IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 8, 2021, accepted January 9, 2022, date of publication January 14, 2022, date of current version January 28, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3143580

Endoscopic Ultrasound Image Recognition Based on Data Mining and Deep Learning

YUFEI XIE[1], YU CAI[2], YANG YU[3], SEN WANG[4], WENLIN WANG[1], AND SHASHA SONG[5]

[1]School of Automation, Wuhan University of Technology, Wuhan, Hubei 430070, China
[2]College of Pharmacy, Hubei University of Chinese Medicine, Wuhan, Hubei 430070, China
[3]Nuclear Power Institute of China, Chengdu, Sichuan 610213, China
[4]Shenzhen Gengfeng Technology Company Ltd., Shenzhen, Guangdong 518001, China
[5]College of Pharmacy, Shenzhen Technology University, Shenzhen, Guangdong 518118, China

Corresponding authors: Wenlin Wang (wlwang0618@126.com) and Shasha Song (songshasha@sztu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 81700056, and in part by the Shenzhen Gengfeng Technology Company Ltd.

ABSTRACT The recognition of medical images, especially endoscopic ultrasound images, has the characteristics of changing images and insignificant gray-scale changes, which requires repeated observation and comparison by medical staff. In view of the above-mentioned characteristics of ultrasound imaging, a system scheme suitable for image processing is proposed, which can analyze the biliary tract, gallbladder, abdominal lymph nodes, liver, descending duodenum, duodenal bulb, stomach, pancreas, pancreatic lymph nodes, there are a total of 10 ultrasonic organs, including 21 kinds of sub-categories and 3510 images. The images are preprocessed using binarization, histogram equalization, median filtering and edge enhancement algorithms. The improved YoloV4 convolutional neural network algorithm is used to train the data set and perform high accuracy is detected in real time. Finally, the average accuracy of this algorithm has reached 91.59%. The algorithm proposed in this paper can make up for the shortcomings of manual detection in the original image detection system, improve the efficiency of detection, and at the same time as an auxiliary system can reduce detection misjudgments, and promote the development of automated and intelligent detection in the medical field.

INDEX TERMS Endoscopic ultrasonography, image processing, data mining, convolutional neural network.

I. INTRODUCTION

Ultrasound medical imaging is widely used in medical diagnosis. With the advent of the era of big data, medical imaging has entered a new development process that combines big data and deep learning. The medical ultrasound image library imports a large number of medical images every day, including various information about normal and abnormal organs in various parts. For the existing technology, the extraction of medical information is very critical and necessary. Based on this key information, professionals can derive the type of patient and the degree of harm from it, and take relevant treatment methods based on the image data. At present, common medical imaging equipment can manually mark images and color specific areas. However, for specific parts and organs, doctors' experience and learning basis are still needed to make judgments. In recent years, the development of computer technology and artificial intelligence has gradually expanded to the field of medicine. In the recognition of medical images, the direct image enhancement technology proposed by Chen Yan *et al.* is different from the indirect image enhancement technology used in the classification of medical images [1]. Li Bo *et al.* proposed a medical image classification algorithm based on multi-feature fusion in scale space, which classifies images by establishing a Gaussian scale space [2]. Li Zuoyong described an improved segmentation algorithm for digital morphology medical images, which segmented medical images by combining morphology and filtering operations [3]. In the field of deep learning, image segmentation algorithms gradually exert their advantages. Wang Li *et al.* introduced an algorithm based on Faster-RCNN neural network, which serves as an auxiliary diagnosis [4]. Bakalo Ran *et al.* proposed a weak semi-supervised detection dual-branch deep learning network algorithm for medical image recognition, which has a significant improvement effect on recognition [5].

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

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

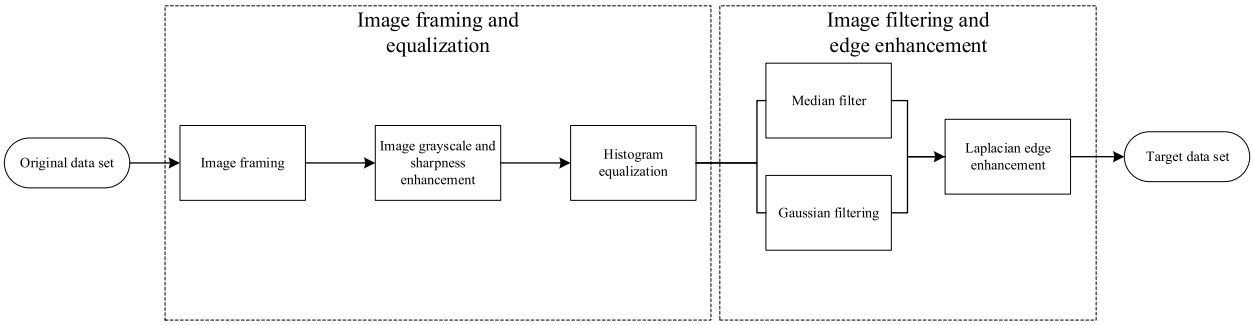

FIGURE 1. Preprocessing process of ultrasound endoscopy data set.

Literatures [6]–[9] introduced a multi-layer boundary perception deep learning network, a detection network based on Spatial Domain, Transform Domain and CNN methods, and a deep learning image classification algorithm using wavelet decomposition to replace the convolution kernel. At present, this type of algorithm is mainly trained on a large amount of data, with a large amount of calculation and high computational cost, and it is difficult to give full play to its advantages in ultrasound endoscopic detection with few key frames, and the detection samples are insufficient. The current research on ultrasound endoscopy mainly has the following problems:

(1) Ultrasound image itself is relatively fuzzy, and the key frames with few features make it difficult to be detected during processing.

(2) Medical ultrasound imaging uses the physical characteristics of ultrasound and converts it into an image. Its imaging is very limited for the detection of organs that contain gas and bones, and the interference is difficult to eliminate. It requires strong professional knowledge to assist in judgment, which is likely to cause misdiagnosis.

(3) The labeling error of medical ultrasound images is large, and the labeling relies on the assistance of professional physicians, and the judgment is difficult to quantify.

Aiming at the above problems, this paper proposes a medical ultrasound detection and recognition algorithm based on data mining and deep learning. Different from previous algorithms, this research uses image methods and data mining techniques to improve image quality and optimize deep learning algorithms. Thus, the fine-grained image classification effect is improved. Based on the existing data set, the algorithm performs key frames of ten organs of the biliary tract, gallbladder, abdominal lymph nodes, liver, descending duodenum, duodenal bulb, stomach, pancreas, pancreatic lymph nodes and esophagus. Subdivision, establishing multiple tag libraries for different parts of special organs, and realizing the recognition of 21 organ subtypes. The improved algorithm greatly reduces the requirements of the scene and solves the training problem caused by the different organ environment. Using image edge enhancement and filtering algorithms to improve image clarity and recognition greatly reduces the impact of image quality issues such as blur on recognition. Through OpenCV-based random image transformation and data mining technology, the problem of fewer key frames is solved, and the sample capacity is expanded. Finally, the image is trained and learned through the improved YoloV4 convolutional neural network, which improves the accuracy of image recognition and solves the problems of low efficiency and bias in the current manual judgment of ultrasound endoscopic images.

II. DATA SET IMAGE PREPROCESSING

Aiming at the following quality problems in endoscopic ultrasound images:

(a) Interference and noise interference caused by the useless burr of the image itself.

(b) Image blur and ghosting caused by image framing.

(c) The local salt and pepper noise and global Gaussian noise of the image.

In this paper, an image preprocessing algorithm based on OpenCV is used to optimize the image data set of medical ultrasound endoscopy, which preserves high-frequency signals, filters out irrelevant noise, improves image quality, and makes the features in the data set easier to be identified.

The main processing flow is as shown in Figure 1.

A. ACQUISITION OF DATA SET

At present, there are few data sets of medical ultrasound annotation image classification. This article is based on the framing and annotation of the medical ultrasound endoscopy training video of the People's Medical Publishing House. The video data contains the medical classification and annotation of dozens of visceral endoscopic images including the gallbladder, pancreas, etc., which provides a scientific and rigorous image source for the construction of the data set. Through the framing of the image and video, enough frames are obtained to form an image data set. When framing, the video framing algorithm is used for frame extraction, and the interval is set to 1∼5 frames. This study selected ten normal organs such as stomach and pancreas as the research and test objects, and finally extracted 3510 pictures with useful frames. The data division table of endoscopic ultrasound images is shown in table 1.

TABLE 1. The partition table of the sub-frame image data set of ultrasound endoscopy.

Categories	Subdivided	Number of frames
Biliary tract	Biliary tract1(good)	0-440
	Biliary tract2(good)	
Gallbladder	Gallbladder(good)	441-810
Abdominal lymph nodes	Abdominal lymph nodes(good)	811-1111
Liver	Liver_mode1(good)	1112-1391
	Liver_mode2(good)	
	Duodenal drop segment mode1(good)	1392-1746
	Duodenal drop segment mode2(good)	
	Duodenal drop segment mode3(good)	
Duodenal bulb	Duodenal bulb mode1(good)	1747-2258
	Duodenal bulb mode2(good)	
	Duodenal bulb mode3(good)	
Stomach	Stomach mode1(good)	2259-2751
	Stomach mode2(good)	
	Stomach mode3(good)	
	Stomach mode4(good)	
	Stomach mode5(good)	
Pancreatic	Pancreatic mode1(good)	2752-3063
	Pancreatic mode1(good)	
Pancreatic lymph nodes	Pancreatic lymph nodes(good)	3064-3310
Esophageal	Esophageal(good)	3311-3510

The obtained images cannot be directly used for image classification training, and certain repairs and image enhancements are required [10], [11]. The data set needs to be further processed to remove the interference frame composed of bubbles (such as large bubbles in the esophageal image) and other interference items in each organ, so that the characteristics of a certain organ in different orientations can still be identified in a complex scene [12]. In this paper, labelimg (a visualization tool for generating data set labels) is used for image labeling and ROI (Region of Interest, ROI) frame selection, and 21 divided images of organs are labelled. By generating the corresponding xml file, a data set in VOC format is obtained. Part of the pancreatic endoscopic ultrasound image set after framing, as shown in Figure 2.

B. IMAGE GRAYING AND ENHANCEMENT

The color interface in the data set is composed of RGB (Red, Green, Blue) three channels, and its characteristics only indicate the optical characteristics of the image, and cannot reflect the morphological characteristics of the target object. RGB image does not help to solve the identification of ultrasonic internal organs, and the redundant information it contains will increase the amount of features and calculations. In order to reduce the amount of calculation in the later stage, the image needs to be unified to the [0, 255] grayscale variation range [13]. The average dimension method is used in this study.

$$Gray(i, j) = (R(i, j) + G(i, j) + B(i, j))/3 \quad (1)$$

In the above formula, $Gray(i,j)$ represents the gray value of row i and column j. R, G and B respectively represent the value of each channel, and the average value is used as the gray value of the pixel.

After the medical ultrasound image is divided into frames, some images will appear blurry and ghosting. This kind of noise will reduce the recognition of the image and increase the difficulty of recognition. To solve the above problems, after the image is grayed, the PIL library in Python is used to enhance the sharpness of the image to improve the quality and definition of the image [14]. As shown in Figure 3,

FIGURE 2. Pancreatic endoscopic ultrasound image data set (partial).

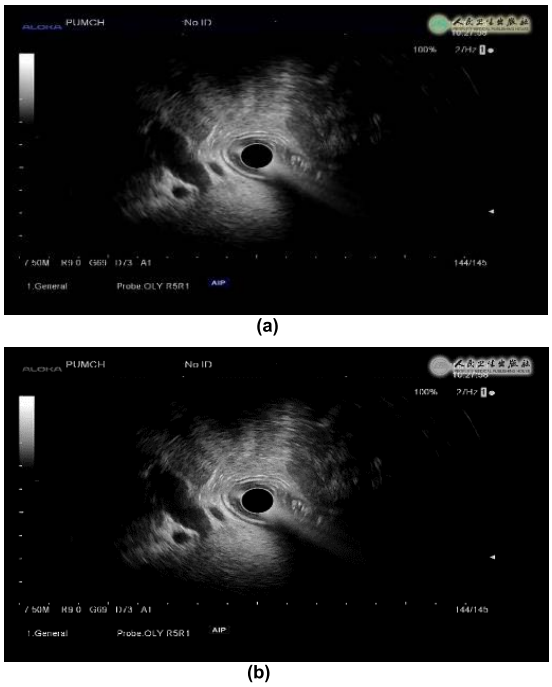

FIGURE 3. (a) Original image of endoscopic image; (b) After sharpening.

the sharpness of the edge portion of the image (b) is significantly enhanced compared to the image (a). It has a better feature enhancement effect for images judged based on internal edge signals of organ endoscopic images.

C. HISTOGRAM EQUALIZATION

The grayscaled image has uneven distribution of brightness. In order to improve the overall pixel grayscale distribution and contrast of the image, it is necessary to transform an image with a known grayscale distribution into a uniformly distributed grayscale image. By improving the irregular distribution of pixels, the range of pixel distribution is enlarged, and the contrast of the image is further improved. The steps to realize the remapping distribution of the histogram are as follows:

Calculate the probability density function at each gray level:

$$P_r\left(r_k\right) = n_k/n \tag{2}$$

In the above formula, $P_r\left(r_k\right)$ is the probability under the gray level of r_k, n_k is the number of pixels under the gray level, and n is the total number of pixels in the image. Use the mapping relationship to get the distribution function value after gray level mapping.

$$S_K = T(\mathrm{r}_k) = \sum_{i=0}^{k} P_r(\mathrm{r}_i) \tag{3}$$

Among them, S_k represents the value of the probability distribution function under the gray level from r_j to r_k.

After converting to the standard gray value through the mapping arrangement, the histogram is equalized, and the number of pixels will no longer only be distributed near the black, but distributed on the gray scale of 0-255, realizing the medical image Rearrangement of gray levels. Histogram before and after equalization, as shown in Figure 4.

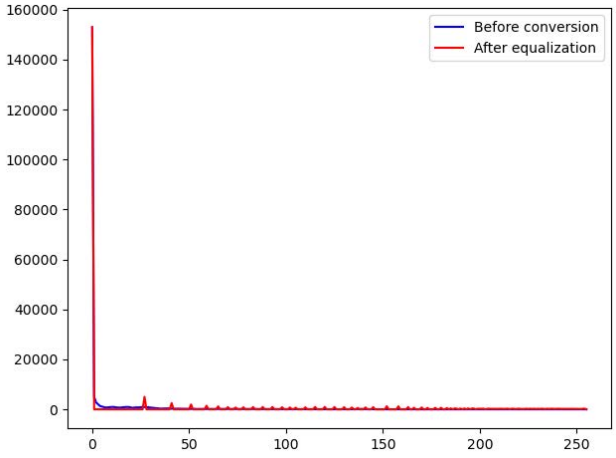

FIGURE 4. Comparison of histograms before and after equalization.

As shown in Figure 4, the abscissa represents the gray level, and the ordinate represents the number of pixels in the gray level. After equalizing the pancreas image, the gray value and pixel statistics are used to obtain a visual gray histogram. Before equalization, most of the pixels are distributed in the 0-50 area. After equalization, it can be clearly seen that the distribution of the grayscale histogram is more balanced, with a distribution in the range of 50∼255, thus achieving grayscale remapping and image equalization [15].

D. IMAGE FILTERING ALGORITHM

Ultrasound imaging mainly uses the acoustic characteristics of ultrasound reflected in different organs and tissues, and can distinguish different organ contours. The image has interference mainly caused by speckle noise, and this kind of noise cannot be eliminated by physical methods, and can only be processed by the method of imaging [16]. In this article, the graininess and glitch interference in the ultrasound endoscopic images are noise points. In order to reduce the impact of this kind of noise, Gaussian denoising is first adopted, and then median filtering is used [17].

Gaussian denoising blurs the full-screen noise, performs assignment calculations through the movement of the module, and uses a two-dimensional Gaussian filter to design a 3*3 mask when setting the template.

$$H_{i,j} = \frac{1}{2\pi\delta^2} e^{\frac{-(\mathrm{i}-\mathrm{k}-\mathrm{i})^2+(\mathrm{j}-\mathrm{k}-1)^2}{2\delta^2}} \tag{4}$$

Among them, k represents the size of the serial port. Through the movement of the window, $\mathrm{H}_{i,j}$ is the value of row i and column j in the mask, and Gaussian filtering is performed on pixels in all positions.

However, Gaussian filtering cannot remove salt and pepper noise. In response to this situation, this paper uses median

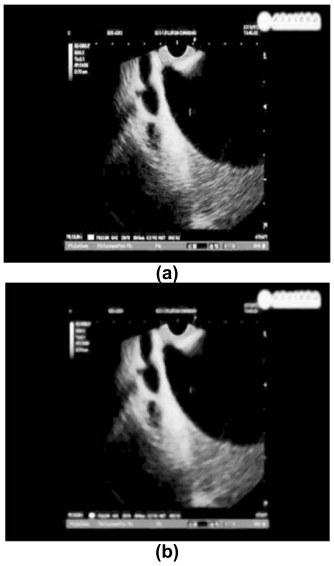

FIGURE 5. (a) Before filtering; (b) After filtering.

filtering as a supplementary method to further remove the salt and pepper noise in the image, while retaining the edge features of the image. When the template is moved, the gray levels of the pixels in the mask are sorted, and the median value of the area covered by the mask is taken as the new gray value of the central pixel. The filtered result is shown in Figure 5. Compared with figure (a), the burr and salt and pepper noise of the processed figure (b) are greatly weakened, and the joint filtering effect is good.

E. EDGE EXTRACTION AND ENHANCEMENT ALGORITHM

The recognition of endoscopic images is mainly to judge the ultrasonic appearance characteristics of different organs and membrane structures. The main high-frequency information is the changes in the pixel gray levels at the edges and contours. Therefore, the optimization and extraction of edge information will play an important role in the recognition effect. In this paper, Laplace operator is used to construct a 3×3 convolution kernel, which is used to calculate the gray jump value of the edge pixels of the image [18], [19].

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{5}$$

Among them, $\nabla^2 f$ represents the second-order gray-scale differential value of the image at (x, y) coordinates.

The simplified expression is:

$$\nabla^2 f = -4f(\mathrm{x}, \mathrm{y}) + \mathrm{f}(\mathrm{x}-1, \mathrm{y}) + \mathrm{f}(\mathrm{x}+1, \mathrm{y}) + \mathrm{f}(\mathrm{x}, \mathrm{y}-1) + \mathrm{f}(\mathrm{x}, \mathrm{y}+1) \tag{6}$$

Further use the Laplace convolution template to realize the calculation:

$$H = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \tag{7}$$

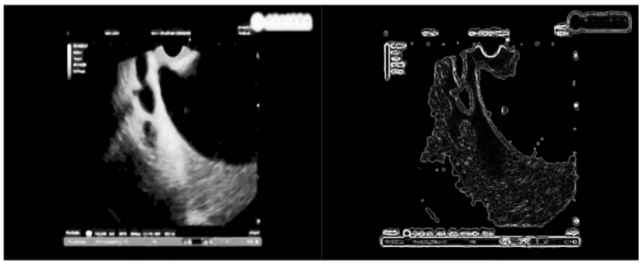

FIGURE 6. (a) Before edge extraction; (b) After edge extraction.

Among them, H represents the Laplace operator convolution template.

The convolution operation can extract the edges of the image, as shown in Figure 6. The edge features of the image are preserved, and the features of the image are further enhanced.

III. CONVOLUTIONAL NEURAL NETWORK TRAINING AND PREDICTION

The research idea based on convolutional neural network is shown in Figure 7.

The research route mainly consists of three parts: data mining and expansion, Yolo V4 detector network framework construction, network evaluation and improvement. In order to improve the robustness of the algorithm, this paper selects the best parameters and regularization methods to debug and optimize the algorithm.

The simulation environment is under the Windows 10 operating system, using the Tensorflow framework to build a convolutional neural network YoloV4. The project uses RTX Quadro 4000 graphics card as the acceleration platform, and finally 50 rounds of training, a total of 2 hours and 35 minutes.

A. CONSTRUCTION OF YOLOV4 NETWORK FRAMEWORK

The construction of YOLO V4 network detector is very important, which mainly includes backbone network, SPP, PANet network and prediction network [20], as show in Figure 8. The backbone network is used to extract features, SPP participates in pooling as an additional part [21], and PANet mainly participates in feature fusion [22]–[25]. The Yolo head part is mainly used for forecasting.

SPP uses 1×1, 5×5, 9×9, 13×13 pooling to check the feature layer convolution and pooling. This structure can increase the receptive field and separate the characteristics of the upper and lower layers as much as possible. PANet is a segmentation algorithm used to improve the features of the target detection object and realize the repeated extraction of features.

The feature layer dimensions designed using Yolo network are respectively (52, 52, 256), (26, 26, 512) and (13, 13, 1024). On this basis, the parameters are adjusted, and at the same time, regularization and data enhancement are introduced. The training forms a convolutional neural network

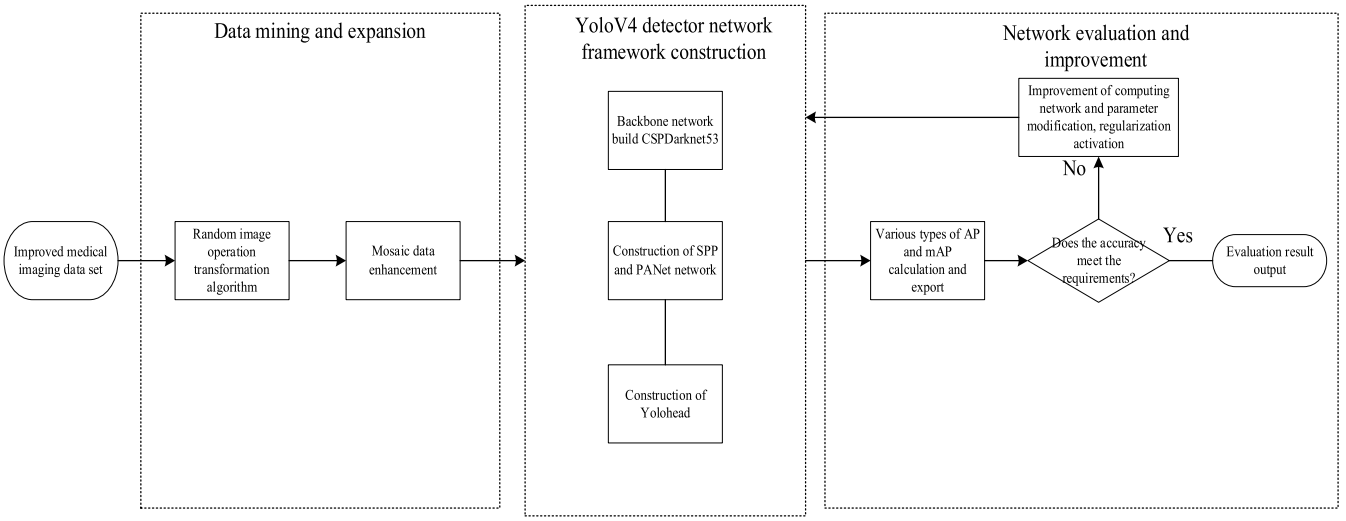

FIGURE 7. Simplified roadmap for convolutional neural network research.

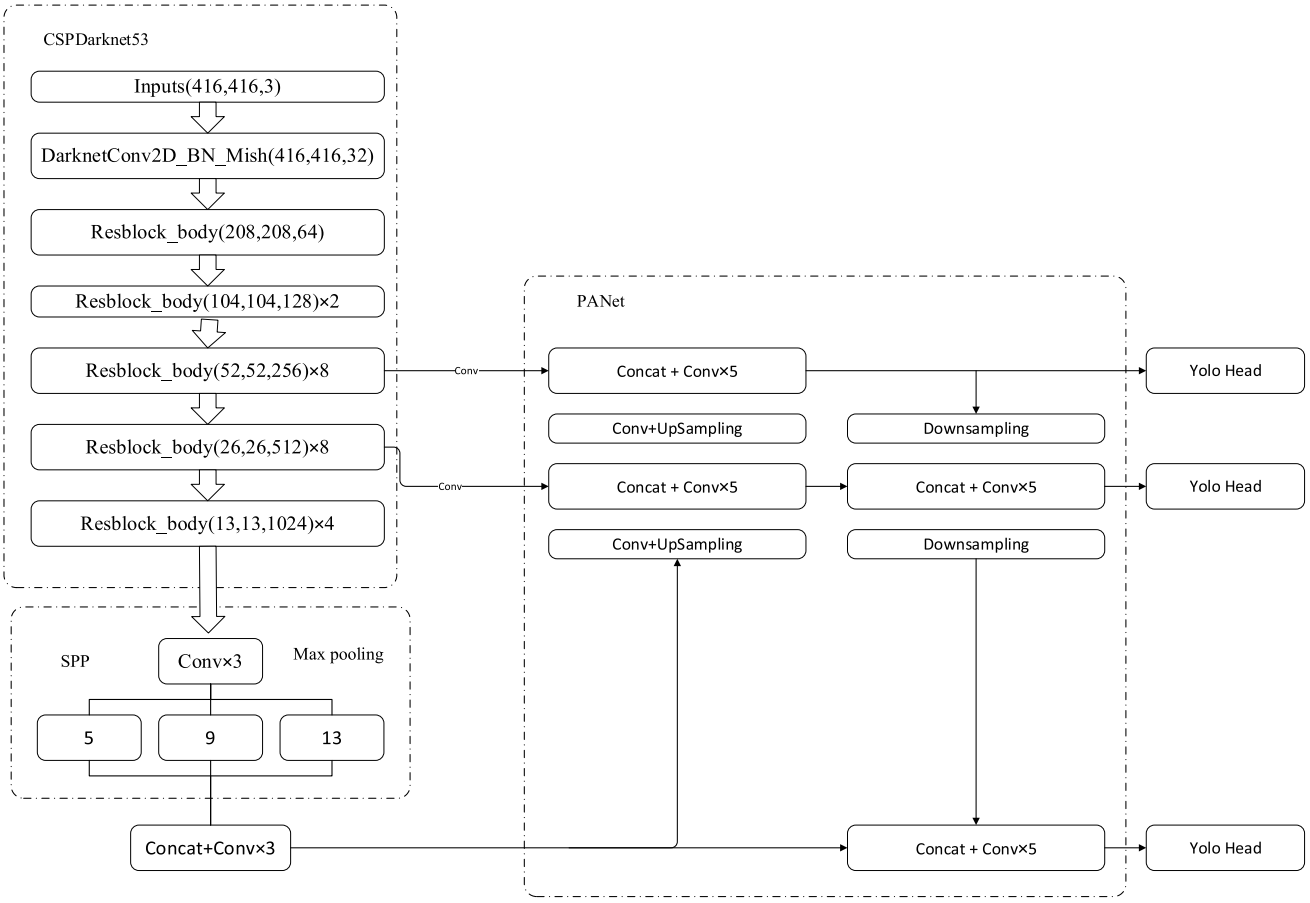

FIGURE 8. Schematic diagram of YoloV4 network framework.

suitable for endoscopic ultrasound samples. The improved network performed well in the experiment and can effectively identify endoscopic image features.

B. TRAINING PREPARATION AND DATA MINING

The endoscopic ultrasound images in this article contain both characteristic frames and useless frames. The characteristic

frames are organs that have been correctly marked in the image, with fewer interference items. However, the number of such frames is relatively small. In order to solve this problem, data enhancement algorithms are used to increase the diversity of image features while avoiding over-fitting of the model [26].

(a) Data enhancement is based on the OpenCV function library to perform random inversion, rotation, scaling and other operations on the training data, and process the frames in the training set by setting the random state flag.

(b) The mosaic data enhancement method is introduced, the principle of which is to use image stitching to cut partial pixel areas of four images and stitch them into one image. That is, four pictures are calculated each time, and so the batch (number of images per round of training) is not large, which can improve the calculation efficiency of the data.

(c) In the process of forward propagation and backward calculation, the neural network will continuously update the model information, calculate the loss value, and determine the optimal solution for target optimization according to the loss function. Use stochastic gradient descent to continuously optimize and determine the optimal solution. However, when encountering multimodal functions, this gradient descent strategy will fall into the trap of local optimality, and this local optimal solution is not the overall optimal solution. To solve this problem, the cosine annealing decay is introduced, that is, it jumps out when it falls into the local optimal solution, and then restarts the calculation until the overall optimal solution is found. The principle is:

$$\eta_t = \eta^j_{\min} + (\eta^j_{\max} - \eta^j_{\min})(1 + \cos\left(\frac{T_{current}}{Ti}\,\pi\right) \quad (8)$$

In the above formula, j represents the jth index value, $\eta^j_{\max}$ is the maximum learning rate, and $\eta^j_{\min}$ is the minimum learning rate. $T_{current}$ indicates the number of epoch rounds currently executed. The learning rate will be changed after each restart, so as to achieve an effect of updating the learning rate. After each restart, it is multiplied by a fixed value to realize automatic increase. In this way, the learning rate will be updated after each restart.

Random image changes, Mosaic data enhancement, and cosine annealing attenuation are used to achieve data expansion of a limited frame of ultrasound images, and reduce the amount of calculation, avoiding over-fitting of the calculation.

C. TRAINING AND IDENTIFICATION OF NEURAL NETWORKS

The endoscopic ultrasound image data set includes processed images and original images, with a total of 3510 effective frames and an image size of 608 × 608. The ratio of the training set to the test set is 9:1, the batch size of each iteration is 2, each round needs to complete 1580 batches, and the total iteration epoch (number of training rounds) is 50 rounds.

The training of the convolutional neural network is mainly the iterative calculation and update of the convolutional layer, the pooling layer, and the fully connected layer. The main steps are:

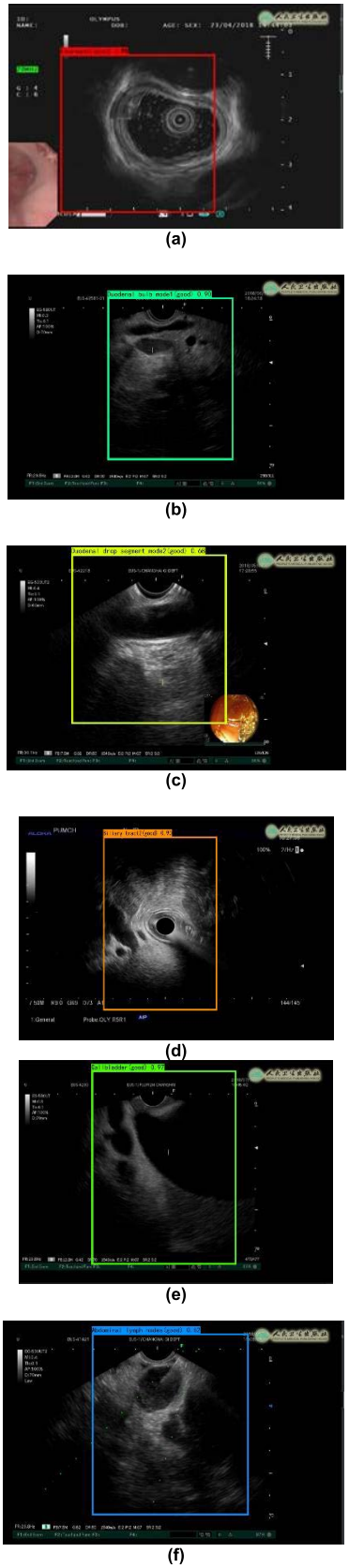

FIGURE 9. Recognition results of organs (partial).

(a) Create an image data set in VOC2007 format, mark the identifiable information ROI, and generate the corresponding XML file.

(b) Convolve the image features through the CSPDarknet53 network, compress and extract useful features, and calculate the loss value under the loss function.

(c) Batch gradient descent is used to calculate the gradient update parameters w and b, and regularization is used to reduce the weight while preventing overfitting.

(d) When the Loss value converges to the expected effect, the weight file is output.

(e) The weight file and the forward propagation process are used to identify the target in the image.

After training according to the above steps, the recognition results of some organs are shown in Figure 9.

As shown in Figure 9, (a) duodenal bulb; (b) esophagus; (c) bile duct; (d) descending duodenum; (e) abdominal lymph nodes; (f) gallbladder.

The recognition result is consistent with the actual situation and has a faster recognition rate. The single image takes about 0.1ms.

IV. RESULT ANALYSIS

In order to verify the effect and accuracy of training and the detection ability of the improved YoloV4 neural network, it is necessary to judge the degree of overlap between the detection frame of the target recognition and the real frame, and the parameter intersection ratio IOU is the main basis.

$$IOU = \frac{s_\Lambda}{s_u} \tag{9}$$

Among them, the numerator represents the intersection area of the detection frame and the real frame, the denominator represents the union area of the detection frame and the real frame, and IOU represents the intersection ratio.

The Precision P of the model is calculated by the following formula:

$$p = \frac{TP}{TP + FP} \tag{10}$$

Among them, TP indicates that the detected result is a positive sample, and FP indicates that the detected result is a negative sample.

The calculation of the recall rate in the model is similar to the Precision:

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

Among them, FN indicates that the detected result is a negative sample, but it is actually a positive sample. Recall reflects the proportion of the correctly classified part of the positive sample to all the positive sample parts.

A single index has certain limitations on the evaluation of the model, and the effect of target detection should be judged under the confidence level. Positive and negative samples are classified, and the corresponding parameter is confidence. The introduction of confidence makes the precision P and the recall rate R unite, and the average precision AP is represented by the P, R curve. The AP statistics of normal pancreatic lymph nodes are shown in Figure 10.

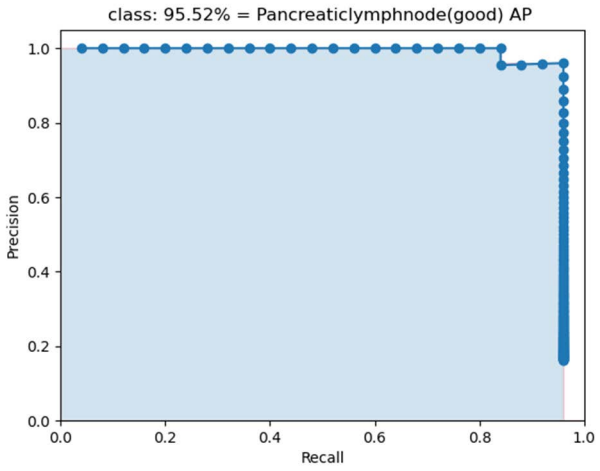

FIGURE 10. AP statistics of normal pancreatic lymph nodes.

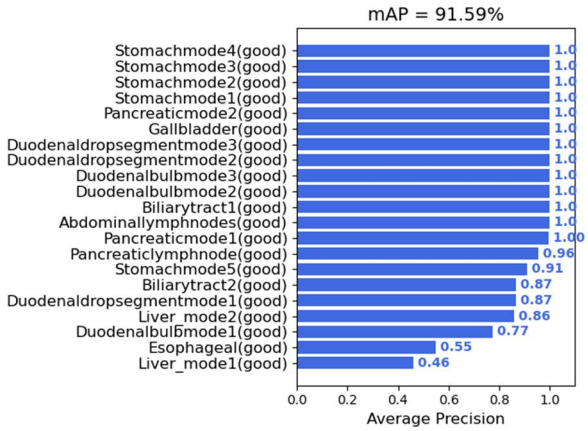

FIGURE 11. mAP statistics for 21 sub-categories.

Among them, the abscissa represents the recall rate, and the ordinate represents the precision rate. It can be found that as the positive sample threshold point shifts to the left, the value of precision is initially very close to 1, and the recall is very close to 0, and AP of normal pancreatic lymph nodes is relatively high, reaching 95.52%.

mAP means average precision, which is a measure of the average precision of all types of objects trained. mAP is an important basis for evaluating the quality of the model. Its calculation formula is:

$$mAP = \frac{\sum_{i=1}^{c} AP_i}{C} \tag{12}$$

Among them, c represents the total number of sample classifications, and AP_i is the AP value of the i-th category.

Use the above principles to evaluate the model in this article and plot the mAP of 21 sub-categories in the data set, as shown in Figure 11.

TABLE 2. Terminology and explanation.

Terminology	Explanation
AP	Average Precision.
Backbone、Neck、Yolo head	Components of the Yolo neural network framework.
Batch	Number of images used for training in each round.
CSPDarknet53	Backbone extraction network in Yolo.
epoch	Number of training rounds.
IOU	Intersection over Union. An indicator describing the degree of coincidence between the detection frame and the real frame.
Labelimg	A visualization tool for generating data set labels.
Laplace operator	A kind of differential operator with rotation invariance.
mAP	Mean Average Precision.
Mosaic	A data augmentation method for training.
P	Precision.
PANet	A feature fusion module in Yolo.
PIL	Python Image Library.
R	Recall rate.
ROI	Region of Interest. Framed target area.
SPP	A feature enhancement network in Yolo.
Tensorflow	A machine learning library, this article programming is based on tensorflow2.0.
VOC	Voice. Data set format for training.
W, b	Parameter weights and biases of neurons.
XML	eXtensible Markup Language. Used to store label data and other information.

Among them, the abscissa represents the AP value, and the ordinate represents the category of ultrasound endoscopic images. The mean average precision mAP is 91.59%. The results show that the recognition algorithm has high accuracy.

V. CONCLUSION

This paper proposes a method of ultrasound endoscopic image recognition system based on data mining and deep learning.

(a) The image preprocessing algorithm is used to improve the quality of ultrasound endoscopic images and solve the image quality problems caused by framing and ultrasound interference.

(b) The data enhancement algorithm in the OpenCV function library is used to expand the number of training sets, improve the robustness of the training model, and solve the problem of too few effective frames in the current training set.

(c) The Mosaic algorithm is introduced to improve its training efficiency, and solves the problems of low recognition efficiency and high computational cost of existing algorithms.

(d) The cosine annealing algorithm is used to avoid overfitting of the results, which further improves the robustness of the algorithm.

Through simulation verification, the mAP of 21 sub-organs reached 91.56%. This algorithm can effectively identify the organs in endoscopic ultrasound images, providing new solutions and solutions for identifying more organs and cancerous organs in the future.

AUTHOR CONTRIBUTIONS

Wenlin Wang designed the research; Yufei Xie and Yu Cai conducted the research; Yufei Xie and Sen Wang analyzed the data; Yufei Xie and Shasha Song wrote the manuscript; Wenlin Wang and Shasha Song had access to all data and had primary responsibility for the final content.

DISCLOSURES

The authors disclosed no conflicts.

APPENDIX 1

To facilitate research, the main terminology used in this article are explained as follows:

REFERENCES

[1] C. Yan and G. Guohua, "Application of a direct image enhancement method in medical image classification," *Comput. Appl. Softw.*, vol. 6, pp. 26–32, Mar. 2007.

[2] L. Bo, C. Peg, L. Wei, and Z. Dazhe, "Medical image classification based on multi-feature fusion in scale space," *Comput. Appl.*, vol. 33, no. 4, pp. 1108–1114, 2013.

[3] H. Jinmei and L. Zuoyong, "Medical image segmentation based on improved mathematical morphology algorithm," *Comput. Simul.*, vol. 28, no. 5, pp. 299–302, 2011.

[4] W. Li, L. Huijuan, Y. Minchao, and Y. Ke, "A cancer image detection method based on Faster-RCNN," *J. China Univ. Metrol.*, vol. 29, no. 2, pp. 136–141, 2018.

[5] R. Bakalo, J. Goldberger, and R. Ben-Ari, "Weakly and semi supervised detection in medical imaging via deep dual branch net," *Neurocomputing*, vol. 421, pp. 15–25, Jan. 2021.

[6] F. An and J. Liu, "Medical image segmentation algorithm based on multilayer boundary perception-self attention deep learning model," *Multimedia Tools Appl.*, vol. 80, pp. 15017–15039, Feb. 2021.

[7] S. Pradeep and P. Nirmaladevi, "A review on speckle noise reduction techniques in ultrasound medical images based on spatial domain, transform domain and CNN methods," in *Proc. IOP Conf., Mater. Sci. Eng.*, vol. 1055, no. 1, 2021, Art. no. 012116.

[8] A. H. Masquelin, N. Cheney, C. M. Kinsey, and J. H. T. Bates, "Wavelet decomposition facilitates training on small datasets for medical image classification by deep learning," *Histochem. Cell Biol.*, vol. 155, no. 2, pp. 309–317, Feb. 2021.

[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[10] L. Huilan and Y. Hui, "Image classification method based on iterative training and ensemble learning," *Comput. Eng. Des.*, vol. 41, no. 5, pp. 1301–1307, 2020.

[11] W. Hao, Z. Ye, S. Honghai, and Z. Jingzhong, "Overview of image enhancement algorithms," *Chin. Opt.*, vol. 10, no. 4, pp. 438–448, 2017.

[12] W. Teng, B. Leping, Y. Zhonglin, L. Qifeng, and Y. Xuanfang, "A flame recognition method based on image similarity of consecutive frames," *J. Naval Univ. Eng.*, vol. 29, no. 4, pp. 48–52, 2017.

[13] Z. Guihua, F. Yanbo, and L. Weidong, "Grayscale of image processing and acquisition of characteristic regions," *Qiqihar Univ. J.*, vol. 4, pp. 49–52, Feb. 2007.

[14] Z. Zunshang, "Research on image enhancement technique," Nat. Univ. Defense Technol., Changsha, China, 2009.

[15] M. Shuiqing, Z. Jing, and H. Changjun, "Research on image enhancement method based on median filtering and histogram equalization," *Wireless Internet Technol.*, vol. 22, pp. 106–107, Aug. 2017.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[17] L. Haifeng, Z. Chao, L. Jiang, and L. Fuliang, "Improved mean division algorithm for median filtering," *Comput. Syst. Appl.*, vol. 26, no. 3, pp. 162–168, 2017.
[18] C. Chuxia, "Research on algorithms of image filtering and edge detection for image enhancement," Hefei Univ. Technol., Hefei, China, 2009.
[19] S. K. Dinkar, K. Deep, S. Mirjalili, and S. Thapliyal, "Opposition-based Laplacian equilibrium optimizer with application in image segmentation using multilevel thresholding," *Expert Syst. Appl.*, vol. 174, Jul. 2021, Art. no. 114766.
[20] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT: Real-time instance segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9157–9166.
[21] Y. Zhengyan, "Change detection for high-resolution SAR images based on NSCT SPP net," XiDian Univ., Xi'an, China, 2018.
[22] L. Yang and G. Hongwei, "Adaptive multi-modal feature fusion for far and hard object detection," *J. Meas. Sci. Instrum.*, vol. 12, no. 2, pp. 232–241, 2021.
[23] S. Hua, "Comparative analysis between regularized neural network algorithm and the early stopping iteration algorithm," *Bull. Sci. Technol.*, vol. 29, no. 10, pp. 112–114, 2013.
[24] S. Zehao, "Target detection algorithm based on feature pyramid network," *Mod. Comput. (Prof. Ed.)*, vol. 3 pp. 42–44, Apr. 2018.
[25] J. Long, K. Xueyuan, H. Haihong, Q. Zhinian, and W. Yehong, "Research on overfitting of artificial neural network forecasting model," *Acta Meteorol. Sinica*, vol. 1, pp. 62–70, Jul. 2004.
[26] W. Bing, L. Hongxia, L. Wenjing, and Z. Menghan, "Mask detection algorithm based on improved YOLO lightweight network," *Comput. Eng. Appl.*, vol. 57, no. 8, pp. 62–69, 2021.

YUFEI XIE is currently pursuing the degree with the Wuhan University of Technology. He participated in a number of scientific research projects, including national innovation and entrepreneurship program for college students. His research interests include computer vision and artificial intelligence.

YU CAI received the M.Phil. degree from the University of Macau and the Ph.D. degree from the Hubei University of Chinese Medicine. She is currently a Lecturer with the Hubei University of Chinese Medicine. Her research interests include material basis and pharmacology of Chinese medicine.

YANG YU received the bachelor's and M.S. degrees from Harbin Engineering University, China, in 2010 and 2013, respectively. He is currently a Senior Engineer with the Nuclear Power Institute of China. His research interests include big data analysis and algorithm design.

SEN WANG received the B.S. degree from Qingdao University, China, in 2013. He is currently a Data Analysis Engineer with Shenzhen Gengfeng Technology Company Ltd. His research interest includes algorithm design and programming.

WENLIN WANG received the B.S. degree in automation from the Harbin Institute of Technology, in 2008, and the M.S. and Ph.D. degrees in nuclear science and technology from Harbin Engineering University, in 2013 and 2016, respectively. He is currently a Lecturer with the School of Automation, Wuhan University of Technology. His research interests include system modeling and simulation, data mining, and analysis.

SHASHA SONG received the M.M. and M.D. degrees from Harbin Medical University, China, in 2013 and 2016, respectively. She is currently an Assistant Professor with Shenzhen Technology University. Her research interests include medical data analysis and disease diagnosis. She has many academic papers were published.