Preprints.org

Review

Not peer-reviewed version

Trackerless 3D Freehand Ultrasound Reconstruction: A Review

Chrissy Adriaans , Mark Wijkhuizen , Lennard van Karnenbeek , Freija Geldof , Behdad Dashtbozorg *

Posted Date: 20 August 2024

doi: 10.20944/preprints202408.1319.v1

Keywords: Freehand ultrasound; 3D reconstruction; Deep learning; Inertial measurement units (IMUs)

Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Trackerless 3D Freehand Ultrasound Reconstruction: A Review

Chrissy A. Adriaans [1,2], Mark Wijkhuizen [1], Lennard M. van Karnenbeek [1], Freija Geldof [1], Behdad Dashtbozorg [1*]

[1] Image-Guided Surgery, Department of Surgery, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

[2] Technical Medicine, Faculty of Mechanical, Maritime, and Materials Engineering, Delft University of Technology, Mekelweg 2, 2628 CD, Delft, The Netherlands

* Correspondence: b.dasht.bozorg@nki.nl;

Abstract: Two-dimensional ultrasound (2D US) is commonly used in clinical settings for its cost-effectiveness and non-invasiveness, but it is limited by spatial orientation and operator dependency. Three-dimensional ultrasound (3D US) overcomes these limitations by adding a third dimension and enhancing integration with other imaging modalities. Advances in deep learning (DL) have further propelled the viability of freehand image-based 3D reconstruction, broadening clinical applications in intraoperative and point-of-care (POC) settings. This review evaluates state-of-the-art freehand 3D US reconstruction methods that eliminate the need for external tracking devices, focusing on experimental setups, data acquisition strategies, and reconstruction methodologies. PubMed, Scopus, and IEEE Xplore were searched for studies since 2014 following PRISMA guidelines, excluding those using additional imaging or tracking systems other than inertial measurement units (IMUs). Fourteen eligible studies were analyzed, showing a shift from traditional speckle decorrelation towards DL-based methods, particularly Convolutional Neural Networks (CNNs). Variability in datasets and evaluation methods hindered a comprehensive quantitative comparison, but notable accuracy improvements were observed with IMUs and integration of contextual and temporal information within CNNs. These advancements enhance freehand 3D US reconstruction feasibility, though variability limits definitive conclusions about the most effective methods. Future research should focus on improving precision in complex trajectories and adaptability across clinical scenarios.

Keywords: Freehand ultrasound; 3D reconstruction; Deep learning; Inertial measurement units (IMUs)

1. Introduction

Two-dimensional ultrasound (2D US) has established itself as an invaluable imaging modality in clinical settings due to its low cost, portability, and non-invasive nature. Its real-time visualization capability offers significant utility in intraoperative procedures, providing clinicians with essential anatomical and functional information. Despite these advantages, 2D US is hindered by limitations such as limited spatial orientation, lacking valuable 3D context, and its dependence on the operator's experience and interpretive skills. These constraints underscore the inherent variability and the potential for subjectivity in 2D US [1]. The introduction of three-dimensional ultrasound (3D US) marks a significant advancement, addressing many of these limitations. Specifically, 3D US provides a more comprehensive visualization of anatomical regions of interest (ROIs) and provides the flexibility of examining the acquired data from multiple viewpoints post-acquisition. This allows for more accurate quantitative measurements and additionally reduces the variability due to operators. Moreover, 3D US allows for integration with other 3D imaging modalities, such as MRI or CT. This enables a multifaceted view of the patient's anatomy, enriching the diagnostic process. With these advancements, 3D US extends the potential of 2D US across a broader spectrum of clinical applications, like surgical planning, monitoring of treatment progression, and a wide range of intraoperative and interventional tasks [2].

© 2024 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

The path to realizing 3D US reconstructions has seen various methodologies, including the use of 3D US transducers, external tracking, and image-based reconstruction methods. 3D US transducers employ 2D crystal arrays rather than traditional 1D crystal arrays. This design facilitates the electronic steering and focusing of the US beam in multiple dimensions, thereby enabling the real-time acquisition of 3D volumetric data. However, since the transducer elements are spread across two dimensions, this configuration results in fewer elements in each dimension. Therefore, to maintain a sufficient spatial resolution the imaging system must limit the coverage area, resulting in a smaller field of view (FOV), particularly at greater depths. External tracking devices such as mechanical, optical, or electromagnetic (EM) systems are designed to position the US probe in 3D space. These systems are prone to artifacts caused by magnetic interference or optical occlusion and have constraints in the range of motion. Considering the high costs and cumbersome set-up of 3D US transducers and external tracking systems, their suitability for certain clinical applications like intraoperative and point-of-care US (POCUS) is diminished [3,4].

Image-based freehand 3D US represents a promising alternative, eliminating the need for both external tracking devices and 3D transducers, thereby capitalizing on the potential for integration with cost-effective, portable and sometimes wireless, US devices [5]. The computational challenge of these systems lies in the reconstruction process, which necessitates an accurate estimation of the probe's trajectory, divisible into in-plane and out-of-plane (elevational) movements. While in-plane movements are more readily quantified, out-of-plane estimations remain complex. Prior research on out-of-plane motion dates back to seminal work by Chen et al. [6] and has been mainly based on speckle noise, the granular gray-scale textures in B-mode US images. Speckle decorrelation methods map the transformation between neighboring US images to the correlation of their speckle patterns, i.e. the higher the speckle correlation, the lower the elevational distance between neighboring frames [7]. Under Rayleigh scattering conditions, where the size of these scatters is much smaller than the wavelength of the sound waves, this speckle pattern is theoretically predictable. However, when applied to dynamic clinical scenarios, these models often fall short, as speckle variability and real tissue movement introduce errors that accumulate, leading to drift and a compromise in accuracy [8–10].

Recent advancements in artificial intelligence (AI) and especially deep learning (DL) techniques have expanded the horizons for extracting detailed information from image data [11]. This has enabled significant progress in overcoming the intrinsic difficulties of image-based 3D US reconstruction. While traditional tracking technologies that require external reference are less suited to point-of-care and intraoperative settings, inertial measurement units (IMUs) have emerged as a viable and less obtrusive alternative for position tracking. These devices integrate a tri-directional magnetometer, a gyroscope, and an accelerometer into compact units. This offers a favorable balance between hardware independence and the need for positional information in trajectory reconstruction. With the growing adoption of IMU technology in clinical devices and developments in DL techniques, image-based reconstruction methods without external tracking are becoming increasingly viable. This approach additionally aligns with the ongoing shift towards more accessible and efficient POCUS applications.

Consequently, the aim of this systematic review is to identify current state-of-the-art methods for freehand 3D US reconstruction that forgo the need for external tracking devices. This review will critically assess various experimental setups, data acquisition strategies, reconstruction methodologies, and their outcomes to define the state-of-the-art in this rapidly advancing field.

2. Materials and Methods

2.1. Literature Search

2.1.1. Search Strategy

A systematic review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [12]. A literature search was carried out across the

PubMed, Scopus, and IEEE Xplore databases on February 18th, 2024. Only studies published from 2014 onwards were taken into account to ensure the focus remained on advanced reconstruction methods, particularly those leveraging more sophisticated machine learning (ML) and DL algorithms instead of rule-based techniques [2,13]. The search string encompassed the main concepts of "freehand", "ultrasound", "three-dimensional", and "Artificial Intelligence", supplemented with synonyms and related keywords. The detailed search string employed for all databases can be found in Appendix A. An additional snowballing method was utilized, examining the references of the selected papers. All duplicates were removed during the search process.

2.1.2. Eligibility Criteria

Studies were considered eligible if they met the following inclusion criteria: (I) focus must be on 3D US reconstructions, encompassing trajectory reconstruction of the US probe, volume reconstructions for regions of interest (ROIs), or both, (II) data employed for 3D reconstruction must be derived from 2D US, (III) 2D US must be freehand or handheld acquired, thereby excluding the use external tracking but permitting IMU.

Studies were excluded with the following criteria: (I) lack of methodological descriptions for the 3D reconstruction process or those that do not specify the tools or software utilized, (II) employing additional imaging modalities or external tracking as input data for reconstruction, with the exception of IMU, (III) preliminary publications conducted by the same research team as subsequent included publications presenting only minor enhancements, such as hyperparameter adjustments or minor dataset expansions, (IV) full text is inaccessible, and (V) full text not written in English. If studies presented multiple datasets or methods, only the results of the datasets and methods aligned with the eligibility criteria were considered.

2.1.3. Study Selection Process

Following the literature search, all identified studies were initially screened based on their titles and abstracts. Papers that did not adhere to the eligibility criteria were excluded. The full text was retrieved for the remaining papers and screened for eligibility based on complete texts. Papers identified through snowballing were similarly assessed. Each assessment was meticulously carried out by a single reviewer. The included studies were categorized based on the focus of 3D reconstruction: estimating the complete 3D trajectory of the US probe, reconstructing a volume within an ROI without tracking the entire sweep trajectory, or addressing both objectives.

2.2. Data Extraction

2.2.1. Study Characteristics

Data extraction of study characteristics from the included articles focused on three main aspects: datasets and acquisition, reconstruction methods, and their evaluation by outcome measures. For the datasets employed, details regarding the clinical anatomy, as well as the number of sequences, frames, and subjects, were extracted. Furthermore, more detailed sweep characteristics related to the included sequences in the datasets were extracted, such as the motion trajectory, length of the sweep, type of US probe, and available acquisition parameters. Additionally, the methodology used to establish ground truth data was extracted. For the 3D reconstruction methods, information regarding the utilized data inputs, preprocessing methods, network architecture and components for trajectory reconstruction or volume reconstruction, and loss functions was extracted. Lastly, information regarding the robustness of the method was extracted.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

2.2.2. Quantitative Outcome Measures Of 3D Reconstruction Methods

For each study, relevant quantitative outcome measures were extracted for the proposed reconstruction method and important ablation experiments. These ablation experiments involve systematically removing or modifying components of a model to evaluate their impact on performance, thereby elucidating the contribution of individual elements within the model. Evaluation of the reconstruction methods entailed extraction of trajectory reconstruction errors, volume reconstruction errors, and the computation time. If none of the following described outcome measures were reported within a study, the main outcome measure deemed relevant was extracted.

Trajectory reconstruction errors

Evaluation of trajectory errors involved two principal groups of metrics:

1) Error between successive frames: The reconstruction of the trajectory relies on the prediction of transformation parameters between successive frames. Consequently, the most direct metric to gauge model performance is the Mean Absolute Error (MAE) between predicted and ground truth transformation parameters across a sequence of N number of frames, as outlined in Equation 1:

$$\mathrm{MAE}(t_x, t_y, t_z, \theta_x, \theta_y, \theta_z) = \frac{1}{N}\sum_{k=1}^{n}\left|\vec{p}\left(T_k(i \to j)^{\mathrm{pred}}, T_k(i \to j)^{\mathrm{GT}}\right)\right| \tag{1}$$

Here, $|\vec{p}(A)|$ computes the element-wise absolute difference between each corresponding parameter of the predicted $T_k(i \to j)^{\mathrm{pred}}$ and ground truth $T_k(i \to j)^{\mathrm{GT}}$ transformation matrices for each pair (k) of successive US frames (i → j). This matrix consists of the translation (t_x, t_y, t_z) and rotation (θ_x, θ_y, θ_z) parameters. The MAE was considered parameter-wise for all transformation parameters or in subsets denoting only translation, orientation, or out-of-plane parameters.

In addition to element-wise parameter analysis, error measurement was conducted based on the spatial location of frames denoted as the frame error. This method calculates the Euclidean distance between the predicted and ground truth frame locations, averaged over the four corners of each frame. This approach offers a broader perspective on the positional accuracy of the frame sequence without focusing solely on individual transformation parameters.

2) Drift measures: Another key metric in trajectory reconstruction is the Final Drift (FD), defined by the Euclidean distance between the positions of the center of the last frame of the estimated trajectory and the last frame of the ground truth trajectory. Since the prediction of the trajectory of the sweep is based on a transformation between successive frames, errors accumulate over the length of the sweep. This accumulated error is reflected in the metric FD. Additionally, the Final Drift Rate (FDR) (%) was extracted which is the FD normalized by the ground truth length of the sweep. Important to note is that a minimal FD does not indicate a satisfactory reconstruction for complex scan strategies such as a loop scan, requiring additional evaluation metrics.

Volume reconstruction errors

For volume reconstruction errors, quantitative metrics such as Dice overlap and the absolute difference in volume, in milliliter, were extracted. These measures assess the degree of overlap and volumetric accuracy between the reconstructed volume and the ground truth, providing a robust framework for evaluating the performance of 3D reconstruction algorithms.

2.2.3. Analysis of Ablation Experiments

Analyses were conducted on two optimization strategies for reconstruction methods to evaluate the impact on outcome measures by ablation experiments. This included one subgroup that focused on optimizing models incorporating multiple data types, such as IMU alongside US frames. Another subgroup aimed to incorporate additional contextual or temporal information by analyzing more extensive US sequences rather than limiting them to just two adjacent frames. The relevant outcome

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

measures previously discussed were evaluated before and after incorporation, quantifying the improvement in terms of percentage.

3. Results

3.1. Study Selection

An initial search across PubMed, IEEE Xplore, and Scopus identified 79 potentially relevant publications. After removing 21 duplicates, eligibility of the remaining 58 articles was assessed based on our predefined inclusion and exclusion criteria. Of these, 29 were excluded after title and abstract screening and 19 were excluded after full-text review, leaving 10 articles for inclusion. Additionally, four studies were identified through snowballing after eligibility assessment, culminating in a total of 14 reports that were included for data extraction. The study selection process is shown in Figure 1.

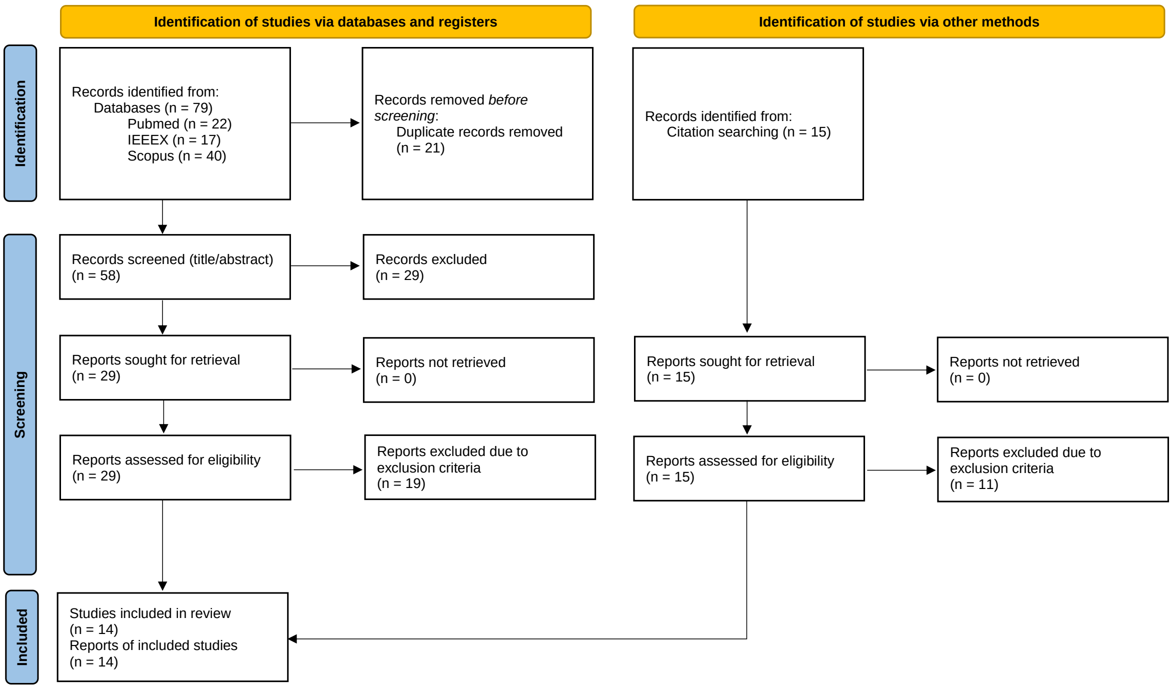
Identification of studies via databases and registers
Identification of studies via other methods
Identification
Screening
Included
Records identified from:
Databases (n = 79)
Pubmed (n = 22)
IEEEX (n = 17)
Scopus (n = 40)
Records removed before screening:
Duplicate records removed
(n = 21)
Records identified from:
Citation searching (n = 15)
Records screened (title/abstract)
(n = 58)
Records excluded
(n = 29)
Reports sought for retrieval
(n = 29)
Reports not retrieved
(n = 0)
Reports sought for retrieval
(n = 15)
Reports not retrieved
(n = 0)
Reports assessed for eligibility
(n = 29)
Reports excluded due to exclusion criteria
(n = 19)
Reports assessed for eligibility
(n = 15)
Reports excluded due to exclusion criteria
(n = 11)
Studies included in review
(n = 14)
Reports of included studies
(n = 14)

Figure 1. Study selection process visualized by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart [12].

3.2. Study Characteristics

Table 1 presents a comprehensive summary of the characteristics extracted concerning the objectives, datasets, acquisition techniques, and reconstruction methods utilized in the studies. The majority of the studies exclusively focus on the evaluation of the full 3D US trajectory reconstruction; despite the ultimate aim being volume reconstruction of a ROI within the trajectory [14–22]. Two studies evaluated 3D trajectory reconstruction and additionally visualized the ROI, specifically arteries, within that reconstruction, albeit lacking quantitative assessments of the volume reconstruction [23,24]. Two others quantitatively assessed both trajectory and ROI volume reconstructions, specifically targeting the thyroid and prostate [25,26]. A singular study focused on directly estimating the volume of the cerebral ventricle system (CVS), bypassing trajectory reconstruction [27].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

Table 1. Summary of the extracted study characteristics of papers included through the study selection process

Study	Aim	Dataset (n=sequences)	Sweep characteristics	US Probe	Data inputs	Reconstruction methods	Loss function	Ground truth
Tetrel et al. [14]	Trajectory reconstruction with speckle decorrelation and graph-based optimization	**Phantom:** **1)** Speckle phantom (n=1) **2)** Phantom (n=9)	**1)** Motorized translation (50 mm), precision of 5 µm **2)** Elevational displacement (20-56 mm)	ATL HDI5000 US scanner, linear 4–7 MHz probe, depth 3 cm	US sequence and phantom based speckle decorrelation curve	GMeA: graph-based trajectory estimation based on speckle decorrelation, optimized by weighted graph edges by learned error of tracked sequence using Gaussian process regressor	N/A	Micron Tracker optical sensor, 97% accuracy volume measurements
Martin et al. [27]	CVS segmentation after landmark based 3D reconstruction of 2D US images	**In vivo:** **1)** Cerebral ventricle (n=15, 14 subjects)	Angular sweep, 136-306 frames per sequence	Siemens Acuson 9L4 probe	2 freehand 2D US sequences in coronal and sagittal orientation	**Reconstruction:** landmark based registration of coronal and sagittal views, model parameters optimization, mapped to 3D grid, voxel-based volume interpolation. **Segmentation:** CNN U-Net	**Reconstruction:** Gradient descent **Segmentation:** Soft-Dice loss	**Segmentation:** manual CVS segmentation in sagittal view
Prevost et al. [23]	Trajectory reconstruction incorporating IMU	**Phantoms:** **1)** BluePhantom US biopsy (n=20, 7168 frames) **In-vivo:** **2)** Forearms (n=88, 41869 frames, 12 subjects) **3)** Calves (n=12, 6647 frames) **4)** Forearms + IMU (n=600, 307200 frames, 15 subjects) **5)** Carotids + IMU (n=100, 21.945 frames, 10 subjects)	**1)** Basic, average length 131 mm **2)** Basic, average length 190 mm **3)** Basic, average length 175 mm **4)** Basic, shift, wave, tilt, average length 202 mm **5)** Basic, tilt, average length 75 mm	Cicada research US machine, linear probe, 128 elements, 5 MHz, 35 FPS	Pairs adjacent 2D US frames, optical flow, IMU orientation data	CNN (ablation experiments with input channels optical flow, IMU and θ based on CNN or IMU)	MSE loss	Optical tracking system Stryker NAV3 Camera, translation accuracy 0.2
Balakrishnan et al. [15]	Trajectory reconstruction with a texture-based similarity metric	**In vivo:** **1)** Forearm (n=13, 7503 frames, 3 subjects)	Varying acquisition speeds, forearm sizes and shapes, axial resolution 400x457	GE Logiq E Ultrasound System, 9L-RS linear probe	US, optical flow, texture-based similarity values	Gaussian (SVM) based regression model, similarity metric TexSimAR	N/A	EM Tracking System Ascension trakSTAR

Continued on next page

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

Table 1. (continued)

Study	Aim	Dataset (n=sequences)	Sweep characteristics	US Probe	Data inputs	Reconstruction methods	Loss function	Ground truth
Miura et al. [17]	Trajectory reconstruction incorporating motion features and consistency loss	**In vivo:** **1)** Forearm (n=190, 30801 frames, 5 subjects) **Phantom:** **2)** Breast phantom (n=60, 8940 frames) **3)** Hypogastric phantom (n=40, 6242 frames)	**1, 2, 3)** Sweeps of 6 seconds	**1, 2)** SONIMAGE HS1, L18-4 linear probe **3)** C5-2 convex probe, 30 FPS	Pairs adjacent 2D US frames, optical flow	CNN (ResNet45) + FlownetS (optical flow), with ablation experiments adding FlowNetS (motion features) and consistency loss	MSE loss, consistency loss	Optical tracking V120: Trio OptiTrack, with 5 markers attached on US probe
Wein et al. [25]	Trajectory reconstruction and thyroid volume segmentation	**In-vivo:** **1)** Thyroid (n=180, 9 subjects).	Variations in acquisition speed, captured anatomy and tilt angles	Cicada research US machine, linear, 128 elements, 5 MHz	Pairs adjacent 2D US frames and optical flow, 1 transverse (TRX) and 1 sagittal (SAG) direction	**Segmentation:** 2D U-Net + union of segmentations in SAG+TRX. **Trajectory:** CNN, optimized by joint sweep reconstruction through co-registration of orthogonal sweeps	MSE loss	**3D:** Based on 2D U-Net, dice 0.73 **Trajectory:** optical tracking system Stryker NAV3 Camera, translation accuracy 0.2
Guo et al. [16]	Trajectory reconstruction with contextual learning	**In vivo:** **1)** Transrectal US (n=640, 640 subjects)[a]	Axial images, steadily through prostate from base to apex	Philips iU22 scanner in varied lengths, end firing C95 transrectal US probe	Transrectal US, N-neighboring frames	Deep contextual learning network (DCL-Net) (3D ResNext) with self-attention module focusing on speckle-rich areas[b]	MSE loss + case-wise correlation loss	EM tracking (mean over N neighboring frames)
Guo et al. [18]	Trajectory reconstruction with two US transducers by domain adaptation techniques	**In vivo:** **1)** Transrectal US for training (n=640, 640 subjects) **2)** Transabdominal US (n=12, 12 subjects)	**1)** Axial images, steadily sweeping through the prostate from base to apex **2)** N/A	**1)** End-firing C95 transrectal US probe **2)** C51 US probe	Videosubsequence of transrectal and transabdominal US	CNN with novel paired-sampling strategy to transfer task specific feature learning from source (transrectal) to target (transabdominal) domain	MSE loss, discrepancy loss (L2 norm between feature outputs of generators of both domains)	EM tracking (mean over N neighboring frames)
Leblanc et al. [24]	3D stretched reconstruction of femoral artery, DL-based	**In vivo:** **1)** Femoral artery (n=111, 40788 frames, 18 subjects)	Thigh to knee following femoral artery, lengths 102-272 mm	Aixplorer echograph, Supersonic Imagine	Pair of 2D US frames, after echograph processing with speckle reduction	In-plane by registration of mask R-CNN based artery segmentation and interpolation, out-of-plane by CNN and linear interpolation to generate final volume	MAE	Optical tracking, NDI Polaris Spectra. Segmentation 2D: manual

Continued on next page

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

Table 1. (continued)

Study	Aim	Dataset (n=sequences)	Sweep characteristics	US Probe	Data inputs	Reconstruction methods	Loss function	Ground truth
Luo et al. [19]	Trajectory reconstruction with IMU and online learning, contextual information	**In vivo:** **1)** Arm (n=250, 41 subjects) **2)** Carotid (n=160, 40 subjects)	**1)** Linear, curved, fast and slow, loop, average length 94.83 mm **2)** linear, average length 53.71 mm	Linear probe, 10 MHz, image depth 3.5 cm	N-neighboring frames US sweep, IMU	MoNet: BK (ResNet + LSTM) + IMU + online learning (adaptive optimization self-supervised by weak IMU labels) with ablation experiments for IMU and online learning	BK: MAE + Pearson correlation loss, online learning: MAE(°) + Pearson correlation loss (acceleration)	EM tracking, resolution of 1.4 mm positioning and 0.5° orientation
Luo et al. [21]	Trajectory reconstruction using 4 IMUs, online learning and contextual information	**In vivo:** **1)** Arm (n=288, 36 subjects) **2)** Carotid (n=216, 36 subjects)	**1)** Linear, curved, loop, sector scan, average length of 323.96 mm **2)** Linear, loop, sector scan, average length of 203.25 mm	Linear probe, 10 MHz, image depth 4 cm	N-length US sweep, 4 IMUs	OSCNet: BK (ResNet + LSTM + IMU) + online learning on modal-level self-supervised (MSS) by weak IMU labels and sequence-level self-consistency strategy (SCS)	BK: MAE + Pearson correlation loss, Online learning: MAE(°) + Pearson correlation loss (acceleration) for single- and multi-IMU, self-consistency loss	EM tracking, resolution of 1.4 mm positioning and 0.5° orientation
Luo et al. [20]	Trajectory reconstruction with online learning and contextual information	**In vivo:** **1)** Spine (n=68, 23 subjects) dataset on a robotic arm with EM positioning	Linear, average length of 186 mm	N/A	US images, canny edge maps and optical flow of 2 adjacent frames	RecON: BK (ResNet + LSTM) + online Self Supervised Learning (SSL): Frame-level Contextual Consistency (FCC), Path-level Similarity Constraint (PSC) and Global Adversarial Shape prior (GAS)	BK: MAE + Pearson correlation loss + motion-weighted training loss. Online learning: MAE + Pearson correlation loss, adversarial loss	Volume based on EM and mechanical tracking
Guo et al. [26]	Trajectory reconstruction utilizing contextual information and volume segmentation prostate	**In vivo:** **1)** Transrectal US (n=618, 618 subjects) + segmentations **2)** Transabdominal (n=100)	**1)** Axial images, steady sweep through prostate from base to apex **2)** Varying lengths and resolutions	**1)** Philips C9-5 transrectal US probe **2)** Philips mc7-2 US probe	Video subsequence of transrectal and transperinal US, n=7 frames	Deep contextual-contrastive network (DC2-Net) (3D ResNext) with self-attention module to focus on speckle-rich areas and a contrastive feature learning strategy	MSE loss, case wise correlation loss and margin ranking loss for contrastive feature learning	EM tracking (mean over N neighboring frames)

Continued on next page

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

Table 1. (continued)

Study	Aim	Dataset (n=sequences)	Sweep characteristics	US Probe	Data inputs	Reconstruction methods	Loss function	Ground truth
Li et al. [22]	Trajectory reconstruction incorporating long-term temporal information	**In vivo:** **1)** Forearm (n=228 19 subjects)[a]	Left and right arms, straight, c-shape, and s-shape, distal-to-proximal, 36-430 frames/sequence, 20 fps, lengths 100-200 mm	Ultrasonix machine, curvilinear probe (4DC7-3/40), 6 MHz, depth 9 cm	US sequence with number M of past (i) and future (j) frames, subsets of trainset to evaluate anatomical and protocol dependency	**1)** RNN + LSTM **2)** Feedforward CNN (EfficientNet) leveraging multi-task learning framework designed to exploit long-term dependencies[b]	Multi-transformation loss based on MSE (consistency at multiple frame intervals)	Optical tracking, NDI Polaris Vicra

Note: Included studies in the table are ordered by publication date, [a] Open-source dataset, [b] Open-source code, *CNN* = Convolutional Neural Network, *MSE* = Mean Squared Error, *BK* = Backbone

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

3.2.1. Data Acquisition And Datasets

From the selected studies, a total of 23 datasets adhered to the eligibility criteria of 2D images and were considered for further evaluation. It is important to note that due to the challenges posed by speckle decorrelation theory in real tissue, some studies included phantom datasets [14,17,23], while the remainder of the studies employed clinical datasets. As presented in Figure 2, the clinical datasets show a variety of different anatomies, paving the way for diverse clinical applications. Across the datasets included, there is a large range in the number of sequences utilized (from 9 to 619, μ=194.08 and σ=213.29), which is also applicable to the number of frames and subjects. This directly influences the data volume available for model training and testing and the anatomical variety between sequences. Seventeen datasets reported on the trajectory motions in the sweep, with seven describing basic linear motions. Four datasets describe more complex sweeps encompassing multiple motion types (like wave, tilt, or shift) [19,21–23]. Additionally, variations in speed and sudden movements were occasionally noted. Another important dataset characteristic concerns the length of the sweeps, which was reported for 13 out of the 23 datasets. These datasets provided either a range or an average of sweep lengths, spanning from 20 mm to 323.96 mm (μ=154.60 and σ=77.06).

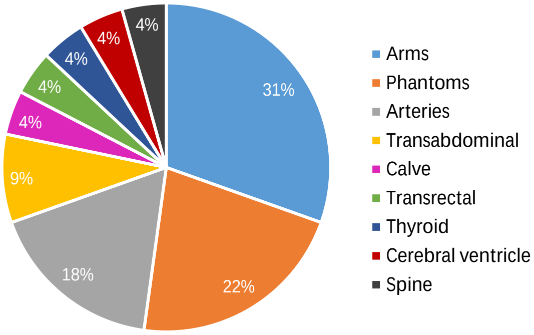

Figure 2. Overview of the distribution of utilized dataset types across the included studies.

3.2.2. Reconstruction Methods

Preprocessing

Minimal preprocessing of US images was undertaken, primarily involving cropping and resizing to suit the specific requirements of the employed network architectures. Prevost et al. [23] conducted experiments with various resolutions and identified an optimal resolution with a pixel size of 0.3 mm, which best accommodates the scale of the speckle pattern. Larger pixel sizes could obscure relevant speckle patterns, whereas smaller pixel sizes might amplify the impact of electronic noise and network demands. Leblanc et al. [24] found that a pixel size of 0.15 mm was optimal for efficient network training, noting that higher resolutions resulted in larger input frames which constrained batch sizes, while lower resolutions reduced precision. These images were obtained post-echograph processing, suggesting that speckle filtering may have occurred; similarly for Li Qi et al. [22] who applied median speckle reduction. Two studies reported no speckle reduction or scanline conversion was employed [17,23], due to a degradation performance using filtered B-mode images [23]. Augmentation strategies employed in four studies included horizontal mirroring, pairing non-consecutive images to enhance robustness against speed variations, and techniques like subsequence intercepting, interval sampling, and sequence inversion [19,21,23,24].

Network architectures

Various methodologies were employed across the studies for 3D US reconstruction. Earlier studies often used traditional modeling approaches. For instance, Tetrel et al. [14] utilized a graph-based trajectory estimation model based on speckle phantom decorrelation curves. This was optimized by

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

modeling measurement errors from an optical tracked sequence through Gaussian process regression, a ML technique that predicts the relationship between the observed motion measurements and their associated errors. This allows for more accurate weighting of the measurements in the graph, improving the overall trajectory estimation. Martin et al. [27] employed a hybrid technique focusing on segmentation of the CVS, combining landmark-based registration of coronal and sagittal sequences with DL-based segmentation. Another study introduced a ML model, a Gaussian SVM regressor, equipped with a novel texture-based similarity metric to capture the dynamic textures of US images for estimating out-of-plane motion [15]. This metric essentially aims to train on speckle similarities. The remaining majority, eleven studies, have adopted DL approaches reflecting more state-of-the-art techniques. Predominantly, feed-forward convolutional neural networks (CNNs) were employed. These included 2D CNNs, like ResNet and EfficientNet [22], for processing 2D US images, and 3D CNNs, such as ResNext [16,26], for handling volumetric data. Some studies further enhanced these architectures by integrating Long Short-Term Memory (LSTM) networks to capture temporal dependencies within sequences, thus learning from sequential dynamics.

Among the studies using CNNs, there was a diverse array of approaches to achieve correct trajectory estimation. For example, Wein et al. [25] utilized two perpendicular US sweeps by co-registration of separate predicted trajectories by a CNN. Combining information and redundancy in overlapping data was exploited to refine the trajectory for a better volume estimation, since utilization of 1 sweep resulted in a volume error of 45%. Le Blanc et al. [24] utilized the CNN merely for predicting out-of-plane motion, while in-plane registration was performed by aligning segmentation masks of the artery. Guo et al. [18] specifically focused on the adaption of the network from a transrectal to transabdominal transducer. In later works, Guo et al. [16,26] introduced a self-attention module that focuses on speckle-rich areas, by utilizing feature maps from the last residual block to produce an attention map highlighting the most informative regions. Lastly, two groups of studies employed comparable methodologies within the CNNs to enhance baseline models, which warrant further evaluation. The first approach involves integrating multiple different input channels in addition to US images, such as optical flow (OF) vector fields and IMU data. The second approach focuses on incorporating more contextual information by analyzing US sequences on a global level, taking into account information from multiple consecutive frames rather than focusing solely on pairs of adjacent frames at a local level. These strategies were assessed through experimental analysis.

Loss functions

Most studies employed Mean Squared Error (MSE) loss over a sequence of N number of frames, as defined in Equation (2).

$$\mathrm{MSE}(t_x, t_y, t_z, \theta_x, \theta_y, \theta_z) = \frac{1}{N} \sum_{k=1}^{n} \left(\vec{p} \left(T_k(i \to j)^{\mathrm{pred}}, T_k(i \to j)^{\mathrm{GT}} \right) \right)^2 \tag{2}$$

In this equation, the matrix $(\vec{p}(A))^2$ computes the element-wise squared difference between each corresponding translation and rotation parameter of the predicted $T_k(i \to j)^{\mathrm{pred}}$ and ground truth $T_k(i \to j)^{\mathrm{GT}}$ transformation matrices for each pair (k) of successive US frames (i → j). The MSE penalizes larger errors more heavily and is therefore valued for its reduced sensitivity to noise during minor frame-to-frame transformations, compared to the Mean Absolute Error (MAE).

However, relying solely on the MSE [23,25] tends to result in the model memorizing general movement patterns of the clinician's probe rather than detecting subtle nuances or variations. To address this issue, several studies incorporated a correlation-based loss in addition to MSE, focusing on the correlation among each transformation parameter (6 DOF) rather than the mean trajectory. Specifically, a case-wise-correlation loss ($\mathrm{L_{cc}}$) as outlined in Equation (3) was adopted [16,23,25,26], which is based on a Pearson correlation loss, sometimes combined with the MAE [19–21].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

$$\mathrm{L}_{\mathrm{cc}}(t_x, t_y, t_z, \theta_x, \theta_y, \theta_z) = \frac{1}{N} \sum_{k=1}^{n} \frac{Cov\left(T_k(i \to j)^{\mathrm{pred}}, T_k(i \to j)^{\mathrm{GT}}\right)}{\sigma\left(T_k(i \to j)^{\mathrm{pred}}\right) \sigma\left(T_k(i \to j)^{\mathrm{GT}}\right)} \tag{3}$$

Here, the correlation coefficients are quantified for each DOF through the covariance (Cov) of the predicted and ground truth transformation parameter normalized by the product of their standard deviation (σ). The case-wise correlation loss stabilizes training and smooths trajectories by emphasizing global correlations among transformation parameters. By preventing overfitting to common scanning styles it enhances generalization, increasing the sensitivity of the model for varying speeds and motions. Further refinements include the introduction of a motion-weighted regularization to the Lcc and MSE, adjusting regularization strength based on the motion speed to address discrepancies more effectively in slower movements [20].

Additional loss functions were based on consistency to further increase robustness; Miura et al. [17] was the first to introduce a consistency loss across two US frames, employing techniques such as flipping and inverse prediction to ensure consistent output regardless of input orientation changes. Luo et al. [21] expanded this concept at a sequence level, creating flipped subsequences and utilized this as a self-consistency strategy (SCS) through online learning and adaptive optimization of the model. Luo et al. [20] applied similar strategies, specifically adopting frame-level contextual consistency (FCC) and path-level similarity constraints (PSC). Li et al. [22] developed a multi-transformation loss based on MSE to ensure consistency between predictions across multiple frame intervals. Lastly, Guo et al. [26] implemented a Margin loss, rooted in contrastive feature learning, which enhances the model's discriminative power by forcing the feature generator to produce similar features for frames with similar transformation parameters, and distinct representations for samples with discrepancies in their transformation parameters.

3.3. Quantitative Outcomes

A detailed overview of the extracted outcome measures for all assessed methods and datasets is presented in Table 2. Frame-to-frame errors were mainly based on the MAE over all transformation parameters [17,23], a subset defined by the out-of-plane transformation parameters [15,24], or the angular transformations [19–21]. Studies focusing on trajectory reconstruction reported either FD or FDR. Given that FD accumulates over the length of the sweep, FDR is a more objective and comparable metric. This rate was reported in six studies, visually represented by different colors in Figure 3 for all ablation methods per dataset. The best performing methods, highlighted in bold, ranged from 5.2% [23] to 15.67% [19], with an average of 10.95%. Outcomes regarding volume reconstruction were reported in three studies. The direct estimation of the CVS volume resulted in a Dice similarity coefficient of 0.82 ± 0.04 [27]. Following trajectory reconstruction, the volume reconstruction of the prostate yielded a Dice of 0.89 ± 0.06 and a volume error of 3.21 ± 1.93 milliliter [26]. For the thyroid, a volume error of 1.15 ± 0.12 milliliter was obtained [25]. Computation time, discussed in four studies, varied; 5 seconds for CVS volume segmentation [27], 0.85 seconds for stretched trajectory reconstruction of the femoral artery [24], 2.58 seconds for trajectory reconstruction of a 100-frame US sweep [16,26], and 3 minutes for combined trajectory reconstruction and volume reconstruction of the thyroid [25].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

Table 2. Quantitative outcomes for included studies per method, evaluated on dataset (n), as described in Table 1.

Method	Final drift (mm)	FDR (%)	Errors between successive frames	Volume measures
Tetrel et al. [14]	**Mean**			
A: Graph based - MeA	**1.1)** 1.803, **1.2)** 0.731, **1.3)** 3.039, **1.4)** 1.995, **1.5)** 3.735, **1.6)** 9.345, **1.7)** 9.154, **1.8)** 9.999	N/A	N/A	N/A
Martin et al. [27]			**MAD between SAG/TRX (mm)**	**HD (mm), Dice**
A: landmark based – CVS segmentation (U-Net)	N/A	N/A	1.55 ± 1.59	13.6 ± 4.7, 0.82 ± 0.04
Prevost et al. [23]	**Median (min - max)**	**Mean**	**MAE of t_x, t_y, t_z, θ_x, θ_y, θ_z (mm/°)**	
A: Standard CNN	**1)** 26.17 (14.31 - 65.10) **2)** 25.16 (3.72 - 63.26) **3)** 54.72 (27.11 - 116.64)	N/A	**1)** 2.25, 5.67, 14.37, 2.13, 1.86, 0.98 **2)** 6.30, 5.97, 6.15, 2.82, 2.78, 2.40 **3)** 4.91, 8.95, 25.89, 2.01, 2.54, 2.90	N/A
B: CNN (OFa + θ by CNN)	**1)** 18.30 (1.70 - 36.90) **2)** 14.44 (3.35 - 41.93 **3)** 19.69 (8.53 - 30.11) **4)** 27.34 (3.22 - 139.02)	**2)** 19	**1)** 1.32, 2.13, 7.79, 2.32, 1.21, 0.90 **2)** 3.54, 3.05, 4.19, 2.63, 2.52, 1.93 **3)** 3.11, 5.86, 5.63, 2.75, 3.17, 5.24 **4)** 8.89, 6.61, 5.73, 5.21, 7.38, 4.01	N/A
C: CNN (IMU + θ by CNN)	**4)** 29.22 (3.12 - 186.83)	N/A	**4)** 6.56, 7.23, 16.70, 0.94, 2.65, 2.80	N/A
D: CNN (OF + IMU + θ by CNN)	**4)** 15.07 (2.54 - 55.20)	N/A	**4)** 5.16, 2.67, 4.43, 0.96, 3.54, 2.85	N/A
E: CNN (OF + θ by IMU)	**4)** 11.43 (1.33 - 42.94)	N/A	**4)** 2.98 2.57 4.79 0.19 0.21 0.13	N/A
F: CNN (OF + IMU + θ by IMU)	**4)** 10.42 (0.76 - 35.22)	5.2	**4)** 2.75 2.41 4.36 0.19 0.21 0.13	N/A
Balakrishnan et al. [15]	**Median (min - max)**		**MAE of t_z, θ_x, θ_y (mm/°)**, out of plane	
A: Gaussian SVM regressor	6.59 (5.550 - 23.02)	N/A	9.11, 1.95, 1.66	N/A
Miura et al. [17]			**MAE of t_x, t_y, t_z, θ_x, θ_y, θ_z (mm/°)**	
A: CNN (ResNet, OF)	N/A	N/A	0.72, 0.18, 0.76, 0.60, 1.26, 0.52	N/A
B: CNN (ResNet, OF + FlowNetS)	N/A	N/A	0.74, 0.18, 0.78, 0.61, 1.28, 0.52	N/A
C: CNN (ResNet, OF + $L_{consistency}$)	N/A	N/A	0.66, 0.15, 0.82, 0.56, 1.23, 0.47	N/A
D: CNN (ResNet, OF + FlowNetS + $L_{consistency}$)	N/A	N/A	0.64, 0.15, 0.80, 0.53, 1.21, 0.47	N/A
Wein et al. [25]			**Relative trajectory error, mean ± SD:** cumulative in-plane translation/length	**Volume error (ml)**

Continued on next page

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

Table 2. (continued)

Method	Final drift (mm)	FDR (%)	Errors between successive frames	Volume measures
A: CNN + joint co-registration 54-DOF	N/A	N/A	0.16 ± 0.09	1.15 ± 0.12
Guo et al. [26]	**Median (min - max), mean**			
A: DCL-Net (attention, n=5, L_{MSE} + L_{CC})	17.40 (1.09 - 55.50), 17.39	N/A	N/A	N/A
B: DCL-Net (attention, n=2 to n=8, L_{MSE} + L_{CC})	Visualized in boxplot per n input frames, n=5 significant improvement to n=2 ($p<0.05$)	N/A	N/A	N/A
Guo et al. [18]	**Median (min - max), mean**			
A: Target transabdominal	21.21 (5.88 - 32.94), 20.01	N/A	N/A	N/A
B: TAUVR	22.02 (6.87 - 32.13), 20.34	N/A	N/A	N/A
Leblanc et al. [24]	**Median (min - max), mean**	**Median**	**MAE (mm)**, out-of-plane translation	
A: CNN + artery alignment	13.42 (0.18 - 68.31), 17.22	8.98	0.28	N/A
Luo et al. [19]		**Mean ± SD**	**MAE (θ_x, θ_y, θ_z) (°), mean ± SD**	
A: BK (ResNet + LSTM)	N/A	**1)** 16.42 ± 14.24, **2)** 20.55 ± 18.73	**1)** 2.29 ± 2.50, **2)** 2.61 ± 1.72	N/A
B: BK + IMU	N/A	**1)** 14.05 ± 10.36, **2)** 17.78 ± 11.50	**1)** 1.75 ± 1.57, **2)** 2.18 ± 1.43	N/A
C: MoNet (CNN + IMU + Online)	N/A	**1)** 12.75 ± 9.05, **2)** 15.67 ± 8.37	**1)** 1.55 ± 1.46, **2)** 1.50 ± 0.98	N/A
Luo et al. [21]		**Mean ± SD**	**MAE (θ_x, θ_y, θ_z) (°), mean ± SD**	
A: BK (ResNet + LSTM)	N/A	**1)** 13.32 ± 8.2, **2)** 12.85 ± 6.5	**1)** 4.32 ± 1.7, **2)** 3.83 ± 2.0	N/A
B: BK + MSS (4 IMU)	N/A	**1)** 10.78 ± 5.6, **2)** 11.31 ± 5.4	**1)** 3.18 ± 2.76, **2)** 3.16 ± 1.8	N/A
C: BK + SCS ($L_{consistency}$)	N/A	**1)** 10.56 ± 5.9, **2)** 11.30 ± 5.4	**1)** 3.65 ± 1.9, **2)** 3.36 ± 1.8	N/A
D: OSCNet (BK + SCS + MSS)	N/A	**1)** 10.01 ± 5.7, **2)** 10.90 ± 5.3	**1)** 2.76 ± 1.3, **2)** 2.60 ± 1.6	N/A
Luo et al. [20]		**Mean ± SD**	**MAE (θ_x, θ_y, θ_z) (°), mean ± SD**	
A: BK (ResNet + LSTM)	N/A	15.54 ± 8.29	1.36 ± 0.71	N/A
B: BK + OF	N/A	14.88 ± 8.83	1.35 ± 0.46	N/A
C: BK + OF + CE	N/A	12.53 ± 6.32	1.33 ± 0.58	N/A
D: BK + OF + CE + Motion	N/A	12.30 ± 6.31	1.30 ± 0.45	N/A

Continued on next page

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

Table 2. (continued)

Method	**Final drift (mm)**	**FDR (%)**	**Errors between successive frames**	**Volume measures**
E: BK + OF + CE + Motion + SSL(PSC+FCC)	N/A	11.36 ± 5.51	1.30 ± 0.44	N/A
F: ReCon (Motion + SSL + GAS)	N/A	10.82 ± 5.36	1.25 ± 0.46	N/A
Guo et al. [26]	**Mean ± SD**	**Mean ± SD**	**Frame error (mm), mean ± SD** Euclidean distance, 4 corner points	**Dice, volume error (ml) mean ± SD**, (prostate)
A: DC2-Net (attention)	**1)** 12.20 ± 10.07	**1)** 11.66 ± 9.77	**1)** 0.93 ± 0.28	**1)** 0.83 ± 0.08
B: DC2-Net (attention + L_{CC})	**1)** 11.92 ± 8.89	**1)** 11.59 ± 9.22	**1)** 0.93 ± 0.28	**1)** 0.86 ± 0.05
C: DC2-Net (attention + L_{CC} + L_{margin})	**1)** 10.20 ± 8.47, **2)** 9.85 ± 5.74	**1)** 9.64 ± 8.14, **2)** 14.58 ± 12.76	**1)** 0.90 ± 0.26, **2)** 1.12 ± 0.26	**1)** 0.89 ± 0.06, **2)** 3.21 ± 1.93
Li et al. [22]	**Mean ± SD**		**Frame error (mm), mean ± SD** Euclidean distance, 4 corner points	**Dice, mean ± SD** (Trajectory)
A: RNN (M = 2)	34.54 ± 18.10	N/A	0.57 ± 0.44	0.41 ± 0.33
B: RNN (M = 100)	6.97 ± 6.79	N/A	0.20 ± 0.07	0.73 ± 0.23
C: ff-CNN (M = 2)	29.59 ± 19.53	N/A	0.53 ± 0.46	0.50 ± 0.29
D: ff-CNN (M = 100)	7.24 ± 8.33	N/A	0.19 ± 0.08	0.77 ± 0.17
E: ff-CNN (M = 100, straight sweeps)	22.30 ± 41.10	N/A	0.48 ± 0.25	0.64 ± 0.26
F: ff-CNN (M = 100, c- and s-shapes)	6.74 ± 7.19	N/A	0.24 ± 0.13	0.80 ± 0.13
G: ff-CNN (M = 100, 25% of subjects)	13.66 ± 15.94	N/A	0.41 ± 0.24	0.75 ± 0.17

OF = Optical flow, *LSTM* = Long Short-Term Memory, *CE* = Canny Edge, *RNN* = Recurrent Neural Network, *ff-CNN* = feed-forward CNN

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

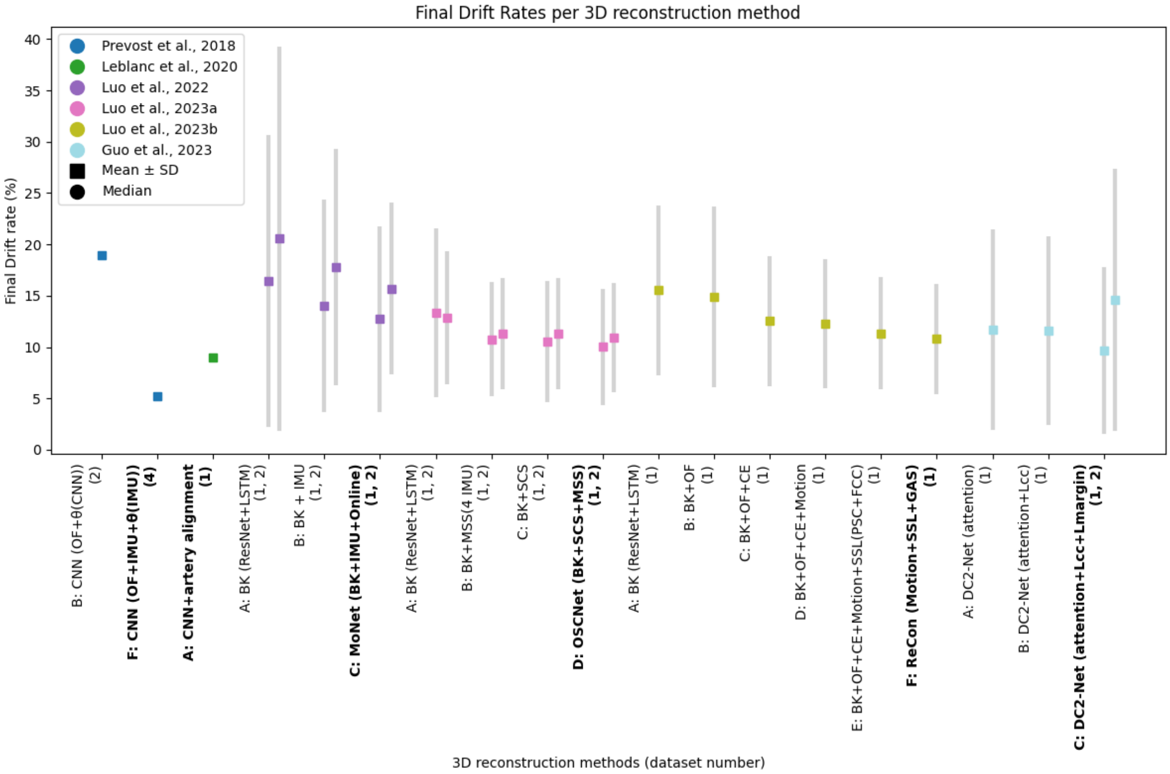

Figure 3. Mean or median Final Drift Rates (%) per reconstruction method across different datasets in the included studies.

3.3.1. Analysis Of Ablation Experiments

Incorporation of multiple data inputs

Analysis of the reconstruction methods indicated a comparable trend among CNNs for optimizing their baseline model through the use of multiple input channels alongside US frames. Optical flow (OF) vector fields were employed by several studies to capture dynamic motion between consecutive frames, thereby providing crucial temporal information about in-plane movements [17,20,23,25]. Luo et al. [20] also integrated Canny edge maps to accentuate anatomical structures, enhancing the model's capability to identify anatomical features and in-plane motions.

Three studies explored the integration of IMU data, which provides orientation and acceleration information. However, after calibration with ground-truth position data, the acceleration typically exhibited high noise levels and was therefore less frequently used. Pioneer in this field was Prevost et al. [23] who integrated the IMU orientation data by experimenting with different integration techniques such as concatenating the Euler rotation angles to the network's penultimate layer, thereby including the IMU orientation data in the training and prediction of the network (Table 2, method D) or directly applying calibrated IMU orientations on top of the CNN trained without IMU (Table 2, method E). Combining these two strategies, i.e. using the CNN with all available input to predict translation and resorting to the IMU orientations for the trajectory estimation, led to a further improvement (Table 2, method F). The pairwise difference between all experiments integrating OF and IMU (methods B-F), was statistically significant. Parameter-wise improvement in MAE yielded by the addition of IMU data is shown in Table 3 with a reduction of final drift by 61.89%.

In addition to the IMU orientation data, Luo et al. [19] first utilized IMU acceleration data despite its high noise. After calibration of IMU with ground truth tracking data, the acceleration showed high noise in absolute values, but a correct trend over a wider range. A temporal and multi-branch structure was employed that effectively utilized the low signal-to-noise ratio (SNR) of IMU acceleration data.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

Additionally, they introduced a multi-modal online self-supervised strategy leveraging IMU data as weak labels to constrain the predicted parameters. In a subsequent study, Luo et al. [21] further tackled the challenges of acceleration noise of an individual sensor by employing four IMUs. This approach integrated a modal-level self-supervised strategy that fused multiple IMU data sources, mitigating the noise effects of one IMU and reducing output discrepancies. The improvement of MAE and reported drift measures of the described approaches is shown in Table 3.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

Table 3. Quantitative improvements of evaluation metrics due to the incorporation of additional inputs in CNNs

Method	Addition of	Dataset	Transformation parameter	Improvement MAE (%)	Improvement drift measures (%)
Prevost et al. (23) B → F	1 IMU	1	t_x	69.07	FD: 61.89
			t_y	63.54	
			t_z	23.91	
			θ_x	96.35	
			θ_y	97.15	
			θ_z	96.76	
Luo et al. [19] A → B	1 IMU	1	$\theta_x, \theta_y, \theta_z$	23.58	FDR: 14.43
		2		42.53	FDR: 13.48
Luo et al. [21] A → B	4 IMUs	1	$\theta_x, \theta_y, \theta_z$	26.39	FDR: 19.07
		2		17.49	FDR: 11.98
Luo et al. [20]	Optical Flow + Canny Edge	1	$\theta_x, \theta_y, \theta_z$	2.21	FDR: 19.37

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

Contextual and temporal information

The enhancement of baseline models through the integration of contextual and temporal information ranged from minor adjustments to substantial architectural changes. For instance, Guo et al. [16,26] transitioned from 2D to 3D CNNs, employing 3D residual blocks and convolutional kernels. This shift allowed for more effective extraction of feature mappings along the axis of the channel, corresponding to the temporal direction of the sweep in this context. This approach showed that a configuration with five frames yielded significant enhancements over two frames [16].

Further integration of contextual and temporal data was achieved through the implementation of various consistency losses within network architectures, previously outlined in section B2. Luo et al. [21] employed a sequence-level self-consistency strategy, while Luo et al. [20] utilized frame-level contextual consistency (FCC) and path-level similarity constraints (PSC). These strategies led to improvements in MAE and FDR, as detailed in Table 4.

More extensive modifications in network architecture were explored by Li et al. [22], who evaluated the use of recurrent neural networks (RNNs) with LSTM modules and feedforward(ff)-CNNs. Those models are known to effectively handle sequential data over long periods. Since no significant difference between these models was found, it was suggested that ff-CNNs are equally competent in modeling US sequences compared to more specialized RNNs. Long-term dependencies were embedded in the network architecture by analyzing sequences of a length M of past (i) and (j) future frames, ranging from 2-100. This methodology allowed the network to capitalize on the consistency of predicting multiple transformations over various frame intervals, significantly enhancing the precision and stability of the US image reconstructions, as evidenced in Table 4.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

Table 4. Quantitative improvements due to the integration of strategies using contextual or temporal information derived from the ultrasound sequence, rather than 2 US frames.

Method	Addition of	Dataset	Transformation parameter	Improvement (%)	Improvement drift measures (%)
Luo et al. [21] A → C	Self Consistency Strategy	1 2	$\theta_x, \theta_y, \theta_z$	MAE: 13.21 MAE: 17.72	FDR: 7.14 FDR: 3.63
Luo et al. [20] D → E	Frame, path and sequence level online learning	1	$\theta_x, \theta_y, \theta_z$	MAE: 3.85	FDR: 4.75
Li et al. [22] C → D	M=2 to M=100 past and future frames	1	N/A	Frame error: 64.15	FD: 75.53

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

3.3.2. Generalization And Robustness Of Methods

Generalization capabilities were assessed by three studies. Prevost et al. [23] used US sweep datasets from forearms and carotids. Networks were trained independently on each dataset, combined, and with fine-tuning of the last two layers or the entire model specifically for the carotid dataset. The findings demonstrated that training exclusively on forearm data produced the least favorable results. Fine-tuning the last two layers for carotid data only enhanced performance, comparably to retraining the network from scratch on solely carotid data. Moreover, extensive fine-tuning of the entire network further improved outcomes, suggesting that initial layers, which handle low-level features, are to some extent anatomy-specific. Training on both datasets involved the greatest amount of data and time but did not significantly surpass the results of comprehensive fine-tuning.

Li et al. [22] evaluated reconstruction performance on the complete original test set with various variance-reduced training sets regarding scanning protocol and anatomical variance. It showed, as depicted in Table 2 Method E-F, that training only on straight sweeps increased the final drift and frame errors extensively. This performance reduction also applies, to a lesser extent, to train sets with a reduction in the number of subjects and thus anatomical variance. Meanwhile, training merely on more complex motions, like c- and s-shape, yielded much less substantial performance losses.

Guo et al. [18] specifically aimed to enhance model adaptability across different US domains, transitioning from a pre-trained network on a large transrectal dataset to a smaller transabdominal dataset. To tailor the model for specific domain characteristics, a paired-sampling strategy was implemented that processed samples from both domains simultaneously during training. This enabled the network to identify domain-invariant features while adapting to nuances specific to the target domain. The fine-tuning phase focused on aligning the learned features more closely with the characteristics of transabdominal scans. The effectiveness of this approach was demonstrated through the improved feature distribution alignment between the domains, indicating robust cross-domain adaptability. Notably, their Targeted Adaptive Ultrasound Volume Reconstruction (TAUVR) approach achieved comparable results to the network trained exclusively on the target domain, as shown in Table 2 Method A and B.

4. Discussion

This systematic review provides an extensive evaluation of advancements in freehand 3D US reconstruction without external tracking, with a focus on DL methodologies that have progressively dominated medical imaging applications. The majority of the studies in this review have implemented DL techniques, predominantly CNNs. These methods have greatly enhanced the precision and usability of 3D US reconstructions, especially in settings involving real tissue. The effectiveness of CNNs can be related back to the speckle decorrelation theory, which formed the basis of previous methodologies. This was substantiated by preprocessing experiments obtaining optimal performance with a resolution correlating to the speckle noise and mitigation of the speckle reduction filter. Additionally, a self-attention module highlighted a focus on speckle-rich areas, indicating their critical role in the model's transformation predictions [16].

Studies using CNNs showed a clear trend toward optimizing baseline models to manage complex motion patterns and anatomical variations. Enhancements were achieved by incorporating more contextual information by adapting network components of architectures and integrating IMU data or channels with optical flow vector fields. These strategies improved the performance of baseline models. Especially the integration of IMUs, providing additional position information, has emerged as a significant advancement. This showed substantial improvements in model accuracy and a reduction in drift errors, exemplified by Prevost's achievement of a state-of-the-art FDR of 5.2% [23]. These enhancements underscore the potential of IMUs to replace traditional, cumbersome tracking systems, although further progress in noise reduction and data fidelity is necessary.

However, this review also identifies several limitations within the included studies impacting outcome measures. There are significant disparities in dataset characteristics, encompassing variations in quality and a variety of clinical applications. These different clinical applications necessitate different relevant outcome measures. For instance, in applications requiring 3D reconstructions for measuring the vessel length, length error is more critical than rotational errors, influencing the choice of metrics used to evaluate reconstruction methods [23]. Additionally, the reported lengths of US sweeps are crucial as they affect metrics such as final drift, which accumulates throughout the sweep. However, lengths are frequently unreported, with only 3 out of 8 papers discussing final drift providing this data, rendering it a subjective measure. Furthermore, the type of motion performed during the sweep also affects model training and generalization capabilities. Simpler motions are easier to learn but might be less representative of clinical scenarios. Complex motions, utilized in only four studies, are more challenging to learn, leading to varied performance across datasets with different anatomy and motion characteristics. This was also observed by Prevost et al. [23], where dataset 4 including more complex sweep motions showed diminished performance compared to dataset 1-3, including basic motion trajectories. These findings are corroborated by generalization experiments where evaluating the methods on different datasets led to decreased performance, indicating dependency on specific protocols, anatomical features, and acquisition parameters. However, the ability to fine-tune or adapt networks to different acquisition parameters or anatomical features using smaller datasets, as demonstrated by Prevost et al. [23] and Guo et al. [18], suggests that retraining from scratch for each new application is generally unnecessary.

Subsequently, due to the diverse range of outcome measures and its dependence on specific dataset characteristics, quantitative outcomes are hard to compare. This makes it challenging to conclusively determine the most effective network architectures. Notably, only four studies have reported computation times, thus providing limited information on the suitability of these methods for real-time applications. Consequently, this review serves more as an overview of existing methods rather than a comparative analysis of quantitative measures for reconstruction methods across studies.

Given the focus of this review on freehand-acquired 2D US images, to facilitate point-of-care use, studies utilizing datasets acquired with 3D US transducers were not considered. These studies sampled the 2D image planes from 3D volumes, to obtain 2D US image data. Nonetheless, knowledge of network fine-tuning suggests that some of these excluded reconstruction methods might have been interesting. Thus, the exclusion of studies using 3D transducers or non-freehand 2D images utilized in reconstruction may have narrowed the range of methodologies evaluated, potentially omitting advanced techniques. As such, the scope and conclusions of this review are confined to directly applicable network architectures for 2D US images used in 3D reconstruction methods.

As previously described, the applicability of the networks is currently constrained to representative datasets utilized in training, influenced by anatomical features, scanning protocols, and acquisition parameters. This is not necessarily considered a significant limitation, as US systems typically include application-specific presets that allow for network fine-tuning or adaptation. However, for further advancements, it is crucial that networks possess the capability to handle more representative sweeps in clinical settings. For example, intraoperative conditions may depend on specific movements where a simple sweep is not always possible. To meet these requirements, it is essential to consider factors such as managing back-and-forth sweeps where models may need to process redundant information. This consideration is vital when the organ of interest does not fit within a single US frame and requires multiple passes. Additionally, many models are direction-dependent, which may result in mirrored images.

Furthermore, assessing the accuracy of trajectory estimation compared to ground truth tracking (either EM or optical) is an initial step toward accurate 3D reconstructions. However, future evaluations of performances should also incorporate more clinical parameters. The focus of the included studies has been predominantly on trajectory estimation and its accuracy, primarily assessed by final drift. While this serves as a preliminary gauge of accuracy, it may not directly correlate to clinical relevance. Therefore, considering the current advancements in trajectory estimation accuracy, future research should aim for more clinical validation tailored to its specific applications. For instance, if the goal is measuring the volume as an important parameter for diagnosis, the Dice coefficient is paramount, even if the final drift error remains suboptimal. Alternatively, evaluating the tracking error of the ROI, rather than the frames within the entire trajectory may offer more clinically relevant insights. These clinically relevant outcomes tailored to the specific application should become more central in the evaluation of the 3D reconstruction methods.

Achievement of high accuracy in the clinical evaluation of 3D US reconstruction methods could have a transformative impact on various medical applications. A model that accurately reconstructs 3D images without relying on external tracking devices or complex setups has the potential to revolutionize fields such as surgical planning and diagnostic imaging. For instance, real-time, detailed 3D visualizations could enhance the precision of surgical procedures, reduce operation times, and improve patient outcomes. In diagnostic imaging, the ability to reconstruct 3D structures from 2D US data allows for quantitative volume measurements, which could lead to earlier detection and better treatment monitoring, e.g. in measurements of the prostate volume, critical for diagnosing and monitoring prostate cancer. This could be achieved in a non-invasive, low-cost, point-of-care setting, making advanced diagnostics more accessible, particularly in primary care environments. Additionally, this technology could democratize access to advanced imaging techniques. As these models continue to evolve, they could seamlessly integrate with other imaging modalities, providing a comprehensive view of patient anatomy and enabling more personalized and effective treatment plans. Therefore, while this review highlights significant advancements in 3D reconstruction methods, the future applications of this technology hold even greater promise for enhancing patient care across diverse clinical contexts.

5. Conclusions

This systematic review has comprehensively evaluated the progress on freehand 3D US reconstruction methodologies that circumvent the need for external tracking devices. This is a dynamically evolving field, particularly influenced by developments in DL. The application of CNNs and the integration of IMUs have notably advanced the capabilities of image-based reconstruction methods compared to conventional speckle decorrelation methods. However, the diversity in study designs and the heterogeneity in datasets present challenges in quantitatively comparing reconstruction methods and deriving definitive conclusions. While there have been significant improvements in trajectory accuracy, clinical validation of these advancements remains important. Ongoing research must address the adaptability of these systems to varied clinical environments and further refine the models to handle more complex US sweeps and anatomical variations. By eliminating the reliance on external tracking and expensive 3D transducers, these advancements provide significant reductions in both cost and complexity while expanding its usability across various medical settings.

Author Contributions: Conceptualization, F.G. and B.D.; methodology, C.A.A., L.M.K., M.W., F.G. and B.D.; software, C.A.A.; validation, C.A.A., L.M.K., M.W., F.G. and B.D.; formal analysis, C.A.A., F.G. and B.D.; investigation, C.A.A. and F.G.; resources, B.D.; data curation, C.A.A. and F.G.; writing—original draft preparation, C.A.A. and F.G.; writing—review and editing, C.A.A., L.M.K., M.W., F.G. and B.D.; visualization, C.A.A. and F.G.; supervision, F.G. and B.D.; project administration, B.D. All authors have read and agreed to the published version of the manuscript.

Funding: Research at the Netherlands Cancer Institute is supported by institutional grants of the Dutch Cancer Society and of the Dutch Ministry of Health, Welfare and Sport.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

Appendix A. Search strings

PubMed:
("3D" OR "three-dimensional" OR "volumetric") AND ("ultrasound" OR "Ultrasonography" [Mesh]) AND ("freehand" OR "handheld") AND ("Artificial Intelligence"[Mesh] OR "deep learning")

IEEE Xplore:
(3D OR three-dimensional OR volumetric OR volume reconstruction) AND (ultrasound) AND (freehand OR handheld) AND (Artificial Intelligence OR deep learning)

Scopus:
("All Metadata":"3D" OR "All Metadata":"three-dimensional" OR "All Metadata":"volumetric" OR "All Metadata":"volume reconstruction") AND ("All Metadata":"ultrasound" OR "Mesh_Terms": "Ultrasonography") AND ("All Metadata":"freehand" OR "All Metadata":"handheld") AND ("All Metadata":"Artificial Intelligence" OR "All Metadata":"deep learning")

References

1. Fenster, A.; Downey, D.B. Three-Dimensional Ultrasound Imaging. *Annual Review of Biomedical Engineering* **2000**, *2*, 457–475. doi:https://doi.org/10.1146/annurev.bioeng.2.1.457.
2. Mozaffari, M.H.; Lee, W.S. Freehand 3-D Ultrasound Imaging: A Systematic Review. *Ultrasound in Medicine & Biology* **2017**, *43*, 2099–2124. doi:10.1016/j.ultrasmedbio.2017.06.009.
3. Lang, A.; Parthasarathy, V.; Jain, A.K. Calibration of EM Sensors for Spatial Tracking of 3 D Ultrasound Probes. In *Data Acquisition Applications*; IntechOpen, 2012; chapter 0.
4. Huang, Q.; Zeng, Z. A Review on Real-Time 3D Ultrasound Imaging Technology. *BioMed Research International* **2017**, *2017*, 6027029. doi:10.1155/2017/6027029.
5. Clarius Website. http://www.clarius.me, 2024.
6. Chen, J.F.; Fowlkes, J.B.; Carson, P.L.; Rubin, J.M. Determination of scan-plane motion using speckle decorrelation: Theoretical considerations and initial test. *International Journal of Imaging Systems and Technology* **1997**, *8*, 38–44. doi:10.1002/(SICI)1098-1098(1997)8:1<38::AID-IMA5>3.0.CO;2-U.
7. Chang, R.F.; Wu, W.J.; Chen, D.R.; Chen, W.M.; Shu, W.; Lee, J.H.; Jeng, L.B. 3-D US frame positioning using speckle decorrelation and image registration. *Ultrasound in Medicine & Biology* **2003**, *29*, 801–812. doi:10.1016/s0301-5629(03)00036-x.
8. Gee, A.H.; James Housden, R.; Hassenpflug, P.; Treece, G.M.; Prager, R.W. Sensorless freehand 3D ultrasound in real tissue: Speckle decorrelation without fully developed speckle. *Medical Image Analysis* **2006**, *10*, 137–149. doi:https://doi.org/10.1016/j.media.2005.08.001.
9. Liang, T.; Yung, L.S.; Yu, W. On Feature Motion Decorrelation in Ultrasound Speckle Tracking. *IEEE Transactions on Medical Imaging* **2013**, *32*, 435–448. doi:10.1109/TMI.2012.2230016.
10. Afsham, N.; Rasoulian, A.; Najafi, M.; Abolmaesumi, P.; Rohling, R. Nonlocal means filter-based speckle tracking. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **2015**, *62*, 1501–1515. doi:10.1109/TUFFC.2015.007134.
11. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. doi:10.1038/nature14539.
12. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; Chou, R.; Glanville, J.; Grimshaw, J.M.; Hróbjartsson, A.; Lalu, M.M.; Li, T.; Loder, E.W.; Mayo-Wilson, E.; McDonald, S.; McGuinness, L.A.; Stewart, L.A.; Thomas, J.; Tricco, A.C.; Welch, V.A.; Whiting, P.; Moher, D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* **2021**, p. n71. doi:10.1136/bmj.n71.
13. Shen, D.; Wu, G.; Suk, H.I. Deep Learning in Medical Image Analysis. *Annual Review of Biomedical Engineering* **2017**, *19*, 221–248. doi:10.1146/annurev-bioeng-071516-044442.
14. Tetrel, L.; Chebrek, H.; Laporte, C. Learning for graph-based sensorless freehand 3D ultrasound. Machine Learning in Medical Imaging (MLMI) 2016, 2016, Vol. 10019 LNCS, pp. 205–212. doi:10.1007/978-3-319-47157-0_25.
15. Balakrishnan, S.; Patel, R.; Illanes, A.; Friebe, M. Novel Similarity Metric for Image-Based Out-Of-Plane Motion Estimation in 3D Ultrasound. Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), 2019, pp. 5739–5742. doi:10.1109/EMBC.2019.8857148.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 August 2024 doi:10.20944/preprints202408.1319.v1

16. Guo, H.; Xu, S.; Wood, B.; Yan, P. Sensorless Freehand 3D Ultrasound Reconstruction via Deep Contextual Learning. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, 2020, pp. 463–472. doi:10.1007/978-3-030-59716-0_44.
17. Miura, K.; Ito, K.; Aoki, T.; Ohmiya, J.; Kondo, S. Localizing 2D Ultrasound Probe from Ultrasound Image Sequences Using Deep Learning for Volume Reconstruction. Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis. ASMUS PIPPI 2020, 2020, pp. 97–105. doi:10.1007/978-3-030-60334-2_10.
18. Guo, H.; Xu, S.; Wood, B.J.; Yan, P. Transducer adaptive ultrasound volume reconstruction. IEEE 18th International Symposium on Biomedical Imaging (ISBI), 2021, pp. 511–515. doi:10.1109/ISBI48211.2021.9433756.
19. Luo, M.; Yang, X.; Wang, H.; Du, L.; Ni, D. Deep Motion Network for Freehand 3D Ultrasound Reconstruction. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2022, pp. 290–299. doi:10.1007/978-3-031-16440-8_28.
20. Luo, M.; Yang, X.; Wang, H.; Dou, H.; Hu, X.; Huang, Y.; Ravikumar, N.; Xu, S.; Zhang, Y.; Xiong, Y.; Xue, W.; Frangi, A.F.; Ni, D.; Sun, L. RecON: Online learning for sensorless freehand 3D ultrasound reconstruction. *Medical Image Analysis* **2023**, *87*. doi:10.1016/j.media.2023.102810.
21. Luo, M.; Yang, X.; Yan, Z.; Li, J.; Zhang, Y.; Chen, J.; Hu, X.; Qian, J.; Cheng, J.; Ni, D. Multi-IMU with online self-consistency for freehand 3D ultrasound reconstruction. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2023, pp. 342–351. doi:10.1007/978-3-031-43907-0_33.
22. Li, Q.; Shen, Z.; Li, Q.; Barratt, D.C.; Dowrick, T.; Clarkson, M.J.; Vercauteren, T.; Hu, Y. Long-Term Dependency for 3D Reconstruction of Freehand Ultrasound Without External Tracker. *IEEE Transactions on Biomedical Engineering* **2024**, *71*, 1033–1042. doi:10.1109/TBME.2023.3325551.
23. Prevost, R.; Salehi, M.; Jagoda, S.; Kumar, N.; Sprung, J.; Ladikos, A.; Bauer, R.; Zettinig, O.; Wein, W. 3D freehand ultrasound without external tracking using deep learning. *Medical Image Analysis* **2018**, *48*, 187–202. doi:10.1016/j.media.2018.06.003.
24. Leblanc, T.; Lalys, F.; Tollenaere, Q.; Kaladji, A.; Lucas, A.; Simon, A. Stretched reconstruction based on 2D freehand ultrasound for peripheral artery imaging. *International Journal of Computer Assisted Radiology and Surgery* **2022**, *17*, 1281–1288. doi:10.1007/s11548-022-02636-w.
25. Wein, W.; Lupetti, M.; Zettinig, O.; Jagoda, S.; Salehi, M.; Markova, V.; Zonoobi, D.; Prevost, R. Three-Dimensional Thyroid Assessment from Untracked 2D Ultrasound Clips. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, 2020, pp. 514–523. doi:10.1007/978-3-030-59716-0_49.
26. Guo, H.; Chao, H.; Xu, S.; Wood, B.J.; Wang, J.; Yan, P. Ultrasound Volume Reconstruction From Freehand Scans Without Tracking. *IEEE Transactions on Biomedical Engineering* **2023**, *70*, 970–979. doi:10.1109/TBME.2022.3206596.
27. Martin, M.; Sciolla, B.; Sdika, M.; Wang, X.; Quetin, P.; Delachartre, P. Automatic Segmentation of the Cerebral Ventricle in Neonates Using Deep Learning with 3D Reconstructed Freehand Ultrasound Imaging. 2018 IEEE International Ultrasonics Symposium (IUS), 2018. doi:10.1109/ULTSYM.2018.8580214.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.