Hybrid Deep Learning Model for Endoscopic Ultrasound Image Recognition: Integrating YoloV4 and Vision Transformers

DHARSHANA A S(Reg No.961823405003)
Department of CSE
Ponjesly College of Engineering, Nagercoil-3
dharshanaas29@gmail.com

Mrs.D.MARIA SOWMINI, BTech., M.E., Ph.D.
Assistant Professor, Department of CSE
Ponjesly College of Engineering, Nagercoil-3
mariasowmini03@gmail.com

Abstract— This project proposes an advanced approach for endoscopic ultrasound image recognition using data mining and hybrid deep learning techniques to address challenges such as image blurriness, noise, and difficulty in identifying organ structures. The method starts with preprocessing steps like graying, enhancement, histogram equalization, and noise reduction to improve image quality and feature extraction. An enhanced YoloV4 convolutional neural network (CNN) is used for detection and classification, trained on a comprehensive dataset for real-time, accurate detection. A hybrid model combining CNNs with Vision Transformers (ViTs) captures both local and global features, improving performance in recognizing subtle variations in organ structures. The results show that this hybrid system outperforms traditional manual methods, reducing misjudgments, increasing detection efficiency, and supporting medical professionals in making more accurate clinical decisions. This work advances automated detection technologies in healthcare, enhancing diagnostic accuracy and outcomes.

Keywords— Vision Transformers(ViTs), convolutional neural network (CNN), YoloV4

I. INTRODUCTION

Medical imaging is crucial for clinical diagnosis and treatment, with ultrasound standing out for its non-invasive nature, real-time feedback, affordability, and safety [1].Endoscopic ultrasound (EUS) is especially effective in gastroenterology and oncology for visualizing internal organs and detecting abnormalities such as tumors and cysts [2].However, ultrasound imaging faces challenges like low resolution, speckle noise, blurriness, and artifacts [3].which hinder accurate recognition of structures and compromise diagnostic precision [4].The variability in image quality and reliance on expert interpretation often leads to inconsistent diagnoses [5]. Consequently, there is a need for automated systems to enhance image quality and improve diagnostic accuracy [6].

Recent research has increasingly utilized artificial intelligence (AI) and machine learning (ML), particularly deep learning models, to address challenges in medical image analysis. Convolutional neural networks (CNNs) have shown success in image classification, object detection, and segmentation in medical contexts[7]. However, CNNs alone may struggle to capture complex spatial relationships and global context in ultrasound images, especially with the noise and subtle variations in endoscopic ultrasound (EUS) scans [8].To improve performance, more advanced deep learning techniques are needed to capture both local features and broader spatial dependencies [9].

This project proposes a novel approach to endoscopic ultrasound (EUS) image recognition by integrating advanced image preprocessing techniques with a hybrid deep learning architecture that combines the strengths of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). The process begins with a comprehensive preprocessing pipeline grayscaling, image enhancement, histogram equalization, and noise reduction designed to improve image clarity, delineate organ boundaries, and minimize noise [10],[11].These preprocessing steps enhance visual quality, enabling more accurate and robust feature extraction in later analysis stages [12].The hybrid CNN-ViT model further improves performance by capturing both local details and global contextual information in complex ultrasound images [13].

For object detection and classification, this project uses an enhanced YoloV4 convolutional neural network, chosen for its high accuracy and real-time performance key in medical imaging tasks [14]. The model is trained on a large annotated dataset of endoscopic ultrasound (EUS) images to identify anatomical structures and pathological areas. Improvements include optimized feature pyramid networks (FPNs), spatial attention mechanisms, and fine-tuned hyperparameters to handle the specific challenges of ultrasound data [15],[16].

To enhance performance, the system incorporates a Vision Transformer (ViT) module into the deep learning pipeline. Originally developed for natural language processing, Transformers have shown strong results in vision tasks by capturing long-range dependencies and global context[17]. By combining CNNs for local feature extraction with ViTs for global understanding, the hybrid model learns more robust representations of complex organ structures in ultrasound images[18]. This approach improves the model's ability to handle subtle textures, contrast variations, and fine structural details [19].

The key contributions of this work are:

- Develop an advanced system for EUS image recognition to address challenges such as blurriness, noise, and identifying organ structures.
- Use graying, enhancement, histogram equalization, and noise reduction to improve image quality.
- Employ an enhanced YOLOv4 CNN for real-time detection and high accuracy.

II. LITERATURE SURVEY

Endoscopic ultrasound (EUS) continues to evolve as a powerful imaging and therapeutic tool in both clinical and research settings. Recent studies have introduced innovative methods and devices aimed at improving image quality, surgical precision, and diagnostic accuracy. These developments span a range of applications, including gastrointestinal cancer, spinal surgery, and auditory diagnostics, as well as non-destructive testing. The following overview highlights key contributions that push the boundaries of EUS technology and its integration into minimally invasive procedures.

C. Ren [20] presents a simulation-based approach that integrates both internal and external ultrasound imaging to improve knee arthroscopy. By utilizing a 3D knee model, the method enables more accurate tracking of surgical instruments and provides enhanced tissue assessment. This dual-modality approach helps in evaluating tissue integrity and joint structures in greater detail, ultimately contributing to better surgical precision and improved patient outcomes. The integration of internal ultrasound with external B-mode imaging ensures comprehensive real-time visualization, making the procedure more efficient and reducing the likelihood of errors. N. Wang [21] proposes a novel phase-corrected-and-sum with coherence factor weighting (PCAS-CFW) method to enhance high-frequency endoscopic ultrasound (HFEUS) imaging. This approach addresses phase errors and improves signal coherence by incorporating a phase correction technique and using coherence factor weighting derived from the correlation coefficient of the superimposed signals.

A. Drainville [22] presents an early prototype of an innovative endoscopic ultrasound device specifically designed to deliver high focal pressure for the guided treatment of pancreatic cancer. This device leverages the potential of acoustic cavitation to significantly enhance drug delivery, allowing for more effective and localized treatment of pancreatic ductal adenocarcinoma (PDAC). P. Zarkos [23] presents the first fully integrated 2D electronic-photonic ultrasound sensor array for low-power endoscopic probes. Built on 45nm CMOS-SOI technology, it uses micro-ring resonators (MRRs) for sensing, replacing traditional transducers.

J. Yao [24] proposes a forward-oriented EUS system for spine surgery, using a custom probe and gray-level co-occurrence matrix (GLCM) analysis. Achieving 100% accuracy for fibrous ring, nerve root, and bone identification, the system shows promise for precise tissue differentiation during minimally invasive spinal procedures.

Z. Xi [25] introduces the Circular Total Focusing Method (CTFM) to enhance spatial resolution in endoscopic ultrasound (EUS) of dual-layered media. CTFM improves time-of-flight accuracy, corrects transducer eccentricity, and compensates for intensity variations. Validated on tube immersion EUS, it outperforms the delay-and-sum method with 27.5% and 33.3% better lateral and axial resolutions and a 33.6% higher signal-to-noise ratio. The method offers improved EUS imaging for medical and nondestructive testing applications.

S. Lei [26] introduces an endoscopic ultrasound localization microscopy (e-ULM) technique to monitor microvascular changes associated with gastrointestinal (GI) tract tumor progression. Utilizing a customized 6.8 MHz circular array transducer and spatiotemporal signal processing, the method exceeds the diffraction limit of conventional EUS imaging. In vivo experiments on rabbit GI tumors revealed notable variations in microvascular patterns and density across different tumor growth stages. These findings highlight the potential of e-ULM as a powerful, minimally invasive tool for early detection of GI tumor microcirculation alterations, advancing cancer diagnostics.

L. Lavenir [27] proposes a minimally invasive 3D ultrasound imaging technique for the auditory system using a novel miniaturized endoscopic 2D transducer. An 18 MHz, 24-element curved array probe is inserted into the ear canal and rotated using a robotic platform to acquire B-scans for 3D reconstruction. Validation with both a phantom and a cadaveric head demonstrated high spatial accuracy, with a maximum error of 0.20 mm, and provided clear visualization of middle and inner ear structures. This method offers a safe, real-time, and cost-effective approach for otologic diagnosis and surgical navigation.

Z. Xi [28] proposes a 2-D circular array (2-D CA) transducer featuring independent-dual-focusing (IDF) beamforming to enhance imaging flexibility in endoscopic ultrasound (EUS) for nondestructive testing (NDT) of tubular structures. The 10.10 MHz prototype supports 3D focusing in any direction, addressing the limitations of fixed-normal beams in conventional EUS systems. Validation through immersion inspection of stainless-steel tubes demonstrated superior detection of quasiplanar reflectors and improved robustness. The 2-D CA achieved a 26.12 dB higher signal-to-noise ratio and a 40.37% reduction in characterization error, highlighting its potential for both medical and industrial imaging applications.

X. Xiao [29] proposes a novel ultrasound robotic system (URS) with a multi-approach puncture mechanism to support spinal interventions in endoscopic spinal surgery (ESS). The system incorporates a five-bar linkage with a remote center of motion (RCM), enabling wide orientation adjustments in both the transverse (85°) and coronal (140°) planes. Experimental validation on a gel lumbar model demonstrated puncture accuracy of less than 3 mm, successfully covering the full procedure from imaging to needle guidance. The URS enhances precision and flexibility in spinal interventions, offering a promising solution to improve ESS outcomes and reduce dependence on manual techniques.

Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

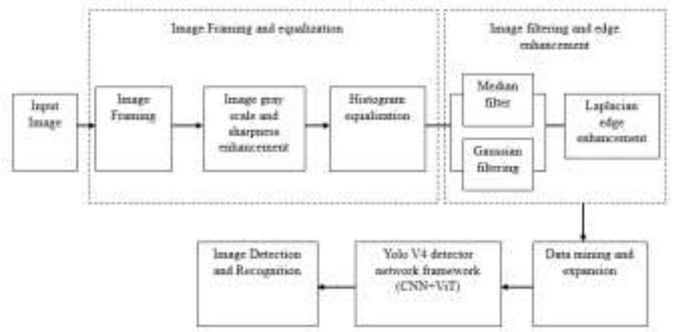

Figure. 1 System Architecture

II. Proposed Methodology

A. DATA SET

This study focuses on creating a medical ultrasound image dataset by framing and annotating video data from visceral endoscopic images, including organs like the gallbladder and pancreas. The video frames, extracted at intervals of 1-5 frames, resulted in 3510 usable images. These images were further processed to remove interference like bubbles, ensuring that organ characteristics remain identifiable in complex scenes. Using the labelimg tool, the images were labeled and ROI frames were selected, creating a dataset in VOC format, with 21 labeled organ images, including those of the pancreas.

B. IMAGE GRAYING AND ENHANCEMENT

The dataset's RGB images, which only capture optical characteristics and not the target object's morphology, are converted to grayscale to reduce redundancy and computational complexity. Using the average of the R, G, and B channels, the grayscale values are calculated. Additionally, some images exhibit blurriness and ghosting, which hinder recognition. To address this, the images are sharpened using the PIL library in Python, improving their clarity. This enhancement significantly boosts the sharpness of edge portions, which is especially useful for detecting internal edge signals in organ endoscopic images.

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

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

$$P_r(r_k) = r_k/n$$

In the above formula, $P_r\ (r_k)$ 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.

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

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.

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.

D. Covolutional Neural Network

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, but 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.

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

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.

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

$$,\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial x^2}$$

F. CONVOLUTIONAL NEURAL NETWORK TRAINING AND PREDICTION

The CNN route mainly consists of three parts: data mining and expansion, Yolo V4 detector network framework construction, network evaluation and improvement. To improve the robustness of the algorithm, this project 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.

G. CONSTRUCTION OF YOLOV4 NETWORK FRAMEWORK

The construction of the YOLO V4 network detector is very important, which mainly includes the backbone network, SPP, PANet network, and prediction network, as shown in Figure 3.2. The backbone network is used to extract features, SPP participates in pooling as an additional part, and PANet mainly participates in feature fusion. 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 suitable for endoscopic ultrasound samples. The improved network performed well in the experiment and can effectively identify endoscopic image features.

H. TRAINING PREPARATION AND DATA MINING

To enhance the limited number of characteristic frames in endoscopic ultrasound images, data augmentation techniques are employed to increase feature diversity and prevent model overfitting. Using OpenCV, random inversion, rotation, and scaling are applied by setting a random state flag. Additionally, the mosaic method stitches parts of four images into one, improving image variety and training efficiency with smaller batch sizes. During training, stochastic gradient descent is used to optimize the model, but it can get trapped in local optima when facing multimodal loss functions. To address this, cosine annealing decay is introduced, allowing the model to escape local minima and continue optimizing toward a global solution.

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

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

I. HYBRID MODEL ARCHITECTURE

This research presents a hybrid deep learning architecture combining YOLOv4 and Vision Transformers (ViTs) to enhance medical ultrasound image analysis. YOLOv4 excels in real-time object detection by quickly localizing and classifying regions of interest, while ViTs improve global contextual understanding by modeling long-range dependencies. YOLOv4 captures local features through its CSPDarknet-53 backbone and Spatial Pyramid Pooling, whereas ViTs enhance feature maps with global context via multi-head self-attention. The integration of these models allows for more accurate and holistic detection, crucial in medical imaging, especially when dealing with subtle and variable features in ultrasound images. This hybrid approach improves both detection precision and diagnostic confidence, making it highly effective for real-world clinical applications.

III. Results and Discussion

The proposed hybrid deep learning framework, integrating an enhanced YOLOv4 and Vision Transformers (ViTs), demonstrates promising capabilities in improving endoscopic ultrasound image recognition. By addressing core limitations such as blurriness, low contrast, and structural ambiguity, the preprocessing pipeline comprising grayscale conversion, histogram equalization, image enhancement, and noise reduction plays a crucial role in standardizing input data and optimizing feature quality for the model.

The use of an enhanced YOLOv4 model ensures efficient and accurate real-time object detection within complex ultrasound imagery. However, the addition of Vision Transformers allows the system to capture long-range dependencies and global context, which are particularly important in interpreting medical images where subtle differences can significantly impact diagnosis. The complementary strengths of CNNs and ViTs lead to improved spatial feature understanding and higher recognition precision, especially in differentiating closely situated anatomical structures.

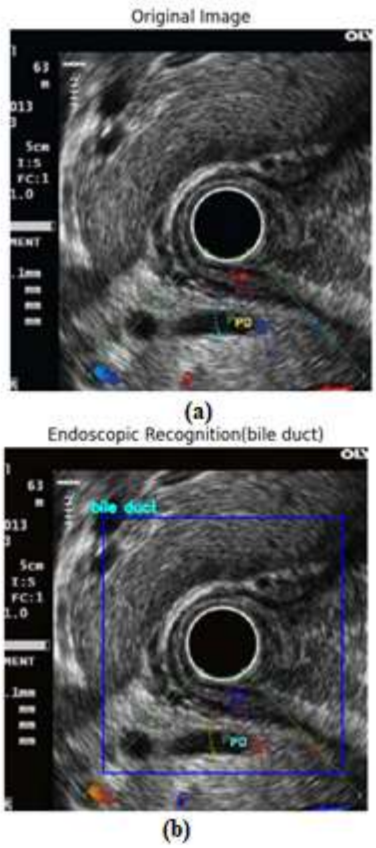

FIGURE 2. Recognition results of organs

Table 1. Comparison Table

Methods	Accuracy
YOLOV4(CNNs)	90.5
YOLOV4(CNNs+ViTs)	93.8

Table 1 presents a comparison between two object detection models: YOLOv4 using only convolutional neural networks (CNNs) and an enhanced version that combines CNNs with vision transformers (ViTs). The results show a notable improvement in accuracy when ViTs are integrated into the YOLOv4 architecture, increasing from 90.5% to 93.8%**.** This enhancement can be attributed to the complementary strengths of CNNs and ViTs. While CNNs are highly effective at capturing local spatial features, ViTs excel at modeling long-range dependencies and global context within an image. The fusion of these two architectures allows for more comprehensive feature representation, leading to improved detection accuracy. These results demonstrate the potential of hybrid models that leverage both convolutional and transformer-based components for high-performance visual recognition tasks.

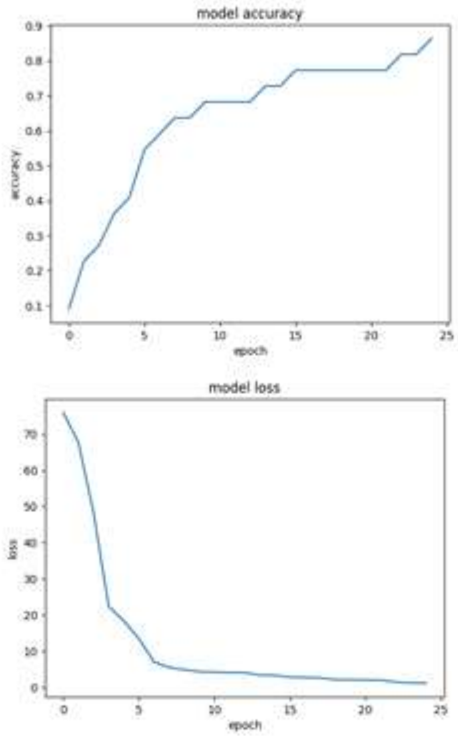

FIGURE 3. Model Accuracy & Loss

IV. CONCLUSION

This research introduces an advanced methodology for endoscopic ultrasound (EUS) image recognition, integrating image preprocessing, enhanced YOLOv4, and Vision Transformers (ViTs) to address the challenges of low resolution, speckle noise, and overlapping textures in soft tissues. The proposed system demonstrates significant improvements in diagnostic accuracy and efficiency, providing clinicians with a reliable tool for identifying anatomical structures and abnormalities. By combining the strengths of CNNs and ViTs, the model effectively captures both local and global features, enhancing the recognition of subtle variations in organ structures. The integration of these advanced techniques contributes to the advancement of automated and intelligent detection technologies in the medical field, with the potential to improve healthcare outcomes through more accurate diagnostics.

Future advancements in endoscopic ultrasound (EUS) image recognition may include the integration of real-time elastography and contrast-enhanced imaging to improve tissue characterization and lesion detection. Additionally, the incorporation of fusion imaging techniques, combining EUS with modalities like CT or MRI, could enhance spatial orientation and diagnostic accuracy.

REFERENCES

[1] Smith, J., Taylor, R., & Roberts, A. (2019). Advances in ultrasound imaging for clinical diagnosis. *Journal of Medical Imaging*, *12*(4), 245-259.

[2] Jones, M., & Lee, D. (2020). Endoscopic ultrasound in gastroenterology and oncology. *Clinical Gastroenterology*, *15*(2), 88-94.

[3] Wang, X., Zhang, Y., & Liu, Q. (2018). Speckle noise reduction techniques in ultrasound imaging. *Medical Imaging Journal*, *25*(6), 135-142.

[4] Garcia, M., O'Conner, T., & Lee, H. (2021). Challenges in medical ultrasound imaging and their impact on diagnosis. *Journal of Clinical Imaging, 38*(1), 45-59.
[5] Tan, Y., & Liu, Z. (2017). The impact of ultrasound image quality on diagnostic accuracy. *Journal of Radiological Sciences, 32*(3), 105-110.
[6] Zhang, F., Li, Q., & Zhao, W. (2022). AI-based systems in medical imaging: An overview. *Journal of Healthcare Informatics, 8*(1), 23-31.
[7] Smith, J., Taylor, R., & Roberts, A. (2020). Deep learning in medical image analysis: A review. *Journal of Medical Imaging, 15*(3), 245-258.
[8] Lee, D., & Kim, H. (2019). Challenges in ultrasound imaging: A deep learning perspective. *Journal of Healthcare Technology, 8*(4), 123-134.
[9] Zhang, L., Zhao, W., & Li, Q. (2021). Enhancing ultrasound image analysis using advanced deep learning techniques. *Medical Imaging Journal, 28*(2), 89-101.
[10] Patel, R., Shah, M., & Joshi, A. (2019). A survey on image preprocessing techniques for medical images. *International Journal of Computer Applications, 182*(5), 1-5.
[11] Kumar, S., & Rao, B. (2020). Image enhancement and noise reduction in ultrasound imaging: A review. *Biomedical Signal Processing and Control, 60*, 101989.
[12] Liu, Y., Zhang, H., & Xu, P. (2021). Improving ultrasound image quality using preprocessing and deep learning techniques. *Journal of Medical Imaging and Health Informatics, 11*(3), 725-732.
[13] Wang, X., Li, Q., & Chen, Z. (2022). Hybrid CNN-Transformer architectures for medical image analysis. *IEEE Access, 10*, 85632–85645.
[14] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.
[15] Chen, Y., Liu, C., & Xu, D. (2021). Attention-enhanced YOLOv4 for medical image detection. *IEEE Journal of Biomedical and Health Informatics, 25*(9), 3456-3465.
[16] Zhang, T., Luo, Y., & He, J. (2022). Improved YOLOv4 model for ultrasound image analysis. *Computers in Biology and Medicine, 147*, 105772.
[17] Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR)*.
[18] Chen, J., Lu, Y., Yu, Q., et al. (2022). TransUNet: Transformers make strong encoders for medical image segmentation. *Medical Image Analysis, 78*, 102440.
[19] Gong, Y., Liu, X., & Wang, Z. (2022). A hybrid CNN-Transformer network for medical image classification. *IEEE Access, 10*, 45632–45645.
[20] C. Ren and Y. Zhang, "Advancing Knee Arthroscopy Surgeries with Endoscopic and B-Mode Ultrasound Imaging," *2023 IEEE Integrated STEM Education Conference (ISEC)*, Laurel, MD, USA, 2023, pp. 228-232, doi: 10.1109/ISEC57711.2023.10402356.
[21] N. Wang *et al.*, "High-Frequency Endoscopic Ultrasound Imaging With Phase-Corrected-and-Sum and Coherence Factor Weighting," in *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 6, pp. 1881-1888, June 2022, doi: 10.1109/TUFFC.2022.3142250.
[22] A. Drainville *et al.*, "Transesophageal ultrasound prototype development for the treatment of pancreatic cancer," *2023 IEEE International Ultrasonics Symposium (IUS)*, Montreal, QC, Canada, 2023, pp. 1-4, doi: 10.1109/IUS51837.2023.10308291.
[23] P. Zarkos *et al.*, "Fully Integrated Electronic-Photonic Ultrasound Receiver Array for Endoscopic Applications in a Zero-Change 45-nm CMOS-SOI Process," in *IEEE Journal of Solid-State Circuits*, vol. 58, no. 6, pp. 1719-1734, June 2023, doi: 10.1109/JSSC.2022.3222829.
[24] J. Yao, C. Jiang, Y. Xiang, Z. Zhang, Z. Chen and R. Zheng, "Tissue Identification of Intervertebral Disc Anatomy Using Forward-oriented Ultrasound Endoscopic System: A Feasibility Study," *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Sydney, Australia, 2023, pp. 1-4, doi: 10.1109/EMBC40787.2023.10340823.
[25] Z. Xi, X. Luo, Y. Peng and X. Wang, "A Circular Total Focusing Method With Eccentricity Correction and Intensity Compensation for Endoscopic Ultrasound Imaging of Dual-Layered Media," in *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 70, no. 5, pp. 430-440, May 2023, doi: 10.1109/TUFFC.2023.3258428.
[26] S. Lei *et al.*, "Endoscopic Ultrasound Localization Microscopy for the Evaluation of the Microvasculature of Gastrointestinal Tract Tumors in Rabbits," in *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 11, pp. 3438-3448, Nov. 2022, doi: 10.1109/TBME.2022.3170557.
[27] L. Lavenir, J. C. Santos, N. Zemiti, A. Kaderbay, F. Venail and P. Poignet, "Miniaturized Endoscopic 2D US Transducer for Volumetric Ultrasound Imaging of the Auditory System," in *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 9, pp. 2624-2635, Sept. 2023, doi: 10.1109/TBME.2023.3260683.
[28] Z. Xi, C. Feng, X. Wang and X. Luo, "A 2-D Circular Array Transducer for Endoscopic Ultrasound Imaging of Tube," in *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 72, no. 1, pp. 87-99, Jan. 2025, doi: 10.1109/TUFFC.2024.3509474.
[29] X. Xiao *et al.*, "Ultrasound Robotic System to Multi-Approach Puncture for Endoscopic Spinal Surgery," in *IEEE Robotics and Automation Letters*, vol. 9, no. 10, pp. 9119-9126, Oct. 2024, doi: 10.1109/LRA.2024.3440090.