This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2021.3126530, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

Deep Learning for Instrumented Ultrasonic Tracking: From synthetic training data to *in vivo* application

Efthymios Maneas, Andreas Hauptmann, *Member, IEEE,* Erwin J. Alles, Wenfeng Xia, Tom Vercauteren, Sebastien Ourselin, Anna L. David, Simon Arridge, and Adrien E. Desjardins

***Abstract*— Instrumented ultrasonic tracking is used to improve needle localisation during ultrasound guidance of minimally-invasive percutaneous procedures. Here, it is implemented with transmitted ultrasound pulses from a clinical ultrasound imaging probe that are detected by a fibre-optic hydrophone integrated into a needle. The detected transmissions are then reconstructed to form the tracking image. Two challenges are considered with the current implementation of ultrasonic tracking. First, tracking transmissions are interleaved with the acquisition of B-mode images and thus, the effective B-mode frame rate is reduced. Second, it is challenging to achieve an accurate localisation of the needle tip when the signal-to-noise ratio is low. To address these challenges, we present a framework based on a convolutional neural network (CNN) to maintain spatial resolution with fewer tracking transmissions and to enhance signal quality. A major component of the framework included the generation of realistic synthetic training data. The trained network was applied to unseen synthetic data and experimental *in vivo* tracking data. The performance of needle localisation was investigated when reconstruction was performed with fewer (up to eight-fold) tracking transmissions. CNN-based processing of conventional reconstructions showed that the axial and lateral spatial resolution could be improved even with an eight-fold reduction in tracking transmissions. The framework presented in this study will significantly improve the performance of ultrasonic tracking, leading to faster image acquisition rates and increased localisation accuracy.**

***Index Terms*— Ultrasonic needle tracking, interventional devices, deep learning, in vivo imaging**

This work was supported by the Wellcome Trust (WT101957; 203145Z/16/Z; 203148/Z/16/Z) and the Engineering and Physical Sciences Research Council (EPSRC) (NS/A000027/1; NS/A000050/1; NS/A000049/1; EP/L016478/1), by a Starting Grant from the European Research Council (ERC-2012-StG, Proposal 310970 MOPHIM) and by a CMIC-EPSRC platform grant (EP/M020533/1), and by the Academy of Finland Project 336796 (Finnish Centre of Excellence in Inverse Modelling and Imaging, 2018–2025) as well as Project 338408. A.L. David is supported by the UCL/UCLH NIHR Comprehensive Biomedical Research Centre.

E. Maneas, E.J. Alles, and A.E. Desjardins are with the Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University College London, London W1W 7TY, U.K. and with the Department of Medical Physics and Biomedical Engineering, University College London, London WC1E 6BT, U.K. (e-mail: efthymios.maneas@ucl.ac.uk.)

A. Hauptmann is with the Research Unit of Mathematical Sciences, University of Oulu, Oulu FI-90014, Finland, and with the Department of Computer Science, University College London, London WC1E 6BT, U.K.

W. Xia, T. Vercauteren, and S. Ourselin are with the School of Biomedical Engineering and Imaging Sciences, King's College London, London SE1 7EH, U.K.

A.L. David is with the Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University College London, London W1W 7TY, U.K., the Institute for Women's Health, University College London, London WC1E 6HX, U.K., and with the NIHR UCLH Biomedical Research Centre, London W1T 7DN, U.K.

S. Arridge is with the Department of Computer Science, University College London, London WC1E 6BT, U.K.

I. Introduction

Ultrasound imaging is frequently used for real-time guidance of minimally-invasive percutaneous procedures in interventional pain management and regional anaethesia [1], interventional oncology [2] and fetal medicine [3]. During these procedures, a needle is inserted into the body and guidance is achieved through alignment of the needle tip with the ultrasound imaging plane. Deviations from the imaging plane and uncertainties about the location of the needle tip can lead to significant complications such as damage of the nerves and pneumothorax during nerve block insertions [4] and, miscarriage or preterm birth during umbilical cord blood sampling [3] and multifetal pregnancy reduction, where needle tip movements with high precision are required. In fetal medicine, an application considered in this study, percutaneous ultrasound-guided uterine access can be particularly challenging when the mother is obese, when there is an amniotic volume that is less than expected for the gestational age (oligohydramnios), and when the practitioner is inexperienced.

Several approaches have been proposed to improve needle localisation and visualisation that can be classified to image processing, modifications to the needle to make it more echogenic and integration with external sensors, changes to ultrasound formation, motion analysis and machine learning [5]. Instrumented ultrasonic tracking involves the integration of a medical device into a needle or catheter which, depending on the configuration, can receive/transmit ultrasound pulses in concert with ultrasound transmission/reception by an external ultrasound imaging probe. Such instrumented ultrasonic tracking methods have received significant academic [6]–[11] and commercial [12] attention.

In this study, we utilise a custom, newly-developed instrumented ultrasonic tracking system [6] that relies on reception of ultrasound pulses by a fibre-optic hydrophone (FOH) integrated into a needle. Tracking transmissions from the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2021.3126530, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

ultrasound probe are interleaved with the acquisition of B-mode ultrasound images. The received ultrasound pulses are then reconstructed to form the tracking image.

Two particular challenges are considered with this implementation of ultrasonic tracking. First, if transmissions performed for tracking are distinct from the time periods used for transmission/reception in B-mode, they can reduce the effective B-mode imaging frame rate. Second, it can be challenging to obtain accurate localisation of the needle tip if the signal-to-noise (SNR) ratio of the tracking image is low. Averaging over multiple tracking images reduces the frame rate and can be confounded by movement artifacts.

Recently, with the advent of Deep Learning (DL) [13], convolutional neural networks (CNNs) have been applied to accelerate reconstruction and to improve image quality in many different medical imaging modalities [14]–[24]. Here, we hypothesised that a CNN is suitable for processing of ultrasound tracking images: to maintain the resolution and improve needle localisation when reconstruction is performed with fewer transmissions for each tracking image.

Our contributions in this study can be summarised as follows. We developed a DL framework based on convolutional neural networks for instrumented ultrasonic tracking that can be reliably employed for *in vivo* applications. Our approach can be separated into four components: First, we formulated ultrasonic tracking as an image enhancement problem in the image domain considering up to eight-fold subsampling of the channel data prior to reconstruction. Second, we developed a realistic simulation pipeline to generate synthetic training, validation and testing data to characterise the performance of needle visualisation and localisation of the trained network. Third, we trained the network solely on synthetic data and evaluated its performance on unseen synthetic testing data and *in vivo* data obtained from a preclinical fetal sheep model.

This paper is organised as follows. Initially, conventional reconstruction for instrumented ultrasonic tracking and formulation as a learned image enhancement problem is introduced. Next, the simulation pipelines to generate synthetic data are described, and the experimental ultrasonic tracking system used to acquire *in vivo* data is presented. We describe the evaluation metrics and we present quantitative results for synthetic and *in vivo* data. Finally, we discuss the findings and potential limitations of the current approach and provide an outlook for future steps.

The application of Deep Learning to ultrasound imaging [25], [26] and to related modalities [27]–[30] is a burgeoning field. Deep neural networks have been proposed to enhance B-mode ultrasound images to improve interpretation of anatomical structures [31] or to reconstruct images directly from channel data without beamforming [32]–[35]. In the context of enhanced needle visualisation and localisation, there is a wide variety of machine learning approaches in the literature [36], [37]. The majority of these approaches rely on identifying the location of the needle tip in B-mode images that do not involve information from external sensors. In the study of Mwikirize et al., a region-based CNN was used for needle detection in 2D B-mode images [38]. The localisation accuracy was significantly improved when temporal information was included in the network architecture [39]. In a follow-on study, needle localisation was achieved using patch classification and semantic segmentation from extracted 2D orthogonal images of a 3D volume [40]. Segmentation for multiple needle localisation during prostate brachytherapy was achieved using U-net variations [41], [42].

The computational problem of point source localisation in photoacoustic imaging is directly related to the reconstruction of instrumented ultrasonic tracking images obtained with transmissions from individual transducer elements, via reciprocity principle (*c.f.* Sec. II-A) [6]. Deep Learning methods have shown promise with identifying point sources in photoacoustic imaging [43]. Allman et al. [44] proposed a method to localise point targets and remove reflection artifacts that have a similar appearance using a method based on region-based CNN. Similarly, Johnstonbaugh et al. [45] proposed to use an encoder-decoder CNN to localise point sources in the presence of strong optical scattering in deep tissue. Finally, for localisation of up to four point sources, Yardano et al. [46] proposed a deep neural network consisting of a shared encoder and two parallel decoders. We note that all the previous approaches were based on the localisation of the point sources directly from the channel data without applying beamforming.

To the best of our knowledge, this study is the first to explore the use of DL for ultrasonic tracking using integrated sensors, with a novel framework based on CNNs for training, evaluation and application to *in vivo* ultrasonic tracking images.

II. Methods

A. Ultrasonic tracking

1) Conventional reconstruction for ultrasonic tracking: The instrumented ultrasonic tracking paradigm used here includes a needle with an embedded ultrasound receiver (i.e. FOH) that detects transmissions from an ultrasound imaging probe (Fig. 1; Application). For the formation of a 2D ultrasonic tracking image, each transducer element emits a signal sequentially, where the FOH within the imaging plane records the time of flight from the transducer element to the needle tip. The recorded time-of-flight data is then combined for all elements and forms the tracking measurement (i.e. channel data) given by $g(r,t)$, where r and t denotes the transducer element location and time, respectively. Using the principle of reciprocity, the ultrasound signal generation is then interpreted as a pressure wave emitted from a point source p_0 located at the needle tip such that:

$$Ap_0 = \tilde{g}, \tag{1}$$

$$\tilde{g} = g + \delta g, \tag{2}$$

where A models the ultrasound wave propagation and δg is the noise term. A conventional tracking image, is an approximation of p_0 that is considered here as a low-quality tracking image p^{LQ} due to limited-view ultrasound detection, noise and undersampling, can then be recovered by Fourier beamforming as:

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2021.3126530, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

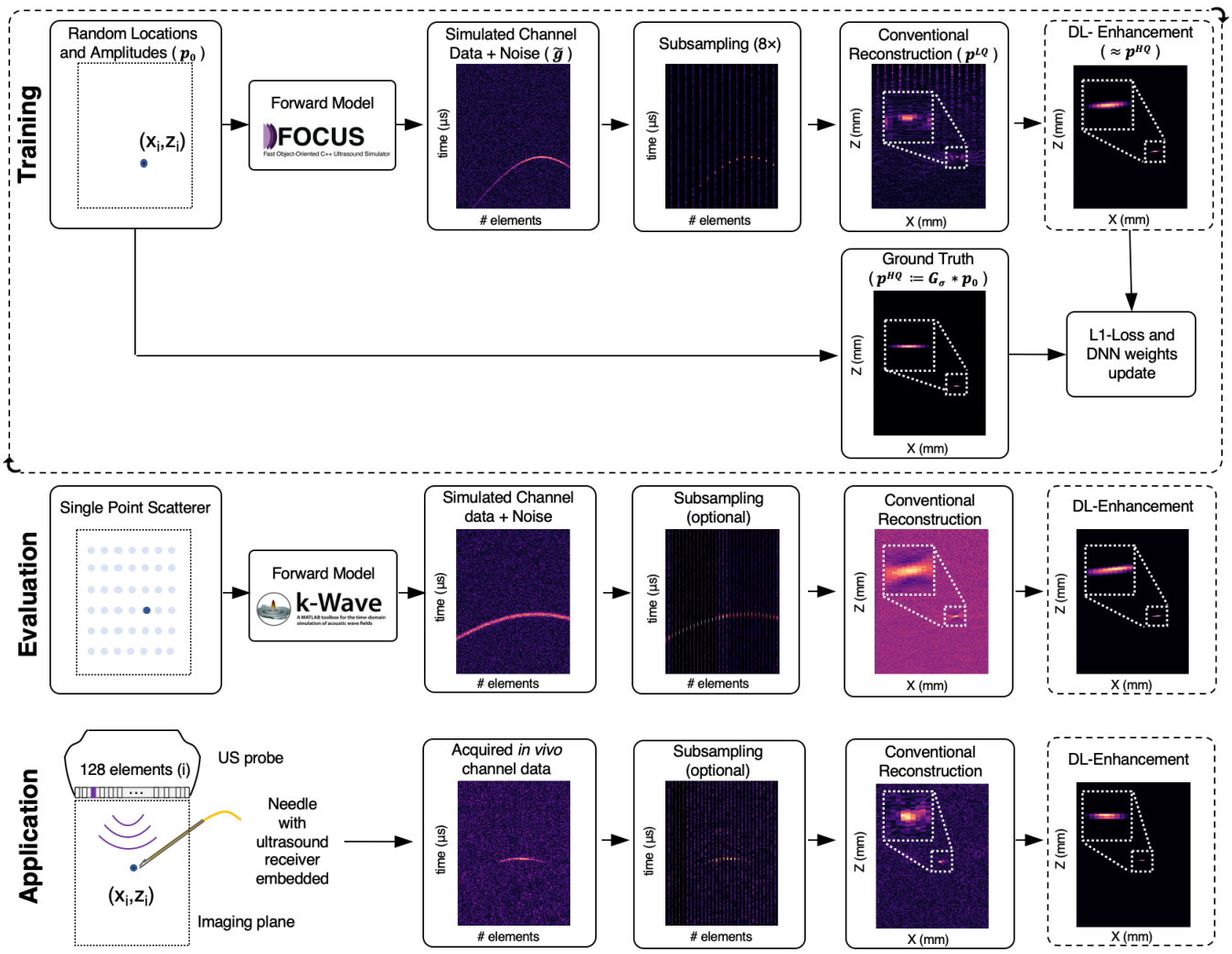

Fig. 1. Main components of the proposed framework for deep learning (DL)-based enhancement of instrumented ultrasonic tracking images. Top row: Synthetic training data generation using FOCUS for modelling of ultrasonic tracking data acquisition. Middle row: Synthetic testing data generation using k-Wave for acoustic wave propagation to evaluate the performance of the trained network. Bottom row: application of the trained network with synthetic data to unseen experimental *in vivo* ultrasonic tracking data obtained from a preclinical model. Note: Envelope detection has been performed to the channel data only for illustration purposes. Subsampling was performed in the elements direction.

$$\mathcal{F}^{\dagger}\tilde{g} = p^{\mathrm{LQ}}. \tag{3}$$

Here, $\mathcal{F}^{\dagger}$ performs the reconstruction in the frequency domain using fast Fourier transforms (FFT) and relies on a periodically spaced transducer array [47]. In this study, we use the computationally efficient implementation of this algorithm found in the k-Wave toolbox [48].

2) Learned image enhancement: The number of tracking measurements and the measurement noise limit how accurately a point source p_0 (i.e. the needle tip) can be resolved in conventional image reconstructions. Here, p_0 is a binary image with the value of one at only one pixel corresponding to the needle tip, and zero elsewhere. From that standpoint, a higher number of tracking measurements is beneficial for increasing ultrasonic tracking image quality. However, this increase comes at the expense of acquisition time. Furthermore, even with many tracking measurements, the limited-view geometry limits the accuracy with which locations deep within the target can be resolved. Therefore, a high-quality image needs to be recovered to accurately determine the needle tip position. This leads to the image enhancement problem considered in this study.

We formulate the problem of recovering the accurate ultrasonic tracking image p^{HQ} from the conventional low-quality reconstruction p^{LQ} as a learning problem. Our aim is to train a CNN, Λ, with parameters θ, such that $\Lambda_\theta(p^{\mathrm{LQ}}) \approx p^{\mathrm{HQ}}$. That means we have to find an optimal set of parameters θ^* that is:

$$\theta^* = \arg\min_{\theta} \sum_{i=1}^{N} \|\Lambda_\theta(p_i^{\mathrm{LQ}}) - p_i^{\mathrm{HQ}}\|_1. \tag{4}$$

We choose to train Λ using the $L1$-loss as it is more robust to outliers [49]. For p^{LQ}, we consider a conventionally reconstructed image which can occur from full or subsampled channel data. To synthesize an ideal image of the needle tip, we define p^{HQ} as the point source p_0 convolved with a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2021.3126530, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

Gaussian kernel:

$$p^{\mathrm{HQ}} := G_\sigma * p_0. \tag{5}$$

where G_σ is a 2D Gaussian smoothing kernel with standard deviation σ. Using a Gaussian kernel is more stable than simply learn a single pixel. As such, we can define the resolution that we are aiming to achieve depending on the kernel size.

A major part of our proposed framework relies on the generation of realistic *in silico* training data, such that the trained network can be applied directly to the desired *in vivo* application without the need for retraining. Thus, the training data generation is of essential importance, which we describe in detail below.

B. Generating realistic synthetic data

1) Training and validation data: For the generation of training and validation data, we used the FOCUS ultrasound simulator [50], [51] (Fig. 1; top row), which is particularly attractive as it uses a computationally efficient numerical implementation of the spatio-temporal impulse response of piston transducers in a homogeneous medium (modelled as water without attenuation and a uniform sound speed of 1500 m/s). To model the tracking transmissions from the ultrasound probe, 128 rectangular planar piston transducers (measuring 0.3 mm laterally by 6 mm elevationally to match the experimental setup, *c.f.* Sec. II-C) were distributed equidistantly across the transducer aperture (measuring 38.4 mm laterally). These transducers were individually excited, where the temporal profile of the transducer excitation was modelled as a tone burst, the centre frequency (6.5 MHz) and duration (3 cycles) of which were adjusted to match the power spectrum of the experimental data. For ultrasound reception, the FOH was modelled as an ideal point receiver with an infinitesimal spatial extent and a uniform frequency response, which corresponds to a single voxel measuring 0.3 mm. The ultrasound time traces detected by the FOH for each transducer at the position (r, t) were computed sequentially to form the channel data g, resulting in a 2D array of 128 x 2048 samples with a corresponding sampling frequency of 100 MHz.

The training set is given by $N = 1000$ and the validation set is given by $N = 200$ high- and low-quality image pairs $\{p_i^{\mathrm{HQ}}, p_i^{\mathrm{LQ}}\}_{i=1}^{N}$ in which, the point source locations ranged from -10 to 10 mm laterally and from 10 to 25 mm axially.

To create p^{LQ}, we started with the synthetic channel data g from 128 transducer elements and we added Gaussian noise δg with zero mean and standard deviation ($\sigma = 0.06$) estimated from experimental preclinical *in vivo* data using a region of interest (3×12 mm) that contained only noise. We then subsampled the channel data 8×, i.e. we only retained data from 16 equidistantly spaced transducer elements and added zeros to the remaining 112 transducer elements. The standard deviation of the noise of the full-channel generated data was varied between 0.5× and 3× of that measured experimentally before subsampling to generate a wide range of training data. The experimental channel data had an average SNR of 17.5; the full-channel generated data had SNR values that ranged from a minimum of 10.5 to a maximum of 56.1. Finally, to obtain p^{LQ}, the noisy and subsampled channel data was reconstructed as in Eq. 3, followed by envelope detection via the Hilbert transform. Using the Hilbert transform, we restricted the solution space from 0-1 so that convergence is achieved faster during training.

The reference ground truth images, p^{HQ}, were generated following Eq. 5 by taking the point FOH receiver location p_0 (i.e. the needle tip) and convolving it with an anisotropic Gaussian kernel ($\sigma = [4, 2]$ pixels), which resulted in FWHM of 0.14 mm and 1.45 mm for the axial and lateral resolution respectively. We chose an anisotropic Gaussian kernel with a larger width in the axial dimension, to account for the unequal axial and lateral sampling rates in the 2D tracking images.

2) Testing data: We generated the testing data with the k-Wave toolbox. Unlike FOCUS, k-Wave is a full-wave method based on a pseudo-spectral approach [52], which requires the entire volume between a source and detector to be discretised. To limit the computational requirements, k-Wave simulations were performed in 2D at the expense of a small reduction in accuracy, which was used to both avoid training bias and assess the robustness against small inaccuracies in the numerical model. We directly simulated Eq. 1 that represents an equivalent but reciprocal tracking experiment to the training data. To generate channel data, ultrasound transmission of a point source, modelled as a single voxel after spatial smoothing, resulting in a full-width at half-maximum (FWHM) of 0.74 mm, was propagated through a homogeneous medium and detected by a linear array with 128 rectangular transducers. Acoustic wave propagation was performed on an isotropic grid with 0.3 mm spacing (measuring 60 mm axially by 38.4 mm laterally) and sampling frequency of 50 MHz to limit computational requirements. As a further test of robustness and bias avoidance, a uniform sound speed of 1540 m/s was used. Subsequently, reconstruction using the FFT-based implementation of Eq. 3 and envelope detection were performed as described previously.

To create the synthetic testing set, a point source was translated along a grid laterally from 7.25 to 32.5 (step size: 2.5 mm) and axially from 5 to 55 mm (step size: 5 mm) respectively. Additionally, Gaussian noise was added to the channel data prior to reconstruction, resulting in different SNR values, which were varied from 3 to 21 in steps of 3. The SNR was estimated using the maximum amplitude value of full channel data divided by the standard deviation of a background region that contained no appreciable signal. The channel data were then subsampled 2×, 4× and 8×, and reconstruction was performed using Eq.3. Subsampling was performed with zeroing, as previously described. For each subsampling case, 847 images were generated that were only used for inference and excluded from the training set.

Although the needle itself was not modelled explicitly in this study, variations in its position within the grids used for training and testing were captured implicitly; it was assumed that these positions were all attainable by the needle tip.

C. Experimental setup and data

1) Ultrasonic tracking system: The ultrasonic tracking system consisted of a clinical ultrasound scanner (Sonix MDP,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2021.3126530, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

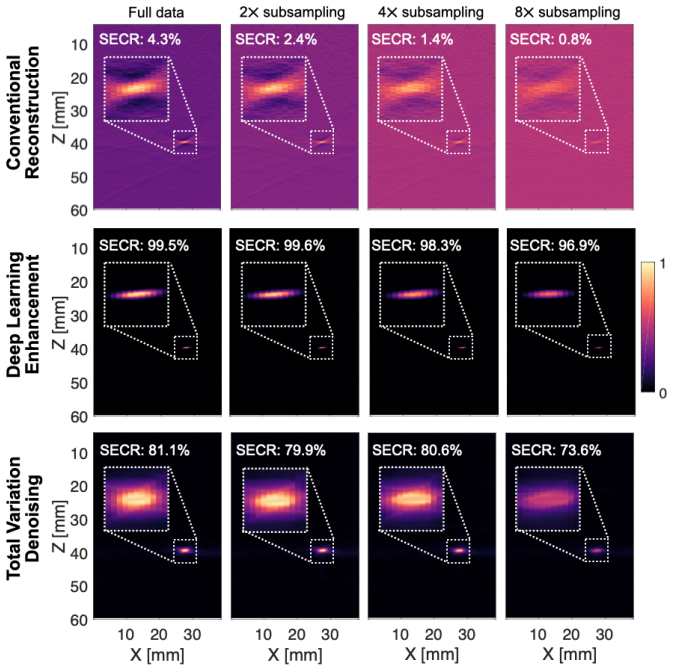

Fig. 2. Evaluation of the trained network using synthetic testing data generated with k-Wave. Conventionally reconstructed (top row), deep learning-enhanced (middle row) and total variation denoised (bottom row) images obtained using full or subsampled channel data. The point source (i.e. needle tip) is located at 40 mm depth. Full channel data had an initial SNR of 3.

Analogic Ultrasound, Richmond, BC, Canada) with a linear array probe (L9-4/38, 128 elements, 9–4 MHz Bandwidth, 300 μm pitch, Analogic Ultrasound, Richmond, BC, Canada), and a FOH for reception of ultrasound pulses. The FOH comprised of a thin film Fabry-Pérot interferometer placed at the distal end face of a single-mode fibre and was integrated into a 20-gauge needle cannula. The ultrasound scanner was operated in research mode, which allowed for control of the transducer element transmissions, and corresponding output triggers that were used for synchronising FOH signal acquisitions with respect to those transmissions. During tracking, each of the 128 transducer elements was excited individually to emit a divergent pressure field. The received signals were reconstructed with the FFT-based implementation of Eq. 3 to obtain the tracking image followed by envelope detection via the Hilbert transform. An imposed delay to limit the data transfer rate between B-mode and tracking acquisitions resulted to an effective frame rate of 1 Hz. The ultrasonic tracking system and the needle geometry in this study were chosen for their relevance to a broad range of percutaneous procedures. More details about the implementation of the ultrasonic tracking system can be found in [6].

2) In vivo data: To evaluate the performance (*c.f.* Sec. II-F) of the trained network for *in vivo* clinically-realistic conditions, insertion of a 22 Gauge needle (Becton Dickinson, UK) into the heart of a fetal sheep in mid-gestation under ultrasound guidance was performed. The procedure was conducted in accordance with the U.K. Home Office regulations and the Guidance for the Operation of Animals (Scientific Procedures) Act (1986). Ethics approval was provided by the joint animal studies committee of the Royal Veterinary College and the University College London, UK.

D. Network implementation

For the network implementation, we used a simple modification of the established residual neural network (ResNet) [53] architecture following [49] due to its recent success in similar tasks such as super-resolution. The particular architecture for our study consists of 16 residual blocks, each consisting of 2 convolutional layers with width of 64 channels and 3×3 convolutional kernels and biases, with a rectified linear unit as nonlinearity between the 2 convolutional layers. We note that convolutions act locally and we found that training on smaller patches instead of the full image size can achieve faster convergence. Thus, we trained with patches of 64×64 pixels, which roughly equals the receptive field of our network, that were randomly extracted from the pairs in the training set. Implementation of our network was performed in Python using TensorFlow v1.13 and Keras v2.2.4. Training was done for 80 epochs that maximised the peak SNR in the validation set using ADAM optimizer [54] (initial step size: 0.001; minibatch size: 16 patches) and an Nvidia GTX 1080Ti GPU with 12 GB memory.

E. Comparison method

We further evaluated the performance of the proposed DL framework by comparing it to a classic analytical method for image enhancement, namely total variation (TV) denoising [55]. That is, we seek a reconstruction as the minimiser of the following penalty functional

$$p^* = \arg\min_p \|p - p^{\mathrm{LR}}\|_2^2 + \alpha\|\nabla p\|_1. \qquad (6)$$

Here, the first term ensures closeness to the initial reconstruction and the second term, the total variation penalty, promotes sparsity in the gradients, i.e. it favours piece-wise constant reconstructions, while edges are preserved. The parameter $\alpha > 0$ balances both terms, where a larger parameter will enforce higher regularity in the reconstruction. A uniform value ($\alpha = 1.0$) that performed best on average for all subsampling cases and SNR values was chosen.

We chose a total variation image enhancement as a comparison method for two reasons. First, the obtained reconstructions exhibit comparable features as the data-driven approach, achieving a high SECR. Second, computationally efficient algorithms are available [56] to solve Eq. 6. In our case, a proximal gradient descent scheme was used.

Finally, the computation times of applying the trained network (i.e. inference) and TV denoising method were calculated by averaging 100 instances.

F. Evaluation protocol and metrics

Three metrics were used to evaluate the performance of the trained network. First, we measured the spatial (i.e. axial and lateral) resolution with which the needle tip could be visualised. A bounding box (3 × 12 mm, axial & lateral) centred around the maximum amplitude of the needle tip

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2021.3126530, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

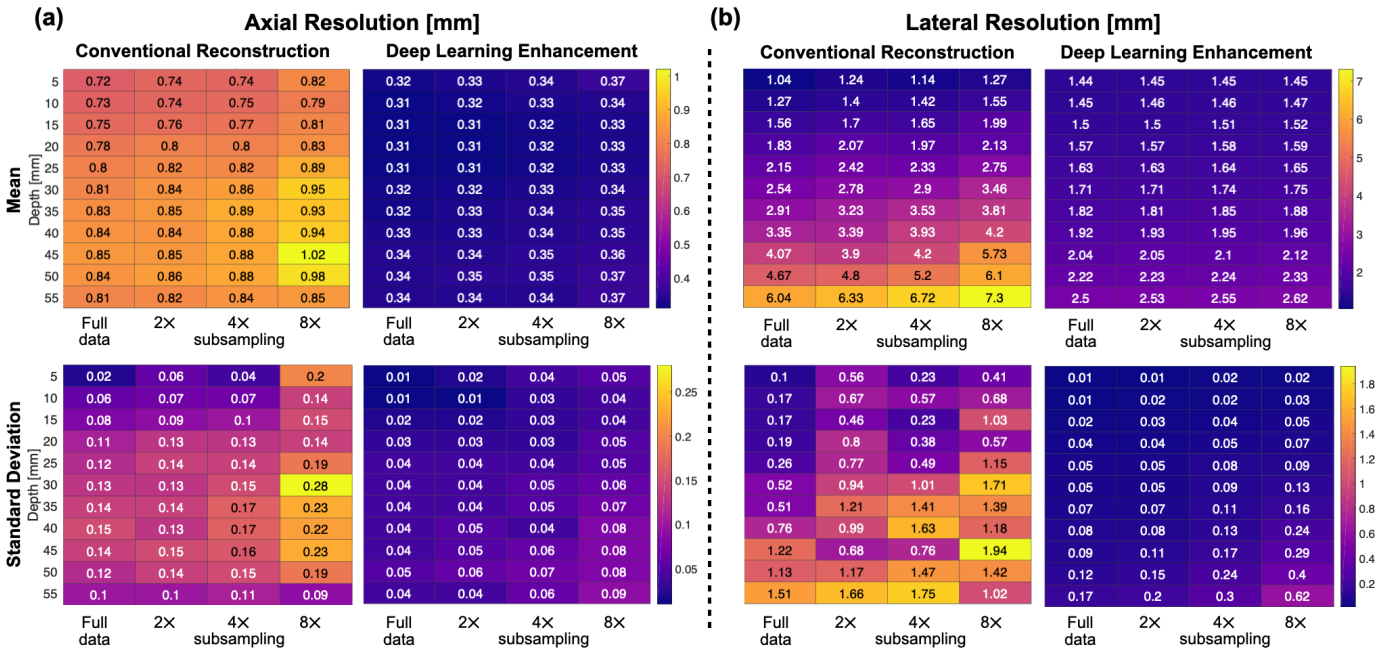

Fig. 3. Spatial resolution assessment: Average and standard deviation values of axial (a) and lateral (b) resolution that correspond to conventional reconstructions (first column) and deep learning-enhanced images (second column) from the synthetic testing dataset. High consistency was observed for both axial and lateral resolution when the trained network was applied to conventionally reconstructed images.

obtained from the ground truth image was used, and maximum amplitude axial and lateral profiles were obtained to calculate the corresponding FWHM values.

Second, as a measure of the extent to which the ultrasonic tracking image was spatially localised (i.e. the amount of the image energy concentrated around the actual needle tip is quantified), we defined the signal energy concentration ratio (SECR) as:

$$\mathrm{SECR}(p) = \frac{\sum_i (p_i^{\mathrm{source}} - \overline{p}^{\mathrm{back}})^2}{\sum_i (p_i - \overline{p}^{\mathrm{back}})^2}, \tag{7}$$

where p^{source} is a 3×3 mm bounding box centred around the maximum amplitude of the needle tip obtained from the ground truth image, $\overline{p}^{\mathrm{back}}$ is the mean value of a region (4.5×18.5 mm) that contains only background noise and is kept fixed for all the images, and p corresponds to either p^{LQ} or p^{HQ}, as required. The SECR values, which are bounded to the interval $[0, 1]$, are reported as percentages.

Third, the localisation error of the needle tip in the synthetic testing dataset was measured. The needle tip deviation is calculated as the 2D Euclidean distance between the true needle tip location (obtained from the ground truth image) and the needle tip location obtained using the maximum intensity. For each needle tip location, the mean, standard deviation and root-mean-square error (RMSE) are calculated.

III. Results

A. Synthetic data

A point source (i.e. needle tip) from the synthetic testing set which is located at 40 mm depth with initial full channel data SNR 3 is shown in Fig. 2. With conventional reconstruction, a worsening of the spatial resolution as we move deeper into the tissue is a typical phenomenon due to limited-angle viewing of the ultrasound probe. Reconstruction from subsampled channel data distorts the spatial resolution and increases the background noise as fewer tracking transmissions are used (Fig. 2; top row). Passing these images through the trained network, the background noise is filtered-out, and the image quality and spatial resolution are clearly improved for all subsampling scenarios (Fig. 2; middle row). Filtering these images using a classical TV denoising method (Fig. 2; bottow row) results in enhancement of the image quality, although the spatial resolution is increased. Particularly, the SECR varies from 4.3 to 0.8% when conventional reconstruction is used, which is significantly increased from 99.6 to 96.9% when the trained network is applied. TV denoising results to an improvement of SECR from 81.1 to 73.6%. In contrast, the spatial resolution was improved after post-processing using the trained network. In particular, for fully sampled data, the axial resolution was decreased from 1.01 to 0.31 mm and the lateral resolution from 4.79 to 1.92 mm - to update them, respectively. Compared to the network, TV denoising resulted in higher axial and lateral resolution values (i.e. 1.18 mm and 2.67 mm). With 8× subsampling, the DL-enhanced image maintained improved spatial resolution (axial: 0.28 mm; lateral: 1.66 mm), although it was not possible to measure it from the conventional image due to poor SNR and despite the large bounding box chosen (*c.f.* Sec. II-F). For the same reasons, no spatial resolution measurements were possible from any subsampled conventionally reconstructed images. Further improvements were observed for less severe subsampling, despite training being performed exclusively on 8× subsampled data.

To quantify the performance of the trained network in the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2021.3126530, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

TABLE I

EVALUATION OF THE DEEP LEARNING FRAMEWORK USING THE SYNTHETIC TESTING DATASET AND THE SIGNAL ENERGY CONCENTRATION RATIO (SECR) AS A PERFORMANCE METRIC. CR: CONVENTIONAL RECONSTRUCTION, CR+DL: CONVENTIONAL RECONSTRUCTION AND DEEP LEARNING-BASED POST-PROCESSING, TV: TOTAL VARIATION.

SECR range (%)	Full data			2× Subsampling			4× Subsampling			8× Subsampling		
	CR	CR+DL	TV	CR	CR+DL	TV	CR	CR+DL	TV	CR	CR+DL	TV
0 – 20	29.4	0	0	42.86	0	0	67.65	0.24	0	97.17	1.77	0
20 – 40	29.63	0	0	35.77	0	0	32.35	0	0	2.83	0	0.24
40 – 60	28.93	0	9.09	21.37	0	9.09	0	0	9.56	0	0.24	10.39
60 – 80	12.04	0	25.86	0	0	26.21	0	0	26.68	0	0.59	42.38
80 – 100	0	100	65.05	0	100	64.7	0	99.76	63.75	0	97.4	46.99

TABLE II

LOCALISATION ERROR USING THE SYNTHETIC TESTING DATASET. NEEDLE TIP DEVIATION IS CALCULATED AS THE 2D EUCLIDEAN DISTANCE BETWEEN THE TRUE TIP LOCATION (OBTAINED FROM THE GROUND TRUTH IMAGE) AND THE TIP LOCATION OBTAINED USING THE MAXIMUM INTENSITY. FOR EACH LOCATION, THE MEAN, STANDARD DEVIATION (STD) AND ROOT-MEAN-SQUARE ERROR (RMSE) ARE CALCULATED. RMSES FOR LATERAL ($\mathbf{RMSE_x}$) AND AXIAL ($\mathbf{RMSE_z}$) DIMENSIONS ARE PROVIDED. CR: CONVENTIONAL RECONSTRUCTION, CR+DL: CONVENTIONAL RECONSTRUCTION AND DEEP LEARNING-BASED POST-PROCESSING.

	Full data		2× Subsampling		4× Subsampling		8× Subsampling	
	CR	CR+DL	CR	CR+DL	CR	CR+DL	CR	CR+DL
MEAN (mm)	0.044	0.085	0.055	0.077	0.552	0.082	2.224	0.372
STD (mm)	0.075	0.133	0.085	0.124	4.630	0.125	9.296	3.650
RMSE (mm)	0.087	0.158	0.101	0.146	4.660	0.149	9.553	3.666
RMSE_x (mm)	0.079	0.155	0.091	0.142	1.758	0.146	3.069	1.276
RMSE_z (mm)	0.036	0.031	0.044	0.031	4.316	0.031	9.047	3.437

spatial resolution reduction, we used the synthetic testing set and measured the FWHM of each conventionally reconstructed and DL-enhanced image. We should note that the images, in which we were able to compute the FWHM, were grouped and averaged according to their depth location (Fig. 3). An image was excluded from the analysis if either the axial or the lateral resolution couldn't be measured due to low SNR. Therefore, for conventional reconstruction, 1.89% of fully sampled images, and 6.73%, 21.49% and 42.38% of 2,4,8$\times$ subsampled images were excluded. On the other hand, for DL-enhanced images, 0.24% and 1.77% of only 4,8$\times$ subsampled images were excluded, which corresponded to images with SNR 3. All images were included when full channel and $2\times$ subsampled data were used. An improvement in axial resolution was observed when the trained network was applied (Fig. 3a; top row). Additionally, higher consistency among the resolution values as evident by lower standard deviations across the different subsampling scenarios was noticed with a maximum standard deviation value of 0.09 mm in the DL-enhanced images (Fig. 3a; bottom row). For the lateral resolution (Fig. 3b; top row), a reduction was achieved after the trained network was applied; however, it was smaller than the reduction in axial resolution values. High consistency was maintained with a maximum standard deviation value of 0.62 mm, which occurred at $8\times$ subsampling and at the highest depth of 55 mm in the DL-enhanced images (Fig. 3b; bottow row).

SECR was used to quantify the enhancement of the trained network and the TV denoising on the image quality (Table I). Overall, by using the trained network, the image quality was greatly improved and the vast majority of post-processed images had an SECR of >97.5%. It was observed that low SECR values <20% in the DL-enhanced images were associated with unsuccessful identification of the needle tip and no spatial resolution measurements were possible. In table I, all of the DL-enhanced images obtained from full channel data reconstructions had a SECR value within the range of $80-100\%$, while no conventional reconstruction had such a value. Additionally, a significant improvement in SECR was noticed for the subsampled cases. When $2\times$ to $8\times$ subsampling was applied, $100, 99.76$ and 97.4% of DL-enhanced data had a SECR value of $80-100\%$, respectively. Low values <20% of SECR in DL-enhanced images corresponded to high depths and SNR of 3. In contrast, $2\times$ and $4\times$ subsampled conventional images had no SECR value >60% and >40%, while 97.17% of $8\times$ subsampled conventional images had an SECR value <20%. TV denoising improved the image quality of the conventionally reconstructed images and provided comparable results with the DL-enhanced images. Although the SECR values were lower compared to the values derived from DL-enhanced images, no TV denoised image had an SECR value from $0-20\%$. By using 2D tracking images of 1948×128, inference of the trained network was performed in 0.15 s, while the processing time for TV denoising was 2.15 s, respectively.

Finally, using the synthetic testing dataset and having access to the ground truth location of the needle tip, the tracking accuracy was evaluated (Table II). In fully sampled and $2\times$ subsampled data, the localisation error exhibited similar performance in conventionally reconstructed and DL-enhanced images; the maximum mean error was 0.085 ± 0.133 mm and the RMSE 0.158 mm. However, when higher subsampling was applied (i.e. $4\times$ and $8\times$), the localisation error of the conventional reconstructions was significantly increased. With $4\times$ subsampling, conventional reconstructions had a mean error of 0.552 ± 4.630 mm while the DL-enhanced images

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TUFFC.2021.3126530, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

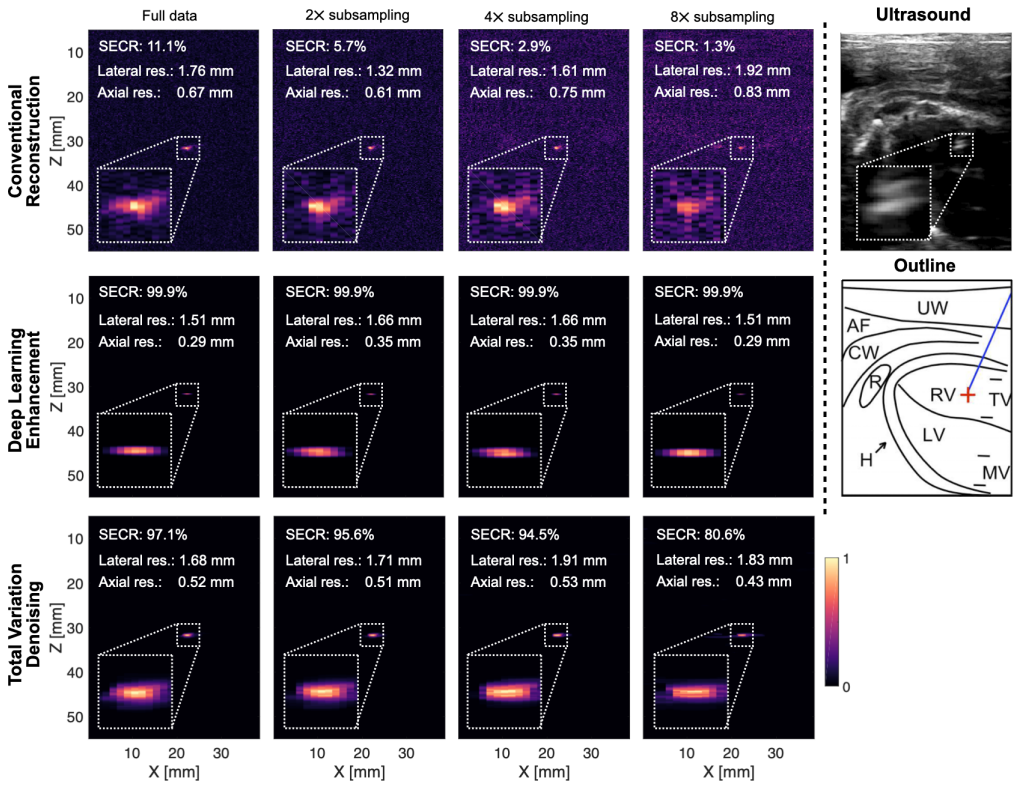

Fig. 4. Application of the trained network to process an *in vivo* needle tracking insertion. Conventional reconstructed, deep learning-enhanced and TV denoised images with subsampling scenarios are shown. The effect on the spatial resolution and image quality improvement of the DL-based post-processing is quite prominent and outperforms a traditional TV denoising method.

maintained a lower mean error of 0.082 ± 0.125 mm, which is similar to the one obtain with fully sampled images. Further to this, RMSE had a similar value (0.149 mm) to the fully sampled data compared to $4\times$ subsampling, which was a 3-fold higher (4.660 mm). In conventional reconstructions, the RMSE in the axial dimension (i.e. $\mathrm{RMSE_z}$) seemed to have the bigger contributions in such an error increase (4.316 mm), compared to the $\mathrm{RMSE_x}$ obtained in the lateral dimension (1.758 mm). With $8\times$ subsampling, the mean localisation error increased for both conventionally and DL-enhanced images; however, after the trained network was applied, the mean error was reduced from 2.224 ± 9.296 mm to 0.372 ± 3.650 mm. Similarly, the RMSE was reduced from 9.553 to 3.666 mm when the images were passed through the trained network.

B. In vivo data

As a demonstration of the network's performance on unseen *in vivo* data, ultrasonic tracking data from a 22 Gauge needle with an integrated FOH placed percutaneously into the right ventricle of a fetal sheep under general anaesthesia were acquired (Fig. 4). The needle tip was visible in the B-mode ultrasound images, although obtaining its locations accurately by visual inspection alone was challenging due to the depth of the fetus within the uterine cavity of the ewe. When the trained network was applied, prominent improvements in spatial resolution and removal of background noise, as compared to conventional reconstruction, were observed (Fig. 4). In particular, the axial resolution was improved by at least 1-fold; furthermore, for all full and subsampled data scenarios, the SECR reached 99.99%. With TV denoising, there were improvements in the axial resolution relative to the conventional reconstructions. However, such improvements were lower compared to the network. Similarly, TV improved the image quality in conventional reconstructions, although the achieved SECRs were lower compared to the applied network.

IV. Discussion

We developed a DL framework that comprised a CNN and synthetic training data to process reconstructed ultrasonic needle tracking images. This was, to our knowledge, the first application of DL to processing *in vivo* ultrasonic needle tracking images. Using the framework developed here, constancy of axial and lateral resolution across depth was achieved with subsampling up to $8\times$ fewer transmissions for tracking. This result will lead directly to faster ultrasound imaging using frames that are interspersed with the more rapidly-acquired tracking frames.

The use of synthetic data is beneficial within the context of ultrasonic tracking, as it removes the burden of acquiring experimental data with manual annotation. Apart from being time-consuming, manual annotation introduces a further challenge to accurately identify the needle tip when there can be low visibility and uncertainty about its true location; defining a ground truth in ultrasonic tracking methods remains an open problem [5].

In contrast to other studies that localise the needle tip or reconstruct the image directly from channel data, we chose to

formulate the learning task in the image domain for several reasons. First, the SNR of the reconstructed image is expected to be higher. This follows from the fact that the noise amplitude after reconstruction is preserved, but the reconstructed point source (i.e. needle tip) is a sum over the measured signals [52]. Second, by formulating the learning task in image domain, we benefit from essential properties of a CNN, namely translation invariance and independence of image size.

Compared to conventional tracking image reconstruction [6], [9], the framework presented here has strong potential to improve needle identification. Using synthetic testing images from $8\times$ subsampled channel data, the percentage of cases in which the FWHM could not be measured due to poor SNR was significantly reduced. Processing of *in vivo* tracking images using Deep Learning outperformed the classical TV denoising framework, both in terms of image quality (spatial resolution and SECR) and computation time.

There are several ways in which this work can be extended. First, the generation of synthetic data assumed a homogeneous non-attenuating medium with single sound speed for ultrasound wave propagation and detection. Variations in the acoustic attenuation and the sound speed of the imaged medium to improve robustness could readily be incorporated using k-Wave, and into the synthetic training dataset. Second, the current framework may enhance out-of-plane signals that could correspond to a false needle tip position. Further studies to assess the impact of reflection and diffraction from hyperechoic structures such as bone [44] and brachytherapy seeds [57] are required. Third, the robustness of needle tip tracking could be improved by incorporating multiple tracking frames, for instance by recurrent neural networks or a filtering approach in a Bayesian framework. Fourth, the current framework could be extended to localise multiple point sources [44] for systems with multiple ultrasonic sensors. Finally, to calculate the localisation accuracy directly to *in vivo* tracking images, the ground truth of needle tip location data could be obtained with the use of motorised stages [9].

The Deep Learning framework presented in this study has strong potential to improve the frame rate and needle tip identification accuracy in ultrasound tracking. This combination of improvements will have broad applicability across multiple clinical fields, leading to improvements in procedural efficiency and reductions in the risk of complications.

References

[1] K. J. Chin, A. Perlas, V. W. Chan, and R. Brull, "Needle visualization in ultrasound-guided regional anesthesia: challenges and solutions," *Regional Anesthesia & Pain Medicine*, vol. 33, no. 6, pp. 532–544, 2008.

[2] T. Helbich, W. Matzek, and M. Fuchsjäger, "Stereotactic and ultrasound-guided breast biopsy," *European Radiology*, vol. 14, no. 3, pp. 383–393, 2004.

[3] F. Daffos, M. Capella-Pavlovsky, and F. Forestier, "Fetal blood, sampling during pregnancy with use of a needle guided by ultrasound: A study of 606 consecutive cases," *American Journal of Obstetrics and Gynecology*, vol. 153, no. 6, pp. 655–660, 1985.

[4] D. Karakitsos *et al.*, "Real-time ultrasound-guided catheterisation of the internal jugular vein: a prospective comparison with the landmark technique in critical care patients," *Critical Care*, vol. 10, no. 6, pp. 1–8, 2006.

[5] P. Beigi, S. E. Salcudean, G. C. Ng, and R. Rohling, "Enhancement of needle visualization and localization in ultrasound," *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–10, 2020.

[6] W. Xia *et al.*, "In-plane ultrasonic needle tracking using a fiber-optic hydrophone," *Medical physics*, vol. 42, no. 10, pp. 5983–5991, 2015.

[7] X. Guo, B. Tavakoli, H.-J. Kang, J. U. Kang, R. Etienne-Cummings, and E. M. Boctor, "Photoacoustic active ultrasound element for catheter tracking," in *Photons Plus Ultrasound: Imaging and Sensing 2014*, vol. 8943. International Society for Optics and Photonics, 2014, p. 89435M.

[8] X. Guo, H.-J. Kang, R. Etienne-Cummings, and E. M. Boctor, "Active ultrasound pattern injection system (AUSPIS) for interventional tool guidance," *PloS one*, vol. 9, no. 10, p. e104262, 2014.

[9] W. Xia *et al.*, "Looking beyond the imaging plane: 3D needle tracking with a linear array ultrasound probe," *Scientific Reports*, vol. 7, no. 1, pp. 1–9, 2017.

[10] A. Cheng *et al.*, "Photoacoustic-based catheter tracking: simulation, phantom, and in vivo studies," *Journal of Medical Imaging*, vol. 5, no. 2, p. 021223, 2018.

[11] M. Graham *et al.*, "In vivo demonstration of photoacoustic image guidance and robotic visual servoing for cardiac catheter-based interventions," *IEEE transactions on medical imaging*, vol. 39, no. 4, pp. 1015–1029, 2019.

[12] J. Mung, F. Vignon, and A. Jain, "A non-disruptive technology for robust 3D tool tracking for ultrasound-guided interventions," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2011, pp. 153–160.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[14] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb, "Solving inverse problems using data-driven models," *Acta Numerica*, vol. 28, pp. 1–174, 2019.

[15] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.

[16] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Medical Physics*, vol. 44, no. 10, 2017.

[17] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," *IEEE transactions on Medical Imaging*, vol. 37, no. 2, pp. 491–503, 2017.

[18] J. Adler and O. Öktem, "Learned primal-dual reconstruction," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1322–1332, 2018.

[19] A. Hauptmann, S. Arridge, F. Lucka, V. Muthurangu, and J. A. Steeden, "Real-time cardiovascular MR with spatio-temporal artifact suppression using deep learning–proof of concept in congenital heart disease," *Magnetic resonance in medicine*, vol. 81, no. 2, pp. 1143–1156, 2019.

[20] A. Hauptmann *et al.*, "Model based learning for accelerated, limited-view 3D photoacoustic tomography," *IEEE Transactions on Medical Imaging*, 2018.

[21] E. M. A. Anas, H. K. Zhang, J. Kang, and E. Boctor, "Enabling fast and high quality led photoacoustic imaging: a recurrent neural networks based approach," *Biomedical Optics Express*, vol. 9, no. 8, pp. 3852–3866, 2018.

[22] S. Antholzer, M. Haltmeier, and J. Schwab, "Deep learning for photoacoustic tomography from sparse data," *Inverse problems in science and engineering*, vol. 27, no. 7, pp. 987–1005, 2019.

[23] S. Guan, A. A. Khan, S. Sikdar, and P. V. Chitnis, "Limited-view and sparse photoacoustic tomography for neuroimaging with deep learning," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.

[24] M. W. Kim, G.-S. Jeng, I. Pelivanov, and M. O'Donnell, "Deep-learning image reconstruction for real-time photoacoustic system," *IEEE Transactions on Medical Imaging*, 2020.

[25] R. J. Van Sloun, R. Cohen, and Y. C. Eldar, "Deep learning in ultrasound imaging," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 11–29, 2019.

[26] M. Mischi, M. A. L. Bell, R. J. van Sloun, and Y. C. Eldar, "Deep learning in medical ultrasound—from image formation to image analysis," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 67, no. 12, pp. 2477–2480, 2020.

[27] A. Hauptmann and B. Cox, "Deep learning in photoacoustic tomography: current approaches and future directions," *Journal of Biomedical Optics*, vol. 25, no. 11, p. 112903, 2020.

[28] J. Gröhl, M. Schellenberg, K. Dreher, and L. Maier-Hein, "Deep learning for biomedical photoacoustic imaging: A review," *Photoacoustics*, vol. 22, p. 100241, 2021.

[29] C. Yang, H. Lan, F. Gao, and F. Gao, "Review of deep learning for photoacoustic imaging," *Photoacoustics*, vol. 21, p. 100215, 2021.

[30] L. Tian *et al.*, "Deep learning in biomedical optics," *Lasers in Surgery and Medicine*, 2021.

[31] S. Liu *et al.*, "Deep learning in medical ultrasound analysis: a review," *Engineering*, vol. 5, no. 2, pp. 261–275, 2019.

[32] A. C. Luchies and B. C. Byram, "Deep neural networks for ultrasound beamforming," *IEEE Transactions on Medical Imaging*, vol. 37, no. 9, pp. 2010–2021, 2018.

[33] Y. H. Yoon, S. Khan, J. Huh, and J. C. Ye, "Efficient B-mode ultrasound image reconstruction from sub-sampled rf data using deep learning," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 325–336, 2018.

[34] W. Simson *et al.*, "End-to-end learning-based ultrasound reconstruction," *arXiv preprint arXiv:1904.04696*, 2019.

[35] S. Khan, J. Huh, and J. C. Ye, "Adaptive and compressive beamforming using deep learning for medical ultrasound," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 67, no. 8, pp. 1558–1572, 2020.

[36] D. J. Gillies *et al.*, "Deep learning segmentation of general interventional tools in two-dimensional ultrasound images," *Medical Physics*, vol. 47, no. 10, pp. 4956–4970, 2020.

[37] H. Yang, C. Shan, A. F. Kolen, and P. H. de With, "Medical instrument detection in ultrasound-guided interventions: A review," *arXiv preprint arXiv:2007.04807*, 2020.

[38] C. Mwikirize, J. L. Nosher, and I. Hacihaliloglu, "Convolution neural networks for real-time needle detection and localization in 2D ultrasound," *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 5, pp. 647–657, 2018.

[39] C. Mwikirize, A. B. Kimbowa, S. Imanirakiza, A. Katumba, J. L. Nosher, and I. Hacihaliloglu, "Time-aware deep neural networks for needle tip localization in 2D ultrasound," *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–9, 2021.

[40] A. Pourtaherian *et al.*, "Robust and semantic needle detection in 3D ultrasound using orthogonal-plane convolutional neural networks," *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 9, pp. 1321–1333, 2018.

[41] Y. Zhang *et al.*, "Multi-needle localization with attention U-Net in us-guided HDR prostate brachytherapy," *Medical Physics*, vol. 47, no. 7, pp. 2735–2745, 2020.

[42] C. Andersén, T. Rydén, P. Thunberg, and J. H. Lagerlöf, "Deep learning-based digitization of prostate brachytherapy needles in ultrasound images," *Medical Physics*, 2020.

[43] A. Reiter and M. A. L. Bell, "A machine learning approach to identifying point source locations in photoacoustic data," in *Photons Plus Ultrasound: Imaging and Sensing 2017*, vol. 10064. International Society for Optics and Photonics, 2017, p. 100643J.

[44] D. Allman, A. Reiter, and M. A. L. Bell, "Photoacoustic source detection and reflection artifact removal enabled by deep learning," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1464–1477, 2018.

[45] K. Johnstonbaugh *et al.*, "A deep learning approach to photoacoustic wavefront localization in deep-tissue medium," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 2020.

[46] A. Yazdani, S. Agrawal, K. Johnstonbaugh, S.-R. Kothapalli, and V. Monga, "Simultaneous denoising and localization network for photoacoustic target localization," *IEEE Transactions on Medical Imaging*, 2021.

[47] K. P. Köstli, M. Frenz, H. Bebie, and H. P. Weber, "Temporal backward projection of optoacoustic pressure transients using Fourier transform methods," *Physics in Medicine & Biology*, vol. 46, no. 7, p. 1863, 2001.

[48] B. E. Treeby and B. T. Cox, "k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave fields," *Journal of Biomedical Optics*, vol. 15, no. 2, p. 021314, 2010.

[49] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 136–144.

[50] R. J. McGough, "Rapid calculations of time-harmonic nearfield pressures produced by rectangular pistons," *The Journal of the Acoustical Society of America*, vol. 115, no. 5, pp. 1934–1941, 2004.

[51] J. F. Kelly and R. J. McGough, "A time-space decomposition method for calculating the nearfield pressure generated by a pulsed circular piston," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 53, no. 6, pp. 1150–1159, 2006.

[52] B. Cox and P. Beard, "Fast calculation of pulsed photoacoustic fields in fluids using k-space methods," *The Journal of the Acoustical Society of America*, vol. 117, no. 6, pp. 3616–3627, 2005.

[53] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[54] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *3rd International Conference for Learning Representations (ICLR), 2015*, 2015.

[55] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[56] M. Benning and M. Burger, "Modern regularization methods for inverse problems," *Acta Numerica*, vol. 27, p. 1, 2018.

[57] M. K. A. Singh, V. Parameshwarappa, E. Hendriksen, W. Steenbergen, and S. Manohar, "Photoacoustic-guided focused ultrasound for accurate visualization of brachytherapy seeds with the photoacoustic needle," *Journal of Biomedical Optics*, vol. 21, no. 12, p. 120501, 2016.