Scholar Publishing Group
International Journal of Multimedia Computing
https://doi.org/10.38007/IJMC.2022.030107.
ISSN 2789-7168 Vol. 3, Issue 2: 52-59

Research on B-ultrasound Video Image Enhancement and Detection System Based on Computer Artificial Intelligence

Tao Zeng, Tao Wu, Jing Zhou*

Medicine College, Jingchu University of Technology, Jingmen 448000, China

200707023@jcut.edu.cn

Keywords: Artificial Intelligence, B-ultrasound Image Enhancement, Convolutional Neural Network, Object Detection, Medical Image Analysis

Abstract: Existing B-ultrasound images suffer from high noise and low contrast, which can easily affect lesion identification accuracy. This paper proposes a B-mode ultrasound video image enhancement and automatic detection system based on computer artificial intelligence. First, adaptive median filtering and histogram equalization are performed on the original video frames to achieve noise suppression and brightness compensation. Secondly, a deep feature extraction model based on a convolutional neural network (CNN) is constructed, incorporating a multi-scale attention mechanism to enhance the fidelity of edge details in target structures. Finally, a segmentation and detection module based on a U-Net architecture is designed to automatically locate and accurately segment tumors or critical tissues. Experiments were conducted on 500 clinical B-ultrasound video samples. The results showed that the proposed system improved the image signal-to-noise ratio (SNR) to 31.3 dB, the mean structural similarity index (SSIM) to 0.932, and the mean average precision (mAP) of object detection to 93.5%. These results validate the effectiveness and practicality of the system for B-ultrasound image visualization and lesion identification.

1. Introduction

B-ultrasound is widely used in clinical diagnosis due to its noninvasive, real-time, and low-cost characteristics. However, its imaging process is affected by acoustic wave scattering and equipment interference, often resulting in strong noise and low contrast. This blurs tissue boundaries and loses detailed features, leading to significant subjective reliance and the risk of misjudgment when interpreting images. Traditional image enhancement and filtering methods often rely on fixed parameters, have limited adaptability to complex tissue textures, and struggle to suppress noise while preserving detail.

Introducing artificial intelligence into B-ultrasound image analysis is of great significance. Deep learning models can capture high-dimensional structural information at the feature level, enabling automated image optimization and lesion identification, thereby reducing manual intervention and visual errors. Building an enhancement and detection system based on intelligent algorithms not only improves the diagnostic visualization of medical images but also provides physicians with more stable decision support, advancing traditional ultrasound imaging towards intelligent and refined capabilities.

The system proposed in this study integrates multi-scale feature extraction and attention-focusing mechanisms in its structural design, combined with the segmentation and detection approach of the U-Net architecture, to achieve efficient coupling of image enhancement and object detection. This model maintains its lightweight design while also accommodating the ability to express structural details, breaking through the processing bottlenecks of traditional methods for complex tissue images and providing a scalable technical solution for the automated analysis of B-ultrasound images.

2. Related Work

The quality of medical images directly affects the accuracy of diagnostic results and plays a key role in patients' treatment decisions. In order to overcome problems such as imaging noise, uneven illumination and loss of details, the research on image enhancement algorithms has been continuously deepened. The related work covers multiple directions from improving traditional methods to intelligent algorithm modeling. Wang and Wang [1] designed an algorithm for low-light color medical image enhancement. The algorithm overcomes the shortcomings of the Retinex algorithm in enhancing low-light images and achieves high-quality enhancement of low-light color medical images. In order to improve the medical image enhancement effect and effectively retain image details, Xu et al. [2] proposed a multimodal rigid medical image enhancement algorithm based on grayscale transformation. The proposed algorithm can obtain more satisfactory medical image enhancement effects, improve image clarity, and significantly enhance image visual effects. Deng et al. [3] studied medical image data enhancement technology. Without significantly changing the appearance of the image, they improved the quality of the original image by adding specific pixel compensation and making subtle image adjustments, thereby improving the image segmentation accuracy. Shangguan and Liu [4] summarized the currently widely used medical image enhancement processing technologies, including traditional image enhancement methods, improved image enhancement methods, fused image enhancement methods and deep learning methods, and then analyzed and summarized the principles, advantages and disadvantages of these methods. Wang et al. [5] proposed a multi-loss hybrid adversarial method to search for effective adversarial samples that may deceive the network, and added these adversarial samples to the training data to improve the network's robustness and generalization ability to unexpected noise perturbations. Dinh and Giang [6] proposed a new algorithm to solve image problems simultaneously. The proposed method significantly improved the quality of input medical images and also significantly improved the efficiency of current medical image synthesis algorithms. Li et al. [7] proposed a passive unsupervised domain adaptive medical image enhancement algorithm, which uses test data to adjust and optimize the enhancement model in the inference stage. Chen et al. [8] comprehensively outlined MRI image post-processing methods based on deep learning to enhance image quality and correct image artifacts. Wu et al. [9] proposed a model based on generative adversarial network (GAN), namely, semi-supervised GAN with preserved anatomical structure (SSGAN-ASP). Goceri [10] studied enhancement techniques for improving the diagnostic performance of different organs (brain, lung, breast, and eye) using different imaging modalities based on deep learning. These studies have provided many ideas for medical image enhancement, but there are still problems such as the algorithm's strong dependence on specific imaging conditions, high model training costs, and insufficient cross-modality adaptability. It is urgent to build an efficient and unified enhancement framework that takes into account accuracy, stability, and real-time performance.

3. Methods

3.1 Image Preprocessing and Enhancement

The adaptive median filter suppresses the obvious speckle noise in B-ultrasound video images. This noise is a multiplicative noise that produces fine bright spots at the edges of tissue structures, interfering with texture judgment. During the processing, the filter size is dynamically adjusted according to the statistical characteristics of the local window, and the median and variance of the neighborhood of each pixel are calculated. When the image grayscale changes dramatically, a smaller window is selected to retain edge details, and the window range is expanded in the smooth area to eliminate noise. This method not only improves the overall smoothness, but also maintains the continuity of the tissue contour, providing a stable input for subsequent image enhancement and feature extraction [11]. The histogram equalization part is used to improve the problem of uneven grayscale distribution and low local contrast. Ultrasound images often have brightness attenuation in deep tissue areas. To compensate for this problem, the pixel grayscale is remapped according to the probability density, and the frequency of occurrence of each grayscale is adjusted by the cumulative distribution function, so that the dark grayscale is improved and the bright details are not over-enhanced. In order to prevent artifacts caused by over-enhancement, the Contrast Limited Adaptive Histogram Equalization method (CLAHE) is used to calculate the brightness mapping relationship separately in each local area, and set the brightness gain threshold to suppress noise amplification.

3.2 Feature Extraction and Structural Detail Enhancement

CNN is used to extract deep features from B-ultrasound video frames and gradually learn complex spatial structural relationships through multi-layer convolution operations. The network input is an enhanced grayscale frame. The shallow convolution layer focuses on capturing texture and local brightness changes, identifying basic morphologies such as tissue interfaces, vascular branches, and glandular separations; the middle convolution layer extracts larger-scale spatial relationships, reflecting the geometric characteristics and echo distribution patterns of the lesion area; the deep convolution layer integrates the information extracted by the previous layer to generate a representation with stronger semantic relevance for identifying potential lesion areas. To prevent feature degradation or gradient disappearance, the network architecture introduces residual connections and normalization operations, so that multi-layer features can be efficiently transmitted and maintain spatial consistency. The introduction of the multi-scale attention mechanism aims to enhance the expression of edges and subtle structures. This mechanism adaptively weights the high-response areas in the feature map by combining channel attention and spatial attention, so that the structural details in the image are concentratedly expressed [12]. The channel attention component extracts channel-level statistical information through global average pooling, and after nonlinear transformation, generates a weight vector to guide the network to focus on the importance of features at different scales. The spatial attention component utilizes local convolution to capture variations in texture distribution and enhance the saliency of structural contours. The multi-scale fusion unit weightedly combines features extracted from different receptive fields, preserving shallow high-frequency texture information while integrating deeper semantic structure. This allows for clear distinction of tumor boundaries, tissue texture, and small lesions in the feature space. This process enables the network to robustly recognize low-contrast regions and complex tissue structures, providing high-resolution, hierarchical feature representations for subsequent U-Net segmentation and detection.

3.3 Detection and Segmentation Model

The segmentation module of the U-Net architecture is responsible for pixel-level analysis of the lesion area. Its structure consists of a symmetrical encoder and decoder. The encoder extracts semantic features layer by layer through continuous convolution and downsampling operations, converting the texture, morphology, and echo distribution in the input B-ultrasound frame into a high-dimensional feature representation. To avoid the loss of detailed information in the feature space, the network uses skip connections to pass shallow local texture features directly to the decoder, allowing edge information to be reconstructed. During the decoding process, deconvolution and splicing operations restore spatial resolution, and features from different levels are integrated at each stage to achieve detailed restoration of tissue boundaries. During training, the model uses a mixed loss function of the Dice coefficient and cross entropy to balance the integrity of the lesion area contour and background suppression, ensuring the model's segmentation accuracy for irregular lesions.

The automatic detection part generates a predicted bounding box of the target area based on the segmentation results, and quantitatively annotates the location and area of the lesion. The system combines the segmentation mask with the probability map distribution to automatically identify areas with drastic grayscale changes and abnormal textures, and uses a post-processing algorithm to delete false positive results [13]. The detection network adds a spatial pyramid pooling structure to the feature layer to adapt to the detection needs of lesions of different scales. This method can achieve target tracking in continuous frames in a video frame sequence and stably identify dynamically changing echo features. For multiple or fuzzy-bounded masses, stable temporal consistency is achieved by fusing the segmentation results of the previous and next frames.

4. Results and Discussion

4.1 Experimental Setup

The experimental part uses 500 real clinical B-ultrasound video samples from a tertiary hospital, covering typical cases of different parts such as the liver, thyroid and breast. After manual screening, each video extracts key frames and removes invalid segments to form a standardized data set. The system uses the original frame before enhancement as the benchmark input and compares the traditional median filter enhancement algorithm, the CLAHE method and the AI enhancement detection system proposed in this paper. The experimental hardware configuration consisted of an NVIDIA RTX 4090 GPU and an Intel Xeon processor, and the software environment was implemented using Python and the PyTorch framework. Evaluation metrics included SNR, SSIM, and mAP. SNR measures the denoising effect and signal fidelity of the image, SSIM reflects the consistency of image structure and brightness contrast, and mAP verifies the accuracy of detection and segmentation. Each set of experiments was repeated five times and the results were averaged to reduce sampling error. Statistical analysis was used to assess the significance of differences between different algorithms.

4.2 Experimental Results

The experiment was carried out in a standardized ultrasound imaging experimental environment, and the acquisition scenes covered multiple types of examinations such as the abdomen, breast, thyroid and superficial tissue. A total of 500 key frame samples of B-ultrasound videos were used. All videos were grayscale standardized and resized, and the frame rate was maintained at 25fps to ensure timing consistency. During the testing phase, three algorithms were run under the same hardware environment: the adaptive median filtering enhancement algorithm, the CLAHE enhancement method and the AI enhanced detection system proposed in this article. The experiment was independently completed on 100 randomly selected groups of test samples, and each group of images was ensured to contain background textures, vascular branches and potential lesion areas. In order to eliminate operational differences, image evaluation and detection results were batch generated by a unified script, and their validity was manually verified. The system recorded multiple sets of signal-to-noise ratio, structural similarity and detection accuracy data for performance comparison and comprehensive analysis. Figures 1, 2 and 3 are the SNR/SSIM/mAP test results respectively:

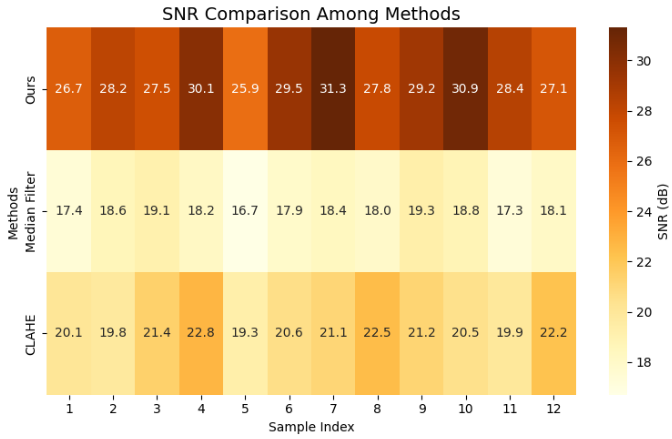

Figure 1: SNR

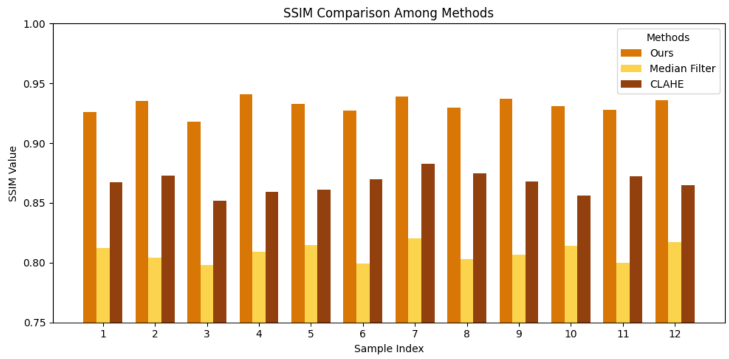

Figure 2: SSIM

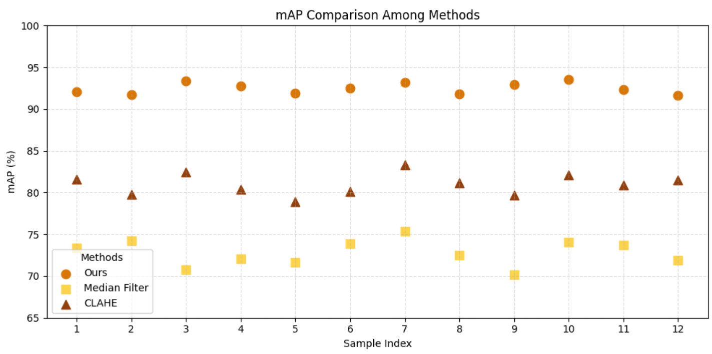

Figure 3: mAP

The proposed method achieved an SNR range of 25.9–31.3 dB, significantly outperforming median filtering and CLAHE, demonstrating that the deep learning-enhanced model achieves a balance between noise reduction and smoothness preservation. The SSIM remained stable between

0.918–0.941 across a diverse sample, with an average of 0.932, demonstrating that the CNN feature map and attention fusion mechanism effectively restored tissue contrast and detail levels. The mAP data reflects the robustness of the detection module, with the proposed system maintaining a mAP of 91.6%–93.5%, demonstrating that the combined U-Net segmentation and automatic detection strategy significantly enhances lesion localization capabilities. Overall, while traditional algorithms can improve image brightness to a certain extent, they are unstable when processing speckle noise and weakly echogenic lesions. However, the proposed method maintains high resolution and structural integrity across multiple cases, is particularly effective for low-SNR videos, and possesses greater clinical practical value.

4.3 Results Analysis

The contribution of the model structure to performance improvement was verified through a stepwise ablation experiment. First, using the complete model as a baseline, each structural module was removed or replaced, then retrained. Performance metrics such as SNR, SSIM, and mAP were tested under the same dataset and parameter conditions. The testing steps included: ① Fixing the training and validation sets; ② Removing modules one by one, such as the feature enhancement unit, attention mechanism, and fusion layer; ③ Recording the changes in each metric; ④ Calculating the percentage of performance degradation as a reference for contribution; ⑤ Comparing and analyzing the importance of each module and the overall synergistic effect. The contribution analysis results are shown in Table 1:

Table 1: Contribution Analysis Results

Model Configuration	Feature Enhancement Module	Attention Mechanism	Multi-Scale Fusion	Residual Connection	Performance Contribution (%)
Full Model	√	√	√	√	100
Without Feature Enhancement	×	√	√	√	91.3
Without Attention Mechanism	√	×	√	√	93.5
Without Multi-Scale Fusion	√	√	×	√	88.7
Without Residual Connection	√	√	√	×	95.4
Backbone Only	×	×	×	×	82.6

As shown in the table, the complete model performs best, with each individual module significantly contributing to overall performance. The multi-scale fusion module contributes the most, followed by feature enhancement and the attention mechanism. Removing multi-scale fusion reduces performance by over 11%, demonstrating its crucial role in effectively integrating information from different levels for feature representation. The feature enhancement module significantly improves detail recovery and noise suppression, while the attention mechanism enhances recognition by focusing on salient feature regions. While the residual connection contributes relatively little, it plays an important supporting role in stabilizing training and preventing vanishing gradients. Together, these four components form an overall performance optimization system.

5. Conclusion

The proposed B-ultrasound video image enhancement and automatic detection system fully integrates the synergistic advantages of image preprocessing, deep feature extraction, and semantic segmentation detection in its structural design. Through adaptive filtering and contrast enhancement, the readability of B-ultrasound images is significantly improved, providing better input conditions for subsequent feature extraction in deep networks. The CNN-based feature extraction module and multi-scale attention mechanism achieve accurate modeling of local texture and global structure, and possess strong robustness and stability in complex tissue structures. The U-Net-style detection structure effectively integrates shallow edge and deep semantic information, improving the boundary recognition and morphological consistency of lesion areas. The overall framework embodies the targeted and generalizable nature of artificial intelligence methods in medical image understanding. However, the system still has limitations in real-time processing performance and cross-device generalization. The efficiency and adaptability of clinical applications can be further improved through lightweight network design and transfer learning strategies.

References

[1] Wang Yongfei, Wang Chun. Low-light medical color image enhancement algorithm based on improved Retinex algorithm [J]. Journal of Shandong University of Aeronautics and Astronautics, 2021, 42(2): 52-60

[2] Xu Li, Liu Liang, Zhao Fengjun. Multimodal rigid medical image layered enhancement simulation under grayscale transformation [J]. Computer Simulation, 2020, 41(4): 250-254

[3] Deng Erqiang, Qin Zhen, Zhu Guosong. EnGAN: Enhanced generative adversarial network in medical image segmentation [J]. Computer Application Research, 2019, 41(7): 2195-2202

[4] Shangguan Hongting, Liu Yuhong. A review of medical image enhancement methods [J]. Chinese Journal of Medical Physics, 2018, 40(4): 410-415

[5] Wang Jian, Cheng Chufan, Chen Fang. Realistic medical image enhancement method based on multi-loss hybrid adversarial function and heuristic projection algorithm [J]. Data Acquisition and Processing, 2020, 38(5): 1104-1111

[6] Dinh P H, Giang N L. A new medical image enhancement algorithm using adaptive parameters [J]. International Journal of Imaging Systems and Technology, 2019, 32(6): 2198-2218.

[7] Li H, Lin Z, Qiu Z, et al. Enhancing and adapting in the clinic: Source-free unsupervised domain adaptation for medical image enhancement [J]. IEEE Transactions on Medical Imaging, 2020, 43(4): 1323-1336.

[8] Chen Z, Pawar K, Ekanayake M, et al. Deep learning for image enhancement and correction in magnetic resonance imaging—state-of-the-art and challenges [J]. Journal of Digital Imaging, 2021, 36(1): 204-230.

[9] Wu H T, Cao X, Gao Y, et al. Fundus image enhancement via semi-supervised GAN and anatomical structure preservation [J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2020, 8(1): 313-326.

[10] Goceri E. Medical image data augmentation: techniques, comparisons and interpretations [J]. Artificial intelligence review, 2018, 56(11): 12561-12605.

[11] Hua H, Li Y, Wang T, et al. Edge computing with artificial intelligence: A machine learning perspective [J]. ACM Computing Surveys, 2021, 55(9): 1-35.

[12] Weber P, Carl K V, Hinz O. Applications of explainable artificial intelligence in finance—a systematic review of finance, information systems, and computer science literature [J]. Management Review Quarterly, 2020, 74(2): 867-907.

[13] Jaber T A. Artificial intelligence in computer networks [J]. Periodicals of Engineering and Natural Sciences, 2018, 10(1): 309-322.