AI-Enabled Vessels Segmentation Model for Real-Time Laparoscopic Ultrasound Imaging

Anonymous Full Paper
Submission 42

Abstract

Laparoscopic ultrasound (LUS) is essential for assessing the liver during laparoscopic liver resections. However, the interpretation of LUS images presents significant challenges due to the steep learning curve and image noise. In this study, we propose an enhanced U-Net-based neural network with a ResNet18 backbone specifically designed for real-time liver vessel segmentation of 2D LUS images. Our approach incorporates five preprocessing steps aimed at maximizing the training information extracted from the ultrasound sonogram region. The modified U-Net model achieved a Dice coefficient of 0.879, demonstrating real-time performance at 40 frames per second and enabling the development of advanced ultrasound-based surgical navigation solutions.

1 Introduction

Liver cancer remains one of the top 10 deadliest cancers worldwide, resulting in approximately 750,000 annual deaths [1]. The reason for its mortality rate is late diagnoses, limited treatment options, and underlying liver disease with aggressive tumor biology [2]. To locate liver tumors and vessels during laparoscopic liver surgery, clinicians are using the laparoscopic ultrasound (LUS), which helps to navigate and to avoid unnecessary damage during liver resection or ablation. LUS is a radiation-free medical device, portable and cost-effective. It provides real-time images by capturing ultrasound reflected pulses from soft tissues and bones [3]. All of these LUS benefits give clinicians the ability to effectively diagnose liver cancer, such as hepatocellular carcinoma and other metastases. Additionally, LUS allows visualization of essential liver structures, the portal vein, hepatic veins, and bile ducts.

While valuable, LUS comes with several drawbacks. A key problem is speckle noise, an artifact from ultrasound waves, that interferes as reflected off tissue microstructures. This effect lowers overall image quality [4]. Vessel boundaries can also appear unclear because of differences in tissue echogenicity — the way tissues reflect sound. When boundaries fade, tracking blood vessels during surgery becomes more difficult. Finally, underlying conditions like fatty liver or cirrhosis are causing liver texture changes which interfere with the interpretation of ultrasound scans [5].

These imaging issues limit how effectively LUS can guide surgeons during liver procedures [6]. Several conventional techniques could be used to account for these challenges. One of the default modes of current ultrasound (US) systems, is Color Doppler mode, which can be used to visualize blood flow by detecting frequency shifts in moving blood cells and to enable real-time assessment of vascularity. However, it has a relatively small region of interest, and its effectiveness is heavily dependent on the operators' skill, which might introduce inconsistency in the interpretation of the LUS data [7]. Another traditional visualization method is a Contrast-Enhanced Ultrasound (CEUS). It uses microbubble contrast agents to improve the visibility of blood vessels. Unfortunately, this method requires careful timing to capture the best blood flow enhancement after the contrast is given, which can be difficult in busy surgical environment [8]. Traditional segmentation algorithms, such as region growing, thresholding, and clustering, have also been employed for tissue segmentation [9]. All of them require manual tuning of thresholds value and seed points, which limits their robustness in handling the complex and heterogeneous tissue structures present in ultrasound images.

Over the recent years, deep learning-based approaches have emerged as a leading method for automated vessel segmentation. These techniques have achieved impressive results, with a Dice similarity coefficient of 0.734 for ultrasound images [10], 0.928 for MRI slices [11], and 0.814 for CT scans [12]. U-Net-based architectures, have been recognized as the gold standard for semantic segmentation tasks [13]. Their encoder-decoder structure and skip connections have made them adaptable to enhancements such as adding residual blocks (ResU-Net) [14], dense connections (DenseU-Net) [6], attention gates (Attention U-Net) [15], or transformers (TransU-Net) [16]. Although these studies have demonstrated competitive results in segmenting various biological tissues from ultrasound data, only a few have explored the performance of real-time segmentation [6, 17].

Real-time ultrasound image segmentation is a complex task due to the noise and inconsistent data. Preprocessing is often employed to suppress speckle

noise and reduce artifacts, but this adds computational overhead. Varying echogenicity makes boundary detection difficult. Therefore, deep learning models need to be carefully optimized to achieve accurate results. Post-processing techniques, such as mask refinement for frame-to-frame consistency, or resizing output to the original resolution, add further computational load. This makes difficult to balance between tuning models for high accuracy or for the speed required in real-time applications.

In a recent study, Smistad et al. [18] used an Artificial Intelligence (AI) model to segment blood vessels, nerves, and bone structures during anesthesia-related procedures, and showed a promising performance. However, the predictions were made on a frame-by-frame basis without considering the temporal information in the sequential ultrasound data, which could have inherent potential information to enhance the performance.

This paper aims to develop an automated AI-enabled LUS model for real-time vessel segmentation, which can significantly improve liver cancer surgery. The contributions of the proposed workflow are summarized as follows:

1. Fully anonymized LUS liver video data was locally acquired and annotated with the assistance of experienced clinicians. The annotations are currently under final review by a radiologist.

2. A dynamic approach was developed to extract the ultrasound sonogram from video frames. It enabled precise masking of the imaging area and prevented the network from learning irrelevant background features, thereby segmentation accuracy got improved without compromising real-time performance.

3. The triplet input setup, similar to ones used for LUS-CT co-registration [19] and for object recognition [20], was integrated into a lightweight ResNet18 U-Net model, enhancing segmentation quality by introducing contextual information between frames.

4. Contrast Limited Adaptive Histogram Equalization (CLAHE) [21] was applied and optimized for our dataset. It enhanced vessel boundaries and improved lumen visibility, which resulted in increased segmentation accuracy.

5. A comprehensive study was conducted to evaluate the performance of different U-Net family encoders, focusing on both segmentation accuracy and real-time inference efficiency.

The AI-generated 2D liver vessel segmentation masks can also be used for 3D vessel reconstruction, aiding in image registration between preoperative and intraoperative stages.

2 Proposed methodology

This paper addresses the segmentation of the vessels in real time in laparoscopic ultrasound using a U-Net–based encoder–decoder with a ResNet18 backbone.

2.1 Pre-processing pipeline

As shown on the left side of Figure 1, all the LUS frames undergo standardization before entering the network to: a) emphasize learning focused on the acoustic sonogram, b) stabilize contrast across different cases, acquisition depths, and sonogram shapes, and c) ensure a consistent input shape while preserving spatial geometry. The specific preprocessing steps (1 to 5) visualized in Figure 1 will be described in detail as follows:

1. **Grayscale conversion**: In LUS video, the RGB channels are identical copies of the same intensity data, so this processing step eliminates redundant channels, resulting in a single-channel grayscale representation.

2. **Sonogram detection:** To locate the US sonogram as a trapezoid in frame coordinates, the dynamic contour-based detection with a fallback heuristic was developed and used. For the fallback, the most recently detected good coordinates were used. The trapezoid is defined by four corner points, stored in a 4×2 matrix $P^{\text{trap}} \in \mathbb{Z}_{\geq 0}^{4\times 2}$, where each row holds the (x, y) coordinates of one corner in a set of intigers $\mathbb{Z}$. By using this method, we managed to successfully detect the sonogram contour even when the acquisition depth was changing.

3. **Tight crop and CLAHE:** After locating the sonogram coordinates, we applied a tight rectangular crop to remove unnecessary background. Two hyperparameters—the contrast limiting threshold (*clipLimit*) and the grid size for histogram equalization (*GridSize*) — were tuned in a small-scale grid search. The chosen values, *clipLimit = 2.0* and *GridSize = 8x8*, which consistently improved image contrast and led to better segmentation performance both quantitatively and qualitatively.

4. **Sonogram masking:** After cropping, the sonogram was isolated using a binary polygon mask (sonogram = 0, outside = 255). Whitening the background region minimized the predictive model from extending beyond the sonogram contour, ensuring the network focuses on anatomical features within the sonogram area.

5. **Resizing and padding:** On the final step, we resized the ultrasound frame by maintaining the aspect ratio to prevent vessel distortion. Lastly,

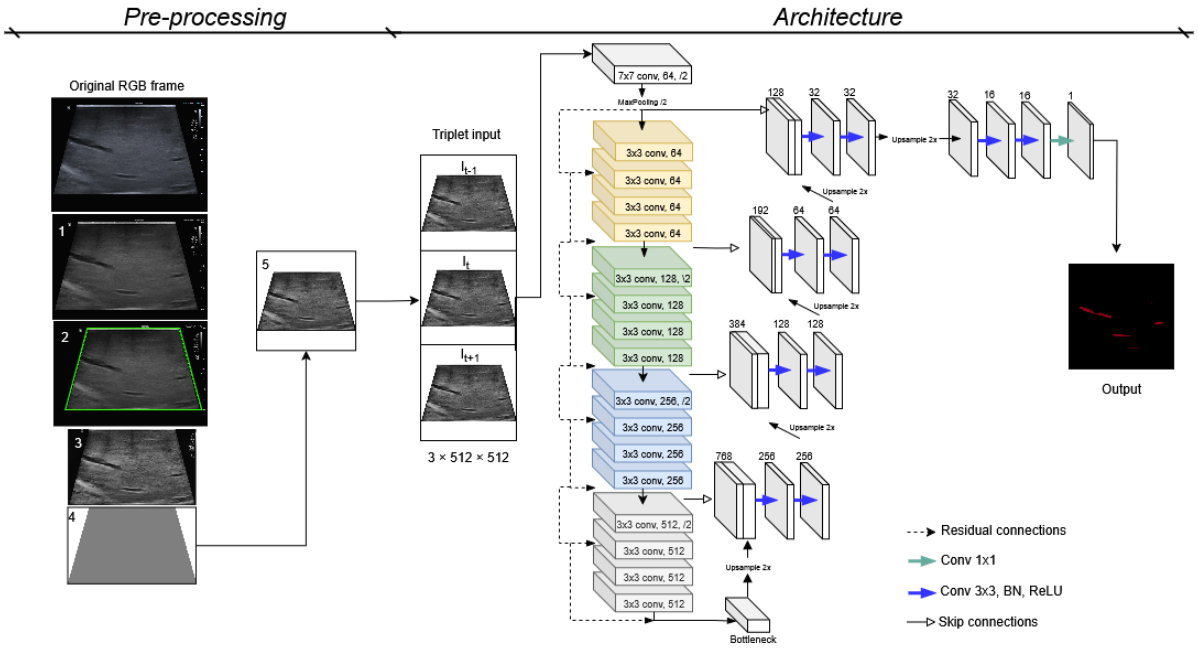

Figure 1. Overview of the proposed vessel segmentation pipeline. The raw RGB laparoscopic ultrasound frames undergo five preprocessing steps, are arranged into temporal triplets, encoded with a ResNet18 backbone, and decoded with a U-Net decoder to generate the final binary mask of the vessels.

performed the symmetric padding to 512 x 512 using bilinear interpolation. This fixed square input ensures compatibility with various network backbones, making the preprocessing pipeline adaptable for different model selections.

By applying these five preprocessing steps to each LUS video frame, we direct the model's focus toward the sonogram area that contains vascular anatomy. CLAHE enhances local contrast for thin, low-contrast vessels, while aspect-ratio-based resizing prevents geometric distortions of tubular structures.

2.2 Network architecture

For this study, we tested and adopted a U-Net-based encoder–decoder network with multiple backbones from the U-Net family, including lightweight ResNet18, MobileNet_v2, DenseNet-121, medium-sized ResNet50, and Vanilla U-Net, as well as a larger model, InceptionNetV2. The selected encoder, ResNet18, and the standard U-Net decoder, used for training and inference, are shown on the right side of Figure 1. ResNet18, pretrained on ImageNet [22], provides residual blocks that support stable training on small medical datasets. To exploit temporal coherence, sequential frames are processed as triplets rather than individually, enabling motion-aware and more consistent predictions. In general, U-Net architecture was selected due to its strong performance in segmentation tasks and its skip connections which help retain fine spatial details that are often lost during the downsampling process.

2.2.1 Modified input layer: temporal triplet setup

Unlike conventional U-Net inputs that use a single 2D frame, we implemented triplets of consecutive frames to leverage the temporal information of sequential ultrasound video data. During training, a symmetric triplet setup $[I_{t-1}, I_t, I_{t+1}]$ was used, with the ground truth mask corresponding to the middle image I_t. This configuration allows the model to learn context from both past and future frames, possibly enhancing vessel continuity and robustness to speckle noise. During inference, since future frames are not available, we switched to a more standard triplet setup: $[I_{t-2}, I_{t-1}, I_t]$, which maintains the benefits of temporal context without compromising real-time performance.

2.2.2 Loss, optimization and training controls

In preliminary experiments with a Vanilla U-Net, we tested several segmentation loss functions commonly used in medical imaging, including Dice loss [23], Focal loss [24], and Binary Cross-Entropy (BCE) loss [25]. Based on this comparison, we selected BCE as the primary training loss for all tested encoder backbones, as it provided the most consistent generalization and improvements. The Dice coefficient was used as the primary evaluation metric during model validation. The BCE loss can be described

as follows:

$$\mathrm{BCE}(p, y) = -\left(y \log(p) + (1 - y) \log(1 - p)\right), \quad (1)$$

where $y \in \{0, 1\}$ is the ground truth, and $p \in [0, 1]$ is the predicted probability (after sigmoid function). We employed Binary Cross-Entropy with logits loss (*BCEWithLogitsLoss* in PyTorch [26]), which is equivalent to applying a sigmoid activation followed by binary cross-entropy, but implemented in a numerically more stable form.

2.2.3 Optimization

All our networks were trained for up to 100 epochs under identical conditions to ensure a fair comparison across backbones. For optimization, we used Adam optimizer with a learning rate of 1×10^{-3} and a learning rate scheduler (factor = 0.5, patience = 5) to improve convergence once the validation loss plateaued. Other data augmentation techniques were excluded from the training. To accelerate training and minimize memory usage, mixed precision training (AMP) was employed, offering lower computational costs without compromising accuracy. The implemented early stopping interrupted the training if the validation Dice score did not improve for 10 epochs, preventing possible overfitting. The batch size was set to 18 to fully utilize the available GPU (RTX 4080, 12GB VRAM), and each channel of the triplet input was normalized to $[-1, 1]$.

2.3 Dataset and data split

The locally acquired dataset consisted of laparoscopic ultrasound videos from 11 separate cases, which were divided into a total of 2,200 frames. Each case folder contained sequential 2D frames along with pixel-wise binary masks that were locally annotated. To maintain independence between training and evaluation, a leave-one-case-out strategy was employed, with the held-out case serving as a fixed test set and the remaining ten cases used for model development. This data was divided into training and validation subsets in a 90/10 split.

3 Results and discussion

Various backbone architectures were tuned and compared within the proposed U-Net framework. Performance was initially evaluated using 5-fold cross-validation across all cases to identify the most promising encoders. Following this, we conducted single-fold training with a fixed leave-one-case-out test set to assess generalization on unseen data.

Table 1. Comparison of vessel segmentation performance of different models: Dice coefficient (DC) with standard deviation (Std) from 5-fold cross-validation (CV), and DC, recall, precision, and Intersection over Union (IoU) on the test set.

Encoder Type	5-Fold CV	Test set			
	DC ± Std	DC	Recall	Precision	IoU
ResNet18	0.916 ± 0.002	0.879	0.840	**0.923**	0.783
ResNet50	0.904 ± 0.002	0.856	0.811	0.912	0.748
MobileNet_v2	0.906 ± 0.002	0.862	0.819	0.917	0.757
Vanilla U-Net	0.918 ± 0.003	0.859	0.827	0.897	0.753
DenseNet121	0.901 ± 0.003	0.849	0.798	0.913	0.738
IncResNetV2	0.926 ± 0.002	**0.888**	**0.867**	0.911	**0.798**

3.1 Quantitative results

Table 1 presents the results of the 5-fold cross-validation experiments, comparing six backbone configurations. All tested encoders achieved Dice scores above 0.9 in 5-fold cross-validation, demonstrating that vessel segmentation in LUS is a learnable task. The consistent results across architectures indicate strong robustness and suggest that encoder selection can be guided by efficiency and deployment considerations rather than segmentation accuracy alone. When looking at the results from the test set, both InceptionResNetV2 and ResNet18 performed well, with InceptionResNetV2 reaching the highest Dice score of 0.888 and recall of 0.867 on the test set. Deep architecture and Inception modules enable InceptionResNetV2 to capture multi-scale rich features, while residual connections help stabilize training similarly as in ResNet18. However, with a large number of parameters, the model is computationally heavy, leading to longer training times and slower inference compared to lighter backbones.

Despite being the lightest model, ResNet18 produced competitive Dice of 0.879 and high precision score of 0.923, benefiting from residual connections that support efficient gradient flow and stable feature learning. Additionally, the smaller parameter count noticeably improved the inference speed, preserving real-time capability even with our additional preprocessing steps. This balance of accuracy and efficiency makes ResNet18 particularly suited for laparoscopic ultrasound vessel segmentation, where reliable performance must be achieved under strict computational constraints.

3.2 Qualitative results

Figure 2 shows representative segmentation examples from three test frames, comparing ground truth annotations with predictions from the lightweight ResNet18, the heavier InceptionResNetV2, and the medium-sized Vanilla U-Net. Across the models, there is a noticeable tendency for slight oversegmentation. While InceptionResNetV2 produced the most visually refined results, its computational complexity makes it less suitable for real-time de-

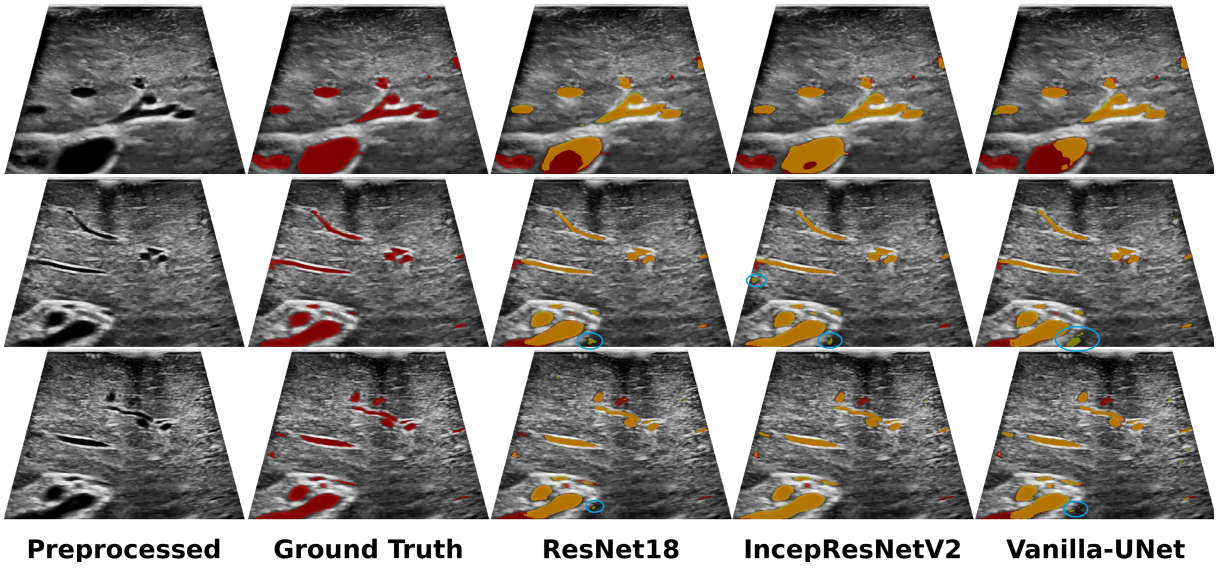

Figure 2. Vessel segmentation results using three models.
Red: ground truth, Yellow: model predictions, Blue circles: over-predicted regions.

ployment. In contrast, ResNet18 provided visually comparable masks, with segmentation quality not significantly inferior to that of Vanilla U-Net, despite being much lighter.

3.3 Impact of the triplet setup

To assess whether the proposed triplet input improves temporal stability compared to single-frame predictions, we trained a ResNet18 model with both single-frame and triplet inputs under otherwise identical settings. We then defined temporal consistency metrics, following recent studies [27, 28], and reported them in Table 2. *Temporal Dice* is defined as the Dice coefficient between consecutive frame predictions, averaged over the sequence, while *Temporal IoU* is defined analogously using the IoU.

In addition, we evaluated prediction stability across time. Following Rebol et al. [29], we measured *Flip-rate,* the proportion of pixels whose labels switch between consecutive frames (e.g., a vessel pixel that disappears and reappears). This captures segmentation "flickering" over time. However, in our experiments, Flip-rate values were consistently close to zero, likely reflecting both the strong class imbalance (vessel vs. background) and the generally high segmentation accuracy. Thus, the Flip-rate confirms the absence of large temporal instabilities in both compared models.

To quantify boundary stability, we used the idea from Perazzi et al. [30]. We define *Boundary jitter* as the average displacement (in pixels) of the predicted vessel boundary between consecutive frames, capturing small shifts of vessel contours. The results indicate that our model performs better when trained with adjacent frames, enhancing all temporal metrics. We also assessed the per-frame standard deviation against the ground truth (GT), which showed slight improvement. Additionally, the triplet input configuration enhanced temporal consistency, yielding masks that were more stable and less prone to flickering compared to the single-frame predictions shown in the provided four consecutive frames in Figure 3. The masks produced by the triple-frame setup had fewer false positive predictions, fewer false negative predictions, and fewer instances of flickering, resulting in a more stable stream during real-time experiments.

Table 2. Comparison of single-frame and triplet input ResNet18 U-Net models. Mean ± standard deviation is reported for all metrics. Arrows indicate whether higher (↑) or lower (↓) values are better.

Metric	ResNet18 U-Net (Single)	ResNet18 U-Net (Triplet)
Per-frame DC vs. GT (↑)	0.875 ± 0.031	**0.879 ± 0.030**
Temporal DC (↑)	0.932 ± 0.056	**0.938 ± 0.045**
Temporal IoU (↑)	0.877 ± 0.091	**0.886 ± 0.073**
Flip rate (↓)	0.008 ± 0.008	**0.007 ± 0.007**
Boundary jitter (px, ↓)	1.334 ± 1.257	**0.980 ± 0.990**

3.4 Real-time experiment and results

To test deployment under realistic conditions, the models were evaluated on an unseen laparoscopic ultrasound video using the full real-time pipeline, including preprocessing, frame-wise inference, and post-processing to mimic the clinician's view. The developed dynamic contour detector enabled the model to adapt to changing imaging depths without introducing artifacts, maintaining high frames per second (fps) even with intensive preprocessing and post-processing.

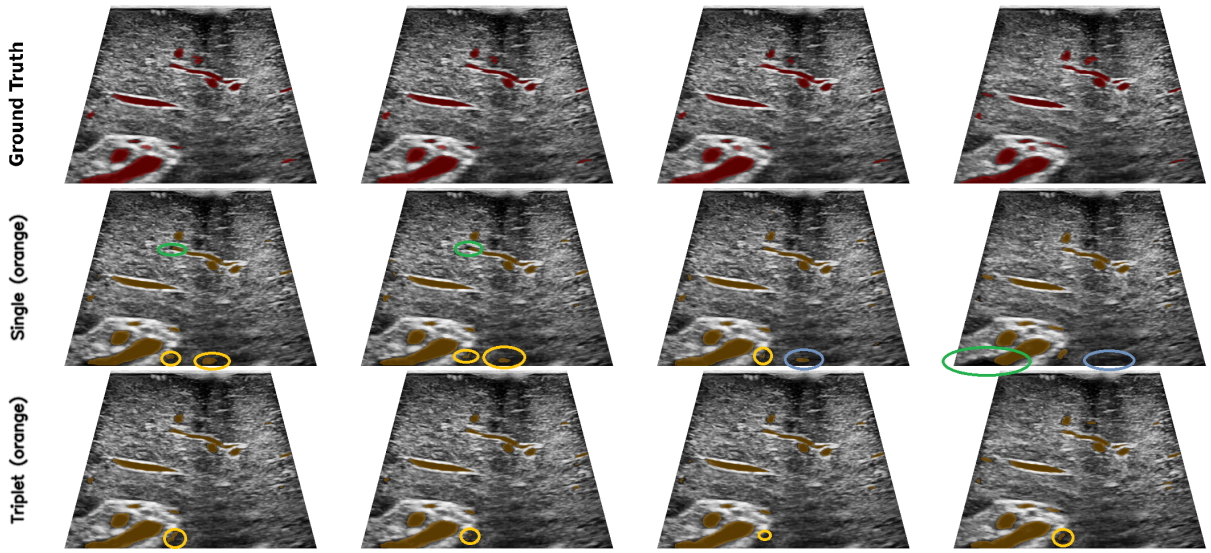

Figure 3. Qualitative comparison of vessel segmentation results from single-frame and triplet-based ResNet18 U-Net models. Green circles highlight false negatives (under-predicted vessels), while yellow circles mark false positives (over-predicted regions), and blue circles indicate temporal flipping across frames.

The real-time performance metrics of the tested models are summarized in Table 3. The final column reflects the results observed during continuous video playback in fps. Despite InceptionResNetV2 achieving the highest Dice score, it was the slowest for inference. The literature suggests that a frame rate of 30 fps is generally sufficient for real-time performance [31, 32]. A couple of tested models, including ResNet18, ResNet50, and MobileNet_v2 reached and sustained this fps target, with ResNet18 also allowing for potential further tuning if needed. Notably, this model delivered segmentation quality comparable to that of InceptionResNetV2 while achieving a stable throughput of 40 fps, making it the most viable candidate for intraoperative segmentation deployment.

Table 3. Comparison of model complexity and performance. Parameter count, forward-pass inference time, model size, and measured real-time fps (including preprocessing and postprocessing in the ultrasound pipeline) are reported.

Encoder name	Parameters	Time (ms)	Model size (MB)	FPS
ResNet18	14,328,209	**5.48**	56.08	**40**
ResNet50	32,521,105	9.03	127.38	36
MobileNet_v2	6,628,945	7.55	26.17	37
Vanilla U-Net	31,037,633	12.63	121.33	29
DenseNet121	13,607,633	16.97	53.78	29
IncResNetV2	62,029,297	29.21	243.04	21

4 Conclusion

In this paper, we presented a five-step preprocessing framework combined with a triplet-based ResNet18 U-Net model for real-time laparoscopic ultrasound image segmentation, achieving competitive Dice scores for liver vessel segmentation. Key contributions include a dynamic contour detector that improved generalization across varying depths and a triplet input setup that enhanced the temporal stability of vessel segmentation. We also evaluated real-time performance and mask quality, confirming the feasibility of deployment. Future work will focus on direct integration with live ultrasound streams and extension to 3D vessel reconstruction.

Acknowledgments

Anonymous Full Paper
Submission 42

References

[1] The World Cancer Research Fund. *Liver cancer statistics*. 2022. URL: https://www.wcrf.org/preventing-cancer/cancer-statistics/liver-cancer-statistics/.

[2] E. P. Weledji, G. Enow Orock, M. N. Ngowe, and D. S. Nsagha. "How grim is hepatocellular carcinoma?" In: *Ann. Med. Surg.* 3.3 (2014), pp. 71–76. ISSN: 20490801. DOI: 10.1016/j.amsu.2014.06.006.

[3] E. Kazam. *Ultrasound Teaching Manual: The Basics of Performing and Interpreting Ultrasound Scans*. 1999. DOI: 10.1016/s0899-7071(99)00133-3.

[4] M. Baad, Z. F. Lu, I. Reiser, and D. Paushter. "Clinical significance of US Artifacts". In: *Radiographics* 37.5 (2017), pp. 1408–1423. ISSN: 15271323. DOI: 10.1148/rg.2017160175.

[5] J. F. Gerstenmaier and R. N. Gibson. "Ultrasound in chronic liver disease". In: *Insights Imaging* 5.4 (2014), pp. 441–455. ISSN: 18694101. DOI: 10.1007/s13244-014-0336-2.

[6] "Dense-PSP-UNet: A neural network for fast inference liver ultrasound segmentation". In: *Comput. Biol. Med.* 153.September 2022 (2023), p. 106478. ISSN: 18790534. DOI: 10.1016/j.compbiomed.2022.106478.

[7] P. R. Hoskins. "Principles of doppler ultrasound". In: *Diagnostic Ultrasound, Third Ed. Phys. Equip.* (2019), pp. 143–158. DOI: 10.1201/b14901-8.

[8] "Clinical applications of contrast-enhanced ultrasound in vascular surgery: State-of-the-art narrative and pictorial review". In: *JVS-Vascular Insights* 3 (2025), p. 100254. ISSN: 29499127. DOI: 10.1016/j.jvsvi.2025.100254.

[9] D. L. Pham, C. Xu, and J. L. Prince. "Current methods in medical image segmentation". In: *Annu. Rev. Biomed. Eng.* 2.1 (2000), pp. 315–337. DOI: 10.1146/annurev.bioeng.2.1.315.

[10] "Segmentation of Liver Blood Vessel in Ultrasound Images Using Mask R-CNN". In: *Adv. Biomed. Eng.* 13 (2024), pp. 379–388. DOI: 10.14326/abe.13.379.

[11] J. Zeng, D. Jha, E. Aktas, E. Keles, A. Medetalibeyoglu, M. Antalek, R. Lewandowski, D. Ladner, A. A. Borhani, G. Durak, and U. Bagci. "A Reverse Mamba Attention Network for Pathological Liver Segmentation". In: (2025), pp. 0–7. DOI: 10.48550/arXiv.2502.18232. arXiv: 2502.18232v2.

[12] H. B. Jenssen, V. Nainamalai, E. Pelanis, R. P. Kumar, A. Abildgaard, F. K. Kolrud, B. Edwin, J. Jiang, J. Vettukattil, O. J. Elle, and Å. smund Avdem Fretland. "Challenges and artificial intelligence solutions for clinically optimal hepatic venous vessel segmentation". In: *Biomedical Signal Processing and Control* 106 (2025), p. 107822. DOI: https://doi.org/10.1016/j.bspc.2025.107822.

[13] W. Weng and X. Zhu. "UNet: Convolutional Networks for Biomedical Image Segmentation". In: *IEEE Access* 9 (May 2015), pp. 16591–16603. ISSN: 21693536. DOI: 10.1109/ACCESS.2021.3053408. arXiv: 1505.04597.

[14] M. Regalado, N. Carreras, C. Mata, A. Oliver, X. Lladó, and T. Agut. "Automatic Segmentation of Sylvian Fissure in Brain Ultrasound Images of Pre-Term Infants Using Deep Learning Models". In: *Ultrasound Med. Biol.* 51.3 (2025), pp. 543–550. ISSN: 1879291X. DOI: 10.1016/j.ultrasmedbio.2024.11.016.

[15] A. Sulaiman, V. Anand, S. Gupta, A. Rajab, H. Alshahrani, M. S. Al Reshan, A. Shaikh, and M. Hamdi. "Attention based UNet model for breast cancer segmentation using BUSI dataset". In: *Sci. Rep.* 14.1 (2024), pp. 1–13. ISSN: 20452322. DOI: 10.1038/s41598-024-72712-5.

[16] "D-TransUNet: A Breast Tumor Ultrasound Image Segmentation Model Based on Deep Feature Fusion". In: *J. Artif. Intell. Med. Sci.* 0.0 (2024), pp. 00–00. DOI: 10.55578/joaims.240226.001.

[17] M. Awais, M. Al Taie, C. S. O'Connor, A. H. Castelo, B. Acidi, H. S. Tran Cao, and K. K. Brock. "Enhancing Surgical Guidance: Deep Learning-Based Liver Vessel Segmentation in Real-Time Ultrasound Video Frames". In: *Cancers (Basel).* 16.21 (2024). ISSN: 20726694. DOI: 10.3390/cancers16213674.

[18] E. Smistad, T. Lie, and K. F. Johansen. "Real-time segmentation of blood vessels, nerves and bone in ultrasound-guided regional anesthesia using deep learning". In: *IEEE Int. Ultrason. Symp. IUS* (2021), pp. 1–4. ISSN: 19485727. DOI: 10.1109/IUS52206.2021.9593525.

[19] J. Ramalhinho, B. Koo, N. Montaña-Brown, S. U. Saeed, E. Bonmati, K. Gurusamy, S. P. Pereira, B. Davidson, Y. Hu, and M. J. Clarkson. "Deep hashing for global registration of untracked 2D laparoscopic ultrasound to CT". In: *Int. J. Comput. Assist. Radiol. Surg.* 17.8 (2022), pp. 1461–1468. ISSN: 18616429. DOI: 10.1007/s11548-022-02605-3.

[20] "Triplet Temporal-based Video Recognition with Multiview for Temporal Action Localization". In: *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.* 2023-June.i (2023), pp. 5428–5434. ISSN: 21607516. DOI: 10.1109/CVPRW59228.2023.00573.

[21] A. M. Reza. "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement". In: *J. VLSI Signal Process. Syst. Signal Image. Video Technol.* 38.1 (2004), pp. 35–44. ISSN: 13875485. DOI: 10.1023/B:VLSI.0000028532.53893.82.

[22] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei. "ImageNet: A large-scale hierarchical image database". In: *2009 IEEE Conf. Comput. Vis. Pattern Recognit.* Vol. 117. 3. IEEE, June 2009, pp. 248–255. ISBN: 978-1-4244-3992-8. DOI: 10.1109/CVPR.2009.5206848.

[23] F. Milletari, N. Navab, and S. A. Ahmadi. "V-Net: Fully convolutional neural networks for volumetric medical image segmentation". In: *Proc. - 2016 4th Int. Conf. 3D Vision, 3DV 2016* (2016), pp. 565–571. DOI: 10.1109/3DV.2016.79. arXiv: 1606.04797.

[24] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar. "Focal Loss for Dense Object Detection". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 42.2 (2020), pp. 318–327. ISSN: 19393539. DOI: 10.1109/TPAMI.2018.2858826. arXiv: 1708.02002.

[25] I. G. Courville, Y. Bengio, and Aaron. *Deep Learning*. MIT Press, 2016, Chapter 6: Deep Feedforward Networks. URL: http://www.deeplearningbook.org.

[26] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. "PyTorch: An imperative style, high-performance deep learning library". In: *Adv. Neural Inf. Process. Syst.* 32.NeurIPS (2019). ISSN: 10495258. DOI: 10.48550/arXiv.1912.01703. arXiv: 1912.01703.

[27] S. Varghese, Y. Bayzidi, A. Bar, N. Kapoor, S. Lahiri, J. D. Schneider, N. Schmidt, P. Schlicht, F. Huger, and T. Fingscheidt. "Unsupervised temporal consistency metric for video segmentation in highly-automated driving". In: *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.* 2020-June (2020), pp. 1369–1378. ISSN: 21607516. DOI: 10.1109/CVPRW50498.2020.00176.

[28] H. Wei, J. Ma, Y. Zhou, W. Xue, and D. Ni. "Co-learning of appearance and shape for precise ejection fraction estimation from echocardiographic sequences". In: *Med. Image Anal.* 84.April 2021 (2023), p. 102686. ISSN: 13618423. DOI: 10.1016/j.media.2022.102686.

[29] M. Rebol and P. Knöbelreiter. "Frame-To-Frame Consistent Semantic Segmentation". In: (2020). DOI: 10.3217/978-3-85125-752-6-18. arXiv: 2008.00948.

[30] F. Perazzi, J. Pont-Tuset, B. McWilliams, L. V. Gool, M. Gross, and A. Sorkine-Hornung. "A benchmark dataset and evaluation methodology for video object segmentation". In: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2016-December (2016), pp. 724–732. ISSN: 10636919. DOI: 10.1109/CVPR.2016.85.

[31] S. Vaze, W. Xie, and A. I. Namburete. "Low-Memory CNNs Enabling Real-Time Ultrasound Segmentation towards Mobile Deployment". In: *IEEE J. Biomed. Heal. Informatics* 24.4 (2020), pp. 1059–1069. ISSN: 21682208. DOI: 10.1109/JBHI.2019.2961264.

[32] T. Natali, A. Zhylka, K. Olthof, J. N. Smit, T. R. Baetens, N. F. M. Kok, K. F. D. Kuhlmann, O. Ivashchenko, T. J. M. Ruers, and M. Fusaglia. "Automatic hepatic tumor segmentation in intra-operative ultrasound: a supervised deep-learning approach". In: *J. Med. Imaging* 11.02 (2024). ISSN: 23294310. DOI: 10.1117/1.jmi.11.2.024501.