



Identifying visible tissue in intraoperative ultrasound: a method and application

Alistair Weld¹ · Luke Dixon² · Michael Dyck³ · Giulio Anichini⁴ · Alex Ranne¹ · Sophie Camp⁴ · Stamatia Giannarou¹

Received: 15 October 2024 / Accepted: 24 April 2025 / Published online: 28 June 2025
© The Author(s) 2025

Abstract

Purpose: Intraoperative ultrasound scanning is a demanding visuotactile task. It requires operators to simultaneously localise the ultrasound perspective and manually perform slight adjustments to the pose of the probe, making sure not to apply excessive force or breaking contact with the tissue, while also characterising the visible tissue.

Method: To analyse the probe-tissue contact, an iterative filtering and topological method is proposed to identify the underlying visible tissue, which can be used to detect acoustic shadow and construct confidence maps of perceptual salience.

Results: For evaluation, datasets containing both in vivo and medical phantom data are created. A suite of evaluations is performed, including an evaluation of acoustic shadow classification. Compared to an ablation, deep learning, and statistical method, the proposed approach achieves superior classification on in vivo data, achieving an F_β score of 0.864, in comparison with 0.838, 0.808, and 0.808. A novel framework for evaluating the confidence estimation of probe-tissue contact is created. The phantom data are captured specifically for this, and comparison is made against two established methods. The proposed method produced the superior response, achieving an average normalised root-mean-square error of 0.168, in comparison with 1.836 and 4.542. Evaluation is also extended to determine the algorithm's robustness to parameter perturbation, speckle noise, data distribution shift, and capability for guiding a robotic scan.

Conclusion: The results of this comprehensive set of experiments justify the potential clinical value of the proposed algorithm, which can be used to support clinical training and robotic ultrasound automation.

Keywords Ultrasound · Probe-tissue contact · Acoustic shadow

Introduction

Intraoperative imaging is an important tool for surgeons in decision-making during surgery. The low cost of ultrasound (US) machines and the ease of integration into surgical settings has led to enthusiasm for intraoperative ultrasound (IOUS) [1]. However, the widespread adoption of IOUS has been limited, and there remains great performance variability between operators [2]. Becoming proficient in IOUS requires

extensive training and hands-on experience, due to complex visuotactile interaction, imaging artefacts, noise, machine settings, and a lack of clear global anatomical key structures, which can cause interpretation ambiguity.

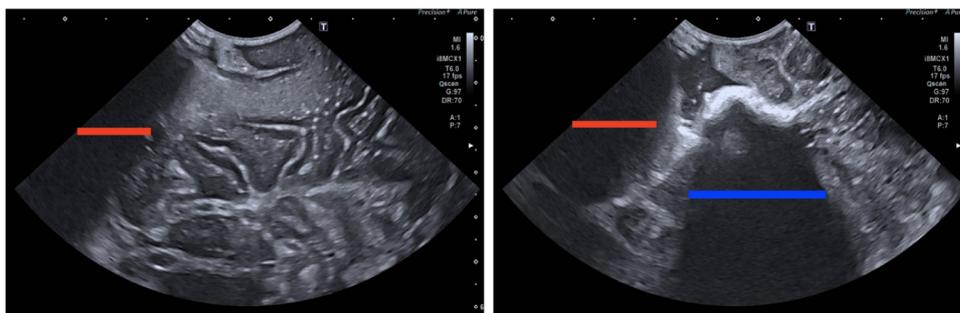
One particularly challenging aspect is balancing optimal probe-tissue contact and orientation, without causing inadvertent pressure-related damage. Contact-related acoustic shadow (AS) can occur due to interposed gas or contact breaking, which reduces image reliability through tissue obscuration. This is particularly challenging in the case of brain surgery [2].

Recent advances in IOUS have led to increased interest in algorithmic analysis of US feedback. Some methods detect shadows caused by bone [3, 4], calcification [5, 6], or general probe-tissue interactions [7–9]. However, these methods do not specifically examine probe-tissue coupling, which is the focus of this work. Figure 1 presents an example of where shadowing does not always equal poor contact. To

✉ Alistair Weld
a.weld20@imperial.ac.uk

¹ The Hamlyn Centre, Imperial College London, London, UK
² Department of Imaging, Charing Cross Hospital, London, UK
³ School of Computation, Information and Technology, Technical University of Munich, Munich, DE, Germany
⁴ Department of Neurosurgery, Charing Cross Hospital, London, UK

Fig. 1 Sample annotations of b-mode ultrasound images. The red line delineates acoustic shadowing secondary to poor contact while the blue line defines acoustic shadowing related to other causes (in this case echogenic blood clot in a resection cavity)



address this limitation, the algorithm in [10–12] produces a confidence map of signal attenuation, using a random walk-based method. The limiting factor for the latter method is the assumption that hyperechoic lines correlate with reduced signal attenuation. However, this is not a well-generalised assumption. In this work, we propose a method for identifying and topologically representing visible tissue in US for the detection of contact-related AS and confidence in perceptual salience. Our contributions are as follows:

1. A method for identifying visible tissue within US images is proposed, using iterative filters and topological representation, that can be used for probe-tissue contact-related AS detection.
2. A novel framework is proposed for the analysis of the confidence estimate of ultrasound probe-tissue contact.

Methodology

Detection of tissue features

To detect areas of high spatial intensity variation, linear US images $I \in \mathbb{R}^{height \times width}$ are iteratively processed using Gaussian filters, with a vertical bias, following:

$$G_f = h_1 * G_{f-1},$$

$$h_1 = \left(\frac{1}{\sqrt{2\pi}\sigma_u} e^{-u^2/2\sigma_u^2} \right) \left(\frac{1}{\sqrt{2\pi}\sigma_v} e^{-v^2/2\sigma_v^2} \right), \quad (1)$$

$$\sigma_u \gg \sigma_v, G_1 = I$$

where $G \in \mathbb{R}^{height \times width}$ denotes an output of the Gaussian filter, f the iteration set, and σ the standard deviation of the Gaussian filter stack. The image coordinates are u and v , where v increases with signal depth along the scan lines (image columns) that are perpendicular to the transducer at the top. The Gaussian smoothing is biased vertically, as anatomical changes are more likely to be exposed in the vertical direction. The function is linear and the iterative processing is associative ($G * h_1$) * $h_1 = G * (h_1 * h_1)$ and therefore equal to a single convolution with a wider standard deviation $\hat{\sigma}_u \approx \sqrt{f}\sigma_u$. Direct edge-detection methods (e.g.

Laplacian or Sobel) are not robust to noise and do not yield a smooth response. Each output G_f of the Gaussian filter sequence is stacked and the standard deviation is taken for each pixel across the stack following:

$$S_{u,v} = \hat{G} * \sqrt{\frac{\sum_{f=1}^{len(f)+1} \left(G_{f,u,v} - \frac{\sum_{f=1}^{len(f)+1} G_{f,u,v}}{len(f)+1} \right)^2}{len(f) + 1}} \quad (2)$$

where $S \in \mathbb{R}^{height \times width}$ is the output of the standard deviation. \hat{G} is an extra smoothing filter for the standard deviation output with $\sigma = 1$, to remove outliers.

Points with high spatial variation correspond to areas that have been most affected by the smoothing process. Hence, these points are separated by thresholding the standard deviation output following:

$$\chi = \{(u, v) \mid S_{u,v} \geq \gamma\}, \gamma = \text{round}(\sigma_u \times \sqrt{len(f)}) \quad (3)$$

The γ formulation is elaborated in Sect. 4.2.2. Points of high spatial variation are represented as a 2D point cloud χ .

To reduce the computational load for building the topology and to remove outliers, the point cloud is sparsified utilising an inverse Gaussian density map $\lambda \in \mathbb{R}^{height \times width}$ following:

$$\lambda = \phi_1 + \phi_2 \left(1 - \frac{\mathbf{1}_{\phi_2 \times \phi_2}}{\phi_2^2} * \frac{(h_2 * (S > \gamma))^2}{\max((h_2 * (S > \gamma))^2)} \right), \quad (4)$$

$$h_2(u, v) = \frac{1}{2\pi\sigma_w^2} \exp\left(-\frac{u^2+v^2}{2\sigma_w^2}\right) \quad (5)$$

where ϕ denotes the weighting parameters, $\mathbf{1}_{\phi_2 \times \phi_2}$ is a matrix $\phi_2 \times \phi_2$ of ones, h_2 is the Gaussian kernel, and $(S > \gamma)$ is a 2D binary array. The point cloud is then iterated over, where Eq. 6 is repeated until all points are separated by their corresponding λ_a distance threshold:

$$A := \{a \in \chi \mid \exists b \in \chi, b \neq a : \text{dist}(a, b) \leq \lambda_a\},$$

$$\chi - A := \{a \in \chi \mid \forall b \in \chi, b \neq a : \text{dist}(a, b) > \lambda_a\} \quad (6)$$

where a, b indexes elements in the point cloud set, and $\text{dist}()$ gives the Euclidean distance between two points (pixels).

To generate the topological representation of the visible tissue, a Vietoris–Rips complex is then constructed on the sparse point cloud, defined as:

$$\Psi := \{d \subseteq \chi \mid \forall a, b \in d : \text{dist}(a, b) \leq \epsilon\} \quad (7)$$

where ϵ defines the maximum distance threshold. The Vietoris–Rips complex method was chosen over alternatives such as the Delaunay method because of its simplicity and robustness to outliers.

Artefact detection and confidence map generation

A 2D simplex δ is made up of three points within the complex Ψ , whose pairwise distances all lie within the threshold ϵ ; let $\Delta \subseteq \Psi$ be the set of all 2D simplices. For each scan line ξ_{sl} (where $sl \in [1 \dots \text{width}]$), we assign a value of 1 (AS) or 0 (visible tissue). Contact-related AS is defined for a scan line if the number of overlapping 2D simplices in Δ along that line is below a specified threshold τ .

The confidence map is a pixel-wise quantification of the perceptual salience, which represents the resolution and contrast of the detected visible tissue. A high score indicates high-quality tissue contrast and resolution. For every pixel in the image, the number of 2D simplexes that overlap is counted, creating an occurrence map O . The confidence map is constructed by taking the log of the occurrence map, $\hat{M} = \log(O + 1)$.

Resources and evaluation setup

Intraoperative data

Data curation

For the AS classification, a data set containing 51 images from 11 different patients was collected. The images were sourced from routine pseudo-anonymised 2D US scans, acquired in vivo during brain surgery, using a Canon i900 US machine, using both a linear and a convex probe. The data were retrospectively reviewed and annotated by a consultant neuroradiologist experienced in neurooncological iUS. The images exhibit varying degrees of AS artefacts secondary to poor probe-tissue contact. The study had full local ethical approval from the HRA and Health and Care Research

Wales (HCRW) authorities¹. For standardisation, all convex images were linearised using a standard approach of taking the inner trapezoid and interpolating. All linear images have the shape 300×600 . To reduce the influence of near-field artefacts, the top 100 rows are cropped from all linearised US images. This choice was based on observations from the in vivo dataset and the maximum point where the near-field artefacts, caused by factors that included sheathing, gel, and cranial cavity interaction, ended.

For performance evaluation, threefold cross-validation was used, and the data set is split into a seen and unseen subset, with the division done on the patient level.

Speckle noise robustness evaluation setup

Speckle noise is the most prominent artefact, created by the interaction of signals with biological tissue, which can cause distortion of anatomical details and edges. As proposed in [13], the observed US image I can be defined as the product of the underlying clean image \hat{I} affected by multiplicative ξ_m and additive ξ_a noise, given by:

$$I = \hat{I} \odot \xi_m \oplus \xi_a \quad (8)$$

where it is commonly assumed that $\|\xi_a\|^2 \ll \|\hat{I} \odot \xi_m\|^2$ [14]. Commonly, the additive term is ignored due to the lesser impact. Assuming sampling from a generalised Gamma distribution (GGD) [15] with the probability density function (pdf):

$$p_Z(z) = \frac{\gamma_{ggd} z^{\gamma_{ggd} \nu - 1}}{\alpha^{\gamma_{ggd} \nu} \Gamma(\nu)} \exp\left\{-\frac{z}{\alpha} \gamma_{ggd}\right\}, z \geq 0, \alpha, \nu, \gamma_{ggd} > 0 \quad (9)$$

To demonstrate robustness to speckle noise, anatomical edges such as bone or layered tissue is models as a square wave \hat{I} :

$$Q(x) = \begin{cases} q_l & \text{if } x < a \\ q_u & \text{if } a \leq x < b, 0 \leq q_l \ll q_u \leq 255 \\ q_l & \text{if } x \geq b \end{cases} \quad (10)$$

where a, b are the transition time steps and q_l, q_u are the lower and upper values. Consequently,

$$(q_l \odot \xi_m \oplus \xi_a) \ll q_u \quad \text{and} \quad \begin{aligned} |q_l - (q_l \odot \xi_m \oplus \xi_a)| &\ll |q_u - q_l|, \\ |q_u - (q_u \odot \xi_m \oplus \xi_a)| &\ll |q_u - q_l|, \end{aligned}$$

¹ Study title—US-CNS: Multiparametric Advanced Ultrasound Imaging of the Central Nervous System Intraoperatively and Through Gaps in the Bone, IRAS project ID: 275556, Protocol number: 22CX7609, REC reference: 22/WA/0259, Sponsor: Research Governance and Integrity Team (RGIT).

which therefore indicates that artificial structures could only be created by sampling from the extreme tail of p_Z

$$P\left(\bigcap_{i=1}^n \{Z_i > \hat{z}\}\right) = \left(\int_{\hat{z}}^{\infty} p_Z(z) dz\right)^n, \quad \hat{z} \gg 0, \quad (11)$$

which is exponentially unlikely.

Experimental evaluation of the robustness of the algorithm to speckle noise is performed by quantifying the topological feature discrepancies when speckle noise is applied. The speckle noise applied via Eqs. 8, 9, with the ν, γ_{ggd} values varied within a range $\{1, 1.5, 2, 2.5, 3\}$, and sampled from the GGD, using rejection sampling [16]. To evaluate the topological difference, the q -Wasserstein distance (with $q = 2$) is used to measure the discrepancies between persistence diagrams, derived from the output point cloud [17–19]. This measures the topological perturbation under varying noise influence. The q -Wasserstein distance can be defined as $W_q(D, \hat{D}) = \inf_{\varphi: D \rightarrow \hat{D}} \left(\sum_{t \in D} \|t - \varphi(t)\|_{\infty}^q\right)^{\frac{1}{q}}$ [20], where φ ranges over bijections between the two persistence diagrams D and \hat{D} (affected by speckle noise). W_q is normalised for interpretability.

Classification evaluation and comparison setup

For the evaluation of the AS classification, the proposed method (*Topol*) is compared to a statistical method that uses the sum of the intensity of a scan line (*Baseline₁*), a statistical method that uses the change in the intensity of the pixels (*Baseline₂*) and a neural network *VGG16* [21]. The *Baseline₁* method was motivated by [5, 6]. The use of the *VGG16* architecture was motivated by [9]. *Baselline₂* is used for the ablation study, the proposed method up to the iterative filtering standard deviation. Confusion matrix analysis is performed, where a positive classification is for AS. The methods are ranked using $F_{\beta} = \frac{(1+\beta^2) \times PPV \times TPR}{\beta^2 \times PPV + TPR}$, where PPV is a positive predictive value, and TPR is the true positive rate $\beta = 0.5$ to prioritise precision over sensitivity.

The parameter tuning for *Topol* is done for each fold using the seen subset exclusively. The best performing set of parameters for *Topol*—found using a grid search—was near identical across the 3 folds [$f = [2, 3, 4, 5, 6]$, $\sigma_u = 1.8$, $\sigma_w = 20$, $\phi_1 = 15$, $\phi_2 = 20$, $\epsilon = 60$]. With the minor exception of the Gaussian filter, folds 1 and 3 used $\sigma_v = 0.2$, while fold 2 used $\sigma_v = 0.3$. The scan line threshold, to determine the AS, was constant for all folds $\tau = 2$. Although the parameters were chosen using a manually defined grid, this process could easily be performed fully automatically using any hyperparameter optimisation method.

Baseline₁ individually classifies each scan line through scan line summing, with a fixed threshold $\sum_{i=1}^{300} \xi_{sl,i} <$

$\kappa_{Baseline_1}$. The threshold is determined using a receiver operating characteristic curve (ROC). For folds $[1, 2, 3]$ - $\kappa_{Baseline_1} = [4900, 5000, 5500]$. The ROC process is repeated for *Baseline₂*, where for folds $[1, 2, 3]$ - $\kappa_{Baseline_2} = [325, 350, 350]$. A deep learning method is implemented to gauge generalisability given data limitations. Linear probing was used with a frozen pretrained *VGG16* feature encoder (ImageNet) with a trainable classification head, where the outputs are the classification scores for each scan line in the image. Training is performed for 10 epochs using Sigmoid cross-entropy. *Topol*, *Baseline₁* and *Baseline₂*, were tuned using accuracy (ACC).

Parameters robustness is explored by individually perturbing the calibrated parameters from fold 3 and measuring the effect on the corresponding unseen data. For the parameter study, a maximum perturbation of 20% was deemed suitable for a valid robustness evaluation. The coefficient of variation $= 100 \times \frac{\sigma_{pert}}{\mu_{pert}}$ is used to measure the impact on performance.

Robotic setup

To extend the AS experiments, a robotic setup is created using the AS output to guide a scanning process—parameters from fold 1 were used. Specifically, AS detection is used to control probe-tissue coupling during the landing, rotation, and the tilting of the probe, optimising probe-tissue (by correcting the in plane rotation through adjustment along the longitudinal axis of the probe and centring the US imaging) contact throughout the scanning process and probe manipulation. The experiments were carried out using a seven DoF DLR MIRO [22], with a *GE LOGIQ e* US machine and a *GE 12L-RS* linear US probe. The phantom is created using a mixture of GELITA@GELATINE *Type Ballistic 3* with water and glycerine, with olives and blueberries added for echo diversity [23].

Through Cartesian forward kinematics $f_{kin} : \mathcal{Q} \rightarrow SE(3) \mid \mathbf{q} \mapsto \mathbf{x} = f_{kin}(\mathbf{q})$, the pose of the US probe $\mathbf{x} \in SE(3)$ can be represented as a function of the generalised configuration coordinates of the robot \mathbf{q} . A Cartesian impedance controller [24, 25] can be constructed to control a robot, by determining desired joint torques $\boldsymbol{\tau} \in \mathbb{R}^n$.

$$\boldsymbol{\tau} = \mathbf{J}^T(\mathbf{q})[\mathbf{K}(\mathbf{x}_d - \mathbf{x}) + \mathbf{D}(\mathbf{q})(\dot{\mathbf{x}}_d - \dot{\mathbf{x}})], \quad (12)$$

where \mathbf{x} and $\dot{\mathbf{x}}$ are the current pose and velocity. This forms a spring-damper system with stiffness and damping parameters $\mathbf{K}, \mathbf{D}(\mathbf{q}) \in \mathbb{R}^{6 \times 6}$. Enabling control of pose \mathbf{x}_d and velocity $\dot{\mathbf{x}}_d$ of a US probe. The geometrical Jacobian $\mathbf{J}(\mathbf{q}) \in \mathbb{R}^{6 \times 7}$ is given by $\dot{\mathbf{x}} = [\mathbf{v} \ \boldsymbol{\omega}]^T = \mathbf{J}(\mathbf{q})\dot{\mathbf{q}}$.

$$m_1 = \frac{1}{width} \sum_{sl=1}^{width} (1 - \xi_{sl}), \quad m_1 \in [0; 1] \quad (13)$$

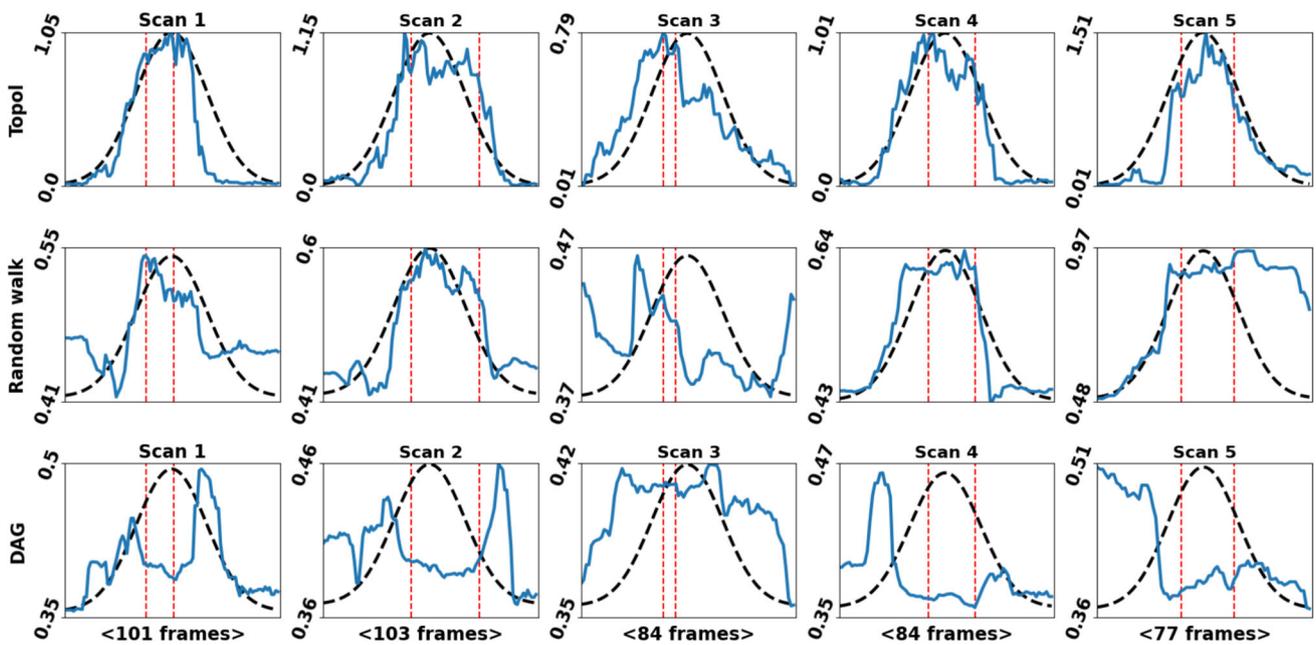


Fig. 2 Per frame average confidence vector outputs—x-axis = frames and y-axis = average confidence. The blue lines show the method’s outputs, *Topol* top, *Randomwalk* middle and *DAG* bottom row. The black dotted lines are the scaled normal curve. The dotted red lines are

where the annotators perceived the best images. What can be observed is that the output of *Topol* follows the desired trajectory, closer than the other methods

$$m_2 = \frac{1}{width/2} \left(\sum_{sl=width/2}^{width} \xi_{sl} - \sum_{sl=1}^{width/2} \xi_{sl} \right),$$

$$m_2 \in [-1; 1], \tag{14}$$

Two metrics are defined to measure the coupling— m_1 is the coverage of contact and m_2 is the symmetry of contact. These metrics can guide the scan by defining desired values $m_{1,d}$ and $m_{2,d}$ and mapping the error to a desired Cartesian velocity expressed in the probe’s coordinate frame, enabling longitudinal correction of the contact.

$$\dot{x}_d = \begin{bmatrix} v_{d,m} \\ \omega_{d,m} + \omega_{d,t} + \omega_{d,r} \end{bmatrix}$$

$$= \begin{bmatrix} (m_{1,d} - m_1) [0 \ 0 \ k_1]^T \\ (m_{2,d} - m_2) [k_2 \ 0 \ 0]^T + \omega_{d,t} + \omega_{d,r} \end{bmatrix}, \tag{15}$$

with two gains $k_1 = 0.005$ m/s and $k_2 = 0.05$ rad/s. The translation velocity is defined by $v_{d,m}$, while the rotation velocities are defined by $[\omega_{d,m}, \omega_{d,t}, \omega_{d,r}]$. Here, the probe coordinate frame is defined so that the y-axis aligns with the scanning direction, the z-axis with the longitudinal axis of the probe and the x-axis orthogonal to both. The integration of Eq. (15) gives the desired pose x_d . If no trajectory is defined ($\omega_{d,t} = \omega_{d,r} = [0, 0, 0]$), so, for example, the probe is landing or maintaining perpendicular contact. Using \dot{x}_d , the robot will automatically adjust the position of the probe

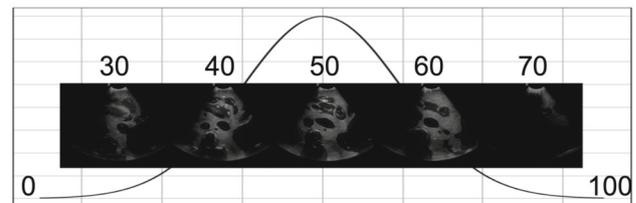


Fig. 3 Image feedback at different frames from scan 1—abdominal. The optimal image is shown at frame 50

along the longitudinal axis until $m_1 = m_{1,d}$, and adapt the orientation until ($m_2 = m_{2,d} = 0$). To perform rotation and tilt, $\omega_{d,t} = [0 \ * \ 0]^T$ and $\omega_{d,r} = [0 \ 0 \ *]^T$ accordingly, where $*$ is defined by the operator.

Medical phantom data

Data curation

Controlled evaluation of confidence is difficult, as optimal scans may produce subpar images due to unavoidable factors such as forced positioning, e.g., limited space within a cavity or the location of the pathological tissue. Data were collected for relative evaluation, which involved tilting the probe from $-90^\circ > 0^\circ$ (aligned with the normal of the tissue surface) $> +90^\circ$ with respect to the normal vector of the tissue surface. Data were captured using a Zonare Z.one US system with

Fig. 4 Sample simplicial complex outputs. These complexes indicate areas of significant spatial intensity variation within the ultrasound images. Areas where there are dense complexes, correlate with high confidence in tissue visibility; correlating with a confidence map output of higher perceptual quality

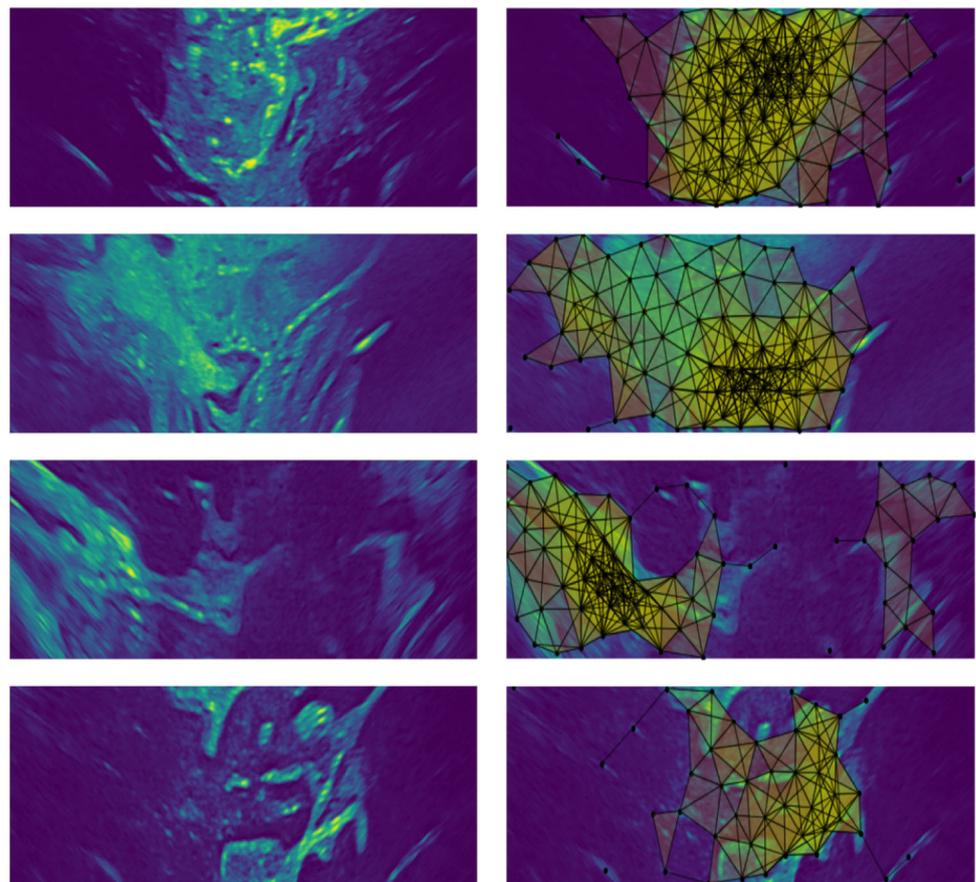


Table 1 Scan line classification results from the intraoperative data

3F Avg	Algo	ACC	TNR	TPR	PPV
Seen/Train	<i>Baseline</i> ₁	0.87	0.88	0.86	0.86
	<i>Baseline</i> ₂	0.88	0.91	0.85	0.90
	<i>VGG16</i>	0.90	0.92	0.90	0.90
	<i>Topol</i>	0.89	0.95	0.83	0.93
Unseen/Test	<i>Baseline</i> ₁	0.84	0.81	0.89	0.79
	<i>Baseline</i> ₂	0.87	0.85	0.87	0.83
	<i>VGG16</i>	0.83	0.82	0.84	0.80
	<i>Topol</i>	0.87	0.90	0.84	0.87

A positive sample corresponds to a scan line affected by AS. Highlighted in grey is the best score for each metric. The results have been averaged over the threefolds

Bold text highlights the best result

a C4-1 transducer. Four scans were taken using the FAST / acute abdomen Phantom. Scans 1 and 2 from the abdomen, scan 3 from the rib cage, and scan 4 from the kidney. A fifth scan was taken using a Kezlex brain phantom. Frames were extracted at 10Hz. For generalisation evaluation of AS classification, 449 frames are extracted and annotated.

Table 2 Rankings (R#) of the methods using the F_{β} scores

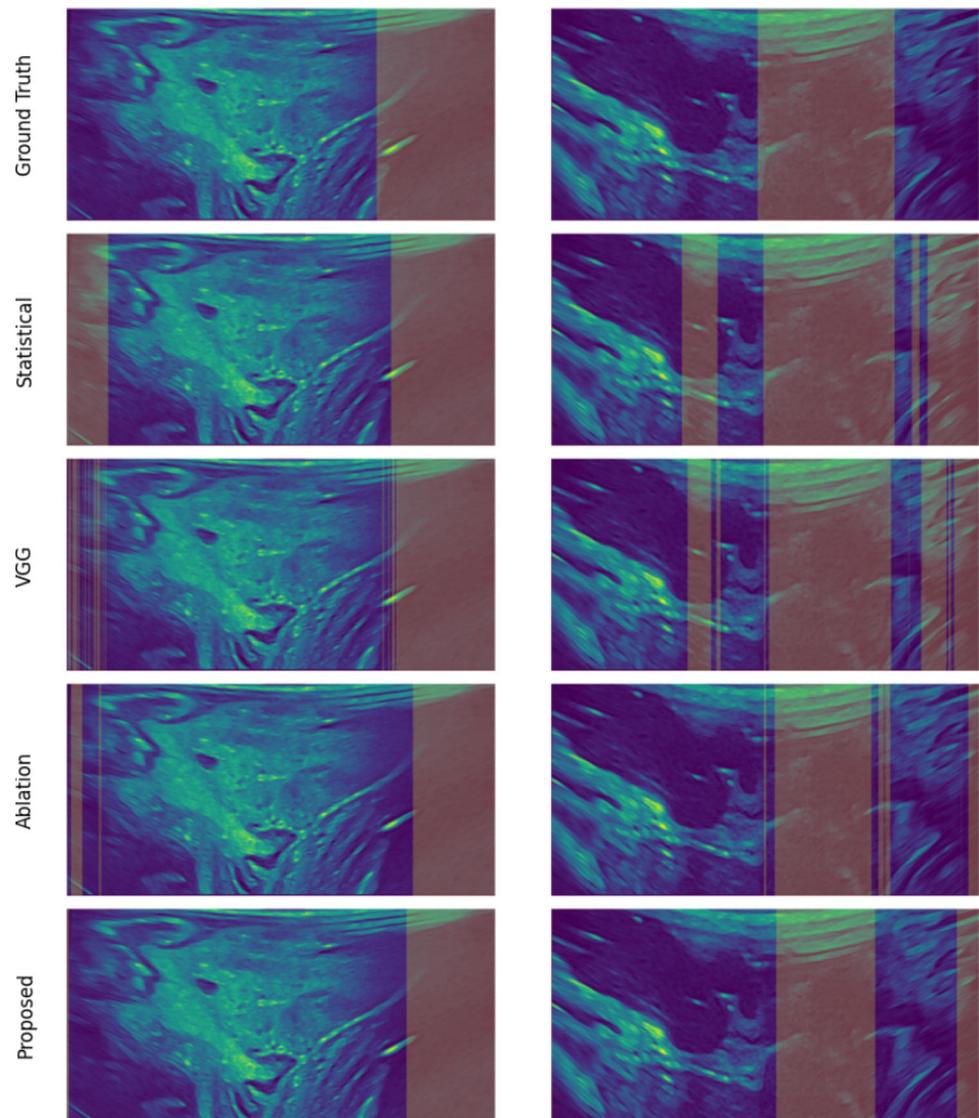
3F Avg	R1	R2	R3	R4
Seen/Train	Topol	<i>VGG16</i>	<i>Baseline</i> ₂	<i>Baseline</i> ₁
	0.908	0.900	0.890	0.860
Unseen/Test	Topol	<i>Baseline</i> ₂	<i>VGG16</i>	<i>Baseline</i> ₁
	0.864	0.838	0.808	0.808

The results are a product of averaging across the threefolds
Bold text highlights the best result

Confidence map evaluation and comparison setup

The proposed method is compared with the *Randomwalk* [10] and *DAG* [26] methods, using their official codes and default parameters. The *Topol* parameters tuned on fold 1 of the AS classification were used for this experiment. Due to the way the data were collected, the target confidence vector is defined as a normal distribution across the frames. Based on [27], $mean(\hat{M})$ is used to determine the confidence in the perceptual saliency of the visible anatomy per frame. The evaluation is performed using the normalised root-mean-squared error, $(NRMSE) = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (T_i - \hat{T}_i)^2}}{\max(T) - \min(T)}$, $\hat{T}_i = mean(\hat{M}_i)$, where T is the target confidence vector, the black

Fig. 5 AS detection on two US images. The top row shows the annotation, the second is from *Baseline*₁, the third is from *VGG16* and the bottom is from *Topol*. The orange areas are detected AS



dotted lines in Fig. 2, fitted to the max/min of \hat{T} ; n is the frame. Figure 3 is from scan 1. The best image is the middle frame, and the black curve is representative of the assumed confidence trajectory. To confirm this, annotations were made to highlight the optimal frame range, highlighted with the red dotted lines in Fig. 2.

Classification generalisability

To test the generalisability and performance of each algorithm on out-of-distribution data, AS scanline classification experiments were extended to phantom data (different machine, anatomy and image texture). Each method was evaluated three times, using the three sets of parameters calibrated for the threefolds described in Sect. 4.2, and the average results across the threefolds were recorded.

Results and discussion

The evaluation was carried out primarily using a Python implementation; results are shown in Fig. 4. The method was also implemented with minimal optimisation using C++, which ran at 20Hz, indicating real-time executability.

Noise robustness

The q-Wasserstein distances, after applying the set speckle noise $\{1, 1.5, 2, 2.5, 3\} \times \{1, 1.5, 2, 2.5, 3\}$ for $\nu \times \gamma_{ggd}$, measured an average result of 0.097. This result gives evidence of the robustness of the topological construction under varying speckle noise. The topological differences are largely uniform across all distributions.

Table 3 Rankings (R#) of the methods' performances on the generalisability test on the probe pose analysis data, using the F_β scores

	R1	R2	R3	R4
Scan 1	Topol 0.923	<i>Baseline₂</i> 0.842	<i>Baseline₁</i> 0.807	<i>VGG16</i> 0.694
Scan 2	Topol 0.920	<i>Baseline₂</i> 0.883	<i>Baseline₁</i> 0.789	<i>VGG16</i> 0.740
Scan 3	Topol 0.833	<i>Baseline₂</i> 0.789	<i>Baseline₁</i> 0.745	<i>VGG16</i> 0.377
Scan 4	Topol 0.891	<i>VGG16</i> 0.847	<i>Baseline₂</i> 0.797	<i>Baseline₁</i> 0.753
Scan 5	Topol 250.830	<i>VGG16</i> 0.777	<i>Baseline₂</i> 0.686	<i>Baseline₁</i> 0.643

For all scans, *Topol* achieves the best rank
Bold text highlights the best result

Acoustic shadow classification

The results are presented in Tables 1 and 2 where TNR is the true negative rate. Outputs are shown in Fig. 5. Firstly, it

Table 4 Parameter perturbation, performance drift measured via coefficient of variation values (%)

Parameter	ACC	TNR	TPR	PPV
f	0.431	1.928	2.204	1.749
σ_u	0.320	1.191	1.455	1.011
σ_v	0.455	2.380	2.290	2.151
γ	0.982	4.499	3.986	3.925
ϕ_1	0.622	1.683	1.453	1.579
ϕ_2	0.512	1.150	1.550	1.031
σ_w	0.358	0.277	0.635	0.347
ϵ	2.270	8.663	5.986	6.306

can be determined that *Topol* is the most robust method for distribution shift with more consistent performance between seen and unseen data. The results produced by *Baseline₁* show that overreliance on the raw pixel intensities leads to oversensitivity. By producing results superior to *Baseline₁*, *Baseline₂* justifies the iterative filtering process, demon-

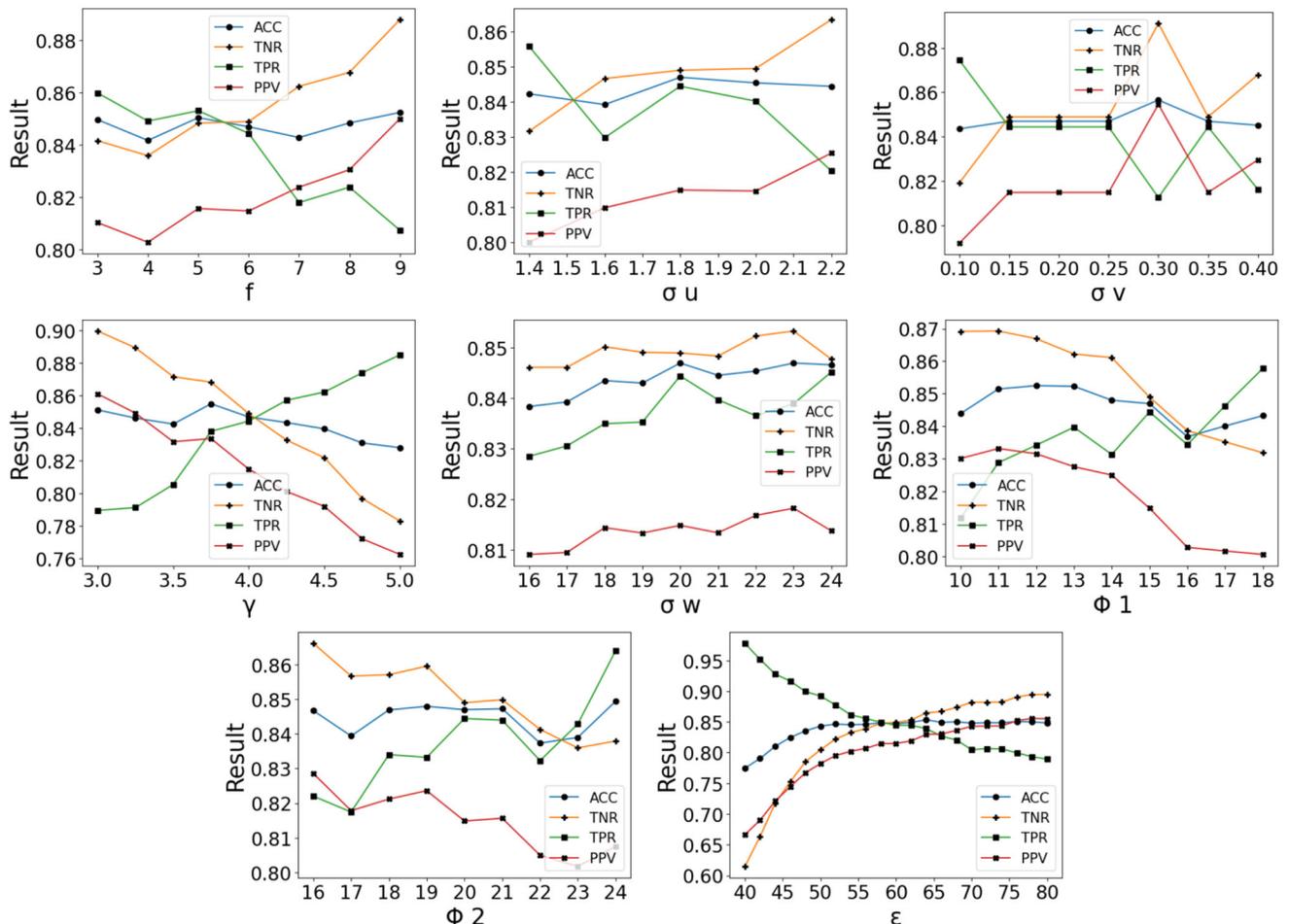
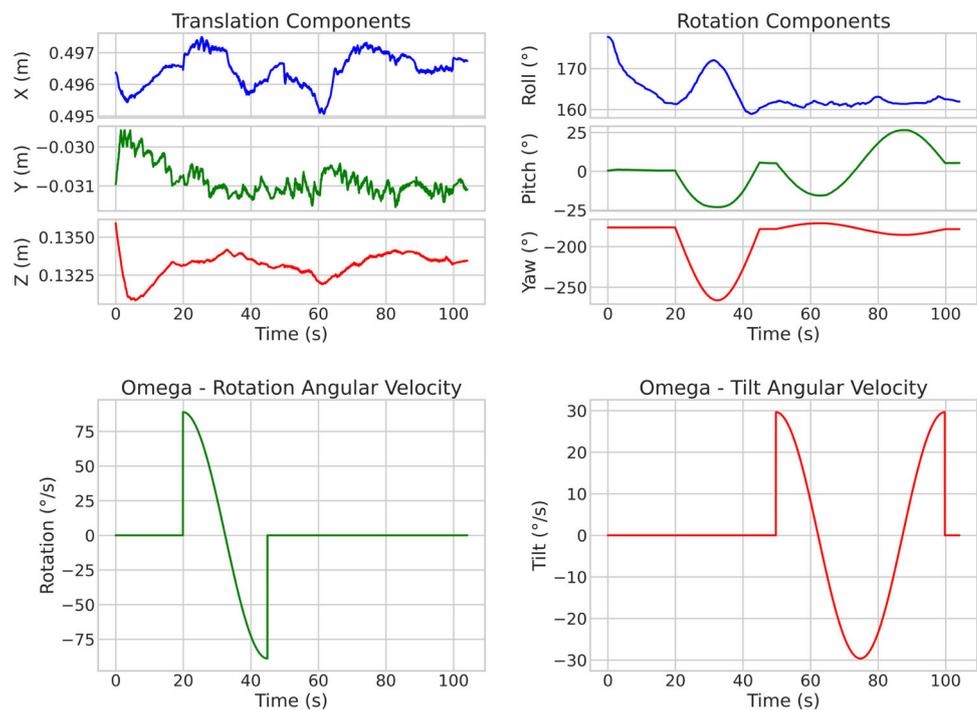


Fig. 6 Graphs displaying the effect of individually varying the parameters on the performance of the whole system—using the unseen fold 3 data. Blue lines = ACC, orange lines = TNR, green lines = TPR and red lines = PPV

Fig. 7 Top row shows the real robot movement, the bottom row shows the desired angular velocities (trajectories)



strating that identifying areas of dynamic change is a strong indicator of where there is AS. *VGG16* shows achieves inadequate generalisation. From the F_β ranks, it is conclusive that *Topol* produces the ideal behaviour.

Generalisability study

The results of the phantom dataset experiments are shown in Table 3. For all scans, *Topol* achieves the best performance. The ranking of *Baseline₂* again highlights the advantage of using the intensity change rather than the raw intensity. The inconsistent performance of *VGG16* highlights the unreliability of deep learning methods when only small datasets are available.

Parameter perturbation study

The results of this experiment are shown in Fig. 6. It is observed that the overall performance is only marginally affected by individually drifting the parameters, with most metrics remaining above 0.8. The coefficient of variation values (%) of the performance drift are shown in Table 4. This is evidence of the robustness of the design of the algorithm, as the performances do not vary largely under parameter uncertainty. We note that the γ value that was decided was initially based on intuition from the iterative filtering process. In particular, calculating the overall standard deviation $\sqrt{5} \times 1.8 \approx 4.02$ yields a value close to $\gamma=4$. This result suggests a potential correlation between the iteration and the optimal threshold choice.

Robotics results

The sequence of robotic action is first the landing and stabilisation of the probe on the phantom surface, followed by a $\pm 90^\circ$ rotation, followed by a $\pm 30^\circ$ tilt. The actual position and orientation, measured via the homogeneous transformation matrix in the robot base frame, as well as the corresponding position and orientation commands (desired trajectories) of the probe, are shown in Fig. 7. Of the translation components, Z is of greatest interest, as the scan was held on a single spot, and Z was the main translation for correction. The real homogeneous transformations matrices show that the robot correctly executed the motion without large deviation. The AS results for $[m_1, m_2]$ are shown in Fig. 8. It can be determined that AS analysis enabled the control framework to successfully anchor the contact to the desired coverage and symmetry.

Confidence map behaviour

Figure 2 shows that for all scans, the behaviour of *Topol* is closer to what is expected. The range of \hat{T} values produced by *Topol* can also be considered more intuitive than *Randomwalk* and *DAG* as the confidence starts near 0 and peaks around 1. Quantitative analysis is also performed and shown in Table 5, produced by measuring the NRMSE of \hat{T} (blue line) versus T (black dotted line). For each scan, the proposed method produces a significantly better NRMSE score than the other methods.

Fig. 8 Acoustic shadow behaviour and optimisation for robot scan

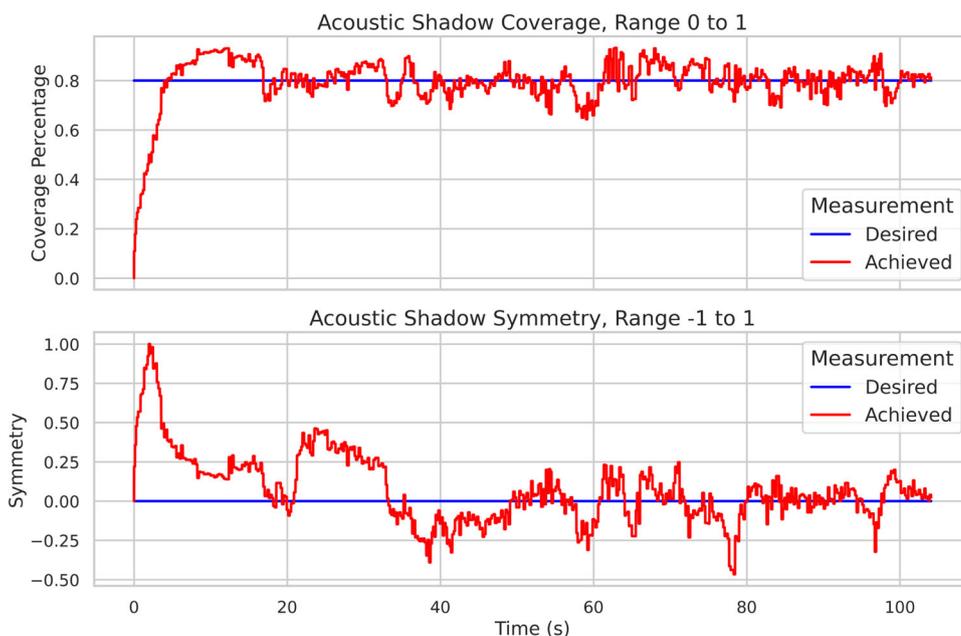


Table 5 Method's NRMSE results are shown for each scan

Method	Scan 1	Scan 2	Scan 3	Scan 4	Scan 5
<i>Topol</i>	0.18	0.13	0.28	0.14	0.11
<i>Randomwalk</i>	1.85	0.94	4.79	0.73	0.87
<i>DAG</i>	2.40	4.96	6.08	5.12	4.15

In all cases, *Topol* produces the best score
 Bold text highlights the best result

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Conclusion

This paper introduces a novel method for evaluating the quality of US feedback. An evaluation of AS scan line classification and confidence behaviour has been performed. In both cases, the proposed method outperforms the compared methods. Further experiments show that this method is capable of guiding a robotic arm in performing an autonomous scan. The explicit evaluation of US probe-tissue coupling is vital for ensuring safe scanning. This method provides a novel perspective on how this can be approached, creating an avenue of future work which can be beneficial for both robotic US and supporting trainee operators.

Acknowledgements This project was supported by the UK Research and Innovation (UKRI) Centre for Doctoral Training in AI for Healthcare (EP/S023283/1), the Royal Society (URF\R\201014), the NIHR Imperial Biomedical Research Centre.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

- Sastry RA, Bi WL, Pieper SD, Frisken SF, Kapur T, Wells WM, Golby AJ (2017) Applications of ultrasound in the resection of brain tumors. *JON* 27(1):5–15
- Dixon L, Lim A, Grech-Sollars M, Nandi D, Camp SJ (2022) Intraoperative ultrasound in brain tumor surgery: a review and implementation guide. *Neurosurg Rev* 45:2503–2515
- Berton F, Cheriet F, Miron M-C, Laporte C (2016) Segmentation of the spinous process and its acoustic shadow in vertebral ultrasound images. *CBM* 72:201–211
- Alsinan AZ, Patel VM, Hacıhaliloglu I (2020) Bone shadow segmentation from ultrasound data for orthopedic surgery using gan. *IJCARS* 15(9):1477–1485
- Santos Filho E, Saijo Y, Yambe T, Tanaka A, Yoshizawa M (2006) Segmentation of calcification regions in intravascular ultrasound images by adaptive thresholding. In: *CBMS*. IEEE, pp 446–454
- Lee JH, Kim GY, Hwang YN, Kim SM (2018) Automatic detection of dense calcium and acoustic shadow in intravascular ultrasound images by dual-threshold-based segmentation approach. *Sens Mater*
- Hellier P, Hellier P, Coupé P, Morandi X, Collins DL (2010) An automatic geometrical and statistical method to detect acous-

- tic shadows in intraoperative ultrasound brain images. *MIA* 14(2):195–204
8. Meng Q, Housden J, Matthew J, Rueckert D, Schnabel JA, Kainz B, Sinclair M, Zimmer V, Hou B, Rajchl M, Toussaint N, Oktay O, Schlemper J, Gómez A (2019) Weakly supervised estimation of shadow confidence maps in fetal ultrasound imaging. *TMI* 38:2755–2767
 9. Melapudi V, Aladahalli C, Shriram KS, Chang H-M (2022) Exploiting acquisition path to detect shadows in ultrasound images. In: *Medical imaging*, vol 12038. SPIE, pp 143–148
 10. Karamalis A, Wein W, Klein T, Navab N (2012) Ultrasound confidence maps using random walks. *MIA* 16(6):1101–12
 11. Chatelain P, Krupa A, Navab N (2017) Confidence-driven control of an ultrasound probe. *TRO* 33(6):1410–1424
 12. Jiang Z, Grimm M, Zhou M, Esteban J, Simson W, Zahnd G, Navab N (2020) Automatic normal positioning of robotic ultrasound probe based only on confidence map optimization and force measurement. *RAL* 5(2):1342–1349
 13. Jain AK (2018) *Fundamentals of digital image processing. Control of Color Imaging Systems*
 14. Zong X, Laine AF, Geiser EA (1998) Speckle reduction and contrast enhancement of echocardiograms via multiscale nonlinear processing. *TMI* 17(4):532–540
 15. Michailovich OV, Tannenbaum A (2006) Despeckling of medical ultrasound images. *TUFFC* 53(1):64–78
 16. Neumann V (1951) Various techniques used in connection with random digits. *Notes by GE Forsythe*, 36–38
 17. Maria C, Boissonnat J-D, Glisse M, Yvinec M (2014) The gudhi library: simplicial complexes and persistent homology. In: *International congress on mathematical software*
 18. Flamary R, Courty N, Gramfort A, Alaya MZ, Boisbunon A, Chambon S, Corenflos A, Fournier N, Gayraud N, Janati H, Redko I, Rolet A, Schutz A, Sutherland DJ, Tavenard R, Tong A, Vayer T, Mueller A (2021) Pot: Python optimal transport. *JMLR* 22(78):1–8
 19. Lacombe T, Cuturi M, Oudot S (2018) Large scale computation of means and clusters for persistence diagrams using optimal transport. *NeurIPS* 31
 20. Mileyko Y, Mukherjee S, Harer J (2011) Probability measures on the space of persistence diagrams. *Inverse Prob* 27(12):124007
 21. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: *ICLR*
 22. Seibold U, Kübler B, Bahls T, Haslinger R, Steidle F (2018) The DLR Mirosurge surgical robotic demonstrator. World Scientific, pp 111–142. https://doi.org/10.1142/9789813232266_0005
 23. Morehouse H, Thaker HP, Persaud C (2007) Addition of metamucil to gelatin for a realistic breast biopsy phantom. *JUM* 26(8):1123–1126
 24. Hogan N (1985) Impedance control: an approach to manipulation: Part I-Theory
 25. Khatib O (1987) A unified approach for motion and force control of robot manipulators: the operational space formulation. *JRA* 3(1):43–53
 26. Hung ALY, Chen W, Galeotti JM (2021) Ultrasound confidence maps of intensity and structure based on directed acyclic graphs and artifact models. In: *ISBI. IEEE*, pp 697–701
 27. Virga S, Zettinig O, Esposito M, Pfister K, Frisch B, Neff T, Navab N, Hennemersperger C (2016) Automatic force-compliant robotic ultrasound screening of abdominal aortic aneurysms. In: *IROS. IEEE*, pp 508–513

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.