

ARTICLE

LAG-U-Net: A Lightweight Attention-Guided Network for Efficient Cardiac Ultrasound Segmentation with Cross-Dataset Evaluation

Nasim Ahmadzadeh Nobari Azar^{1,2}, Abdullahi Umar Ibrahim^{3,4,*}, Ammara Talib⁵, Kamil Dimililer⁶ and Mohamad Ziad Altabel⁷

¹Computer Information Systems, Faculty of Artificial Intelligence & Informatics, Near East University, Nicosia, Cyprus, Türkiye

²Computer Information Systems Research and Technology Center, Near East University, Mersin 10, Türkiye

³Department of Biomedical Engineering, Near East University, Mersin 10, Türkiye

⁴Department of Medical Biochemistry, Kaduna State University, Kaduna, Nigeria

⁵The Research Center for Science Technology and Engineering, Department of Biomedical Engineering, Near East University, Mersin 10, Türkiye

⁶Department of Electrical and Electronic Engineering, Applied Artificial Intelligence Research Centre (AAIRC), Center for Technology and Engineering (BILTEM), Near East University, Near East Boulevard, Nicosia, North Cyprus, via Mersin 10, Türkiye

⁷Department of Artificial Intelligence Engineering, Faculty of Artificial Intelligence & Informatics, Near East University, Near East Boulevard, Nicosia, North Cyprus-TRNC, via Mersin 10, Türkiye

*Corresponding Author: Abdullahi Umar Ibrahim. Email: abdullahi.umaribrahim@neu.edu.tr

Received: 23 June 2026; Accepted: 10 September 2026; Published: 21 October 2026

ABSTRACT: Accurate cardiac structure segmentation from echocardiographic images remains challenging because of speckle noise, low contrast, anatomical variability, and differences in image acquisition. At the same time, many high-performing segmentation networks impose substantial computational requirements. This study presents LAG-U-Net, a lightweight attention-guided encoder-decoder architecture designed to examine the trade-off between segmentation accuracy and computational efficiency in echocardiography. The model uses a compact U-Net-style backbone with attention-guided skip connections and was evaluated for four-class segmentation of the background, left-ventricular cavity, myocardium, and left atrium on the CAMUS dataset. Experiments used the official patient-level training, validation, and test partitions, synchronized training-only augmentation, class-aware loss optimization, multiple CNN and transformer-based baselines, and three independent random seeds. Across the three runs, LAG-U-Net achieved a mean foreground Dice Similarity Coefficient (DSC) of 0.8717 ± 0.0013 and an Intersection-over-Union (IoU) of 0.7837 ± 0.0017 . Its DSC was slightly lower than that of conventional U-Net (0.8769 ± 0.0024) but slightly higher than that of the parameter-matched LiteUNet (0.8687 ± 0.0025). LAG-U-Net required 1.95 million trainable parameters compared with 7.79 million for U-Net and achieved an NVIDIA L4 inference latency of 3.38 ms per image compared with 4.92 ms for U-Net, demonstrating a favorable accuracy-efficiency trade-off. External evaluation was additionally performed on the left-ventricular class of EchoNet-Dynamic without fine-tuning. DSC decreased from 0.9201 ± 0.0008 for the CAMUS left-ventricular class to 0.1578 ± 0.0087 on EchoNet-Dynamic, indicating substantial cross-dataset domain shift and limited zero-shot transfer. Overall, LAG-U-Net provides a compact segmentation architecture with competitive within-dataset performance and reduced computational requirements, while the external results emphasize the need for domain adaptation and broader multi-center validation before clinical deployment.

KEYWORDS: Cardiac ultrasound segmentation; echocardiography; lightweight deep learning; attention mechanism; medical image segmentation; CAMUS dataset

1 Introduction

Cardiovascular imaging has an important role in the diagnosis, prevention, and clinical assessment of cardiovascular diseases [1–3]. Recent studies have also discussed broader developments and innovations in cardiovascular and cardiac imaging [4–6]. Among the available cardiac imaging methods, echocardiography is widely used for the structural and functional assessment of the heart [3]. Echocardiographic examination provides ultrasound images covering the cardiac cycle, including the end-diastolic (ED) and end-systolic (ES) phases, which are important for evaluating cardiac function [3]. The analysis of cardiac structures and their boundaries is necessary for obtaining clinical measurements such as ventricular volume and ejection fraction. However, manual assessment can be time-consuming and dependent on the specialist's experience. Therefore, increasing attention has been given to automated analysis of echocardiographic images and computer-aided methods.

Artificial intelligence (AI) and deep learning have increasingly been applied to cardiac imaging and echocardiographic analysis, including segmentation and automated functional assessment [5,7–10]. Deep learning methods have also been investigated for ultrasound medical image segmentation [11]. The availability of public echocardiographic datasets has also supported the development and evaluation of these methods. CAMUS contains annotated echocardiographic data from 500 patients [12], while EchoNet-Dynamic contains more than 10,000 echocardiographic videos with expert annotations and associated cardiac measurements [13]. Automatic segmentation of structures of the heart that can be evaluated via ultrasound examination has become a popular topic in the sphere of scientific research [12]. Nevertheless, ultrasound images remain difficult to analyze due to speckle noise, acoustic shadowing, low contrast of tissues, differences in the angles from which these images were obtained, and the quality of the images. These aspects give rise to the issue of finding the borders of the structures of the heart. As a consequence, several specialists have been seeking ways of improving the results obtained through segmentation, starting from model-based technologies and ending with deep convolutional networks and transformer-based methods [14–16].

Before the wide application of DL technology, segmentation of the heart via ultrasound examination was usually performed through the use of either model-based methods or techniques developed by specialists. With the implementation of fully convolutional encoder-decoder architectures, the methodology for medical image segmentation underwent a major transformation [17]. The U-Net's ability to contract and expand using skip connections to capture both semantic context and spatial detail has made it the benchmark for image segmentation [17]. Many subsequent models based on the same architecture, including UNet++ and nnU-Net, resulted in improved multi-scale feature fusion, improved decoder refinement, and improved standardization of training protocols that enabled a very high level of performance across numerous biological image segmentation tasks [18]. Publicly available benchmarks in echocardiography, such as CAMUS and EchoNet-Dynamic, have accelerated the development of reproducible evaluations of deep models used for the purposes of left ventricle segmentation and functional assessment of the cardiac cycle [19]. While these and other advances provide clinically meaningful benefits to the patient, many CNN-based frameworks achieve improved accuracy through increasing network depth, increasing the width of channels, or increasing the complexity of architecture, all of which tend to result in increased computational cost and memory requirements associated with the models.

In recent years, transformer-based architectures have emerged to address the challenges associated with convolutional networks, including their mainly local receptive fields and inability to capture longer-range relationships [20]. Hybrid and pure transformer-based segmentation models, such as TransUNet [21], Swin-Unet [22], UNETR [23], and SegFormer [24], use self-attention mechanisms to improve global context modeling and multiscale representation learning for segmentation tasks. Although transformer-based

approaches generally produce an improvement over prior representation quality, they also incur a greater computational burden within the context of working with higher-resolution medical images or volumetric data [24]. Beyond medical image segmentation, computationally efficient learning approaches have also been investigated in other cardiovascular biomedical applications. Gerdroodbary and Salavatidezfouli [25] investigated surrogate modeling for cardiovascular fluid–structure interaction in the abdominal aorta and compared Proper Orthogonal Decomposition combined with Long Short-Term Memory (POD–LSTM) with a nonlinear CNN–LSTM approach. Their study demonstrated that reduced-order and deep learning surrogate models can decrease the computational burden of complex cardiovascular simulations while maintaining useful predictive capability under different physiological conditions. Although this application differs from echocardiographic image segmentation, it illustrates the broader importance of balancing predictive performance and computational efficiency in cardiovascular computational modeling. The computational cost associated with these models can have ramifications for echocardiography, where it may be necessary to achieve near real-time inference for clinical workflow processes or as part of a bedside system or in portable settings [26]. Therefore, while final model development within segmentation based on transformers has yielded a considerable amount of work in this area, the results have not translated to routine use or implementation when limited resources are available in an ultrasound-based environment, as the clinical experience required for this work is not comparable to that of non-clinical work instances [27].

In response to the ever-increasing computational requirements of modern segmentation networks, there has been an increasing amount of literature documenting lightweight architectures that attempt to provide similar levels of segmentation quality while also reducing the number of parameters, FLOPs, memory, and inference time [28,29]. These works demonstrate the value of maintaining both accuracy and efficiency in a segmentation model rather than just optimizing the overlap metrics [30]. Existing approaches include compact convolutional networks, knowledge distillation, mobile vision transformers, and lightweight attention mechanisms. However, many of these methods have been developed mainly for CT, MRI, dermoscopic images, or general segmentation tasks. Echocardiography presents a different set of difficulties, including speckle noise, low tissue contrast, variations in image acquisition, and differences in cardiac anatomy between patients.

It is evident from the recent works in the field of U-Net technology that this model is still at the extraction level of medical image segmentation, as challenges such as diverse types of data, generalizability, expenses, and clinical application still persist [31]. There has been an emergence of architectures that use attention-based models to make the process more effective. One representative example of an attention-modulated U-Net is Attention U-Net, which uses attention gates on the skip connections in order to make certain that the weights of the encoder are calculated as they are transmitted to the decoder [32]. In the context of ultrasound imaging, such a solution may be particularly useful because ultrasound images contain distracting factors such as noise and artifacts that result in the emergence of unwanted information in the output feature maps. Even though attention gating is an already-established approach, this research aims at evaluating its performance by means of directly comparing Attention U-Net with other models and carrying out ablation tests that will provide an answer on the importance of this approach.

If we turn our attention to the broader literature connected with cardiovascular AI, we will see that the importance of balancing efficiency with predictive performance in the area can be observed here as well. Kumar and Rekha [33] examined the possibility of using deep learning and optimization methods in the prediction of cardiovascular disorders and emphasized the importance of the use of effective learning strategies in this field. In another investigation, Gerdroodbary and Salavatidezfouli [25] compared linear and nonlinear surrogate models in the context of cardiovascular fluid-structure interaction using a number of methods, such as POD-LSTM and CNN-LSTM, and focusing on both the accuracy of predictions and the

overall expenses incurred. Although both works do not touch upon segmentation issues in echocardiography, their findings still support the idea of the necessity of implementing computational models that are both effective and easy in application.

Despite these developments, several challenges remain in echocardiographic segmentation. High-capacity CNN and transformer architectures can provide strong feature representations but may impose substantial parameter, memory, and inference requirements, whereas aggressive network compression can reduce representational capacity. Attention mechanisms provide one possible means of selectively refining skip-connected features, but their benefit must be considered together with the additional computational overhead and verified experimentally. Furthermore, many studies report results under a single experimental setting or dataset, making it difficult to distinguish architectural benefit from dataset-specific effects. Against this background, the present study investigates a compact three-level U-Net-based architecture with attention-gated skip connections and evaluates it using multiple segmentation and boundary metrics, three independent runs, controlled ablation, computational-efficiency profiling, statistical analysis, and zero-shot cross-dataset LV evaluation. Accordingly, the intended contribution is not a fundamentally new attention operation, but an experimentally evaluated lightweight attention-guided configuration for echocardiographic segmentation with explicit consideration of the accuracy-efficiency trade-off. This is the basis of the lightweight attention-based model proposed in this study. The principal contributions of the present work include:

- Development of a compact three-level encoder-decoder architecture with attention-gated skip connections to investigate the trade-off between segmentation accuracy and computational efficiency in echocardiographic segmentation.
- Integration of hierarchically learned convolutional features with multi-level attention-guided skip filtering to examine selective feature fusion within a lightweight segmentation framework.
- Evaluation on CAMUS together with LV-only zero-shot external evaluation on EchoNet-Dynamic to characterize cross-dataset transfer behavior and domain shift rather than to claim universal clinical generalizability.
- Implementation of a comprehensive evaluation protocol incorporating multiple overlap and boundary metrics, three independent training runs, paired statistical testing, computational-efficiency profiling, controlled ablation experiments, qualitative assessment, and failure-case analysis.
- Demonstration of reduced model complexity and favorable inference efficiency relative to conventional U-Net under the evaluated hardware environment, providing evidence of computational feasibility while not establishing real-time clinical or edge-device deployment readiness.

2 Methodology

CAMUS and EchoNet-Dynamic were chosen because they allow the proposed framework to be examined in two different settings. CAMUS served as the main dataset for model development and in-domain evaluation because it provides expert pixel-level annotations of the left ventricular cavity, myocardium, and left atrium at both end-diastolic and end-systolic phases. This makes it suitable for controlled multi-class segmentation training and evaluation. EchoNet-Dynamic, on the other hand, contains a much larger collection of echocardiographic videos collected under different imaging conditions and provides expert left-ventricular endocardial tracings. For this reason, it was used only for external cross-dataset evaluation of the left-ventricular class. The model trained on CAMUS was applied directly to EchoNet-Dynamic without fine-tuning so that the effect of changes in dataset and acquisition conditions could be examined. In this way, CAMUS provides the main in-domain benchmark, while EchoNet-Dynamic offers a more difficult external transfer setting. The results from this external evaluation are therefore used to examine domain shift and

should not be interpreted as proof of general clinical generalizability across different centers, devices, or patient populations. The overall methodology is summarized in Fig. 1.

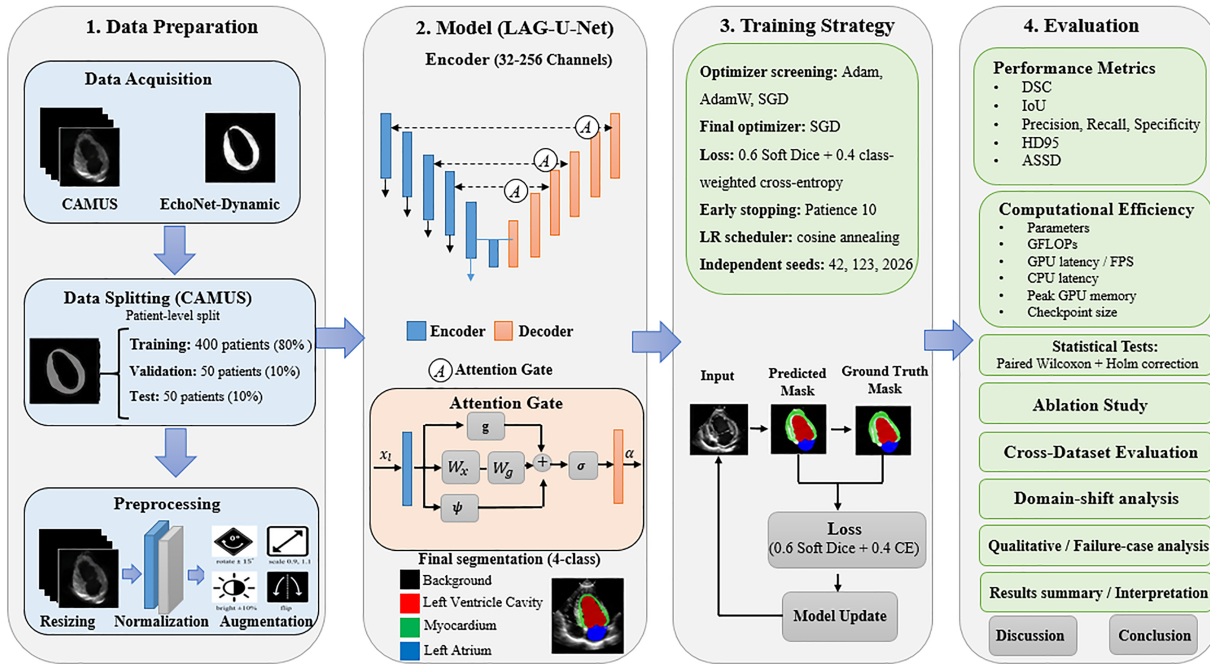


Figure 1: Overall workflow of the proposed LAG-U-Net framework. The pipeline includes data preparation, preprocessing, attention-guided segmentation architecture, training strategy, loss optimization, and evaluation methodology.

2.1 Datasets

2.1.1 CAMUS Dataset

The CAMUS (Cardiac Acquisitions for Multi-structure Ultrasound Segmentation) dataset contains echocardiographic data of the heart from 500 patients using the same echocardiographic views [12]. Each patient's dataset includes pixel-based annotations by an expert cardiologist of the left ventricle, myocardium, and left atrium at both end-diastole (ED) and end-systole (ES). Filtered frames from the ED and ES phases were then selected, resulting in 2000 images (visuals of each patient's heart) and their associated masks (the part of the frame where the heart is located). The CAMUS dataset was collected at the University Hospital of Saint-Etienne, France, and was anonymized in accordance with the requirements of the local ethics committee [12].

2.1.2 EchoNet-Dynamic Dataset

The EchoNet-Dynamic dataset is a large-scale echocardiography dataset consisting of more than 10,000 cardiac ultrasound videos collected from routine clinical examinations [13]. Each video is accompanied by expert annotations of the left ventricular endocardial border and associated clinical measurements such as ejection fraction (EF). In this study, EchoNet-Dynamic was not used as the primary training dataset for segmentation. Instead, it was employed to evaluate zero-shot cross-dataset transfer of the LAG-U-Net model under different echocardiographic acquisition and image-distribution characteristics. The EchoNet-Dynamic dataset was derived from echocardiographic examinations performed at Stanford Health Care under Stanford University Institutional Review Board oversight (Protocol 43721), with a waiver of individual

informed consent for the retrospective use of de-identified data [13]. A summary of the datasets, including their annotations and usage, is presented in Table 1.

Table 1: Summary of datasets used in this study (number of patients/videos, annotations, and usage).

Dataset	Patients/Videos	Annotations	Usage
CAMUS	500 patients	LV, Myocardium, LA masks	Training and segmentation evaluation
EchoNet-Dynamic	10,000+ videos	LV tracings, EF labels	Cross-dataset evaluation

Representative samples from the datasets, including echocardiographic images and their corresponding ground truth segmentation masks. As shown in Fig. 2, the echocardiographic images exhibit variability in shape and intensity, highlighting the challenges in accurate segmentation.

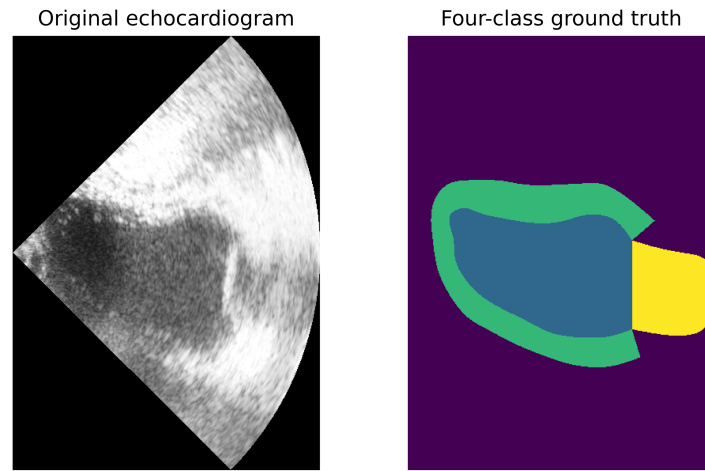


Figure 2: Representative CAMUS echocardiographic image and corresponding four-class ground-truth segmentation mask. The segmentation labels represent background, left-ventricular cavity, myocardium, and left atrium.

The present study used only publicly available, de-identified datasets and did not involve the recruitment of new participants or the collection of new identifiable clinical data.

2.2 Dataset Partitioning

The CAMUS dataset was partitioned at the patient level to prevent data leakage between the training, validation, and test sets. All images belonging to a given patient were assigned exclusively to one subset, ensuring that no patient appeared in more than one partition. This strategy provides a more reliable estimate of patient-independent performance and avoids potentially optimistic results that may arise when frames from the same patient are distributed across different subsets. The dataset has been separated into three sets after preprocessing and extracting frames, as follows:

- Training: 1600 samples (80%)
- Validation: 200 samples (10%)
- Test: 200 samples (10%)

A single fixed patient-level partition was used throughout all CAMUS experiments. The dataset was divided into 400 training patients (1600 images), 50 validation patients (200 images), and 50 test patients (200 images), corresponding to an 80%/10%/10% split. Patients were mutually exclusive across the three subsets,

ensuring that images from the same patient could not occur in more than one partition. The same training, validation, and test sets were retained for all evaluated models and experimental repetitions; therefore, no repeated random train–test resampling was performed. To assess variability due to model training rather than data partitioning, each main experiment was independently repeated three times using random seeds 42, 123, and 2026, and the resulting performance was summarized across these runs.

2.3 Data Preprocessing

Echocardiographic images exhibit substantial variability in intensity, contrast, anatomical appearance, and image quality, while speckle noise and acquisition-related artifacts can further complicate cardiac structure delineation. A standardized preprocessing pipeline was therefore applied to reduce unnecessary input variation while preserving the anatomical labels required for segmentation. For CAMUS, all ultrasound frames and their corresponding segmentation masks were resized to 256×256 pixels. Linear interpolation was used for ultrasound images, whereas nearest-neighbor interpolation was used for segmentation masks to preserve discrete class labels. Image intensities were normalized independently for each frame by first applying min-max normalization and subsequently rescaling the normalized values to approximately $[-1, 1]$.

Data augmentation was applied only to the training subset [34]. Spatial transformations were applied synchronously to each image and its corresponding mask to maintain pixel-level anatomical correspondence. The augmentation pipeline included random affine transformations with scaling between 0.90 and 1.10, rotation between -15° and $+15^\circ$, and translation of up to $\pm 3\%$, applied with a probability of 0.50. Random horizontal flipping was applied with a probability of 0.50. Image-only intensity augmentation included random brightness and contrast variation of up to $\pm 10\%$ with a probability of 0.30 and Gaussian noise augmentation with a probability of 0.20. The validation and test subsets were not randomly augmented and underwent only deterministic resizing and intensity normalization [35]. Because CAMUS and EchoNet-Dynamic differ in image acquisition, annotation format, and intensity distribution, the EchoNet-Dynamic frames used for zero-shot external evaluation were processed using the same deterministic evaluation pipeline rather than the training augmentation pipeline. The selected video frames and their reconstructed left-ventricular masks were jointly resized to 256×256 pixels, after which the images underwent the same intensity normalization used for CAMUS evaluation. No EchoNet-Dynamic images were used for model training or fine-tuning. Since EchoNet-Dynamic provides left-ventricular endocardial tracings rather than CAMUS-style multiclass masks, cross-dataset evaluation was restricted to the left-ventricular cavity, which represents the common anatomical target between the two datasets. A summary of the preprocessing pipeline applied in this study can be found in Table 2.

Table 2: Preprocessing and training-only data augmentation applied in the experimental pipeline. Geometric transformations were synchronized between images and segmentation masks, whereas stochastic augmentation was disabled during validation, CAMUS testing, and EchoNet-Dynamic external evaluation.

Operation	Setting	Applied to
Resize	256×256	Train/validation/test
Image interpolation	Linear	Images
Mask interpolation	Nearest-neighbor	Masks
Intensity normalization	Min-max, then rescale to $[-1, 1]$	All images
Random affine	Scale 0.90–1.10; rotation $\pm 15^\circ$; translation $\pm 3\%$; $p = 0.50$	Training only
Horizontal flip	$p = 0.50$	Training only
Brightness/contrast	$\pm 10\%$; $p = 0.30$	Training images only
Gaussian noise	Variance 2–12; $p = 0.20$	Training images only
Validation/test augmentation	None	Validation/test

Data augmentation was applied only to the CAMUS training set to increase the effective variability of the limited annotated training data and to reduce overfitting. The augmentation pipeline consisted of a random affine transformation with scaling in the range of 0.90–1.10, rotation between -15° and $+15^\circ$, and translation of up to $\pm 3\%$ of the image dimensions (probability = 0.50); horizontal flipping (probability = 0.50); random brightness and contrast adjustment within $\pm 10\%$ (probability = 0.30); and Gaussian noise perturbation (probability = 0.20). For geometric transformations, identical spatial operations were applied synchronously to each echocardiographic image and its corresponding segmentation mask so that anatomical correspondence was preserved. Bilinear interpolation was used for images, whereas nearest-neighbor interpolation was used for masks to avoid creating non-integer class labels. The relatively small geometric ranges were selected to reproduce plausible variations in image scale, probe orientation, and frame positioning without introducing substantial anatomical distortion. Brightness, contrast, and noise perturbations were included to reduce sensitivity to variations in ultrasound acquisition conditions and image quality. No stochastic augmentation was applied to the validation set, CAMUS test set, or EchoNet-Dynamic external evaluation data, thereby ensuring that performance was measured on unmodified evaluation samples.

To illustrate the preprocessing procedure, representative outputs from the executed training-time augmentation pipeline are presented in Fig. 3. The preprocessing pipeline resizes each image and mask to 256×256 pixels, normalizes image intensity, and applies synchronized geometric transformations to the image-mask pairs together with stochastic image-level appearance perturbations during training. The examples illustrate the range of transformations generated by the implemented augmentation procedure. No stochastic augmentation was applied during validation, CAMUS testing, or EchoNet-Dynamic external evaluation.

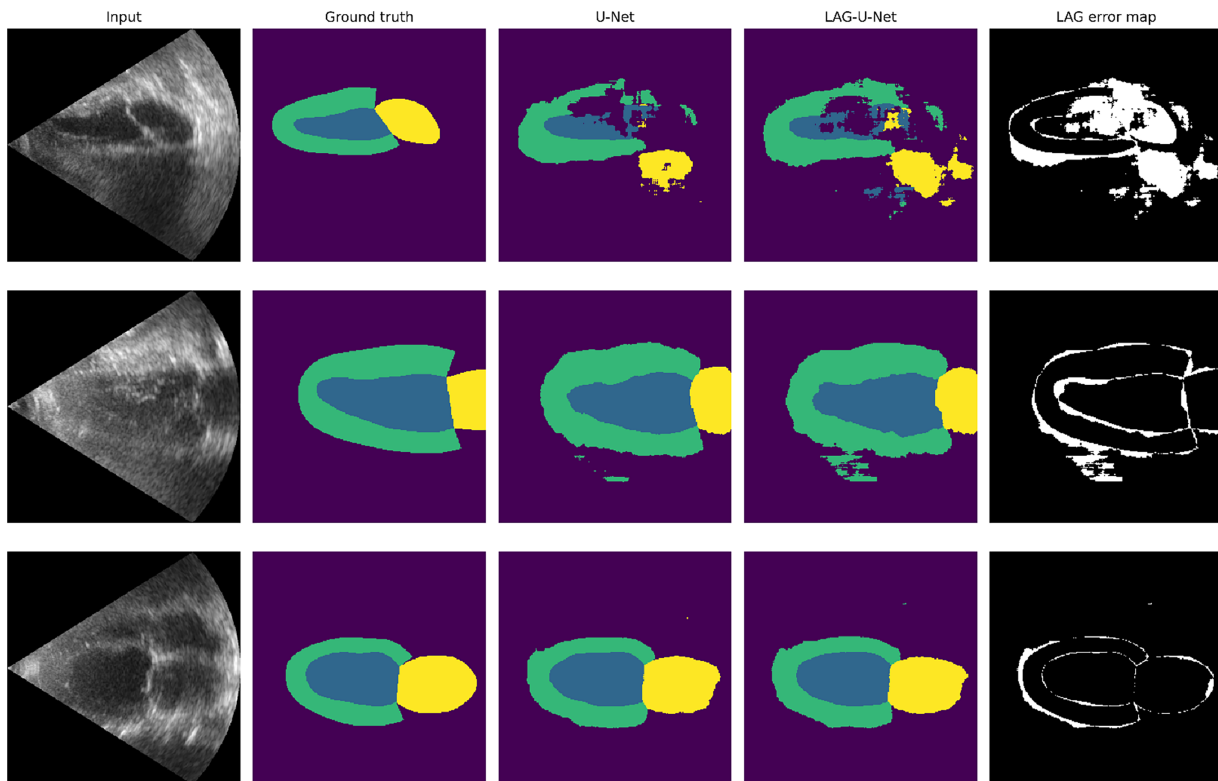


Figure 3: Representative outputs from the implemented preprocessing and training-time augmentation pipeline. The examples illustrate stochastic geometric and appearance transformations applied during CAMUS training together with the corresponding synchronized four-class segmentation masks. Validation, test, and external-evaluation images were processed deterministically without stochastic augmentation.

The overall training pipeline of the LAG-U-Net framework, including preprocessing, segmentation architecture, training strategy, and evaluation methodology, is illustrated in Fig. 4.

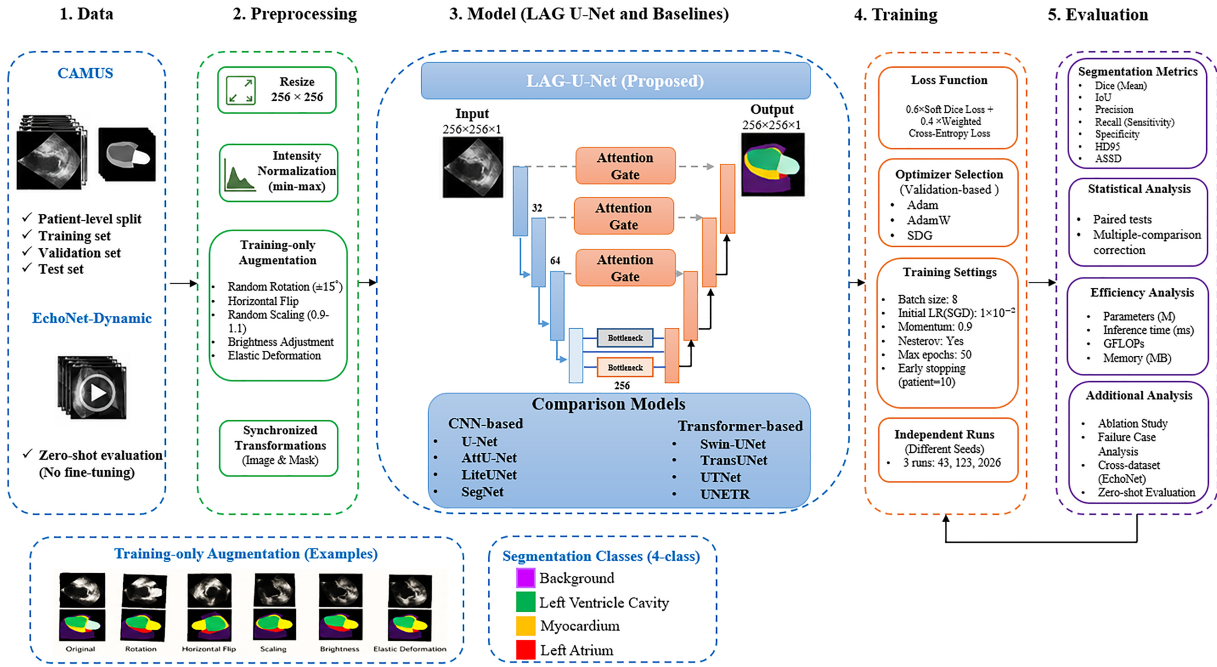


Figure 4: Detailed end-to-end workflow of the proposed LAG-U-Net experimental framework. The pipeline comprises CAMUS patient-level partitioning and EchoNet-Dynamic external evaluation, resizing to 256×256 pixels, intensity normalization, training-only synchronized augmentation, LAG-U-Net and comparison-model training, validation-based optimizer selection, combined soft Dice and class-weighted cross-entropy loss, three independent random seeds, multi-metric segmentation assessment, statistical analysis, computational-efficiency profiling, controlled ablation, qualitative and failure-case analysis, and zero-shot cross-dataset evaluation. The overall architecture of the implemented LAG-U-Net model is illustrated in Fig. 5.

2.4 LAG-U-Net Model Architecture

The proposed LAG-U-Net is a lightweight attention-guided encoder-decoder network developed for multi-class segmentation of echocardiographic images. The framework is designed to retain the spatial localization capability of U-Net while reducing model complexity and selectively filtering the encoder features transferred to the decoder. Given a preprocessed 256×256 grayscale echocardiographic image, the network produces pixel-wise predictions for four CAMUS classes: background, left-ventricular cavity, myocardium, and left atrium. Feature extraction in the proposed framework is performed automatically within the end-to-end convolutional network rather than through a separate handcrafted statistical-feature extraction stage. The convolutional encoder progressively transforms the preprocessed echocardiographic images into hierarchical feature maps, while the deeper encoder and bottleneck layers provide higher-level representations used for segmentation. The attention gates subsequently weight the encoder feature maps using the corresponding decoder gating signals before skip-connection fusion. Thus, manually defined statistical descriptors such as the mean, variance, skewness, kurtosis, entropy, or texture statistics are not used as input features in the proposed LAG-U-Net. Statistical quantities reported in the experimental analysis are calculated from the segmentation-performance measurements and their distributions and are therefore part of the evaluation procedure rather than the feature-extraction pipeline.

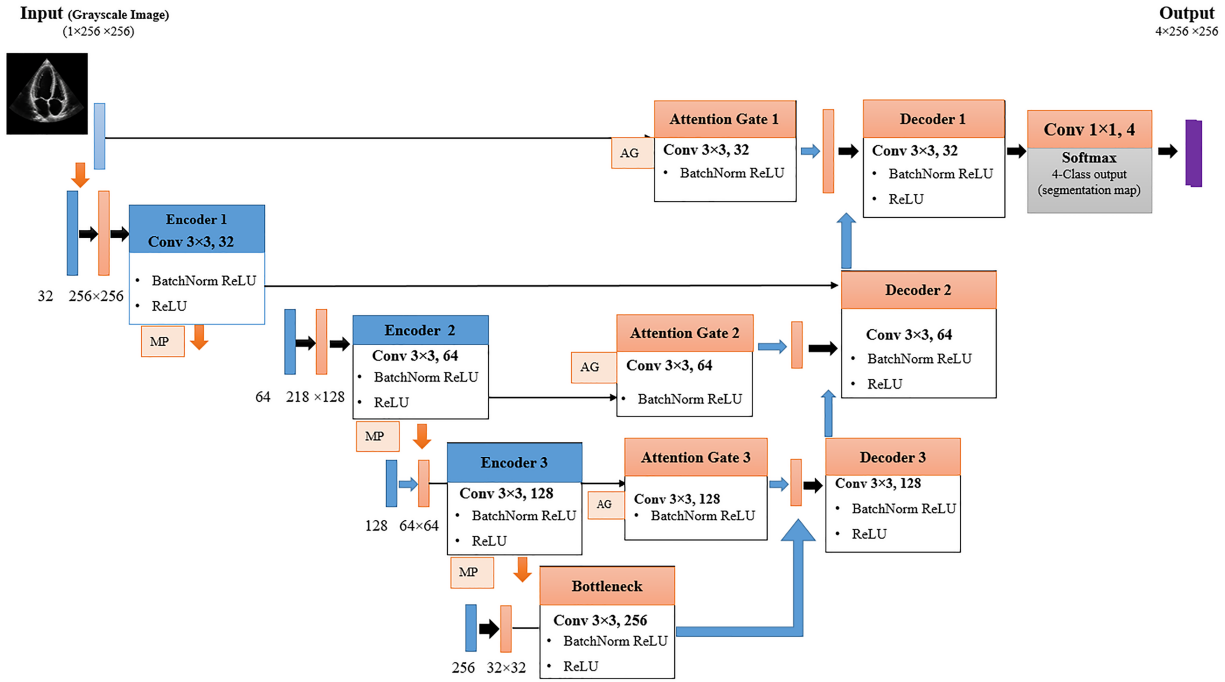


Figure 5: Architecture of the implemented LAG-U-Net for four-class echocardiographic segmentation. The network consists of a three-level lightweight encoder with 32, 64, and 128 feature channels, a 256-channel bottleneck, three decoder stages, and attention gates at all encoder-decoder skip connections. A final 1×1 convolution produces four output channels corresponding to background, left-ventricular cavity, myocardium, and left atrium.

Echocardiographic images present a distinctive segmentation challenge because clinically relevant anatomical boundaries may have weak local contrast, while encoder feature maps may simultaneously contain responses produced by speckle patterns, acoustic artifacts, and surrounding non-target structures. Conventional U-Net skip connections transfer these high-resolution encoder features directly to the decoder. Although this preserves fine spatial information, it can also transmit features that are not useful for identifying the target anatomy.

In LAG-U-Net, each skip connection is therefore modulated by an attention gate before feature concatenation. For an encoder feature map x and the corresponding decoder gating signal g , both representations are projected to an intermediate feature space and combined to estimate a spatial attention coefficient. The resulting coefficient map is applied multiplicatively to the encoder feature map before it is fused with the decoder representation. Consequently, skip features that are more compatible with the higher-level decoder context receive greater emphasis, whereas spatially less relevant responses are attenuated.

This mechanism is particularly relevant to low-contrast echocardiography because local intensity information alone may be insufficient to distinguish an anatomical boundary from the surrounding tissue. The decoder gating signal provides coarser semantic context that can guide selection of spatially relevant encoder information during reconstruction. Similarly, in the presence of speckle and other local image variability, the attention gate does not function as an explicit denoising operation; rather, it provides a learned mechanism for reducing the contribution of spatially irrelevant encoder responses before feature fusion. Because the attention coefficients are estimated independently for each input image, the gate is also able to adapt its spatial emphasis to variation in cardiac shape, size, and position instead of relying on a fixed anatomical region. Applying this filtering at multiple skip levels allows feature selection to occur across different spatial resolutions while preserving the compact convolutional backbone.

The attention-gating operation used in LAG-U-Net follows the established concept of attention-gated skip connections introduced in Attention U-Net [32], in which encoder features are selectively weighted using contextual information from the decoder before feature fusion. Therefore, the present study does not claim the invention of a fundamentally new mathematical attention operator. Instead, the contribution of LAG-U-Net lies in integrating multi-level attention-guided skip filtering into a deliberately reduced-width U-Net architecture and evaluating whether this combination can recover part of the representational capability lost through network compression while retaining low computational complexity.

More specifically, the implemented LAG-U-Net uses a base channel width of 32 and applies attention gating at all three encoder–decoder skip levels. This differs experimentally from the conventional Attention U-Net baseline evaluated in this study, which uses a wider base channel configuration of 64 with attention at the same three skip levels. The controlled ablation framework additionally contains a lightweight U-Net without attention and a lightweight single-attention variant. These configurations allow the effects of network width and the extent of attention-guided feature fusion to be examined separately. Accordingly, the architectural contribution of LAG-U-Net should be understood as a lightweight attention-guided configuration designed for an accuracy–efficiency trade-off in echocardiographic segmentation rather than as a new attention mechanism in isolation.

The encoder contains three resolution stages followed by a bottleneck. The numbers of feature channels are 32, 64, and 128 in the three encoder stages, respectively, with 256 channels in the bottleneck. Each convolutional block consists of two 3×3 convolutional layers, each followed by batch normalization and ReLU activation. A 2×2 max-pooling operation is used between successive encoder stages to reduce spatial resolution and progressively increase the receptive field.

The main architectural distinction of LAG-U-Net is the integration of attention gates into all three encoder-to-decoder skip connections. For each skip connection, the encoder feature map and the corresponding decoder gating signal are projected using 1×1 convolutions. Their transformed representations are combined and passed through ReLU activation, followed by a 1×1 convolution and sigmoid activation to generate spatial attention coefficients. These coefficients weight the encoder features before they are transferred to the decoder. Consequently, spatial regions considered more relevant to the current decoder representation receive greater emphasis, whereas less relevant background responses are attenuated before feature fusion. The attention mechanism therefore operates on the skip features rather than replacing the convolutional feature-extraction pathway.

The decoder contains three corresponding reconstruction stages. At each stage, the decoder feature map is upsampled using a 2×2 transposed convolution with stride 2. The upsampled decoder representation is then concatenated with the attention-filtered feature map from the corresponding encoder stage. The fused representation is refined using two 3×3 convolutional layers with batch normalization and ReLU activation. This process progressively restores the original spatial resolution while combining high-level contextual information with localized anatomical information from the encoder.

Finally, a 1×1 convolution maps the final decoder representation to four output channels. These channels represent class logits for the four anatomical categories, and the final segmentation label for each pixel is obtained from the class with the highest predicted score. Thus, the output layer does not perform binary sigmoid segmentation; the implemented model performs four-class segmentation.

Overall, the proposed workflow consists of: (1) preprocessing and normalization of the echocardiographic image; (2) hierarchical feature extraction through the lightweight encoder; (3) attention-based filtering of encoder features at each skip connection; (4) progressive reconstruction and feature fusion in the decoder; and (5) generation of a four-class cardiac segmentation map. The combination of reduced

channel dimensions and attention-guided skip connections is intended to investigate whether selective feature transfer can provide a favorable accuracy-efficiency trade-off without relying on a substantially larger segmentation architecture. The overall experimental workflow of the proposed LAG-U-Net framework, including dataset preparation, preprocessing, model training, optimizer selection, multi-metric evaluation, computational-efficiency analysis, ablation, and cross-dataset evaluation, is illustrated in Fig. 4.

2.5 Training Strategy

The training and model parameters were specified through a combination of predefined architectural settings and validation-based optimizer selection. The input resolution was fixed at 256×256 pixels and the batch size at 8. The proposed LAG-U-Net used a compact three-level encoder-decoder configuration with feature widths of 32, 64, and 128 channels and a 256-channel bottleneck. These architectural settings were fixed for the main experiments to implement the intended lightweight design rather than being selected through test-set optimization.

For optimization, Adam, AdamW, and stochastic gradient descent (SGD) were first compared under the same CAMUS training and validation partitions, augmentation pipeline, loss function, random seed, and 20-epoch evaluation limit. Adam and AdamW were initialized with a learning rate of 1×10^{-4} , whereas SGD used an initial learning rate of 1×10^{-2} with momentum of 0.9 and Nesterov acceleration. The optimizer achieving the highest validation DSC was selected for the subsequent main experiments. Under this protocol, SGD provided the highest validation DSC and was therefore used for the final training runs.

The main models were trained for a maximum of 50 epochs using a weight decay of 1×10^{-5} . A cosine-annealing learning-rate scheduler was applied throughout training, with the learning rate decreasing toward 1×10^{-6} . Early stopping with a patience of 10 epochs was based on validation foreground DSC, and the checkpoint corresponding to the highest validation DSC was retained for test evaluation. To reduce dependence on a single initialization, the principal experiments were independently repeated using random seeds 42, 123, and 2026.

The segmentation objective combined soft Dice loss and class-weighted cross-entropy loss with weights of 0.6 and 0.4, respectively. The Dice component optimized foreground overlap, whereas the weighted cross-entropy component accounted for differences in class representation. Importantly, the CAMUS test set was not used for optimizer or checkpoint selection. Table 3 summarizes the final training configuration.

Table 3: Final training and hyperparameter configuration used in the executed experiments.

Parameter	Final Setting
Framework	PyTorch
Input size	256×256
Number of output classes	4
LAG-U-Net encoder widths	32, 64, 128
Bottleneck width	256
Batch size	8
Maximum epochs	50
Optimizers compared	Adam, AdamW, SGD
Adam initial learning rate	1×10^{-4}
AdamW initial learning rate	1×10^{-4}
SGD initial learning rate	1×10^{-2}
Selected optimizer	SGD

(Continued)

Table 3 (continued)

Parameter	Final Setting
SGD momentum	0.9
SGD Nesterov acceleration	Yes
Weight decay	1×10^{-5}
Learning-rate scheduler	Cosine annealing
Minimum learning rate	1×10^{-6}
Early-stopping patience	10 epochs
Checkpoint criterion	Highest validation foreground DSC
Loss function	0.6 Soft Dice + 0.4 class-weighted cross-entropy
Independent random seeds	42, 123, 2026
Training device	NVIDIA L4 GPU

The validation-based optimizer comparison is summarized numerically in [Table 4](#), while the corresponding validation DSC trajectories are shown in [Fig. 6](#).

Table 4: Validation-based comparison of optimization algorithms for LAG-U-Net.

Optimizer	Initial Learning Rate	Best Validation DSC	Best Epoch
SGD	1×10^{-2}	0.862842	20
Adam	1×10^{-4}	0.862694	20
AdamW	1×10^{-4}	0.862425	20

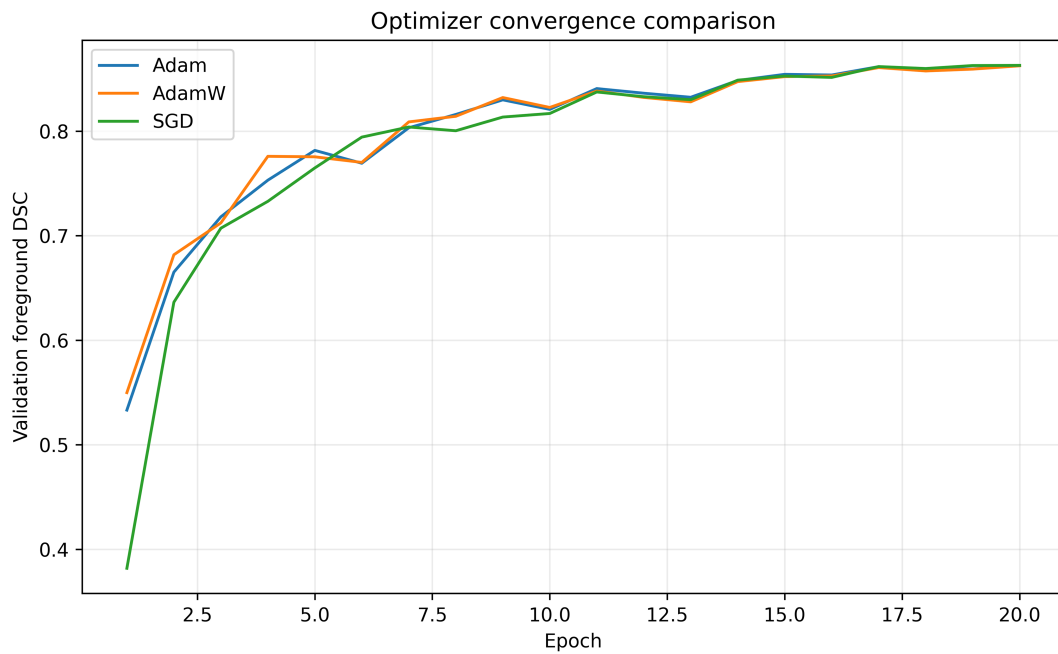


Figure 6: Convergence comparison of Adam, AdamW, and SGD for LAG-U-Net during the optimizer-screening experiment. Validation foreground DSC is plotted across epochs under identical data partitioning, augmentation, loss, and model settings.

SGD produced the highest validation DSC (0.862842), followed closely by Adam (0.862694) and AdamW (0.862425). The small differences indicate that LAG-U-Net was relatively insensitive to the three tested optimizers under the evaluated settings. Nevertheless, because optimizer selection was predefined according to the highest validation DSC, SGD was used in the subsequent experiments. No test-set information was used during optimizer selection.

2.6 Class Imbalance Handling

Pixel-level class frequencies were calculated from the CAMUS training masks before model optimization. The analysis showed a marked imbalance between the background and anatomical classes: background pixels represented approximately 75.75% of the training pixels, whereas the LV cavity, myocardium, and left atrium represented approximately 9.17%, 10.03%, and 5.06%, respectively. To reduce the tendency of the optimization process to favor the dominant background class, class-specific weights were calculated from the inverse square root of the training-set class frequencies and normalized to have a mean of one. The resulting weights were 0.3812 for background, 1.0958 for the LV cavity, 1.0478 for the myocardium, and 1.4752 for the left atrium.

The final training objective combined foreground soft Dice loss with weighted multiclass cross-entropy. Specifically, the total loss was defined as $0.6 \times \text{soft Dice loss} + 0.4 \times \text{weighted cross-entropy loss}$. The Dice component directly optimizes overlap for the foreground cardiac structures, whereas the class-weighted cross-entropy component assigns greater penalty to errors involving under-represented anatomical classes. The same class-weighting strategy and loss formulation were applied consistently to all models included in the main comparative experiments.

The loss function is defined as:

$$\mathcal{L} = 0.6 \mathcal{L}_{Dice} + 0.4 \mathcal{L}_{WCE} \quad (1)$$

$\mathcal{L}_{WCE}$ denotes the class-weighted multiclass cross-entropy loss. To quantify training variability while maintaining reproducibility, the principal experiments were independently repeated using random seeds 42, 123, and 2026. Reported performance values represent the mean $\pm$ standard deviation across the three runs.

2.7 Evaluation Metrics

Dice Similarity Coefficient (DSC)

The DSC measures the overlap between the predicted segmentation and the ground truth.

$$DSC = \frac{2|P \cap G|}{|P| + |G|} \quad (2)$$

where P represents the predicted mask and G represents the ground truth mask.

Intersection over Union (IoU)

The IoU metric measures the ratio between the intersection and union of predicted and ground truth segmentation regions.

$$IoU = \frac{|P \cap G|}{|P \cup G|} \quad (3)$$

In addition to DSC and IoU, precision, recall (sensitivity), specificity, the 95th-percentile Hausdorff distance (HD95), and the average symmetric surface distance (ASSD) were calculated to provide a more comprehensive assessment of segmentation performance.

Precision measures the proportion of pixels predicted as belonging to a cardiac structure that are correctly classified:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

Recall, also referred to as sensitivity, measures the proportion of ground-truth foreground pixels that are correctly identified:

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

Specificity quantifies the proportion of background pixels correctly classified as background:

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$$

Boundary accuracy was evaluated using HD95 and ASSD. HD95 represents the 95th percentile of the bidirectional distances between the predicted and ground-truth segmentation boundaries. Compared with the maximum Hausdorff distance, HD95 is less sensitive to isolated outlier pixels while still reflecting substantial boundary disagreement. ASSD represents the mean bidirectional surface distance between the predicted and ground-truth contours and therefore provides a complementary measure of average boundary agreement. Because the masks were evaluated after resizing to 256×256 pixels without physical pixel-spacing calibration, HD95 and ASSD are reported in pixel units.

For the primary CAMUS comparison, metrics were calculated for the foreground cardiac classes, averaged within each independent run, and subsequently summarized as mean $\pm$ standard deviation across the three random seeds.

Each experiment was repeated three times, and average results were reported. Cross-dataset evaluation was conducted by training on CAMUS and testing on EchoNet-Dynamic without fine-tuning. Model complexity was evaluated using parameter count and FLOPs computed with THOP.

2.8 Cross-Dataset Evaluation and Domain-Shift Analysis

To evaluate sensitivity to a change in data distribution, the LAG-U-Net models trained exclusively on CAMUS were applied to EchoNet-Dynamic without additional fine-tuning. This constitutes a zero-shot cross-dataset evaluation rather than a second within-dataset training experiment. Because CAMUS provides multi-class annotations of the left ventricular cavity, myocardium, and left atrium, whereas EchoNet-Dynamic provides left ventricular endocardial tracings, the cross-dataset comparison was restricted to the left ventricular class to maintain a compatible segmentation target.

Cross-dataset performance was evaluated using Dice similarity coefficient, Intersection-over-Union, precision, recall, specificity, HD95, and ASSD across the three independently trained CAMUS models. The results demonstrate a clear reduction in segmentation performance when the model is transferred from the CAMUS test distribution to EchoNet-Dynamic. This degradation indicates that the change in data distribution has a measurable influence on segmentation performance and therefore does not support an interpretation of domain-invariant generalization.

To characterize this domain gap beyond segmentation accuracy alone, image-level descriptors were calculated separately for CAMUS and EchoNet-Dynamic. These descriptors included mean image intensity, intensity standard deviation, image entropy, Laplacian variance, and foreground-area ratio. Their distributions are compared in Fig. 7. Differences across these descriptors indicate changes in intensity distribution, image texture and information content, local image sharpness, and the relative size of the annotated cardiac region. These properties provide quantitative evidence that the two datasets do not represent identical imaging distributions and offer a plausible context for the observed cross-dataset performance degradation.

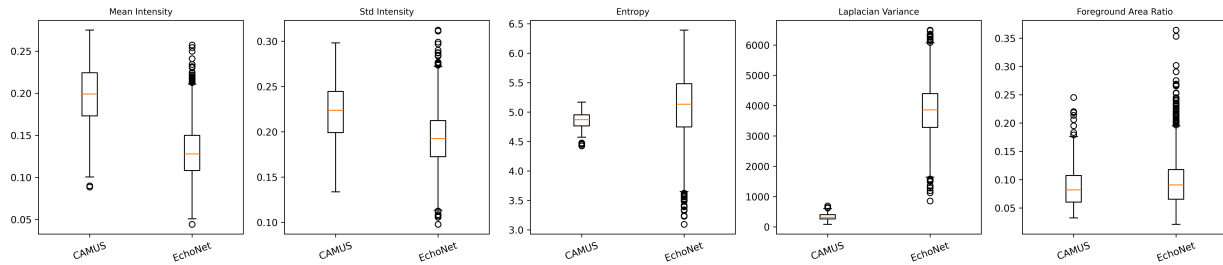


Figure 7: Domain-shift analysis between CAMUS and EchoNet-Dynamic. Boxplots compare normalized mean intensity, intensity standard deviation, image entropy, Laplacian variance, and LV foreground-area ratio across the two datasets. The observed distributional differences provide quantitative evidence of a substantial source-to-target domain shift that contributes to the deterioration in cross-dataset segmentation performance.

Failure-case analysis was additionally performed using the EchoNet-Dynamic samples with the lowest seed-averaged Dice scores. Representative examples are presented in Fig. 8 together with the input image, reference mask, predicted mask, and pixel-wise error map. The failure cases demonstrate that cross-dataset errors are not uniform across samples and become more pronounced in difficult images where boundaries are weakly defined, image appearance differs from the CAMUS training distribution, or the spatial extent of the left ventricle differs from common training examples. These observations should be interpreted as descriptive failure analysis rather than proof that a single image characteristic causes each segmentation error.

Taken together, the zero-shot segmentation results, quantitative image-distribution descriptors, and failure-case analysis demonstrate that domain shift is a substantive limitation of the current framework. Accordingly, the revised manuscript does not interpret EchoNet-Dynamic performance as evidence of robust cross-center or device-independent generalization. Instead, the experiment is used as a stress test that identifies the need for future domain adaptation, multi-center validation, and evaluation across heterogeneous ultrasound acquisition systems.

2.9 Computational Efficiency and Inference-Time Measurement

Computational efficiency was evaluated using trainable parameter count, floating-point operations (GFLOPs), GPU inference latency, CPU inference latency, GPU throughput in frames per second (FPS), peak GPU memory usage, and model checkpoint size. All models were profiled using a single grayscale input tensor with dimensions $1 \times 1 \times 256 \times 256$, corresponding to a batch size of one image. Parameter count was computed from trainable model parameters, while computational complexity was estimated using THOP. The number of multiply-accumulate operations (MACs) returned by the profiler was converted to GFLOPs as $2 \times \frac{MACs}{10^9}$.

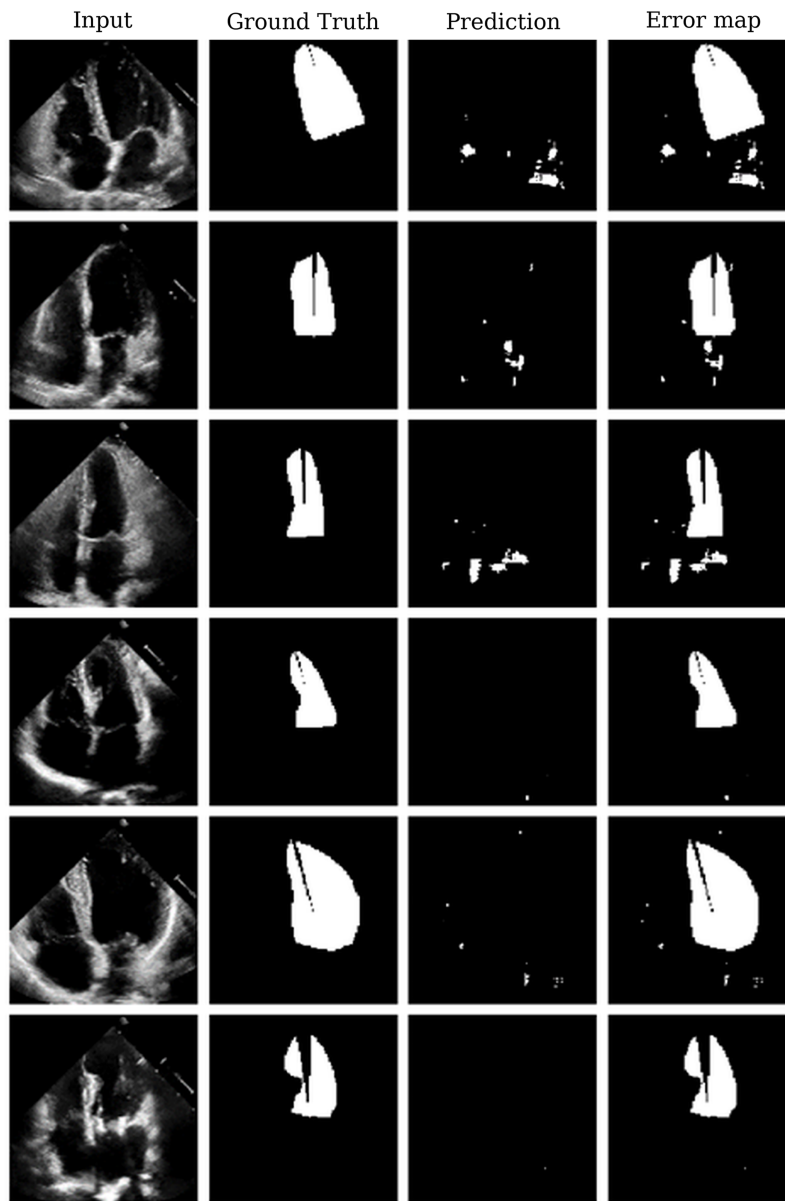


Figure 8: Representative failure cases from the CAMUS-to-EchoNet-Dynamic cross-dataset evaluation. Each row shows an EchoNet-Dynamic input frame, reconstructed ground-truth LV mask, LAG-U-Net prediction, and corresponding pixel-wise error map. Cases were selected from the lowest seed-averaged DSC values and illustrate segmentation errors associated with the substantial domain difference between the training and external evaluation datasets.

GPU inference latency was measured after placing the trained model in evaluation mode on the GPU. Twenty untimed forward passes were first performed as warm-up iterations. GPU timing was then measured over 100 forward passes using CUDA events. CUDA synchronization was performed before and during timing to ensure that asynchronous GPU operations were completed before elapsed time was recorded. The reported GPU latency represents the arithmetic mean of the 100 measured forward-pass times in milliseconds per image. GPU throughput was subsequently calculated as:

$$FPS = \frac{1000}{\text{mean GPU latency (ms)}}$$

CPU inference latency was measured separately using an identical model instance loaded on the CPU. Five untimed warm-up forward passes were followed by 30 timed forward passes using a high-resolution performance timer, and the mean elapsed time was reported in milliseconds per image. Peak GPU memory was measured after resetting the CUDA peak-memory statistics and was reported as the maximum allocated GPU memory during profiling. Model checkpoint size was obtained directly from the stored checkpoint file. All latency measurements represent model forward-pass inference only and therefore exclude disk input/output, image preprocessing, CPU-to-GPU data transfer, and subsequent clinical workflow operations. The experiments were executed in a CUDA-enabled Google Colab environment using an NVIDIA L4 GPU, PyTorch 2.11.0+cu128, CUDA 12.8, and Python 3.12.13.

3 Results

The CAMUS dataset was split into training, validation, and test sets per the official protocol. For training, all input images were resized to the dimension of 256×256 pixels and normalized. Following the validation-based optimizer screening, SGD was selected for the main experiments and used with an initial learning rate of 1×10^{-2} , momentum of 0.9, Nesterov acceleration, a batch size of 8, and a maximum training duration of 50 epochs. To examine optimization behavior and select the training optimizer independently of the final test results, LAG-U-Net was trained using Adam, AdamW, and SGD under an identical 20-epoch screening protocol. The best validation DSC values were 0.862694 for Adam, 0.862425 for AdamW, and 0.862842 for SGD. All three optimizers reached their highest recorded validation DSC at epoch 20, and their validation trajectories did not indicate unstable or divergent optimization during the comparison period. The numerical differences among the three optimizers were small; therefore, these results should not be interpreted as evidence of a substantial superiority of one optimization algorithm over the others. Nevertheless, following the predefined validation-based selection criterion, SGD provided the highest validation DSC and was subsequently used for the main experiments. The corresponding validation DSC trajectories are shown in [Fig. 6](#).

To provide a broader architectural evaluation, the proposed LAG-U-Net was compared with multiple convolutional and efficient segmentation architectures under the same data partitioning, preprocessing, optimization, and evaluation protocol. The comparison included the conventional U-Net, Attention U-Net, Lite U-Net, FCN-Simple, U-Net++, and DeepLabV3+, while SegFormer-B0 was additionally included as an efficient transformer-based comparator. Each architecture was independently trained using three random seeds, and the reported results represent the mean $\pm$ standard deviation across the three runs.

The expanded architectural comparison provides a more balanced assessment of the proposed model. U-Net++ and DeepLabV3+ achieved the highest overlap accuracy, with mean DSC values of 0.8978 and 0.8965, respectively, whereas LAG-U-Net achieved a mean DSC of 0.8717 and an IoU of 0.7837. Thus, the revised results do not indicate that LAG-U-Net universally outperforms larger segmentation architectures in terms of segmentation accuracy. Instead, its principal advantage is computational compactness while retaining competitive segmentation performance. LAG-U-Net contains approximately 1.95 million trainable parameters, compared with 7.79 million for U-Net, 12.32 million for DeepLabV3+, and 15.96 million for U-Net++. The comparison with Lite U-Net is particularly informative because both models contain approximately the same number of parameters; the addition of attention increased the mean DSC from

0.8687 to 0.8717. These findings therefore support interpreting LAG-U-Net as an accuracy-efficiency trade-off rather than as an architecture that achieves the highest accuracy across all evaluated models. The complete quantitative comparison of the evaluated segmentation architectures is summarized in [Table 5](#).

Table 5: Quantitative segmentation performance on the CAMUS test set.

Model	DSC ↑	IoU ↑	Precision ↑	Recall ↑	Specificity ↑	HD95 ↓	ASSD ↓
U-Net++	0.8978 ± 0.0013	0.8204 ± 0.0023	0.8840 ± 0.0022	0.9208 ± 0.0015	0.9885 ± 0.0004	7.1503 ± 0.2535	2.8049 ± 0.3053
DeepLabV3+	0.8965 ± 0.0009	0.8176 ± 0.0012	0.8795 ± 0.0029	0.9233 ± 0.0016	0.9881 ± 0.0003	6.8527 ± 0.1010	2.6659 ± 0.0205
Attention U-Net	0.8806 ± 0.0011	0.7966 ± 0.0012	0.8601 ± 0.0004	0.9136 ± 0.0024	0.9865 ± 0.0002	12.1746 ± 0.9170	3.6064 ± 0.1125
U-Net	0.8769 ± 0.0024	0.7914 ± 0.0034	0.8568 ± 0.0025	0.9097 ± 0.0018	0.9863 ± 0.0002	11.7967 ± 0.8050	3.6008 ± 0.1497
LAG-U-Net	0.8717 ± 0.0013	0.7837 ± 0.0017	0.8483 ± 0.0024	0.9093 ± 0.0010	0.9854 ± 0.0003	13.9835 ± 0.9142	4.0452 ± 0.1521
Lite U-Net	0.8687 ± 0.0025	0.7792 ± 0.0036	0.8459 ± 0.0034	0.9064 ± 0.0023	0.9854 ± 0.0002	13.4649 ± 0.9381	3.9815 ± 0.2009
SegFormer-B0	0.8630 ± 0.0037	0.7667 ± 0.0052	0.8381 ± 0.0045	0.9045 ± 0.0021	0.9844 ± 0.0003	9.7103 ± 0.3468	3.5805 ± 0.1066
FCN-Simple	0.7652 ± 0.0188	0.6376 ± 0.0250	0.7280 ± 0.0313	0.8404 ± 0.0010	0.9732 ± 0.0052	25.7448 ± 4.6118	7.6188 ± 1.0382

[Table 5](#) summarizes performance using complementary overlap, classification, and boundary-based metrics. U-Net++ and DeepLabV3+ achieved the strongest overall segmentation accuracy, whereas LAG-U-Net produced a mean DSC of 0.8717 ± 0.0013 and IoU of 0.7837 ± 0.0017 . Its precision, recall, and specificity were 0.8483 ± 0.0024 , 0.9093 ± 0.0010 , and 0.9854 ± 0.0003 , respectively. For boundary-based evaluation, LAG-U-Net obtained an HD95 of 13.9835 ± 0.9142 pixels and an ASSD of 4.0452 ± 0.1521 pixels. These findings show that the lightweight architecture does not provide the highest segmentation accuracy among the evaluated models. Its value should therefore be interpreted together with the computational-efficiency results rather than as evidence of superior segmentation performance alone.

To examine whether the relative performance of U-Net and LAG-U-Net differed across anatomical structures, class-wise DSC was additionally evaluated. [Table 6](#) presents the mean DSC values across the three independent training runs together with the paired statistical comparisons. LAG-U-Net achieved slightly lower mean DSC than U-Net for the background, left-ventricular cavity, myocardium, and left atrium, resulting in negative Δ DSC values for all four classes. The largest difference was observed for the myocardium, whereas the differences for the background and left-ventricular cavity were smaller. These findings indicate that the computational reduction achieved by LAG-U-Net is accompanied by a modest reduction in class-wise segmentation accuracy relative to the conventional U-Net rather than a uniform improvement in segmentation performance.

Table 6: Class-wise DSC comparison between U-Net and LAG-U-Net on the CAMUS test set.

Anatomical Class	U-Net DSC (Mean ± SD)	LAG-U-Net DSC (Mean ± SD)	Δ DSC (LAG-U-Net – U-Net)	Raw <i>p</i> -Value	Holm-Adjusted <i>p</i> -Value	Significant
Background	0.976801 ± 0.000309	0.975151 ± 0.000704	−0.001650	4.47×10^{-13}	1.79×10^{-12}	Yes
LV cavity	0.923697 ± 0.001441	0.920072 ± 0.000756	−0.003625	1.68×10^{-4}	3.36×10^{-4}	Yes
Myocardium	0.831723 ± 0.002692	0.825620 ± 0.003295	−0.006104	1.68×10^{-8}	5.04×10^{-8}	Yes
Left atrium	0.875151 ± 0.003452	0.869268 ± 0.001630	−0.005883	2.07×10^{-3}	2.07×10^{-3}	Yes

Note: DSC values are reported as mean ± standard deviation across three independent training runs. Class-wise comparisons between U-Net and LAG-U-Net were performed using paired Wilcoxon signed-rank tests because the paired DSC differences did not satisfy the normality criterion. Holm correction was applied to account for multiple class-wise comparisons. Statistical significance was defined as a Holm-adjusted *p*-value <0.05. Negative Δ DSC values indicate lower DSC for LAG-U-Net relative to U-Net.

Class-wise statistical comparisons were performed using paired Wilcoxon signed-rank tests because the distributions of paired DSC differences did not satisfy the normality criterion. Holm correction was applied to account for multiple comparisons, with statistical significance defined as an adjusted p -value < 0.05 . The class-wise differences reported in Table 6 were statistically significant after Holm correction; however, statistical significance should not be interpreted as evidence of a practically large performance difference. The absolute DSC differences between the two models remained modest. In all four classes, the negative Δ DSC values indicate that U-Net achieved slightly higher segmentation accuracy than LAG-U-Net. Accordingly, these results support interpretation of LAG-U-Net primarily in terms of computational efficiency and model compactness rather than superior class-wise segmentation accuracy.

3.1 Cross-Dataset Evaluation and Domain-Shift Analysis

To assess cross-dataset transferability under a strict external-evaluation setting, the LAG-U-Net models trained exclusively on CAMUS were evaluated on EchoNet-Dynamic without fine-tuning or target-domain adaptation. Because EchoNet-Dynamic provides left-ventricular endocardial tracings rather than the four-class masks available in CAMUS, the comparison was restricted to the left-ventricular cavity in both datasets. The results therefore represent an anatomically matched LV-only comparison rather than a four-class cross-dataset evaluation.

As shown in Table 7, LAG-U-Net achieved an LV DSC of 0.9201 ± 0.0008 on the CAMUS test set but only 0.1578 ± 0.0087 when transferred directly to EchoNet-Dynamic. Similar reductions were observed for IoU, precision, recall, HD95, and ASSD. The substantial deterioration in overlap and boundary metrics indicates a pronounced domain shift between the two datasets and does not support a claim of robust zero-shot generalization. Although specificity remained high on EchoNet-Dynamic, this metric is influenced by the large proportion of background pixels and should therefore be interpreted together with the low DSC, IoU, and recall rather than in isolation.

Table 7: Cross-dataset LV segmentation performance of LAG-U-Net.

Dataset	DSC $\uparrow$	IoU $\uparrow$	Precision $\uparrow$	Recall $\uparrow$	Specificity $\uparrow$	HD95 $\downarrow$	ASSD $\downarrow$
CAMUS LV	0.9201 ± 0.0008	0.8559 ± 0.0010	0.9182 ± 0.0011	0.9274 ± 0.0006	0.9922 ± 0.0001	10.1419 ± 0.8873	3.1107 ± 0.1861
EchoNet-Dynamic LV	0.1578 ± 0.0087	0.0957 ± 0.0047	0.3772 ± 0.0150	0.1272 ± 0.0122	0.9812 ± 0.0042	75.9466 ± 9.4228	36.2345 ± 8.6510

Note: Values are mean $\pm$ standard deviation across three independently trained LAG-U-Net models. Models were trained only on CAMUS and evaluated on EchoNet-Dynamic without fine-tuning. For anatomical comparability, both datasets are evaluated for the LV cavity only. HD95 and ASSD are reported in pixels. $\uparrow$ indicates higher is better; $\downarrow$ indicates lower is better.

To investigate the source of this performance reduction, image-domain characteristics were additionally compared between CAMUS and EchoNet-Dynamic. Differences were observed in normalized intensity distributions, image entropy, local edge/texture variation, and anatomical foreground proportion. These differences are visualized in Fig. 7 and indicate substantial variation in acquisition and image characteristics between the source and target datasets. Representative low-performing EchoNet-Dynamic cases are further presented in Fig. 8, where prediction errors illustrate the practical effect of this domain mismatch on LV boundary delineation.

The failure-case analysis shows that the reduction in cross-dataset performance is not limited to small boundary deviations. In several difficult EchoNet-Dynamic examples, the model incompletely identifies the LV cavity or produces substantial boundary disagreement. These observations are consistent with the marked deterioration in DSC, IoU, HD95, and ASSD and further indicate that domain-specific adaptation or training

on more heterogeneous multi-center data would be required before reliable deployment across different echocardiographic sources.

Computational efficiency was evaluated using parameter count, GFLOPs, GPU and CPU inference latency, GPU throughput, peak GPU memory allocation, and checkpoint size. All latency measurements correspond to forward-pass inference for a single grayscale input of size $1 \times 1 \times 256 \times 256$. For GPU profiling, 20 warm-up forward passes were first performed to stabilize execution. Inference latency was then measured over 100 repetitions using CUDA events with explicit device synchronization, and the reported value represents the mean latency in milliseconds per image. GPU throughput was calculated as 1000 divided by the mean GPU latency. For CPU profiling, a separate CPU model instance was used, followed by five warm-up passes and 30 forward-pass measurements using a high-resolution performance timer. Peak GPU memory was obtained from the maximum CUDA memory allocation during profiling. GFLOPs were estimated from the number of multiply-accumulate operations for a single forward pass. All GPU measurements were obtained using an NVIDIA L4 GPU. The resulting efficiency measurements are summarized in Table 8.

Table 8: Computational-efficiency profiling of the evaluated segmentation models. GPU and CPU latency represent mean forward-pass inference time for a single $1 \times 1 \times 256 \times 256$ input after device-specific warm-up. GPU latency was averaged across 100 CUDA-event measurements, CPU latency across 30 timed measurements, and FPS was calculated as 1000 divided by mean GPU latency in milliseconds.

Model	Parameters (M)	GFLOPs	GPU Latency (ms/image)	GPU FPS	CPU Latency (ms/image)	Peak GPU Memory (MB)	Checkpoint Size (MB)
U-Net	7.786	83.299	4.924	203.09	171.510	240.508	29.776
Attention U-Net	7.786	84.976	5.620	177.95	189.835	268.508	29.779
LiteUNet	1.949	20.913	2.276	439.37	52.802	117.531	7.497
FCNSimple	0.361	5.534	0.973	1027.72	18.109	60.229	1.397
U-Net++	15.965	31.906	6.349	157.51	105.651	226.469	61.058
DeepLabV3+	12.324	8.997	4.262	234.64	34.875	161.149	47.137
SegFormer-B0	3.712	3.356	16.771	59.63	37.115	114.762	14.244
LAG-U-Net	1.949	21.348	3.382	295.65	81.980	131.531	7.496

3.2 Computational Feasibility Analysis

To assess the computational feasibility of the proposed framework, we profiled model complexity, inference latency, throughput, memory consumption, and model-storage requirements using 256×256 single-frame inputs. LAG-U-Net contained 1.95 million trainable parameters and required approximately 21.35 GFLOPs per forward pass. On the NVIDIA L4 GPU used for the experiments, its mean inference latency was 3.38 ms per image, corresponding to approximately 295.65 frames per second. CPU inference required approximately 81.98 ms per image. The measured peak GPU memory allocation was approximately 131.53 MB, and the model checkpoint occupied approximately 7.50 MB. For comparison, the conventional U-Net required 7.79 million parameters, 83.30 GFLOPs, 4.92 ms GPU inference time, 171.51 ms CPU inference time, approximately 240.51 MB of peak GPU memory, and a 29.78 MB checkpoint. These results indicate that LAG-U-Net substantially reduces model size and computational requirements relative to the conventional U-Net. However, LAG-U-Net was not the fastest architecture among all evaluated models;

therefore, these findings support a compact accuracy–efficiency trade-off rather than a claim of universally superior computational performance. The measurements provide controlled computational-feasibility evidence, but they should not be interpreted as validation of deployment on dedicated mobile ultrasound or edge hardware.

The computational measurements provide evidence of model compactness and inference efficiency under the evaluated hardware environment, but they should not be interpreted as direct validation of deployment on portable ultrasound systems, smartphones, embedded processors, or other edge platforms. In particular, actual edge-device performance depends on processor architecture, available memory, numerical precision, inference framework, operator support, power constraints, thermal behavior, and hardware-specific optimization. No hardware-specific quantization, pruning, TensorRT/ONNX acceleration, mobile inference framework, or embedded-device benchmark was performed in the present study. Consequently, the reported parameter count, GFLOPs, CPU/GPU latency, memory consumption, and checkpoint size characterize computational feasibility and provide a basis for future deployment studies, but do not establish deployment readiness. Future work should benchmark LAG-U-Net directly on representative low-resource platforms and portable ultrasound hardware, including measurements of end-to-end latency, throughput, memory footprint, power consumption, and segmentation accuracy after hardware-specific optimization.

The training and validation curves obtained under the final experimental configuration demonstrate the optimization behavior of the evaluated segmentation models over the predefined maximum training duration of 50 epochs. The corresponding training-loss and validation-DSC trajectories are shown in Fig. 9.

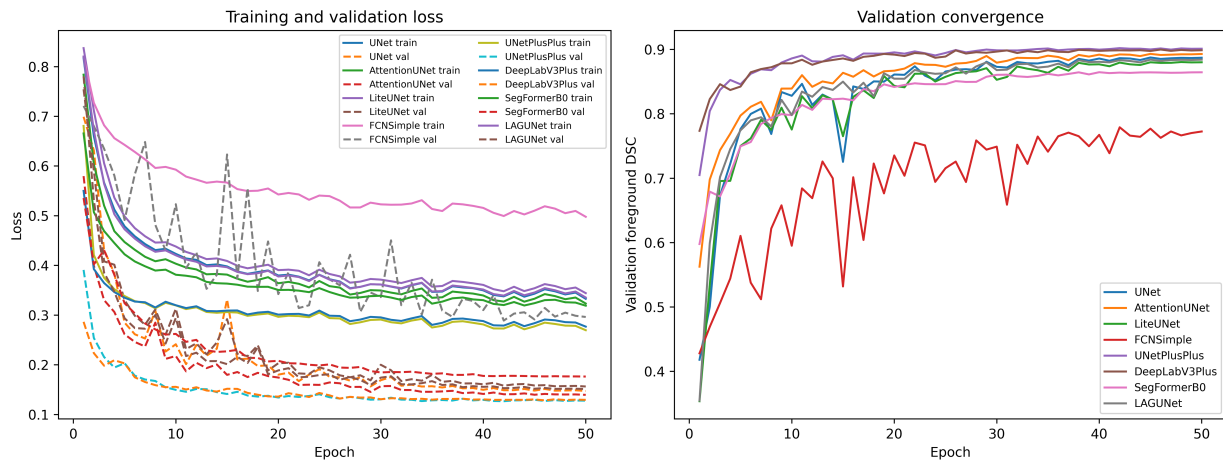


Figure 9: Training-loss and validation-DSC convergence of the evaluated segmentation models under the final experimental configuration. Training was performed for a maximum of 50 epochs with validation-based early stopping and best-checkpoint selection.

To complement the quantitative evaluation, representative CAMUS test cases were visually examined to assess spatial agreement between the predicted and ground-truth segmentations. Fig. 10 compares the input images, ground-truth masks, U-Net predictions, LAG-U-Net predictions, and corresponding LAG-U-Net error maps.

To examine sample-level variation in the relative performance of U-Net and LAG-U-Net, the per-sample DSC values of the two models are compared in Fig. 11.

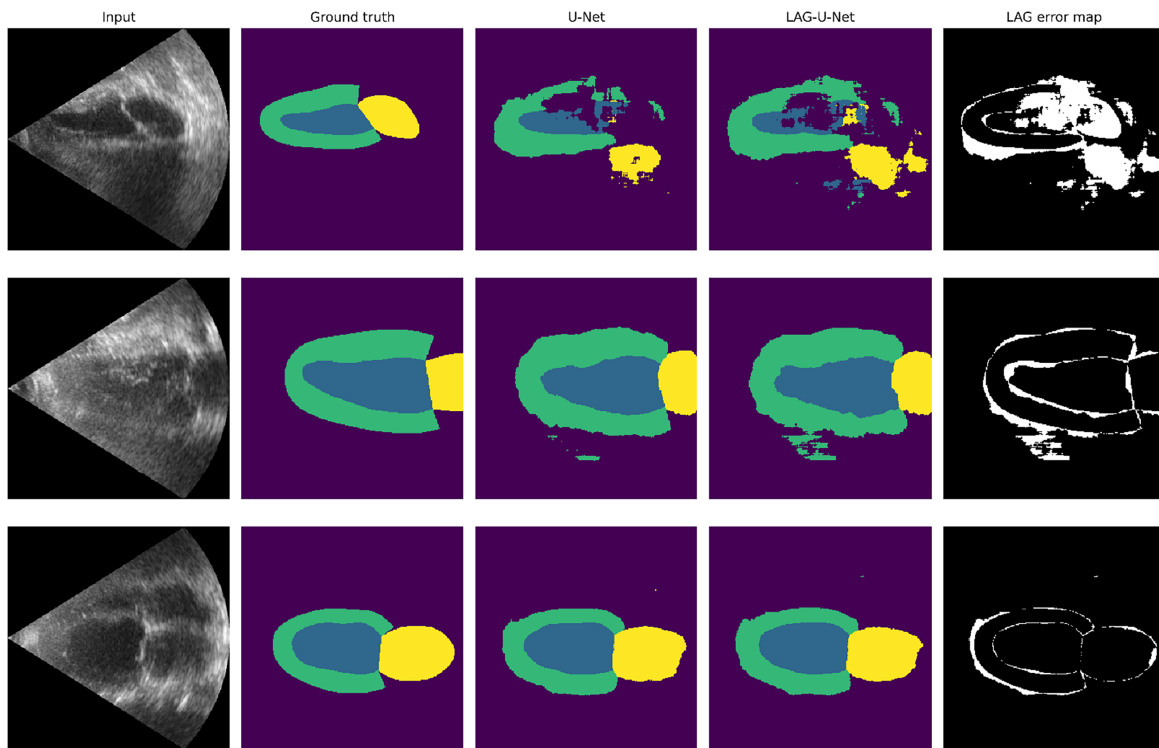


Figure 10: Representative qualitative segmentation results on the CAMUS test set. Each row shows the echocardiographic input image, four-class ground-truth mask, U-Net prediction, LAG-U-Net prediction, and corresponding LAG-U-Net error map. The examples illustrate both successful anatomical delineation and remaining segmentation discrepancies in challenging image regions.

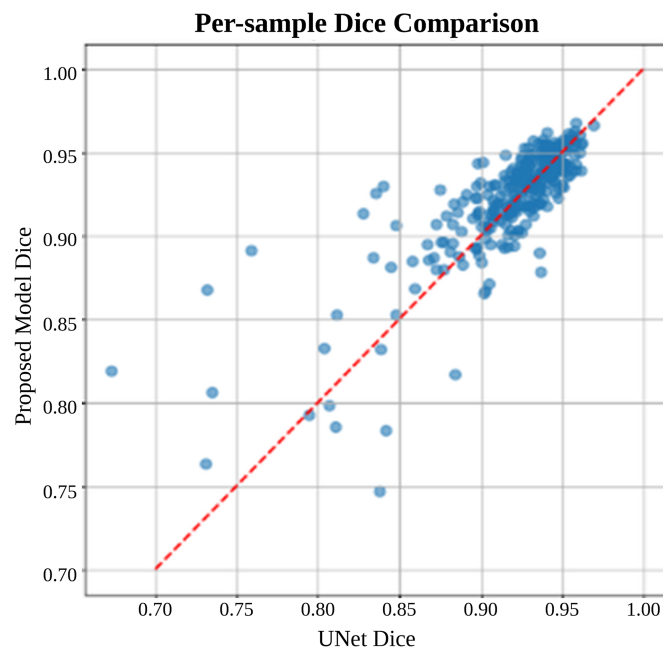


Figure 11: Per-sample DSC comparison between U-Net and LAG-U-Net on the CAMUS test set. Each point represents one test sample. Points above the identity line indicate higher DSC for LAG-U-Net, whereas points below the line indicate higher DSC for U-Net.

3.3 Ablation Study

A controlled ablation study was performed to examine the separate effects of model width and attention-gated skip connections within the proposed framework. Four configurations were evaluated under the same CAMUS data partitions, preprocessing pipeline, augmentation strategy, loss function, optimizer-selection procedure, training conditions, and three independent random seeds. The standard U-Net used the larger base channel width without attention gates. LiteUNet reduced the base channel width from 64 to 32 while retaining conventional skip connections. LiteOneAttention used the same lightweight architecture but introduced an attention gate at one decoder skip level. Finally, LAG-U-Net retained the lightweight channel configuration and applied attention gates at all three skip levels.

Table 9 summarizes the mean and standard deviation of the ablation results across the three independent runs. Reducing the network width substantially decreased the parameter count from approximately 7.79 million parameters for U-Net to approximately 1.95 million parameters for all lightweight variants. This reduction was accompanied by a modest decrease in segmentation accuracy, with LiteUNet obtaining a mean DSC of 0.8687 compared with 0.8769 for the standard U-Net. Introducing a single attention level partially recovered this reduction, increasing the mean DSC to 0.8728 and the mean IoU to 0.7854. The complete LAG-U-Net, with attention applied at all three skip levels, achieved a mean DSC of 0.8717 and a mean IoU of 0.7837.

Table 9: Controlled ablation analysis of model width and attention-gated skip connections on the CAMUS test set. Results are reported as mean $\pm$ standard deviation across three independent runs. Lower HD95 and ASSD values indicate better boundary agreement.

Model Variant	DSC, Mean $\pm$ SD	IoU, Mean $\pm$ SD	HD95, Mean $\pm$ SD	ASSD, Mean $\pm$ SD	Parameters (M)
U-Net	0.8769 $\pm$ 0.0024	0.7914 $\pm$ 0.0034	11.7967 $\pm$ 0.8050	3.6008 $\pm$ 0.1497	7.786
LiteUNet	0.8687 $\pm$ 0.0025	0.7792 $\pm$ 0.0036	13.4649 $\pm$ 0.9381	3.9815 $\pm$ 0.2009	1.949
LiteOneAttention	0.8728 $\pm$ 0.0022	0.7854 $\pm$ 0.0033	13.4585 $\pm$ 1.2986	3.9062 $\pm$ 0.2414	1.949
LAG-U-Net	0.8717 $\pm$ 0.0013	0.7837 $\pm$ 0.0017	13.9835 $\pm$ 0.9142	4.0452 $\pm$ 0.1521	1.949

The boundary-based metrics provide a similarly nuanced result. U-Net achieved the lowest mean HD95 and ASSD among the four ablation configurations, whereas the lightweight variants showed higher boundary errors. LiteOneAttention produced slightly better overlap and ASSD than LiteUNet, suggesting that selective attention can partially compensate for the representational reduction caused by channel compression. However, extending attention to all three skip levels did not provide an additional improvement over the single-attention configuration in this experiment.

These findings indicate that the primary contribution of the lightweight design is a substantial reduction in model size rather than a consistent improvement in segmentation accuracy. Attention-gated skip connections can recover part of the performance lost through channel reduction, but the executed ablation study does not support the stronger claim that increasing the number of attention gates necessarily produces higher segmentation accuracy. Therefore, LAG-U-Net should be interpreted as an efficiency-oriented architecture that provides a compact segmentation solution with competitive, although not uniformly superior, accuracy.

Fig. 12 visualizes the DSC and IoU results of the four ablation configurations reported in Table 9. Compared with LiteUNet, introducing a single attention level produced a modest improvement in DSC and IoU. However, extending attention to all three skip levels did not provide an additional improvement over the

single-attention configuration. These findings indicate a configuration-dependent benefit of attention rather than a uniform increase in segmentation performance.

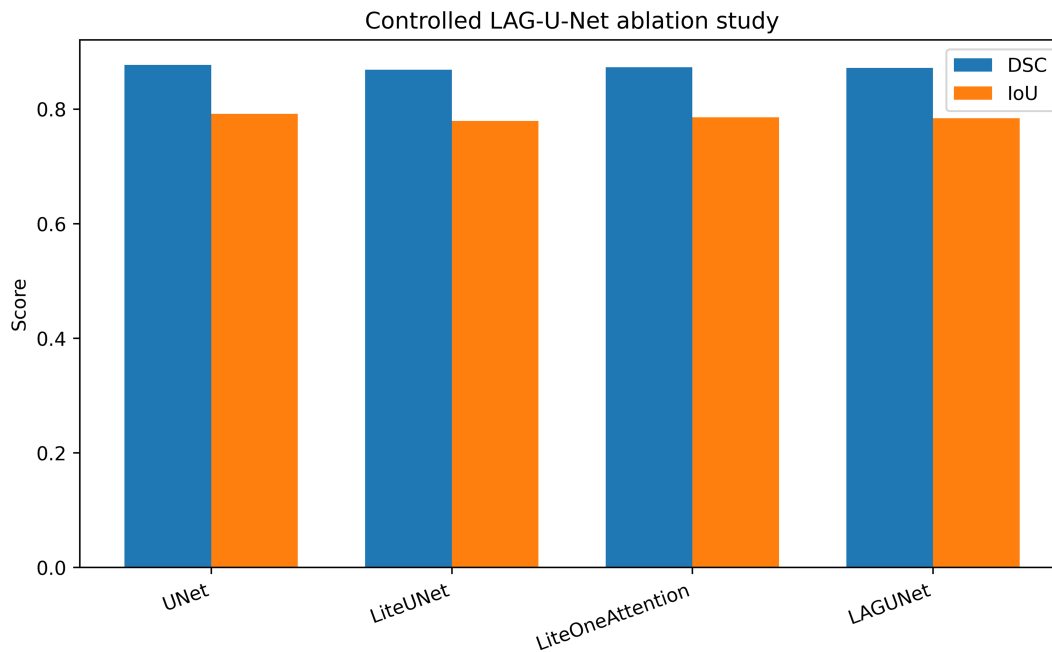


Figure 12: Controlled ablation analysis of the LAG-U-Net framework on the CAMUS dataset. Mean DSC and IoU across the independent experimental runs are shown for the baseline U-Net, lightweight U-Net, lightweight U-Net with a single attention level, and the complete LAG-U-Net architecture, enabling the individual and combined effects of lightweight design and attention-gated feature fusion to be assessed.

Fig. 12 summarizes the controlled ablation analysis under identical evaluation conditions. Reducing the channel dimensions substantially decreases model complexity, while the intermediate attention configuration and complete LAG-U-Net allow the contribution of attention-guided feature fusion to be examined separately from the lightweight architectural changes. The results therefore characterize the accuracy–efficiency trade-off of the proposed design rather than assuming that each architectural modification produces a uniform performance improvement.

The controlled ablation experiments were designed to examine whether the attention mechanism contributes information beyond the reduction in network width. The lightweight U-Net provides the compact encoder–decoder architecture without attention, the intermediate LiteOneAttention configuration introduces attention only at the deepest skip connection, and the complete LAG-U-Net applies attention to all three skip levels. Because all variants use the same lightweight base architecture and are tested with the same patient-level data split and repeated training procedure, the comparison helps separate the effect of attention from the effect of reducing model size. The observed differences are relatively small, suggesting that attention may improve the balance between segmentation accuracy and efficiency, but it does not lead to consistently better performance in every setting.

Qualitative and failure-case analyses were also carried out using representative CAMUS examples, including cases with lower segmentation performance. These analyses are useful in echocardiography because overlap metrics alone cannot show where the errors occur, especially near poorly defined anatomical boundaries or in difficult image regions. Together with the ablation results, the qualitative findings provide

additional insight into the contribution of attention-guided feature selection, while also showing that attention alone cannot guarantee robustness against ultrasound noise, low contrast, or anatomical variation.

3.4 Comparison with Related Work

Recent advances in medical image segmentation research indicate that there is a clear relationship between the complexity of the model and the computational costs incurred. Although there exist some newer models like U-Net++, DeepLabV3+, and Attention U-Net, which have introduced several innovations, the U-Net model is still in demand. Transformer-based models are more efficient when it comes to assessing the global context; some of them are lightweight, like SegFormer [24]. Thus, we have used the U-Net, Attention U-Net, U-Net++, DeepLabV3+, FCN, lightweight U-Net, and SegFormer-B0 models in our experiments with a single data set.

One of the recent models, Attention U-Net [32], exploits the advantages of attention gates applied in the skip connections for highlighting the encoder characteristics that pass on to the decoder. LAG-U-Net employs the above-stated technology and presents some innovations of its own. Therefore, the thing that distinguishes it from the previous model is the application of an attention-enabled skip connection and a much thinner convolutional backbone. The experimental results show that although Attention U-Net achieves slightly better segmentation performance, it has substantially greater model complexity than LAG-U-Net.

When discussing the issue of new lightweight segmentation techniques, it becomes evident that the studies are mostly oriented towards decreasing the complexity of the models without suffering in terms of segmentation performance. For instance, the RepSegNet model relies on the technique of reparameterization [29], while LiteFANet employs lightweight fusion and spatial-channel attention techniques [30]. Nevertheless, even the abovementioned models cannot be fully comparable to CAMUS since they were developed under different conditions and varying settings.

Speaking about echocardiological studies, there is a body of research that contributes to the improvement of segmentation performance and helps in bringing down the annotation costs while discussing the generalization possibilities of deep learning systems. The research by Sanjeevi et al. [36] proves that there are a number of challenges that still have to be taken into account while implementing deep learning for echocardiography purposes. Awasthi et al. [37] introduced LVNet, a lightweight model for left-ventricular segmentation with lower computational requirements. Alajrami et al. [38] examined active-learning methods aimed at reducing the amount of expert annotation needed for left-ventricular segmentation. Maani et al. [39] proposed SimLVSeg, which uses self- and weakly supervised learning for left-ventricular segmentation in 2-D+time echocardiography. Together, these studies show different ways of improving efficiency and data use in cardiac ultrasound segmentation and provide useful context for the accuracy-efficiency focus of LAG-U-Net.

Research in cardiac ultrasound segmentation has also moved beyond conventional CNN-based models. Foundation-model approaches have been adapted for left-ventricular segmentation and evaluated on EchoNet-Dynamic, CAMUS, and other ultrasound datasets [7]. Transformer-based and temporal models have also been used to capture broader spatial and sequential information in echocardiographic images [8,9]. Although these methods can achieve strong segmentation results, their performance must be considered together with differences in datasets, annotation procedures, computational requirements, and external validation.

The results of the present study therefore do not suggest that LAG-U-Net consistently outperforms all recent segmentation methods. Instead, it was evaluated under the same experimental protocol against

CNN-based, attention-based, lightweight, and efficient transformer baselines. Larger architectures such as U-Net++ and DeepLabV3+ achieved higher segmentation accuracy, while LAG-U-Net used a considerably smaller model and still maintained competitive overlap performance. The contribution of LAG-U-Net is therefore better understood as a compact attention-guided architecture that aims to balance segmentation performance with computational efficiency in echocardiographic segmentation.

3.5 Discussion

Through new experiments, we can evaluate LAG-U-Net thoroughly compared to previous versions of assessment based on either one measure of overlap or a single training trial. The model was evaluated on three different occasions using conventional overlap, classification, and boundary metrics. In addition to this analysis, we have employed various paired statistical techniques, profiling, and both qualitative analysis and cross-dataset assessment.

LAG-U-Net does not aim to achieve better segmentation performance due to higher architectural complexity. Instead, it strives to keep relevant anatomical features within a lightweight encoder-decoder architecture. Although decreasing the width of channels allows for reducing the number of learnable parameters and computational requirements, it also narrows down the representation ability of the model. Therefore, attention gates are introduced in the skip connections and are used to choose which encoder features will be fed into the decoder. Since, in echocardiographic imaging, cardiac boundaries might be distorted by various artifacts, including speckle noise and weak contrast, this mechanism is capable of limiting the impact of all irrelevant parts of the spatial information. Nonetheless, the attention technique shouldn't be treated as a technique that guarantees the model's robustness against ultrasound artifacts.

Ablation studies provide a better understanding of the nature of the impacts brought about by specific design decisions. By distinguishing between channel compression and attention-gating factors, the analysis demonstrates what advantages are gained from the lightweight architecture and what advantages come from using the attention mechanism. The outcomes reveal that compressing the model and employing attention is a matter of trade-off, since there is a trade-off between segmentation accuracy and computational efficiency rather than a systematic improvement in terms of all criteria. For this reason, LAG-U-Net is best interpreted as an architecture that aims to balance segmentation performance and computational demand, rather than one that is universally superior to larger architectures.

The revised evaluation also considers variability that is not adequately represented by mean Dice or IoU alone. Boundary-based measures, including HD95 and ASSD, complement region-overlap metrics by identifying cases in which apparently acceptable overlap may coexist with localized contour errors. Class-wise and per-sample analyses further demonstrate that segmentation difficulty is not uniform across anatomical structures or echocardiographic images. The qualitative and failure-case analyses show that challenging examples may arise from weak anatomical boundaries, reduced contrast, image artifacts, atypical structure appearance, and differences in foreground size. These observations are clinically relevant because average dataset-level performance can obscure individual cases in which segmentation quality deteriorates.

Cross-dataset evaluation provides an additional stress test of model transferability. The CAMUS experiments evaluate multi-class segmentation of the left-ventricular cavity, myocardium, and left atrium, whereas the EchoNet-Dynamic evaluation is restricted to the left-ventricular class because equivalent multi-class annotations are not available in EchoNet-Dynamic. Consequently, the EchoNet experiment should not be interpreted as a direct replication of the full CAMUS task. Instead, it evaluates zero-shot transfer of the shared left-ventricular segmentation target under different acquisition and image-distribution characteristics. The accompanying domain-shift analysis demonstrates measurable differences in image characteristics between the datasets and helps explain why cross-dataset performance can differ from within-dataset evaluation.

These findings therefore provide evidence about transfer under domain shift but do not establish universal clinical generalizability.

Computational efficiency was examined using complementary measures including parameter count, computational complexity, inference latency, throughput, memory consumption, and model size. These measurements are more informative for the lightweight-design objective than parameter count alone. However, measurements obtained on the experimental hardware should not be interpreted as direct evidence of real-time performance on mobile ultrasound equipment or other edge devices. Device-specific benchmarking and optimization remain necessary before clinical deployment can be established.

Overall, the revised findings support a cautious interpretation of LAG-U-Net as a lightweight attention-guided segmentation architecture with a potentially useful accuracy-efficiency trade-off. The expanded evaluation also identifies conditions under which this advantage becomes smaller and highlights the effects of anatomical complexity and dataset shift. Therefore, the present results support further investigation of the architecture but do not establish superiority across all models, datasets, ultrasound devices, or clinical environments.

3.5.1 Limitations

Several limitations should be considered when interpreting the present findings. First, although CAMUS provides a well-established benchmark for cardiac ultrasound segmentation, evaluation on benchmark datasets cannot reproduce the full variability encountered in routine clinical practice. The present study did not include prospective multi-center data, explicit stratification across different ultrasound devices, acquisition protocols, institutions, or patient populations. Consequently, the reported results should not be interpreted as evidence of clinical generalizability across all echocardiographic settings.

Second, the segmentation tasks available in CAMUS and EchoNet-Dynamic are not directly equivalent. LAG-U-Net was trained on CAMUS as a four-class segmentation model comprising background, left-ventricular cavity, myocardium, and left atrium, whereas EchoNet-Dynamic provides left-ventricular endocardial tracings. Accordingly, the zero-shot EchoNet-Dynamic evaluation was restricted to the LV class and cannot establish external generalization for the myocardium or left atrium. In addition, the model was transferred from CAMUS to EchoNet-Dynamic without fine-tuning. The observed differences in image intensity, contrast, texture, image sharpness, and foreground characteristics between the datasets therefore represent a genuine domain shift. Although this study explicitly quantified several of these dataset-level differences and examined external failure cases, evaluation on two public datasets remains insufficient to establish robustness across the broader diversity of clinical echocardiography.

Third, the present experiments are based on individual annotated echocardiographic frames rather than temporal modeling of complete cardiac cycles. Therefore, LAG-U-Net does not exploit motion information or temporal consistency between consecutive frames. Incorporating temporal information from full echocardiographic sequences may improve consistency during dynamic cardiac imaging and represents an important direction for future investigation.

Fourth, the computational analysis provides hardware-specific feasibility evidence rather than direct clinical deployment validation. Parameters, GFLOPs, GPU and CPU inference latency, throughput, peak GPU memory, and checkpoint size were measured under controlled computational conditions; however, the model was not deployed on a mobile ultrasound platform, embedded processor, or dedicated edge device. Therefore, claims regarding real-time or resource-constrained clinical deployment should be interpreted as potential feasibility rather than demonstrated clinical implementation.

Finally, the ablation and comparative experiments indicate that the lightweight design provides a substantial reduction in model complexity, but this reduction does not yield uniformly superior segmentation accuracy relative to all comparison architectures. Likewise, increasing the number of attention-gated skip connections did not consistently improve all segmentation metrics. The contribution of LAG-U-Net should therefore be interpreted primarily in terms of its compact accuracy-efficiency trade-off rather than as universal performance superiority. Further validation using larger multi-center cohorts, device-diverse acquisitions, additional external datasets, temporal echocardiographic sequences, and deployment-specific hardware is required before broader clinical conclusions can be made. The computational-feasibility measurements were obtained under a controlled software and hardware environment and should not be interpreted as direct evidence of clinical or edge-device deployment readiness. The model was not evaluated on portable ultrasound equipment, mobile processors, embedded accelerators, or other low-resource clinical hardware. Consequently, device-specific latency, energy consumption, thermal constraints, memory sharing, image-acquisition overhead, and integration with real-time clinical workflows remain to be evaluated prospectively. A further limitation is that the present study does not constitute a prospective multi-center or device-stratified clinical validation. Although CAMUS and EchoNet-Dynamic provide data acquired under different dataset-specific conditions and the latter was used for zero-shot external evaluation, the available experiments were not designed to separately quantify performance according to hospital site, ultrasound manufacturer, scanner model, acquisition protocol, operator, demographic subgroup, or clinical condition. Consequently, the observed cross-dataset performance should be interpreted as a domain-shift stress test rather than evidence of device-independent or population-independent clinical robustness. Future validation should therefore include prospectively collected multi-center cohorts containing heterogeneous ultrasound systems and patient populations, with performance reported separately across acquisition sites, scanner vendors, demographic and clinical subgroups, and image-quality strata.

3.5.2 Future Work

Future work will focus on several directions. First, temporal information from echocardiographic video could be incorporated through recurrent or transformer-based architectures to investigate whether sequential information improves segmentation consistency. Second, validation should be extended to larger multi-center datasets representing different patient populations, acquisition protocols, and ultrasound systems. Third, deployment on dedicated edge or mobile hardware should be evaluated directly, including latency, memory consumption, and hardware-specific optimization. In the process, this research needs to assess the generalizability of the proposed framework on larger and more diverse multi-center datasets in order to determine whether the proposed framework can generalize to different populations and different imaging methods. The third area of investigation is the development of hybrid architectures using lightweight convolutional backbones, as well as efficient transformer-based self-attention methods. Subsequently, such hybrid architectures could result in significant increases in performance without significantly increasing computational cost for training and testing. This research will also explore ways to develop the proposed framework for deployment on edge devices and/or real-time clinical systems using model compression and model acceleration methods that take into account the specific hardware used. Overall, all of these areas of research will result in a reduction in the gap between research and clinical application, thus facilitating the automation of segmentation processes as part of each echocardiographic workflow.

4 Conclusion

This study presented LAG-U-Net, a lightweight attention-guided encoder-decoder framework for multi-class segmentation of cardiac structures in echocardiographic images. The proposed architecture was

designed around two complementary objectives: reducing computational complexity through a compact convolutional backbone and improving feature selection through attention-guided skip connections. In the implemented model, encoder features are progressively extracted using lightweight convolutional blocks, while attention gates modulate the skip-connected feature maps before their fusion with decoder features. The attention coefficients are jointly determined from encoder and decoder information, allowing spatially relevant features to receive greater emphasis while less informative responses are suppressed before reconstruction of the segmentation map. The final experimental evaluation was designed to assess this architecture more rigorously than in the original submission. LAG-U-Net was evaluated using a fixed patient-independent CAMUS partition and three independent training runs, together with multiple overlap and boundary metrics, stronger CNN- and transformer-based baselines, paired statistical testing, computational-efficiency analysis, controlled ablation experiments, qualitative assessment, failure-case analysis, and zero-shot evaluation on EchoNet-Dynamic. The resulting evidence indicates that the principal value of the proposed design lies in its accuracy–efficiency trade-off rather than in a uniform performance advantage over every comparator or anatomical structure. The controlled ablation analysis further separates the effects of channel reduction and attention-guided feature fusion, providing a more direct assessment of the contribution of the proposed architectural components. The CAMUS experiments address multi-class segmentation of the left ventricular cavity, myocardium, and left atrium, whereas the EchoNet-Dynamic evaluation provides a more challenging external assessment under a different data distribution and annotation setting. The observed cross-dataset degradation emphasizes that differences in acquisition characteristics, image quality, annotation protocols, and patient populations remain important limitations to generalization. Accordingly, the present results should not be interpreted as evidence of deployment-ready clinical robustness. Future work should therefore evaluate the framework on prospective multi-center data and heterogeneous ultrasound systems, incorporate temporal information from echocardiographic video, and perform hardware-specific validation on portable and resource-constrained devices. Overall, the revised evidence supports LAG-U-Net as a compact attention-guided segmentation framework whose main contribution is the integration of lightweight feature extraction and gated skip-feature fusion within a systematically evaluated echocardiographic segmentation pipeline.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: conceptualization, Nasim Ahmadzadeh Nobari Azar and Abdullahi Umar Ibrahim; methodology, Nasim Ahmadzadeh Nobari Azar and Abdullahi Umar Ibrahim; software, Abdullahi Umar Ibrahim; validation, Ammara Talib and Mohamad Ziad Altabel; investigation, Ammara Talib; data curation, Nasim Ahmadzadeh Nobari Azar; writing—original draft preparation, Nasim Ahmadzadeh Nobari Azar and Abdullahi Umar Ibrahim; writing—review and editing, Abdullahi Umar Ibrahim and Mohamad Ziad Altabel; supervision, Kamil Dimililer; project administration, Kamil Dimililer. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The datasets used in this study are publicly available. The CAMUS dataset is available through the official CAMUS challenge website at <https://camus.creatis.insa-lyon.fr/challenge/>. The EchoNet-Dynamic dataset is available through the Stanford University EchoNet-Dynamic data portal at <https://echonet.github.io/dynamic/>. No new clinical datasets were collected for this study.

Ethics Approval: This study involved secondary analysis of publicly available, de-identified echocardiographic datasets and did not involve the recruitment of new participants or the collection of new identifiable clinical data. The CAMUS dataset was collected at the University Hospital of Saint-Etienne, France, and was anonymized in accordance with the requirements of the local ethics committee [12]. The EchoNet-Dynamic dataset was derived from echocardiographic

examinations performed at Stanford Health Care under Stanford University Institutional Review Board oversight (Protocol 43721), with a waiver of individual informed consent for the retrospective use of de-identified data [13]. Near East University is the institutional affiliation of the authors and was not involved in the original collection of these datasets. The present study used only the publicly released, de-identified datasets.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Definition
CAMUS	Cardiac Acquisitions for Multi-structure Ultrasound Segmentation
DSC	Dice Similarity Coefficient
IoU	Intersection over Union
LV	Left Ventricle
LA	Left Atrium
CNN	Convolutional Neural Network
DL	Deep Learning
HD95	95th-percentile Hausdorff Distance
ASSD	Average Symmetric Surface Distance
FPS	Frames Per Second

References

1. Barison A, Timoteo AT, El Messaoudi S, Borodzicz-Jazdyk S, Moscatelli S, Mandoli GE, et al. Cardiovascular imaging in 2024: review of current research and innovations. *Eur Heart J Imaging Meth Pract*. 2025;3(1):qyaf066. doi:10.1093/ehjimp/qyaf066.
2. Carerj ML, Restelli D, Poleggi C, di Bella G, Zito C, Manganaro R, et al. The role of imaging in cardiovascular prevention: a comprehensive review. *J Cardiovasc Echogr*. 2025;35(1):8–18. doi:10.4103/jcecho.jcecho_26_25.
3. Noelck NJ, Perry HA, Talley PL, Le DE. A narrative review of the clinical applications of echocardiography in right heart failure. *J Clin Med*. 2025;14(15):5505. doi:10.3390/jcm14155505.
4. Liga R, Korosoglou G. Editorial: case reports in cardiovascular imaging 2023. *Front Cardiovasc Med*. 2024;11:1424893. doi:10.3389/fcvm.2024.1424893.
5. Kajrolkar A. Artificial intelligence in cardiac imaging: a comprehensive review of clinical implementation and validation. *Prem J Cardiol*. 2025;2:100003. doi:10.70389/pjc.100003.
6. Figliozzi S, Hutt E, Gimelli A, Jaber WA. Cardiac imaging highlights from European Society of Cardiology 2024: the future is within our grasp! *Eur Heart J Imaging Methods Pract*. 2025;3(1):qyaf009. doi:10.1093/ehjimp/qyaf009.
7. Chao CJ, Gu YR, Kumar W, Xiang T, Appari L, Wu J, et al. Foundation versus domain-specific models for left ventricular segmentation on cardiac ultrasound. *npj Digit Med*. 2025;8(1):341. doi:10.1038/s41746-025-01730-y.
8. Huang KC, Lin CE, Lin DS, Lin TT, Wu CK, Jeng GS, et al. Video transformer for segmentation of echocardiography images in myocardial strain measurement. *J Imaging Inform Med*. 2026;39(3):2661–79. doi:10.1007/s10278-025-01682-5.
9. Lin Z, Cui R, Ning L, Peng J. Temporal features-fused vision retentive network for echocardiography image segmentation. *Sensors*. 2025;25(6):1909. doi:10.3390/s25061909.
10. Majchrowska S, Hildeman A, Mokhtari R, Diethe T, Teare P. Exploring interpretable echo analysis using self-supervised parcels. *Comput Biol Med*. 2025;192(9):110322. doi:10.1016/j.compbimed.2025.110322.
11. Li H, Yang J, Qu M, Feng Y. A semi-supervised multi-task assisted method for ultrasound medical image segmentation. *Neurocomputing*. 2025;639(2):130217. doi:10.1016/j.neucom.2025.130217.

12. Leclerc S, Smistad E, Pedrosa J, Ostvik A, Cervenansky F, Espinosa F, et al. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. *IEEE Trans Med Imaging*. 2019;38(9):2198–210. doi:10.1109/TMI.2019.2900516.
13. Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, et al. Video-based AI for beat-to-beat assessment of cardiac function. *Nature*. 2020;580(7802):252–6. doi:10.1038/s41586-020-2145-8.
14. Cacao GF, Du D, Nair N. Unsupervised image segmentation on 2D echocardiogram. *Algorithms*. 2024;17(11):515. doi:10.3390/a17110515.
15. Sharma P, Kumar A, Jain SK. Ultrasound image segmentation: a systematic review of deformable models from classical techniques to intelligent advancements. *Inf Fusion*. 2026;127(8):103768. doi:10.1016/j.inffus.2025.103768.
16. Manghera PS, Lall J. Artificial intelligence in cardiac imaging: a systematic review of diagnostic, prognostic, and workflow enhancements. *Eur J Cardiovasc Med*. 2025;15(11):581–91. doi:10.61336/ejcm/25-11-85.
17. Rainio O, Roshan E, Hosseini SM, Rehman R, Okenwa J, Klén R. Deep learning for medical ultrasound image segmentation: a systematic review of the current research. *J Imaging Inform Med*. 2026. doi:10.1007/s10278-026-01931-1.
18. Isensee F, Wald T, Ulrich C, Baumgartner M, Roy S, Maier-Hein K, et al. nnU-Net revisited: a call for rigorous validation in 3D medical image segmentation. In: *Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2024*; 2024 Oct 6–10; Marrakesh, Morocco. doi:10.1007/978-3-031-72114-4_47.
19. Ullah Z, Hilal M, Lee E, Pamucar D, Kim J. Unified review and benchmark of deep segmentation architectures for cardiac ultrasound on CAMUS. *arXiv:2601.00839*. 2026.
20. He K, Gan C, Li Z, Rekik I, Yin Z, Ji W, et al. Transformers in medical image analysis. *Intell Med*. 2023;3(1):59–78. doi:10.1016/j.imed.2022.07.002.
21. Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, et al. TransUNet: transformers make strong encoders for medical image segmentation. *arXiv:2102.04306*. 2021.
22. Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, et al. Swin-Unet: Unet-like pure transformer for medical image segmentation. In: *Proceedings of the Computer Vision–ECCV 2022 Workshops*; 2023 Oct 23–27; Tel Aviv, Israel. doi:10.1007/978-3-031-25066-8_9.
23. Hatamizadeh A, Tang Y, Nath V, Yang D, Myronenko A, Landman B, et al. UNETR: transformers for 3D medical image segmentation. In: *Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*; 2022 Jan 3–8; Waikoloa, HI, USA. doi:10.1109/wacv51458.2022.00181.
24. Xie E, Wang W, Yu Z, Anandkumar A, Alvarez JM, Luo P. SegFormer: simple and efficient design for semantic segmentation with transformers. *Adv Neural Inf Process Syst*. 2021;34:12077–90.
25. Gerdroodbary MB, Salavatidezfouli S. A predictive surrogate model based on linear and nonlinear solution manifold reduction in cardiovascular FSI: a comparative study. *Comput Biol Med*. 2025;189(9):109959. doi:10.1016/j.compbimed.2025.109959.
26. Sabo S, Pettersen H, Bøen GC, Jakobsen EO, Langøy PK, Nilsen HO, et al. Real-time guidance and automated measurements using deep learning to improve echocardiographic assessment of left ventricular size and function. *Eur Heart J Imaging Meth Pract*. 2025;3(2):qyaf094. doi:10.1093/ehjimp/qyaf094.
27. Bennett S, Johnson CL, Fisher G, Erskine F, Krasner S, Fletcher AJ, et al. Development and validation of echocardiography artificial intelligence models: a narrative review. *J Clin Med*. 2025;14(19):7066. doi:10.3390/jcm14197066.
28. Liu X, Song L, Liu S, Zhang Y. A review of deep-learning-based medical image segmentation methods. *Sustainability*. 2021;13(3):1224. doi:10.3390/su13031224.
29. Juraev R, Kim IM, Yun S, Kang JM. Efficient medical image segmentation using RepSegNet lightweight reparameterized neural network. *Sci Rep*. 2026;16(1):4682. doi:10.1038/s41598-025-34973-6.
30. Xu K, Guo X, Pan B, Zhang Y, Liu Y, Zhang X, et al. LiteFANet: a lightweight UNet-based fusion-attention segmentation network for 2D and 3D medical images. *Biomed Signal Process Control*. 2026;112:108632. doi:10.1016/j.bspc.2025.108632.
31. Neha F, Bhati D, Shukla DK, Dalvi SM, Mantzou N, Shubbar S. An analytics-driven review of U-Net for medical image segmentation. *Healthc Anal*. 2025;8(1):100416. doi:10.1016/j.health.2025.100416.

32. Oktay O, Schlemper J, Le Folgoc L, Lee M, Heinrich M, Misawa K, et al. Attention U-net: learning where to look for the pancreas. arXiv:1804.03999. 2018.
33. Kumar AS, Rekha R. A dense network approach with Gaussian optimizer for cardiovascular disease prediction. *New Gener Comput.* 2023;41(4):859–78. doi:10.1007/s00354-023-00234-1.
34. Islam T, Hafiz MS, Jim JR, Kabir MM, Mridha MF. A systematic review of deep learning data augmentation in medical imaging: recent advances and future research directions. *Healthc Anal.* 2024;5(1):100340. doi:10.1016/j.health.2024.100340.
35. Tupper A, Gagné C. Revisiting data augmentation for ultrasound images. *Trans Mach Learn Res.* 2025.
36. Sanjeevi G, Gopalakrishnan U, Parthinarupothi RK, Madathil T. Deep learning supported echocardiogram analysis: a comprehensive review. *Artif Intell Med.* 2024;151(25):102866. doi:10.1016/j.artmed.2024.102866.
37. Awasthi N, Vermeer L, Fixsen LS, Lopata RGP, Pluim JPW. LVNet: lightweight model for left ventricle segmentation for short axis views in echocardiographic imaging. *IEEE Trans Ultrason Ferroelect Freq Contr.* 2022;69(6):2115–28. doi:10.1109/tuffc.2022.3169684.
38. Alajrami E, Ng T, Jevsikov J, Naidoo P, Fernandes P, Azarmehr N, et al. Active learning for left ventricle segmentation in echocardiography. *Comput Meth Programs Biomed.* 2024;248(3):108111. doi:10.1016/j.cmpb.2024.108111.
39. Maani F, Ukaye A, Saadi N, Saeed N, Yaqub M. SimLVSeg: simplifying left ventricular segmentation in 2-D+Time echocardiograms with self- and weakly supervised learning. *Ultrasound Med Biol.* 2024;50(12):1945–54. doi:10.1016/j.ultrasmedbio.2024.08.023.