the non-intestinal wall area on the image, $|\mathbf{P}|$ calculates the total number of the pixels belong to the set $\mathbf{P}$, $\mathbf{P}$ is a set which consists all pixels satisfy $G(\mathbf{x}) > t_1$. As in Eq. (9), we calculate the mean intensity value of the intestinal wall-like area. Like in Eq. (4), we get the $Fm(\mathbf{x})$ by suppressing regions that are not likely to be specular reflections. The parameters Fm_{min} and Fm_{max} are the minimum and maximum values on the image Fm, respectively.

Since the ideal alternative should be with the highest truth degree and with the lowest indeterminacy and falsity degree, the ideal specular alternatives under the local and global suppression are set to be $A_{l^*} = \{\langle C_l, 1, 0, 0\rangle\}$, $A_{g^*} = \{\langle C_g, 1, 0, 0\rangle\}$, respectively.

Based on the cosine similarity measure in the SVNS domain (Eq. (2)), the neutrosophic similarity score at the position of **x** can be obtained as follows:

$$\begin{aligned} Ns(\mathbf{x}) &= w_l \frac{T_l(\mathbf{x})T_{l^*} + I_l(\mathbf{x})I_{l^*} + F_l(\mathbf{x})F_{l^*}}{\sqrt{T_l(\mathbf{x})^2 + I_l(\mathbf{x})^2 + F_l(\mathbf{x})^2}\sqrt{T_{l^*}{}^2 + I_{l^*}{}^2 + F_{l^*}{}^2}} \\ &+ w_g \frac{T_g(\mathbf{x})T_{g^*} + I_g(\mathbf{x})I_{g^*} + F_g(\mathbf{x})F_{g^*}}{\sqrt{T_g(\mathbf{x})^2 + I_g(\mathbf{x})^2 + F_g(\mathbf{x})^2}\sqrt{T_{g^*}{}^2 + I_{g^*}{}^2 + F_{g^*}{}^2}} \\ &= \frac{w_l T_l(\mathbf{x})}{\sqrt{T_l(\mathbf{x})^2 + I_l(\mathbf{x})^2 + F_l(\mathbf{x})^2}} + \frac{w_g T_g(\mathbf{x})}{\sqrt{T_g(\mathbf{x})^2 + I_g(\mathbf{x})^2 + F_g(\mathbf{x})^2}} \end{aligned}, \tag{12}$$

where w_l and w_g are the weights for the local and global suppression criteria, respectively, and $w_l + w_g = 1$.

By using Eq. (12), we can construct the neutrosophic image ***Ns***. As in Fig. 2, the ***Ns*** highlights the specular regions. Thus, the fixed thresholding method is simply applied for ***Ns***, and pixels with relatively high values are judged as regions of specular reflections.

3) **Suppressing via Image Inpainting**

The method of dynamic searching based inpainting [55] is applied for suppressing the specular highlights here. A dynamic window is employed for searching the non-specular pixels near each specular pixel. The eight-neighbors window is first considered. As in Fig. 3, the searching window with a specific size is rotated in counter-clockwise order. The search process will not stop until the required amount of information is covered. For example, at least six non-specular or painted pixels are covered by the search window. The window size will be increased from 3 to 5, 7, and 9, and the same rotating search procedure will be executed until enough non-specular or painted information is collected. As there always exists non-specular or painted pixels nearby a specular pixel, the iteration always stops on a step when the search window size is set 3 or 5. Finally, the average value of the collected pixels in RGB color channels is used to paint the corresponding specular pixel. The resulting samples are illustrated in Fig. 2(f).

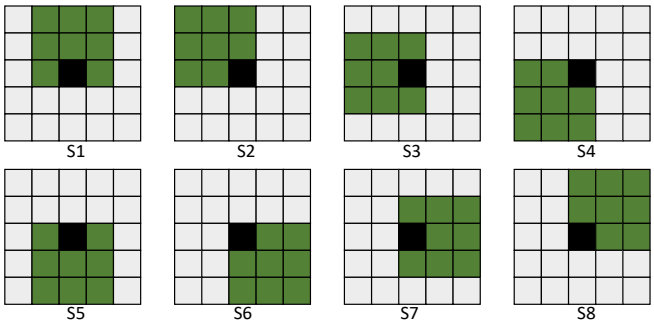


Fig. 3. Illustration of each rotating step for a 3x3 search window. The search window is adapted from Step S1 to S8 successively.

B. NeutSS-PLP Network Architecture

Salient object detection is quite challenging. A good salient object detector should have the capability of extracting the visually

distinctive regions from an image. In other words, it should be capable of fusing the multi-level and inherent features. A deep enough network can learn multi-level features, and the strategy of multiple stages with different strides helps the network learn inherent features from different scales [1, 5, 6]. Figure 4 shows that Hypercolumn network [1] simply fuses information from skip layers with different scales. HED network [6] employs side-output layers at each stage of the VGG network [2]. It has been proved that the HED architecture is not suitable for salient object detection [5]. To deal with the problem of information fusion from both shallow and deep side-outputs, short connections are applied for the DSS[5] network. As in Fig. 4(c), for each stage of the VGG network, features are directly fed back to the low-level layers behind side networks. Thus, such a top-down DSS approach can yield improved performances for saliency detection.

The UNet [56] architecture has been demonstrated to be efficient in handling the problem of medical images segmentation. The U shape method provides an effective way to transmit the information from deeper layers to shallower layers, and this approach helps the network jointly fuse multi-level and inherent features before making the final prediction. Motivated by the scheme of the DSS and UNet network, we introduce short connections on both sides of the side networks in each stage of the VGG network.

The architecture of our NeutSS-PLP network is shown in Fig. 1 and Fig. 4(d). In the following sections, we will explain the DSS architecture and introduce the scheme of two-level short connections, which is the key of the NeutSS-PLP network.

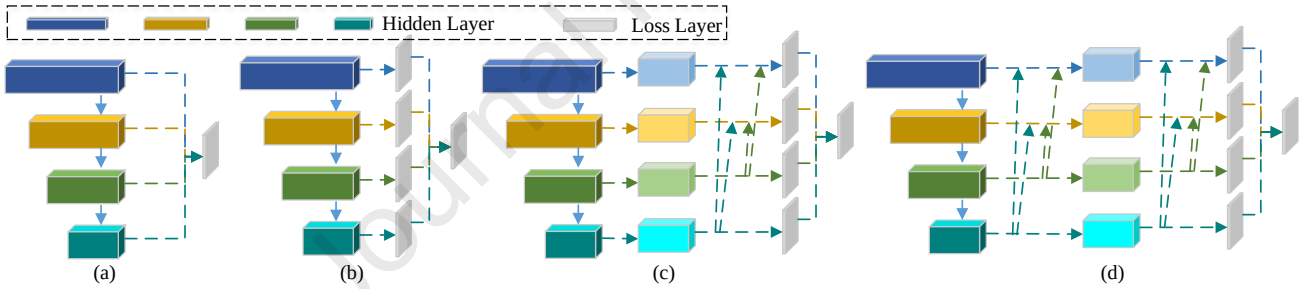


Fig. 4. Illustration of main architectures of different networks. (a) Hypercolumn [1], (b) HED [6], (c) DSS [5], (d) Ours. Short connections are introduced into our NeutSS-PLP network on both sides of side networks to take full advantage of features extracted from different stages and layers of the network.

1) **DSS architecture**

Motivated by the HED architecture to better utilize the multi-level information, the center idea of fusing information of side-output at the last convolution layer in each block of the VGG network is incorporated into the DSS network. In addition, side networks are utilized to further process the features in each side-output before the operation of short connections, as in Fig. 4(c).

Unlike the HED network that employs five side-outputs of the VGG network, the DSS architecture utilizes the side-output of the last pooling layer, since the feature produced by this layer retains some relevant location information. Suppose $\mathbf{W}$ denotes the collection of all standard network layer parameters. Each side-output is with a classifier, whose weights parameters can be represented by $\mathbf{w} = (\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \ldots, \mathbf{w}^{(M)})$, where M is the total number of the side-outputs. The loss function of the side-output in the DSS network is represented by [5]

$$L_{side}(\mathbf{W},\mathbf{w},\mathbf{r}) = \sum_{m=1}^{M} \alpha_m l_{side}^{(m)}(\mathbf{W},\mathbf{w}^{(m)},\mathbf{r}), \tag{13}$$

where α_m is the weight of the *m*th side-output loss, and $\mathbf{r}=\{r_i^m\}$, $i > m$, which is a control vector for the construction of the short connections. The side activation for the DSS network at the *m*th side-output is [5]

$$R_{side}^{(m)} = \begin{cases} \sum_{i=m+1}^{M} r_i^m R_{side}^{(i)} + A_{side}^{(m)} & else \\ A_{side}^{(m)} & if\,(m=M) \end{cases}, \tag{14}$$

where $A_{side}^{(m)}$ are activations of the *m*th side-output before connected to deeper layers with short connections, and r_i^m is the weight of the short connection from the *i*th side-output to the *m*th side-output. r_i^m can be set to be 0, if there is no need to construct the corresponding short connection. The standard cross-entropy loss is applied for computing the loss function over all pixels in an image, and the *m*th side-output loss $l_{side}^{(m)}$ in Eq. (13) is defined as

$$l_{side}^{(m)}(\mathbf{W},\mathbf{w}^{(m)},\mathbf{r}) = -\sum_{z_j \in \mathbf{Z}} z_j \log \Pr(z_j = 1 \mid X;\mathbf{W},\mathbf{w}^{(m)},\mathbf{r}) + (1-z_j)\log \Pr(z_j = 0 \mid X;\mathbf{W},\mathbf{w}^{(m)},\mathbf{r}) \tag{15}$$

where a training image $\mathbf{X}=\{\mathbf{x}_j, j=1,...,|\mathbf{X}|\}$, the corresponding ground truth saliency map $\mathbf{Z}=\{z_j, j=1,...,|\mathbf{Z}|\}$, $\Pr(z_j = 1 \mid X;\mathbf{W},\mathbf{w}^{(m)},\mathbf{r})$ denotes the probability of the activation value at location *j* in the *m*th side-output, which can be calculated by using the sigmoid function on the activations of $R_{side}^{(m)}$ at location *j*. Both the loss of the side-outputs and the final output after weighted fusion are considered in the DSS network. The fusion loss function is [5]:

$$L_{fuse}(\mathbf{W},\mathbf{w},\mathbf{f},\mathbf{r}) = \sigma(Z, \sum_{m=1}^{M} f_m R_{side}^{(m)}), \tag{16}$$

where the fusion weights $\mathbf{f}=\{f_1,...,f_M\}$, and $\sigma(\cdot,\cdot)$ denotes the function for calculating the distance between the fused predictions and the corresponding ground truth saliency map. The loss function defined in Eq. (15) is employed for obtaining $\sigma(\cdot,\cdot)$. Therefore, the final loss of the DSS saliency detection network is [5]:

$$L_{final}(\mathbf{W},\mathbf{w},\mathbf{f},\mathbf{r}) = L_{side}(\mathbf{W},\mathbf{w},\mathbf{r}) + L_{fuse}(\mathbf{W},\mathbf{w},\mathbf{f},\mathbf{r}). \tag{17}$$

2) **Two-level Short Connections**

As seen from Fig. 1 and Fig. 4, the proposed NeutSS-PLP network owns two-level short connections. Considering the deeper side-outputs are with more global information, we develop the scheme of short connection in the DSS for transmitting the high-level features to the shallower layers, whose outputs contain more details. The side networks architecture used in the DSS is employed to handle the demanding salient object detection task. Each side network contains two convolutional layers with different channels and kernel sizes. The details can be found in Table 1. Compared to the DSS network, this side network is with

half channels corresponding to each convolutional layer in DSS.

Table 1
Details Of Each Side Network.

Network	1	2
CONVN-1	3x3, 64	3x3, 64
CONVN-2	3x3, 64	3x3, 64
CONVN-3	5x5, 128	5x5, 128
CONVN-4	5x5, 256	5x5, 256
CONVN-5	5x5, 256	5x5, 256
CONVN-6	7x7, 256	7x7, 256

(k×k, N) represents the convolutional layer with N channels, and the kernel size is K. "1" and "2" denote two convolutional layers in a side network. Each convolutional layer Is followed by a Relu layer.

As shown in Fig. 1, there are two parts of short connections in our network. If we split the network into the left and right parts by each side network block, the right part corresponds to the DSS short connections architecture. When ignoring all the short connections, we can discover that the information flow goes along the side networks from the left side to the right side. In other words, features processed on the right side can be considered as high-level features. Thus, for the two-level short connections scheme, we call the left part the low-level short connections and the right part the high-level short connections.

Our new side activations $\tilde{R}_{side}^{(m)}$ at the *m*th side-output of the network can be defined as

$$\tilde{R}_{side}^{(m)} = \begin{cases} \sum_{i=m+1}^{M} \tilde{r}_i^m \tilde{R}_{side}^{(i)} + \tilde{A}_{side}^{(m)} & else \\ \tilde{A}_{side}^{(m)} & if\,(m = M) \end{cases}, \tag{18}$$

where *M* is the total number of the side-outputs. The activations $\tilde{A}_{side}^{(m)}$ at the *m*th side-output after the processing of the high-level connection is given by

$$\tilde{A}_{side}^{(m)} = \begin{cases} \sum_{i=m+1}^{M} \hat{r}_i^m \tilde{A}_{side}^{(i)} + \hat{A}_{side}^{(m)} & else \\ \hat{A}_{side}^{(m)} & if\,(m = M) \end{cases}, \tag{19}$$

where $\hat{A}_{side}^{(m)}$ is the activation at the last convolutional layer of each block of the VGG network as well as the last pooling layer. $\tilde{r}_i^m$ and $\hat{r}_i^m$ are the weights of short connections from the *i*th to the *m*th side layer for the high-level and low-level short connections, respectively. The weight of the short connections can be set to 0, if we want to drop out the corresponding connections. To further extract relatively high-level information before transmitting it to the shallower layers, we add a convolutional layer at the left side of each side network. Each output of the low-level short connections is followed by a convolutional layer noted with a dashed bounding box in Fig. 1. The dashed bounding box means that the convolutional layer only exists when there are the corresponding short connections. The same bilinear interpolation operation as in the DSS network for up-sampling is used here. Suppose the weights parameters of the classifier $\tilde{\mathbf{w}}$ at each side-output is denoted by $\tilde{\mathbf{w}} = (\tilde{\mathbf{w}}^{(1)}, \tilde{\mathbf{w}}^{(2)}, ..., \tilde{\mathbf{w}}^{(M)})$. The loss of the side-outputs of

our NeutSS-PLP network is

$$\tilde{L}_{side}(\mathbf{W},\tilde{\mathbf{w}},\tilde{\mathbf{r}},\hat{\mathbf{r}}) = \sum_{m=1}^{M} \alpha_m \tilde{l}_{side}^{(m)}(\mathbf{W},\tilde{\mathbf{w}}^{(m)},\tilde{\mathbf{r}},\hat{\mathbf{r}}), \tag{20}$$

where $\tilde{\mathbf{r}} = \{\tilde{r}_i^m\}$, $\hat{\mathbf{r}} = \{\hat{r}_i^m\}$, $i > m$, which are the control vectors for constructing the two-level short connections, α_m is the weight of the *m*th side-output loss,. Considering the imbalance problem for the polyp detection issue, the loss of each side-output is computed by employing the pixel-level class-balanced cross-entropy loss function. The loss function of the *m*th side-output is given by

$$l_{side}^{(m)}(\mathbf{W},\tilde{\mathbf{w}}^{(m)},\tilde{\mathbf{r}},\hat{\mathbf{r}}) = -\sum_{z_j \in \mathbf{Z}} z_j \frac{|\mathbf{Z}_-|}{|\mathbf{Z}|} \log \Pr(z_j = 1 \mid X;\mathbf{W},\tilde{\mathbf{w}}^{(m)},\tilde{\mathbf{r}},\hat{\mathbf{r}}) + (1-z_j)\frac{|\mathbf{Z}_+|}{|\mathbf{Z}|} \log \Pr(z_j = 0 \mid X;\mathbf{W},\tilde{\mathbf{w}}^{(m)},\tilde{\mathbf{r}},\hat{\mathbf{r}}), \tag{21}$$

where $\mathbf{Z}_+$ and $\mathbf{Z}_-$ denote the polyp and non-polyp ground truth label sets, respectively, |.| function calculates the number of labels in a specific set, other symbols are with similar definitions in Eq. (15).

The fusion loss function of our network can be represented by

$$\tilde{L}_{fuse}(\mathbf{W},\tilde{\mathbf{w}},\mathbf{f},\tilde{\mathbf{r}},\hat{\mathbf{r}}) = \sigma(Z, \sum_{m=1}^{M} f_m \tilde{R}_{side}^{(m)}). \tag{22}$$

By combining the side-outputs loss and the fusion loss, we can get the final loss:

$$\tilde{L}_{final}(\mathbf{W},\tilde{\mathbf{w}},\mathbf{f},\tilde{\mathbf{r}},\hat{\mathbf{r}}) = \tilde{L}_{side}(\mathbf{W},\tilde{\mathbf{w}},\tilde{\mathbf{r}},\hat{\mathbf{r}}) + \tilde{L}_{fuse}(\mathbf{W},\tilde{\mathbf{w}},\mathbf{f},\tilde{\mathbf{r}},\hat{\mathbf{r}}). \tag{23}$$

IV. Experimental Results And Discussions

In this section, we first introduce the test dataset and parameter setting. Next, we demonstrate the results of the specular reflections suppression. A comprehensive study on the colorectal polyp extracting network is further conducted, and the detailed results of the polyp region extraction are illustrated. Our methods are compared with some of the *state-of-the-art* saliency detection schemes.

A. *Dataset and Settings*

1) **Dataset**

Two datasets are employed. We use the EndoScene [57] and the Kvasir-SEG[58] dataset for evaluating the effectiveness of our approach. The EndoScene dataset is combined with the CVC-ColonDB and CVC-ClinicDB datasets. The CVC-ColonDB dataset is built up by extracting 300 images from 13 polyp video sequences acquired from 13 patients, and the corresponding polyp masks are given. The images in the CVC-ColonDB dataset are with the same resolution of 500 X 573. The CVC-ClinicDB dataset includes a total of 612 images with associated polyp masks obtained from 31 polyp video sequences of 23 patients. The image resolution is 384 X 288 for the CVC-ClinicDB dataset. The Kvasir-SEG dataset consists of 1000 gastrointestinal polyp images with various resolutions from 332 X 487 to 1920 X 1072. All the images are downsampled to 256 X 256 in this work. The

associated specular masks are also included in the EndoScene dataset. We split the EndoScene dataset into training, validation, and testing sets according to [33, 57], which contain 547, 183, and 182 images, respectively. For the Kvasir-SEG dataset, we randomly select 800 images for training, and the left 200 images are utilized as the testing set. To reduce the risk of overfitting in training the model, we use vertical flip, horizontal flip, rescaling, random rotation, and color jittering for data augmentation.

2) **Experimental settings**

Based on Eq. (18) and Eq. (19), short connections can be selectively dropped by setting the parameters $\tilde{r}_i^m$, $\hat{r}_i^m$ to be 0. In this paper, the network pattern of the NeutSS-PLP is given as follows:

$$\tilde{R}_{side}^{(m)} = \begin{cases} \sum_{i=3}^{6} \tilde{r}_i^m \tilde{R}_{side}^{(i)} + \tilde{A}_{side}^{(m)} & if\ (m = 1,2) \\ \tilde{r}_5^m \tilde{R}_{side}^{(5)} + \tilde{r}_6^m \tilde{R}_{side}^{(6)} + \tilde{A}_{side}^{(m)} & if\ (m = 3,4) \\ \tilde{A}_{side}^{(m)} & if\ (m = 5,6) \end{cases}, \tag{24}$$

$$\tilde{A}_{side}^{(m)} = \begin{cases} \hat{A}_{side}^{(m)} & if\ (m = 1,2,5,6) \\ \hat{r}_5^m \tilde{A}_{side}^{(5)} + \hat{r}_6^m \tilde{A}_{side}^{(6)} + \hat{A}_{side}^{(m)} & if\ (m = 3,4) \end{cases}, \tag{25}$$

where the parameters $\tilde{r}_i^m$, $\hat{r}_i^m$ in Eq. (24) and Eq. (25) are all set to 1. We also construct another pattern of short connections of the NeutSS-PLP for ablation study. As in Table 2, we call such a network NeutSS-PLP-P1 for short, which has the same high-level short connections as the NeutSS-PLP. The low-level short connections in NeutSS-PLP-P1 are defined as

$$\tilde{A}_{side}^{(m)} = \begin{cases} \hat{A}_{side}^{(m)} & if\ (m = 1,2,3,5,6) \\ \hat{r}_5^m \tilde{A}_{side}^{(5)} + \hat{r}_6^m \tilde{A}_{side}^{(6)} + \hat{A}_{side}^{(m)} & if\ (m = 4) \end{cases}. \tag{26}$$

For the specular reflections suppression, each colonoscopy image is normalized between 0 and 1. The thresholding parameter t_0 is set to be 0.65 in order to filter the regions that do not consist of specular pixels, and t_1 is set to be 0.15 to filter the non-intestinal wall area. The weighting parameters w_l and w_g for calculating the neutrosophic similarity score are chosen to be 0.5, and the final threshold for extracting the specular reflection region is set to be 0.4.

At the training stage, the Adam optimizer is used as the optimized method, and a mini-batch size of six is applied for model training. The initial learning rate is set to be 5e-5, and it is decreased to 5e-6 at epoch 50. All the models are implemented using the PyTorch package and trained with 100 epochs. Specifically, the CUDA version is 10.2, the cuDNN version is 10.1.243, the PyTorch version is 1.3.1, and the python version is 3.7.4.

3) **Evaluation metrics**

We use nine evaluation metrics to evaluate the performance of our method: F1 score, F2 score, Precision, Sensitivity, specificity, overall accuracy, Intersection-over-Union for Polyp region (IoU_P), IoU for Background region (IoU_B), and Mean IoU (mIoU).

Table 2
Ablation study of the proposed method in EndoScene dataset.

Model	F1	F2	Precision	Sensitivity	Specificity	Accuracy	mIoU	IoU_B	IoU_P
DSS[5]	0.862	0.840	0.903	0.825	0.992	0.964	0.853	0.962	0.744
DSS-LC	0.867	0.847	0.902	0.834	0.992	0.966	0.856	0.963	0.750
DSS-LF-LC	0.871	0.855	0.900	0.844	0.993	0.969	0.865	0.966	0.763
DSS-Neut	0.869	0.843	0.917	0.826	0.994	0.966	0.861	0.963	0.758
DSS-LC-Neut	0.867	0.848	0.900	0.837	0.993	0.967	0.861	0.965	0.758
DSS-LF-LC-Neut	0.875	0.858	0.907	0.846	0.993	0.968	0.867	0.966	0.769
NeutSS-PLP-P1	0.876	0.859	0.906	0.848	0.993	0.968	0.866	0.966	0.767
NeutSS-PLP	**0.886**	**0.872**	**0.919**	**0.864**	**0.994**	**0.971**	**0.877**	**0.968**	**0.787**

Abbreviations - LC: light channel; Neut: training with specular reflections suppressing images via SVNS; LF: layer fusion.

B. Ablation Study

In this section, several ablation studies are given to evaluate the key components of the proposed approach. The experimental results are summarized in Table 2.

1) Baseline: The DSS network with the VGG16 backbone is adopted as the baseline. The recommended short connection pattern is applied. The smoothing method of the fully connected Conditional Random Field (CRF) is not applied here, because we discover that the CRF strategy cannot improve the performance in dealing with the polyp region extraction issue.

2) Side network: Each side network contains two convolutional layers in both the NeutSS-PLP and DSS networks. The only difference is that each side network is the half size of the channels of the DSS network. For convenience, we call the half-size side-network Light Channel (LC). As in Table 2, DSS-LC performs better on all evaluation metrics except for Precision when compared with the DSS. This result demonstrates that the side network with the light channel is appropriate for extracting the polyp-like features.

3) Layer fusion: Two-level short connections are introduced into the NeutSS-PLP network. The method with Layer Fusion (LF) settings means that the low-level short connections are included, and the corresponding pattern of the short connections follows the NeutSS-PLP network. The LF is added, and the performance of DSS-LF-LC is shown in Row 3 in Table 2. A slight decrease occurs for the Precision metric, and the other evaluation metrics increase. Especially, Sensitivity increases by 1%, and IoU_P increases by 1.3%, compared to the results of the DSS-LC, which prove the effectiveness of the two-level short connections. As can be seen in rows 7 and 8, the NeutSS-PLP network outperforms the NeutSS-PLP-P1 network on all the evaluation metrics. A gain of 2% is achieved when more low-level short connections are introduced into the network. This result demonstrates that the network can benefit from the information interactions.

4) Neutrosophic specular reflections suppression: In Fig. 1, besides the source colonoscopy images, we use the images after specular reflection suppression as the input of the polyp region extraction network. To investigate the effect of introducing the suppressed images into the training stage, we add this scheme to the DSS, DSS-LC, and DSS-LF-LC networks. Table 2 shows that almost all the evaluation metrics increase when the suppressed images are applied for training, especially, the IoU_P increases by 1.9% for the DSS network. These results indicate that the suppressed images can provide another dimension of input information,

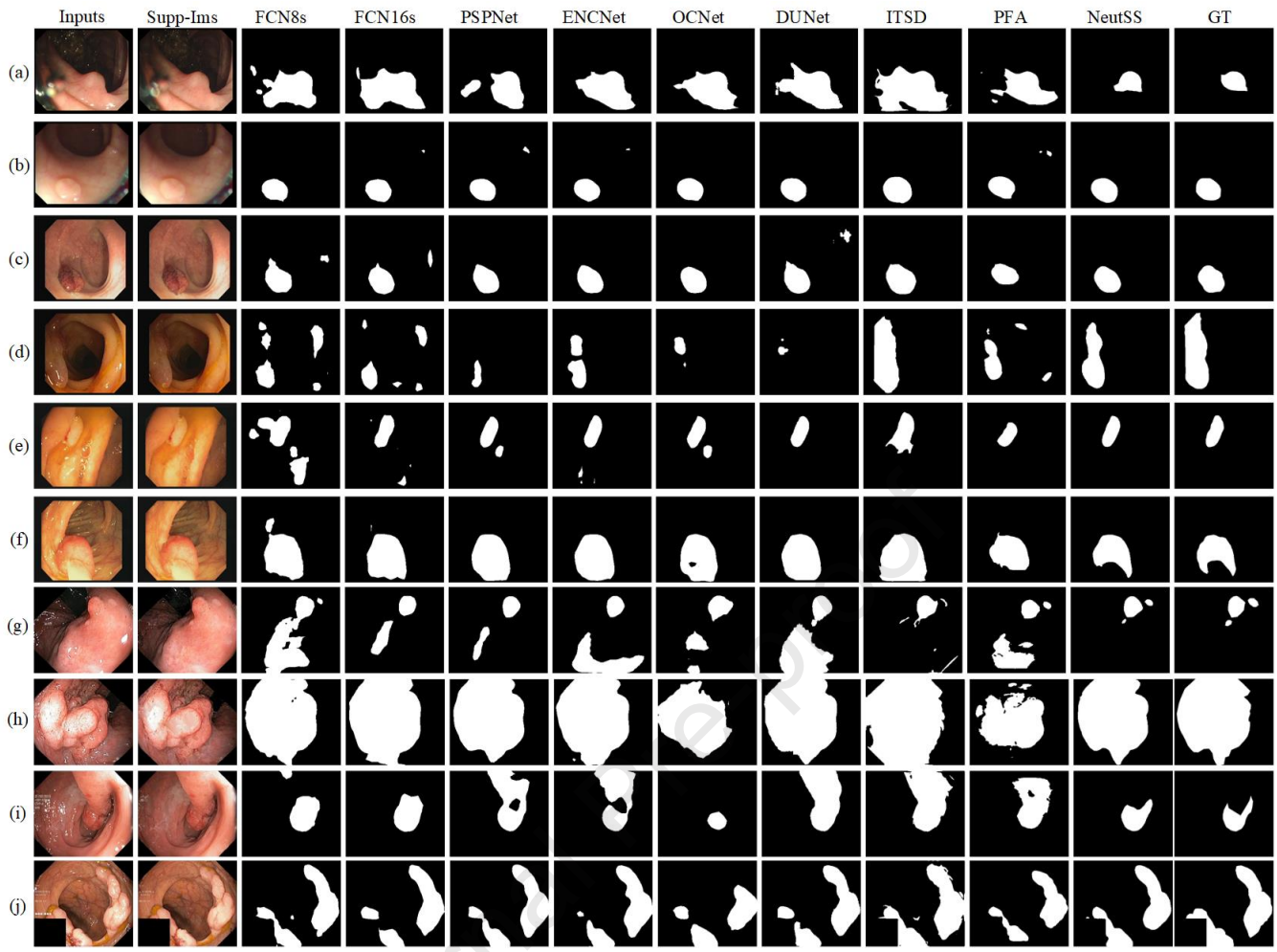


Fig. 5. Qualitative comparison of polyp region extraction results by different methods for the EndoScene dataset. From left to right are the test images, specular reflections suppressing images via SVNS, results of segmentation models FCN8S[3], FCN16S[3], PSPNet[8], ENCNet[9], OCNet[11], DUNet[12], results of saliency detection models ITSD[4], PFA[7], results of the NeutSS-PLP network, and ground truth. Examples (a-b) are from the CVC-ColonDB dataset, (c-f) are from the CVC-ClinicDB dataset, (g-i) are from the Kvasir-SEG dataset.

since the flattened specular regions on the image plane are restored. The network is capable of learning more robust features with little disturbing of the specular problem.

5) Loss function: For the class-imbalance problem, the weighted loss function Eq. (21) is applied for the NeutSS-PLP and the NeutSS-PLP-P1 network, and the other networks in Table 2 employ the traditional pixel-level cross-entropy loss function. The loss function is the only difference between the DSS-LF-LC-Neut and NeutSS-PLP networks. All the evaluation metrics increase with the weighted loss function. In particular, IoU_P, mIoU, Sensitivity, Precision, F2, and F1 increase by 1.8%, 1.1%, 1.8%, 1.2%, 1.4%, and 1.1% respectively. In addition, comparing the results of the DSS-LF-LC-Neut network and the NeutSS-PLP-P1 network, we find out that the performance difference is rather small, even the NeutSS-PLP-P1 network employs simpler low-level short connections.

In summary, compared to the traditional DSS network, the NeutSS-PLP gains 2.4%, 3.2%, 1.6%, 3.9%, 0.2%, 0.7%, 2.4%, 0.6%, and 4.3% for the F1 score, F2 score, Precision, Sensitivity, Specificity, Overall Accuracy, mIoU, IoU_B, and IoU_P, respectively.

Table 3
Comparison of polyp region extraction results between our proposed NeutSS-PLP and other saliency detection models in the EndoScene dataset.

Model	F1	F2	Precision	Sensitivity	Specificity	Accuracy	mIoU	IoU_B	IoU_P
ITSD[4]	0.881	**0.887**	0.872	**0.891**	0.991	**0.975**	0.816	**0.973**	0.659
PFA[7]	0.86	0.866	0.851	0.869	0.991	0.973	0.852	0.971	0.734
NeutSS-PLP	**0.886**	0.872	**0.919**	0.864	**0.994**	0.971	**0.877**	0.968	**0.787**

Table 4
Comparison of polyp region extraction results between our proposed NeutSS-PLP and other saliency detection models in the Kvasir-SEG dataset.

Model	F1	F2	Precision	Sensitivity	Specificity	Accuracy	mIoU	IoU_B	IoU_P
ITSD[4]	0.9176	0.9153	0.9216	0.9137	0.9859	0.9688	0.8479	0.9636	0.7323
PFA[7]	0.8957	0.902	0.8854	0.9063	0.9874	0.965	0.8806	0.9571	0.8042
NeutSS-PLP	**0.9286**	**0.9239**	**0.9366**	**0.9208**	**0.9903**	**0.9738**	**0.9135**	**0.964**	**0.863**

C. Comparison with Baseline Models

In this section, we compare the proposed method with two saliency detection models ITSD [4] and PFA [7], as well as six segmentation models FCN8S [3], FCN16S [3], PSPNet [8], ENCNet [9], OCNet [11], and DUNet [12]. The qualitative comparison results are shown in Fig. 5, and the quantitative comparison results are given in Tables 3-6.

Because our two-level short connections scheme concentrates on enhancing the interactions among the low-level and high-level features, we select the comparison algorithms by considering whether they make use of multi-scale or information interaction between different layers. The PFA network employs the attention scheme to focus on the effective context and spatial structural features, and achieves *state-of-the-art* performances on saliency detection. The ITSD network considers the correlation between saliency and contour in which a saliency branch and a contour branch are applied, and interactive feature transmission between brunches is utilized to enhance the ability of saliency detection. FCN is a conventional fully connected segmentation model, and the FCN8s and FCN16s networks are two basic versions of the FCN approach. Similar to the PFA network, PSPNet also uses the pyramid strategy. PSPNet utilizes the global information by applying the pyramid pooling module, and it can yield good performances on various datasets. Similarly, ENCNet deploys the context encoding module to capture the global contextual information for semantic segmentation. To capture even richer context information, OCNet combines the pyramid pooling scheme in PSPNet, the atrous spatial pyramid pooling, and the interlaced sparse self-attention strategy. DUNet is a deformable UNet, in which the deformable convolution and the upsampling operators are applied for adaptively extracting context information and combing the low-level and high-level features. DUNet recovers the pixel-wise predictions from the low-resolution images and outperforms a lot of conventional algorithms.

As shown in Fig. 5, Table 5, and Table 6, the proposed NeutSS-PLP network performs better than the six segmentation algorithms in all the evaluation metrics for both datasets. In Table 3, when compared with these two saliency detection networks, the ITSD network shows better performance than the NeutSS-PLP network in F2 score, Sensitivity, Accuracy, and IoU_B metric.

Table 5
Comparison of polyp region extraction results between our proposed NeutSS-PLP and other segmentation models in the EndoScene dataset.

Model	F1	F2	Precision	Sensitivity	Specificity	Accuracy	mIoU	IoU_B	IoU_P
FCN8s[3]	0.78	0.793	0.837	0.819	0.985	0.953	0.818	0.954	0.682
FCN16s[3]	0.789	0.807	0.829	0.831	0.986	0.954	0.826	0.956	0.694
PSPNet[8]	0.842	0.838	0.917	0.843	0.994	0.966	0.866	0.964	0.769
ENCNet[9]	0.842	0.838	0.912	0.845	0.993	0.965	0.865	0.963	0.766
OCNet[11]	0.834	0.829	0.919	0.833	0.994	0.965	0.862	0.963	0.761
DUNet[12]	0.842	0.843	0.906	0.853	0.992	0.963	0.868	0.963	0.772
NeutSS-PLP	**0.886**	**0.872**	**0.919**	**0.864**	**0.994**	**0.971**	**0.877**	**0.968**	**0.787**

Table 6
Comparison of polyp region extraction results between our proposed NeutSS-PLP and other segmentation models in the Kvasir-SEG dataset.

Model	F1	F2	Precision	Sensitivity	Specificity	Accuracy	mIoU	IoU_B	IoU_P
FCN8s[3]	0.8809	0.8838	0.8985	0.8925	0.9831	0.9636	0.8808	0.9526	0.8089
FCN16s[3]	0.8756	0.8866	0.8767	0.9015	0.9775	0.961	0.8735	0.949	0.798
PSPNet[8]	0.8919	0.8951	0.903	0.9021	0.9854	0.9702	0.8952	0.96	0.8303
ENCNet[9]	0.8812	0.8804	0.9101	0.8861	0.9825	0.9583	0.8853	0.9535	0.817
OCNet[11]	0.808	0.7862	0.8768	0.7827	0.984	0.9326	0.8144	0.9306	0.6983
DUNet[12]	0.8958	0.9027	0.9003	0.9137	0.9787	0.9564	0.8927	0.9557	0.8295
NeutSS-PLP	**0.924**	**0.9243**	**0.9235**	**0.9245**	**0.9887**	**0.9742**	**0.9137**	**0.9653**	**0.8621**

We hypothesize the ITSD network can work well for polyp region extraction, due to the fact that the contour information fusion and different layer features are considered. However, the NeutSS-PLP network achieves 12.8%, 6.1%, 4.7% gains than the ITSD network in IoU_P, mIoU, and Precision, respectively. For the issue of polyp region extraction, the mIoU metric and IoU_P metric are quite important. Our approach performs the best among all the comparison algorithms with regard to the following metrics: mIoU, IoU_P, F1 score, Precision, and Specificity. Although the ITSD network outperforms our approach by 2.7% in Sensitivity, it performs worse in the other metrics like Precision and IoU_P. This is mainly because the ITSD network always predicts a larger polyp region than the ground truth, as in Fig. 5, especially in rows (a), (e), (f), and (h-j). The PFA network brings the spatial attention scheme into the shallower layers and applies the channel attention strategy for the deeper layers. Both features in these two levels are fused to predict the salient region. This simple and effective architecture helps the PFA network achieve a slightly better performance in mIoU and IoU_P than the ITSD network. However, the PFA network may be disturbed by the polyp-like regions. The relevant results are given in rows (b), (d) and (g) of Fig. 5. Owe to the input of images before or after specular suppressing and the short connection scheme, the NeutSS-PLP network outperforms the PFA network in several key evaluation metrics. With the NeutSS-PLP network, the Precision, mIoU, and IoU_P increase by 6.8%, 2.5%, and 5.3%, respectively. When testing in the Kvasir-SEG dataset, benefit from more training data, the advantage of feature fusion capacity helps the NeutSS-PLP network outperform both saliency detection networks in each evaluation metrics.

Tables 5 and 6 provides the quantitative results in comparison with the segmentation baseline networks. The NeutSS-PLP network is the best one among all the segmentation algorithms in terms of all the evaluation metrics for both datasets. Particularly, the NeutSS-PLP network gains 3.18% more than the best performance of other segmentation algorithms in the IoU_P evaluation metric. Both the FCN8s and the FCN16s networks use the VGG16 network as their encoder backbone, and the coarse decoders are

utilized. For the FCN16s network, information from the final and stride-16 layers is applied for calculating the final prediction. For the FCN8s network, the information from the stride-8 layer is added. Although the FCN8s network utilizes more information and generates further Precision, FCN16s still outperforms it in most of the evaluation metrics for the EndoScene dataset. However, FCN8s outperforms FCN16s for the other dataset. Due to the coarse decoder and the lack of a strategy for efficiently passing the low-level information to the decoder, the performance of these two networks is worse than that of the other networks except OCNet in the Kvasir-SEG dataset. The ResNet-50 is applied as the backbone for PSPNet and ENCNet, and OCNet uses ResNet-100 as its backbone. Both PSPNet and ENCNet employ global information as well as a simple scheme for transmitting the information from the shallow layer to the deeper layer. As in Table 5 and Table 6, their performances are with quite a small gap. Two different strategies of pyramid pooling are used by OCNet to capture richer context information. OCNet has a more complicated architecture, and its performance is worse than that of the previous two networks with regard to the evaluation metrics of F1, F2, Sensitivity, mIoU, and IoU_P. In addition, OCNet performs worst in the Kvasir-SEG dataset. This can be ascribed to the complicated concatenations making the network difficult to train. Due to the use of deformation convolution and the U-shape architecture, DUNet is superior in F2, Sensitivity, mIoU, and IoU_P compared to PSPNet, ENCNet, and OCNet when testing in the EndoScene dataset. However, its Precision score is relatively low in both datasets.

In Fig. 5, the NeutSS-PLP network performs well in some challenging circumstances, such as polyp with confusing boundaries or bubbles. As shown in row (a), the boundary of this polyp is rather confusing, and this poses a challenge to all the algorithms, which fail to detect the right polyp region except for the proposed approach. Since the surroundings are very similar to the polyp, a small area of non-polyp area is wrongly recognized as a polyp region by our method. Those results in rows (f), and (g-i) also indicate the robustness of the NeutSS-PLP network. Three polyps exist on the image shown in row (g), and one of these polyps is quite small. All other networks fail to detect the smallest polyp except the NeutSS-PLP network. When the polyp region extraction models encounter bubbles or specular reflections, some of them predict wrong results, as in row (b), (e), (h), and (j). Our algorithm performs well due to the short-connections architecture and the suppressed images.

D. Limitations and Future Work

There are two main limitations of our work. First, it may miss some polyp regions or misidentify non-polyp regions, e.g., Fig. 5(a), (d), (f), and (j). In Fig. 5(d), it produces a relatively poorer result than the ITSD network. The polyp area near the boundary of the image is predicted as a non-polyp area. Introducing an appropriate scheme into the polyp detection network for achieving a more accurate region-boundary is essential to deal with this issue. Therefore, exploring the strategy of contour restriction into the model will be an interesting topic for the future work. Second, images are all downsampled to small sizes for our model in this work. Although the colonoscopes with relatively low resolutions are commonly used, using high resolution is a trend. The strategy of cropping the high-resolution images into small pieces and feeding them into the models is often employed. However, such a

strategy may lose global information when judging each piece of the image, especially for the regions near the boundary of these images. We will utilize effective strategies for polyp detection, e.g., small polyp detection in high-resolution colonoscopy images by integrating global and local contextual information.

V. CONCLUSIONS

In this paper, a new network integrating both the neutrosophic theory and saliency detection strategy is proposed for polyp region extraction in colonoscopy images. Two kinds of neutrosophic criteria are developed, and the corresponding *T*, *I*, and *F* functions are given. The neutrosophic images are used to detect the specular reflections, and the images after specular suppressing are fed into our polyp extraction technique. For the polyp extraction model, two-level short connections are introduced into the network for transmitting the features from the deep layers to shallow layers at different stages. Comprehensive experiments have verified the effectiveness of our scheme, which outperforms several *state-of-the-art* algorithms in both saliency detection and image segmentation.

VI. ACKNOWLEDGMENTS

This work was supported in part by the Natural Science Foundation of Zhejiang Province under Grant LY20F020011, the Social Sciences and Humanities Youth Foundation of Ministry of Education under Grant 21YJCZH039, the Natural Science Foundation of Zhejiang Province under Grant LY19F020015, TY22F025548, the National Natural Science Foundation of China under Grant 61603258, 61871289, 61802347, 62002227, and in part by the Key scientific research project of Shaoxing University under Grant 2020LG1004.

REFERENCES

[1] B. Hariharan, P. Arbelaez, R. Girshick, J. Malik, Hypercolumns for object segmentation and fine-grained localization, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[2] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, (2014).

[3] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431-3440.

[4] H. Zhou, X. Xie, J.-H. Lai, Z. Chen, L. Yang, Interactive two-stream decoder for accurate and fast saliency detection, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 9141-9150.

[5] Q. Hou, M. Cheng, X. Hu, A. Borji, Z. Tu, P.H. Torr, Deeply Supervised Salient Object Detection with Short Connections, IEEE Trans Pattern Anal Mach Intell, 4 (2019), pp. 815-828.

[6] S. Xie, Z. Tu, Holistically-nested edge detection, Int J Comput Vision, 125 (2017), pp. 3-18.

[7] T. Zhao, X. Wu, Pyramid Feature Attention Network for Saliency Detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 3085-3094.

[8] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2881-2890.

[9] H. Zhang, K. Dana, J. Shi, Z. Zhang, X. Wang, A. Tyagi, A. Agrawal, Context encoding for semantic segmentation, Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2018, pp. 7151-7160.

[10] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA: a cancer journal for clinicians, 71 (2021), pp. 209-249.

[11] Y. Yuan, L. Huang, J. Guo, C. Zhang, X. Chen, J. Wang, OCNet: Object Context for Semantic Segmentation, Int J Comput Vision, (2021), pp. 1-24.

[12] Q. Jin, Z. Meng, T.D. Pham, Q. Chen, L. Wei, R. Su, DUNet: A deformable network for retinal vessel segmentation, Knowl-based Syst, 178 (2019), pp. 149-162.

[13] R.L. Siegel, K.D. Miller, H.E. Fuchs, A. Jemal, Cancer statistics, 2021, CA: a cancer journal for clinicians, 71 (2021), pp. 7-33.

[14] D.A. Corley, C.D. Jensen, A.R. Marks, W.K. Zhao, J.K. Lee, C.A. Doubeni, A.G. Zauber, J. de Boer, B.H. Fireman, J.E. Schottinger, Adenoma detection rate and risk of colorectal cancer and death, New Engl J Med, 370 (2014), pp. 1298-1306.

[15] S.B. Ahn, D.S. Han, J.H. Bae, T.J. Byun, J.P. Kim, C.S. Eun, The miss rate for colorectal adenoma determined by quality-adjusted, back-to-back colonoscopies, Gut and Liver, 6 (2012), pp. 64.

[16] C.M. le Clercq, M.W. Bouwens, E.J. Rondagh, C.M. Bakker, E.T. Keulen, R.J. de Ridder, B. Winkens, A.A. Masclee, S. Sanduleanu, Postcolonoscopy colorectal cancers are preventable: a population-based study, Gut, 63 (2014), pp. 957-963.

[17] X. Jia, X. Mai, Y. Cui, Y. Yuan, X. Xing, H. Seo, L. Xing, M.Q.-H. Meng, Automatic polyp recognition in colonoscopy images using deep learning and two-stage pyramidal feature prediction, IEEE Transactions on Automation Science and Engineering, 17 (2020), pp. 1570-1584.

[18] Y. Yuan, D. Li, M.Q.H. Meng, Automatic polyp detection via a novel unified bottom-up and top-down saliency approach, IEEE Journal of Biomedical and Health Informatics, 22 (2018), pp. 1250-1260.

[19] M. Ganz, X. Yang, G. Slabaugh, Automatic segmentation of polyps in colonoscopic narrow-band imaging data, IEEE T Bio-med Eng, 59 (2012), pp. 2144-2151.

[20] J. Bernal, F.J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, F. Vilariño, WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians, Comput Med Imag Grap, 43 (2015), pp. 99-111.

[21] N. Tajbakhsh, S.R. Gurudu, J. Liang, Automated polyp detection in colonoscopy videos using shape and context information, IEEE T Med Imaging, 35 (2016), pp. 630-644.

[22] A. Sanchez Gonzalez, B. Garcia Zapirain, D. Sierra Sosa, A. Elmaghraby, Automatized colon polyp segmentation via contour region analysis, Comput Biol Med, 100 (2018), pp. 152-164.

[23] P.N. Figueiredo, I.N. Figueiredo, L. Pinto, S. Kumar, Y.-H.R. Tsai, A.V. Mamonov, Polyp detection with computer-aided diagnosis in white light colonoscopy: comparison of three different methods, Endoscopy international open, 7 (2019), pp. E209-E215.

[24] Q. Li, L. Li, W. Wang, Q. Li, J. Zhong, A comprehensive exploration of semantic relation extraction via pre-trained CNNs, Knowledge-Based Systems, 194 (2020), pp. 105488.

[25] S. Tang, F. Yu, Construction and verification of retinal vessel segmentation algorithm for color fundus image under BP neural network model, The Journal of Supercomputing, 77 (2021), pp. 3870-3884.

[26] S. Wang, Y. Cong, H. Zhu, X. Chen, L. Qu, H. Fan, Q. Zhang, M. Liu, Multi-scale context-guided deep network for automated lesion segmentation with endoscopy images of gastrointestinal tract, IEEE Journal of Biomedical and Health Informatics, 25 (2020), pp. 514-525.

[27] S. Zhang, W. Ren, X. Tan, Z.-J. Wang, Y. Liu, J. Zhang, X. Zhang, X. Cao, Semantic-aware dehazing network with adaptive feature fusion, IEEE Transactions on Cybernetics, (2021).

[28] O.F. Ahmad, A.S. Soares, E. Mazomenos, P. Brandao, R. Vega, E. Seward, D. Stoyanov, M. Chand, L.B. Lovat, Artificial intelligence and computer-aided diagnosis in colonoscopy: current evidence and future directions, The Lancet Gastroenterology & Hepatology, 4 (2019), pp. 71-80.

[29] P. Brandao, O. Zisimopoulos, E. Mazomenos, G. Ciuti, J. Bernal, M. Visentini-Scarzanella, A. Menciassi, P. Dario, A. Koulaouzidis, A. Arezzo, Towards a computed-aided diagnosis system in colonoscopy: automatic polyp segmentation using convolution neural networks, Journal of Medical Robotics Research, 3 (2018), pp. 1840002.

[30] P. Wang, X. Xiao, J.R.G. Brown, T.M. Berzin, M. Tu, F. Xiong, X. Hu, P. Liu, Y. Song, D. Zhang, Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy, Nature Biomedical Engineering, 2 (2018), pp. 741-748.

[31] D. Banik, K. Roy, D. Bhattacharjee, M. Nasipuri, O. Krejcar, Polyp-Net: A multimodel fusion network for polyp segmentation, IEEE T Instrum Meas, 70 (2020), pp. 1-12.

[32] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, L. Shao, Pranet: Parallel reverse attention network for polyp segmentation, International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 263-273.

[33] Y. Fang, D. Zhu, J. Yao, Y. Yuan, K.-y. Tong, ABC-Net: Area-Boundary Constraint Network with Dynamical Feature Selection for Colorectal Polyp Segmentation, IEEE Sensors Journal, (2020).

[34] D. Jha, P.H. Smedsrud, D. Johansen, T. de Lange, H.D. Johansen, P. Halvorsen, M.A. Riegler, A comprehensive study on colorectal polyp segmentation with ResUNet++, conditional random field and test-time augmentation, IEEE journal of biomedical and health informatics, 25 (2021), pp. 2029-2040.

[35] P. Sasmal, M. Bhuyan, S. Gupta, Y. Iwahori, Detection of Polyps in Colonoscopic Videos Using Saliency Map-Based Modified Particle Filter, IEEE T Instrum Meas, 70 (2021), pp. 1-9.

[36] F. Deeba, F.M. Bui, K.A. Wahid, Computer-aided polyp detection based on image enhancement and saliency-based selection, Biomedical Signal Processing and Control, 55 (2020), pp. 101530.

[37] F. Smarandache, Neutrosophy: neutrosophic probability, set, and logic: analytic synthesis & synthetic analysis, American Research Press1998.

[38] X. Peng, J. Dai, A bibliometric analysis of neutrosophic set: two decades review from 1998 to 2017, Artif Intell Rev, (2018), pp. 1-57.

[39] Y. Yuan, W. Qin, B. Ibragimov, B. Han, L. Xing, RIIS-DenseNet: Rotation-Invariant and Image Similarity Constrained Densely Connected Convolutional Network for Polyp Detection, International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 620-628.

[40] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, Proceedings of the European conference on computer vision (ECCV), 2018, pp. 801-818.

[41] I. Ullah, M. Jian, S. Hussain, J. Guo, H. Yu, X. Wang, Y. Yin, A brief survey of visual saliency detection, Multimed Tools Appl, 79 (2020), pp. 34605-34645.

[42] J. Ye, Multicriteria decision-making method using the correlation coefficient under single-valued neutrosophic environment, International Journal of General Systems, 42 (2013), pp. 386-394.

[43] S. Song, Z. Jia, J. Yang, N.K. Kasabov, A Fast Image Segmentation Algorithm Based on Saliency Map and Neutrosophic Set Theory, IEEE Photonics Journal, 12 (2020), pp. 1-16.

[44] S. Dhar, M.K. Kundu, Accurate multi-class image segmentation using weak continuity constraints and neutrosophic set, Applied Soft Computing, 112 (2021), pp. 107759.

[45] K. Hu, J. Ye, E. Fan, S. Shen, L. Huang, J. Pi, A novel object tracking algorithm by fusing color and depth information based on single valued neutrosophic cross-entropy, Journal of Intelligent and Fuzzy Systems, 32 (2017), pp. 1775-1786.

[46] K. Hu, W. He, J. Ye, L. Zhao, H. Peng, J. Pi, Online Visual Tracking of Weighted Multiple Instance Learning via Neutrosophic Similarity-Based Objectness Estimation, Symmetry, 11 (2019), pp. 832-855.

[47] A. Rashno, D.D. Koozekanani, P.M. Drayna, B. Nazari, S. Sadri, H. Rabbani, K.K. Parhi, Fully Automated Segmentation of Fluid/Cyst Regions in Optical Coherence Tomography Images With Diabetic Macular Edema Using Neutrosophic Sets and Graph Algorithms, IEEE T Bio-med Eng, 65 (2018), pp. 989-1001.

[48] N.E.M. Khalifa, F. Smarandache, G. Manogaran, M. Loey, A study of the neutrosophic set significance on deep transfer learning models: An experimental case on a limited covid-19 chest x-ray dataset, Cognitive Computation, (2021), pp. 1-10.

[49] W.-H. Cui, J. Ye, J. Fu, Cotangent similarity measure of single-valued neutrosophic interval sets with confidence level for risk-grade evaluation of prostate cancer, Soft Computing, 24 (2020), pp. 18521-18530.

[50] M. Sharma, I. Kandasamy, W. Vasantha, Comparison of neutrosophic approach to various deep learning models for sentiment analysis, Knowl-based Syst, 223 (2021), pp. 107058.

[51] S. Bhat, D. Koundal, Multi-focus Image Fusion using Neutrosophic based Wavelet Transform, Applied Soft Computing, 106 (2021), pp. 107307.

[52] C. Fan, K. Hu, S. Feng, J. Ye, E. Fan, Heronian mean operators of linguistic neutrosophic multisets and their multiple attribute decision-making methods, International Journal of Distributed Sensor Networks, 15 (2019), pp. 1550147719843059.

[53] W.-H. Cui, J. Ye, Logarithmic similarity measure of dynamic neutrosophic cubic sets and its application in medical diagnosis, Computers in Industry, 111 (2019), pp. 198-206.

[54] J. Ye, Vector similarity measures of simplified neutrosophic sets and their application in multicriteria decision making, International Journal of Fuzzy Systems, 16 (2014), pp. 204-211.

[55] S.M. Alsaleh, A.I. Aviles, P. Sobrevilla, A. Casals, J.K. Hahn, Automatic and robust single-camera specular highlight removal in cardiac images, Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, IEEE, 2015, pp. 675-678.

[56] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234-241.

[57] D. Vázquez, J. Bernal, F.J. Sánchez, G. Fernández-Esparrach, A.M. López, A. Romero, M. Drozdzal, A. Courville, A benchmark for endoluminal scene segmentation of colonoscopy images, Journal of healthcare engineering, 2017 (2017).

[58] D. Jha, P.H. Smedsrud, M.A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, H.D. Johansen, Kvasir-seg: A segmented polyp dataset, International Conference on Multimedia Modeling, Springer, 2020, pp. 451-462.

Keli Hu received the Ph.D. degree in communication and information systems from the Shanghai Institute of Microsystem and Information Technology, University of Chinese Academic and Sciences, Shanghai, China, in 2014. In July 2014, he joined Shaoxing University, Shaoxing, China, where he is currently an Associate Professor with the Department of Computer Science and Engineering. He was also a visiting scholar in Faculty of Engineering, National University of Singapore from September 2018 to March 2019. His research interests include computer vision, machine learning, object tracking, image processing, and neutrosophic theory.

Liping Zhao received the B.S. degree in computer science and technology from Hengyang Normal University in 2006, the M.S. degree in computer science and technology from Hunan University in 2009, and the Ph.D. degree from Tongji University, Shanghai, China, in 2017. She was with the College of Mathematics, Physics, and Information Engineering, Jiaxing University, from July 2009 to June 2017. She is currently with the Department of Computer Science and Engineering, Shaoxing University, Shaoxing, China. Her research interests include computer vision and video coding.

Sheng Feng received the B.S. degree in computing from Greenwich University in 2004, the M.S. degree in software engineering in 2013, and the Ph.D. degree in pattern recognition and intelligent system from Northeastern University, Shenyang, China, in 2017. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Shaoxing University, Shaoxing, China. His research interests include intelligent robot, computer vision, and wireless sensor networks.

Shengdong Zhang received the B.S. degree in computer science and technology from Shanxi University, Taiyuan, China, in 2009, the M.S. degree in mechanics from Peking University, Beijing, China, in 2012, and the Ph.D. degree from Wuhan University, Wuhan, China, in 2020. He is currently an Research Scientist with the Department of Computer Science and Engineering, Shaoxing University, Shaoxing, China. His research interests include deep leaning, computer vision, and image processing.

Qianwei Zhou received the Ph.D. degree in communication and information systems from the Shanghai Institute of Microsystem and Information Technology, University of Chinese Academic and Sciences, Shanghai, China in 2014. In July 2014, he joined Zhejiang University of Technology, Hangzhou, China, where he is currently a Research Scientist in the College of Computer Science. He was also a visiting scholar in Imaging Research Division, Department of Radiology, University of Pittsburgh, USA from August 2018 to August 2019. His research interests include the crossing field of machine learning and computer-aided design, Internet-of-Thing-related signal processing, pattern recognition and medical image understanding.

Xiaozhi Gao received his B.Sc. and M.Sc. degrees from the Harbin Institute of Technology, China in 1993 and 1996, respectively. He obtained his D.Sc. (Tech.) degree from the Helsinki University of Technology (now Aalto University), Finland in 1999. He was appointed as a Docent at the same university in 2004. He has been working as a professor at the University of Eastern Finland, Finland since 2018. His current research interests are nature-inspired computing methods with their applications in optimization, data mining, machine learning, control, signal processing, and industrial electronics.

Yanhui Guo received his B. S. degree in Automatic Control from Zhengzhou University, China, M.S. degree in Pattern Recognition and Intelligence System from Harbin Institute of Technology, China, and a Ph.D. degree in the Department of Computer Science, Utah State University, USA. He was a research fellow in the Department of Radiology at the University of Michigan and an assistant professor at St. Thomas University. Dr. Guo is currently an associated professor in the Department of Computer Science at the University of Illinois Springfield. His research area includes computer vision, machine learning, big data analytics, and computer-aided detection/diagnosis.

In this paper, we present an effective polyp region extraction strategy by employing the scheme of saliency detection and the theory of neutrosophic. The key contributions of our work are given as follows:

- We define the local and global threshold criteria as well as the corresponding T, I, and F functions to transfer the specular detection problem to the SVNS-based multi-criteria decision-making problem.
- We apply the CNN-based saliency detection network to deal with the colorectal polyp region extraction problem.
- We propose a saliency detection-based polyp extraction network with two-level short connections to take advantage of the multi-level and multi-scale features extracted from different stages of the network.

All other authors have no competing interests to disclose.