Evaluating Denoising Methods for 3D Ultrasound Uterus Segmentation with DynUNet

Marija Habijan, Juraj Perić, Irena Galić
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Croatia
marija.habijan@ferit.hr

2025 International Symposium ELMAR | 979-8-3315-9679-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/ELMAR66948.2025.11193753

***Abstract*—The uterus is a key reproductive organ whose morphology affects fertility and pregnancy outcomes. Recent advances in 3D transvaginal ultrasound (US) enable detailed volumetric imaging of uterine anatomy. However, automated analysis of 3D US volumes remains challenging since the uterus exhibits large inter-patient variability, US images suffer from speckle noise, low contrast, and artifacts. To address this, we propose a deep-learning (DL) pipeline for fully automatic 3D uterus segmentation using a DynUNet model. We investigate several denoising methods including contrast enhancement, filtering, wavelet denoising and learning-based denoising to evaluate their effect on the final segmentation accuracy. Our experiments demonstrate that denoising with DnCNN significantly improves segmentation performance, achieving a Dice Similarity Coefficient (DSC) of 71.95% on the test set and 11% improvement over the baseline. These results show the importance of denoising methods for robust segmentation in noisy ultrasound images.**

***Keywords*—DynUNet; Medical Segmentation; Ultrasound Volumes; Ultrasound Denoising**

I. Introduction

The uterus plays a central role in female reproductive health because its shape and volume are closely related to fertility and pregnancy outcomes. Three-dimensional (3D) ultrasound has significantly improved visualization of the internal female genital tract, allowing non-invasive imaging of uterine anatomy. Despite this, large-scale quantitative studies of uterine morphology are still lacking. A key obstacle is the difficulty of obtaining accurate 3D ultrasound segmentation of the uterus. US images suffer from speckle noise and shadowing, the uterine cavity can change position and shape between scans while variations in bladder filling introduce noise and organ motion. All of these factors make the uterine boundary difficult to delineate.

Some prior work has tackled 3D ultrasound uterus segmentation. Early studies focused on semi-automatic approaches that require manual initialization. More recently, deep learning (DL) methods have been commonly applied. Behboodi et al. [1] used 2D U-Net models trained on multiple planes (sagittal, coronal, axial) of 3D ultrasound volumes, addressing challenges such as positional variations and low image quality. Singh et al. [2] utilized a multi-modal approach, combining B-mode and power Doppler ultrasound scans to exploit complementary imaging information. Their method used different fusion strategies: early, intermediate, and late fusion, together with U-Net and U-Net++ architectures to segment the placenta in 3D volumes. Qureshi et al. [3] inspected different DL models (U-Net, FD-U-Net, and Attention U-Net) for automated segmentation of the Levator Ani Muscle (LAM) from 3D endovaginal ultrasound images. Looney et al. [4] developed a fully automated technique for placenta segmentation from 3D ultrasound volumes using a convolutional neural network called OxNNet. Bone et al. [5] gathered a large UterUS dataset [6] of 3D transvaginal scans and trained an nnU-Net-based model to segment the uterine cavities. Their work combined segmentation with a shape-alignment analysis pipeline.

In this work, we focus on uterus segmentation and impact of different denoising methods on segmentation accuracy. We apply a 3D DynUNet model [7] to the task of uterus segmentation in 3D ultrasound and we evaluate different denoising methods. We hypothesize that denoising can improve final segmentation accuracy. Our main contributions are:

- We demonstrate the use of a DynUNet model for fully automatic segmentation of the uterus in 3D ultrasound.
- We implement and compare the following denoising methods (1) CLAHE + Butterworth, (2) DWT + Threshold, (3) CLAHE + Median Filter, (4) Median-Mean Hybrid Filter, (5) Unsarp-Masking + Bilateral Filter, (6) Perona-Malik Anisotropic Diffusion and (7) 3D DnCNN denoising to show their influence on segmentation result.
- We provide quantitative analysis of how each denoising method affects segmentation performance, reporting mean Dice Similarity Coefficient (DSC) and improvements over a baseline.

The remainder of the paper is organized as follows. Section II describes used dataset, denoising methods and DynUnet model. Section III presents the experimental results and comparisons. Section IV concludes with a brief discussion and future work.

II. Method

In this work, we employ a 3D DynUNet model for uterus segmentation in 3D ultrasound volumes. We evaluate different image denoising methods prior to segmentation with DynUNet. Each 3D ultrasound volume was preprocessed by one of the following pipelines, as illustrated in Figure 1, and a separate DynUNet model was trained on each preprocessed dataset under identical training setup.

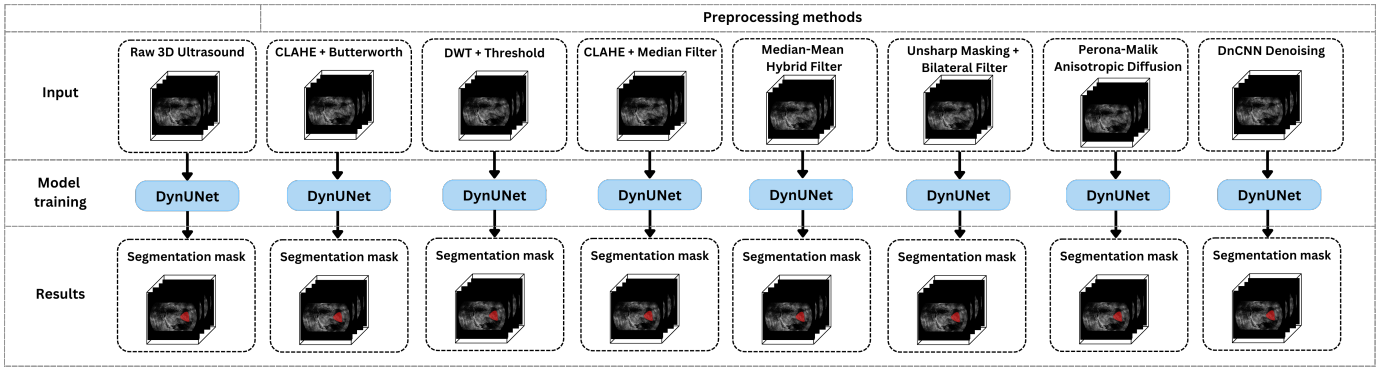

Figure 1. Multi-branch preprocessing and segmentation pipeline for 3D uterus segmentation using DynUNet. A raw 3D ultrasound image is processed through seven pipelines: (1) no preprocessing (baseline), (2) CLAHE + Butterworth filtering, (3) DWT + Thresholding, (4) CLAHE + Median filtering, (5) Median–Mean hybrid filtering, (6) Unsharp Masking + Bilateral filtering, (7) Perona–Malik diffusion, and (8) 3D DnCNN denoising model. Each branch's output is fed into a DynUNet model (blue) for segmentation, producing a mask of the uterus at the output.

A. Dataset Description

We use the publicly available UterUS dataset [6], which contains in total 315 3D transvaginal ultrasound volumes, including 141 with expert-annotated uterine cavity masks and 174 unannotated cases. Subjects are categorized into four clinical groups: (G) general population - women aged 18–28 with no pregnancy history; (I) unexplained infertility - at least six months of infertility with normal clinical assessments, (M) recurrent miscarriages - two or more spontaneous losses with normal karyotype and no thrombophilia and (RIF) recurrent implantation failure - two or more unsuccessful embryo transfers. All scans were acquired during the proliferative phase of the menstrual cycle and were acquired using General Electric ultrasound machines and Samsung devices.

B. Preprocessing Pipelines for DynUNet Segmentation

In this work, we evaluated different preprocessing methods to assess their influence on DynUNet-based 3D uterus segmentation. The used pipelines for denoising were: (1) no preprocessing (baseline), (2) CLAHE + Butterworth, (3) DWT + Threshold, (4) CLAHE + Median Filter, (5) Median-Mean Hybrid Filter, (6) Unsharp Masking + Bilateral Filter, (7) Perona–Malik Anisotropic Diffusion and (8) 3D DnCNN denoising model.

1) Baseline: The raw ultrasound images were fed directly into the DynUNet, serving as a baseline with no denoising or enhancement.

2) CLAHE + Butterworth: Contrast-Limited Adaptive Histogram Equalization (CLAHE) [8] was applied slice-wise (clip limit = 0.01) to enhance local contrast, followed by a Butterworth filter [9] (2nd-order, cutoff frequency = 0.05) to emphasize high-frequency components (edges and fine structures). This combination aims to improve uterus boundary visibility while suppressing low-frequency background.

3) DWT + Thresholding: A 3D Discrete Wavelet Transform [10] was used to decompose the input volume. The high frequency detail coefficients were thresholded (coefficients below 1% of the maximum were set to zero) and the volume was reconstructed with the inverse wavelet transform. This aims to reduce speckle noise while preserving essential structural details.

4) CLAHE + Median Filter: The CLAHE + Median Filter method enhances local contrast using CLAHE (clip limit = 0.01), followed by a $3 \times 3 \times 3$ median filter to reduce speckle noise while preserving anatomical edges.

5) Median–Mean Hybrid Filter: The Median–Mean Hybrid Filter [11] combines a $3 \times 3 \times 3$ median filter and a $3 \times 3 \times 3$ mean filter, aiming to balance edge preservation and global smoothing by averaging their outputs into a single denoised volume.

6) Unsharp Masking + Bilateral Filter: Unsharp Masking was first applied to enhance edge contrast by subtracting a Gaussian-blurred version of the image ($\sigma = 1.0$, amount = 1.0) from the original. The resulting sharpened image was then denoised using a bilateral filter (neighborhood diameter $d = 5$, intensity $\sigma = 75$, $spatial\sigma = 75$), which smooths homogeneous regions while preserving edge gradients. This combination aims to highlight subtle anatomical boundaries while controlling speckle noise without blurring critical regions.

7) Perona–Malik Anisotropic Diffusion: The Perona–Malik Anisotropic Diffusion [12] is a well-established method for reducing speckle noise in ultrasound images. We apply it over 100 iterative diffusion steps (conductance parameter $\kappa = 10$, integration constant $\gamma = 0.05$) to difuese homogeneous regions, while preserving edge discontinuities.

8) DnCNN Denoising: We trained a 3D DnCNN-based denoising model [13] to learn the speckle noise characteristics of our ultrasound data. The DnCNN model consisted of 10 convolutional layers (with 32 feature maps each) and was trained on the training set volumes to map noisy inputs to clean outputs. The trained DnCNN was applied to each volume to produce a denoised output, which was then used for DynUNet segmentation.

C. DynUNet Model

DynUNet is an advanced variant of U-Net that supports anisotropic convolution kernels and residual connections, dynamically adapting its architecture to the input spacing and

desired receptive field. The DynUNet follows a typical U-Net–style encoder–decoder with skip connections between corresponding resolution levels. In our configuration, there are five resolution stages in the encoder, including an initial input block, three down-sampling blocks, and a bottleneck block at the lowest resolution. The network takes a single-channel 3D input patch of size $128 \times 128 \times 128$ voxels. The encoder begins at full spatial resolution and progressively downsamples the feature maps by a factor of 2 at each stage. Specifically, the number of feature channels is doubled at each down-sampling level, from 16 channels in the first stage to 32, 64, 128, and finally 256 channels in the bottleneck stage. Convolution kernel sizes are specified per layer in an anisotropic manner: the first *conv block* uses a $3 \times 3 \times 3$ kernel, the next two use $5 \times 5 \times 5$ kernels, and the final encoder/bottleneck uses a $3 \times 3 \times 3$ kernel. Down-sampling is implemented by *strided convolutions* (stride 2) within the *conv blocks*, rather than separate pooling operations. The first encoder block operates at the full 128^3 resolution with stride 1 (no downsampling), while each subsequent encoder block applies a stride 2 convolution to halve the spatial dimensions. This yields feature maps at $1/2$, $1/4$, and $1/8$ of the original size, until the $1/8$ resolution bottleneck. Corresponding skip connections are forwarded from each encoder stage to the matching decoder stage to preserve spatial context.

All encoder and bottleneck blocks are implemented as residual units (res-blocks). Each such block consists of two successive 3D convolution layers, each followed by an instance normalization and leaky ReLU activation (negative slope 0.01) for nonlinear transformation. The residual design means that the input to the block is added to the output of the second convolution before the final activation. This skip connection within the block helps gradient flow and stabilizes training. When the number of channels increases (e.g., at downsampling stages), the identity path is adjusted with a $1\times1\times1$ convolution and appropriate normalization so that it can be summed with the output. Thus, each encoder stage (including the bottleneck) is a stack of two $3\times3\times3$ convolutions with a built-in residual shortcut.

The decoder mirrors the encoder with three up-sampling stages that restore the spatial resolution step by step. Each decoder stage begins with a transposed convolution using a kernel size of $2 \times 2 \times 2$ and stride 2, which up-samples the feature map by a factor of 2. The up-sampled feature map is then concatenated with the skip-connected features from the corresponding encoder level (channel dimensions are equal by design). After concatenation, a convolutional block processes the fused features. In DynUNet, each up-sampling block uses a *UnetUpBlock* consisting of the transposed convolution followed by a *UnetBasicBlock* convolution module. In our case, the decoder convolution block (after concatenation) is a pair of $3\times3\times3$ convolutions with instance norm and leaky ReLU that refine the features. This yields the decoded feature map for the next higher resolution. Through three such up-sampling stages, the decoder recovers to the full 128^3 resolution. Finally, the network ends with an output convolution layer of kernel size $1 \times 1 \times 1$ that maps the decoder feature channels at the highest resolution to the desired output channels (in this case, one probability map). This $1 \times 1 \times 1$ convolution is effectively the segmentation head that produces the voxel-wise prediction. We apply a sigmoid function to this output during inference to obtain a probability map for the foreground (uterus) class, since a single output channel is used to represent the foreground versus background.

D. Training Setup

Each DynUNet model was trained independently for segmentation using a 3D patch-based approach. Input volumes were resampled and cropped to a uniform size of $128 \times 128 \times 128$ voxels. Dataset was splited into training and validaion set (121 volumes), while test set consisted of 20 volumes. We used 5-fold cross-validation where each fold was used as a test, while the remaining four were used for training and validation, ensuring subject-level separation and balanced evaluation across the dataset. The network was optimized using the Adam optimizer with an initial learning rate of 10^{-4} and training was conducted for 100 epochs. The model was trained on a single foreground class (uterus), with a sigmoid activation applied to the final output layer. For inference, we used MONAI's sliding window inference with a region of interest size of 128^3, overlap factor of 0.6, and a batch size of 1. The best model checkpoint was selected based on the highest Dice Similarity Coefficient (DSC) observed on the validation set during training. All experiments were performed using PyTorch and MONAI on a Nvidia A6000 GPU. [1].

III. Results

All eight DynUNet models were trained using the same framework and hyperparameters to ensure a fair comparison. During training and validation, we computed the Dice Similarity Coefficient (DSC) for uterus segmentation. After training, each model was evaluated on a hold-out test set to assess generalization. Obtained results are shown in Table 1.

The baseline model (no preprocessing) achieved a DSC of 83.72% on the validation set and 60.60% on the test set. Several pipelines improved test performance, with DnCNN denoising yielding the best result (71.95% test DSC), outperforming the baseline by over 11 percentage points (Figure 2.). This highlights DnCNN's strength in suppressing speckle noise while preserving anatomical detail. The second-best method, wavelet-based DWT + thresholding, reached 69.61% test DSC, confirming the effectiveness of multi-resolution denoising. Moderate gains (4–5%) were seen with CLAHE + Median and CLAHE + Butterworth filters, which enhanced contrast and moderately reduced noise. Perona–Malik Anisotropic Diffusion achieved similar results by smoothing homogeneous regions while preserving edges. In contrast, Median–Mean Hybrid filtering and Unsharp Masking + Bilateral filtering underperformed, with test DSCs of 50.43% and 56.70%, respectively, due to over-smoothing or noise

[1]https://github.com/medi-train/Evaluating-Denoising-Methods-for-3D-Ultrasound-Uterus-Segmentation-with-DynUNet

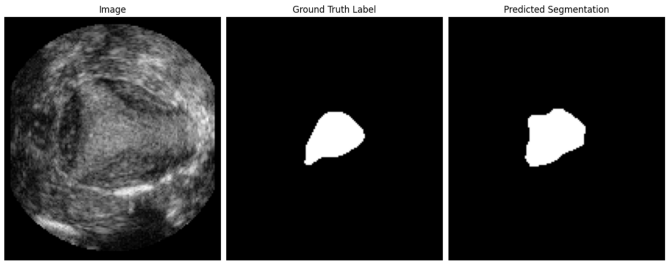

Figure 2. An example of test image, ground truth and obtained segmentation mask with DSC 92.03 %.

amplification. These results suggest that excessive filtering can degrade critical edge information or introduce artifacts which has bad influence on generalization. Overall, the most effective methods balanced noise reduction with edge preservation, with learned denoising (DnCNN) proving especially powerful for improving segmentation accuracy in 3D ultrasound.

TABLE I. Evaluation of the DynUNet segmentation models without and with different pre-processing pipelines using the DSC metric on validation and test set.

	3D DynUNet [DSC %]	
	Validation Set	*Test Set*
Baseline	83.72	60.60
CLAHE + Butterworth	84.50	64.36
DWT + Threshold	84.76	69.61
CLAHE + Median Filter	84.81	65.42
Median-Mean Hybrid Filter	86.43	50.43
Unsharp Maskin + Bilateral Filter	85.60	56.70
Perona-Malik	84.97	64.40
DNCNN	**85.76**	**71.95**

IV. Discussion and Conclusion

In this work, we evaluated the impact of seven denoising methods for improving 3D uterus segmentation using DynUNet model. The baseline model achieved a test Dice score of 60.60%, while the best performing method, 3DnCNN denoising, improved performance to 71.95%.

The comparative analysis shows that denoising methods have significantly different impact on segmentation performance, probably due to how well they balance noise suppression and edge preservation. Traditional methods like wavelet thresholding (69.61%) and CLAHE-based filters (up to 65%) provided moderate gains, while over-smoothing or edge-distorting methods, such as Median–Mean Hybrid and Unsharp Masking filtering, led to reduced accuracy. DnCNN, as a learning-based denoiser, outperformed all others because it adaptively removes speckle noise without blurring anatomical boundaries. Wavelet thresholding (DWT + threshold) also performed strongly, likely because it selectively reduces high-frequency noise while retaining uterus contours. In contrast, methods like CLAHE + Median or Anisotropic Diffusion provided modest gains by improving contrast or reducing noise uniformly, but lacked the adaptability needed to handle ultrasound's complex artifacts. The Median–Mean hybrid and Unsharp + Bilateral approaches performed poorly due to their tendency to either oversmooth critical textures or amplify high-frequency noise, which confused the segmentation network. These findings indicate that successful denoising methods for ultrasound segmentation must enhance discriminative features while avoiding distortion or oversimplification of the uterus structure.

Although we conducted experiments with transformer-based segmentation architectures (Swin-UNTER), DynUNet consistently outperformed them on our dataset. This suggests that further improvements may come from enhancing the input quality rather than increasing segmentation model complexity. Therefore, the future work will focus on exploring how use of advanced denoising models affect final segmentation performance.

Acknowledgment

This work was supported by the Croatian Science Foundation under the project number IP-2024-05-9492. The authors would like to thank UL-FRI-LGM for providing UterUS dataset.

References

[1] B. Behboodi, H. Rivaz, S. Lalondrelle, and E. J. Harris, "Automatic 3d ultrasound segmentation of uterus using deep learning," *2021 IEEE International Ultrasonics Symposium (IUS)*, pp. 1–4, 2021.

[2] S. Singh, G. N. Stevenson, B. Mein, A. Welsh, and A. Sowmya, "Automatic 3d multi-modal ultrasound segmentation of human placenta using fusion strategies and deep learning," *ArXiv*, vol. abs/2401.09638, 2024.

[3] A. Qureshi, K.-T. Hsu, Z. Asif, P. V. Chitnis, A. S. Shobeiri, and Q. Wei, "Deep learning based automatic segmentation of the levator ani muscle from 3d endovaginal ultrasound images," 2023.

[4] P. T. Looney, G. N. Stevenson, K. H. Nicolaides, W. Plasencia, M. Molloholli, S. Natsis, and S. L. Collins, "Fully automated, real-time 3d ultrasound segmentation to estimate first trimester placental volume using deep learning.," *JCI insight*, vol. 3 11, 2018.

[5] E. Boneš, M. Gergolet, C. Bohak, Žiga Lesar, and M. Marolt, "Automatic segmentation and alignment of uterine shapes from 3d ultrasound data," *Computers in biology and medicine*, vol. 178, p. 108794, 2024.

[6] E. Boneš, M. Gergolet, C. Bohak, Žiga Lesar, and M. Marolt, "Uterus: An annotated dataset of uteri in volumetric ultrasound data." https://github.com/UL-FRI-LGM/UterUS, 2024. Accessed: 2025-06-01.

[7] S. code for monai.networks.nets.dynunet, "Monai." https://docs.monai.io/en/0.3.0/_modules/monai/networks/nets/dynunet.html, 2020. Accessed: 2025-06-01.

[8] S. Pizer, R. Johnston, J. Ericksen, B. Yankaskas, and K. Muller, "Contrast-limited adaptive histogram equalization: speed and effectiveness," in *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*, pp. 337–345, 1990.

[9] A. Dogra and P. Bhalla, "Image sharpening by gaussian and butterworth high pass filter," *Biomedical and Pharmacology Journal*, vol. 7, pp. 707–713, 2014.

[10] M. Weeks and M. Bayoumi, "3d discrete wavelet transform architectures," in *ISCAS '98. Proceedings of the 1998 IEEE International Symposium on Circuits and Systems (Cat. No.98CH36187)*, vol. 4, pp. 57–60 vol.4, 1998.

[11] A. Taguchi and Y. Murata, "The median and mean hybrid filters," in *1991 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 93–96 vol.1, 1991.

[12] "Ieee transactions on pattern analysis and machine intelligence, vol. 12. no. 7. july 1990 scale-space and edge detection using anisotropic diffusion,"

[13] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.