



Article Title: Endoscopic Ultrasound Image Recognition Using Improved You Only Look Once (Yolov4) Convolutional Neural Network

Endoscopic Ultrasound Image Recognition Using Improved You Only Look Once (Yolov4) Convolutional Neural Network

Akhila KS¹, Anuja SB²

¹Student, Department of MCA, Narayanaguru College of Engineering, Anna University

² Asst. Professor, Department of MCA, Narayanaguru College of Engineering, Anna University.

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 analyse 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 pre-processed 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.

Keywords: Endoscopic, Histogram Equalization, Yolov4, Convolutional Neural Network, Image Enhancement, Ultrasound medical imaging.

1 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



Article Title: Endoscopic Ultrasound Image Recognition Using Improved You Only Look Once (Yolov4) Convolutional Neural Network

image enhancement technology proposed by Chen Yan et al. is different from the indirect image enhancement technology used in the classification of medical images. 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. Li Zuoyong described an improved segmentation algorithm for digital morphology medical images, which segmented medical images by combining morphology and filtering operations. 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. 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. Literatures 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.

2 Recent Work

2.1 Opposition-Based Laplacian Equilibrium Optimizer with Application in Image Segmentation Using Multilevel Thresholding, Shail Kumar Dinkar, K. Deep, Shail Kumar Dinkar, K. Deep, 2021.

This paper proposes a modified version of freshly developed Equilibrium Optimizer (EO) for segmentation of gray-scale images using multi-level thresholding. Laplace distribution based random walk is utilized to update the concentration of search agents around equilibrium candidates (best solution) towards to attain optimal position (equilibrium state) for achieving better diversification of search space. An opposition based learning (OBL) mechanism is then applied with hybridization of the varying acceleration coefficient to the best solution for accelerating exploitation at a later phase of each iteration. The performance of proposed Opposition-based Laplacian Equilibrium Optimizer (OB-L-EO) is validated using test suites containing benchmark problems of wide varieties of complexities. Various analyses are conducted including Wilcoxon ranksum test for statistical significance, convergence curves and distance between solution before and after applying modification strategies. Finally, the proposed OB-L-EO is employed for image segmentation by utilizing Otsu's interclass variance function to obtain optimum threshold values for image segmentation. Several new optimization algorithms have been proposed which states that no particular optimization algorithm is capable of solving all problems of all kind of complexities. It inspires the researchers to design new metaheuristic algorithms. The performance of the proposed algorithm is verified by



Article Title: Endoscopic Ultrasound Image Recognition Using Improved You Only Look Once (Yolov4) Convolutional Neural Network

determining mean value of interclass variance and peak signal to noise ratio (PSNR). The obtained results are then compared and analysed with other metaheuristics algorithms to show superiority proposed.

2.2 A Review on Speckle Noise Reduction Techniques in Ultrasound Medical images based on Spatial Domain, Transform Domain and CNN Methods, S. Pradeep and P. Nirmaladevi, 2021.

Ultrasonography is non-invasive and painless. In Ultrasonography the images are often affected with Speckle noise. It is a multiplicative noise. To help the doctors to identify the abnormalities properly there are several methods to diagnose as speckle is a major problem. This paper gives details about popular spatial domain, transform domain, CNN techniques for despeckling in ultrasound images. Transform domain methods like Wavelet methods, Curvelet methods, Bayes Shrink methods are prominent among many researches. Wavelet transforms is a tool used now a day for removal of speckle in an image. CNN based image denoising model where both image and noise are modelled at a time with Image decomposition concept for better performance of the model. A deep convolutional neural network which has unsampled shearlet transform to decompose the image to various sub bands of low, high frequency.

Wavelet Decomposition Facilitates Training On Small Datasets for Medical Image Classification by Deep Learning, A. H. Masquelin, The adoption of low-dose computed tomography (LDCT) as the standard of care for lung cancer screening results in decreased mortality rates in high-risk population while increasing false-positive rate. Convolutional neural networks provide an ideal opportunity to improve malignant nodule detection; however, due to the lack of large adjudicated medical datasets these networks suffer from poor generalizability and overfitting. Using computed tomography images of the thorax from the National Lung Screening Trial (NLST), discrete wavelet transforms (DWTs) is compared against convolutional layers.

2.3 Medical Image Segmentation Algorithm Based On Multilayer Boundary Perception-Self Attention Deep Learning Model, F. An and J. Liu, 2021

Traditional medical image segmentation methods have problems such as low segmentation accuracy and low adaptive ability. Therefore, many scholars have proposed a medical image segmentation method based on deep learning, which has achieved good results in the field of medical image segmentation. However, this type of method has the following problems in the application process: (1) Medical image segmentation target boundary positioning problem. Constrained by factors such as medical image contrast, heterogeneity, and boundary resolution, existing convolution models still cannot accurately locate boundaries. (2) Deep adaptability of deep learning network structure to medical images. Because medical images have more distinct and different feature information than natural images, the current deep learning-based medical segmentation methods have not fully considered this feature. In view of this, this paper proposes a multi-level boundary-aware RUNet segmentation model. The network structure



Article Title: Endoscopic Ultrasound Image Recognition Using Improved You Only Look Once (Yolov4) Convolutional Neural Network

consists of a U-Net-based segmentation network and a multi-level boundary detection network. It can solve the problem of boundary positioning. At the same time, in order to solve the problem of poor adaptability of deep learning network structures to medical images, this paper proposes to introduce a new interactive self-attention module into deep learning models.

2.4 Weakly and Semi Supervised Detection in Medical Imaging via Deep Dual Branch Net, R. Bakalo, J. Goldberger, 2021.

This study presents a novel deep learning architecture for multi-class classification and localization of abnormalities in medical imaging illustrated through experiments on mammograms. The proposed network combines two learning branches. One branch is for region classification with a newly added normal-region class. Second branch is region detection branch for ranking regions relative to one another. This method enables detection of abnormalities at full mammogram resolution for both weakly and semi-supervised settings. A novel objective function allows for the incorporation of local annotations into the model. It presents the impact of the schemes on several performance measures for classification and localization, to evaluate the cost effectiveness of the lesion annotation effort. Evaluation was primarily conducted over a large multi-center mammography dataset of ~3,000 mammograms with various findings. The results for weakly supervised learning showed significant improvement compared to previous approaches. The result shows that the time consuming local annotations involved in supervised learning can be addressed by a weakly supervised method that can leverage a subset of locally annotated data. Weakly and semi-supervised methods coupled with detection can produce a cost effective and explainable model to be adopted by radiologists in the field.

3 Proposed Methodology

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 lymphnodes, 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



Article Title: Endoscopic Ultrasound Image Recognition Using Improved You Only Look Once (Yolov4) Convolutional Neural Network

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.

4 System Architecture

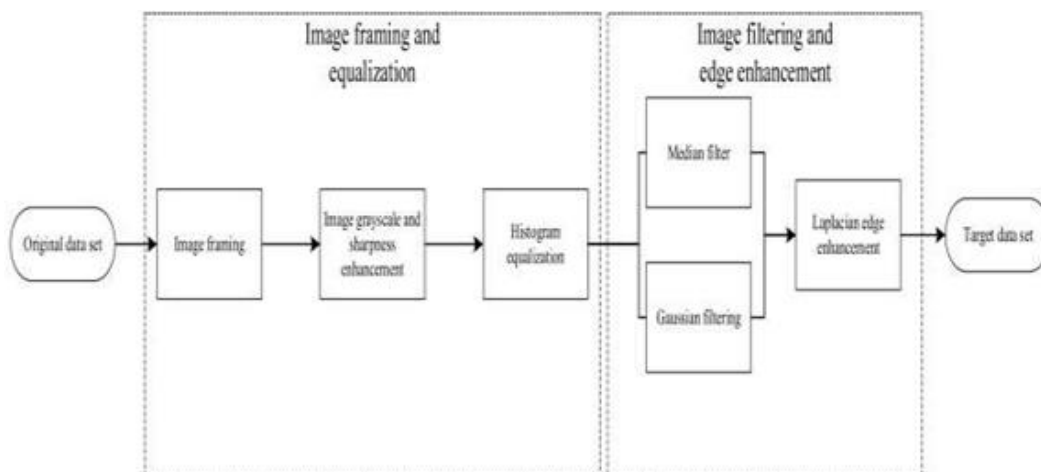


Figure 1: System Architecture

5 Modules

5.1 Acquisition of Data Set

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 obtained images cannot be directly used for image classification training, and certain repairs and image enhancements are required. The data set needs to be further processed to remove the interference frame composed of bubbles) and other interference items in each organ, so that the characteristics of a certain organ indifferent orientations can still be identified in a complex scene. In this paper, labeling (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.



Article Title: Endoscopic Ultrasound Image Recognition Using Improved You Only Look Once (Yolov4) Convolutional Neural Network

5.2 Image Graying and Enhancement

The color interface in the data set is composed of RGB (Red, Green, Blue) threechannels, 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. 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, (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. 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.

5.3 Histogram Equalization

The gray scaled 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

$$P_r(r_k) = n_k/n \quad (2)$$

In the above formula, $P_r()$ 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.

$$S_K = T(r_k) = \sum_{i=0}^k P_r(r_i) \quad (3)$$

Among them, S_k represents the value of the probability distribution function under the gray



Article Title: Endoscopic Ultrasound Image Recognition Using Improved You Only Look Once (Yolov4) Convolutional Neural Network

level from r_j to. 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. 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.

5.4 Image Filtering Algorithm

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. 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. 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.

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

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

5.5 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.

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (5)$$

Among them, $\nabla^2 f$ represents the second-order gray-scale differential value of the image at



Article Title: Endoscopic Ultrasound Image Recognition Using Improved You Only Look Once (Yolov4) Convolutional Neural Network

(x, y) coordinates. The simplified expression is:

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

(6)

Further use the Laplace convolution template to realize the calculation :(7)

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

(7)

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.

5.6 Convolutional Neural Network Training and Prediction

5.6.1 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, as show in Figure 3.3. The backbone network is used to extract features, SPPparticipates in pooling as an additional part, and PANet mainly participates in featurefusion. 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. 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, and 256), (26, 26, and 512) and (13, 13, 1024). On this basis, the parameters are adjusted, andat the same time, regularization and data enhancement are introduced. The training forms a convolutional neural network suitable for endoscopic ultrasound samples. Theimproved network performed well in the experiment and can effectively identify endoscopic image features.

5.6.2 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, thenumber 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.

The principle is



Article Title: Endoscopic Ultrasound Image Recognition Using Improved You Only Look Once (Yolov4) Convolutional Neural Network

$$\eta_t = \eta_{\min}^j + (\eta_{\max}^j - \eta_{\min}^j) \left(1 + \cos \left(\frac{T_{\text{current}}}{T_i} \pi \right)\right) \quad (8)$$

In the above formula, j represents the j th index value, η^j is the maximum learning rate, and $\eta_{\min}^j$ is the minimum learning rate. 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.

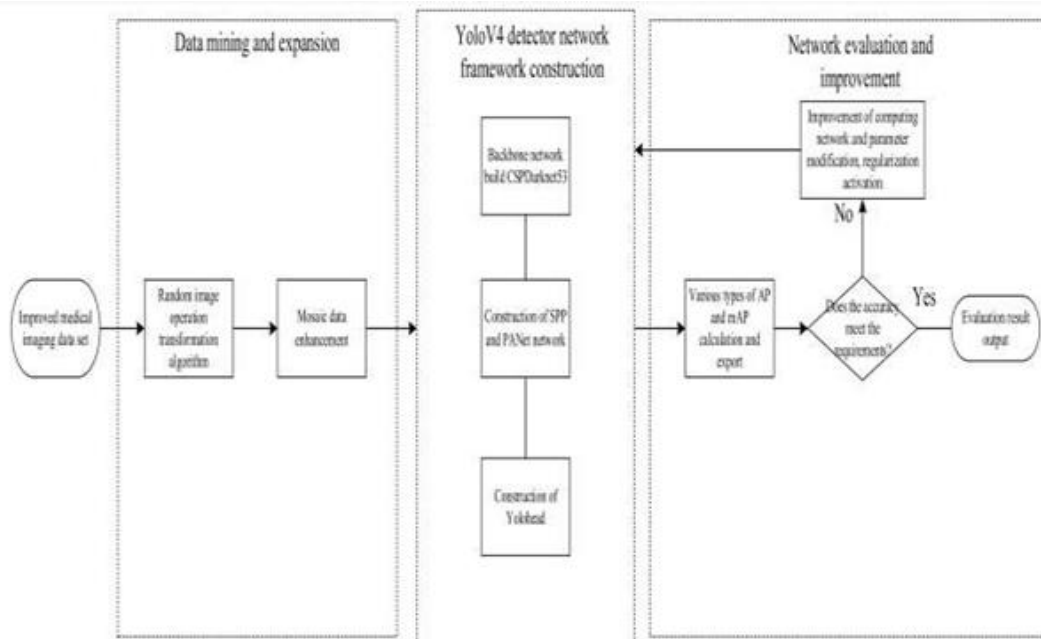


Figure 2: Simplified roadmap for convolutional neural network research

5.6.3 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.



Article Title: Endoscopic Ultrasound Image Recognition Using Improved You Only Look Once (Yolov4) Convolutional Neural Network

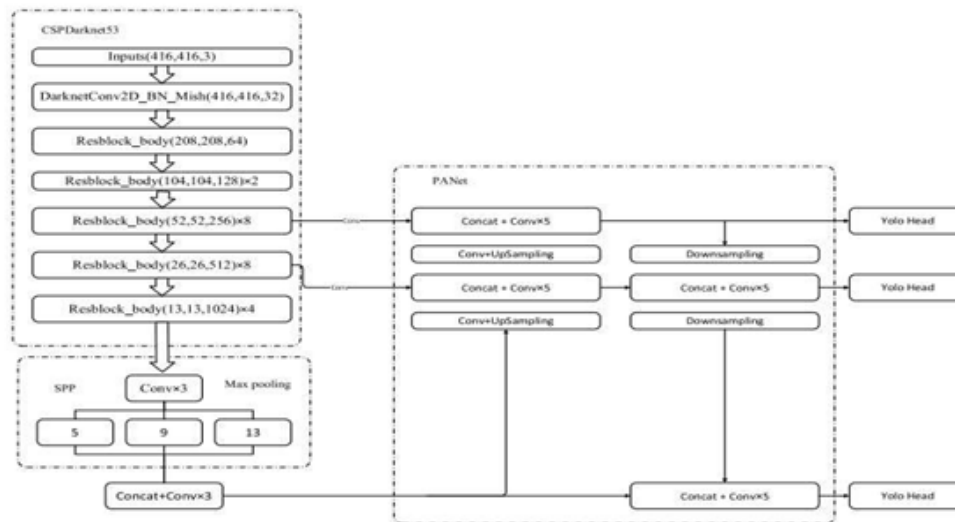


Figure 3: Schematic Diagram of Yolov4 Network Framework.

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:

- Create an image data set in VOC2007 format, mark the identifiable information ROI, and generate the corresponding XML file.
- Convolve the image features through the CSPDarknet53 network, compress and extract useful features, and calculate the loss value under the loss function.
- 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.
- When the Loss value converges to the expected effect, the weight file is output.
- 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.

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

5 Conclusion

This Paper proposes a method of ultrasound endoscopic image recognition system based on data mining and deep learning. 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. 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. 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. The cosine annealing algorithm is



Article Title: Endoscopic Ultrasound Image Recognition Using Improved You Only Look Once (Yolov4) Convolutional Neural Network

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%.

References

1. Yan Chen; Geng Guohua, Year: 2007, "Application of a direct image enhancement method in medical image classification," Computer. Applications. Software, Vol: 6, pp. 26–32.
2. Bo Li; C. Peg; Li Wei; Zhao Dazhe, Year: 2013, "Medical image classification based on multi-feature fusion in scale space", Computer. Applications, Vol: 33, no. 4, pp. 1108–1114.
3. H. Jinmei; L. Zuoyong, Year: 2011, "Medical image segmentation based on improved mathematical morphology algorithm", Computer. Simulations, Vol: 28, no. 5, pp. 299– 302.
4. Li Wang; Lu Huijuan; Ye Minchao; Y. Ke, Year: 2018, "A cancer image detection method based on Faster-RCNN", J. China University. Metrology, Vol: 29, no. 2, pp. 136–141.
5. Bakalo Ran; Jacob Goldberger; Rami Ben-Ari, Year: 2021, "Weakly and semi supervised detection in medical imaging via deep dual branch net", Neurocomputing, Vol: 421, pp. 15–25.
6. An Feng-Ping; Jun-E. Liu, Year: 2021, "Medical image segmentation algorithm based on multilayer boundary perception-self attention deep learning model", Multimedia Tools Applications, Vol: 80, pp. 15017–15039.
7. Pradeep S P. Nirmaladevi, Year: 2021, "A review on speckle noise reduction techniques in ultrasound medical images based on spatial domain, transform domain and CNN methods", in Process. IOP Conference, Material. Science. Engineering, Vol: 1055, no. 1.
8. A. H. Masquelin; N. Cheney; C. M. Kinsey; J. H. T. Bates, Year: 2021, "Wavelet decomposition facilitates training on small datasets for medical image classification by deep learning", Histochem. Cell Biology, Vol: 155, no. 2, pp. 309– 317.
9. S. Ren; K. He; R. Girshick; J. Sun, Year: 2017, "Faster R-CNN: Towards realtime object detection with region proposal networks," IEEE Trans. Pattern Analysis. Machine. Intelligence, Vol: 39, no. 6, pp. 1137–1149.
10. L. Huilan; Y. Hui, Year: 2020, "Image classification method based on iterative training and ensemble learning", Computer. Engineering. Design, Vol: 41, no. 5, pp. 1301–1307.
11. W. Hao; Z. Ye; S. Honghai; Z. Jingzhong, Year: 2017, "Overview of image enhancement algorithms," China, Vol: 10, no. 4, pp. 438–448.



Article Title: Endoscopic Ultrasound Image Recognition Using Improved You Only Look Once (Yolov4) Convolutional Neural Network

12. W. Teng; B. Leping; Y. Zhonglin; L. Qifeng; Y. Xuanfang, Year: 2017, “A flame recognition method based on image similarity of consecutive frames,” J. Naval Universal. Engineering, Vol: 29, no. 4, pp. 48–52.
13. Z. Guihua; F. Yanbo; L. Weidong, Year: 2007, “Grayscale of image processing and acquisition of characteristic regions,” Qiqihar University. J, Vol: 4, pp. 49–52.
14. K. He; X. Zhang; S. Ren; J. Sun, Year: 2015, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” IEEE Transactions. Pattern Analysis. Machine.Intelligence, Vol: 37, no. 9, pp. 1904–1916.
15. M. Shuiqing; Z. Jing; H. Changjun, Year: 2017, “Research on image enhancement method based on median filtering and histogram equalization,” Wireless InternetTechnology, Vol: 22, pp. 106–107.