Computer Methods and Programs in Biomedicine 269 (2025) 108869

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: https://www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine

echoGAN: Extending the field of view in Transthoracic Echocardiography through conditional GAN-based outpainting

Matej Gazda [a], Jakub Gazda [b], Samuel Kadoury [c,d], Robert Kanasz [a], Peter Drotar [a,*]

[a] Technical University of Kosice, Slovakia, Letna 9, Kosice, 040 01, Slovakia
[b] Pavol Jozef Safarik University and Louis Pasteur University Hospital, Srobarova 2, Kosice, 041 90, Slovakia
[c] Universite de Montreal, boulevard Édouard-Montpetit, Montreal, H3T 1J4, Canada
[d] Polytechnique Montréal, 2500, chemin de Polytechnique, Montreal, H3T 1J4, Canada

ARTICLE INFO

Keywords:
Field of view
Ultrasound
cGAN
Outpainting
Domain-aware augmentation

ABSTRACT

Background and Objective: Transthoracic Echocardiography (TTE) is a fundamental, non-invasive diagnostic tool in cardiovascular medicine, enabling detailed visualization of cardiac structures that is crucial for diagnosing various heart conditions. Despite its widespread use, TTE ultrasound imaging faces inherent limitations, notably a trade-off between field of view (FoV) and resolution.

Methods: This paper introduces a novel conditional Generative Adversarial Network (cGAN), incorporating a domain-aware augmentation technique that simulates the typical cone-shaped FoV in ultrasound. This approach is specifically designed to enable effective outpainting of occluded areas, setting the foundation for our cGAN architecture, termed echoGAN.

Results: The results, obtained on two different datasets, confirm that echoGAN demonstrates the capability to generate realistic anatomical structures through outpainting, effectively broadening the viewable area in medical imaging.

Conclusions: This advancement has the potential to enhance both automatic and manual ultrasound navigation, offering a more comprehensive view that could significantly reduce the learning curve associated with ultrasound imaging.

1. Introduction

Transthoracic Echocardiography (TTE) Ultrasound (USG) is a cornerstone diagnostic tool in cardiovascular medicine, valued for its non-invasive approach and ability to comprehensively assess cardiac anatomy and function [1,2]. It enables detailed visualization of cardiac structures, such as chambers, valves, and myocardium, and is essential in diagnosing conditions like heart failure, valvular heart disease, and cardiomyopathies. One of TTE USG's key advantages is its capacity for real-time dynamic imaging, allowing clinicians to observe cardiac motion and providing insights beyond what static imaging can offer.

Different ultrasound probes are tailored for specific medical purposes. Linear array probes are effective for shallow structures, while curvilinear array probes are suited for deeper tissue imaging. Endocavitary probes excel in specialized fields such as urology and gynecology, offering high-resolution images of internal structures. Sector array probes, often used in cardiac imaging, balance field of view (FoV) with depth. However, these probes still confront the challenge of balancing FoV, resolution, and penetration depth.

Despite its wide applications, TTE USG has inherent limitations, particularly in the trade-off between FoV and resolution. Achieving high resolution in a focused area typically results in a narrower FoV, providing detailed imagery at the cost of a broader view. Conversely, a wider FoV can reduce the resolution, compromising detail for a more comprehensive perspective. This balance between FoV and resolution is a crucial consideration in clinical ultrasound imaging.

Although systematic, large-scale data are still scarce, the American Society of Echocardiography notes that a narrower sector angle may be appropriate in some circumstances to enhance image quality while larger sectors result in reduced image quality [3]. Moreover, small single-center audits suggest that a non-trivial proportion (20%–30%) of TTE examinations remains suboptimal [4,5].

Clinicians have the capability to manipulate the field of view to enhance image resolution. This is achieved through the focusing of the ultrasound beam, which can be narrowed to increase the lateral resolution at specific depths, known as the focal zone. By adjusting the focus and consequently decreasing the FoV, the ultrasound system can

* Corresponding author.
E-mail address: peter.drotar@tuke.sk (P. Drotar).

https://doi.org/10.1016/j.cmpb.2025.108869
Received 26 February 2025; Received in revised form 28 April 2025; Accepted 19 May 2025
Available online 7 June 2025
0169-2607/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

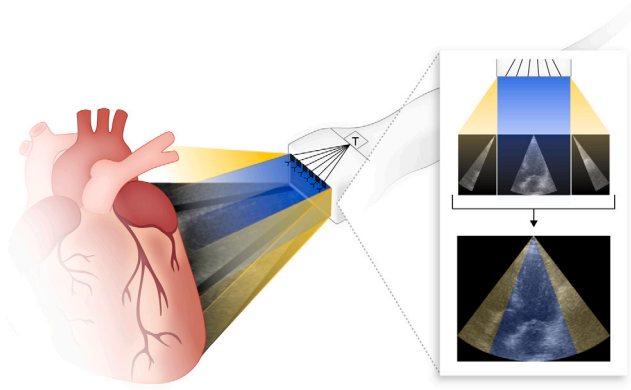

Fig. 1. Extending field of view (yellow) vs. basic (teal).

provide a more detailed view of the area of interest. This technique is particularly useful in cardiac imaging, where high resolution is critical for accurate diagnosis. The adjustment of the FoV and focus is an integral part of ultrasound knobology, allowing for optimization of image quality based on the clinical requirements of the examination [6].

In this context, our research has explored the application of a conditional Generative Adversarial Network (cGAN) to extend the FoV in TTE USG imaging while maintaining high resolution. The proposed cGAN architecture, denoted as echoGAN, is capable of generating realistic anatomical structures through outpainting, thereby broadening the viewable area in medical imaging. The idea of the approach proposed in this paper is depicted in Fig. 1. This approach aims to address the limitations in FoV and resolution trade-offs, potentially enhancing both automatic and manual ultrasound navigation. An expanded FoV is expected to provide a more comprehensive view for automatic navigation [7]; our preliminary results suggest it could improve efficiency and accuracy, though this requires confirmation in a wider range of clinical scenarios. Similar challenges in expanding the FoV have been addressed in related fields, using advanced lens models and flexible array technology to enhance the visual scope and resolution [8,9].

Conventional strategies for widening the FoV of TTE focused on hardware-based solutions, such as diverging acoustic lenses [9,10] or wide-aperture/flexible arrays [11]. However, these modifications invariably increase device bulk and cost, while compromising signal-to-noise performance and ergonomic flexibility, and yield only modest improvements in FoV (typically under 30°). To our knowledge no prior study has addressed FoV enlargement in TTE via a conditional GAN that is explicitly aware of the probe's cone geometry. By introducing echoGAN—with domain-aware mask augmentation that mimics the sector shape–we provide the first probe-agnostic solution that extends the view while preserving quantitative cardiac measurements, thereby overcoming the fundamental FoV-versus-resolution trade-off without any change to acquisition hardware.

Furthermore, this extended FoV could be particularly beneficial for less experienced physicians in manual navigation. The study by Chisholm et al. [12] and the work of Vinodth et al. [13] focus on cardiac ultrasound training, specifically evaluating the duration of training necessary to achieve proficiency in TTE. These studies demonstrate that mastering TTE requires a considerable amount of time. By offering a wider yet detailed view, our cGAN-based approach has the potential to shorten the learning curve associated with ultrasound imaging. This process is often not straightforward and demands a high level of dexterity and spatial awareness. Given these challenges, there is a continuous search for technological advancements that can aid in simplifying the navigation process, making echocardiography more accessible and reliable, especially for less experienced practitioners. Moreover, the utilization of semi-synthetic data to enhance training models for medical device segmentation [2] underscores the potential for augmented datasets to refine AI-driven tools.

In the domain of medical imaging, particularly ultrasound imaging, outpainting introduces unique challenges that are markedly distinct from those encountered in the outpainting of general or natural images [14–16]. The paramount requirement for anatomical accuracy in medical images necessitates that outpainting algorithms not only generate visually plausible extensions but also ensure these extensions accurately represent the expected anatomy. This is crucial, as any deviation could compromise the clinical utility of the image. Moreover, ultrasound images are characterized by specific textures and patterns that denote various tissues, fluids, and pathological conditions.

Our contributions are twofold. (i) Domain-aware cone masking: we propose an augmentation that blanks out angular sectors matching the phased-array scan geometry, compelling the network to learn anatomically faithful completions. (ii) EchoGAN architecture: we present the first conditional GAN specifically designed for TTE FoV extension, implemented in a lightweight generator–discriminator pair that can be trained on a single workstation GPU and supports real-time inference on the embedded GPUs typical of ultrasound systems.

The paper is organized as follows. In the subsequent section, we provide a description of the proposed approach, detailing the echoGAN framework which leverages domain-aware augmentation and conditional GAN architecture. The Section 3 describes the datasets utilized to validate the proposed approach. The Section 4 discusses the results obtained from our experiments. Following the discussion, the paper concludes with final remarks on the study's findings and implications.

Our preprocessing scripts and implementation are available at https://github.com/BraveDistribution/echogan.[1]

2. Related works

Recent advances in image outpainting have progressed along two complementary lines: (i) generative backbone design and (ii) auxiliary guidance that constrains the newly generated content. Early GAN-based approaches such as Painting from Part [17] employed multi-stage generators that refine coarse extrapolations with local patch-level detail, effectively unifying in-painting and out-painting in a single pipeline. Subsequent work sought greater diversity and long-range coherence: CoModGAN [18] injects the partially observed image as a co-modulation signal into a StyleGAN2 backbone, bridging conditional in-painting and unconditional generation to complete very large holes, while InOut [19] frames out-painting as GAN inversion, optimizing multiple latent codes to avoid texture repetition. Transformer backbones were then explored—QueryOTR [20] uses learnable queries to extend context in all directions and TFill [21] models image completion as a sequence-to-sequence problem that captures global dependencies before a convolutional decoder restores high-frequency detail. Most recently, diffusion-based models such as PQDiff [22] have demonstrated arbitrary-scale expansion by learning arbitrary, continuous spatial relationships from randomly cropped view pairs encoded as positional queries, enabling single-step generation.

Orthogonal to backbone choice, researchers have developed increasingly sophisticated guidance mechanisms to enforce semantic plausibility. Layout- or scene-driven methods first generate a high-level representation and then decode it into pixels: Scene-Graph Expansion [23] enlarges the scene graph via a transformer before rendering the missing area, while CTO-GAN [24] combines a predicted semantic layout with an object-level contrastive loss to keep foreground and synthesized background consistent. Physical cues are also proving valuable: Shadow-Enlightened Out-painting [25] leverages cast shadows to infer off-frame objects, ensuring that extrapolated content respects real

[1] https://github.com/BraveDistribution/echogan.

lighting. Taken together, these studies indicate a clear trajectory, moving from unconditional GANs that learn a generic prior to architectures that embed geometric, semantic, or physical constraints.

Limited work addresses outpainting in the biomedical domain, and most solutions rely on multi-frame stitching rather than single-frame extrapolation. SynStitch [26] and its 3-D extension LOTUS [27] register partially overlapping ultrasound sweeps and use diffusion to blend the seams, but they require multiple acquisitions and assume appreciable overlap. Diffusion-based CT-Palette [28] recovers truncated CT slices, yet operates on Cartesian cross-sections with well-defined Hounsfield ranges, not on ultrasound's fan-shaped, low-SNR data. Generative editing frameworks—for example, ESCM-PGGAN for fetal four-chamber views [20] or nucleus-to-actin synthesis in fluorescence microscopy [29]—alter or predict structures within the recorded image rather than extending its spatial coverage. Consequently, there is still no published method that enlarges the field of view of a single transthoracic-echocardiography frame by generating anatomically coherent cardiac tissue outside the recorded sector. These gaps underscore the need for a completion strategy that respects the cone-shaped polar sampling of cardiac ultrasound and can operate on a single frame without auxiliary acquisitions, while remaining computationally lean to fit the hardware envelope of clinical scanners.

3. Materials and methods

The present study proposes GAN for extending the FoV in echocardiography to enhance both manual and automatic navigation. From a computer vision perspective, FoV extension is formulated as an outpainting problem. Outpainting seeks a semantically consistent extension of the input image beyond its available content. The outpainting process involves the GAN generating new pixels and features in a way that seamlessly integrates with the existing content, maintaining the anatomical and contextual coherence of the cardiac structures.

Initial results indicate that our approach may refine navigation accuracy and could facilitate a more comprehensive visualization of cardiac structures. In our experiments, FoV enhancement showed potential to reduce navigation difficulty by providing a more comprehensive view. The proposed approach aims to address longstanding challenges of incomplete cardiac visualization during echocardiography.

3.1. Problem formulation

Given an input image $x \in \mathbb{R}^{H\times W}$, where H and W represent the height and width of the original image, and a binary mask $m \in \{0,1\}^{H\times W}$ that defines the region to be outpainted (with 1 indicating pixels to be retained and 0 indicating the masked region), the task is to generate an extended output image $\hat{y} \in \mathbb{R}^{H\times(W+\Delta W)}$ that fills in the missing content in the masked area while preserving visual and anatomical coherence.

The problem can be expressed as learning a mapping function $G : (\mathbb{R}^{H\times W} \times \{0,1\}^{H\times W}) \to \mathbb{R}^{H\times(W+\Delta W)}$ such that:

$$\hat{y} = G(x \odot (1-m), m) \tag{1}$$

where $\odot$ denotes the element-wise multiplication operator and $x\odot(1-m)$ is the input image with the masked region removed, leaving only the known parts.

The objective is to train G such that the generated image $\hat{y}$ is indistinguishable from the true extended image y under a discriminator D.

3.2. echoGAN

To effectively address the outpainting task, we propose echoGAN, which employs a novel domain-aware augmentation strategy. The augmentation method, network architecture, and training procedure are described in the following sections.

3.2.1. Domain-aware augmentation

We propose a novel domain-aware augmentation technique specifically designed for ultrasound imaging. This augmentation simulates the cone-shaped FOV typical in ultrasound by masking the side regions of the cone. The augmented mask requires the generator to reconstruct the occluded areas.

For ultrasound image x, the cone-like FOV is defined by a binary mask $m_{\text{cone}} \in \{0,1\}^{H\times W}$, where

$$m_{\text{cone}}(i,j) = \begin{cases} 1, & \text{if } \tan^{-1}\left(\frac{|j-c_x|}{H-i}\right) \le \frac{\theta}{2}, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Here, $m_{\text{cone}}(i,j) = 1$ for pixels inside the cone-like region and $m_{\text{cone}}(i,j) = 0$ for those outside the cone (cut-out regions). The cone is parameterized by the center of the cone (c_x, c_y) and the angular spread θ.

The augmentation simulates missing information by masking the side regions of the cone, requiring the generator to infer and reconstruct the missing areas by utilizing an augmented mask $m_{\text{augment}} \in \{0,1\}^{H\times W}$ defined as

$$m_{\text{augment}}(i,j) = \begin{cases} 1, & \text{if } \tan^{-1}\left(\frac{|j-c_x|}{H-i}\right) \le \frac{\theta_{\text{shrink}}}{2}, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

where $\theta_{\text{shrink}} < \theta$ controls the extent of the occlusion on the sides.

Then, the augmented input image is defined as

$$z_{\text{aug}} = \{x \odot m_{\text{augment}}, m_{\text{augment}}\}, \tag{4}$$

where $x \odot m_{\text{augment}}$ represents an ultrasound masked image retaining only the visible region.

Finally, the generator G takes the augmented input z_{aug} and produces the extended image $\hat{y}$:

$$\hat{y} = G(z_{\text{aug}}). \tag{5}$$

3.2.2. Network architecture and training

The proposed GAN architecture consists of two neural networks: generator and discriminator. As a generator we employ U-Net architecture, as introduced by Ronneberger et al. [30]. It features two input channels: one for the masked image and the other for the binary mask, which indicates the regions for outpainting. It is configured with four downsampling layers, each characterized by specific kernel sizes [7, 5, 5, 5] and strides [1, 2, 2, 2], optimizing the network's capacity for feature extraction and image reconstruction within the designated areas.

The Discriminator employs a conventional Convolutional Neural Network (CNN) design, structured with a sequence of channels [1, 32, 64, 128, 128] and uniform strides [1, 2, 2, 2] across layers, with a kernel size consistently set to 3. This configuration enables effective differentiation between generated and authentic images, ensuring the generative model's outputs closely mimic the target distribution. This generator–discriminator architecture represents a lightweight solution that runs easily on the workstation-class or embedded GPUs common in ultrasound systems.

This network, depicted in Fig. 2, is trained by a combination of two different loss functions: adversarial loss function and learned perceptual image patch similarity. The adversarial loss consists of two components: one for the discriminator ($\mathcal{L}_D$) and one for the generator ($\mathcal{L}_G$). The discriminator loss $\mathcal{L}_D$ is designed to correctly distinguish between real and fake images, while the generator loss $\mathcal{L}_G$ aims to fool the discriminator into classifying fake images as real. The adversarial loss for the discriminator can be formulated as:

$$\mathcal{L}_D = -\mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] - \mathbb{E}_{x\sim p_{\text{data}}(x),m}[\log(1 - D(G(x\odot(1-m), m)))] \tag{6}$$

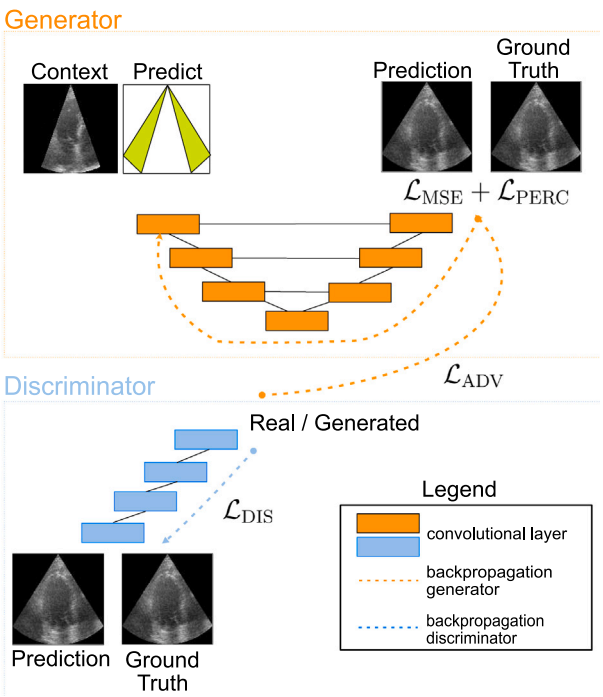

Fig. 2. Proposed echoGAN architecture.

where x represents the real images sampled from the data distribution, m is the mask, and $x \odot (1-m)$ denotes the image with the masked area removed. The generator loss is then denoted as

$$\mathcal{L}_G = -\mathbb{E}_{x \sim p_{\mathrm{data}}(x), m}[\log D(G(x \odot (1-m), m))]. \tag{7}$$

The Learned Perceptual Image Patch Similarity (LPIPS) metric evaluates the perceptual similarity between two images in a way that is more aligned with human visual perception than traditional metrics like MSE. The general form of the LPIPS equation involves computing differences in deep feature representations extracted from the two images being compared, then weighting and summing these differences to produce a final similarity score. The LPIPS is defined as

$$\mathcal{L}_{\mathrm{LPIPS}} = \sum_l w_l \cdot \frac{1}{H_l W_l C_l} \sum_{h,w,c} \|\phi_l(I^1)_{h,w,c} - \phi_l(I^2)_{h,w,c}\|_2^2 \tag{8}$$

where l is the layer of the deep feature extractor, H accounts for height, W for width, C for the number of channels, ϕ is the feature extractor (pretrained VGG16).

Final loss for generation is then a summation of both losses

$$\mathcal{L} = \mathcal{L}_{\mathrm{G}} + \mathcal{L}_{\mathrm{LPIPS}}. \tag{9}$$

The objective is to find the optimal generator G^* that minimizes the combined loss $\mathcal{L}$ while simultaneously training the discriminator D to maximize $\mathcal{L}_D$:

$$G^* = \arg\min_G \max_D \left[\mathcal{L}_D + \mathcal{L}\right]. \tag{10}$$

This formulation ensures that the generator produces realistic outpainted images that closely resemble true TTE ultrasound images.

3.3. Data

We utilized two datasets in this paper: the CAMUS dataset [31] collected under the CAMUS project [31] at the University Hospital of St. Étienne and the EchoNet Dynamic echocardiography dataset [32].

CAMUS dataset was designed to facilitate measurements of Left Ventricular Ejection Fraction (LVEF). For each patient, 2D apical four-chamber and two-chamber view sequences were recorded. These sequences were extracted from video recordings, yielding an average of 20 images per patient (minimum: 10, maximum: 42). In total, the dataset comprises 9964 images, of which 7917 images from 400 patients were allocated for training a generative adversarial network. To prevent data leakage, all data from a given patient were confined to either the training set or the testing set, ensuring no overlap.

To further validate the echoGAN, we incorporated also the EchoNet Dynamic, consisting of 10,030 apical four-chamber echocardiography videos obtained from routine clinical care at Stanford University Hospital between 2016 and 2018. Each video provides a visualization of the heart from various angles and positions using different image acquisition techniques. To standardize the dataset, all videos were cropped and masked to exclude textual information and irrelevant content outside the scanning sector. The resulting frames were subsequently downsampled using cubic interpolation to a uniform resolution of 112 × 112 pixels. Similarly, as for the CAMUS dataset, we kept data from 20% of patients for validation.

Due to the absence of US data in raw format within the datasets, our analysis is confined to the final, scan-converted images. A similar approach might be implemented in raw data as well. We segmented the cone from the image data through the application of conventional image processing techniques. Initially, we located the first non-zero pixel from the top as a means to identify the tip of the probe. Subsequently, we delineated the furthest left and right boundaries emanating from the probe to the edge of the image. By determining the angular relationship between these boundaries, we systematically removed segments measuring 15, 23, 30, and 40 degrees from each side in a symmetrical fashion. The resulting masks, corresponding to the excised segments, were preserved for further analysis.

As for the preprocessing during the training, we just scaled the intensity into [0, 1] interval.

3.4. Experimental setup

All experiments were conducted on a Debian-based Linux operating system using Python version 3.11 and PyTorch version 2.0.0. The training was performed on an NVIDIA RTX 3090 Ti GPU with 24 GB of VRAM and a 12th-generation Intel Core i9-12900K CPU with 64 GB of system RAM. To ensure reproducibility and consistency, we maintained this hardware configuration throughout all experiments.

Training was conducted for a fixed 140 epochs using mixed precision training (16-bit floating point precision) to accelerate computation and reduce memory footprint. We did not apply gradient clipping or other advanced optimization techniques beyond standard Adam optimizer settings. The generator model contained approximately 5.1 million parameters, while the discriminator consisted of approximately 854,000 parameters. With a batch size of 128, the total GPU memory utilization reached approximately 11.45 GB.

During inference, the processing time per image was consistently below 0.3 ms, demonstrating efficiency and suitability for practical deployment.

3.4.1. Training statistics

Fig. 3 compares two training configurations of EchoGAN on the CAMUS dataset for cut of size 80. Specifically, it contrasts the effect of using an adversarial loss weight of 10^{-5} versus 10^{-3}. The figure reports training adversarial loss, discriminator loss, training and validation perceptual losses, as well as validation metrics, including peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) loss.

For the adversarial loss weight of 10^{-5}, the training adversarial and discriminator losses remain consistently low and stable near zero, indicating that the discriminator strongly overpowers the generator early

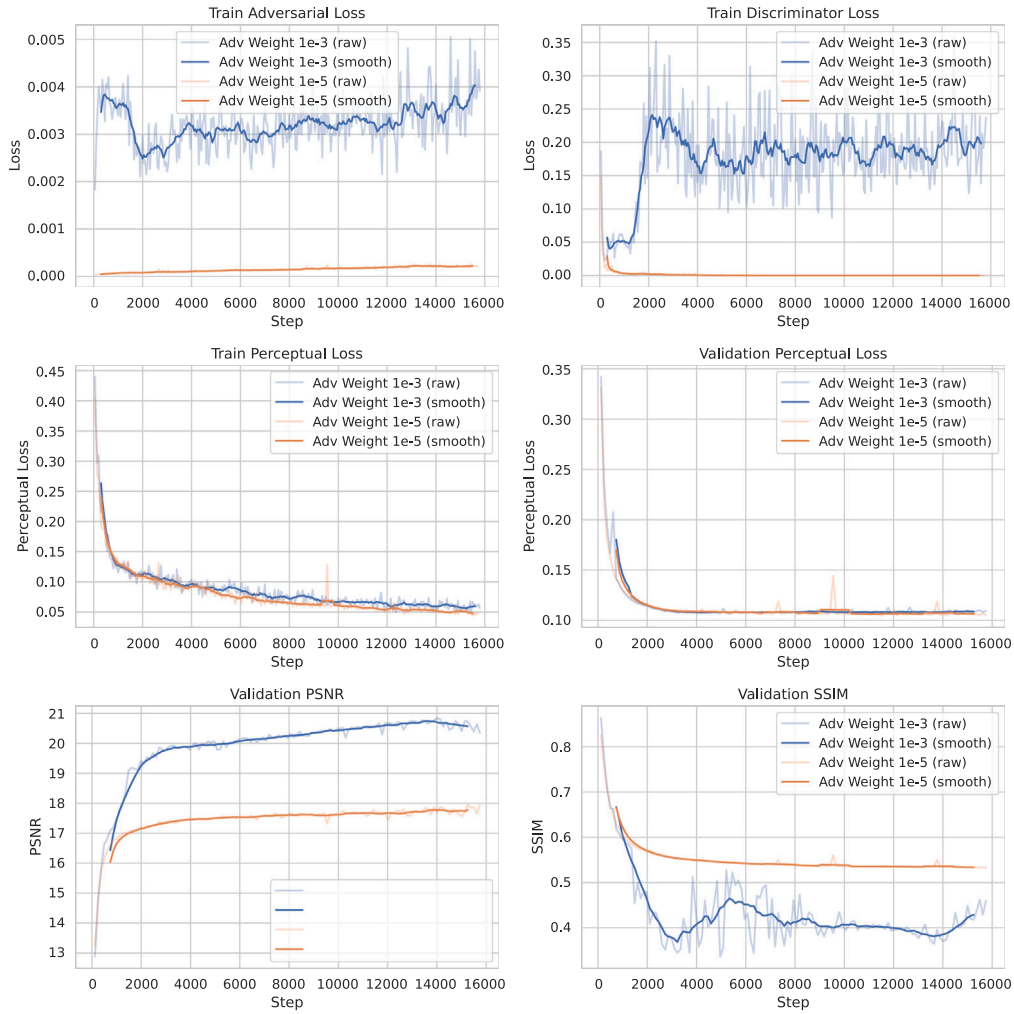

Fig. 3. Training statistics.

in training. This imbalance effectively neutralizes the adversarial min–max dynamic, causing the adversarial objective to become negligible. Consequently, performance deteriorates, evidenced by a decrease in validation PSNR from approximately 20 to 18, an increase (worsening) of the SSIM loss from approximately 0.45 to 0.55, while perceptual loss remains relatively stable.

In contrast, using an adversarial weight of 10^{-3} maintains a robust adversarial interplay, yielding superior validation results (higher PSNR and lower SSIM loss). These findings highlight the critical importance of carefully balancing the adversarial loss weight to achieve optimal reconstruction quality.

3.4.2. Hyperparameter tuning

We performed a hyperparameter search evaluating two parameters: the adversarial loss weight (values: 10^{-1}, 10^{-3}, 10^{-5}) and batch size (values: 64, 128, 256). The experiments were conducted on the CAMUS dataset frames with cuts at 30 and 46. Results are summarized in Fig. 4, where the y-axis represents different adversarial loss weights, the x-axis represents batch sizes, and the cell color intensity ranges from dark blue (best score) to light blue (worst score).

For the PSNR metric, the highest values were consistently achieved with an adversarial weight of 10^{-1}, while changes in batch size exhibited negligible influence on the results. For LPIPS, neither adversarial weight nor batch size had a substantial impact on the final loss, aligning with observations from Fig. 3. Lastly, the SSIM loss achieved optimal results at an adversarial weight of 10^{-1}, with no clear trend emerging concerning the batch size's effect.

4. Results

To quantitatively assess the quality of generated images, we use classic per-pixel measure such as Peak Signal-to-Noise Ratio (PSNR), but also metrics proposed recently to measure how similar two images are in a way that coincides with human judgment, such as Fréchet Inception Distance (FID) [33], Learned Perceptual Image Patch Similarity (LPIPS) [34] and structural similarity (SSIM) [35].

In Fig. 5, we provide visual samples of outpainted images generated by the echoGAN. These examples show the network's ability to extrapolate relevant cardiac features beyond the visible boundaries, demonstrating the ability of the proposed GAN architecture to enhance the overall imaging capabilities in echocardiography. The visual representations serve to illustrate the effectiveness of the echoGAN in creating realistic and contextually coherent extensions of the original ultrasound images, thereby contributing to the advancement of FoV extension techniques in cardiac imaging applications.

To quantitatively measure the quality of outpainted images, the next sections provide results for different sizes of cone cuts that were used to generate the sides of the cone.

4.1. Different size of extended view

We evaluated the influence of the training cone size on the quality of outpainted sections in ultrasound images. Fig. 6 illustrates two distinct scenarios: the first utilizes a narrower 10-degree cone cut as the

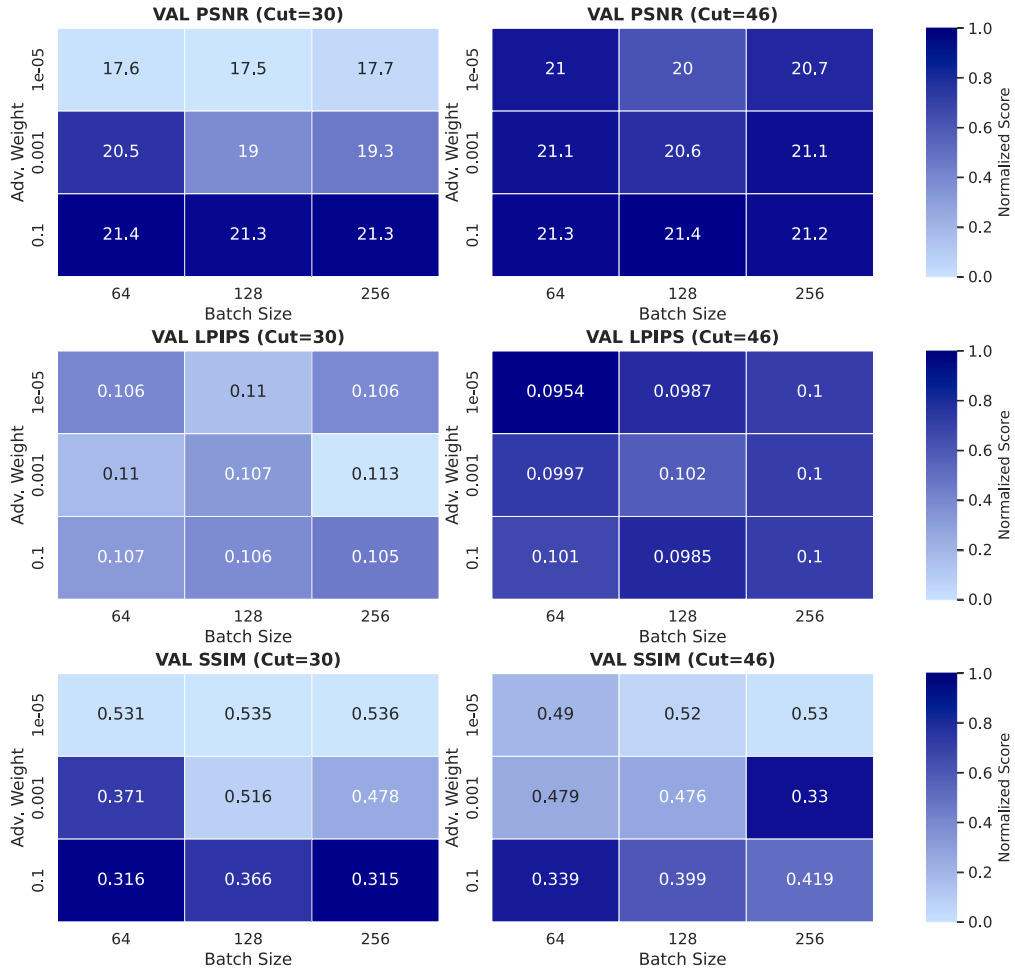

Fig. 4. Training statistics.

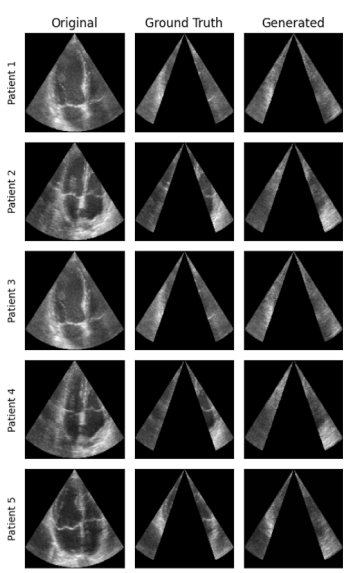

Fig. 5. Several examples from different patients on ground truth image and outpainted image.

inference input, with the echoGAN algorithm tasked with generating outpainted extensions that span 40 degrees on both sides; the second employs a 60-degree cone cut for inference, with outpainting applied to 15-degree segments on each side. Notably, the generated images appear highly realistic, even when echoGAN is interfaced with only a 10-degree cut. The algorithm demonstrates the ability to reproduce cardiac muscle structures with high fidelity. This is particularly evident in the first, second, and fourth rows, where echoGAN successfully outpaints a complete right atrium; that is an encouraging result given that the right atrium was not included in the 10-degree cut, suggesting that echoGAN can reconstruct it entirely. In the case of the left ventricle and left atrium, we observe that echoGAN is capable of outpainting the missing boundary parts for the 30-degree cut and can significantly reconstruct substantial portions of these structures for the 10-degree cut. Again, to quantitatively evaluate the quality of outpainted USG images, we measure similarity metrics for different sizes of outpainted regions. The results are provided in Table 1. To provide also a state-of-the-art comparison, we employed CoMoDGAN [18] and TFill [21]. The 30-degree cut means that networks generate 15 degrees on both sides. Similarly, in the case of 80 degrees, networks generate 40 degrees on both sides, using only a very narrow 10-degree cone as an input.

As expected, smaller FID values, indicating better similarity, are achieved for the cases when a broader cut is used as an input for network inference. The trend is similar for other metrics, with the exception of PSNR. However, in this case, a higher value indicates a better image.

To further verify the effectiveness of our method, we applied the permutation test between echoGAN and both competitive methods for all four considered metrics. Although the Wilcoxon signed-rank test is commonly recommended for the comparison of multiple methods over several datasets [36], in our case, the datasets “Cut 30”, “Cut

Table 1
The results for different sizes of outpainted area across CAMUS and EchoNet datasets.

Method	Metric	CAMUS				EchoNet dynamic			
		Cut 30	Cut 46	Cut 60	Cut 80	Cut 30	Cut 46	Cut 60	Cut 80
echoGAN	**FID ↓**	122.08	**143.83**	**145.74**	**149.63**	**60.24**	**74.33**	89.21	**110.42**
	LPIPS ↓	**0.053**	**0.09**	**0.13**	**0.12**	0.16	**0.16**	**0.16**	**0.22**
	PSNR ↑	**23.91**	**21.46**	**19.77**	**19.95**	20.15	**20.14**	**20.13**	**17.89**
	SSIM ↓	**0.28**	**0.31**	**0.47**	**0.46**	0.48	0.48	0.48	0.59
CoMoDGAN [18]	**FID ↓**	109.59	158.68	208.59	194.03	75.94	88.67	106.77	132.81
	LPIPS ↓	0.1642	0.21	0.27	0.26	0.21	0.25	0.29	0.35
	PSNR ↑	11.59	9.31	7.54	7.79	7.04	7.03	7.03	6.99
	SSIM ↓	0.57	0.70	0.80	0.80	0.78	0.82	0.85	0.90
TFill [21]	**FID ↓**	**99.80**	163.05	240.32	273.60	62.49	77.94	**78.28**	112.28
	LPIPS ↓	0.26	0.37	0.46	0.51	**0.14**	0.21	0.258	0.33
	PSNR ↑	15.57	12.90	11.58	10.24	**22.34**	19.84	17.85	15.74
	SSIM ↓	0.34	0.48	0.6	0.63	**0.19**	**0.29**	**0.39**	**0.53**

Table 2
p-values of the permutation tests for the comparison between echoGAN and each of the other approaches.

	FID	LPIPS	PSNR	SSIM
CoMoDGAN	0.0156[a]	0.0078[a]	0.0078[a]	0.0078[a]
TFill	0.2812	0.0156[a]	0.0312[a]	0.8594

[a] Denotes results where echoGAN is significantly better ($p < 0.05$).

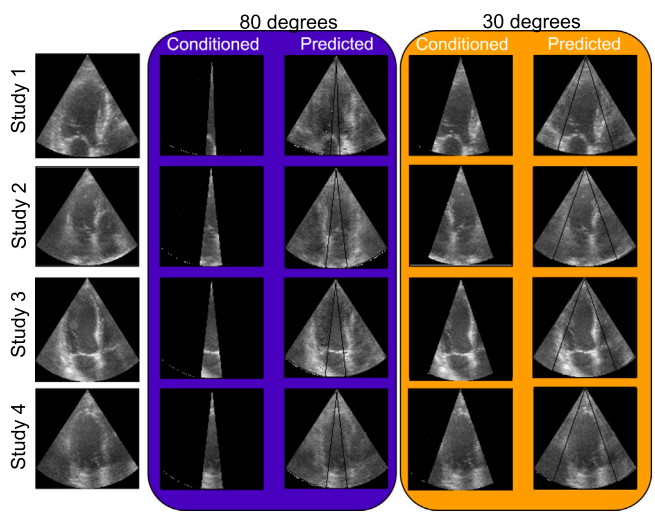

Fig. 6. Extending field of view (yellow) vs. basic (teal) for 30 and 80 degree outpainting.

46", "Cut 60", and "Cut 90" are derived from the same source dataset. Therefore, the assumption of independence between these data points is questionable, so we employ permutation tests. As shown in Table 2, echoGAN achieves statistically significant performance gains in most of the cases over the competing methods CoMoDGAN and TFill.

4.2. Right ventricle area

The majority of diagnostic, prognostic, and monitoring efforts in cardiology traditionally emphasize the analysis of the left heart due to its direct role in systemic circulation. However, comprehensive echocardiographic assessment necessitates detailed examination of the entire heart, including the right ventricle (RV), which plays a pivotal role in pulmonary circulation and systemic venous return [37]. During standard echocardiography, capturing the complete anatomy of the right heart can be challenging. The RV, in particular, may be partially or entirely outside the ultrasound beam's FoV, limiting the accuracy and completeness of the examination. Despite the primary focus on the left heart in clinical practice, quantification of right heart parameters is crucial in various clinical scenarios.

Our experiments are specifically aimed at the quantification of the RV area from both original and outpainted echocardiograms. The rationale for focusing on the RV area is influenced by the established practice in cardiology where, for instance, the measurement of left ventricular ejection fraction incorporates the left ventricular area as a crucial parameter in its evaluation [38]. Drawing inspiration from such methodologies, we hypothesize that if the RV areas derived from generated (outpainted) and real echocardiograms are similar, it would substantiate the utility and validity of the generated images for clinical purposes.

Moreover, accurate assessment of RV size and function is paramount in conditions such as pulmonary hypertension, various cardiomyopathies, and right ventricular infarction, where the right heart's performance directly influences patient outcomes [37].

In our experimental setup, both the original and the artificially generated echocardiograms were manually annotated to obtain segmentation masks. This process allowed for precise comparison of RV volumes, assessing the efficacy and accuracy of GAN-based outpainting in reproducing anatomically coherent extensions of the cardiac structures. By comparison of the RV volumes derived from the conventional and outpainted echocardiograms, this experiment aims to validate the potential of echoGANs in overcoming the intrinsic limitations of current echocardiography practices. Comparison of segmentation for RV in the original and in the outpainted image is in Figs. 7 and 8.

To statistically evaluate the difference between the real and AI-generated right ventricle area measurements, we first assessed whether each group of measurements conformed to a normal distribution. This preliminary step involved applying the Shapiro–Wilk test [39], a widely accepted method for testing normality. The results of the Shapiro–Wilk test indicated that the data do not follow a normal distribution, thereby necessitating the selection of a non-parametric statistical test for further analysis. Consequently, we opted for the permutation test, a robust non-parametric approach that does not rely on assumptions of normality or symmetry. This test was employed to examine our hypothesis that there is no significant statistical difference between the areas determined from the AI-generated and real echocardiography data. The choice of the permutation test was appropriate given the non-normal distribution of our data, allowing us to conduct a reliable comparison of the paired differences between the two groups. The results of this test confirmed the null hypothesis ($T = 8.804$, $p = 0.762$), indicating that there is no statistically significant difference between the RV areas measured in the original echocardiographic images and those derived from images extended through echoGAN outpainting. This finding validates the accuracy and effectiveness of the GAN-based outpainting technique in echocardiography. The ability of GANs to generate anatomically coherent extensions of the cardiac structures without compromising the integrity of the images suggests that this approach can reliably extend the field of view in echocardiographic navigation.

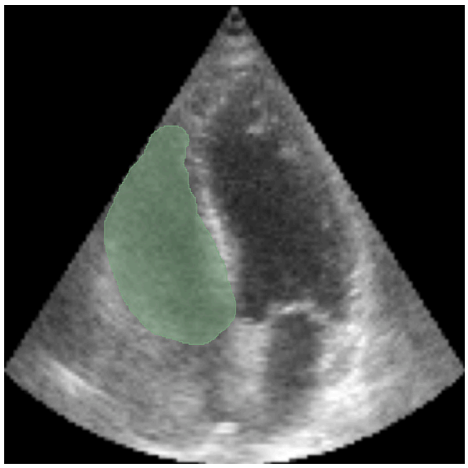

Fig. 7. Segmentation over Right Ventricle for GT 4CH view.

Fig. 8. Segmentation of outpainted RV with 80 degrees cut for the same 4CH view.

5. Discussion

One of the main limitations of our study is limited data. The present study relies on two publicly available datasets—CAMUS and EchoNet-Dynamic—whose patients are predominantly middle-aged adults with preserved image quality and limited pathological variety. This limitation could affect the generalizability of our findings, as the performance of the echoGAN might vary with different or more heterogeneous datasets. Consequently, our results may not translate directly to populations under-represented in these sets, such as paediatric or elderly patients, individuals with congenital or valvular heart disease, or scans acquired with different vendors, probes or gain settings. Moreover, while the echoGAN architecture showed potential to generate anatomically plausible images, there remains a risk of generating artifacts, especially in more complex or ambiguous cardiac regions. These potential inaccuracies necessitate further refinement of the model and algorithms to ensure reliability and clinical utility.

Additionally, a prevalent limitation of publicly available datasets is their focus on specific traditional views like apical two-chamber (A2C) and apical four-chamber (A4C), while neglecting other standard views such as parasternal (parasternal long axis (PLAX), parasternal short axis (PSAX)) and subcostal views (subcostal 4CH and subcostal short axis). Moreover, these datasets completely omit images captured from non-standard views, which are often not even recorded as they are primarily used for navigation purposes.

To address this limitation, future work could incorporate shape priors derived from medical imaging modalities that capture comprehensive chest information, such as CT or MRI. Alternatively, the dataset could be expanded with images from non-standard views generated through real-time ultrasound simulation solutions from CTs or MRIs, as demonstrated in [40]. In addition to physics-based simulators, unpaired image-to-image translation methods like mUNIT [41] or CycleGAN [42] could be employed. While paired image-to-image translation might yield superior results, the required data is rarely available.

Because echoGAN synthesizes image content beyond the acoustically acquired sector, it can occasionally introduce visual artifacts, particularly in regions with sparse or highly attenuated signals. For this reason, we position echoGAN chiefly as a navigation aid. Practical safeguards include user-controlled blending that allows the operator to fade out the synthetic margins at will or future integration of uncertainty maps that automatically flags pixels with low generative confidence.

Directly outpainting ultrasound views using scanline data could be a faster and more precise approach, provided that access to the probe hardware and controls were available, which is currently not feasible. However, oversampling the area near the probe and heavily extrapolating at the end of the scanning volume could introduce errors. Should this capability become available, it would offer a promising avenue for further exploration.

6. Conclusion

This paper introduced echoGAN, an innovative approach leveraging cGANs to extend FoV in Transthoracic Echocardiography ultrasound imaging. By addressing the inherent trade-offs between FoV and resolution in standard ultrasound imaging, echoGAN is capable of producing realistic outpainted regions, enhancing the visualization of cardiac structures while maintaining high-resolution detail. echoGAN shows promise in aiding both manual and automated ultrasound navigation, reducing the learning curve for less experienced operators and facilitating more accurate diagnoses.

Quantitative analysis demonstrated the robustness of echoGAN, with a $FID = 122.08$ and $LPIPS = 0.053$ for a 30-degree outpainting scenario and slightly higher FID values for larger outpainted regions, such as 143.83 for a 46-degree cut and 149.63 for an 80-degree cut. These results indicate that echoGAN can effectively maintain anatomical coherence and quality even as the scope of outpainting increases. Clinical validation through the comparison of right ventricular area measurements in original and outpainted images further reinforced the reliability of this approach, with no statistically significant differences found, as verified by the statistical permutation test.

While echoGAN shows significant potential, limitations include the need for more diverse datasets and improvements in generating artifact-free extensions, particularly in complex anatomical areas. Occasional GAN-induced artifacts remain a limitation; the out-painted sector should therefore be regarded as an adjunct navigational overlay, not a stand-alone diagnostic image, until larger-scale validation confirms its reliability in routine clinical practice. Future work should explore integrating shape priors from CT or MRI scans and include non-standard cardiac views to enhance anatomical fidelity.

CRediT authorship contribution statement

Matej Gazda: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Conceptualization. **Jakub Gazda:** Writing – review & editing, Validation, Methodology. **Samuel Kadoury:** Writing – review & editing, Supervision. **Robert Kanasz:** Writing – review & editing, Visualization, Software. **Peter Drotar:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00394 and by the Slovak Research and Development Agency under contract No. APVV-23-0411.

References

[1] K. Subramaniam, H. Subramanian, J. Knight, D. Mandell, S.M. McHugh, An approach to standard perioperative transthoracic echocardiography practice for Anesthesiologists—Perioperative transthoracic echocardiography protocols, J. Cardiothorac. Vasc. Anesth. 36 (2) (2022) 367–386.

[2] C. Feng, T. Luo, Y. Luo, N. Zhao, K. Huang, C. Xiao, Contrast-enhanced transthoracic echocardiography applied in evaluation of pulmonary right-to-left shunt: A preliminary study, Comput. Med. Imaging Graph. 68 (2018) 55–60.

[3] C. Mitchell, P.S. Rahko, L.A. Blauwet, B. Canaday, J.A. Finstuen, M.C. Foster, K. Horton, H. Kitaoka, B. Kurtz, P.M. Lepper, K. Maganti, B.S. Upadhya, S.E. Wiegers, B. Kurt, Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: Recommendations from the American Society of Echocardiography, J. Am. Soc. Echocardiogr. 32 (1) (2019) 1–64.

[4] R. Amyot, E. Yu, G. Honos, J. Choy, G. Schnell, H. Leong-Poi, Contrast echocardiography: Putting things into perspective – a Canadian cardiovascular society/Canadian society of echocardiography joint commentary, Can. J. Cardiol. 24 (11) (2008) 835–837.

[5] C.R. Usry, S.R. Shin, J.K. Aden, K.M. Ziada, K. Addetia, R.M. Lang, V. Mor-Avi, Optimizing Contrast-Enhanced echocardiography by employing a sonographer driven protocol, J. Echocardiogr. 19 (2021) 173–178.

[6] A. Ng, J. Swanevelder, Resolution in ultrasound imaging, Contin. Educ. Anaesth. Crit. Care Pain 11 (5) (2011) 186–192.

[7] R. Soemantoro, A. Kardos, G. Tang, An AI-powered navigation framework to achieve an automated acquisition of cardiac ultrasound images, Sci. Rep. 13 (2023).

[8] A. Salari, M. Audoin, B.G. Tomov, B.Y.S. Yiu, E.V. Thomsen, J.A. Jensen, Beamformer for a lensed Row-Column array in 3D ultrasound imaging, IEEE Trans. Ultrason. Ferroelectr. Freq. Control (2025) 1.

[9] J. Chen, B. Zhuang, J. Peng, Z. Zhang, B. Wang, C. Dai, D. Wu, Variable curvature flexible transducer for abdominal expandable-imaging, IEEE Trans. Ultrason. Ferroelectr. Freq. Control (2025) 1.

[10] M. Audoin, A. Salari, B.G. Tomov, K.F. Pedersen, J.A. Jensen, E.V. Thomsen, Diverging polymer acoustic lens design for High-Resolution Row-Column array ultrasound transducers, IEEE Trans. Ultrason. Ferroelectr. Freq. Control 71 (1) (2024) 202–213, Epub 2024 Jan 9.

[11] J. Elloian, J. Jadwiszczak, V. Arslan, Y. Qiu, M. Singh, F. Adib, S. Sonkusale, Flexible ultrasound transceiver array for Non-Invasive Surface-Conformable imaging enabled by geometric phase correction, Sci. Rep. 12 (2022) 16184.

[12] C.B. Chisholm, W.R. Dodge, R.R. Balise, S.R. Williams, L. Gharahbaghian, A.-S. Beraud, Focused cardiac ultrasound training: how much is enough? J. Emerg. Med. 44 (4) (2013) 818–822.

[13] V.B. Nanjayya, S. Orde, A. Hilton, Y. Yang, C. Costello, J. Evans, M. Nalos, K. Yastrebov, Levels of training in critical care echocardiography in adults. Recommendations from the college of intensive care medicine ultrasound special interest group, Australas. J. Ultrasound Med. 22 (1) (2019) 73–79.

[14] Y.-C. Cheng, C.H. Lin, H.-Y. Lee, J. Ren, S. Tulyakov, M.-H. Yang, InOut: Diverse image outpainting via GAN inversion, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2022, pp. 11421–11430.

[15] J. Li, C. Chen, Z. Xiong, Contextual outpainting with Object-Level contrastive learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2022, pp. 11451–11460.

[16] Y. Wang, Y. Wei, X. Qian, L. Zhu, Y. Yang, ReGO: Reference-guided outpainting for scenery image, IEEE Trans. Image Process. 33 (2024) 1375–1388.

[17] D. Guo, Y. Li, D. Liu, J. Wu, Q. Fan, Y. Xu, J. Pan, Painting From Part: Image Completion With Multi-stage GANs, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, ICCV, 2021, pp. 9276–9285.

[18] S. Zhao, J. Cui, Y. Sheng, Y. Dong, X. Liang, E.I. Chang, Y. Xu, Large scale image completion via co-modulated generative adversarial networks, 2021, arXiv preprint arXiv:2103.10428.

[19] Y.-C. Cheng, Y.-J. Chou, Y.-H. Tsai, H.-T. Chen, InOut: Diverse Image Outpainting via GAN Inversion, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2022, pp. 14832–14841.

[20] S. Qiao, X. Zhou, J. Liu, Y. Zhang, Y. Wang, Z. Tian, A progressive growing generative adversarial network composed of enhanced style-consistent modulation for fetal ultrasound four-chamber view editing synthesis, Eng. Appl. Artif. Intell. 133 (2024) 108438.

[21] C. Zheng, T.-J. Cham, J. Cai, D. Phung, Bridging global context interactions for high-fidelity image completion, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 11512–11522.

[22] S. Zhang, J. Huang, Q. Zhou, Z. Wang, F. Wang, J. Luo, J. Yan, Continuous-multiple image outpainting in one-step via positional query and a diffusion-based approach, 2024, arXiv preprint arXiv:2401.15652.

[23] C.-A. Yang, I.-C. Chiu, T.-Y. Wang, J.-B. Huang, Scene Graph Expansion for Semantics-Guided Image Outpainting, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2022, pp. 3374–3384.

[24] J. Li, M. Qiang, H. Tang, D. Xu, N. Sebe, E. Ricci, Contextual Outpainting with Object-Level Contrastive Learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2022, pp. 9254–9263.

[25] H. Yu, S. Lu, C. Liu, J. Hu, F. Lin, Shadow-Enlightened Image Outpainting, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2024, pp. 7850–7860.

[26] X. Yao, M. Sinclair, J.A. Noble, SynStitch: A Self-Supervised Learning Network for Ultrasound Image Stitching Using Synthetic Training Pairs and Indirect Supervision, in: Proceedings of the Medical Imaging with Deep Learning Conference, MIDL, 2023, pp. 1–15.

[27] X. Yao, Z. Cui, Y. Hu, A. Noble, LOTUS: Latent Outpainting Diffusion Model for Three-Dimensional Ultrasound Stitching, in: Proceedings of the Medical Image Deep Learning Conference, MIDL, 2025, in press.

[28] M.E. Liman, D. Rueckert, F.J. Fintelmann, P. Müller, Diffusion-based generative image outpainting for recovery of fov-truncated ct images, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2024, pp. 14–23.

[29] J. Vasudevan, X. Li, C. Wang, T.R. Torgrimson, E.A. Vitriol, From qualitative data to correlation using deep generative networks: Demonstrating the relation of nuclear position with the arrangement of actin filaments, PLoS One 17 (7) (2022) e0271056.

[30] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer, 2015, pp. 234–241.

[31] S. Leclerc, E. Smistad, J. Pedrosa, A. Østvik, F. Cervenansky, F. Espinosa, T. Espeland, E.A.R. Berg, P.-M. Jodoin, T. Grenier, C. Lartizien, J. D'hooge, L. Lovstakken, O. Bernard, Deep learning for segmentation using an open Large-Scale Dataset in 2D echocardiography, IEEE Trans. Med. Imaging 38 (9) (2019) 2198–2210.

[32] D. Ouyang, B. He, A. Ghorbani, M.P. Lungren, E.A. Ashley, D.H. Liang, J.Y. Zou, Echonet-dynamic: a large new cardiac motion video data resource for medical machine learning, in: NeurIPS ML4H Workshop, 2019, pp. 1–11.

[33] D. Dowson, B. Landau, The Fréchet distance between multivariate normal distributions, J. Multivariate Anal. 12 (3) (1982) 450–455.

[34] R. Zhang, P. Isola, A.A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep features as a perceptual metric, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2018.

[35] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.

[36] A. Benavoli, G. Corani, F. Mangili, Should we really use post-hoc tests based on mean-ranks? J. Mach. Learn. Res. 17 (5) (2016) 1–10.

[37] A. Zaidi, D. Knight, D. Augustine, A. Harkness, D. Oxborough, L. Pearce, S. Robinson, M. Stout, J. Willis, V. Sharma, Echocardiographic assessment of the right heart in adults: a practical guideline from the British Society of Echocardiography, Echo Res. Pr. 7 (2020).

[38] P. Motazedian, J. Marbach, G. Prosperi-Porta, S. Parlow, P. Di Santo, Diagnostic accuracy of point-of-care ultrasound with artificial intelligence-assisted assessment of left ventricular ejection fraction, Npj Digit. Med. 6 (2023).

[39] N.M. Razali, Y.B. Wah, Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, lilliefors and Anderson-Darling tests, J. Stat. Model. Anal. 2 (2011).

[40] R. Shams, R. Hartley, N. Navab, Real-time simulation of medical ultrasound from CT images, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008: 11th International Conference, New York, NY, USA, September 6-10, 2008, Proceedings, Part II 11, Springer, 2008, pp. 734–741.

[41] X. Huang, M.-Y. Liu, S. Belongie, J. Kautz, Multimodal unsupervised image-to-image translation, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 172–189.

[42] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.