Heliyon 9 (2023) e22662

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

CCRA: A colon cleanliness rating algorithm based on colonoscopy video analysis☆

Yu Bo, Shao Wei *, Yao Dengju, Wang Yunhao, Zhang Heyi

School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China

ARTICLE INFO

Keywords:
Deep learning
Colon cleanliness
Bowel preparation quality
U-net

ABSTRACT

Objective: A Colon Cleanliness Rating Algorithm (CCRA) based on colonoscopy image analysis is proposed in this paper, in order to solve the problem that the results of Colon Cleanliness (or Bowel Preparation Quality) rating caused by manual inspection are inconsistent.
Methods: Firstly, CCRA intercepts images from the colonoscopy video. Secondly, each colonoscopy image's stool area is segmented by U-Net to obtain the 2-classification segmentation results. Finally, the colon cleanliness is obtained by comparing the average area of the stool area with the standard proportion.
Results: After testing, the pixel accuracy of the U-Net model is 97.02 %, IoU is 83.67 %, accuracy is 92.17 %, recall is 90.21 %, F1-Score is 90.95 %. The accuracy of CCRA is 92.45 %–99.275 %
Conclusion: The experimental results show that the CCRA proposed in this paper can quickly and accurately output the colon cleanliness rating of patients without manpower.

1. Introduction

Colon Cleanliness (or bowel preparation quality) is an indicator used to evaluate the level of cleanliness of a patient's intestines, which mostly depends on the degree of feces or food residue in the patient's intestines [1,2]. Colon Cleanliness is often used in the preparation for colonoscopy, and to some extent, determines whether the subsequent examination will be successful. If there is too much fecal residue in the patient's intestines, it is likely to cause problems such as the coverage of the lesion area or obstruction of the endoscopic lens, thereby severely affecting the efficiency and quality of the examination [1–3]. Therefore, colon cleanliness testing has become an essential part of colonoscopy diagnosis and treatment. Currently, one of the mainstream methods for rating colon cleanliness is through manual inspection by clinical endoscopists. The clinicians evaluate the score by observing the images transmitted by the endoscope and combining them with the Boston Bowel Preparation Scale [4]. Then, according to the standard of colon cleanliness, the results are classified into Levels I, II, III, and IV. Level I indicates that the intestinal cavity is clean with no fecal water or only a small amount of clear liquid, and the view is very clear. Level II indicates that there is a small amount of feces in the intestinal cavity, and the view is relatively clear. Level III indicates that there are more fecal residues adhering to the intestines, and the view is relatively blurry. Level IV means that there is a large amount of feces attached to the intestinal cavity, and the next step of colonoscopy cannot be performed [5,6].

☆ Bo Yu reports was provided by The Natural Science Foundation of Heilongjiang Province. Bo Yu reports was provided by The National Natural Science Foundation of China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

* Corresponding author.
E-mail address: davidblue@foxmail.com (S. Wei).

https://doi.org/10.1016/j.heliyon.2023.e22662
Received 24 May 2023; Received in revised form 15 November 2023; Accepted 16 November 2023
Available online 18 November 2023

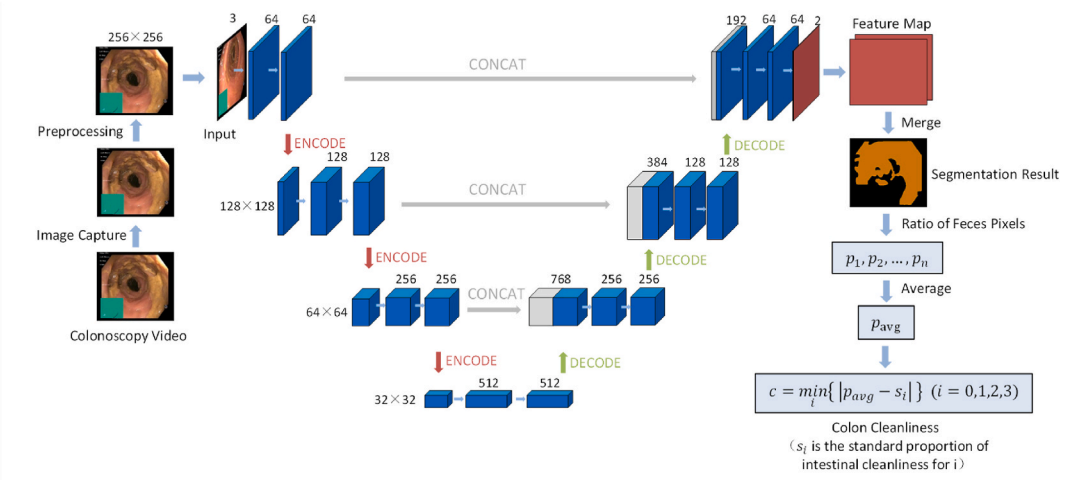

Fig. 1. Ccra process.

Manual evaluation of colon cleanliness has the advantage that clinicians can flexibly evaluate according to their rich clinical experience. However, the disadvantage is that the evaluation results are highly subjective and are influenced to some extent by the clinician's experience and diagnostic level. Therefore, different observers may obtain different evaluation results, and even the same observer may obtain different results after two observations, which affects the uniqueness and credibility of the evaluation results [7]. According to the work of Gao Yan et al. [8] on the reliability assessment of the Boston bowel preparation scale, the within-group correlation coefficient for scoring between different observers was 0.987, and the within-group correlation coefficient for scoring twice by the same observer was 0.713, which indicates the occurrence of the above situation. At the same time, the manual examination also brings more workload and learning burden to clinicians. Therefore, using deep learning algorithms to rate colon cleanliness is a solution to the above problems. Automatic evaluation of colon cleanliness by computers can help overcome problems such as insufficient clinical experience of doctors and subjective evaluation results and optimize the use of medical resources. This article proposes the CCRA, which can automatically obtain colon cleanliness based on colonoscopy videos to alleviate the burden on clinicians and overcome the problem of subjective evaluation results influenced by clinicians. By testing the algorithm on the Nerthus dataset, the evaluation accuracy reached 92.45 %–99.275 %, proving that the proposed algorithm can be used for colon cleanliness evaluation.

This article summarizes its contributions into three points: (1) Proposed a computer vision-based colon cleanliness rating algorithm. The automation of rating can alleviate the burden on clinicians, as well as overcome the problem of subjective evaluation results influenced by clinicians. To our knowledge, this is the first complete rating algorithm that targets colon cleanliness using colonoscopy image analysis in the computer vision field, providing a reference for future researchers who study the automatic evaluation of colon cleanliness. (2) Proposed data augmentation methods specific to colonoscopy images, providing helpful references to other researchers studying colonoscopy image-related problems in the computer vision field. (3) Proved through experiments on the public dataset Nerthus that there is a strong correlation between colon cleanliness and the ratio of feces in the colonoscopy image area.

2. Related work

Currently, there are plenty of research and applications combining deep learning with colonoscopic image analysis, especially in the detection of colorectal polyps, which has gradually become mature. The combination of medical care and artificial intelligence has become a trend for the future [9]. DONG Xu et al. [10] used 5 popular networks, VGG16, ResNet 18, GoogleNet, EfficientNet, and SeNet for binary classification of colonoscopy polyp images and achieved the highest accuracy of 99.84 % using GoogleNet. Yu Jieyao [11] proposed a grouped fully convolutional dense network by adding dense block strategies and grouping convolution methods to existing deep learning networks. Using this network to segment polyp images can reduce model parameters while ensuring segmentation accuracy. Souaidi et al. [12] first encoded the small polyp regions at multiple scales, then extracted features using some layers of the Inception v4 network, and finally fed the feature maps to a multi-box detector, resulting in an mAP of 93.29 %.

However, there is little research on combining deep learning with colon cleanliness. Through relevant literature review, only Ayimukedisi et al. [13] were found to rate colon cleanliness before classifying polyp images. They used an 8-layer convolutional neural network to perform a 4-class classification of colonoscopy images, with the classification results corresponding to the 4 levels of colon cleanliness standards. However, they only achieved an accuracy of 74.67 %. Additionally, relying on a single image to obtain colon cleanliness has a high probability of contingency, and the results of a single image cannot represent true colon cleanliness.

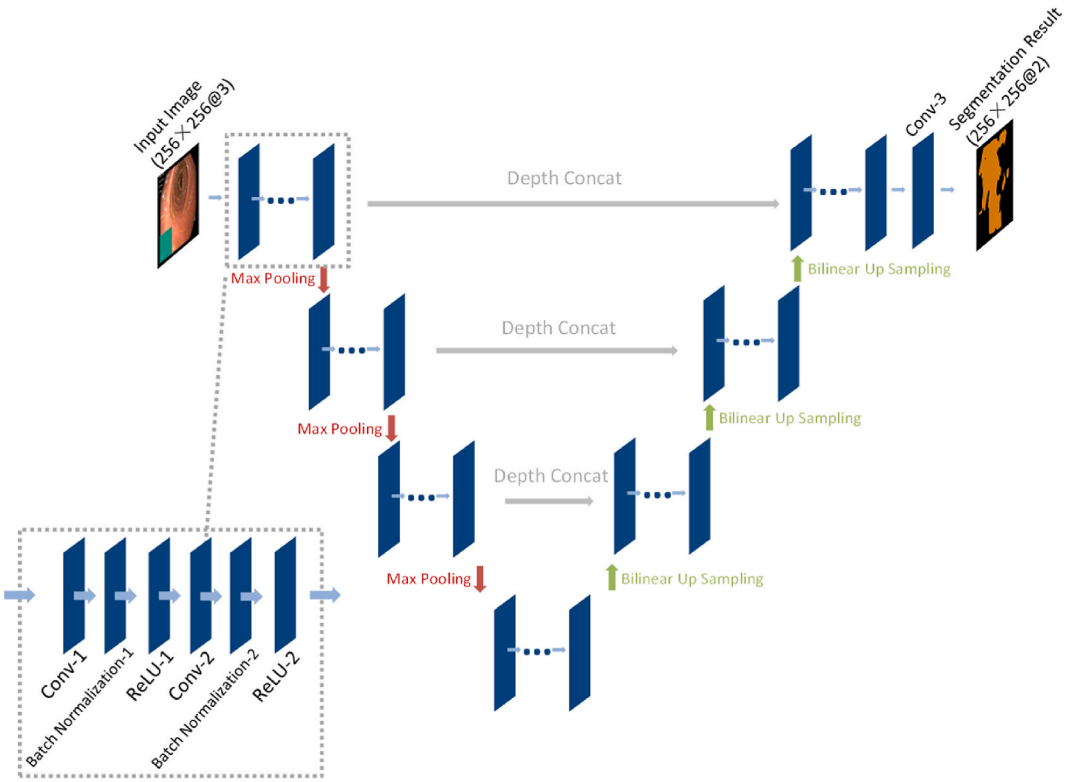

Fig. 2. The U-Net model structure used by CCRA.

3. Proposed method

3.1. CCRA process

Aiming at the problem of inconsistent cleanliness rating results caused by manual inspection of colon cleanliness, we propose the CCRA based on colonoscopy image analysis. The overall process of the algorithm is shown in Fig. 1.

CCRA consists of 5 steps. The input of the algorithm is a colonoscopy video of a broad region of a colon, and the output is an integer c, which can have values of 0, 1, 2, and 3, indicating the colon cleanliness according to the Boston scoring standard. In step 1, the video is uniformly intercepted in the time dimension to obtain several images. In step 2, preprocess the image. In step 3, input each image into the U-Net model for feces region segmentation, and obtain the segmentation result of each image. In step 4, obtain the average value of the feces area proportion. In step 5, compare with the standard proportion s (the reference value of s is given in section 4.5), and then output the colon cleanliness.

3.2. Image capture

This section elaborates on step 1 described in 3.1. For input video, if each frame of the image in the video is processed, the algorithm will take too long, which is not conducive to clinical application. Therefore, it is necessary to capture screenshots and use images to represent the entire colon. Considering the high similarity of adjacent frame images in colonoscopy videos, we adopt a strategy of evenly capturing images in the temporal dimension, believing that each image represents the image of its neighboring frame. We presented experimental results for four different capture frequencies in section 4.5, including detection time and accuracy. Based on these results, we recommend capturing 10 images per video of a broad region of a colon.

3.3. Image preprocessing

This section elaborates on step 2 described in 3.1. To achieve image standardization and enhance fine details, this step involves employing Laplacian filtering for image sharpening and subsequently resizing the image's dimensions to 256x256.

3.4. Image segmentation based on U-net

This section elaborates on step 3 described in 3.1. In order to obtain the area of the feces area, it is necessary to segment the feces

area in the image. According to the technical requirements of detection efficiency and comparative experiments, we chose the U-Net model for image segmentation. U-Net is a convolutional neural network commonly used for image segmentation. It is now widely used in the medical image segmentation field. It was first proposed by Olaf Ronneberger et al., in 2015 [14]. They improved on the Fully Convolutional Networks (FCN) by changing the network to an encoder-decoder structure. During the forward propagation of the network, it can refer to features obtained from different convolutional operations, allowing it to combine context information during the prediction process.

The U-Net model structure used by CCRA is shown in Fig. 2. The network has a total of 44 layers and 38,188,864 parameters that need to be trained. The top-left of the figure shows the input layer, which takes a 3-channel RGB image of size 256 × 256. The network first needs to complete the encoding part on the left, which consists of four combination structures and three max-pooling layers. The max-pooling layers are responsible for downsampling the feature maps. The components of the structure are shown in the gray dashed box on the bottom-left of the figure, which contains six layers. Conv-1 and Conv-2 are two convolutional layers responsible for extracting features from the colonoscopy images. ReLU-1 and ReLU-2 are two nonlinear activation layers that add nonlinear relationships during the data propagation process. Batch Normalization-1 and Batch Normalization-2 are two batch normalization layers that disperse the input data, preventing the output of the ReLU layers from being centered around 0 and thus avoiding the problem of gradient disappearance. The decoding part consists of three combination structures, three bilinear interpolation layers, and one Conv-3 layer. The bilinear interpolation layers are responsible for upsampling the feature maps. The combination structures and decoding part work in the same way as the encoding part. The Conv-3 layer is the layer immediately before the output layer. It uses two 1x1 convolutional kernels to convert the number of feature maps to 2. After decoding, the output goes to the output layer, which contains two 256x256 feature maps. The values in the first feature map represent the likelihood scores of the corresponding pixels being in the fecal area, while the values in the second feature map represent the opposite. It is also worth noting that in the combination structure of the first layer (topmost) of the figure, the number of convolutional kernels in each layer is 64, in the second layer it is 128, in the third layer it is 256, and in the fourth layer it is 512.

3.5. Determination of colon cleanliness based on segmentation results

This section elaborates on steps four and five described in section 3.1. After obtaining the segmentation result of the colonoscopy image, further analysis is needed to determine the level of bowel cleanliness. This section receives the two feature maps obtained in step 3. First, the two feature maps are combined and converted into a binary image. For each pixel, a value of 1 indicates that the pixel belongs to the fecal area, and 0 indicates the opposite. The proportion of fecal area is then calculated for each image, and the average is taken. This value is compared with the standard proportion, and the level of bowel cleanliness that corresponds to the closest standard proportion value is chosen as the final result. This process can be written as follows:

$$c = \min_i \left\{ \left| \frac{\sum_{j=0}^{n-1} \frac{c_{j,1}}{c_{j,0}+c_{j,1}}}{n} - s_i \right| \right\} (i = 0, 1, 2, 3) \tag{1}$$

In Eq. (1), c represents the level of bowel cleanliness, $c_{j,1}$ represents the number of pixels with a value of 1 in the j-th binary image, $c_{j,0}$ represents the number of pixels with a value of 0 in the j-th binary image, n represents the total number of binary images, s_i represents the standard proportion value that corresponds to the i-th level of bowel cleanliness, the computation of s_i is shown in Eq. (2):

$$s_i = \frac{\sum_{j=0}^{N_i-1} \frac{d_{i,j,1}}{d_{i,j,0}+d_{i,j,1}}}{N_i} (i = 0, 1, 2, 3) \tag{2}$$

In Eq. (2), $d_{i,j,0}$ represents the number of pixels with a value of 0 in the ground truth of intestinal cleanliness at level i for the j-th image. $d_{i,j,1}$ represents the number of pixels with a value of 1 in the ground truth of intestinal cleanliness at level i for the j-th image , N_i represents the number of images in the dataset with the intestinal cleanliness level i.

4. Experiment and analysis

4.1. Data labeling

We utilize the publicly available dataset, Nerthus [15], as training data. The dataset comprises 5525 colonoscopy images, which are frames extracted from colonoscopy videos. Each image is a 720 × 576 RGB image and is annotated with the Boston Bowel Preparation Scale (BBPS). However, since the dataset lacks the ground truth required for image segmentation tasks as training benchmarks, the corresponding fecal regions for each image need to be manually labeled by humans.

Given the substantial resemblance between neighboring frames in the dataset, and similar images do not significantly contribute to the model training process, we refined the dataset as follows: for images derived from the same video source, we retained 25 % of the images using systematic sampling. Then, we eliminate the blurry images from the dataset. Next, we used labelme (version 5.0.2) as the label tool to manually label the remaining colonoscopy images. Our label process was conducted as follows (excluding the interface

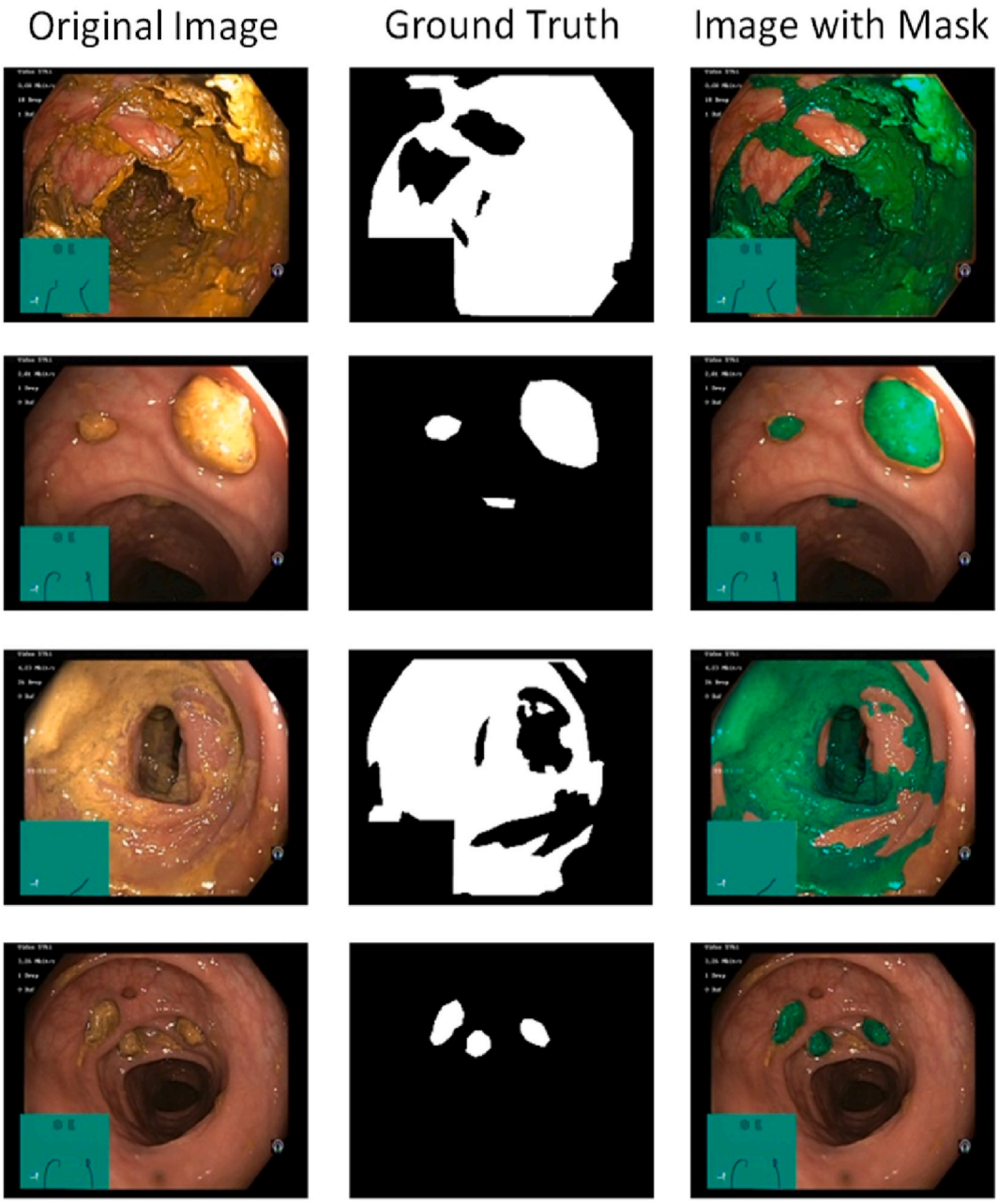

Fig. 3. Partially manually labeled results of the dataset.

parts of the images): (1) For unclear areas caused by insufficient lighting or excessive exposure, they were considered non-fecal regions. (2) For areas obstructed by opaque fluids, they were considered fecal regions. (3) For areas outside of (1) and (2), if the colon mucosa was clearly visible without residual staining, they were considered non-fecal regions; otherwise, they were considered fecal regions. The resulting binary images generated from the JSON label results were utilized as the training benchmarks. A total of 1070 useable images were ultimately labeled for training, including 36 images without any target regions. Part of the labeled results for some images is illustrated in Fig. 3.

4.2. Data augmentation

During the training process of the network model, the model learns the common features of the training data. However, some of these features may not be what we want the model to learn, because they are not reflected in all or most of the data. This can seriously affect the model's generalization ability, and this problem is called overfitting. For this specific dataset, the images are all captured by

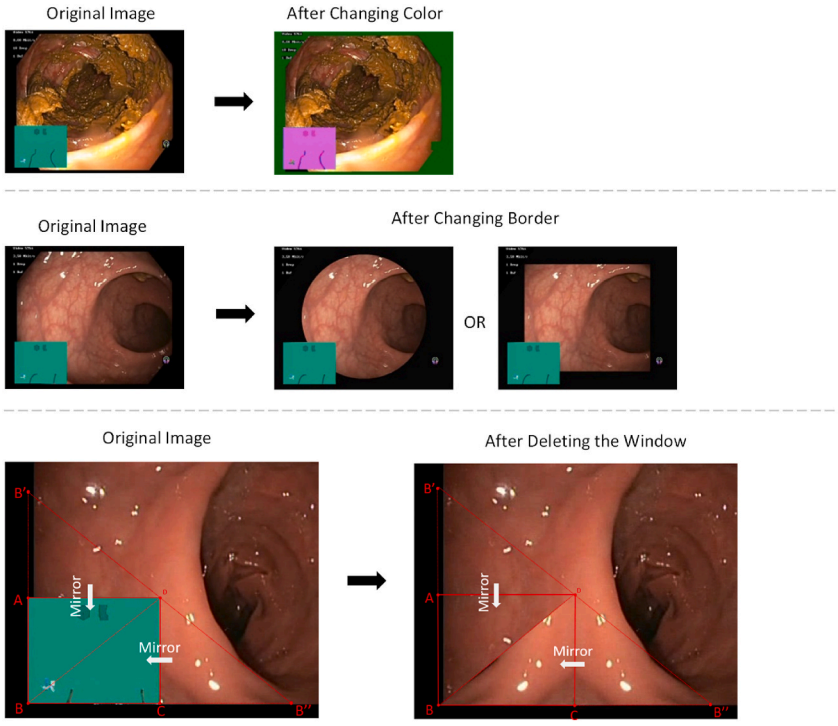

Fig. 4. Shape and color change of the interface.

the same colonoscope instrument, so the colonoscopy interface in the training set images is almost identical. This type of data is likely to cause overfitting in the interface part of the model, resulting in poor generalization capability of the trained model on other colonoscopy interface styles.

Data augmentation techniques are a commonly used solution to address overfitting problems. It refers to a method of making limited modifications to the training data to expand the training set [16]. Inspired by Refs. [17–19], we propose three pre-processing operations for the colon interface part: (1) color change of the interface (2) shape change of the interface (3) deletion of the small windows. Each of the above three changes is applied to the images in every batch of training. To ensure that 50 % of the training images are original images, the same probability value p is set for the three changes, which needs to satisfy the following equation:

$$(1-p)^n = 50\% \tag{3}$$

Where n is the number of times the change occurs, when n equals 3, it can be calculated that p is approximately 0.2063, which means that the probability of change for each occurrence is 20.63 %.

For the changes in interface color, we divide the interface into two parts for processing. The first part is the black interface around the image, and the second part is the green window in the lower left corner. Random color changes were applied to each part separately. The resulting image after the color changes is shown at the top of Fig. 4.

For the changes in interface shape, we have applied shape changes to the black interface border by filling black pixels to transform the interface into a square or circle shape. The resulting image after the shape changes is shown in the middle of Fig. 4.

"Deleting small windows" refers to removing the green rectangular region located in the lower right corner of the original image. To achieve the effect of deletion, the window area needs to be covered based on the surrounding pixels. In this study, we used a mirror-covering approach, as shown in the lower part of Fig. 4.

Regarding the removal of small window changes, a detailed explanation is required. As shown in the lower part of Fig. 4, first, the four vertices of the rectangular window are determined to correspond to points A, B, C, and D. Then, a diagonal line is drawn from the bottom left corner point B to the upper right corner point D, dividing the window into the upper triangle part (△ABD) and the lower triangle part (△BCD). Next, symmetrical triangles (△AB'D and △CB''D) are constructed using the upper boundary (AD) and the right boundary (CD) of the window as the axes of symmetry for the upper and lower triangles, respectively. The pixels in the triangle region are then mirrored and overlaid onto the small window position using the upper and right boundaries as axes of symmetry, achieving the effect of removing the small window.

Due to the possibility that this operation could lead the model to learn incorrect features, we conducted a comparative experiment, and the results showed that the model's performance did not deteriorate as a result.

Table 1
distribution of images in 3 datasets.

BBPS	train dataset	validation dataset	test dataset
I	114	14	14
II	372	46	46
III	294	37	37
IV	76	10	10

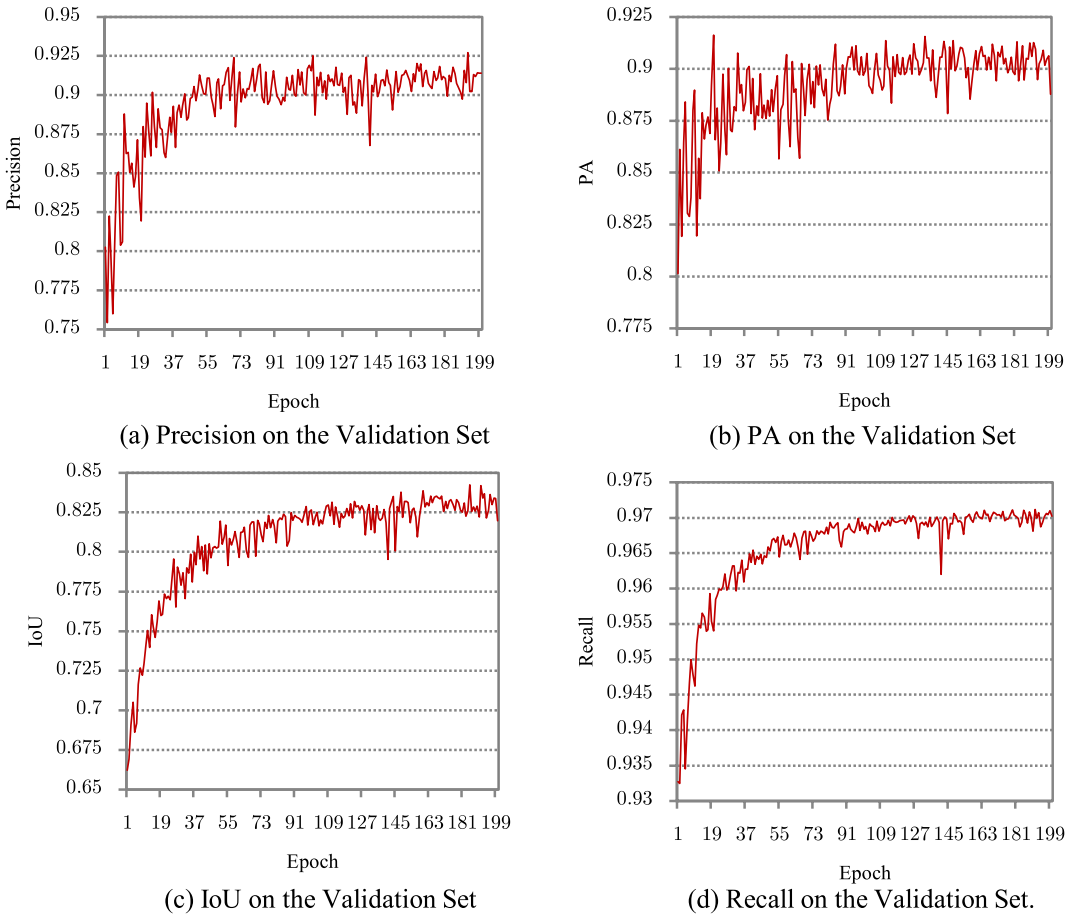

Fig. 5. Variations in various metrics during the training process of U-Net.

4.3. U-net model training and statistical indicators

The 1070 samples described in Section 4.1 were used as the data for training, validation, and testing. These samples were randomly divided into training, validation, and testing sets, with proportions of 80 %, 10 %, and 10 %, respectively. During the process of dividing the dataset, to avoid excessive similarity between the validation/test set and the training set, we ensure that the images in the same dataset are from consecutive frames, rather than using random sampling methods. The final training set consisted of 856 samples, and the validation and testing sets consisted of 214 samples each. The proportions of images corresponding to each BBPS level are approximately the same across the 3 datasets, Table 1 presents the number of images corresponding to each BBPS level in each dataset. It should be noted that in order to ensure the inclusion of 4 BBPS in each set, there may be a small percentage (not exceeding 10 %) of samples from the same video source distributed across different sets. Consequently, this could potentially introduce slight biases in the evaluation metrics of the segmentation model.

The GPU used for training was the NVIDIA GeForce RTX 3070 Laptop GPU, with 16 GB of memory and 5120 CUDA cores.

We employ 5 common statistical indicators in the field of medical image segmentation to assess the performance of the model. The following are their respective definitions and explanations.

Pixel Accuracy (PA) primarily focuses on the correctness of pixel predictions. PA is calculated using the formula shown in Eq. (4).

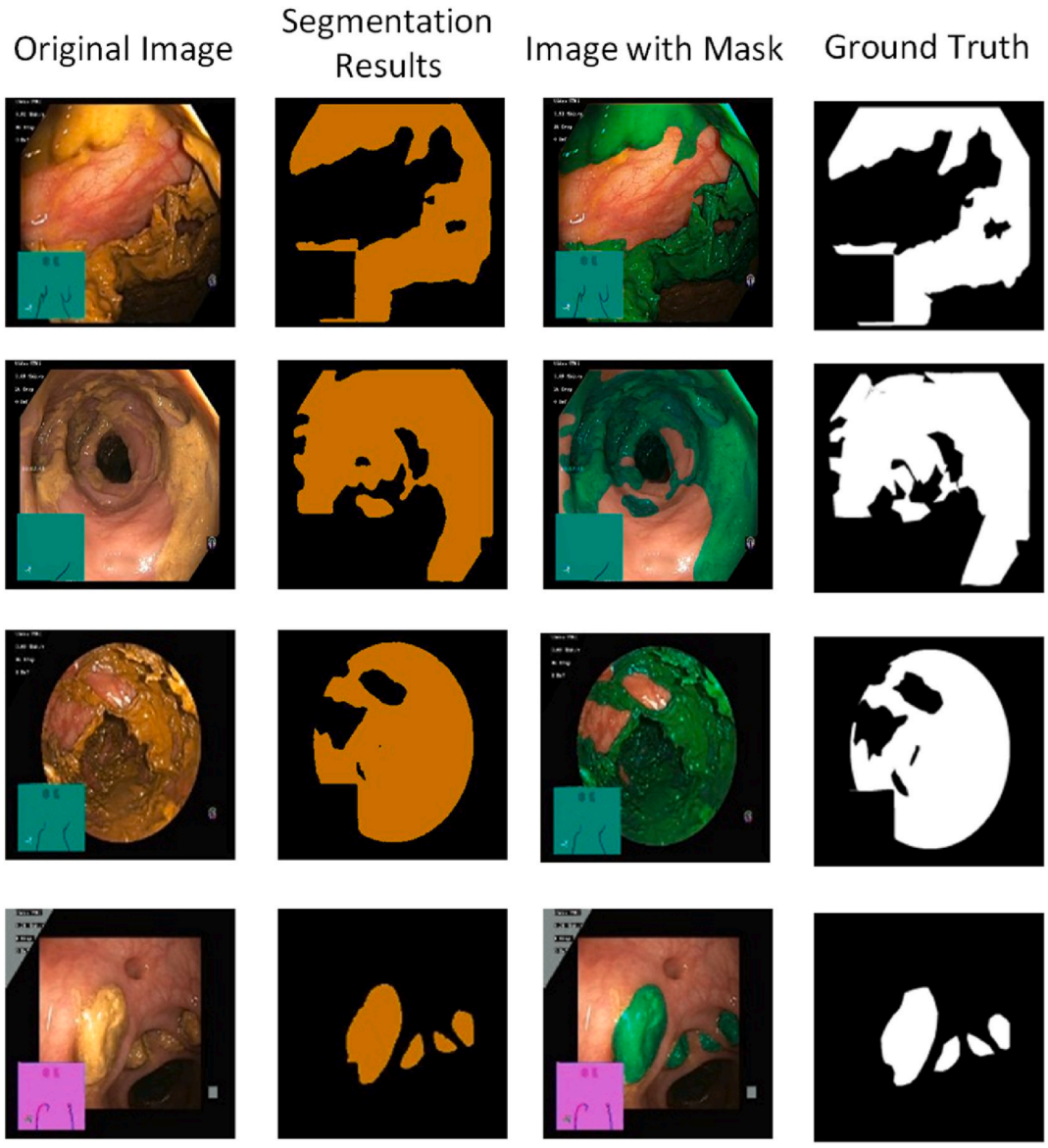

Fig. 6. Segmentation results of colonoscopy images.

$$\mathrm{PA}(\mathrm{X},\mathrm{Y}) = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

Intersection over Union (IoU) evaluates the model by measuring the ratio of overlapping pixels between the predicted region and the ground truth region, divided by the union of the two. IoU is calculated using the formula shown in Eq. (5).

$$\mathrm{IoU}(X,Y) = \frac{X \cap Y}{X \cup Y} = \frac{TP}{TP + FP + FN} \tag{5}$$

Precision represents the proportion of correctly predicted positive class pixels among all predicted positive class pixels. It indicates the percentage of correctly predicted positive class pixels. Precision is calculated using the formula shown in Eq. (6).

$$\mathrm{Precision}(X,Y) = \frac{TP}{FP + TP} \tag{6}$$

Recall represents the proportion of correctly predicted positive class pixels among the ground truth positive class pixels. It indicates the percentage of positive class pixels that are detected. Recall is calculated using the formula shown in Eq. (7).

Table 2
Segmentation performance of different models.

Model	PA	IoU	Precision	Recall Rate	F1-Score
U-Net	0.9702	**0.8367**	**0.9217**	0.9021	**0.9118**
deeplabv3	0.9706	0.8360	0.9090	**0.9146**	**0.9118**
FCN	**0.9717**	0.8343	0.9155	0.9042	0.9098
LR-ASPP	0.9645	0.8029	0.8936	0.8883	0.8909

Table 3
Calculation results of the Pearson correlation coefficient.

The i-th calculation	1	2	3	4	5	6	7	8	9	10
$\rho_{x,y}$	−0.9827	−0.9630	−0.9585	−0.9590	−0.9444	−0.9649	−0.9672	−0.9707	−0.9547	−0.9612

The results demonstrate that the coefficients of the 10 calculated values are all distributed between −0.9 and −1.0, indicating a strong negative correlation between the fecal area ratio and colon cleanliness on the dataset. Therefore, colon cleanliness can be estimated based on the fecal area ratio.

$$\mathrm{Recall}(X,Y)=\frac{TP}{FN+TP} \tag{7}$$

F1-Score is a commonly used evaluation metric for binary classification models. It is the result of the harmonic mean of Precision and Recall, adjusted and combined. F1-Score is calculated using the formula shown in Eq. (8).

$$\mathrm{F}_1(X,Y)=\frac{2\times\mathrm{Precision}(X,Y)\times\mathrm{Recall}(X,Y)}{\mathrm{Precision}(X,Y)+\mathrm{Recall}(X,Y)} \tag{8}$$

The model hyperparameters were set as follows: the training epoch was 200, the initial learning rate was set to 0.0001, the batch size was set to 4 to prevent memory overflow during training, and the L2 regularization coefficient was set to 0.0002. In addition, data augmentation was used to improve the model's generalization ability. We incorporated data augmentation methods into the image preprocessing stage, where the images were resized, rotated (at 90-degree intervals with a 50 % probability), and flipped (with a 30 % probability for both vertical and horizontal flipping) before each mini-batch training. In addition, the data augmentation methods proposed in Section 3.2 for colonoscopy images were also applied in the preprocessing stage.

The U-Net was then trained, and the variations of PA, IoU, Precision, and Recall of the model during the training process on the validation set are shown in Fig. 5.

Fig. 5 shows that the U-Net model's various metrics on the validation set converge to a certain range, indicating that the model can stop training. Fig. 5(a) shows the variation curve of model Precision, Fig. 5(b) shows the variation curve of model PA, Fig. 5(c) shows the variation curve of model IoU, and Fig. 5(d) shows the variation curve of model Recall. All of the above metrics are the results of the model tested on the validation set. The trained model was then tested on the test set, and metrics including PA, IoU, precision, recall, and F1-Score were calculated: PA is 0.9702, IoU is 0.8367, Precision is 0.9217, Recall is 0.9021, and F1-Score is 0.9095.

The result above demonstrates that on average, 97.02 % of pixels are correctly classified, and an IoU of 83.67 % is achieved, while ensuring a high F1-Score. These results demonstrate the good segmentation performance and generalization ability of U-Net on the given problem, indicating that the model is capable of completing the task of fecal area segmentation. The trained model was then applied to four selected images (pre-processed for interface variation), and the results are shown in Fig. 6.

In the process of selecting image segmentation models, we trained several commonly used models, and the training results are shown in Table 2.

From Tables 2 and it can be observed the performance of U-net, Deeplabv3, and FCN is comparable, but U-net shows slightly higher F1-Score, IoU, and precision than the other models. Meanwhile, this study considers that for the feces recognition task, precision and recall are equally weighted, and F1-Score represents the balance between the two. Therefore, this study chooses U-net with a higher F1-Score and IoU for images segmentation task.

4.4. Correlation analysis of colon cleanliness and proportion of fecal area

The Pearson correlation coefficient is commonly used to quantify the correlation between two sets of data [18,19]. In this study, to demonstrate the correlation between the proportion of fecal area and colon cleanliness, the Pearson correlation coefficient was employed to estimate the degree of correlation between them. The calculation formula of the Pearson correlation coefficient is shown in Eq. (9).

$$\rho_{x,y}=\frac{\sum_i(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_i(x_i-\bar{x})^2\sum_i(y_i-\bar{y})^2}} \tag{9}$$

Table 4
CCRA test results for each BBPS (%).

BBPS	images captured number			
	n = 5	n = 10	n = 15	n = 20
0	97.5	99.5	99.7	99.8
1	82.5	92.1	95.8	97.7
2	89.8	98.3	99.4	99.6
3	100.0	100.0	100.0	100.0

Table 5
CCRA test results.

n	5	10	15	20
Accuracy/%	92.450	97.475	98.725	99.275
Average Time Consumed per Rating/s	0.0761	0.1373	0.2155	0.2892

In which, $\rho_{x,y}$ represents the Pearson correlation coefficient between the population of x and the population of y, $\overline{x}$ represents the sample mean of x, $\overline{y}$ represents the sample mean of y, x_i represents the i-th sample in the population x, and y_i represents the i-th sample in the population y. Let x_i be the average proportion of fecal area in 20 randomly sampled images from data with colon cleanliness level i, let $y_i = i$, representing the colon cleanliness. The Pearson correlation coefficient can be obtained by substituting the values of x_i and $y_i = i$ into the formula. As the calculation process involves random sampling, different results may be obtained from repeated calculations. In order to reduce accidental variations, ten calculations were performed, the calculation results are shown in Table 3.

4.5. Calculation of standard proportions and testing with the CCRA

We calculate the proportion of fecal regions corresponding to each level of colon cleanliness across all data. The average value for each level is then taken as the standard proportion. The resulting standard proportion values obtained are as follows: $s_0 = 0$, $s_1 = 0.0568$, $s_2 = 0.2227$, $s_3 = 0.4819$.

CCRA was tested using all test dataset in section 4.1. The testing methodology involved randomly selecting n images from the dataset corresponding to each colon cleanliness level. The selected n images were input into the rating system, which outputs a cleanliness rating. If the rating was the same as the cleanliness level of the image source, it was considered a correct rating; otherwise, it was considered incorrect This process was repeated 4000 times (1000 times for each cleanliness level) and the accuracy was calculated by taking the average of the correct ratings. The testing was performed for n values of 5, 10, 15, and 20, the test results corresponding to each BBPS are shown in Table 4. The average value of the data in Table 4 is taken as the final CCRA test result, as shown in Table 5.

From Tables 5 and it can be observed that as n increases, the accuracy of the rating algorithm improves. When n equals 10, the accuracy can reach over 97 %, with an average rating time of only 0.1373 s. Comparing the cases of n = 10 and n = 15, although the accuracy increased by 1.25 %, the time required increased by almost 1.5 times. Similarly, comparing n = 10 and n = 20, the accuracy increased by 1.8 %, but the time required increased by nearly 2 times. Therefore, in practical applications, a balance between rating time and accuracy should be considered. If strict rating time requirements are needed, n can be appropriately reduced; if higher accuracy is required, n can be appropriately increased.

5. Summary

We propose a colon cleanliness rating algorithm based on colonoscopy image analysis. The algorithm achieves a PA of 97.02 % and an IoU of 83.67 % in image segmentation tasks, demonstrating good segmentation performance. Testing of the Colon Cleanliness Rating Algorithm (CCRA) yields an accuracy of 92.45 %–99.275 % (depending on the number of images captured, n), demonstrating that the CCRA has high efficiency and accuracy, providing a reference for automated colon cleanliness rating. In Section 4.1.2, we cover the deletion method for small windows using a mirroring approach. The disadvantage of this method is that it can lead to the presence of redundant features at the junctions, which has a slight impact on model training. This problem has not been solved and remains a future research direction. Additionally, only three U-Net structures have been tested in this study, and there are many other structures that have not been experimented with, such as modifying the convolution counts before each sampling in U-Net. These avenues can be explored as future research directions. Thirdly, during the process of dividing the dataset, due to the inability to divide the dataset by video, there may be a small number of samples from the same video source, resulting in slight biases in the evaluation. In the future, we hope to find a partitioning method that can solve this problem. We also expect that more publicly available datasets on bowel cleanliness will be released by scholars for relevant research.

Ethics declarations

All participants/patients (or their proxies/legal guardians) provided informed consent for the publication of their anonymized case

details and images.

Data availability statement

Data associated with the study has not been deposited into a publicly available repository and data will be made available on request.

CRediT authorship contribution statement

Yu Bo: Supervision, Resources, Methodology, Funding acquisition, Data curation, Conceptualization, Writing - review & editing. **Shao Wei:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Yao Dengju:** Supervision, Resources, Funding acquisition, Formal analysis. **Wang Yunhao:** Software, Investigation, Data curation. **Zhang Heyi:** Software, Investigation, Data curation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

Acknowledgements

The authors are grateful for the anonymous reviewers who made constructive comments. This work was supported by the Natural Science Foundation of Heilongjiang Province (NO. F2017014) and the National Natural Science Foundation of China (NO. 62172128).

References

[1] J. Athreya Prasad, N. Owen Gareth, Shing W. Wong, et al., Achieving quality in colonoscopy: bowel preparation timing and colon cleanliness, ANZ J. Surg. 81 (4) (2011) 261–265, https://doi.org/10.1111/j.1445-2197.2010.05429.x.
[2] B. Shomron, B. Simon, A. Benjamin, The impact of colon cleanliness assessment on endoscopists' recommendations for follow-up colonoscopy, Am. J. Gastroenterol. 102 (12) (2007) 2680–2685, https://doi.org/10.1111/j.1572-0241.2007.01486.x.
[3] E. Sherer, T. Imler, T. Imperiale, The effect of colonoscopy preparation quality on adenoma detection rates, Gastrointest. Endosc. 75 (3) (2012) 545–553, https://doi.org/10.1016/j.gie.2011.09.022.
[4] E. Lai, A. Calderwood, G. Doros, et al., The Boston bowel preparation scale: a valid and reliable instrument for colonoscopy-oriented research, Gastrointest. Endosc. 69 (3) (2008) 620–625, https://doi.org/10.1016/j.gie.2008.05.057.
[5] Yan Gao, Jusheng Lin, Houde Zhang, et al., Pilot validation of the Boston bowel preparation scale in China, Dig. Endosc. : official journal of the Japan Gastroenterological Endoscopy Society 25 (2) (2013) 167–173, https://doi.org/10.1111/j.1443-1661.2012.01356.x.
[6] Junqiang Li, Ziye Zhao, Enda Yu, Application status of bowel preparation quality assessment scale, Chinese Journal of Digestive Endoscopy 31 (9) (2014) 539–542, https://doi.org/10.3760/cma.j.issn.1007-5232.2014.09.021.
[7] S. Sofia, R. Isadora, D. António, Use of the Boston Bowel Preparation Scale in the real life setting: what affects it? Rev. Esp. Enferm. Dig. : organo oficial de la Sociedad Espanola de Patologia Digestiva 113 (8) (2021) 625, https://doi.org/10.17235/REED.2020.7678/2020, 625.
[8] Yan Gao, Houde Zhang, Muxian Lin, et al., Reliability evaluation of the application of Boston bowel preparation scale, Chinese Journal of Digestive Endoscopy 29 (2) (2012) 78–80, https://doi.org/10.3760/cma.j.issn.1007-5232.2012.02.006.
[9] O.F. Ahmad, A.S. Soares, E. Mazomenos, et al., Artificial intelligence and computer-aided diagnosis in colonoscopy: current evidence and future directions, The Lancet Gastroenterology & Hepatology 4 (1) (2019) 71–80, https://doi.org/10.1016/S2468-1253(18)30282-6.
[10] Xu Dong, Keli Hu, Fangyu Hong, Polyp recognition in colonoscopy image based on convolution neural network, Journal of Shaoxing University(natural science) 42 (4) (2022) 47–52, https://doi.org/10.16169/j.issn.1008-293x.k.2022.10.007.
[11] Jieyao Yu, The research on segmentation method of polyp image in colonoscopy[D], [Master dissertation]. Harbin Engineering University (2020), https://doi.org/10.27060/d.cnki.ghbcu.2020.000394.
[12] S. Meryem, E.A. Mohamed, Multi-scale hybrid network for polyp detection in wireless capsule endoscopy and colonoscopy images, Diagnostics 12 (8) (2022) 2030, https://doi.org/10.3390/DIAGNOSTICS12082030, 2030.
[13] Yalikong Ayimukedisi, Huijun Zhuang, Shilun Cai, et al., Application of artificial intelligence based on deep learning in colonoscopy, Chinese Journal of Practical Surgery 40 (3) (2020) 353–357, https://doi.org/10.19538/j.cjps.issn1005-2208.2020.03.28.
[14] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional Networks for Biomedical Image segmentation[C]. 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 2015, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.
[15] K. Pogorelov, R.K. Randel, D.T. Lange, et al., Nerthus[P], Multimedia Systems Conference, 2017, https://doi.org/10.1145/3083187.3083216.
[16] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, Journal of Big Data 6 (1) (2019) 1–48, https://doi.org/10.1186/s40537-019-0197-0.
[17] Yudong Zhang, Zhengchao Dong, Xianqing Chen, et al., Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation, Multimed. Tool. Appl. 78 (3) (2019) 3613–3632, https://doi.org/10.1007/s11042-017-5243-3.
[18] Zhiqun Li, Lusheng Wang, Dandan Wang, Analysis of music similarity based on Pearson correlation coefficient, J]. Art and Performance Letters 2 (5) (2021), https://doi.org/10.23977/ARTPL.2021.020510.
[19] Wei Zeng, Ting Lu, Zeliang Liu, et al., Research on a laser ultrasonic visualization detection method for human skin tumors based on pearson correlation coefficient, J]. Optics and Laser Technology (2021) 141, https://doi.org/10.1016/J.OPTLASTEC.2021.107117.