Application of Structural Similarity Analysis of Visually Salient Areas and Hierarchical Clustering in the Screening of Similar Wireless Capsule Endoscopic Images

Rui Nie1,2, Huan Yang2,3
1 Chongqing Jinshan Science & Technology (Group) Co., Ltd.
Chongqing, China
2 Department of Gastroenterology, Second Affiliated Hospital, Army Medical University
Chongqing, China
3 Both authors contribute equally to this work

Hejuan Peng3
3 Chongqing Jinshan Science & Technology (Group) Co., Ltd.
Chongqing, China

Weiya Fan5
5 Chongqing Jinshan Science & Technology (Group) Co., Ltd.
Chongqing, China

Jing Liao7
7 Chongqing Jinshan Science & Technology (Group) Co., Ltd.
Chongqing, China

Fang Huang8,a,*
8 Chongqing Jinshan Science & Technology (Group) Co., Ltd.
Chongqing, China

* Corresponding author: a modinaxi@163.com

Abstract—Small intestinal capsule endoscopy is the mainstream method for inspecting small intestinal lesions, but a single small intestinal capsule endoscopy will produce 40,000–120,000 images, the majority of which are similar and have no diagnostic value. It takes 1–3 hours for doctors to identify lesions from these images. This is time-consuming and increases the possibility of misdiagnosis and missed diagnosis, because doctors may experience visual fatigue while paying attention to a large number of similar images for a long time. In order to solve these problems, we proposed a similar wireless capsule endoscope (WCE) image screening method based on structural similarity analysis and the hierarchical clustering of visually salient sub-image blocks. The similarity clustering of images was automatically identified by hierarchical clustering based on the hue, saturation, value (HSV) spatial color characteristics of the images, and the keyframe images were extracted based on the structural similarity of the visually salient sub-image blocks, in order to accurately identify and screen out similar small intestinal capsule endoscopic images. Subsequently, the proposed method was applied to the capsule endoscope imaging workstation. After screening out similar images in the complete data gathered by the Type I OMOM Small Intestinal Capsule Endoscope from 52 cases covering 17 common types of small intestinal lesions, we obtained a lesion recall of 100% and an average similar image reduction ratio of 76%. With similar images screened out, the average play time of the OMOM image workstation was 18 minutes, which greatly reduced the time spent by doctors viewing the images.

Keywords—Visually salient; Agglomerative hierarchical clustering; Wireless capsule endoscopy (WCE); Screening similar images

I. INTRODUCTION

The small intestine capsule endoscope moves along with intestinal peristalsis, capturing images at a fixed or variable frame rate. It takes about 12 hours for the endoscope to capture all the required images. Each case involves about 40,000–120,000 small intestinal capsule endoscopic images, and it takes an average of 1–3 hours for a doctor to diagnose a patient, which means that the doctor suffers from a huge workload when viewing these images. High cost of manpower and time hinders the promotion of medical small intestine capsule endoscopy. With the continuous development of technology, in order to capture the images more comprehensively and reduce missed detection, medical small intestine capsule endoscope has achieved a higher frame rate and image resolution, which means
that the number of images of each patient will increase exponentially, and the workload of doctors to view images will also increase. Since the captured images are spatiotemporally continuous, a large number of adjacent images are highly similar. If similar endoscopic images of small intestine capsule could be accurately identified and screened, and the integrity of the small intestine mucosa could be maintained without losing the abnormal lesion images, the workload of doctors would be greatly reduced.

Previous researchers have done a lot of related work on the screening of small intestine capsule endoscopy images. Zhou et al. [1-7] compared the similarity of the capsule endoscopy images by extracting image features, and removes redundant image data, so as to reduce the workload of doctors reading images. These works have made some achievements in identifying the similarity of endoscopic images of small intestinal capsules, and many algorithms have been tried. For example, key frame extraction method, optical flow method, image data dimension reduction, image registration and motion model estimation algorithm are adopted.

Deep learning techniques rely on a large amount of training set data for learning. Xiao et al. [8-12] used convolution neural network technology to classify or recognize the endoscopic images of small intestine capsules, thus identifying and prompt one or more lesions. The limitation of the deep learning method is that, on the one hand, due to the limited abnormal data set of training and learning, there will be some deviation in the recognition ability of the model in practical application scenes. On the other hand, there are many abnormal types of the small intestine, so it is difficult to exhaust all the abnormal types in morphology and individual differences. Therefore, at present, most of this research work only recognizes one or several types of abnormal images.

Li et al. [13, 14] attempted to identify abnormal images of small intestinal capsule endoscopy by traditional machine learning methods. Although the method based on machine learning did not need a lot of training data, the recognition effect of the algorithm was not as good as that of deep learning.

At present, the methods based on traditional machine learning methods or deep learning techniques lesion recognition are facing the practical problem of designing and verifying a single lesion or multiple lesions. However, doctors must diagnose a wide range of small intestinal lesions, so the best image screening scheme for doctors should keep the integrity of intestinal mucosa as much as possible without missing any lesions. Therefore, the screening scheme based on similarity is superior to the abnormal lesion recognition scheme based on deep learning or machine learning.

In this study, the hierarchical clustering method [15, 16] was used to cluster the small intestinal capsule endoscopic images, and the visual saliency model [17–19] was introduced to identify the visually salient areas in small intestinal capsule endoscopic images. Based on the principle of structural similarity index (SSIM) [20], a keyframe image extraction method of small intestine capsule endoscopy based on hierarchical clustering and the structural similarity of visually significant sub-image blocks is proposed, which can keep continuous small intestine capsule endoscopy images for doctors to review without losing diseased images. Section III introduced the implementation theory of the algorithm. In Section IV, subjective evaluation and objective evaluation were used to evaluate and adjust the parameters of the algorithm. A total of 3,967,945 original images of 52 cases of small intestinal capsule endoscopy was used to verify the performance of the algorithm, covering 17 kinds of common abnormal lesions of the small intestine. The experimental results were discussed in Section V. Section VI was a summary of the study.

II. ALGORITHMIC THEORY

A. Block diagram of the algorithm

![Figure 1. Block diagram of the algorithm](image)

Figure 1. Block diagram of the algorithm (a) Input images; (b) clustering of similar small intestinal capsule endoscopic images based on agglomerative hierarchical clustering; (c) Saliency-SSIM; (d) Choose keyframe images)

It proposed a similar image screening method for wireless capsule endoscopy based on the structural similarity of visually significant sub-images and hierarchical clustering. Figure 1 is a block diagram of the algorithm presented in this study. The algorithm proposed in this study is mainly divided into two steps. The first part is the similar grouping of small intestine capsule endoscopy images based on the aggregation hierarchical clustering method. HSV color features of the whole image are extracted, and similar images are grouped by hierarchical clustering. The second part is the structure similarity keyframe extraction algorithm based on the visual saliency region (Saliency-SSIM). The visually significant sub-image blocks are extracted from the images in the group, the structural similarity of adjacent sub-image blocks is compared, the keyframe images are accurately identified, and other non-keyframe images are screened. The following sections will briefly introduce the steps and theory of the algorithm scheme in this paper.
B. Clustering of similar small intestinal capsule endoscopic images based on agglomerative hierarchical clustering

Algorithm  Clustering of similar small intestinal capsule endoscopic images based on agglomerative hierarchical clustering.

Input:  \( D = \{x_1, x_2, \ldots, x_n\} \), \( x_n \) was a single wireless capsule endoscopy image, and \( D \) was a wireless capsule endoscopy image data set.

\( T_1 \) was clustering distance parameter. Distance measure function \( d(x,y) \) was \( L_2 \) distance.

Out: Similar image grouping, \( S = \{s_1 [x_1, x_2, \ldots], s_2 [x_1, x_2, \ldots], \ldots, s_m [x_1, x_2, \ldots]\} \).

1) Extraction of color features

The image color space was transformed into hue, saturation, and value (HSV), histogram features were extracted from three channels in HSV space, and the values of each channel was quantized into ten intervals, the number of each interval was counted, the values of each interval was normalized, and finally 1*1000-dimensional color feature vector \( q \) was extracted.

2) Clustering of similar images

Every \( n \) small intestine capsule endoscope images in \( \{x_1, x_2, \ldots, x_N\} \) were divided into a group according to the time sequence and named as \( D = \{x_1, x_2, \ldots, x_n\} \), and the images in \( D \) are converted into HSV color space and color feature vectors \( Q = \{q_1, q_2, \ldots, q_n\} \) are extracted. The method comprised the following steps: firstly, taking \( n \) images in image group \( D \) as \( n \) categories, calculating \( L_2 \) distance between color feature vectors, merging two images with the smallest distance into the same category, taking each image as a leaf node of the distance Tree in the merging process, and then recalculating the distance between the merged category and all remaining categories; Repeat the above steps until all the images were merged into one category, and then the distance Tree is constructed; Mark the maximum distance from the Tree as \( D_{\text{max}} \), take the threshold \( T_1 \), and take the images whose distance was less than \( T_1 \times D_{\text{max}} \) in the tree as the similar image group.

C. Saliency-SSIM theory

Algorithm  Saliency-SSIM

Input: Similar image grouping, \( S = \{s_1 [x_1, x_2, \ldots], s_2 [x_1, x_2, \ldots], \ldots, s_m [x_1, x_2, \ldots]\} \).

Out: Keyframe images \( K = \{x_{k1}, x_{k2}, \ldots, x_{km}\} \).

**class saliency-pix (img)**

Transferred the image \( \text{img} \) to the LAB color space. After gaussian filtering, the three-channel LAB images were \( L_{\text{g}}, A_{\text{g}}, B_{\text{g}} \), respectively. The difference operation between each channel and the pre-filtering channel was done respectively. Combined three-channel values to get edge information image \( \text{Img}_g \).

Calculated the sum of the mean and standard deviation of \( \text{Img}_g \) as \( T_d \).

Using \( T_d \) as the threshold, \( \text{Img}_g \) was binarized to get \( \text{img}_\text{saliency} \).

**end**

**class saliency-area (img_saliency)**

The image was divided into sub-image blocks. Counted the sum of pixel values in corresponding \( \text{img}_\text{saliency} \) sub-image blocks.

Sorted the sum of the pixel values of the sub-image blocks, and take the three sub-image blocks with the largest pixel values, \( s_1, s_2, s_3 \).

**end**

**class saliency-SSIM (S [x_1,x_2,...,x_n], T_saliency-ssim)**

for \( i = 1, 2, \ldots, n \)

\( \text{im}_1 = x_i \)

\( \text{im}_2 = x_{i+1} \)

\( \text{img}_\text{saliency}_1 = \text{saliency-pix} (\text{im}_1) \)

\( \text{img}_\text{saliency}_2 = \text{saliency-pix} (\text{im}_2) \)

\( s_{i1}s_{i2}s_{i3} = \text{saliency-area} (\text{img}_\text{saliency}_1) \)

\( s_{i+1}s_{i+2}s_{i+3} = \text{saliency-area} (\text{img}_\text{saliency}_2) \)

\( s_{\text{im}1} = \text{merge} (s_{i1}, s_{i2}, s_{i3}) \)

\( s_{\text{im}2} = \text{merge} (s_{i+1}, s_{i+2}, s_{i+3}) \)

\( T = \text{mssim} (s_{\text{im}1}, s_{\text{im}2}) \)

if \( T > T_s\text{saliency-ssim} \)

K.append (\( \text{im}_1 \))

end if

end for

end
The three filtered channels are \( L_g, A_g, \) and \( B_g. \)

On the basis of the saliency mathematical model based on the Lab color model, the following equation was obtained:

\[
S(x, y) = ||I_x - I_{lab}(x, y)||
\]  \hspace{1cm} (2)

We calculated the significance of the Lab color model for \( L_g, A_g, \) and \( B_g, \) respectively:

\[
S_L = ||L - L_g|| \]  \hspace{1cm} (3)

\[
S_A = ||A - A_g|| \]  \hspace{1cm} (4)

\[
S_B = ||B - B_g|| \]  \hspace{1cm} (5)

The values of the three channels, i.e., \( S_L, S_A, \) and \( S_B, \) were combined to obtain the image edge information \( Img_g. \) The sum of the mean and standard deviation of \( Img_g \) was calculated and recorded as \( Td. \) With \( Td \) as the threshold, \( Img_g \) was binarized to obtain the visually salient pixels in the image. \( Img_g \) was divided into \( 40 \times 40 \)-pixel sub-image blocks denoted \( Img_{g1}, Img_{g2}, \ldots Img_{g36}. \) By counting the number of significant pixels in each sub-image block, the regional coordinates of the three sub-image blocks containing the largest number of the most significant pixels were extracted, i.e., \( s1(w1, h1, w2, h2), s2(w1, h1, w2, h2), \) and \( s3(w1, h1, w2, h2). \) Based on the regional coordinates of the sub-image blocks, the visually salient sub-image blocks of the original images \( img_i \) and \( img_{i+1} \) were extracted and denoted as \( s_{i1}, s_{i2}, s_{i3}, \) and \( s_{i1+1}, s_{i2+1}, s_{i3+1}, \) respectively (Figure 2).

![Figure 2. Visually salient sub-image blocks](image)

Evaluated the similarity value \( T \) between the visually salient sub-image blocks \( s_{i1} \) and \( s_{i2} \) of the current image \( img_i \) and the next adjacent image \( img_{i+1} \) of the same cluster group with reference to MSSIM algorithm. \( T \) was compared with the threshold \( T_{saliency-ssim}, \) and if \( T \) was larger, the current image \( img_i \) was taken as the keyframe of the cluster.

III. RESULTS

A. Small intestinal capsule endoscopic image test dataset

In this study, a WCE image test dataset including various types of small intestinal lesions (hereafter referred to as “the test set”) was established. Among the complete data of 52 cases gathered by the second affiliated Hospital of Army Medical University from January–December 2019 using a Type I OMOM Small Intestinal Capsule Endoscope produced by Chongqing Jinshan Science & Technology (Group) Co., Ltd., the doctors labeled all images with lesions, marked the sequence numbers of those images, and intercepted a total of 68,400 images in the lesion and the normal continuous video segment to form the test set. The test set comprised common types of small intestinal lesions, i.e., flat lesions-plaque(white), excavated lesion-ulcer, protruding lesions-venous structure, content-blood, protruding lesions-nodule, protruding lesions-Mass/tumor, protruding lesions-polyp(s), flat lesions-plaque(red), flat lesions-spot, excavated lesion-erosion, excavated lesion-aphtha, excavated lesion-diverticulum, mucosa-granular, mucosa-erythematous, mucosa-edematous, mucosa-pale, and content-parasites (classified according to the CEST classification standard [21]). Figure 3 shows images of various small intestinal lesions.

![Figure 3. CEST classification of the test set](image)
The similarity between images can be evaluated in two ways, i.e. algorithmic evaluation and subjective algorithmic evaluation. Accordingly, the image reduction rate based on image similarity principle can also be divided into the subjective image reduction rate and the algorithmic image reduction rate. The subjective image reduction rate is based on subjective judgment of image similarity by gastroenterologists. Similar images in the same slice of the test set were independently grouped by five doctors, and no lesions were missed after clustering. In this study, the average value of the subjective reduction rate created by the five doctors was used as the subjective reduction rate of the test set. The subjective reduction ratio (subjective_reduction_ratio) was defined and calculated as follows:

$$\text{subjective\_reduction\_ratio} = 1 - \sum \left( \frac{N_{\text{cluster}}}{N_{\text{total}}} \right)$$

where $N_{\text{cluster}}$ is the number of clusters of the test set selected by a single doctor, and $N_{\text{total}}$ is the total number of images in the test set.

Correspondingly, the reduction ratio of the algorithm for the test set was recorded as reduction_ratio, which was calculated as follows:

$$\text{reduction\_ratio} = 1 - \sum \left( \frac{N_{\text{key}}}{N_{\text{total}}} \right)$$

where $N_{\text{key}}$ is the number of keyframe images in the test set selected by this algorithm.

Each image in the test set was represented by a unique sequence number. The doctors selected the images with lesions in the test set and classified the images per the capsule endoscopy structured terminology (CEST) classification standard [21]. $A_b(\text{img}_1, \text{img}_2, \ldots, \text{img}_n)$ denotes the set of sequence numbers of images with the same type of lesions from the same case, $A_b(A_b_1, A_b_2, A_b_3, \ldots, A_b_k)$ represents the set of lesions in the test set, $K_n$ is the number of elements in the lesion set $A_b$, $S\{x_{k_1}, x_{k_2}, \ldots, x_{k_m}\}$ denotes the set of sequence numbers of the keyframe images selected by the algorithm, and $SD$ denotes the number of lesion categories included in the set $S$:

$$SD = \sum_{j=1}^{k} A_j, A_j = \begin{cases} 1 & \text{if } A_b \bigcap S \neq \phi \\ 0 & \text{else} \end{cases}$$

The abnormal recall was defined as:

$$\text{abnormal\_recall} = \frac{SD}{K_n}$$

$T_1$ and $T_{\text{saliency-ssim}}$ are parameters of the algorithm.

The HSV color features of the test set images were extracted to cluster the image data according to the agglomerative hierarchical clustering method, and the similarity between the images was evaluated by the clustering threshold $T_1$, and the images were divided into different groups. Saliency-SSIM keyframe images were extracted on the basis of clustering. The keyframe images in the test set were identified by using the threshold value $T_{\text{saliency-ssim}}$, and the abnormal recall rate and the similar image screening rate reduction_ratio are calculated. Figure 4 records the curve of abnormal_recall with image screening rate reduction_ratio in the test set with different parameters $T_1$ and $T_{\text{saliency-ssim}}$. The parameter $T_1$ ranged from 0.18 to 0.88, and the parameter $T_{\text{saliency-ssim}}$ ranged from 0.03 to 0.12.

Figure 4. Variation curves of image reduction ratio and abnormal recall with different algorithm parameters ($P_1: T_{\text{saliency-ssim}}=0.03; P_2: T_{\text{saliency-ssim}}=0.04; P_3: T_{\text{saliency-ssim}}=0.05; P_4: T_{\text{saliency-ssim}}=0.06; P_5: T_{\text{saliency-ssim}}=0.07; P_6: T_{\text{saliency-ssim}}=0.08; P_7: T_{\text{saliency-ssim}}=0.09; P_8: T_{\text{saliency-ssim}}=0.1; P_9: T_{\text{saliency-ssim}}=0.12$).

B. Results

As for the subjective evaluation conducted in this study, the five doctors independently divided the test set into 14,672, 13,556, 12,987, 13,376, and 14,390 clusters. The subjective_reduction_ratio of the test set was 79.83%, indicating that the test set could screen out 79.83% of the similar images per the subjective evaluation principle of similarity, with the abnormal recall being 100%.
The method proposed in this study screened out 66.72% of the redundant images without missing lesions, i.e., flat lesions-plaque(white), excavated lesion-ulcer, protruding lesions-venous structure, content-blood, protruding lesions-nodule, flat lesions-plaque(red), protruding lesions-Mass/tumor, mucosa-edematous, protruding lesions-polyp(s), flat lesions-spot, excavated lesion-erosion, excavated lesion-aphtha, excavated lesion-diverticulum, mucosa-granular, mucosa-pale, mucosa-edematous, and content-parasites. Of the redundant images, 79.09% were screened out without missing protruding lesions-venous structure, excavated lesion-aphtha, protruding lesions-polyp(s), flat lesions-plaque(red), mucosa-edematous, flat lesions-plaque(white), excavated lesion-erosion, excavated lesion-diverticulum, mucosa-granular, content-parasites, mucosa-pale, mucosa-edematous(congested), and content-blood. The current color structure similarity (CSS) method [5] has an image reduction rate of 93.87%, without missing the detection of bleeding focus. The motion-estimate method [7] can remove 68% of redundant images, although there is no description of lesion retention. The non-negative matrix factorization (NMF) method [2] has achieved 85% image reduction rate, and there are no missed diagnoses of ulcers, venous abnormalities, and white spots. Table 1 shows the comparison of the experimental results in this study and the current methods.

Table 1. Comparison of experimental results with those of current methods (*:Parameter1:T1=0.48,T_saliency-ssim=0.03;Parameter2:T1=0.88,T_saliency-ssim=0.03)

| Reduction Score(%) | Subjective work evaluation (Parameter) | CSS [Motion estimation] | NMF [2] |
|--------------------|----------------------------------------|-------------------------|---------|
| ER-rate            | 79.83                                  | 66.72                   | 79.09   |
| abnormal-recall    | 100                                    | 100                     | 94.23   |
| Flat lesions       | 100                                    | 100                     | 100     |
| Plaque(white)      |                                        |                         |         |
| Excavated lesion   | 100                                    | 100                     | 85.71   |
| Ulcer              |                                        |                         |         |
| Protruding lesions-| 100                                    | 100                     | 100     |
| venous structure   |                                        |                         |         |
| Content-Blood      | 100                                    | 100                     | 100     |
| Protruding lesions-| 100                                    | 100                     | 83.78   |
| Nodule             |                                        |                         |         |
| Protruding lesions-| 100                                    | 100                     | 92.86   |
| Mass/tumor         |                                        |                         |         |
| Flat lesions       | 100                                    | 100                     | 100     |
| Plaque(red)        |                                        |                         |         |
| Protruding lesions-| 100                                    | 100                     | 100     |
| Polyp(s)           |                                        |                         |         |
| Flat lesions-Spot  | 100                                    | 100                     | 100     |
| Excavated lesion   | 100                                    | 100                     | 100     |
| Erosion            |                                        |                         |         |

With parameter settings based on an abnormal recall of 100%, the proposed similar WCE image screening method based on structural similarity analysis and the hierarchical clustering of visually salient sub-image blocks was introduced into the “redundant image screening” function of the OMOM image workstation produced by Chongqing Jinshan Science & Technology (Group) Co., Ltd. With 4 pictures on 1 screen, the images were played at the speed of 10 frames/second. Parameter settings:T1=0.48,T_saliency-ssim=0.03.

Table 2 lists the average play time and image reduction ratio of all 52 cases in the test set before and after applying the screening of redundant images without missed diagnosis. The average playing time is 78 minutes before applying the redundant images filtering without missed diagnosis, but it is reduced to 18 minutes after applying the filtering. The average reduction rate of the similar image of the 52 cases reached 76%, and the results of 52 cases show that the average total number of images is 76,307, the average number of screened images is 18,096.

Table 2. Statistics of screened images of 52 cases without missed diagnosis in the test set (Parameter settings:T1=0.48,T_saliency-ssim=0.03, played at the speed of 10 frames/second, with 4 pictures per screen in the OMOM image workstation)

| Case No. | No. of original images | No. of keyframe images | Reduction ratio(%) | Original mode play time(minutes) | Redundant image screening mode play time(minutes) |
|----------|------------------------|------------------------|-------------------|---------------------------------|-----------------------------------------------|
| 1        | 50218                  | 15939                  | 68.3              | 51.03                           | 16.20                                         |
| 2        | 89824                  | 14213                  | 84.2              | 91.28                           | 14.44                                         |
| 3        | 76718                  | 14692                  | 80.8              | 77.96                           | 14.93                                         |
| 4        | 76447                  | 16991                  | 77.8              | 77.69                           | 17.27                                         |
| 5        | 72031                  | 19839                  | 72.5              | 73.20                           | 20.16                                         |
| 6        | 61464                  | 17289                  | 71.9              | 62.46                           | 17.57                                         |
| 7        | 105282                 | 22566                  | 78.6              | 106.99                          | 22.93                                         |
| 8        | 72209                  | 17138                  | 76.3              | 73.38                           | 17.42                                         |
| 9        | 17062                  | 4006                   | 76.5              | 17.34                           | 4.07                                          |
| 10       | 44260                  | 8580                   | 80.6              | 44.98                           | 8.72                                          |
| 11       | 86851                  | 22852                  | 73.7              | 88.26                           | 23.22                                         |
| 12       | 105453                 | 23856                  | 77.4              | 107.17                          | 24.24                                         |
| 13       | 47519                  | 11887                  | 75.0              | 48.29                           | 12.08                                         |

509
Therefore, the identification of visually significant sub-image block based on visually significant sub-image block value may help screen out the results of 52 cases show that the average total study was introduced into the “redundant image screening” to a certain extent, the workload of doctors in reading and redundant image screening mode of playing time is 18 minutes. The similar WCE image screening method proposed in this study was introduced into the “redundant image screening” function of the OMOM image workstation. With similar images screened out, the results of 52 cases show that the average total number of images is 76,307, the average number of screened images is 18,096, the average screening rate is 76%, the average original mode of playing time is 78 minutes, and the average redundant image screening mode of playing time is 18 minutes. To a certain extent, the workload of doctors in reading and subjective evaluation, the performance of the algorithm was evaluated based on its similarity reduction of a test set. In the objective evaluation, the performance of the algorithm was evaluated under different parameters, with the similar image reduction rate and abnormal recall rate as indicators. In this way, besides the abnormal recall of the algorithm when approaching the subjective reduction rate, the image reduction rate that the algorithm can achieve without missing any diagnosis of lesions can also be calculated. The experimental results show that when the abnormal recall rate in the test set was 100%, the reduction rate of the test set in this study was 79.83%. Therefore, the performance of the proposed algorithm was evaluated with 79.83% as the maximum achievable reduction rate. In the subjective evaluation, a subjective reduction rate was calculated based on the evaluation standard of gastroenterologists. The subjective reduction rate represents the maximum reduction rate based on image similarity to some extent. The subjective reduction rate of the test set in this study was 79.83%. Therefore, the performance of the proposed algorithm was evaluated with 79.83% as the maximum achievable reduction rate. In the objective evaluation, the performance of the algorithm was evaluated under different parameters, with the similar image reduction rate and abnormal recall rate as indicators. In this way, besides the abnormal recall of the algorithm when approaching the subjective reduction rate, the image reduction rate that the algorithm can achieve without missing any diagnosis of lesions can also be calculated. The experimental results show that when the abnormal recall rate in the test set was 100%, the reduction rate of similar image was 66.72%; when the similar image reduction ratio was 79.09%, which was close to the subjective reduction ratio, the abnormal recall was 94.23%.

The similar WCE image screening method proposed in this study was introduced into the “redundant image screening” function of the OMOM image workstation. With similar images screened out, the results of 52 cases show that the average total number of images is 76,307, the average number of screened images is 18,096, the average screening rate is 76%, the average original mode of playing time is 78 minutes, and the average redundant image screening mode of playing time is 18 minutes. To a certain extent, the workload of doctors in reading and

### IV. DISCUSSION

A similar image screening method for wireless capsule endoscope based on visually significant sub-image block structural similarity and hierarchical clustering was proposed. In this algorithm, images with certain similarities were divided into the same group by condensed hierarchical clustering, and then sub-image blocks with more subtle changes were identified by visual saliency method in the same group of data, thus achieving the purpose of identifying keyframes more accurately. On the other hand, through experiments, it was found that abnormal lesions were more likely to appear in the sub-image blocks of visually significant images. Therefore, the identification of visual significance sub-image block played an important role in better retention of abnormal lesions.

Secondly, a test set of WCE images with common small intestinal lesions, i.e., flat lesions-plaque(white), excavated lesion-ulcer, protruding lesions-venous structure, content-blood, protruding lesions-nodule, flat lesions-plaque(red), protruding lesions-Mass/tumor, mucosa-edematous, protruding lesions-polyp(s), flat lesions-spot, excavated lesion-erosion, excavated lesion-apthha, excavated lesion-diverticulum, mucosa-granular, mucosa-pale, mucosa-erythematous, and content-parasites, was established. Different from most studies which generally verify a single disease type at present, the test data set in this study can evaluate the performance of similar image reduction in abnormal recall rate more objectively and comprehensively.

In this study, the performance of the proposed algorithm, which is more comprehensive and objective than the currently available evaluation methods, was subjectively and objectively evaluated based on its similarity reduction of a test set. In the subjective evaluation, a subjective reduction rate was calculated based on the subjective reduction rate represents the maximum reduction rate based on image similarity to some extent. The subjective reduction rate of the test set in this study was 79.83%. Therefore, the performance of the proposed algorithm was evaluated with 79.83% as the maximum achievable reduction rate. In the objective evaluation, the performance of the algorithm was evaluated under different parameters, with the similar image reduction rate and abnormal recall rate as indicators. In this way, besides the abnormal recall of the algorithm when approaching the subjective reduction rate, the image reduction rate that the algorithm can achieve without missing any diagnosis of lesions can also be calculated. The experimental results show that when the abnormal recall rate in the test set was 100%, the reduction rate of similar image was 66.72%; when the similar image reduction ratio was 79.09%, which was close to the subjective reduction ratio, the abnormal recall was 94.23%.

The similar WCE image screening method proposed in this study was introduced into the “redundant image screening” function of the OMOM image workstation. With similar images screened out, the results of 52 cases show that the average total number of images is 76,307, the average number of screened images is 18,096, the average screening rate is 76%, the average original mode of playing time is 78 minutes, and the average redundant image screening mode of playing time is 18 minutes. To a certain extent, the workload of doctors in reading and
diagnosing endoscopic images of small intestine capsules is greatly reduced.

Limitations: Although the abnormal recall rate of the proposed algorithm is as high as 94.23% when approaching the subjective reduction rate, and 14 diagnostic abnormalities are not missed in the test set. The algorithm based on image similarity screening still has a low probability of missing lesions, so in practical application, the parameter setting with a smaller screening rate is adopted to reduce the risk of missing lesions. Balancing the relationship between them is a problem that needs to be further studied based on the similarity exclusion scheme. Targeted design features to identify abnormal images, further improve the image screening rate, and reduce the reading workload of doctors while maintaining the continuity of the small intestine capsule endoscopy image. In the future, we should analyze the missing lesions in a large-scale test set in order to design features to better identify those lesions that are easy to miss.

V. CONCLUSIONS

A method for screening similar WCE images was proposed in this study, based on structural similarity analysis and the hierarchical clustering of visually salient sub-image blocks. On the basis of clustering, the method of extracting keyframes with similar structure from visually significant sub-image blocks is adopted to identify keyframes in the image data set, so as to identify different images more accurately. When the image reduction rate was 66.72%, the abnormal recall rate of 17 abnormalities in the test set was 100%. An image reduction rate of 79.09% was close to the screening rate of the human eyes, and the abnormal recall rate of the reduced amount was 94.23%. The proposed algorithm was applied to the small intestinal capsule endoscopy imaging workstation developed by Chongqing Jinshan Science & Technology (Group) Co. Ltd. After testing 52 complete cases and screening out the redundant images with retaining an abnormal recall of 100%, the average play time was reduced from 78 minutes to 18 minutes, thereby effectively relieving the doctors' workload for viewing images and improving their work efficiency, enabling them to serve more patients, as well as better identifying and treating small intestinal lesions.

ACKNOWLEDGMENT

Thanks for the translation work of LetPub.

REFERENCES

[1] Spyros Tsevas, Dimitrios K. Iakovidis, Dimitrios E. Maroulis, Emmanuel Pavlakis. Automatic frame reduction of wireless capsule endoscopy video. 8th IEEE International Conference on Bioinformatics and BioEngineering. (2008). doi:10.1109/bIBE.2008.4696805

[2] Dimitris K.Iakovidis, Spyros Tsevas, Andreas Polydorou. Reduction of capsule endoscopy reading times by unsupervised image mining. Computerized Medical Imaging and Graphics (2010) 34(6): 471-478. doi:10.1016/j.compmedimag.2009.11.005

[3] Hyun-Gyu Lee, Min-Kook Choi, Sang-Chul Lee. Motion analysis for duplicate frame removal in wireless capsule endoscopy. Progress in Biomedical Optics and Imaging - Proceedings of SPIE. (2011). doi:10.1117/12.877627

[4] Hyun-Gyu Lee, Min-Kook Choi, Byoung-Seok Shin, Sang-Chul Lee. Reducing redundancy in wireless capsule endoscopy videos. Computers in Biology and Medicine (2013). 43(6): 670-682. doi:10.1117/12.877627

[5] Qasim Al-sheban, Prashan Premaratne, Darryl J. McAndrew, Peter J.Vial, Shehan Abey. A frame reduction system based on a color structural similarity (CSS) method and Bayer images analysis for capsule endoscopy. Artificial Intelligence in Medicine (2019).94: 18-27. doi:10.1016/j.artmed.2018.12.008

[6] Alexandros Karargyris,Nikolaos Bourbakis. A video-frame based registration using segmentation and graph connectivity for wireless capsule endoscopy. IEEE/NIIH Life Science Systems and Applications Workshop (LiSSA 2009). doi:10.1109/LISSA.2009.4906713

[7] Hong Liu, Ning Pan, Heng Lu, Eminn Song , Qian Wang , Chih-Cheng Hung. Wireless capsule endoscopy video reduction based on camera motion estimation. J Digit Imaging (2013). 26(2): 287-301. doi:10.1007/s10278-012-9519-x

[8] Xiao Jia, Meng MQ. A deep convolutional neural network for bleeding detection in Wireless Capsule Endoscopy images. Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). (2016).38th doi:10.1109/embc.2016.7597583

[9] Aoki, Tomonori Yamada, Atsuo Aoyama, Kazuharu Saito, Hiroaki Tsuiboi, Akiyoshi Nakada, Ayako Niikura, Ryota Fujishiro, Mitsuhiro Oka, Shiro Ishihara, Soichiro Matsuda, Tomoki Tanaka, Shinji Koike, Kazuhiro Tada, Tomohiro. Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. Gastrointestinal Endoscopy.2018. doi:10.1016/j.gie.2018.10.027

[10] Sekuboyina, A. K, Devarakonda. A convolutional neural network approach for abnormality detection in Wireless Capsule Endoscopy. 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). doi:10.1109/isbi.2017.7950698

[11] Sharif, Muhammad, Attique Khan, Muhammad. Rashid. Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images. Journal of Experimental & Theoretical Artificial Intelligence, (2019).1–23. doi:10.1080/0952813X.2019.1572657

[12] Xiao Jia, Meng MQ. Gastrointestinal bleeding detection in wireless capsule endoscopy images using handcrafted and CNN features. Annual International Conference of the IEEE Engineering in Medicine and Biology Society.(2017).39th doi:10.1109/embc.2017.8037526

[13] Yuan, Yixuan Li, Baopu Meng, Qinghu Bu. Bleeding frame and region detection in the wireless capsule endoscopy video. IEEE Journal of Biomedical and Health Informatics. (2015). 20(2):624–630. doi:10.1109/jbhi.2015.2539502

[14] Baopu Li, Max Q.-H. Meng. Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments. Computers in Biology and Medicine, (2009).39(2), 141–147. doi:10.1016/j.compmed.2008.11.007

[15] Zhihua Zhou. Machine Learning [M].(2016).197-217

[16] Leonard Kaufman , Peter J Rousseeuw. Finding groups in data: an introduction to cluster analysis. Biometrics (1991). 47(2):788. doi:10.2307/2532178

[17] Laurent Itti, Christof Kohl, Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, (1998).20(11):1254-1259. doi:10.1109/34.730558

[18] Itti L, Baldi P. Bayesian surprise attracts human attention[J]. Vision Research, (2009)49(10): 1295-1306. doi:10.1016/j.visionresearch.2008.09.007

[19] Achanta R, Hemami S, Estrada F, et al. Frequency-tuned salient region detection[C]/ IEEE International Conference on Computer Vision and Pattern Recognition, Miami, (2009):1597-1604. doi:10.1109/cvprw.2009.526596

[20] Zhou Wang, Alan Conrad Bovik, Alan C. Lim, Guohua Zhou. Image quality qssessment: from erroremeasurements to subjective reduction. IEEE TRANSACTIONS ON IMAGE PROCESSING. (2004). 13. NO. 4, APRIL 2004:600-611. doi:10.1109/tip.2003.819861

[21] L. Y. Korman,M. Delvaux,G. Gay, F. Hagenmuller,M. Keuchel,S. Friedman,M. Weinstein,M. Shetzline,D. Cave,R. de Francis?is
Capsule Endoscopy Standard Working Group, Capsule Endoscopy Structured Terminology (CEST): Proposal of a Standardized and Structured Terminology for Reporting Capsule Endoscopy Procedures, *Endoscopy* (2005). 37 (10): 951-959. doi:10.1055/s-2005-870329