

Received March 24, 2020, accepted April 22, 2020, date of publication April 28, 2020, date of current version May 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2991115

Deep Model-Based Semi-Supervised Learning Way for Outlier Detection in Wireless Capsule Endoscopy Images

YAN GAO¹, WEINING LU^{1,2}, XIAOBEI SI¹, AND YU LAN¹

¹Department of Gastroenterology, Beijing Jishuitan Hospital, Beijing 100035, China

²Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China

Corresponding author: Yu Lan (lany_2020@163.com)

This work was supported by the Beijing JST Research Funding under Grant YGQ-201911, which is provided by Beijing Jishuitan (JST) Hospital, China.

ABSTRACT Wireless capsule endoscopy (WCE) has become an irreplaceable tool for diagnosing small intestinal diseases, and detecting the outliers in WCE images automatically remains as a hot research topic. Considering the difficulties in obtaining sufficient labeled WCE data, it is necessary to develop the diagnosis model which works well with only little labeled or even unlabeled training samples. In this paper, a novel semi-supervised deep-structured framework is introduced to solve the problem of outlier detection in WCE images. The key idea of our model is to mine the anomalous graphical patterns existed in the image by analyzing the spatial-scale trends of sequential image regions. Three main contributions are concluded: 1) we integrate a convolutional neural network into long short term memory network, so that the intrinsic differences between outliers and normal instances could be captured. Besides, 2) a assessment model is built by using various signs of anomaly occurrence and fake outliers knowledge learned during the training stage, which enhances the outlier alarm accuracy significantly. Furthermore, 3) a nest-structured training method is proposed, which helps our model achieving efficient training process. Experimental results on the real WCE images demonstrate the effectiveness of our model.

INDEX TERMS Convolutional neural network, long short term memory network, outlier detection, semi-supervised, wireless capsule endoscopy.


I. INTRODUCTION

Small intestinal disease is one of the most common gastrointestinal disorders seen in clinical practice, which including cancer, polyp, infectious inflammation and the like. If not diagnosed promptly in the early stage, these diseases are likely to develop into poor long-term prognosis or even death. Therefore, early detection of small intestinal disease becomes more crucial. However due to the special position and impressive length of small bowel, it is challenging to utilize wired endoscopes through mouth or anus to the lesion area directly.

Owing to the advent of wireless capsule endoscopy (WCE), patients avoid suffering a uncomfortable and lengthy procedure by swallowing this sensor simply. A typical WCE mainly consists of image sensor, lens, LED and wireless

transmitter module. While moving along the intestine, it can capture images of small bowel periodically and send them to an outside storage device. Afterwards, physicians can make diagnosis conclusion by reviewing the whole WCE video. For now, WCE has become an irreplaceable tool for diagnosing small intestinal diseases. However, the application of WCE suffers from one obvious shortcoming. About fifty to sixty thousands images are generated for each examination, whereas images containing lesion areas occupy only less than ten percent. Therefore it is a time-consuming and tedious task for physicians to check through all the WCE images frame by frame. Precisely because of this situation, lots of findings have been proposed to build the computer-aided diagnosis systems for analyzing the WCE images automatically.

The majority of previously published works focus on solving detection problems for two important symbols of various small intestinal diseases: bleeding and polyp. Traditional

The associate editor coordinating the review of this manuscript and approving it for publication was Carmelo Militello .

machine learning methods are the common choices to detect these abnormalities [1]–[9]. For bleeding detection problem, lots of studies take color [1], texture [2] and image statistical information [4] as the characteristic representation of WCE data, and these manual features were used to construct classifiers with various machine learning methods, such as support vector machine (SVM), k-nearest neighbors. [5] conducted an empirical evaluation of four feature descriptors for bleeding recognition on WCE video. Authors in [6] calculated local binary pattern from HIS color space of image as features, and then chose SVM as a classifier. Reference [7] proposed a polyp detection model based on adaptive neuro fuzzy algorithm. Similar to methods for bleeding detection, Color, texture and shape features were also used to discriminate the polyp regions from normal part [8], [9].

Recently, deep learning technologies represented by convolutional neural network (CNN) show outstanding performance in various image processing tasks [10]–[13]. Many researchers have applied deep learning methods to build intelligent diagnosis systems for WCE video and gotten many exciting results [14]–[18]. In [15], convolution neural network is utilized to extract segmentation information of bleeding regions, and extracted features are fed into a SVM to realize the binary classification (bleeding or non-bleeding). Yuan, et.al. proposed a stacked sparse auto-encoder with image manifold constraint for recognizing polyps in the WCE images [16]. In [18], researchers took advantage of pre-trained deep architectures to improve the detection accuracy.

In spite of these advances of autonomous diagnosis systems, one fact can not be ignored is that most of achievements work under a general precondition: there are plenty of labeled samples to construct the discrimination model. In truth, however, obtaining sufficient labeled WCE data is still subject to the following difficulties: 1) Unlike bleeding, most of small intestinal diseases such as polyp, ulcer, vascular malformation have a great diversity of spatial and color nature, which vary in texture, size, and surroundings. It is almost impossible to collect labeled training samples covering all abnormal forms. 2) For rare diseases in small intestine, the number of cases can not meet the need of data volume for constructing a classifier.

Therefore, it is necessary to develop the disease diagnosis model which works well with only little labeled or even unlabeled training samples. Based on the above analysis, we can regard small intestinal diseases detection task as a outlier detection problem. Outlier detection denotes the task of detecting patterns that do not accord with the expected normal state in data. Those non-conforming patterns are outliers or anomalies [19]. Generally, only normal data are available for training stage in outlier detection problem. In this work, normal data means WCE images without any diseases. Polyp, ulcer and the like contained in WCE images are outliers needed to be recognized.

Theoretically, graphical features of arbitrary region in normal WCE image strongly depend on those of adjacent areas. It is because that these areas have similar texture and color.

Contrary to this correlation, outlier parts in WCE image show obvious differences from surrounding areas. Based on this premise, we propose a novel model called Semi-supervised **Outlier Detection Model**, **SODM** for short. The key idea of model is to mine the anomalous graphical patterns existed in the image by analyzing the spatial-scale trends of sequential image regions. It is a semi-supervised learning process since only data of normal category and limited expert knowledge are adopted during the training stage. Our model works well by integrating the following three key ingredients.

- 1) For any selected block in the WCE image, we take its surrounding regions as corresponding previous context, so that we can predict its expected state. The deviation between real state and inferred one will serve as the indicator of small intestinal outliers.
- 2) A deviation generator between real block and expected one is designed by combining CNN and long short memory network (LSTM), called CNN-LSTM for short. CNN with a designed architecture is utilized to learn the intrinsic features from blocks. The state estimator is constructed with the LSTM network, whose inputs are the features drawn from CNN. Afterwards, the deviation can be calculated with the output of state estimator.
- 3) A outlier assessment model (OAM) with two different indexes is built to confirm the anomaly occurrence while reaching a pre-defined confidence level.

Accordingly, the main contributions of our literature are summarized as below.

- 1) Block selection method of this work ensures that any area in the WCE image can be estimated whether the outlier is present. It is a more accurate outlier detection method than taking the WCE image as a whole.
- 2) The proposed LSTM-based state estimator is conditioned on the intrinsic CNN-based features from normal WCE data. Therefore, a proper state estimator for normal patterns could be obtained, which makes our model output a distinct deviation value while outlier occurs.
- 3) We propose a nest-structured model training method, in order that the loss of LSTM-based state estimator could be used to fine-tuning CNN parameters. By doing this, CNN could learn a better pattern to represent the correlation hidden among the sequential blocks.
- 4) The outlier assessment model enhances the outlier alarm accuracy by considering two critical factors, one is excluding the fake outliers by the knowledge learned in the training stage, the other is making a overall evaluation for deviations generated by surrounding blocks in all directions.
- 5) Promising results on real WCE images demonstrate the effectiveness of this work.

The rest of this paper is organized as follows. Section II illustrates the preliminaries; Section III states the proposed

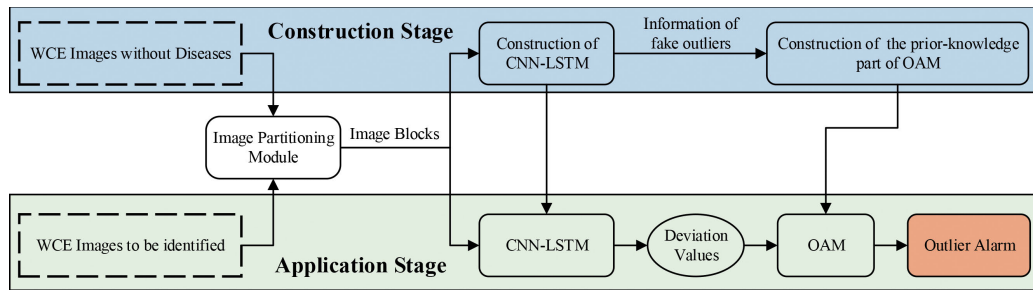


FIGURE 1. Usage of the proposed SODM.

methods, and several experiments are conducted in Section IV. Finally, Section V makes a conclusion.

II. PRELIMINARIES

A. CONVOLUTIONAL NEURAL NETWORK

CNN is one of the most famous technologies of deep neural network, which has obtained many remarkable achievements, especially in the field of image processing. CNN is a kind of neural network with deep architectures, which consists of hierarchically well-trained layers to learn the features from basic to advanced [20].

A typical CNN architecture consists of three types layers, that is convolutional, pooling and fully connected layer. Convolutional layer employs the convolution operation as the projection method between input and output. The function of pooling layer is downsampling the convolutional layer outputs. Fully connected layer connects all outputs of the previous layer.

Nowadays, researchers have proposed several state-of-the-art CNN models, such as AlexNet [21], VGG Net [22], GoogleNet [23] and ResNet [24]. AlexNet shows the powerful ability of deep neural network in dealing with image classification task for the first time. Besides, ResNet is another landmark model in the progress of deep architecture development, which overcomes the difficulty of training the very deep network.

B. LONG SHORT TERM MEMORY NETWORK

LSTM has been demonstrated having the ability of capturing the sequential patterns contained in the time-series or spatial-series data [25], [26]. To obtain the output u_t , the input x_t is processed through input, forget and output gate functions. These gate functions control the information to be remained or discarded for the cell of next step. A standard LSTM cell is defined by the following functions, where i , f , and o represent input gate, forget gate and output gate respectively.

$$i_t = \sigma(W_{xi}x_t + W_{ui}u_{t-1} + W_{ci}c_{t-1} + b_i), \quad (1)$$

$$f_t = \sigma(W_{xf}x_t + W_{uf}u_{t-1} + W_{cf}c_{t-1} + b_f), \quad (2)$$

$$c_t = f_t * c_{t-1} + i_t * \tanh(W_{xc}x_t + W_{uc}u_{t-1} + b_c), \quad (3)$$

$$o_t = \sigma(W_{xo}x_t + W_{uo}u_{t-1} + W_{co}c_t + b_o), \quad (4)$$

$$u_t = o_t * \tanh(c_t). \quad (5)$$

III. PROPOSED METHODS

Our proposed SODM consists of two parts, CNN-LSTM and OAM. The model instruction of construction and application stage is shown in Fig. 1. During the construction stage, WCE images without diseases are collected and then parted into sub-blocks with same size. These blocks are utilized to train the CNN-LSTM part, which generates the deviation value between real image block and expected one. Meanwhile, the prior-knowledge part of OAM is built with the information of fake outliers learned in the training process. In the application stage, firstly, blocks are also sampled from the WCE image to be identified. These blocks are fed into the CNN-LSTM part to calculate deviation values, which constitute the whole OAM combining with the prior-knowledge part. Afterwards, the confidence level of outlier can be concluded by OAM. The following sections will detail the structure and construction algorithms of SODM.

A. IMAGE PARTITIONING MODULE

In this part, procedures of partitioning WCE images are clarified. Fig. 2 shows the entire workflow. The image displayed in Fig. 2 is standard acquisition data of WCE camera. In the first step, one fixed-size block is selected as the target region, which is to be determined whether diseases occur or not. Through setting this fixed-size window to traverse the entire image, we can fulfill the outlier detection in every region. For the sake of illustration, central region (red box) is taken as the example in the following. In the second step, a circular area with the center of red box is bounded, which shown in yellow circle. This surrounding area is chosen as the spatial correlative input to infer the expected state of target region. For the purpose of building the deep model, it requires dividing this circular area into several blocks. Therefore, in step 3, blocks with the same size of target area are chosen sequentially in the radius direction every 10 degree. We mark one of them with the blue dotted rectangle as the example. There are regional overlap among these sub-blocks. To explain, blocks generated from one radius direction are listed in step 4. Evidently, blocks in the blue rectangle belong to surrounding areas (simply b-blocks), and block in the red rectangle represents target region (simply r-block). In the use process, b-blocks and r-block constitute one sample, the expected state of r-block will be inferred from

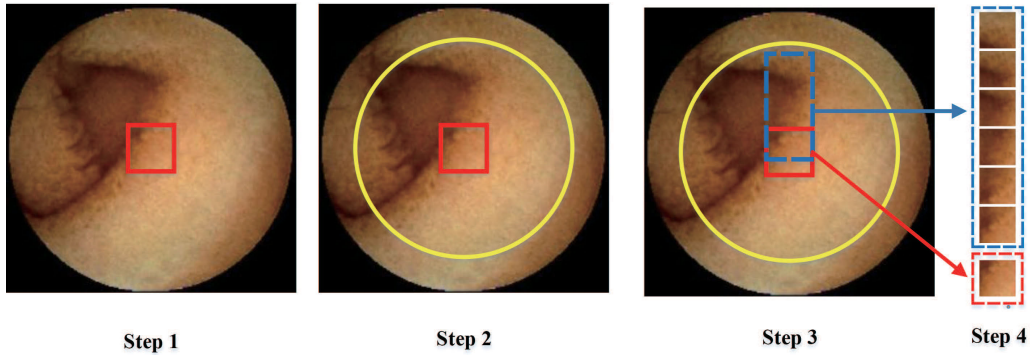


FIGURE 2. Procedures of partitioning WCE images.

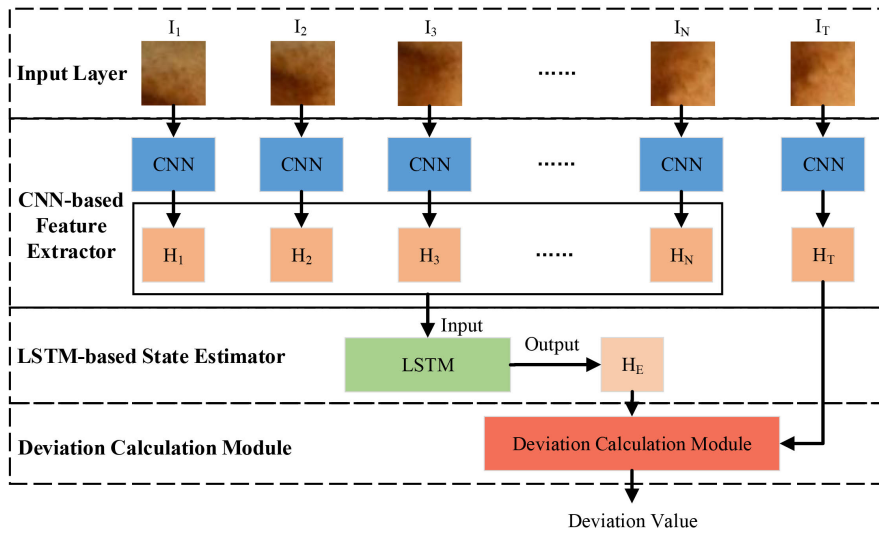


FIGURE 3. The whole workflow of CNN-LSTM.

the spatial pattern of b-blocks. The training and test samples are generated by executing above steps for each WCE image.

B. CNN-LSTM

1) MODEL FRAMEWORK

CNN-LSTM

is made up of two functional modules, the CNN-based feature extractor (CFE) and conditioned LSTM-based state estimator (LSE). Fig. 3 shows the whole workflow of CNN-LSTM, which including four parts, input layer, CFE, LSE and deviation calculation module.

In the input layer, samples are generated by following the procedures in Section III-A at first. To make this easier to follow, we choose only one sample in this section, where I_1 to I_N (b-blocks) represent the sequential data for correlation modeling and I_T (r-block) denotes the real image block to be predicted.

CFE is a deep architecture by combining part of one well-trained 50-layer ResNet (ResNet50, which is pre-trained with data from the ImageNet dataset) and one feature transition layer (FTL). Specifically, the last layer of ResNet50 is removed, replaced by FTL to make a bridge between the

feature extraction part and state estimator part. FTL performs two functions, one is to enhance the model flexibility, which allows the input of LSE to have the desirable size without limit of output layer dimensionality of ResNet50, the other is to map features learned from ResNet50 architecture to more appropriate patterns for representing the sequential correlation. The implementation formula of FTL is defined as follows:

$$H_n = W_{c-l}h_n + b_{c-l}, \quad n = 1, 2, \dots, N, T, \quad (6)$$

where h_n denotes features extracted by part of ResNet50 from input I_n , $n = 1, 2, \dots, N, T$ and W_{c-l}, b_{c-l} are the parameters of linear projection. Naturally, H_n is the output of CFE part. Since ResNet50 weights are fixed in this part, the parameters set to be determined of CFE is written as $\Theta_{CFE} = \{W_{c-l}, b_{c-l}\}$.

LSE is built with a standard LSTM network, which is used for predicting the state of target feature block along the spatial axis. Input for LSE is the data sequence H_1 to H_N , and the output is denoted as H_E . The parameters set of LSE is $\Theta_{LSE} = \{W_{uc}, W_{ui}, W_{uf}, W_{uo}, W_{ci}, W_{cf}, W_{co}, b_c, b_i, b_f, b_o\}$.

Obviously, H_E and H_T are the input to the deviation calculation module for calculating the deviation value. We choose

the root mean square error (RMSE) of H_E and H_T as the deviation value representation, denote as δ_{C-L} , the formula can be written as:

$$\delta_{C-L} = RMSE(H_E, H_T) \quad (7)$$

2) MODEL TRAINING

Only normal WCE images are adopted during the construction stage, therefore, the goal of training CNN-LSTM can be set as minimizing the δ_{C-L} value. By doing this, CNN-LSTM is to be built as a pure normal state estimator, so that the significant deviation will occur while a outlier is detected. As illustrated in Section III-B-1), the training goal is demanded for achieving by calculating the parameters set $\Theta_{CNN-LSTM} = \{\Theta_{CFE}, \Theta_{LSE}\}$, however, it is a hard problem to propose a global objective function for training Θ_{CFE} and Θ_{LSE} at the same time. Therefore, a nest-structured training method is proposed to update these parameters.

The core idea of this training method is to perform alternative cooperative training of CFE and LSE by defining the appropriate objective function respectively. More concretely, LSE is trained by minimizing the loss function in the following:

$$\mathcal{L}_{LSE} = \delta_{C-L}. \quad (8)$$

As for CFE part, a classification layer is added after FTL temporarily, whose parameters are denoted as $\Theta_{C-Layer}$. Therefore, we can regard CFE as a supervised classification model during the training stage. The typical cross-entropy loss function of a supervised classification model can be written as:

$$\begin{aligned} \mathcal{L}_{cross-entropy} &= -\frac{1}{M} \sum_{i=1}^M [y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)], \quad (9) \end{aligned}$$

where M is the number of samples, y_i represents the real label and p_i is the predicted label. Considering that the impact of CFE output on deviation value, the predicted label can be replaced by item relating to deviation value, specially, $p_i = sigmoid(\delta_{C-L})$ is used in our work. The closer that the value of $sigmoid(\delta_{C-L})$ is to 1, the more significant the outlier is. Ideally, this value should be closer to 0 while handling with the normal data. There are no abnormal samples while training model, so y_i can be set as 0 all the time. The cost function of CFE is rewritten as:

$$\mathcal{L}_{CFE} = -\frac{1}{M} \sum_{i=1}^M \log(1 - p_i), \quad (10)$$

Procedures of training method are described as follows, while taking one sample as the example.

C. OAM

1) MODEL FRAMEWORK

OAM proposes a novel empirical strategy for declaring the outliers with as less false detections as possible. In this work, OAM is designed by considering the following rules.

Algorithm 1 Procedures of Model Construction

- 1: **Input:** Sequential data $\{I_n\}$, Training epoch: N_{CFE}, N_{LSE}
- 2: **Output:** $\Theta_{CNN-LSTM} = \{\Theta_{CFE}, \Theta_{LSE}\}$
- 3: Initialize Θ_{CFE} , $\Theta_{C-Layer}$ and Θ_{LSE} with small random values
- 4: By following the workflow of CNN-LSTM, calculate the deviation value with H_E and H_T
- 5: **for** $count_{CFE} = 1 : N_{CFE}$ **do**
- 6: **for** $count_{LSE} = 1 : N_{LSE}$ **do**
- 7: Update Θ_{LSE} with \mathcal{L}_{LSE} , while fixing Θ_{CFE} and $\Theta_{C-Layer}$
- 8: **end for**
- 9: Update Θ_{CFE} and $\Theta_{C-Layer}$ with \mathcal{L}_{CFE} , while fixing Θ_{LSE}
- 10: **end for**

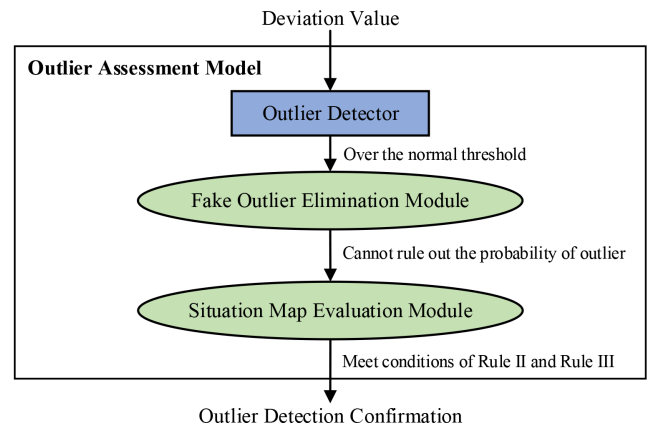


FIGURE 4. The structure of OAM.

a: RULE I

Most of small intestinal diseases show a wide variety of spatial and color natures. Even if only normal WCE images are used, the circumstance that predicted state does not match the real state happens occasionally. Therefore, it is necessary to put features of these fake outliers into consideration while detecting outliers during the practical application.

b: RULE II

The target block would be identified as a outlier, if deviation values measured from the majority of surrounding directions are above than normal level.

c: Rule III

While abnormally high deviation values occur, the confidence level of outlier detection should be high.

On the basis of above factors, OAM consists of one outlier detector and two functional cells, outlier detector, fake outlier elimination module (following Rule I) and situation map evaluation module (following Rule II and Rule III) respectively. The structure of OAM is shown in Fig. 4.

Outlier detector module receives and determines whether δ_{C-L} is an outlier or not. And cells evaluates the confidence level of outlier from various aspects. Particularly, **fake outlier elimination module** is used to remove the false outlier alarm by using the knowledge learned in the training stage. Besides, for each sample, **situation map evaluation module** is composed of the collection by deviation values generated from all directions.

The relationships of various modules in OAM are illustrated in Fig. 4. Switching conditions are detailed in the following.

- 1) *Outlier detector to Fake outlier elimination module*: δ_{C-L} is out of the normal range.
- 2) *Fake outlier elimination module to Situation map evaluation module*: It can not be determined that the target block is not the outlier by only using the learned knowledge of fake outlier.
- 3) *Situation map evaluation module to Outlier Detection Confirmation*: Meet conditions illustrated by Rule II and Rule III.

2) FORMULATION OF OAM

This subsection introduces the construction methods for modules in OAM.

a: OUTLIER DETECTOR

This module need to screen out the potential outliers by using δ_{C-L} values. Therefore, the crucial step is to calculate the normal range of deviation value. We remove the ten percent largest δ_{C-L} obtained in the training stage and then measure the average of left deviation values. This average value would be set to the threshold of the acceptable range of deviation.

b: FAKE OUTLIER ELIMINATION MODULE

In this module, the knowledge of fake outliers should be defined at first. In general, deviation value exceeding the threshold of normal range is one of the most important characteristics of fake outlier. Therefore, we collect all these feature blocks (H_E) with abnormal δ_{C-L} to constitute a fake outlier knowledge base. In the application stage, each feature block from outlier detector would measure the similarity with all elements of this knowledge base, details of similarity computing algorithm can be referred in [27]. Blocks of high similarity with known fake outliers will be ruled out the possibility of anomaly.

c: SITUATION MAP EVALUATION MODULE

This module should realize functions by following Rule II and Rule III. Therefore, two evaluation parts are included. One is to calculate the average of deviation values from all directions, denoted as μ_δ . The other is tallying the proportion of δ_{C-L} with abnormally high deviation values, denoted as AH_δ . The threshold for deciding the abnormally high δ_{C-L} is twice the threshold of normal range in outlier detector.

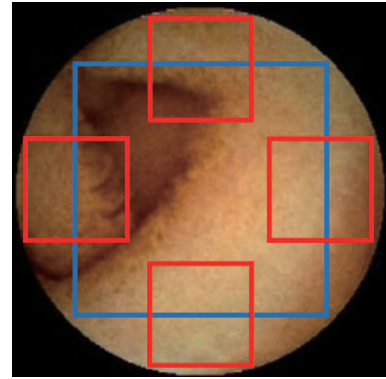


FIGURE 5. The area of target blocks.

The formulation of final confidence level is written as below.

$$Score_{outlier} = \omega_1 * \mu_\delta + \omega_2 * AH_\delta, \quad (11)$$

where ω_1 and ω_2 are weight coefficients for balancing the importance of two indexes.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. DATA PREPARATION

In order to investigate the effectiveness of our proposed SODM for small intestinal diseases detection, datasets consist of normal and abnormal parts are considered.

d: NORMAL WCE IMAGES

This part is provided by the BeijingJiShuiTan Hospital. In all, there are about 22 thousand WCE images collecting from twenty different men. All personal information of them are removed with privacy considerations.

e: ABNORMAL WCE IMAGES

In order to cover a variety of small intestinal diseases, we obtain WCE images with outliers from the public source published in [28], which lists various sorts of small intestinal diseases. We choose five representative symptoms to validate the proposed algorithm, which including ulcers, erythema, protruding lesions, polyp and vascular malformation. For each symptom, three to five samples are selected.

f: PREPROCESSING PROCEDURES

Several procedures are performed to complete the establishment of final data set.

1) *Annotation*: The outliers in WCE images are needed to annotate. In this work, annotation of each abnormal image is fulfilled by three doctors, which consists of two items, categories of diseases and lesion locations.

2) *Image Partitioning*: Target blocks should be selected from each WCE image before executing the partitioning procedures shown in Section III-A. Considering that the view field of WCE image is a round shape, the area of the largest square within the circle and four rectangle regions around the edge is set to generate target blocks. The sketch in Fig. 5

displays this area, where the size of red rectangle is equal with that of the target block. The typical view field is a circle with about 230 pixels in diameter, hence the size of the largest square is about 160×160 pixels. In this square, target blocks of 40×40 pixels size are sampled with 50% overlap in the transverse direction and longitudinal direction, so 49 blocks can be obtained. In all, adding four extra rectangles, 53 target blocks are extracted from each WCE image. Afterwards, the data samples are created by following the partitioning procedures stated in Section III-A. In General, 36 samples would be generated for each target block, apart from those in the edge area. For more information, the radius of surrounding area is 80 pixels and the overlap of sequential sampling is 30 pixels, hence there would be six input blocks and one target block constituting one sample.

3) *Datasets Construction*: We would construct two datasets in this work. One is composed of samples generated in *Image Partitioning* part, denoted as **Sequential Dataset**, which is used for building SODM and further analysis of the model performance. In order to verify the effectiveness of our model, several published outlier detection methods are selected to complete performance comparison. For most of these methods, sequential samples for input are not needed. Hence, the other dataset, denoted as **Discrete Dataset**, is designed by collecting all the image blocks created in *Image Partitioning* part. In this dataset, each image block is regarded as one input sample. Besides, samples are divided into two categories, normal and abnormal. Abnormal ones are defined by blocks with outliers. It is important to note that the information contained in data are exactly the same between two datasets, just varying in input mode for different models.

The division way of training and application parts for two datasets are the same. More concretely, samples of normal WCE images from BeijingJiShuiTan Hospital are used in the training stage, meanwhile, abnormal images from [28] provide samples for the application stage.

B. PERFORMANCE ON OUTLIER DETECTION OF SMALL INTESTINAL DISEASES

1) COMPARISON METHODS

Our proposed model is compared with several outlier detection methods as below.

- 1) ResNet50 + K-nearest neighbors algorithms (KNN);
- 2) ResNet50 + Local outlier factor (LOF) [29];
- 3) ResNet50 + One-class SVM (OC-SVM) [30];
- 4) ResNet50 + Support Vector Data Description (SVDD) [31];
- 5) ResNet50 + One-class Conditional Random Field (OC-CRF) [32];
- 6) Deep Structured Energy based Model (DSEBM) [33];
- 7) Unsupervised Outlier Detection Model (UODA) [34];
- 8) Early Fault Detection Model (FDDA) [35];
- 9) A simple version of SODM (SODM-S1)

Among the above Methods, 1-5 represent one of the most typical frameworks in the field of outlier detection, which

contains feature extraction stage and outlier detection stage. For making a relative fair comparison, Methods 1-5 adopt a pre-trained ResNet50 architecture as the feature extractor, which remains consistent with our work. Besides, classic outlier detection methods, KNN, LOF, OC-SVM and SVDD are utilized to construct the discrimination model for normal samples. Methods 5-8 are four effective methods proposed for solving the outlier detection problems of sequential data. OC-CRF learns the dependence from one-class data by using CRF. DSEBM proposes a solution to outlier detection problem through making use of deep generative model with energy function. Frameworks of UODA and FDDA are most closer to ours, which also consists of two parts: feature extraction module and system state predictor. However, there are still significant differences between our model and these two. The differences lie in two aspects. Firstly, our model brings a more powerful capacity of feature extraction and sequence modeling, which is mainly attributed to the model structure and effective training method. Secondly, OAM in our model supplies a better solution to detecting the small intestinal diseases in high accuracy. The last method is a simple version of SODM, which differs in the training method with original SODM. Particularly, we construct the CFE part of SODM-S1 by using the normal data, the form of loss function is equal to (10), whereas p_i is generated by additional classification layer directly instead of $\text{sigmoid}(\delta_{C-L})$. By this way, contributions of the proposed training method can be observed clearly.

2) IMPLEMENTATION DETAILS

Firstly, input features are extracted by the pre-trained Resnet50 from normal data of discrete dataset, which are fed into the outlier detection parts of method 1-4. In Method 1, KNN splits normal features into k clusters. The criteria of determining outlier is that sample to be identified does not belong to any of k clusters. Besides, the optimal number of cluster is selected by searching $\{1, 2, 4, 8, 16, 32, 64\}$. While using LOF-based method, the maximum of LOF values for input features is chosen as the threshold for detecting the outlier. As for method 3 and 4, input features are directly used for constructing one-class classification model. The implementation procedures and parameters for OC-CRF, DSEBM, UODA and FDDA can refer to the contents in corresponding articles [32]–[35].

As illustrated in Section III, our SODM consists of CNN-LSTM and OAM. Parameter values are shown in Table 1.

3) EVALUATION METRIC

Three types of evaluation metrics are utilized: specificity, sensitivity and accuracy. These metrics are measured by true positive (TP), true negative (TN), false positive (FP) and false negative (FN). TP represents the numbers of samples that detecting outliers correctly, TN is the samples count for detecting normalities correctly. Accordingly, FP and FN are the numbers of normalities and outliers that

TABLE 1. Parameters of SODM.

Parameters	Value
Input size of FTL (Output size of ResNet50 part)	2048
Output size of FTL	1024
Input size of LSTM	1024
Output size of LSTM	1024
ω_1 in situation map evaluation module	0.6
ω_2 in situation map evaluation module	0.4
Threshold for normal range of δ_{C-L}	0.4
Threshold for confirming the outlier	0.5

classified incorrect. The calculation formula of these metrics are as below.

$$Sensitivity = \frac{TP}{TP + FN} \tag{12}$$

$$Specificity = \frac{TN}{TN + FP} \tag{13}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

4) OUTLIER DETECTION RESULTS ON THE DATA SET

Twenty-three WCE images are collected for application stage. Therefore, 1219 target blocks are created, including 937 normal blocks and 282 blocks with outliers according to the annotations from doctors. The values of TP, TN, FP, FN, Accuracy, Sensitivity and Specificity generated by different methods are detailed in Table 2.

Based on the simple observation on results in Table 2, TP value of SODM outperforms those of other comparison methods. It will lead to a low missed diagnosis rate, which is the most important guideline for medical diagnosis process. For the purpose of a complete and in-depth analysis of our model, three evaluation metrics are drawn in Fig. 6.

The index of accuracy indicates the overall performance of model, the accuracy of our model is 93.27%, which is higher than others. KNN obtains the best accuracy performance 90.40% among the methods without considering the input data correlation (Methods 1-4), which is 2.87% less than our model. Among methods with sequential data analyzing technique (Method 5-9), FDDA with 91.30% accuracy is the highest, nevertheless, accuracy of our model is still 1.97% higher than it.

What’s more, three more facts are worthy of attention in Fig. 6.

TABLE 2. Values of TP, TN, FP, FN, Accuracy, Sensitivity and Specificity generated by different methods.

ID	Method	TP	TN	FP	FN	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	KNN	201	901	36	81	90.40	71.28	96.16
2	LOF	178	890	47	104	87.61	63.12	94.98
3	OC-SVM	175	887	50	107	87.12	62.06	94.66
4	SVDD	196	892	45	86	89.25	69.50	95.20
5	OC-CRF	214	872	65	68	89.09	75.89	93.06
6	DSEBM	228	880	57	54	90.89	80.85	93.92
7	UODA	225	885	52	57	91.06	79.79	94.45
8	FDDA	234	879	58	48	91.30	82.98	93.81
9	SODM-S1	222	882	55	60	90.57	78.72	94.13
10	SODM	243	894	43	39	93.27	86.17	95.41

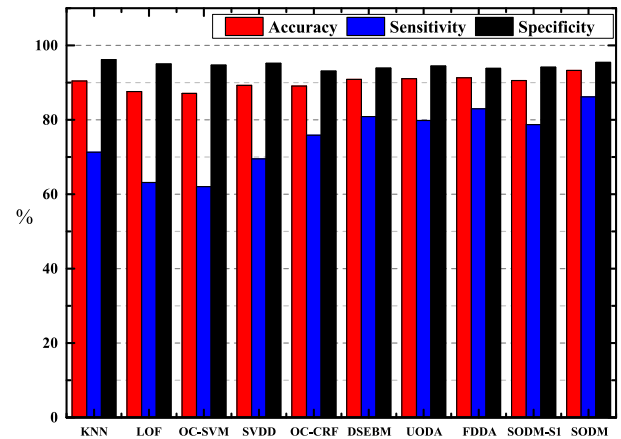


FIGURE 6. Values of accuracy, sensitivity and specificity for various methods.

Firstly, the performance of our model is not significant superior to other methods only upon indexes of accuracy (red bar) and specificity (black bar). The cause of this situation can be explained by the composition of dataset for application. The normal blocks account for the large part, about 76.9%. As noted earlier, accuracy is a overall performance index, and specificity measures the model capacity of classifying normal blocks correctly. Hence, even if the model can not screen outliers out effectively, it still has a relative good accuracy and specificity by grouping outliers into normal category. For instance, sensitivity of KNN is only 71.28%, but accuracy and specificity are 90.40% and 96.16% respectively. In this circumstance, sensitivity become a more important index for indicating the model performance, which reflects the capacity of outlier detection. SODM has 86.17% sensitivity, which is obviously higher than those of other methods (the second-highest sensitivity is 82.98% of FDDA). The sensitivity value suggests a outstanding ability of SODM for detecting outliers.

Secondly, accuracy and sensitivity indexes of methods 5-10 are generally superior to those of methods 1-4, which shows the advantages of considering data correlation while handling with this kind of outlier detection problem.

Thirdly, comparison between SODM-S1 and SODM proves that our proposed training method can truly boost the model performance by adjusting features to strengthen the

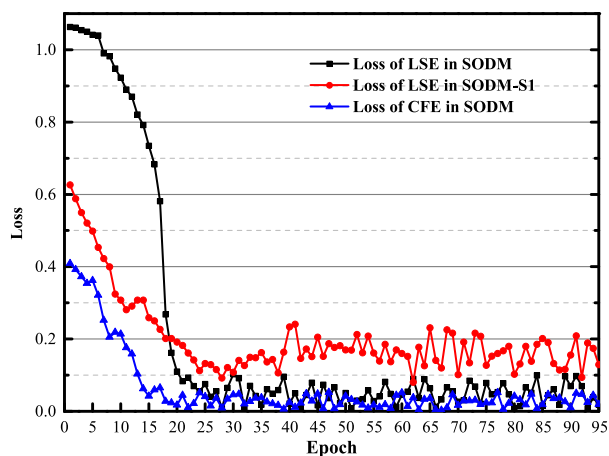


FIGURE 7. Loss curves of SODM and SODM-S1.

spatial dependence. In order to analyze this issue further, loss curves of CFE (blue line) and LSE (black line) in SODM, loss curve of LSE (red line) in SODM-S1 are illustrated in Fig. 7. Several observations can be concluded as follows: 1) In the early stage of training process (epoch 1 to 10), the descent trend of black line is smoother than those of other lines. It is because that CFE is seeking the optimal parameters during this stage (blue line descends quickly), which affects the quality of features used as the LSE module input. 2) Once the CFE of SODM is well-trained, the descent trend of black line becomes sharper than others and approaches the desirable level promptly, we can observe this phenomenon from the variation trends of black and blue lines between epoch 10 to 20. 3) The CFE part of SODM-S1 has completed the construction before used as the input source for LSE, hence the loss curve of LSE shows a slowly declined trend. 4) Most of all, LSE in SODM achieves a better learning result than SODM-S1 does (black line is lower than red line), which represents the usefulness of our proposed training method.

5) ROC CURVES OF DIFFERENT MODELS

Receiver Operating Characteristic (ROC) curve is one of the most crucial metric for evaluating the model practical performance. According to results in Section IV-B-4), KNN and FDDA are the performance optimal models with different types of input. Therefore, we would calculate ROC curves of SODM, KNN and FDDA for further analysis in this part. The specific outcomes are displayed in Fig. 8. x-axis is True Positive Rate (TPR, equal to sensitivity), y-axis is False Positive Rate (FPR), whose calculation formula is $FPR = FP / (FP + TN)$.

Area under the curve (AUC) of different methods are computed by using their respective ROC curves. The value of AUC is 0.942 for SODM, 0.902 for FDDA and 0.847 for KNN. This result reflects the practicability of our proposed model.

What’s more, it can be concluded that SODM is more sensitive to outlier occurrence by analysing the curves displayed

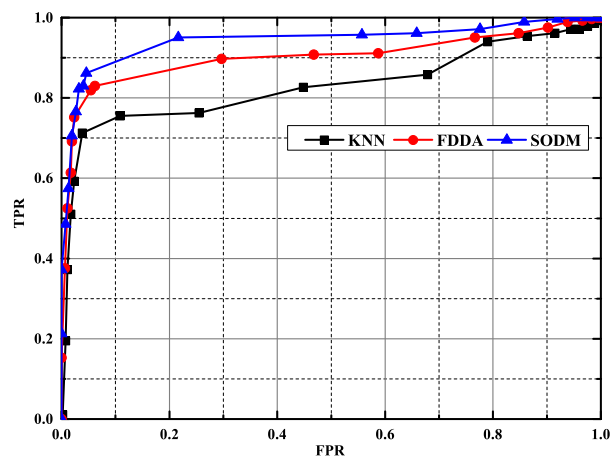


FIGURE 8. ROC curves of KNN, FDDA and SODM.

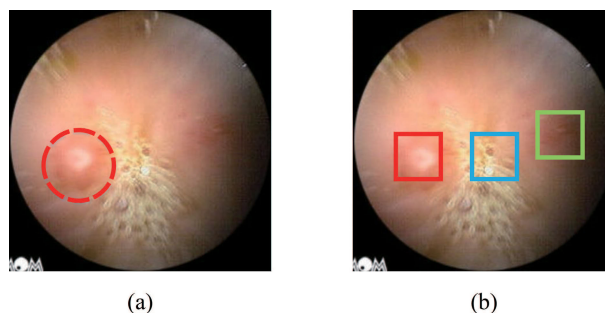


FIGURE 9. Image for illustration of outlier generation process. (a) WCE image with ulcer, whose lesion area is annotated with the red dashed circle. (b) Candidate outlier blocks generated by Outlier Detector.

TABLE 3. Similarity degree of three candidate blocks.

Red block	Blue block	Green block
0.128	0.209	0.892

in Fig. 8. The ROC calculation method for these three models can be roughly expressed in one way: outlier indicators for samples are compared with a variable threshold to determine its category. Particularly, the indicator of SODM and FDDA is deviation value of real input and predicted state, and distance between input sample and cluster center is the indicator for KNN. For all these indicators, larger value suggests higher probability of a outlier. Apparently, quality of indicators can directly affect the performance of ROC curve. Therefore, only when indicator is sensitive to outlier and generates a high indicator value, TPR can keep an appropriate level. It can be observed that TPR of SODM is superior to those of other two models (blue curve is above red and black curves), which distinctly shows the sensitivity of SODM while receiving the outlier input.

6) EMPIRICAL ANALYSIS

In this section, we will discuss the outlier generation process of SODM. One representative image drawn from small intestinal ulcers is taken as the example. The chosen image

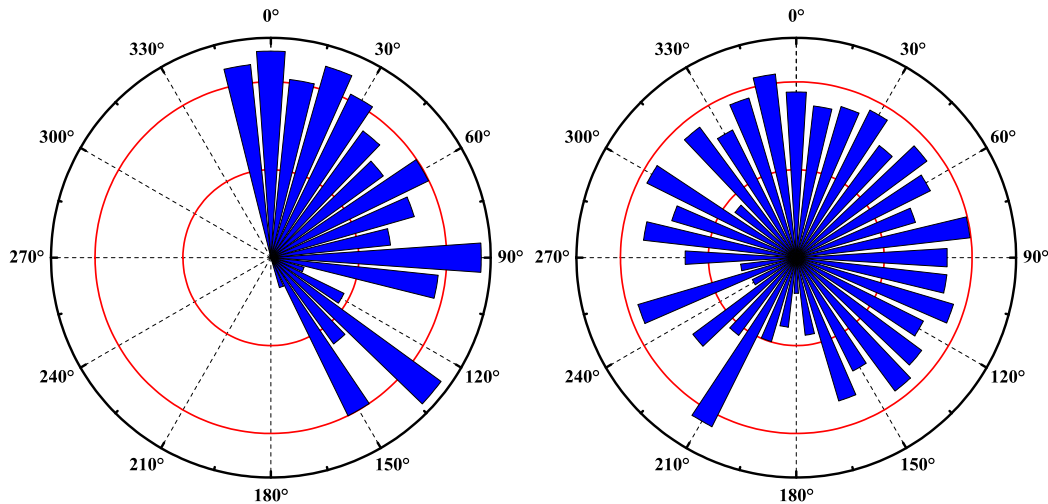


FIGURE 10. Situation map for target block. Left: target block in red color; Right: target block in blue color.

with doctors' annotation is shown in Fig. 9(a). The area contained ulcer is surrounded by the red dashed circle. First of all, blocks with δ_{C-L} greater than normal threshold are output by the *Outlier Detector*. The lesion area covers multiple blocks and each target block is evaluated by sequences from all directions, so it would screen out very similar blocks as the outlier sign for one region. For the convenience of the analysis, representative blocks are selected from each candidate block clusters. In this example, three representative blocks are displayed in red, blue and green respectively in Fig. 9(b). It is obvious that red block detects the outlier correctly.

Afterwards, *Fake Outlier Elimination Module* receives these three blocks to calculate the similarity index degree with every elements in the fake outlier knowledge base. The maximums of similarity for each block are detailed in Table 3. As set in this article, the corresponding block would be classified into fake outlier category if similarity degree is greater than 0.7. Therefore, the green block will be eliminated during this session.

The next stage is that *Situation Map Evaluation Module* makes the final determination by the surrounding information of target block. We utilize polar diagram to display the deviation values generated by sequence data from all directions. The details are shown in Fig. 10. The left part denotes the target block in red color, and the right part is the situation map for target block in blue color. The blue bar pointing various directions in Fig. 10 represents δ_{C-L} generated from the corresponding direction, and length of blue bar is the value of δ_{C-L} . Radius of the smaller red circle is the threshold of normal range, and radius of the larger red circle represents the abnormally high δ_{C-L} threshold.

In the situation map of blue target block, deviation values are missing in the zone from 170 degree to 340 degree, it is because that there is no sufficient space to create the sequential data. From the observation of δ_{C-L} bars generated from -20 degree to 160 degree, three factors can be concluded: 1) most of δ_{C-L} are greater than normal range threshold

(exceeds the smaller circle), which fits the circumstance described in Rule II. Hence, μ_δ can be calculated, the value comes out at 0.69. 2) Seven δ_{C-L} have higher values than abnormally high threshold (exceeds the larger red circle), which is the situation mentioned in Rule III. And the result of AH_δ is 0.39 3) the existence of bubble area (on the right side of target block) destroys the spatial correlation of sequence data in these directions, so that the predicted states by input from these directions are disturbed, which lead to δ_{C-L} with very small values occur in the zone from 100 degree to 110 degree and 160 degree. Factors 1)-2) can help us classifying the target block into outlier category correctly ($Score_{outlier} = \omega_1 * \mu_\delta + \omega_2 * AH_\delta = 0.6 * 0.69 + 0.4 * 0.39 = 0.57 > OutlierThreshold$). Although factor 3) has no effect on the final determination, it is still worthy of attention to designing a more robust model to avoid this kind of disturbance.

As for the situation map in the right part, two things deserve attention. One is that although values of δ_{C-L} satisfy the constraints proposed in Rule II and Rule III, we still rule out the possibility of outlier with a overall evaluation ($Score_{outlier} = \omega_1 * \mu_\delta + \omega_2 * AH_\delta = 0.6 * 0.61 + 0.4 * 0.06 = 0.39 < OutlierThreshold$). The other is that δ_{C-L} has small values in the directions of bubble areas (170 degree to 200 degree). The first one makes us classify blue target block correctly. It proves that functional modules can be used for handling the complex situations. The second thing reflects the ability of our model for discovering spatial correlation.

V. CONCLUSION

The article proposes a novel mode, SODM, to address the problem of outlier detection in WCE images. SODM aims to learn the appropriate features from WCE images blocks to construct the state predictor, the difference between real and expected state can be used to screen out the outlier occurrence. Furthermore, SODM also contains a outlier assessment model to make the outlier declaration with less false alarms. The experiments with real WCE images confirm the

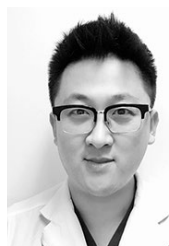
superiority of proposed model. Besides, the further research directions are also discussed, including designing a robust model to avoid the disturbance existed in WCE image and a more sensitive module to detect real outliers.

ACKNOWLEDGMENT

(Yan Gao and Weining Lu contributed equally to this work.)

REFERENCES

- [1] Y. Fu, W. Zhang, M. Mandal, and M. Q.-H. Meng, "Computer-aided bleeding detection in WCE video," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 2, pp. 636–642, Mar. 2014.
- [2] M. Mathew and V. P. Gopi, "Transform based bleeding detection technique for endoscopic images," in *Proc. 2nd Int. Conf. Electron. Commun. Syst. (ICECS)*, Feb. 2015, pp. 1730–1734.
- [3] M. Souaidi, A. A. Abdelouahed, and M. El Ansari, "Multi-scale completed local binary patterns for ulcer detection in wireless capsule endoscopy images," *Multimedia Tools Appl.*, vol. 78, no. 10, pp. 13091–13108, May 2019.
- [4] T. Ghosh, S. K. Bashar, M. S. Alam, K. Wahid, and S. A. Fattah, "A statistical feature based novel method to detect bleeding in wireless capsule endoscopy images," in *Proc. Int. Conf. Informat., Electron. Vis. (ICIEV)*, May 2014, pp. 1–4.
- [5] O. Bchir and M. M. B. Ismail, "Empirical comparison of visual descriptors for ulcer recognition in wireless capsule endoscopy video," in *Proc. Comput. Sci. Inf. Technol.*, York, U.K., Apr. 2018, pp. 402–407.
- [6] E. Tuba, M. Tuba, and R. Jovanovic, "An algorithm for automated segmentation for bleeding detection in endoscopic images," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 4579–4586.
- [7] V. Kodogiannis and M. Boulougoura, "An adaptive neurofuzzy approach for the diagnosis in wireless capsule endoscopy imaging," *Int. J. Inf. Technol.*, vol. 13, no. 1, pp. 46–56, 2007.
- [8] L. Alexandre, N. Nobre, and J. Casteleiro, "Color and position versus texture features for endoscopic polyp detection," in *Proc. Int. Conf. Biomed. Eng. Informat.*, May 2008, pp. 38–42.
- [9] B. Li, Y. Fan, M. Q.-H. Meng, and L. Qi, "Intestinal polyp recognition in capsule endoscopy images using color and shape features," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2009, pp. 1490–1494.
- [10] B. Zhao, J. Feng, X. Wu, and S. Yan, "A survey on deep learning-based fine-grained object classification and semantic segmentation," *Int. J. Autom. Comput.*, vol. 14, no. 2, pp. 119–135, Apr. 2017.
- [11] D. Shen, G. Wu, and H. I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.
- [12] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: A review of literature," *Multimodal Technol. Interact.*, vol. 2, no. 3, p. 47, 2018.
- [13] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019.
- [14] X. Jia and M. Q.-H. Meng, "A deep convolutional neural network for bleeding detection in wireless capsule endoscopy images," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Orlando, FL, USA, Aug. 2016, pp. 639–642.
- [15] X. Jia and M. Q.-H. Meng, "A study on automated segmentation of blood regions in wireless capsule endoscopy images using fully convolutional networks," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 179–182.
- [16] Y. Yuan and M. Q.-H. Meng, "Deep learning for polyp recognition in wireless capsule endoscopy images," *Med. Phys.*, vol. 44, no. 4, pp. 1379–1389, Apr. 2017.
- [17] X. Liu, J. Bai, G. Liao, Z. Luo, and C. Wang, "Detection of protruding lesion in wireless capsule endoscopy videos of small intestine," *Proc. SPIE*, vol. 10575, Feb. 2018, Art. no. 1057513.
- [18] S. Fan, L. Xu, Y. Fan, K. Wei, and L. Li, "Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images," *Phys. Med. Biol.*, vol. 63, no. 16, 2018, Art. no. 165001.
- [19] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 15.1–15.58, 2009.
- [20] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, p. 436, May 2015.
- [21] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012, pp. 1097–1105.
- [22] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," 2014, *arXiv:1405.3531*. [Online]. Available: <https://arxiv.org/abs/1405.3531>
- [23] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
- [24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [26] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A critical review of recurrent neural networks for sequence learning," 2015, *arXiv:1506.00019*. [Online]. Available: <https://arxiv.org/abs/1506.00019>
- [27] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [28] *Clinical Diseases of Capsule Endoscopy*. Accessed: 2011. [Online]. Available: <https://wenku.baidu.com/view/2d9dfc85caaed3382c4d306.html>
- [29] H. Ma, Y. Hu, and H. Shi, "Fault detection and identification based on the neighborhood standardized local outlier factor method," *Ind. Eng. Chem. Res.*, vol. 52, no. 6, pp. 2389–2402, Feb. 2013.
- [30] D. Fernández-Francos, D. Martínez-Rego, O. Fontenla-Romero, and A. Alonso-Betanzos, "Automatic bearing fault diagnosis based on one-class v -SVM," *Comput. Ind. Eng.*, vol. 64, no. 1, pp. 357–365, Jan. 2013.
- [31] J. Huang and X. Yan, "Related and independent variable fault detection based on KPCA and SVDD," *J. Process Control*, vol. 39, pp. 88–99, Mar. 2016.
- [32] Y. Song, Z. Wen, C.-Y. Lin, and R. Davis, "One-class conditional random fields for sequential anomaly detection," in *Proc. 23rd Int. Joint Conf. Artif. Intell.*, 2013, pp. 1685–1691.
- [33] S. Zhai, Y. Cheng, W. Lu, and Z. Zhang, "Deep structured energy based models for anomaly detection," in *Proc. 33rd Int. Conf. Mach. Learn. (ICML)*, New York City, NY, USA, 2016, pp. 1100–1109.
- [34] W. Lu, Y. Cheng, C. Xiao, S. Chang, S. Huang, B. Liang, and T. Huang, "Unsupervised sequential outlier detection with deep architectures," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4321–4330, Sep. 2017.
- [35] W. Lu, Y. Li, Y. Cheng, D. Meng, B. Liang, and P. Zhou, "Early fault detection approach with deep architectures," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 7, pp. 1679–1689, Jul. 2018.



YAN GAO received the B.S. and master's degrees from the Department of Clinical Medicine, Capital Medical University, Beijing, China, in 2007 and 2010, respectively. He currently serves in the Department of Gastroenterology, Beijing Jishuitan Hospital and is a member of the Digestive Endoscopy Group, Chinese Medical Doctors Association. His research interests include digestive endoscopy, and the diagnosis and treatment of gastrointestinal diseases.



WEINING LU received the B.S. degree from the Department of Physics, Fudan University, Shanghai, China, in 2007, and the Ph.D. degree from the Department of Automation, Tsinghua University, in 2017. He is currently an Assistant Professor with the Beijing National Research Center for Information Science and Technology, Tsinghua University. His current research interest includes solving anomaly detection problems by using deep architecture networks, computer vision, and data mining.



XIAOBEI SI received the B.S. and master's degrees from China Medical University, Shenyang, China, in 2011 and 2013, respectively. His research interests include helicobacter pylori infection, and the diagnosis and treatment of gastrointestinal diseases.



YU LAN received the B.S. degree from the Department of Clinical Medicine, Xinjiang Medical University, Xinjiang, China, in 1985, and the Ph.D. degree from the Chinese Academy of Medical Sciences and the Peking Union Medical College, Beijing, China, in 1999. She currently serves in the Department of Gastroenterology, Beijing Jishuitan Hospital. Meanwhile, she is a Professor with the Peking University School of Medicine and a member of the Digestive Branch of the Chinese Medical Association. Her current research interests include digestive endoscopy and the diagnosis and treatment of gastrointestinal motility disorders.

• • •