International Journal of Research in Engineering, Science and Management
Volume 5, Issue 12, December 2022
https://www.ijresm.com | ISSN (Online): 2581-5792

CNN and Data Mining Based Endoscopic Ultrasound Image Recognition

Siva Prasad Patnayakuni*
Senior Data Engineer, HEB, India

***Abstract*: Therapeutic staff must repeatedly observe and compare endoscopic ultrasound images, which have the individuality of changing images and irrelevant gray-scale changes. A system proposal that is fit for image processing is proposed in light of the aforementioned uniqueness of ultrasound imaging. It can analyze the biliary region, gallbladder, abdominal lymph nodes, liver, descending duodenum, duodenal bulb, stomach, pancreas, and pancreatic lymph nodes, with a total of 10 ultrasonic organs, 21 types of sub-categories, and 3498 images. Binarization, histogram (HS) equalization, median filtering, and edge enhancement algorithms are used to preprocess the images. The data set is trained with the enhanced YoloV4 convolutional neural network (CNN) algorithm, and high precision is detected in real time. At long last, the normal precision of this calculation has reached 91.59%. This paper's algorithm has the potential to make up for the shortcomings of the original image detection system's manual detection, increase detection efficiency, reduce detection errors, and encourage the development of automated and intellectual detection in the health field.**

***Keywords*: Convolutional neural network, image processing, data mining, endoscopic ultrasonography.**

1. Introduction

Therapeutic ultrasound (US) imaging is widely used for diagnosis. Therapeutic imaging has entered a new advance process that combines deep learning and big data since the dawn of the big data era. Every day, a large number of therapeutic images and in sequence about both regular and irregular organs in various parts are imported into the therapeutic US image library. The extraction of therapeutic data from the existing technology is very important and necessary. Professionals are able to determine the kind of patient they are treating and the degree of harm they have caused based on this crucial information, as well as the appropriate treatment strategies based on the image data. Currently, common therapeutic imaging equipment can manually color and mark images. However, doctors still need to have experience and knowledge to make decisions about specific organs and parts. Computer technology and artificial intelligence have made steady inroads into the therapeutic field in recent years. The direct image enhancement technology that was proposed by Chen Yan et al. can recognize therapeutic images. It is distinct from the technology used for indirect image enhancement in the classification of therapeutic images [1]. Li Bo and co established a Gaussian scale space to classify images in a therapeutic image categorization algorithm based on multi-feature combination in extent space [2]. An improved segmentation algorithm for digital morphology therapeutic images that combined morphology and filtering operations was described by Li Zuoyong [3]. Picture segmentation algorithms gradually gain advantages in the field of deep learning. Wang Li and cointroduced a Faster-RCNN neural network-based algorithm as an additional diagnosis [4]. Bakalo Ran and others for therapeutic image recognition, a weak semi-supervised detection dual-branch deep learning network algorithm was proposed that significantly improves recognition [5].

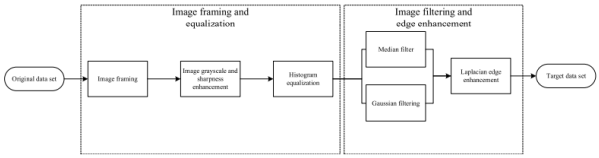

Fig. 1. Preparing the US endoscopy data set for processing

A border perception deep learning system with multiple layers, a recognition system based on the Frequency Domain, Time Domain, and CNN methods, and a deep learning image classification algorithm that substitutes wavelet decomposition for the convolution(conv) kernel are all described in the literature [6]-[9]. This kind of algorithm is currently mostly trained on a lot of data, requires a lot of calculations, and is expensive to run. It's also hard to use all of its advantages in US endoscopic detection with only a few key frames because there aren't enough detection samples. The most prevalent issues in the current US endoscopy research are as follows: (1) US picture itself is generally fluffy, and the critical edges with not many elements make it hard to be recognized during handling. (2) Therapeutic US imaging creates an image by making use of the US's substantial properties. It is difficult to get rid of the interference and its imaging is inadequate for detecting organs to contain gas and bones. It takes a lot of professional knowledge to help make a decision, which is likely to lead to a wrong diagnosis. (3) Therapeutic US images have a significant labeling error, which is difficult to quantify and necessitates the assistance of trained therapeutic professionals. A therapeutic US detection and recognition algorithm based on CNN and data mining is proposed in this paper to address the aforementioned

*Corresponding author: ravindra439@gmail.com

issues. This study uses data mining and image methods to improve image quality and optimize deep learning algorithms, in contrast to previous ones. As a result, effect of fine-grained image classification is enhanced. This algorithm performs key frames on 10 organs in the biliary region, including the gallbladder, liver, descending duodenum, duodenal bulb, stomach, pancreas, and esophagus, using the existing data set. Division, the creation of multiple tag libraries for various parts of particular organs, In addition to the realization of 21 organ subtype recognition. The improved algorithm eliminates the training issue brought on by the diverse organ environment and significantly reduces the scene's requirements. The impact of issues with image quality like blur on recognition is significantly reduced when image boundary improvement with filtering algorithms are used to get better image transparency and identification. The issue of having fewer key frames is resolved and sample capacity is increased by employing OpenCV-based random image transformation and data mining technology. Finally, the improved YoloV4 CNN is used to train and learn the image. This improves image recognition accuracy and addresses the low efficiency and bias of manual US endoscopic image judgment.

Table 1
The sub-frame image data set of US 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

2. Data Set Image Preprocessing

Focuses on the subsequent endoscopic US image quality issues: a) The noise and interference brought on by the image's ineffective burr. b) Image framing-induced ghosting and blur. c) The image's global Gaussian noise and local salt-and-pepper noise. The therapeutic US endoscopy image data set is optimized in this paper using an OpenCV-based image preprocessing algorithm to preserve large frequency signals, eliminate unrelated noise, raise image excellence, also make it simpler to identify the data set's features. Fig. 1 depicts the primary processing flow.

A. Attainment of Data Set

Here currently only some data sets for therapeutic US image classification annotation. The popular Therapeutic Publishing House's therapeutic US endoscopy training video serves as the basis for the framing and annotations in this article.

Dutzendes of primitive endoscopic images, together with the pancreas, gallbladder, and others, have been therapeutic ly classified and annotated in the video data, which serves as a reliable and scientific image source for the data set's building. A sufficient number of frames are obtained through framing the video and image to generate an image data set. The period is set to 1 to 5 frames and the videotape framing algorithm is used for frame taking out during framing. As the research and test objects, ten normal organs, including the pancreas and stomach, were chosen for this study, and 3498 pictures with valuable frames were eventually extracted. Table 1 displays the endoscopic US image data division Table.

In order to use the obtained images for image categorization training, definite image enhancements and repairs are essential [10], [11]. In order to still identify the individuality of a specific organ in dissimilar orientations in a complex scene, data set should be further processed toward eliminate the obstruction frame made up of suds and additional interference substance in each organ [12]. Twenty-one separated similes of organs are labeled using labeling, a revelation instrument for generating information locate labels, and ROI (Region of Interest) frame selection in this study. A data set in the VOC format can be obtained by creating the corresponding xml file. Fig. 2 depicts a portion of the pancreatic endoscopic US image set following framing.

B. Image Graying and Enrichment

Data set's color boundary consists of three RGB channels. Its uniqueness only indicates the optical uniqueness of picture and cannot replicate the target object's morphology. The redundant information in the RGB image will result in an increase in the number of features and calculations wanted to identify ultrasonic internal organs. The picture must be integrated to the GS(GS) variation sequence in the direction of diminish the quantity of estimate required later [13]. In this study, the average dimension method is used.

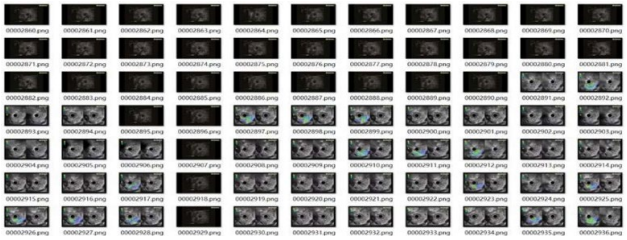

Fig. 2. Endoscopic ultrasound image set of the pancreas

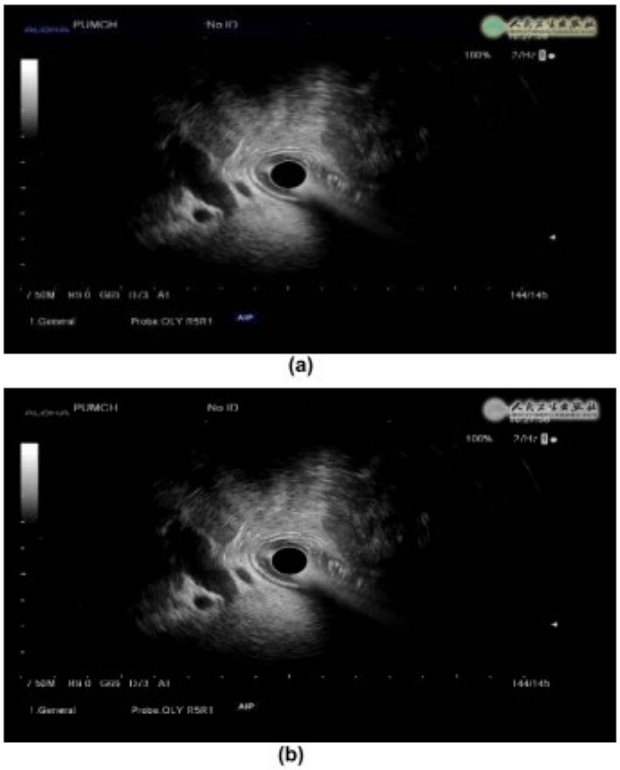

Fig. 3. a) The original endoscopic image; b) Following sharpening

On top of recipe, gray(i,j) addresses dim worth of line I with section j. R, G & B individually address the worth of every channel, along with the typical worth is utilized as the dim worth of the pixel. Some of the therapeutic US images will appear ghosted and blurry after they have been broken up into frames. The recognition of the image will be hampered and made more difficult by this kind of noise.

Later picture is grayed absent, the PIL collection in Python is used to improve the image's sharpness and quality to address the aforementioned issues [14]. When compared to the image in Fig 3, the roughness of the border portion of the picture (b) is considerably improved. Pro images judged based on inside border signals of organ endoscopic picture, it has a better feature enhancement effect.

C. Histogram Equalization

Brightness distribution in a GSd image is uneven. It is essential to convert a picture with a recognized GL allocation into an equally scattered GS picture in order to enhance the contrast and overall distribution of pixel GLs in the image. The series of pixel allocation is expanded, the dissimilarity of the picture is enhanced, and the irregular distribution of pixels is improved. The following steps are required to obtain the HS's remapping distribution: At each gray level, calculate the probability density function as follows:

$$\Pr(r_k) = n_k/n \quad (1)$$

In the preceding formula, Pr (r_k) is the probability that rk falls below the gray level, n_k is the quantity of pixels below the gray level, in addition to n is the total pixels in the picture as a intact. After gray level mapping, get the worth of distribution function by using the mapping relationship.

$$S_K = T(r_k) = \sum_{i=0}^{k} P_r(r_i) \quad (2)$$

S_k is one of them; it denotes the probability distribution function's value below gray level starting r_j to r_k. The therapeutic image Rearrangement of Gray Levels (GL) is realized when the HS is equalized later than converting to the regular Gray value (GV) during the mapping collection. This means that the quantity of pixels will no longer be distributed near the black line but rather on the GS of 0-255. Fig. 4 shows a HS before and after equalization. The gray level is represented by the abscissa, and the number of pixels in the gray level is represented by the ordinate, as exposed in Fig. 4.

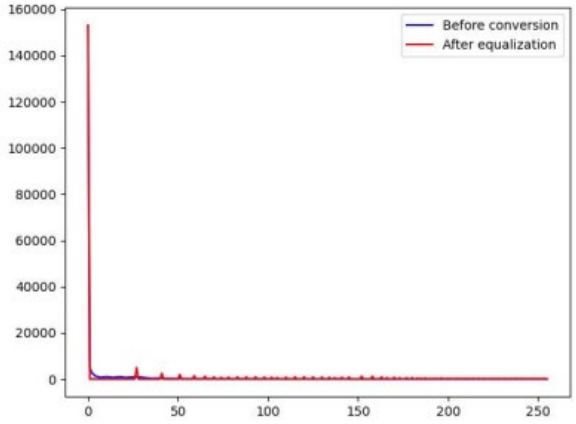

Fig. 4. Evaluation of HSs previous and later equalization

The GV and pixel information are used to create a visual gray HS later the pancreas image has been equalized. The majority of pixels are scattered in the 0 to 50 region prior to equalization. GL remapping and image equalization are achieved when equalization results in a circulation of the GL HS that is more reasonable, ranging from 50 - 255 [15].

D. Picture Filtering Algorithm

US imaging primarily makes use of the acoustic uniqueness of US that are reflected in various tissues and organs. As a result, it is possible to distinguish distinct organ contours. Speckle noise is the primary cause of interference in the image; this category of noise cannot be removed physically and must be processed using imaging [16]. Noise points in this article are the US endoscopic images' graininess and glitch interference.

Gaussian denoising and median filtering are used so as to lessen impact of this category of noise [17].

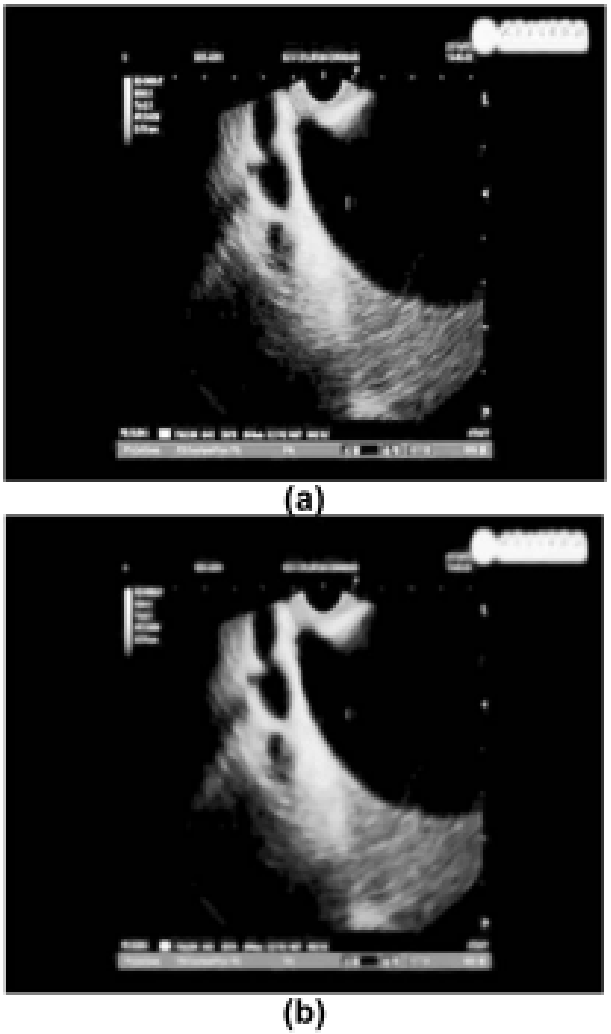

Fig. 5. a) Prior to removing; b) Following removing

Using a 2-D Gaussian filter to create 3x3 filter when located the template, Gaussian denoising blurs the screen blare, performs task calculations while the module is moving, along with performs all of these tasks simultaneously.

$$H_{i,j} = \frac{1}{2\pi\delta^2} e^{\frac{-(i-k-i)^2+(j-k-1)^2}{2\delta^2}} \tag{3}$$

Among them, the serial port's size is represented by k. $H_{i,j}$ is the significance of rows i along with j in the mask, and Gaussian filtering is applied to all positions of the pixels as the window moves. However, noise of any kind cannot be removed using Gaussian filtering.

In return to this situation, this article employs median altering like an additional technique to further remove the image's salt and pepper(S&P) noise while preserving the image's edge features. The GL of the pixels in the mask are sorted when the template is moved, and the new GV of the central pixel is determined by taking the median importance of the region enclosed by the filter. Fig. 5 depicts the refined outcome. The processed Fig. (b)'s burr and S & P noise are significantly reduced in comparison to Fig (a), and the combined masking effect is satisfactory.

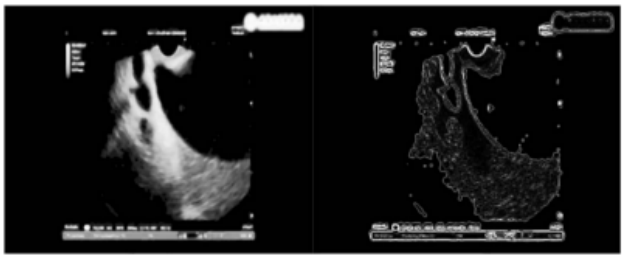

Fig. 6. (a) Prior to cutting the edge; (b) Following edge extraction

E. Enhancement and Edge Extraction Algorithm

The main purpose of endoscopic image recognition is to evaluate the ultrasonic appearance uniqueness of various organs and casing structures. Changes in the GL of pixels along the edges and contours are the primary high-frequency information. As a result, the recognition effect will rely heavily on the optimization and pulling out of border information. The gray jump value of the image's edge pixels is calculated using a 3x3 conv kernel constructed with the Laplace operator in this paper [18], [19].

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{4}$$

2nd order GS differential value of the picture at (x, y) coordinates is represented by 2 f. The shortened version is:

$$\nabla^2 f = -4f(x, y) + f(x-1, y) + f(x+1, y) + f(x, y-1) + f(x, y+1) \tag{5}$$

The calculation can also be carried out using the Laplace conv filter:

$$H = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \tag{6}$$

The Laplace operator conv template is represented by H. it is seen in Fig. 6; the image's edges can be extracted using the conv process. The edge elements of the picture are protected, and the highlights of the picture are additionally improved.

3. Convolutional Neural Network Training and Prediction

The research concept depicted in Fig.7. There are three main parts to the research path: Yolo V4 detector set-up structure, data mining and expansion, and system evaluation and enhancement. This article selects the greatest parameters and

regularization techniques to debug and optimize the algorithm in order to enhance its robustness. By the Tensor flow frame work to create a YoloV4 CNN, the simulation environment runs on Windows 10. The project's acceleration platform is a RTX Quadro 4000 graphics card, and fifty iterations of training take 155 minutes each.

A. Building of the Yolov4 System Framework

As depicted in Fig. 8, the building of the YOLO V4 system detector, which primarily consists of the backbone system, SPP system, PANet system, and prediction system [20], is crucial. The spine system is utilized to separate highlights, SPP takes part in pooling as an extra part [21], and PANet primarily partakes in include combination [22]-[25]. The primary purpose of the Yolo head component is forecasting. SPP utilizes 1 * 1, 5 * 5, 9 * 9, 13 * 13 pooling to check the element layer conv and pooling. The receptive field can be increased and the upper- and lower-layers' uniqueness can be as much as possible separated by this structure. PANet is a segmentation algorithm used to achieve repeated feature extraction and enhance the features of the target detection object. The Yolo system was used to create the following feature layer dimensions: (52, 52, 256), (26, 26, 512), and (13, 13, 1024). On this basis, the parameters are changed, and regularization and data improvement are added at the same time.

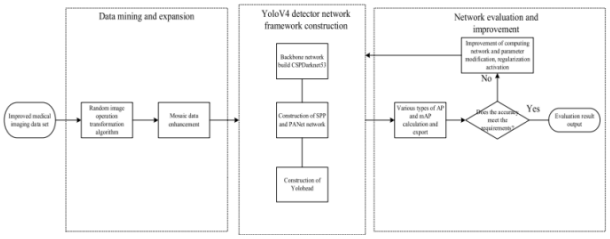

Fig. 7. Basic roadmap for CNN research

A CNN suitable for endoscopic US samples is created through training. The improved system was able to successfully identify endoscopic image features and performed well in the experiment.

B. Preparation for Training along with Data Mining

The endoscopic US images in this paper include together useful and quality frames. Organs that have been correctly marked in the image and have fewer interference items are the characteristic frames. However, there are only a few of these frames. Data enhancement algorithms are used to increase the diversity of image features and avoid over-fitting the model in order to solve this issue [26]. a) Data enhancement makes use of the OpenCV function library to process the frames in the training set by setting the random state flag and perform random inversion, rotation, scaling, and other operations on the training data. b) The mosaic data enhancement method is explained, in which image stitching is used to combine four images' partial pixel areas into a single image. That is, four images are calculated each time, resulting in a small batch (number of images per training round), which can increase data calculation efficiency. c) The neural network will constantly update the model information, calculate the loss value, and choose the best target optimization solution based on the loss function during forward propagation and backward calculation. Continuously optimize and determine the best solution by employing stochastic gradient descent. However, this gradient descent strategy will fall into the trap of local optimality when confronted with multimodal functions, and the local optimal solution is not the overall optimal solution. The cosine annealing decay, which appears in Fig. 8, is introduced to resolve this issue. YoloV4's network framework's schematic diagram.

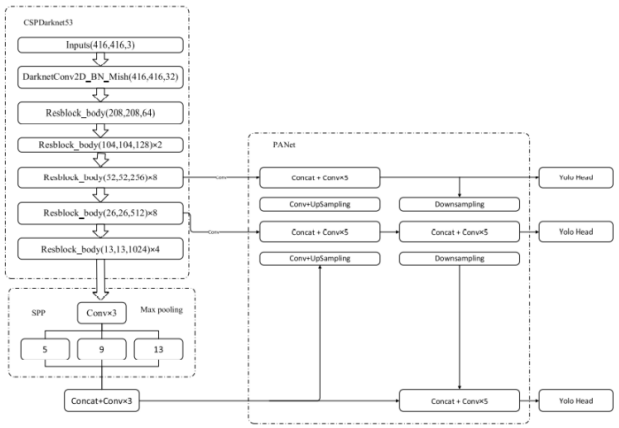

Fig. 8. Plan of YoloV4 classification framework

When it enters the local optimal solution, the computation is restarted pending the overall optimal solution is discovered. The basic idea is:

$$\eta_t = \eta_{\min}^j + (\eta_{\max}^j - \eta_{\min}^j)(1 + \cos\left(\frac{T_{current}}{Ti}\pi\right) \quad (7)$$

The jth key value, j max, and j min of the preceding formula represent the maximum and minimum learning rates, respectively. The no. of epoch rounds at this time being performed is shown by $T_{current}$. After each restart, the learning rate will be adjusted to reflect the latest information. It is increased automatically by multiplying it by a predetermined value after each restart. After each restart, the learning rate will be adjusted in this manner. To enlarge the data of a inadequate frame of US images, random image changes, mosaic data enhancement, and cosine annealing shrinking are utilized to reduce the quantity of result and prevent over-fitting of the result.

C. Training and Detection of NN's

The endoscopic US image data set consists of both processed and unprocessed pictures, each with a total of 3498 effective frames and picture dimension of 608 x 608 pixels. The ratio of the training set to the test set is 8:2, the number of batches in each iteration be 2, and the total no. of iterations is shown in Fig 9.

The number of training rounds in an epoch is fifty. The CNN level, the pooling level, and the fully connected level are mostly updated and calculated iteratively during the CNN training

process. The principal steps are:

a) Produce an XML file for the associated image data set by marking the identifiable information ROI in VOC2007 format.
b) Utilize the CSPDarknet53 system to combine the image features, reduce and extract valuable features, and compute the loss significance using the loss function.
c) The gradient update parameters w and b are calculated using batch gradient descent, and regularization is used to decrease weight and prevent above fitting.
d) Weight file is output while the Loss significance converges to the estimated effect.
e) Forward propagation process and the weight file are used to locate the image's target.

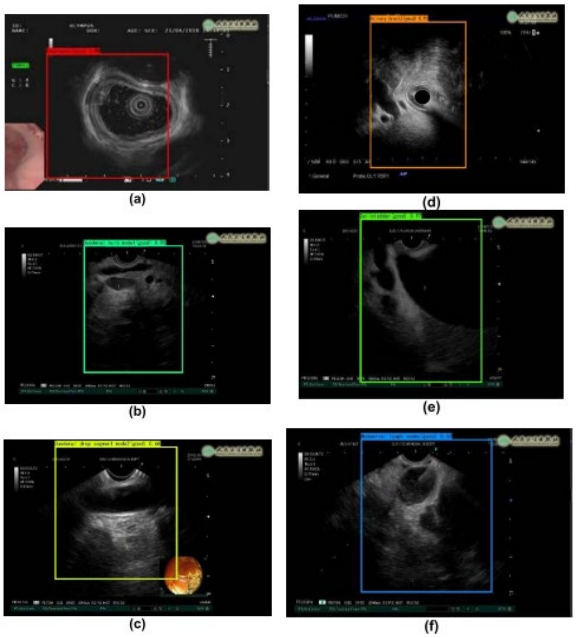

Fig. 9. Recognition results of organs (partial)

Fig. 9 depicts identification results of a few organs following training in accordance with the preceding steps. Fig. 9 depicts a) the duodenal bulb; b) the stomach; c) the bile duct; d) the duodenum that descends; e) lymph nodes in the abdomen; f) The liver. The identification result is quicker to recognize and matches the actual situation. The solitary picture takes on 0.1 milliseconds.

4. Result Analysis

The parameter junction ratio IOU serves as primary foundation for determining the quantity of overlie that exists among the target recognition frame and the actual frame in order to evaluate the improved YoloV4 NN's detection performance and accuracy.

$$IOU = \frac{s_{\Lambda}}{s_u} \tag{8}$$

The denominator represents the combination region of the recognition edge and the real edge, the IOU represents the intersection ratio, and numerator represents the junction area of the detection edge and the real edge. The following formula can be used to determine the model's Precision P:

$$p = \frac{TP}{TP + FP} \tag{9}$$

TP indicates that a positive sample has been detected, while FP indicates that a negative sample has been detected. Similar to the Precision, the model calculates the recall rate as follows:

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

Along with, FN indicates that detected outcome is actually a +ve sample despite being a negative sample. The recall reflects quantity of the positive sample's correctly classified portion to the total number of positive sample parts. The model can only be evaluated using a single index, and the effect of target detection should be evaluated using the confidence level. Confidence is the corresponding parameter for the classification of positive and negative samples. The precision(P) is depicted by the Recall, P curve, and the introduction of confidence results in the union of the precision P and the recall rate R. Fig.10 depicts the P statistics for healthy pancreatic lymph nodes. Among them, the abscissa addresses the review rate, and the ordinate addresses the accuracy rate. The P of normal pancreatic lymph nodes is relatively high, reaching 95.52 percent, and as the positive sample threshold point moves to the left, the value of precision is initially very close to 1, while the value of recall is very close to 0. The term "average precision" (mAP) refers to the mAP of all trained objects. mAP serves as an important foundation for determining the model's quality. Its formula is as follows:

$$mAP = \frac{\sum_{i=1}^{c} AP_i}{C} \tag{11}$$

Among them, P is the P value of the i-th category, and c is the total number of sample classifications. Using the aforementioned guidelines, evaluate the model in this article and plot the mAP of 21 data set subcategories in Fig. 11.

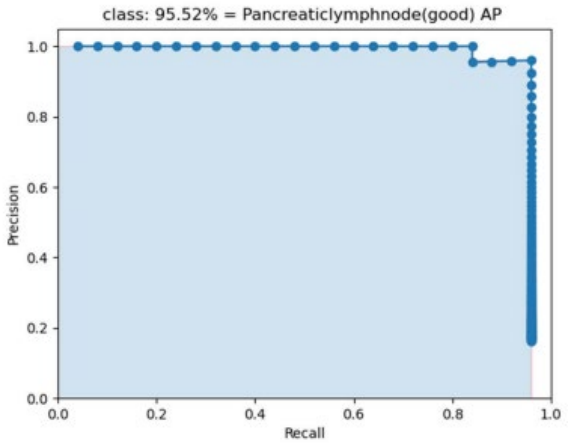

Fig. 10. P statistics of normal pancreatic lymph nodes

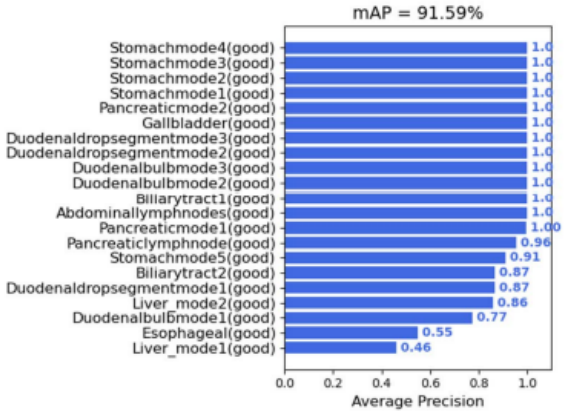

Fig. 11. mAP statistics for 21 sub-categories

The P value is represented by the abscissa, and the category of US endoscopic images is represented by the ordinate. The mAP has a mean precision of 91.59 percent. The outcomes demonstrate that the recognition algorithm is extremely accurate.

5. Conclusion

Based on data mining and deep learning, this manuscript proposes a system for US endoscopic image recognition. a) The image preprocessing algorithm is utilized to address issues with image quality brought about by US interference and framing in endoscopic US images. b) The OpenCV function library's data enhancement algorithm is utilized to increase the number of training sets, boost the exercise model's strength, and address issue of else only some valuable frames in the existing exercise set. c) The Mosaic algorithm is introduced to address the issues of existing algorithms' low recognition efficiency and high computational costs while also increasing its training efficiency. d) To avoid overfitting the results, the cosine annealing algorithm is utilized, further strengthening the algorithm's robustness. The mAP of 21 sub organs reached 91.56% through simulation verification. This algorithm is capable of accurately identifying organs in endoscopic US images, providing new methods for identifying additional organs and cancerous organs in the future.

Appendix

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.

References

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

[2] L. Bo, C. Peg, L. Wei, and Z. Dazhe, ‘‘Therapeutic 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, ‘‘Therapeutic 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 therapeutic imaging via deep dual branch net,’’ Neurocomputing, vol. 421, pp. 15–25, Jan. 2021.

[6] F. An and J. Liu, ‘‘Therapeutic 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 therapeutic 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 therapeutic 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 realtime 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, ‘‘GL 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.